

## Temporal and spatial estimation of groundwater electrical conductivity using soft computing approaches: Analysis of East Azerbaijan Province, Iran

Sarvin Zamanzad-Ghavidel<sup>a,\*</sup>, Reza Sobhani<sup>b,c</sup>, Sina Fazeli<sup>d</sup>, Leonardo Valerio Noto<sup>e</sup>, Carlo De Michele<sup>c</sup> and Dario Pumo<sup>e</sup>

<sup>a</sup> Urmia University-Iran's National Elites Foundation, Daneshvaran Omran-Ab (DOA) Consulting Company, Urmia, Iran

<sup>b</sup> Sari Agricultural Sciences and Natural Resources University, Sari, Iran

<sup>c</sup> Politecnico di Milano, Milano, Italy

<sup>d</sup> University of Tehran, Tehran, Iran

<sup>e</sup> University of Palermo, Palermo, Italy

\*Corresponding author. E-mail: snzghavidel@gmail.com

### ABSTRACT

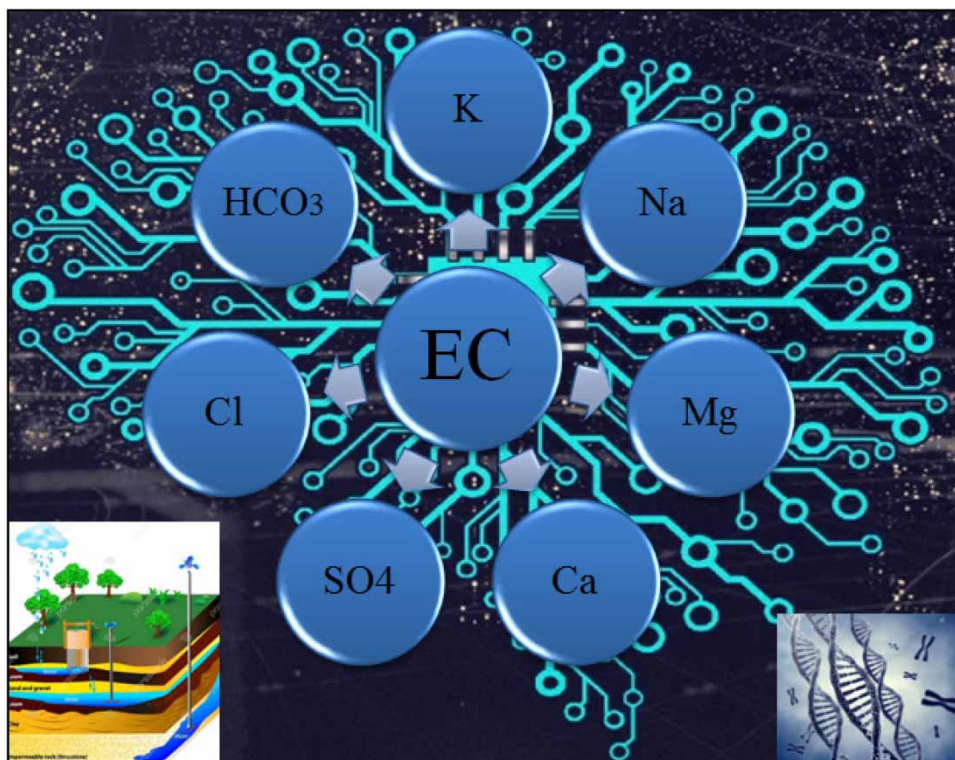
The evaluation of groundwater quality plays an important role in the management of groundwater. The main objectives of the present work are to develop a novel soft computing framework including Adaptive Neuro-Fuzzy Inference System (ANFIS), Wavelet-ANFIS (WANFIS), Gene Expression Programming (GEP), and Wavelet-GEP (WGEP) for the temporal and spatial estimation of groundwater electrical conductivity (EC) in the East Azerbaijan province, Iran over 2001–2020. The results demonstrate the importance of wavelet transform application; the performance percentage enhancement of the WANFIS and WGEP models compared to the ANFIS and GEP, using the RMSE criterion, ranged from 15.48 to 51.09% and from 5.06 to 86.95%, respectively. All the developed models showed the WGEP superior compared to others. The impact of land use characteristics, climatic conditions, and geological features on groundwater quality showed that there is a direct relationship between the extent of agricultural land, semi-arid climate conditions and groundwater EC amounts. The results demonstrated that the values of EC increase from east to west, indicating the direct exchange of surface and groundwater in the study area. Moreover, groundwater quality changes significantly across the width of the fault, with groundwater EC in the northern part of the fault higher than that in the southern part.

**Key words:** climate, geological structure, groundwater resources, land use feature, wavelet theory

### HIGHLIGHTS

- The EC variable of groundwater resources was estimated via single and hybrid-wavelet soft computing approaches.
- The impact of land use characteristics, climatic conditions, and geological features on groundwater quality was investigated.
- The data de-noising by wavelet approaches have the ability to improve the performance of EC estimation at spatial-temporal scales.

## GRAPHICAL ABSTRACT



## 1. INTRODUCTION

Water is a vital resource for every biological and human phenomenon. Climate change and urbanization are causing a substantial alteration of the hydrological cycle at a global level (Noto *et al.* 2023). For this reason, the management and protection of water are of capital importance not only in developing countries but also in developed countries (Yang *et al.* 2017; Poursaeid *et al.* 2021), and it is necessary to formulate new plans for the protection and sustainable use of water resources and incorporate inevitable solutions (Ketchemen-Tandia *et al.* 2017; Li *et al.* 2020). Zoning of qualitative variables of groundwater resources is very important for drinking consumes and agricultural purposes (Jafari *et al.* 2019; Pannell & Rogers 2022; Noto *et al.* 2023).

The groundwater electrical conductivity (EC) as a water quality variable depends on the amount of dissolved substances in the water, often referred to as total dissolved solids (TDS). Groundwater is naturally contaminated and its level of contamination can vary depending on various factors such as climatic conditions, geologic structure, land use characteristics. Contamination in groundwater is caused, for instance, by rock weathering, soil washing, and dissolution of aerosol particles, and it is also particularly affected by human activities such as mining in industry, use of chemical fertilizers in agriculture, and urbanization (Akoachere *et al.* 2019; Lu *et al.* 2019; Komba *et al.* 2020).

Badeenezhad *et al.* (2021) assessed the effects of land use and climatic variability on non-carcinogenic health risks from nitrate contamination of groundwater. Their outcomes showed that changes in land use, especially urban and residential development, significantly affect nitrate values in groundwater. In addition, rising temperatures and lower annual precipitation may also increase the magnitude of non-carcinogenic health risks. Sadeghi *et al.* (2021) evaluated the change in groundwater quality, considering  $\text{HCO}_3$ ,  $\text{SO}_4$ , SAR, and EC, in response to land use changes in the Zrebar Lake basin (Iran). The results showed that, despite increased agricultural land use, water quality remained suitable for consumption as drinking water and in agriculture. The quality of groundwater is strongly influenced by geological formations and saline rivers (Aftab *et al.* 2018; Honarbakhsh *et al.* 2019). Barzegar *et al.* (2016) assessed the hydrogeochemistry and water quality of the Aji-Chay River in Iran. The main processes controlling the water quality of the Aji-Chay River are salinization and rock weathering. Maghrebi *et al.* (2021) found a functional relationship between the geological-geomorphological features and the

conditions of groundwater quality in Iran. Chitsazan & Manshadi (2021) studied the role of the Mehriz fault on hydrochemistry and groundwater flow in the Yazd aquifer, Iran. Their outcomes reflected that the Mehriz fault appears locally as a channel through which poor quality freshwater is discharged into the area that provides drinking water for the city of Yazd.

In recent years, soft computing approaches such as Artificial Neural Network (ANN), Adaptive Neuro-Fuzzy Inference System (ANFIS), Gene Expression Programming (GEP), Support Vector Machine (SVM), and Wavelet Theory (WT) are increasingly used in water and environmental sciences (Hrnjica & Danandeh Mehr 2018; Montaseri *et al.* 2018; Motevalli *et al.* 2019; Nearing *et al.* 2021; Pumo & Noto 2023). Some past studies have indicated the practicality of using soft computing approaches to estimate groundwater quality variables. Nordin *et al.* (2021), for instance, reviewed the four most commonly used soft computing methods, including ANN, ANFIS, Evolutionary Algorithm (EA), and SVM for modeling groundwater quality. Esmaeilbeiki *et al.* (2020) evaluated the spatial variability of EC and total hardness (TH) using a Dynamic Evolving Neural-Fuzzy Inference System (DENFIS), Group Method of Data Handling (GMDH), Multivariate Adaptive Regression Spline (MARS), M5 Tree model (M5 Tree), and GEP approaches in the Chhattisgarh state, India. The results indicated that the zoning maps of the data-driven approaches have reasonable accuracy compared to the observed values. Al-Adhaileh *et al.* (2022) expanded Single Exponential Smoothing (SES) with Bidirectional Long Short-Term Memory (BiLSTM) and an ANFIS to analyze groundwater quality in the Al-Baha region of Saudi Arabia with an arid climate, demonstrating high performance of both SES-BiLSTM and SES-ANFIS approaches in groundwater quality estimation. Scientific literature also reports several examples of soft computing applications to groundwater quality analyses in Iran (Khashei-Siuki & Sarbazi 2015; Aghajari *et al.* 2019; Khalaj *et al.* 2019; Ghobadi *et al.* 2022; Jalalkamali & Jalalkamali 2022).

The present work proposes an innovative framework for EC estimation based on the wavelet approach and is able to eliminate uncertainties and noises that frequently affect groundwater quality data. More specifically, four soft computing approaches, i.e. ANFIS, GEP, Wavelet-ANFIS (WANFIS) and Wavelet-GEP (WGEP), for the temporal and spatial estimation of EC in the East Azerbaijan province (Iran) are developed and compared, exploiting a large data set of the Iran Water Resources Management Company (IWRMC) including sampled values of the major cations (i.e., sodium  $\text{Na}^+$ , calcium  $\text{Ca}^{+2}$ , potassium  $\text{K}^+$ , and magnesium  $\text{Mg}^{+2}$ ) and anions (i.e., chloride  $\text{Cl}^-$ , sulfate  $\text{SO}_4^{-2}$ , and bicarbonate ( $\text{HCO}_3^-$ )) in water and simultaneous measures of EC.

The East Azerbaijan province is one of the most important economic centers in the fields of industry, tourism, and agriculture in Iran. Due to its large extension, this province has a very diverse climate. In recent years, East Azerbaijan province has been exposed to climatic changes and also heavy pollution due to high population density, industrial development, and urbanization (Andaryani *et al.* 2019; Sadeqi & Dinpashoh 2020; Ren *et al.* 2021; Ghazi & Jeihouni 2022; Kadkhodazadeh & Farzin 2022). Using the wavelet tool only by focusing on improving the performance of models to estimate different variables cannot be practical. The innovative feature of the models developed in this research is the identification and elimination of uncertainties and noises, resulting from natural and human factors, by the wavelet tool. Also, previous research has examined a limited number of factors affecting the quality of groundwater resources in different regions, which can create gaps in the areas such as East Azerbaijan province, where the quality of groundwater resources is affected by various factors. Therefore, the present research, by summarizing the majority of factors of surface water salinity, land use characteristics, climatic conditions, geological features, and faults and also examining their impact on groundwater resources' EC in 14 study areas in East Azerbaijan province, makes it possible to compare and analyze the effects of each factor on EC values of groundwater resources. The analysis of the temporal and spatial variability of groundwater EC arising from the proposed framework, allowed for inferring and identifying possible sources of pollution focusing on the factors of surface water salinity, land use characteristics, climatic conditions, geological features, and faults. Overall, current research follows three main objectives: (1) de-noising and removal of uncertainty in groundwater quality variables by WT, (2) development of temporal and spatial estimation of EC amount via four soft computing approaches, i.e. ANFIS, GEP, WANFIS, and WGEP in the East Azerbaijan province (Iran), and (3) evaluation of surface water salinity, land use characteristics, climatic conditions, geological features, and faults on groundwater EC amounts.

## 2. MATERIALS AND METHODS

### 2.1. Study area and used data

The East Azerbaijan province is located in northern Iran with a special geographical location and high population. Its extension is 45,491 km<sup>2</sup>, which is about 2.8% of the total area of the country. Due to topographical diversity, the province is

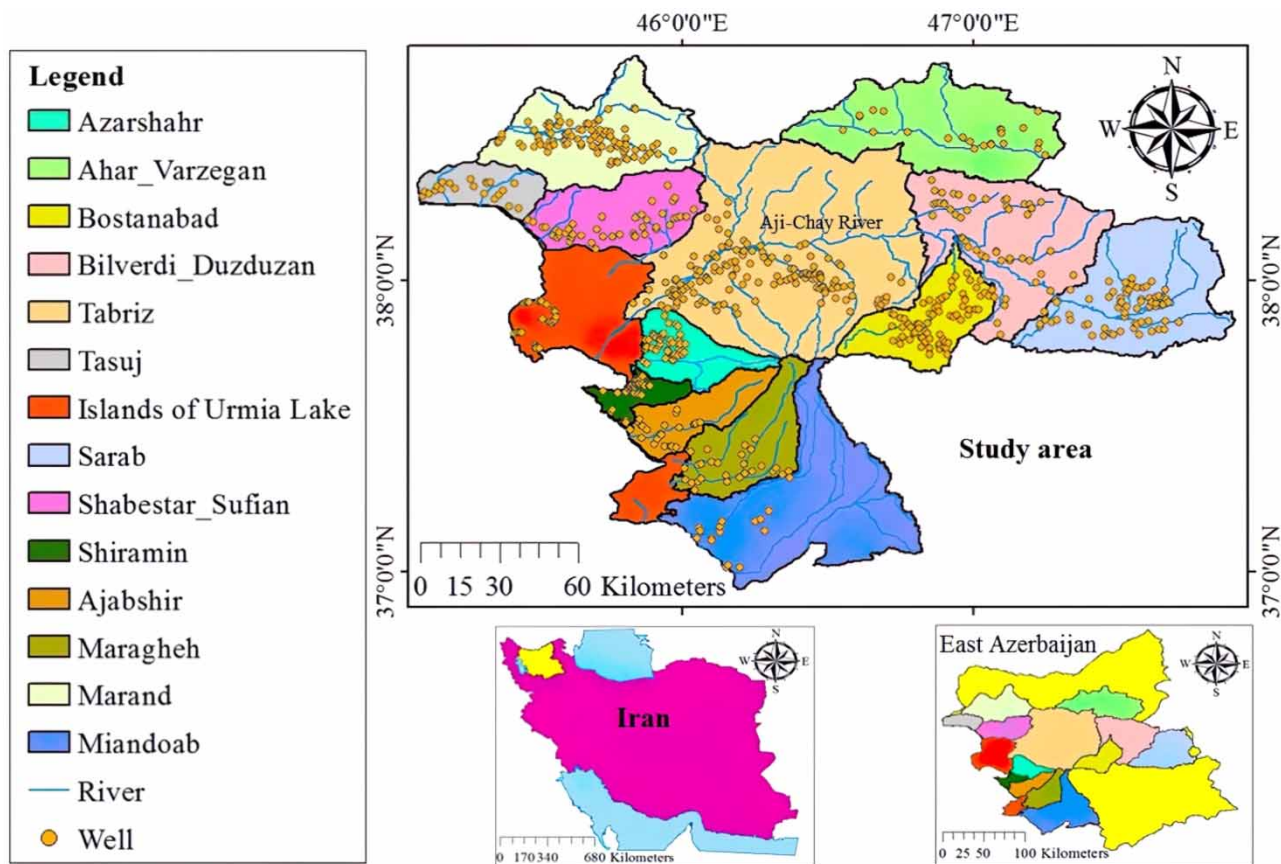
characterized by a significant spatial variability of its climatic conditions, with an average annual rainfall of 250–300 mm. The Aji-Chai River, crossing the Tabriz plain, is one of the most important rivers of the province.

Fourteen study areas, located within the East Azerbaijan province (Figure 1), were selected and considered for this study. The studied area of the province is located in the catchment area of Lake Urmia. Also, all the selected wells are located in the alluvial fan aquifer and in line with standard operation. In addition, the dominant geological structure of the study area is dark grey shale-sand stone and limestone-marl.

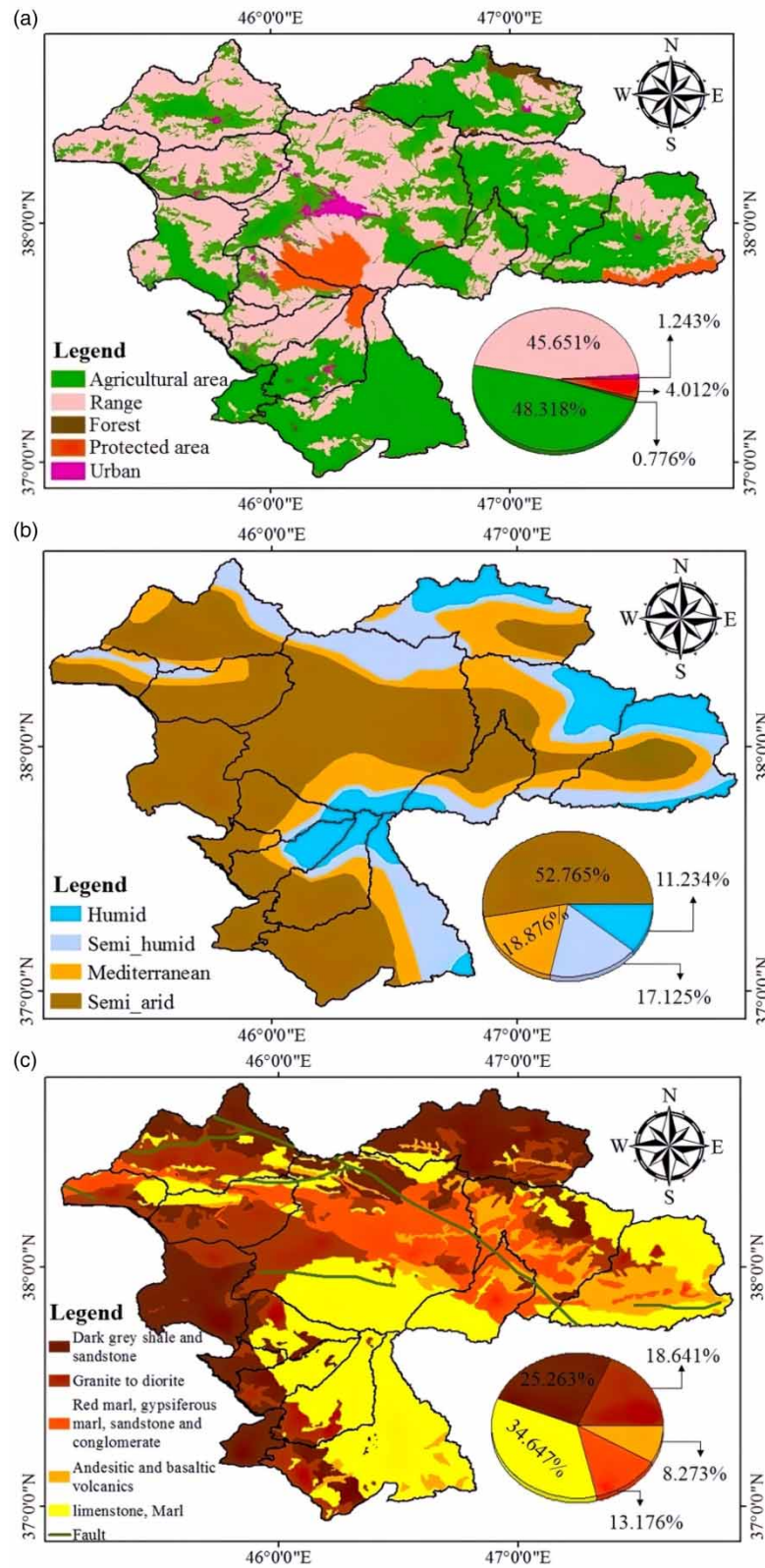
Figure 2 shows the characteristics of land use, climatic conditions, geological structure, and faults in the study areas that are developed in GIS based on the Statistical Centre of Iran's information for the 2020 year. In the studied area, the agricultural lands have the largest area with the value of 48.32% (Figure 2(a)). Figure 2(b) shows the climatic classification of the studied area characterized by semi-arid conditions (BSk) with values of 52.77% based on the Köppen Geiger Climate Classification (<https://www.wrm.ir/>). Figure 2(c) shows the geological structure of the study area; one of the most important faults of Iran, named 'Tabriz', is located in the study area, with a length of more than 100 km. BSk is a climate that its mean annual temperature is less than 18 °C and is too dry to support a forest, but not dry enough to be a desert, usually consisting of grassland plains.

In the current research, K, Na, Mg, Ca, SO<sub>4</sub>, Cl, and HCO<sub>3</sub> variables over 2001–2020 were applied for EC estimation of groundwater quality that was provided by IWRMC. The pre-process and analysis of data reliability was managed by IWRMC. The number of wells and samples for each study area are presented in Table 1. The measurement of groundwater quality variables is followed by the Institute of Standards and Industrial Research of Iran (<https://www.inso.gov.ir/>).

The statistical characteristics of the used data are listed in Table 1, which reports the mean and coefficient variation (Cv) values of each monitored variable over all the wells in each study area. It can be noticed how the average EC ranges from 768 (Sarab) µmhos/cm to 8,373 µmhos/cm (Shiramin) while the CV of EC varies from 0.408 (Miandoab) to 1.174 (Shabestar-Sufian), showing a high variability of the statistics relative to the EC of groundwater across the province.



**Figure 1** | Geographical characteristics of the study area.



**Figure 2** | Characteristics of land use (a), climatic conditions (b), and geological structure-location of faults (c) in the studied area.

**Table 1** | Statistical characteristics of the selected variables for the different regions of the study area

Region ( $i = 1, \dots, 12$ )	Variables (unit)	K (meq/L)	Na (meq/L)	Mg (meq/L)	Ca (meq/L)	SO <sub>4</sub> (meq/L)	Cl (meq/L)	HCO <sub>3</sub> (meq/L)	EC ( $\mu$ mhos/cm)	The number of wells	The number of samples
Azarshahr ( $i = 1$ )	Mean	0.235	7.534	6.237	12.230	5.026	15.136	6.014	2,593.128	50	771
	Cv	0.541	0.957	1.014	0.982	0.813	1.451	0.566	0.909		
Ahar-Varzeghan ( $i = 2$ )	Mean	0.174	6.375	3.854	7.316	6.794	5.191	5.734	1,775.537	23	378
	Cv	2.462	0.933	0.793	0.709	0.901	1.478	0.370	0.749		
Bostanabad ( $i = 3$ )	Mean	0.082	3.793	2.280	4.926	4.301	2.701	4.039	1,108.672	76	1,410
	Cv	0.729	0.773	0.811	0.658	0.912	1.283	0.405	0.673		
Bilverdi-Duzduzan ( $i = 4$ )	Mean	0.092	6.727	4.225	5.786	5.906	6.090	4.790	1,678.780	50	1,012
	Cv	0.854	1.539	0.685	0.706	0.966	1.722	0.230	0.923		
Tabriz ( $i = 5$ )	Mean	0.209	11.535	4.665	6.415	4.248	14.058	4.412	2,279.379	142	2,130
	Cv	0.668	1.390	0.980	1.188	0.967	1.507	0.522	1.088		
Tasuj ( $i = 6$ )	Mean	0.176	6.547	4.783	4.897	3.122	7.059	6.163	1,635.823	21	519
	Cv	0.715	0.923	0.654	0.583	0.635	1.087	0.867	0.582		
Islands of Urmia Lake ( $i = 7$ )	Mean	1.183	18.860	13.142	18.866	3.012	46.144	2.814	5,152.397	25	287
	Cv	0.951	0.883	0.888	0.967	1.106	0.862	0.458	0.807		
Sarab ( $i = 8$ )	Mean	0.089	2.101	1.633	3.831	1.941	1.665	3.994	767.711	66	1,350
	Cv	1.114	1.148	0.960	0.742	1.397	1.899	0.378	0.811		
Shabestar-Sufian ( $i = 9$ )	Mean	0.086	7.525	4.741	4.460	1.824	11.527	3.369	1,682.796	44	825
	Cv	1.144	1.511	1.176	1.146	0.751	1.615	0.360	1.174		
Shiramin ( $i = 10$ )	Mean	0.159	31.686	21.480	36.720	7.963	72.503	9.566	8,373.188	41	394
	Cv	0.747	0.882	0.849	0.896	0.751	0.905	0.449	0.757		
Ajabshir ( $i = 11$ )	Mean	0.070	3.809	3.032	4.744	1.633	3.188	6.816	1,155.738	25	455
	Cv	1.492	1.494	0.792	0.459	1.300	1.745	0.430	0.737		
Maragheh ( $i = 12$ )	Mean	0.135	6.094	3.176	5.243	3.011	5.532	6.103	1,448.921	26	458
	Cv	0.719	2.331	0.910	0.522	1.066	2.751	0.341	1.132		
Marand ( $i = 13$ )	Mean	0.158	8.123	4.920	4.336	2.844	9.045	5.603	1,759.206	90	1,865
	Cv	0.758	0.800	0.709	0.772	0.853	1.094	0.507	0.677		
Miandoab ( $i = 14$ )	Mean	0.150	2.897	3.445	4.880	2.113	2.052	7.198	1,139.332	17	365
	Cv	1.090	0.977	0.376	0.336	0.758	1.228	0.277	0.408		

## 2.2. Wavelet-Db4

The quantities that can be measured in time or space are called signals. In signal analysis, mathematical converters are used to obtain information that is not readily available from raw signals. As one of the most efficient mathematical converters, the wavelet transforms signals into a series of basis and partial functions at the scale of the mother wavelet. The major advantage of the wavelet transform is that it effectively extracts time and frequency information from a time-varying signal (Rhif *et al.* 2019; Emadi *et al.* 2021). Starting from the so-called mother wavelet  $\Psi(t)$ , the wavelet function of  $\Psi(t)$  is defined by the following mathematical form (Equation (1)):

$$\int_{-\infty}^{+\infty} \Psi(t) dt = 0 \quad (1)$$

The basic function  $\Psi_{(a,b)}(t)$  can be obtained from the following relationship by using the delay and rescaling of the mother wavelet (Equation (2)).

$$\Psi_{(a,b)}(t) = |a|^{-0.5} \Psi\left(\frac{t-b}{a}\right), \quad a \in R, b \in R, a \neq 0 \quad (2)$$

where  $a$  is the scale parameter and  $b$  determines the location of the wavelet.

In the case of Continuous Wavelet Transform (CWT), given the signal  $f(t)$ , the wavelet coefficients at any point of the signal ( $b$ ) and for any value of the scale ( $a$ ) can be calculated with the following relationship (Equation (3)):

$$CWT(a,b) = Wf(a,b) = \int_{-\infty}^{+\infty} f(t) \Psi_{a,b} dt = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \Psi\left(\frac{t-b}{a}\right) dt \quad (3)$$

Another form of WT called Discrete Wavelet Transform, abbreviated as DWT, is also used in signal analysis. In DWT, the transmission and scale parameters are chosen discontinuously, so that (Equation (4)):

$$a = 2^{-j}, \quad b = 2^{-j}k \quad (4)$$

where  $j$  and  $k$  are integers. As a result, by replacing  $a$  and  $b$ , the following relationship is obtained (Equation (5)).

$$\Psi_{(j,k)}(t) = 2^{\frac{j}{2}} \Psi(2^j t - k) \quad (5)$$

In recent years, many wavelet functions have been developed, each of which has unique characteristics and has been used in a specific field. The wavelet tool analyzes the data in two main ( $A$ ) and detailed ( $D$ ) sets. In the current research, based on the shape of the variables time series the mother wavelet of Daubechies-4 (db-4) is applied for analyzing the groundwater quality variables at level 1 by coding in MATLAB environment (Emadi *et al.* 2021).

The data selected as input variables were normalized by Equation (6):

$$X_N = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (6)$$

where  $X_N$ ,  $X$ ,  $X_{\min}$ , and  $X_{\max}$  denoted normalized observational data, observed data, minimal and maximal of the used data-base, respectively.

## 2.3. ANFIS and WANFIS

The ANFIS has the ability to combine the two methods of ANN and fuzzy logic system (FIS). This method (Nordin *et al.* 2021) uses learning algorithms of neural networks and fuzzy logic to design nonlinear mappings between the input space and the output space and also has the special ability to combine the qualitative performance of a FIS with the numerical

performance of an ANN in modeling processes. In the ANFIS approach, Takagi Sugeno’s (TS) inference method is used considering Back-Propagation (BP) approach. In the current research, the ANFIS with a type of Subtractive Clustering (SC) model is trained using an optimization algorithm, which changes the radius values in different iterations in order to optimize the value of a selected fitness function, usually measuring model error (Khashei-Siuki & Sarbazi 2015).

After determining the optimal radius values, fuzzy rules should be determined. Then, the Root Mean Square Error (RMSE) is used to determine the result of the output membership functions (Emadi *et al.* 2021). The fuzzy inference process is developed in five layers: (1) Fuzzification of input variables, (2) Using operators (AND; OR) in the initialization section, (3) Deducing the results from the previous steps, (4) Combining the results of fuzzy rules, and (5) De-fuzzing. In this study, if we assume that a fuzzy inference system has two inputs of Ca and Na and an output of EC (Figure 3), TS first-order fuzzy model with two if-then fuzzy rules are expressed as follows:

Rule 1: If Ca is in a LOW state and Na is in a HIGH state, then (Equation (7)):

$$EC = p_1Ca + q_1Na + r_1 \tag{7}$$

Rule 2: If Ca is in a LOW state and Na is in a MEDIUM state, then (Equation (8)):

$$EC = p_2Ca + q_2Na + r_2 \tag{8}$$

where LOW, HIGH, and MEDIUM are the MFs for inputs Ca and Na. Also,  $p_1, p_2, q_1, q_2, r_1,$  and  $r_2$  are the parameters of the output function. In the present work, the ANFIS and WANFIS models were developed in the MATLAB environment (Emadi *et al.* 2021). In the ANFIS models, the observed data were used to develop the EC estimation model, while in the WANFIS model, the data that were analyzed by WT (A-series) were applied for EC estimation models.

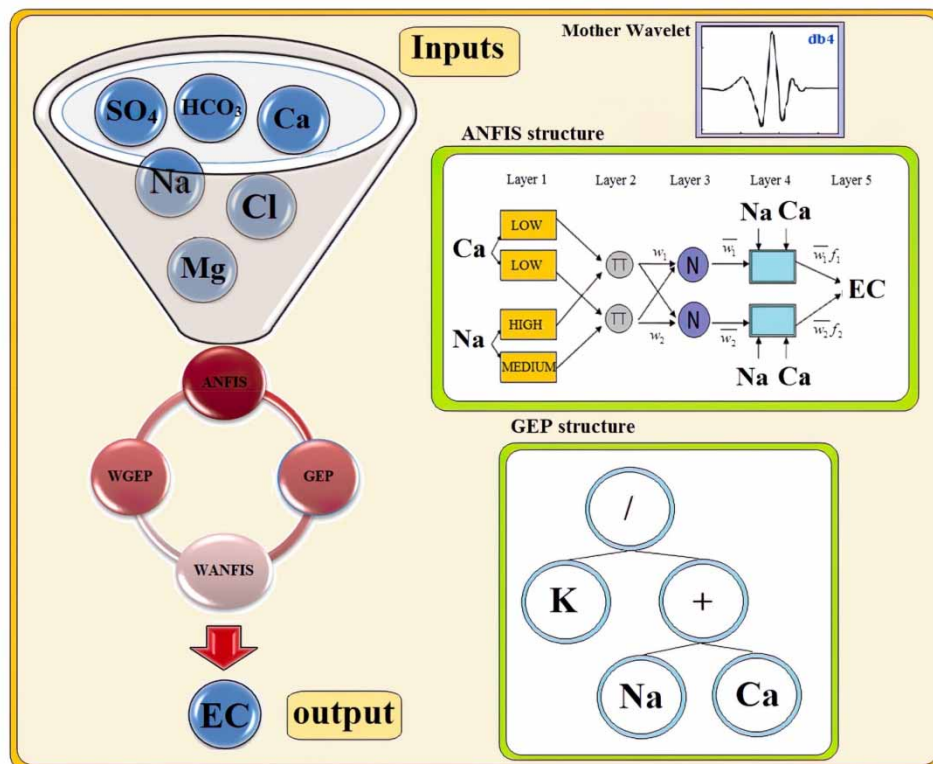


Figure 3 | The structure of the used approaches.



## 2.4. GEP and WGEP

GEP, which originated in the evolution of intelligent models, is considered one of the circular algorithm methods based on Darwin's theory of evolution (Esmailbeiki *et al.* 2020). GEP uses the population of individuals (chromosomes), improves according to a prefixed fitness function using an iterative search method, and derives the best solution through one or more genetic operators. The step-by-step process to solve a problem using GEP consists of the following five steps (Ferreira 2006; Montaseri *et al.* 2018): (1) Selection of the terminal set: it is the independent variable of the system. This step involves the selection of the fitting function, which usually uses the RMSE; (2) Selection of a set of functions: which includes a selection of different mathematical operators; (3) index to measure the model accuracy: these indices can be used to determine the ability of the model to solve a particular problem; (4) Control components: values of numerical components and qualitative variables used to control the execution of the model. The number of training and testing data, the number of chromosomes, the size of the head, the number of genes, the selection of the linkage operator, which can be set with the four options of addition, subtraction, multiplication and division; (5) Criteria for stopping the program: these are criteria to evaluate convergence and/or stopping the execution of the model, such as reaching prefixed limits for the number of population produced, the computational time, the maximum compliance, and so on. In the current study, GeneXpro Tool 4.0 software (Emadi *et al.* 2021) was used to develop the EC estimation models. In the GEP models, the observed data were used to develop the EC estimation model, while in the WGEP model, the data that were analyzed by WT (A-series) were applied for EC estimation models. The structure of the used approaches is presented in Figure 3.

In the next step, two single approaches of ANFIS and GEP were implemented to estimate EC in the 14 selected regions. To remove the uncertainty in the qualitative groundwater data, all variables were slowed down to A and D series using the Wavelet-Db4 approach and, after deleting the noise (D series), used to estimate the EC variables that formed the WANFIS and WGEP approaches. The K, Na, Mg, Ca, SO<sub>4</sub>, Cl, and HCO<sub>3</sub> variables with their normalization status were considered as input variables for EC estimation. Also, the ratio of training (70%) and testing (30%) of the data set was selected based on trial and error for all applied models. The applied data are aggregated at the level of study areas (14 study areas) then the EC estimation models were developed via ANFIS, GEP, WANFIS, and WGEP approaches. In the last step, the performance of developed models was assessed by statistics evaluation indicator that is discussed in the next section.

## 2.5. Statistical indicators

Here, four indicators were employed to assess the performance of the implemented approaches, namely the Correlation Coefficient ( $R$ ), the RMSE, the Mean Absolute Error (MAE), and the Nash–Sutcliffe Efficiency (NSE) (Equations (9)–(12)).

$$R = \frac{\sum_{i=1}^N (EC_{io} - \overline{EC_o})(EC_{ie} - \overline{EC_e})}{\sqrt{\sum_{i=1}^N (EC_{io} - \overline{EC_o})^2 \sum_{i=1}^N (EC_{ie} - \overline{EC_e})^2}} \quad (9)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (EC_{io} - EC_{ie})^2} \quad (10)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |EC_{io} - EC_{ie}| \quad (11)$$

$$NSE = 1 - \frac{(EC_{ie} - EC_{io})^2}{(EC_{io} - \overline{EC_o})^2} \quad (12)$$

where  $EC_{io}$  and  $EC_{ie}$  are the observed and estimated EC values for the  $i$ th sample;  $N$  is the total number of samples; and  $\overline{EC_o}$ ,  $\overline{EC_e}$  are the average of observed and estimated EC values.

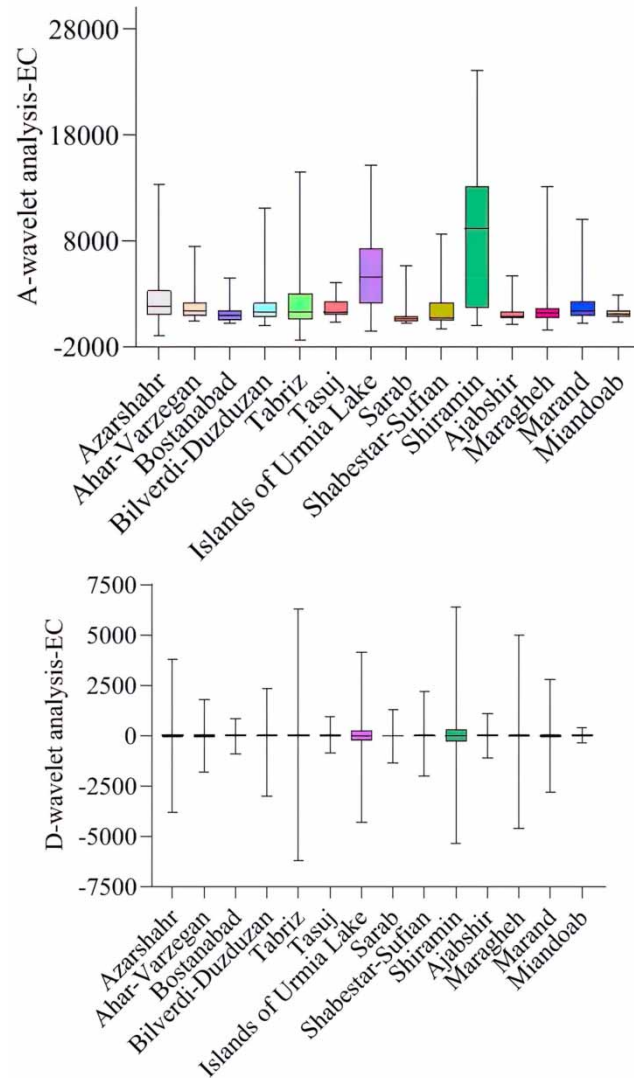
The correlation coefficient between two variables indicates the ability to predict the value of a series in relation to the other. The  $R$ -values vary between 0 and 1, with 1 denoting perfect agreement between observation and estimation points. RMSE and MAE are metrics calculated based on error measurements, where the lowest values denote best model performances. NSE is

a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance; it ranges from minus infinite to one, with values close to one suggesting a model with high predictive ability.

### 3. RESULTS AND DISCUSSION

#### 3.1. Wavelet decomposition

The groundwater quality data were decomposed in two data sets of the main (A) and detailed (D) data using the wavelet tool at level 1. The noise in the data that leads to uncertainties was removed from the modeling by removing the (D) data sets. The wavelet analysis of EC values for the A and D subseries is shown in Figure 4. The outcomes of the A and D wavelet analysis for the variables used are listed in Supplementary material, Tables S1 and S2, respectively. The A values ranged between  $-1,352.404$  (Tabriz) and  $24,125.842$  (Shiramin) for the EC variable. The number of detailed sets depends on the level of decomposition that can be varied based on various factors such as the number of samples, the type of data, and the purpose of analysis. Since in this research the main purpose of using the wavelet tool is to identify the noises of the uncertainties of natural and human factors, therefore, decomposition at level 1 can be appropriate.



**Figure 4** | Wavelet analysis of EC values for A and D subseries.

**3.2. ANFIS and GEP approaches**

ANFIS and GEP approaches, as single models, were preliminarily considered for estimating the EC of groundwater resources in the different study areas. The radius values for ANFIS models were determined via trial-error technique.

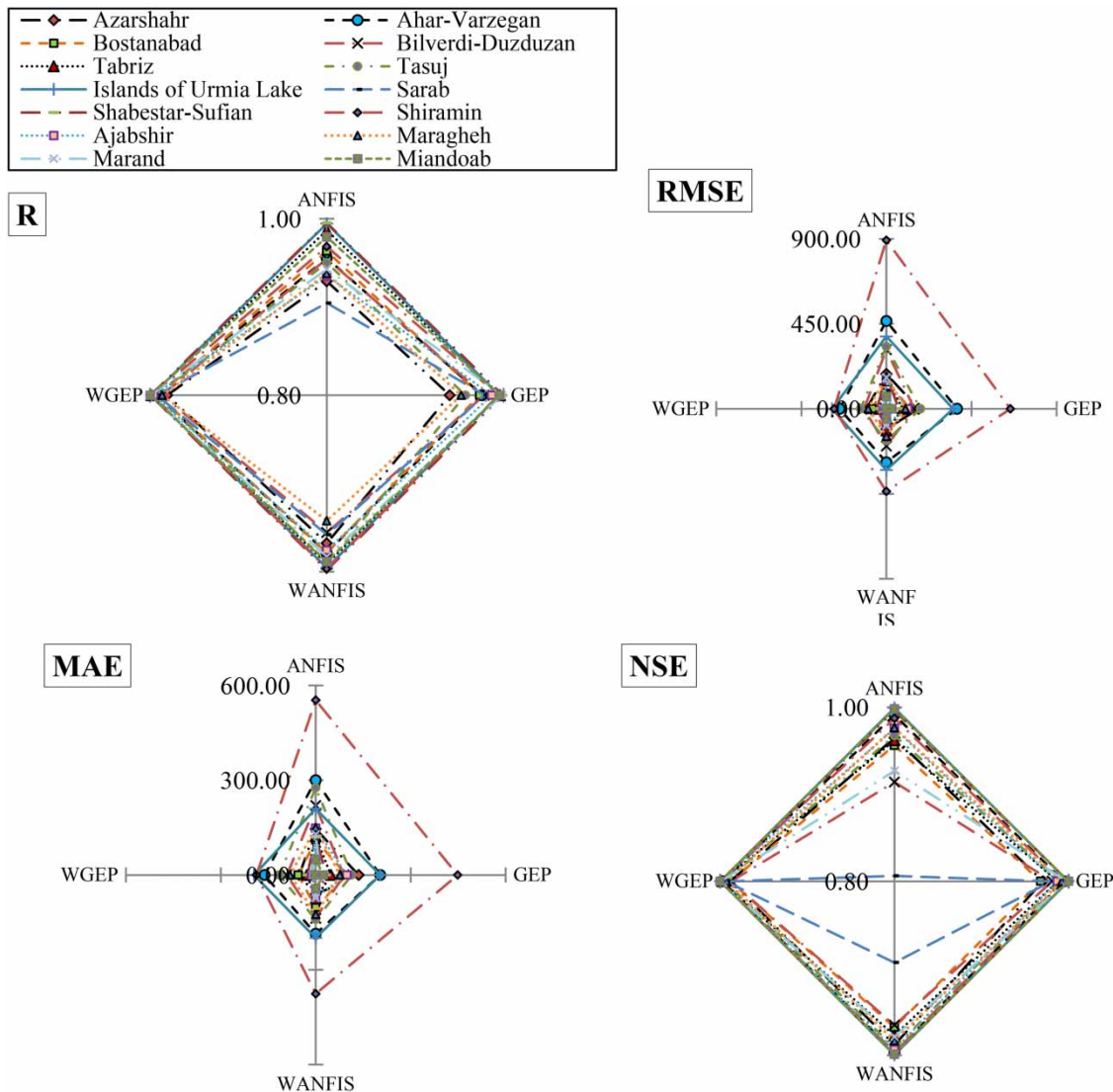
The optimal ANFIS and GEP approach' characteristics and associated performances at each study area are listed in Table 2 and depicted in Figure 5. The optimal models in the ANFIS method with SC type depend on the size of the radius values, which varies between 0.1 and 0.9 (Emadi et al. 2021). The obtained optimal radius values of the ANFIS approach ranged from 0.26 (Tabriz) to 0.51 (Bilverdi-Duzduzan and Tasuj), with performances of 151.965 (Tabriz), 313.523, and 329.199 (Bilverdi-Duzduzan and Tasuj)  $\mu\text{mhos/cm}$  for RMSE values. The general characteristics of the linking function and fitness function error type were addition and RRSE, respectively. Also, the values of mutation rate, inversion rate, IS transposition rate, RIS transposition rate, one point recombination rate, two-point recombination rate, gene recombination rate, and gene transposition rate were 0.044, 0.1, 0.1, 0.1, 0.3, 0.3, 0.1, and 0.1, respectively. The used mathematical operators in the present work were +, -,  $\times$ ,  $\div$ ,  $\sqrt[3]{\phantom{x}}$ ,  $\sqrt{\phantom{x}}$ ,  $\ln$ ,  $e^x$ ,  $x^2$ ,  $x^3$ ,  $\sin x$ ,  $\cos x$ , and  $\arctan x$  for expanding the EC estimation models. The values of R and RMSE for the GEP models ranged from 0.939 (Azarshahr) to 0.999 (Shabestar-Sufian), and from 28.277 (Miandoab) to 655.781 (Shabestar-Sufian)  $\mu\text{mhos/cm}$ , respectively.

**3.3. Wavelet-combined (WANFIS and WGEP) approaches**

The main purpose of developing wavelet-combined models is to eliminate the influence of human factors in measuring data samples and compound events and floods of natural factors on the uncertainty of groundwater quality variables. The performance indicators of the optimal WANFIS and WGEP approaches are listed in Table 2. The best approaches for each study area

**Table 2** | Results of optimal models and their specifications for estimating EC values for the different regions (given in the first column) of the study area

<i>i</i>	Approaches	R	RMSE ( $\mu\text{mhos/cm}$ )	MAE ( $\mu\text{mhos/cm}$ )	NSE	<i>i</i>	Approaches	R	RMSE ( $\mu\text{mhos/cm}$ )	MAE ( $\mu\text{mhos/cm}$ )	NSE
1	ANFIS (0.35)	0.929	184.956	145.922	0.965	8	ANFIS (0.34)	0.905	98.057	89.178	0.806
	GEP	0.939	170.139	135.417	0.970		GEP	0.979	31.380	22.730	0.980
	WANFIS (0.42)	0.968	107.117	61.307	0.988		WANFIS (0.41)	0.958	71.248	47.579	0.893
	WGEP	0.981	105.679	52.228	0.988		WGEP	0.992	17.521	13.440	0.994
2	ANFIS (0.40)	0.955	465.271	300.162	0.991	9	ANFIS (0.38)	0.995	180.474	120.595	0.996
	GEP	0.976	373.670	203.258	0.994		GEP	0.999	64.718	33.243	1.000
	WANFIS (0.31)	0.979	282.773	186.106	0.997		WANFIS (0.58)	0.998	97.123	69.431	0.999
	WGEP	0.993	247.814	162.193	0.997		WGEP	1.000	50.674	26.782	1.000
3	ANFIS (0.35)	0.963	147.417	110.780	0.956	10	ANFIS (0.35)	0.968	893.146	553.473	0.987
	GEP	0.973	126.463	97.884	0.968		GEP	0.994	655.781	448.044	0.993
	WANFIS (0.36)	0.978	123.236	102.651	0.968		WANFIS (0.36)	0.997	436.793	375.814	0.997
	WGEP	0.991	74.603	55.559	0.988		WGEP	0.997	276.510	187.765	0.999
4	ANFIS (0.51)	0.955	313.523	219.712	0.914	11	ANFIS (0.34)	0.935	154.473	147.918	0.976
	GEP	0.978	147.934	88.954	0.981		GEP	0.987	114.015	98.437	0.987
	WANFIS (0.48)	0.956	197.726	137.493	0.965		WANFIS (0.39)	0.976	79.832	69.216	0.993
	WGEP	0.986	118.752	91.441	0.988		WGEP	0.996	14.879	12.416	1.000
5	ANFIS (0.26)	0.987	151.965	122.367	0.962	12	ANFIS (0.37)	0.938	172.132	149.993	0.976
	GEP	0.996	88.204	47.534	0.987		GEP	0.952	100.895	76.017	0.992
	WANFIS (0.34)	0.991	114.231	75.186	0.977		WANFIS (0.47)	0.942	145.394	124.591	0.983
	WGEP	0.999	36.937	28.115	0.998		WGEP	0.986	95.793	80.985	0.993
6	ANFIS (0.51)	0.950	329.199	275.783	0.967	13	ANFIS (0.31)	0.943	165.426	129.001	0.927
	GEP	0.957	180.391	114.197	0.990		GEP	0.998	28.588	13.491	0.998
	WANFIS (0.42)	0.991	168.140	140.673	0.991		WANFIS (0.42)	0.983	88.113	72.489	0.979
	WGEP	0.991	136.526	115.582	0.994		WGEP	0.999	18.047	8.585	0.999
7	ANFIS (0.33)	0.995	383.616	209.837	0.999	14	ANFIS (0.38)	0.979	65.672	49.467	0.998
	GEP	0.995	353.012	203.454	0.999		GEP	0.995	28.277	23.843	1.000
	WANFIS (0.47)	0.996	324.237	198.816	0.999		WANFIS (0.46)	0.989	53.085	43.100	0.999
	WGEP	0.998	287.147	195.096	1.000		WGEP	1.000	6.731	3.983	1.000



**Figure 5** | Comparison of evaluation statistics of applied approaches for estimating EC variables.

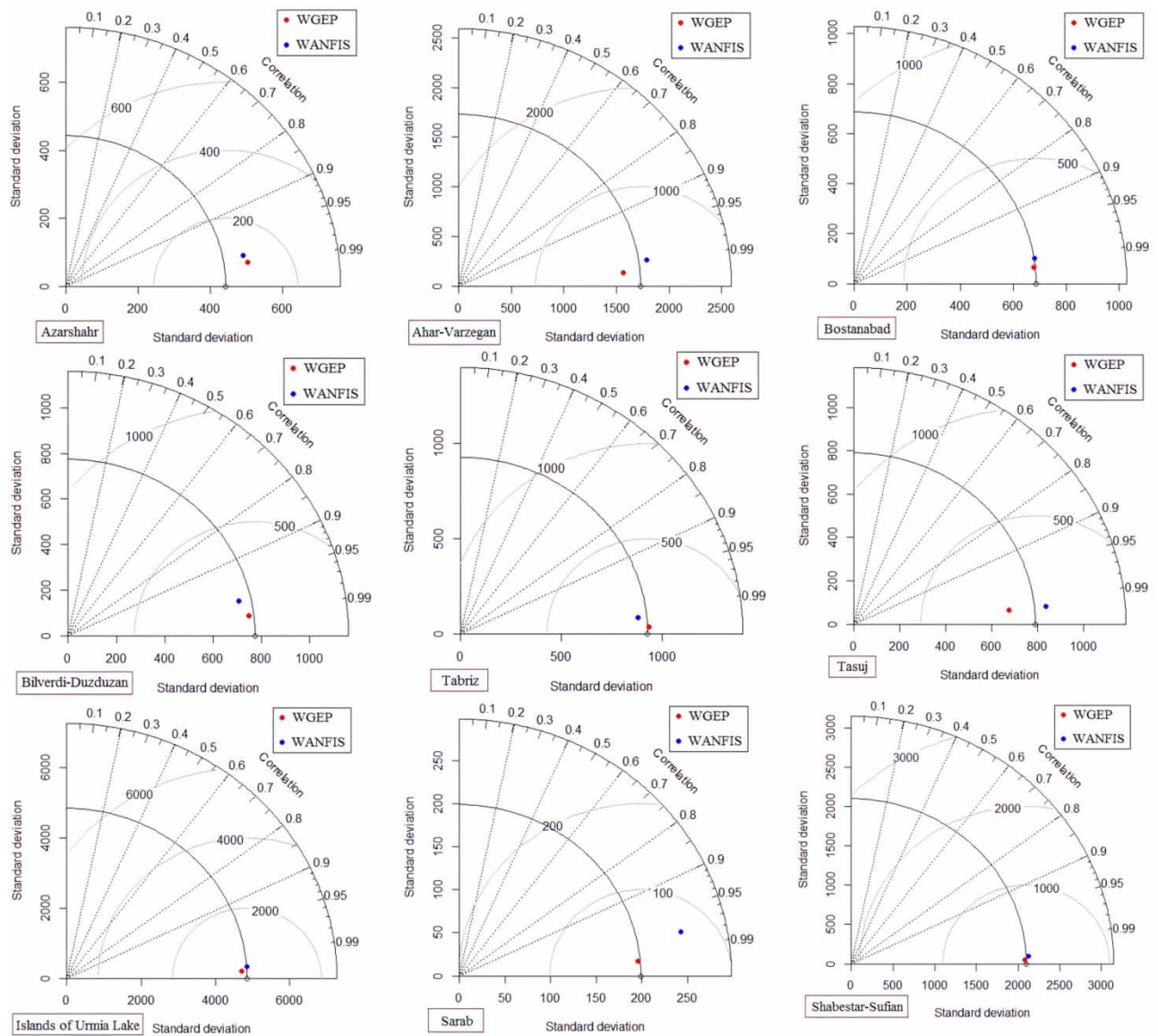
were selected based on *R*, RMSE, MAE, and NSE values for estimating the EC values of groundwater resources in the East Azerbaijan province.

The optimal radius values for WGANFIS models ranged from 0.31 (Ahar-Varzeghan) to 0.58 (Shabestar-Sufian). The values of *R* and RMSE corresponding to the WGANFIS models ranged from 0.942 to 0.998 and from 53.085 to 436.793  $\mu\text{mhos/cm}$ , respectively. The values of *R* and RMSE for WGEP models ranged from 0.981 to 1 and from 6.731 to 287.147  $\mu\text{mhos/cm}$ , respectively.

Figure 5 shows the comparison among all the single and wavelet-combined models developed at each studied area in terms of the statistical indicators *R*, RMSE, MAE, and NSE. Although the *R*-values for the single and wavelet-combined soft computing approaches are close to 1, the wavelet-combined approaches had higher performance than the singular.

Taylor diagram is one of the most essential graphical forms of comparing the statistical results of models, considering *R*, RMSE, and standard deviation. Taylor diagrams of WGANFIS and WGEP approaches for estimating the EC variable are shown in Figure 6. The results showed that the red point (WGEP models' representative) is closer to the observed point than the blue one (WGANFIS models' representative) in the all reigns.

The WGEP has the highest performance among other applied approaches for EC estimation. The scatter plots, confidence and prediction bands (95%), and histograms for the WGEP approaches in 14 study areas are demonstrated in



**Figure 6** | Taylor diagram for WGFIS and WGEF approaches. (continued).

**Figure 7.** The results indicated that more than 95% of estimated EC values are placed in the confidence and prediction bands for all reigns.

**Figure 8** presents the Violin diagram for applied approaches in the study areas. The violin diagram displays the estimated data distributions in various ranges for minimum, first quartile, median, third quartile, and maximum. The results indicated that the difference between the minimum and maximum estimated values is the lowest and highest in the Ajabshir and Shiramin reigns, respectively.

The percent improvement in the performance of the WGFIS and WGEF approaches compared with the ANFIS and GEP approaches according to the RMSE criterion (Table 3) was calculated to be an average of 51.09% (Shiramin) and 86.95% (Ajabshir) as the highest.

Emadi *et al.* (2021) estimated the EC variable of deep wells, semi-deep wells, and aqueducts in the various catchments of Iran using ANN, ANFIS, GEP, and WGEF approaches. The results show that the combination of the wavelet approach with GEP can increase the performance of the model compared to its single form (GEP), which ranges from 17 to 35%, 13 to 32%,

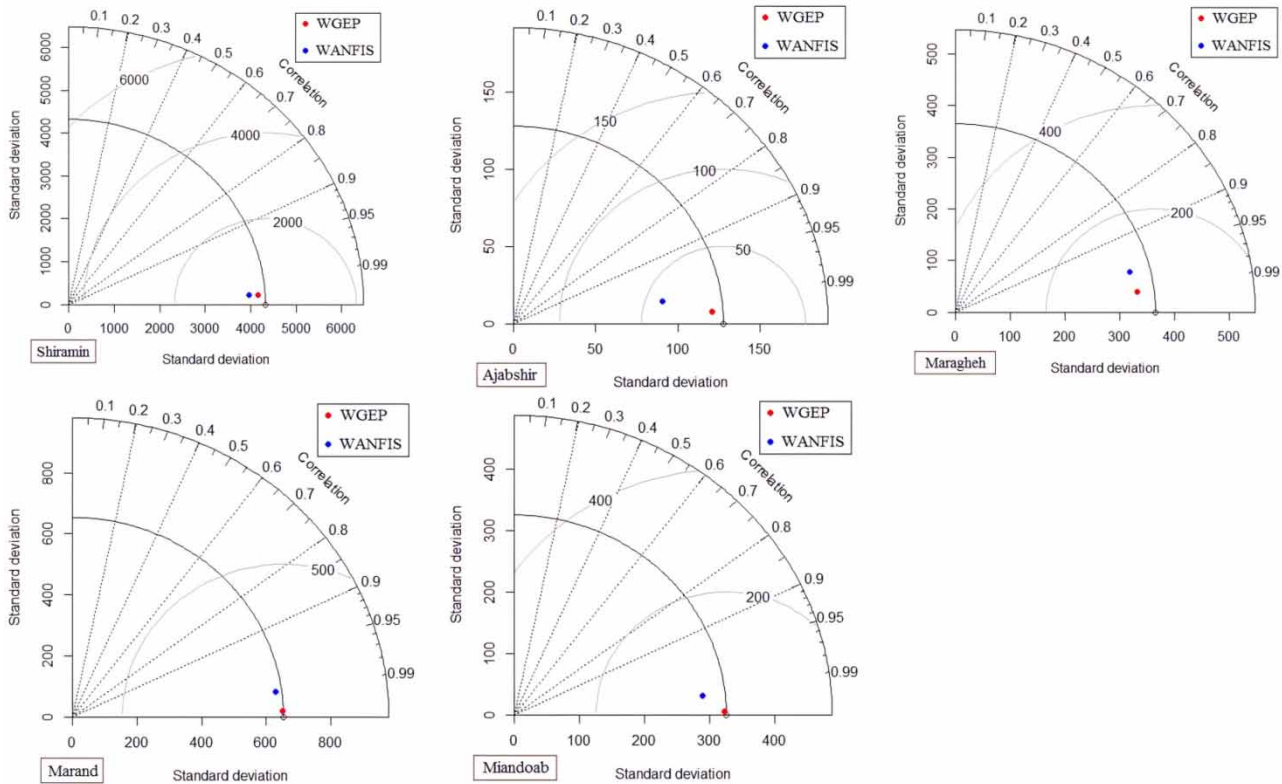
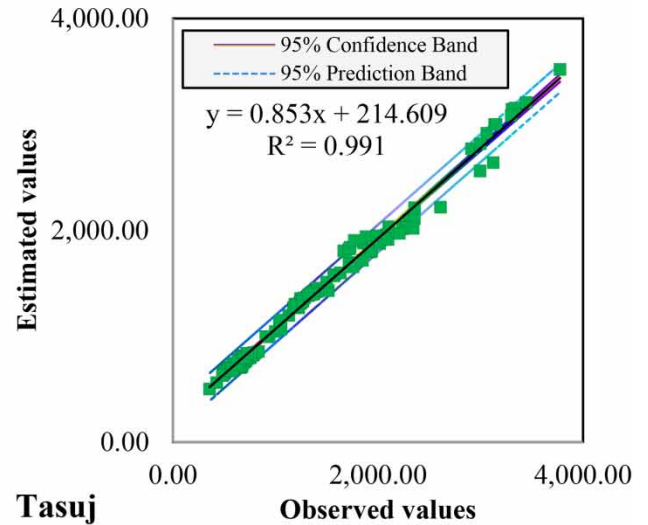
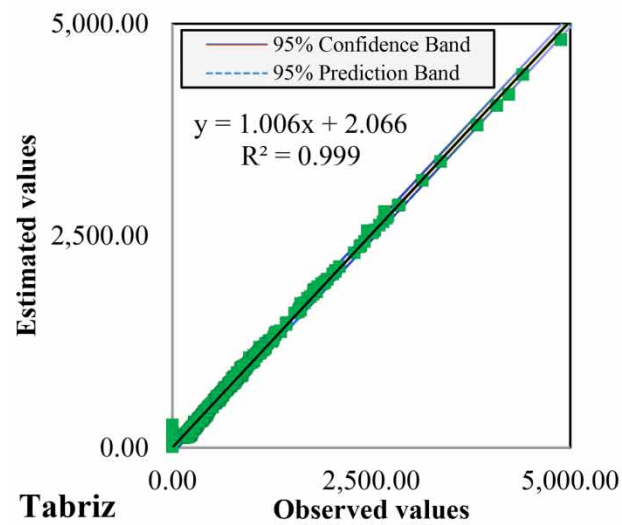
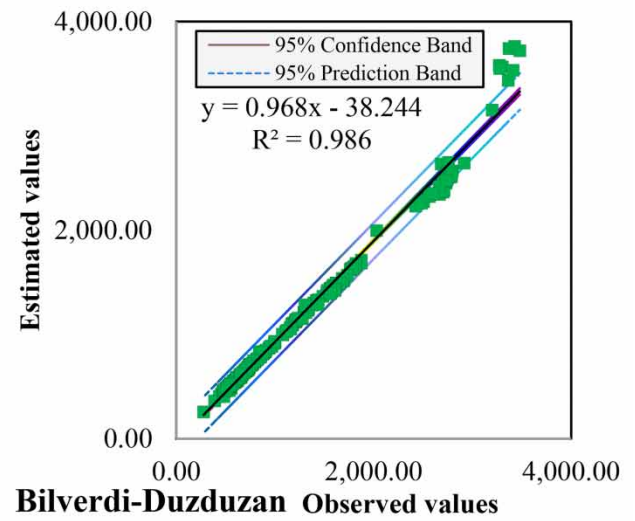
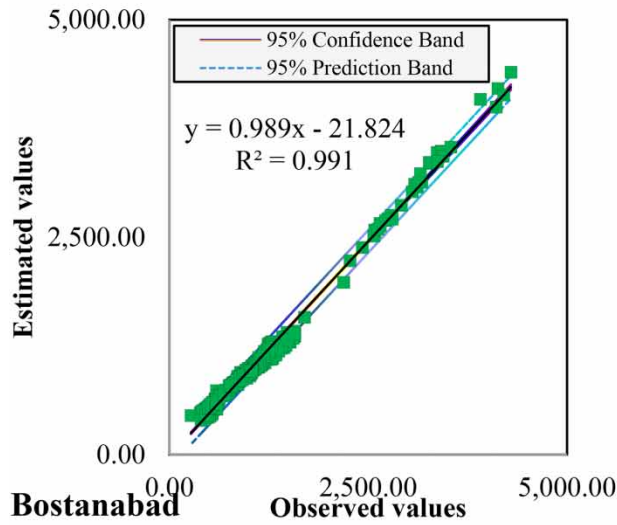
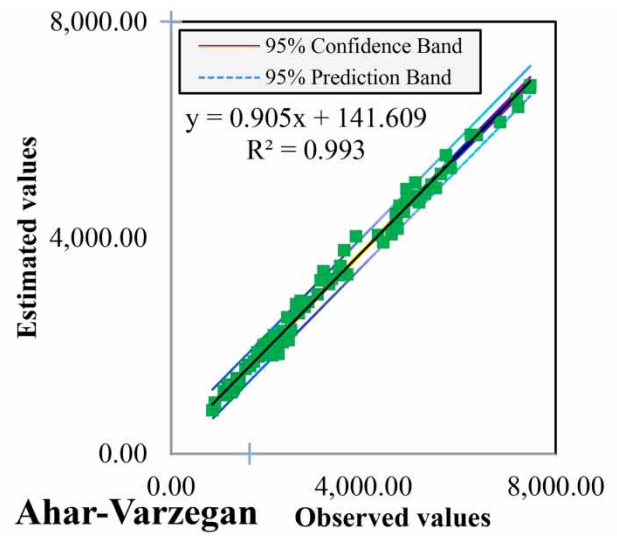
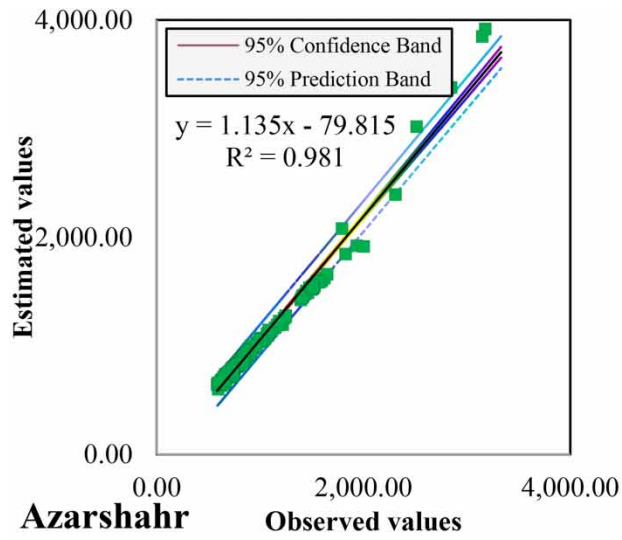


Figure 6 | Continued.

17 to 46% for deep wells, semi-deep wells, and aqueducts, respectively, in the study areas. Therefore, their results are in line with our results regarding the application of soft computing approaches for groundwater EC estimation. Also, their results confirm the better performance of Wavelet-combined approaches compared to their singles with various frameworks.

The performance of the optimal model (WGEP) at different intervals including 30%<sub>Min</sub>, 40%<sub>Mid</sub>, and 30%<sub>Max</sub> of EC values, based on the *R*-values is shown in Table 4. The minimum and maximum *R*-values of 0.613 and 1 are related to the intervals 30%<sub>Min</sub> and 30%<sub>Max</sub> for Shiramin, Shabestar-Sufian, and Miandoab, respectively. The results indicated the best performance of the WGEP model for estimating the extreme maximum EC values in the Shabestar-Sufian and Miandoab regions. The results indicated that due to de-noising and removal of uncertainty in the groundwater quality variables via WT, combined-wavelet models (WANFIS, WGEP) have provided better results than their singular form (ANFIS, GEP). In addition, GEP and WGEP models can automatically select the input variables that have the greatest effect on EC estimation. The GEP and WGEP methods are random search methods and it is less possible to lock the solution of the parameters in the local optimal points. Also, these models have a better performance than ANFIS and WANFIS due to their chromosomal structure, ability to generate generation, and de-noising of the used data. One of the most important advantages of the GEP and WGEP methods over other soft computing methods is to provide the mathematical relationship that governs the problem. The extracted equations from WGEP approaches for the EC estimation of groundwater resources are presented in Supplementary material, Table S3. As an example, the K, Na, Mg, Ca, SO<sub>4</sub>, Cl, and HCO<sub>3</sub> variables were considered as input variables for EC estimation, and the best evolved WGEP model identified only K, SO<sub>4</sub>, Cl, and HCO<sub>3</sub> variables as the effective ones to estimate EC values in the Azarshahr region. Therefore, for future analysis, we will be able to estimate the EC with the mentioned variables in the Azarshahr study area. As a result, developing soft computing approaches could provide practical and efficient tools for quickly estimating the EC variable of groundwater with higher accuracy than traditional approaches. EC is a property that simultaneously takes into account the concentration of different ions in water; in this regard, EC is perfectly suitable as a global water quality variable that can be used as a control parameter in decision support systems, opportunely implemented to help water management by regulatory and control bodies of water



**Figure 7** | Scatter plot, 95% confidence band, 95% prediction band, and histogram for WGEF approaches. (continued.).

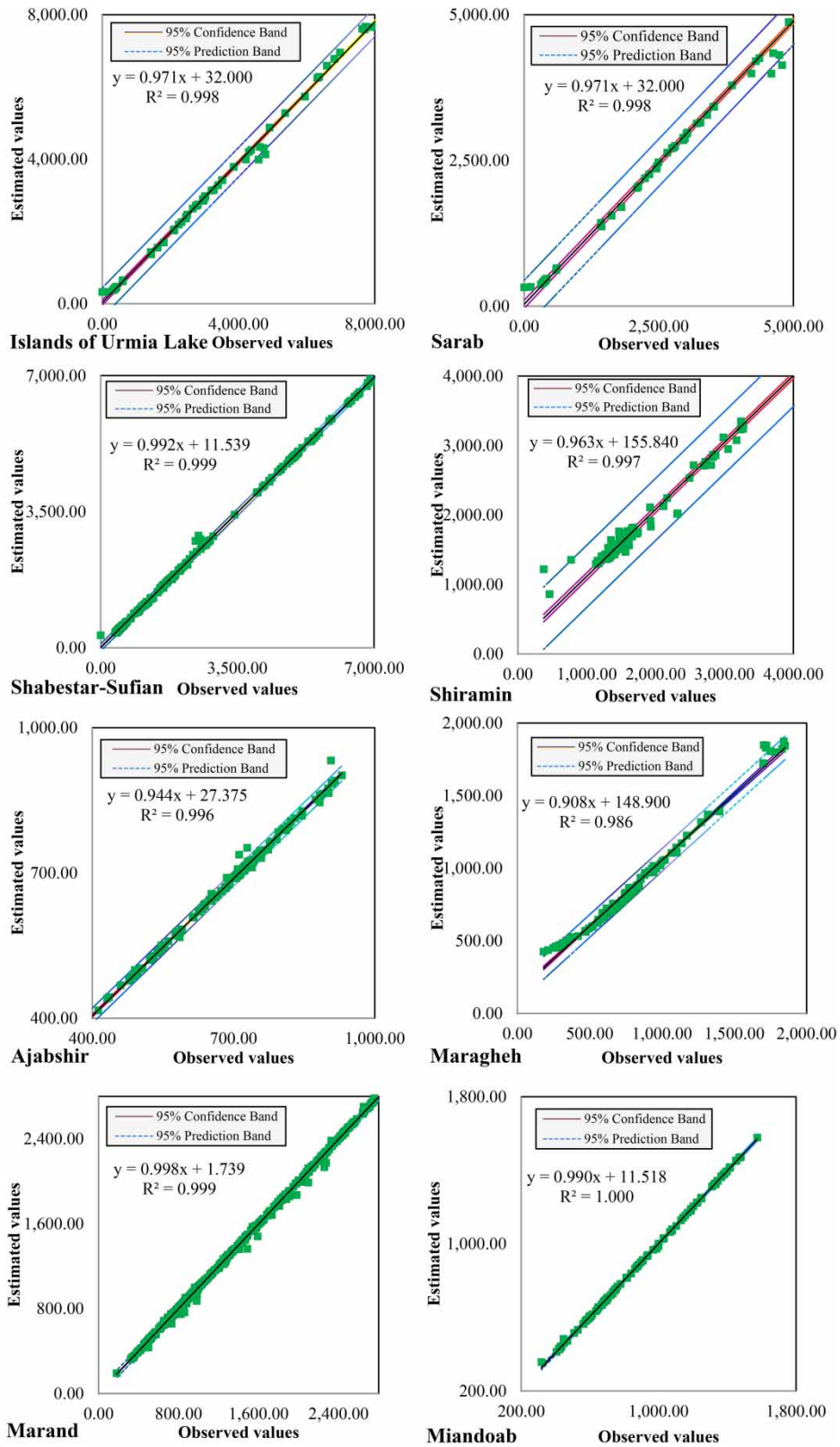
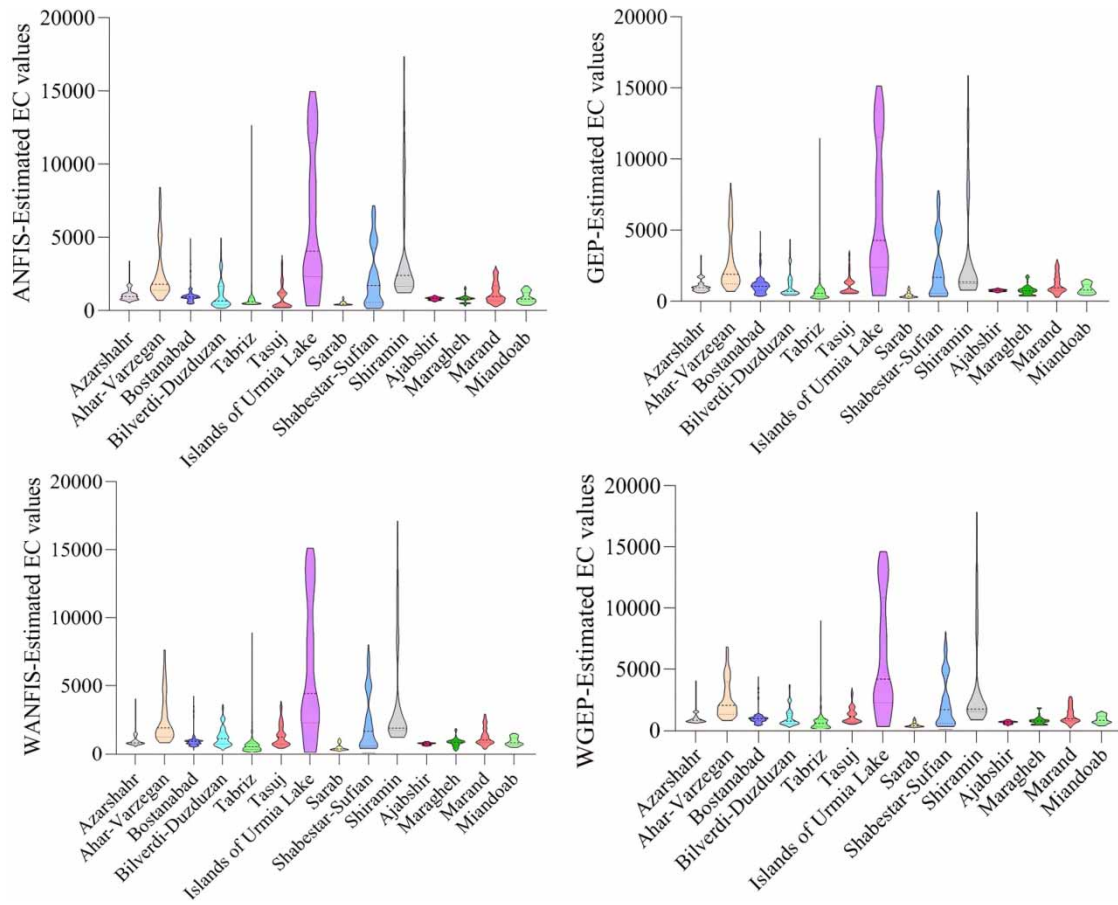


Figure 7 | Continued.





**Figure 8** | Violin diagram for applied approaches in different regions of the study area.

**Table 3** | The percent improvement in the performance of the WANFIS and WGEP approaches compared with the ANFIS and GEP approaches according to the RMSE criterion

<i>i</i>	WANFIS/ANFIS (%)	WGEP/GEP (%)
1	42.09	37.89
2	39.22	33.68
3	16.40	41.01
4	36.93	19.73
5	24.83	58.12
6	48.92	24.32
7	15.48	18.66
8	27.34	44.16
9	46.18	21.70
10	51.09	57.83
11	48.32	86.95
12	15.53	5.06
13	46.74	36.87
14	19.17	76.20

**Table 4** | Performance of the optimal model (WGEP) in various intervals based on *R*-values for the different regions of the study area

Study areas	Azarshahr	Ahar-Varzegan	Bostanabad	Bilverdi-Duzduzan	Tabriz	Tasuj	Islands of Urmia Lake
30% <sub>Min</sub>	0.760	0.873	0.998	0.999	0.783	0.823	0.999
40% <sub>Mid</sub>	0.953	0.927	0.975	0.999	0.985	0.989	0.990
30% <sub>Max</sub>	0.980	0.972	0.926	0.981	0.999	0.960	0.973
Study areas	Sarab	Shabestar-Sufian	Shiramin	Ajabshir	Maragheh	Marand	Miandoab
30% <sub>Min</sub>	0.999	0.816	0.613	0.992	0.972	0.995	0.994
40% <sub>Mid</sub>	0.992	0.990	0.954	0.936	0.947	0.994	0.999
30% <sub>Max</sub>	0.973	1.000	0.991	0.975	0.993	0.996	1.000

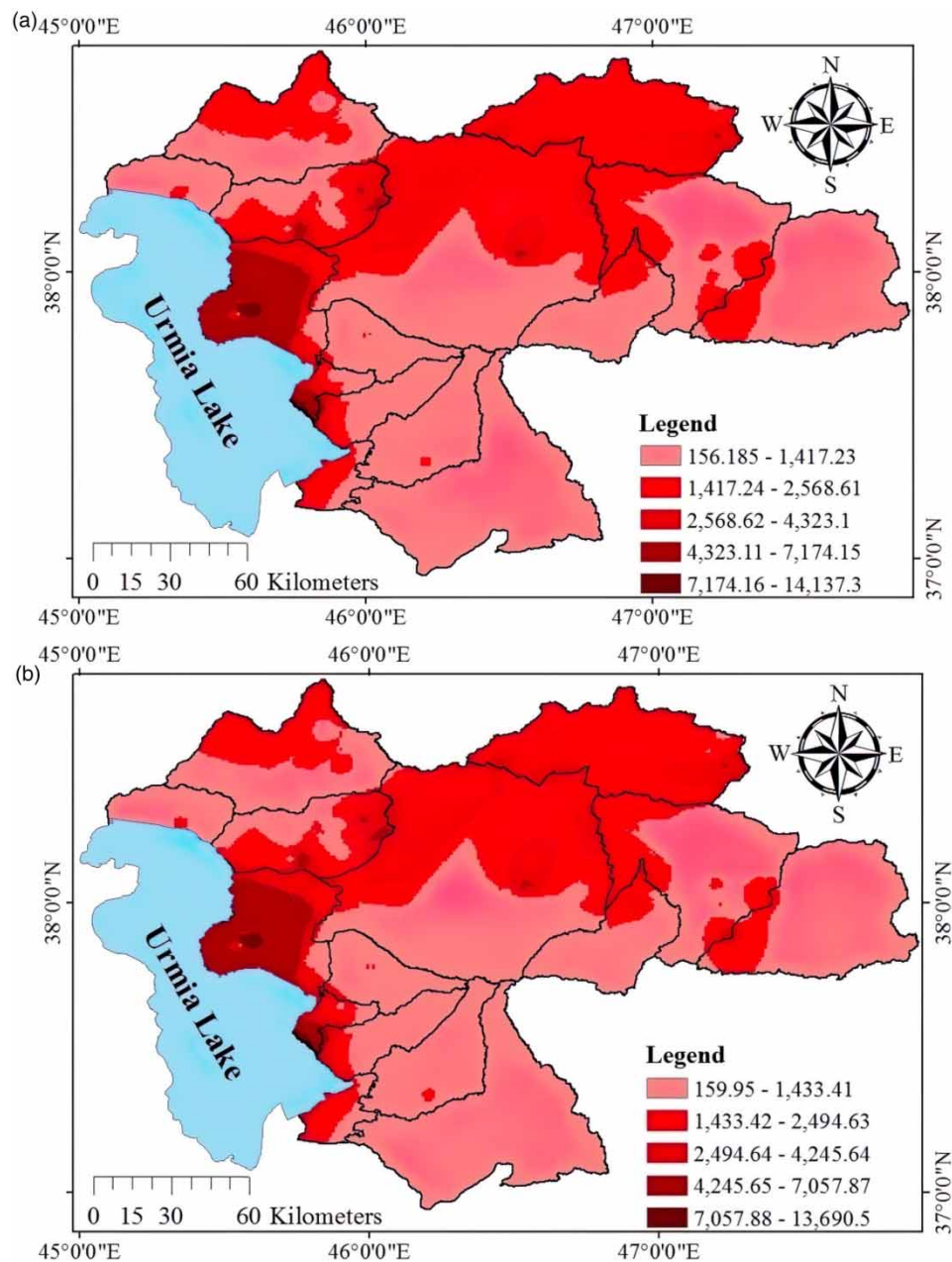
resources. Also, the results showed that by considering the derived equation for Azarshahr as a benchmark, the EC values calculated in other regions indicated high efficiency and the correlation coefficient was higher than 0.90 in all regions.

A comparison of the spatial variations of EC values measured by IWRMC and estimated based on the optimal WGEP approaches is shown in Figure 9. The zoning map of EC values is provided in a GIS environment (Honarbakhsh *et al.* 2019) using Inverse Distance Weighting (IDW) method for interpolation. The results show that the WGEP approach has high performance in determining the spatial distribution of the values of EC.

Actually monitoring EC is usually simpler and less expensive than monitoring K, Na, Mg, Ca, SO<sub>4</sub>, Cl, and HCO<sub>3</sub>. For this reason, physical properties such as EC, as well as TDS and pH, are frequently used as proxy variables to assess the concentration of cations and anions in groundwater. As it is rather known, in fact, EC is highly affected by major cations (i.e., sodium Na<sup>+</sup>, calcium Ca<sup>+2</sup>, potassium K<sup>+</sup>, and magnesium Mg<sup>+2</sup>) and anions (i.e., chloride Cl<sup>-</sup>, sulfate SO<sub>4</sub><sup>-2</sup>, and bicarbonate (HCO<sub>3</sub><sup>-</sup>). The relationship between physical properties and ions' concentration is strongly not linear and deserves deep investigation as it emerges from recent literature. In this regard, the availability of a wide historical database, such as that here analyzed, represents a precious resource that allows us to explore such a relationship, in order to find out how the EC variable is dependent on the different anions and cations, singularly or in a combined manner. Moreover, the use of soft computing techniques to explore the existing pattern between ion concentrations and EC allows for removing some disturbance effects, such as potential noise that inevitably affects this kind of sampling data. The results of our application, for instance, showed that some anions and cations are scarcely and not significantly correlated to EC values. This could also suggest which ion concentration deserves monitoring for water quality assessment and which one could be discarded, optimizing the costs of monitoring. The results of Ekemen Keskin *et al.* (2020) and Utomo *et al.* (2021) are in line with the achieved results in this case. Ekemen Keskin *et al.* (2020) have applied Na, K, Ca, Mg, HCO<sub>3</sub> and CO<sub>3</sub>, Cl, SO<sub>4</sub>, Fe, Mn, Al, and NO<sub>3</sub> variables for EC estimation via ANN and MLR approaches. Utomo *et al.* (2021) investigated the performance of soft computing algorithms, namely ANN, Gaussian processes (GP), and MLR to physical variables of groundwater such as TDS, pH, and EC estimation. The soft computing approaches used seven hydrochemical variables of K, Ca, Mg, Na, SO<sub>4</sub>, Cl, and HCO<sub>3</sub> as inputs to estimate TDS, pH, and EC amounts.

### 3.4. Impact of land use characteristics, climatic conditions, and geological features on EC values

The results show that the EC of groundwater increases from the north-northeast of the studied areas to the west (Figure 9), indicating a possible similar trend also for salinity concentration. In particular, the EC level of groundwater is rather high in the Ahar-Varzegan, the western part of Bilverdi-Duzduzan and Sarab, northeast of Bostanabad study areas, as well as the northern part of the Tabriz study area. The study of land use conditions (Figure 2(a)) shows that in these areas there is a direct relationship between the extent of agricultural land and the high values of EC. In the study areas of Sarab, Ahar-Varzegan, Marand and the western part of Tabriz study area, treated wastewater is used to irrigate fields, which could explain the increase in the salinity of groundwater, and thus in EC, for these areas, as demonstrated by some past studies (El Ayni *et al.* 2011; Shtull-Trauring *et al.* 2020). In the northwestern and central parts of the province, urban-industrial use could have played an important role in raising the EC level of groundwater through the release of industrial materials, urban solid wastes, and the spread of pollution. One of the characteristics of domestic wastewater is in fact the high salt concentrations. In addition, the organic matter in wastewater contains organic sulfur, which is transformed into



**Figure 9** | Comparison of spatial variations of EC values for observed (a) and estimated (b) values based on the optimal WGEP approaches.

sulfate by oxidation. Figure 9 shows high EC of groundwater across the Ajabshir, Urmia Lake Islands, Shiramin, and Shabestar-Sufian study areas, which could be due to their location on the coast of Urmia Lake, the sixth largest salt lake in the world, where the salinity of the lake water and its progress in the groundwater could have played an important role. The groundwater of the Azarshahr region in the southeastern part, the feeding area of the plain, has low salt concentrations, while the water salinity increases significantly toward the northwest-west of the plain. Due to the great depth of the aquifer, the absence of excessive exploitation, the small number of wells and also the feeding of the water table by the alluvial bed of the river, the EC of groundwater in the Miandoab study area is lower compared to other areas.

On the contrary, in most of the plains of East Azerbaijan province, including Marand, Maragheh Bonab, and Ajabshir, a decrease in the groundwater level was observed. The main reasons for this are the decrease in the amount of precipitation, the lack of adequate feeding of aquifers, and the large withdrawal of water from wells due to the lack of surface water resources.

Therefore, the result of groundwater decline is reflected in the deterioration of groundwater quality and the increase in its EC. The comparison of the results of EC zoning (Figure 9) and the land use characteristics shows that the values of EC in the protected area, depicted in Figure 2(a), are low, which seems reasonable due to the lack of land use and the absence of wells for water extraction for various purposes.

The values of calcium and bicarbonate ions are high in the eastern part of Lake Urmia, where limestone outcrops are located (Figure 2(c)). The western parts of the aquifer toward Lake Urmia have sodium chloride due to the salt marshes of the lake. The aforementioned parts are the outflow or discharge points of the groundwater. In recent decades, desertification has gradually increased in the Sarab Plain, which changes the chemical quality of groundwater and reduces the amount of freshwater. Salt and limestone domes may be one of the causes of salinity in the central and northeastern parts of the study area.

The geographical location and climate of the region have a great influence on the quality of groundwater. Water in mountainous regions has lower TDS content, while TDS content in plains and deserts is high. Low precipitation, high temperature and intensity of evaporation increase the solute concentration (Famiglietti 2014). Our results show that in the western part of the studied area with the prevailing semi-arid climate (Figure 2(b)), the EC of groundwater and the concentration of salinity have also increased.

In addition, the Aji-Chai River dissolves various salts along its path and has a relatively high salinity. Our results show that the EC in the areas along the mainstream of the river increases from east to west (outlet of the Aji-Chai River from the watershed). Thus, the results clearly show the direct exchange of surface and groundwater in the involved study areas.

Moreover, the faults can change the elevation and depression of the earth's surface and affect the distribution and quality of natural resources such as groundwater. These huge subsurface fractures sometimes change the course of rivers and create different water courses and springs. Our results show that the quality of groundwater changes greatly across the width of the fault. Also, the EC of groundwater in the northern part of the fault is higher than that of the southern part. Therefore, the groundwater in the northern part of the fault is not suitable for drinking and industrial purposes. A high level of EC is also clearly observed at the intersection points of the faults.

### 3.5. Limitations, recommendations, and perspectives for future studies

The analysis discussed in the previous section showed how there is a clear impact of land use characteristics, climatic conditions, and geological features on groundwater quality variables. Due to the scarce availability of information on land use characteristics, climatic conditions, and geological features, in the current research, it was not possible to consider them as further potential input variables to train soft computing models to assess groundwater EC. Given the influence that such drivers could have on EC, this could be considered as a clear limitation of the proposed framework and it could be recommended for future studies in other territorial contexts or in Iran, using alternative data sources from remote sensing. An interesting perspective for future study is to evaluate the effect of concentrations of other minerals and heavy metals (e.g. iron, strontium, manganese, fluorine, barium, cadmium, and aluminum), on EC, which could be useful for health risk assessments. To this aim it would be possible to explore novel complex-hybrid soft computing approaches at various spatial scales and under different climatic conditions.

## 4. CONCLUSION

The current research was carried out with the aim of exploring the potentialities of soft computing based methods for estimating the EC of groundwater using single and combined-wavelet approaches. In this study, different models were developed in several study areas located in the East Azerbaijan province (Iran). The results of this research may enable the effective use of soft computing methods for temporal and spatial estimation of groundwater resources quality variables, highlighting the ability of WT to remove the uncertainties of hydrological phenomena. In the present study, it was proposed an analysis of possible drivers for groundwater quality characteristics, such as land use, climatic conditions, and geologic structure, the location of faults, the exchange of surface and groundwater resources, which can have a direct influence on the values of dissolved ions in groundwater that is reflected in the spatial variability of the EC across the study area. The use of various hybrid soft computing models for temporal and spatial estimation of groundwater resources quality variables along with the study of the impact of point pollution factors, including industrial effluents, on groundwater resources quality can be considered a priority in the proposals of this research, and the general outcomes as well as the developed framework could be easily extended to

other territorial contexts. Moreover, the obtained results may effectively assist managers and policy-makers to adopt appropriate sustainable management plans to address future challenges in Iran.

## ACKNOWLEDGEMENTS

The authors would like to thank Daneshvaran Omran-Ab (DOA) Consulting Company for its financial support to meet a part of data providing costs [Grant No. 01-03].

## AUTHORS' CONTRIBUTIONS

S.Z.-G. conceptualized the study, did data curation, did software analysis, performed the methodology, prepared the original draft, reviewed and edited the article. R.S. did the modeling, coding, validation, wrote the original draft, prepared figures, and reviewed the article. S.F. prepared the methodology, validated the study, prepared figures, wrote and reviewed the article. L.V.N. prepared the methodology, reviewed and edited the article. C.D.M. prepared the methodology, reviewed and edited the article. D.P. conceptualized the study, reviewed, and edited the article.

## DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

## CONFLICT OF INTEREST

The authors declare there is no conflict.

## REFERENCES

- Aftab, S. M., Maqsood, T., Hassan, S., Hannan, A., Zaidi, A. R. & Tahir, R. 2018 Hypothetical geological model affecting groundwater quality in doabs of Indus Basin, Punjab, Pakistan. *International Journal of Economic and Environmental Geology* **9** (4), 1–11. <https://doi.org/10.46660/ijeeg.Vol9.Iss4.2018.168>.
- Aghajari, M., Mozayyan, M., Mokarram, M. & Chekan, A. A. 2019 Relationship between groundwater quality and distance to fault using adaptive neuro fuzzy inference system (ANFIS) and geostatistical methods (case study: North of Fars Province). *Spatial Information Research* **27** (5), 529–538. <https://doi.org/10.1007/s41324-019-00253-5>.
- Akoachere, R. A. I., Eyong, T. A., Egbe, S. E., Wotany, R. E., Nwude, M. O. & Yaya, O. O. 2019 Geogenic imprint on groundwater and Its quality in parts of the Mamfe Basin, Manyu Division, Cameroon. *Journal of Geoscience and Environment Protection* **7** (05), 184. doi:10.4236/gep.2019.75016.
- Al-Adhaileh, M. H., Aldhyani, T. H., Alsaade, F. W., Al-Yaari, M. & Albaggar, A. K. A. 2022 Groundwater quality: The application of artificial intelligence. *Journal of Environmental and Public Health*. doi:10.1155/2022/8425798.
- Andaryani, S., Trolle, D., Nikjoo, M. R., Moghadam, M. H. & Mokhtari, D. 2019 Forecasting near-future impacts of land use and climate change on the Zilbier river hydrological regime, northwestern Iran. *Environmental Earth Sciences* **78** (6), 1–14. <https://doi.org/10.1007/s12665-019-8193-4>.
- Badeenezhad, A., Radfard, M., Abbasi, F., Jurado, A., Bozorginia, M., Jalili, M. & Soleimani, H. 2021 Effect of land use changes on non-carcinogenic health risks due to nitrate exposure to drinking groundwater. *Environmental Science and Pollution Research* **28** (31), 41937–41947. <https://doi.org/10.1007/s11356-021-13753-5>.
- Barzegar, R., Asghari Moghaddam, A. & Tziritis, E. 2016 Assessing the hydrogeochemistry and water quality of the Aji-Chay River, northwest of Iran. *Environmental Earth Sciences* **75** (23), 1–15. <https://doi.org/10.1007/s12665-016-6302-1>.
- Chitsazan, M. & Manshadi, B. D. 2021 Role of Mehriz Fault in hydrochemical evolution and groundwater flow of Yazd aquifer, central Iran. *Arabian Journal of Geosciences* **14** (7), 1–16. <https://doi.org/10.1007/s12517-020-06395-3>.
- Ekemen Keskin, T., Özler, E., Şander, E., Dügenci, M. & Ahmed, M. Y. 2020 Prediction of electrical conductivity using ANN and MLR: A case study from Turkey. *Acta Geophysica* **68**, 811–820. <https://doi.org/10.1007/s11600-020-00424-1>.
- El Ayni, F., Cherif, S., Jrad, A. & Trabelsi-Ayadi, M. 2011 Impact of treated wastewater reuse on agriculture and aquifer recharge in a coastal area: Korba case study. *Water Resources Management* **25** (9), 2251–2265.
- Emadi, A., Zamanzad-Ghavidel, S., Sobhani, R. & Rashid-Niaghi, A. 2021 Multivariate modeling of groundwater quality using hybrid evolutionary soft-computing methods in various climatic condition areas of Iran. *Journal of Water Supply: Research and Technology-Aqua* **70** (3), 328–341. <https://doi.org/10.2166/aqua.2021.150>.
- Esmailbeiki, F., Nikpour, M. R., Singh, V. K., Kisi, O., Sihag, P. & Sanikhani, H. 2020 Exploring the application of soft computing techniques for spatial evaluation of groundwater quality variables. *Journal of Cleaner Production* **276**, 124206. <https://doi.org/10.1016/j.jclepro.2020.124206>.
- Famiglietti, J. S. 2014 The global groundwater crisis. *Nature Climate Change* **4** (11), 945–948.
- Ferreira, C. 2006 *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*, Vol. 21. Springer, Berlin, Germany.

- Ghazi, B. & Jeihouni, E. 2022 Projection of temperature and precipitation under climate change in Tabriz, Iran. *Arabian Journal of Geosciences* **15** (7), 1–11. <https://doi.org/10.1007/s12517-022-09848-z>.
- Ghobadi, A., Cheraghi, M., Sobhanardakani, S., Lorestani, B. & Merrikhpour, H. 2022 Groundwater quality modeling using a novel hybrid data-intelligence model based on gray wolf optimization algorithm and multi-layer perceptron artificial neural network: A case study in Asadabad Plain, Hamedan, Iran. *Environmental Science and Pollution Research* **29** (6), 8716–8730. <https://doi.org/10.1007/s11356-021-16300-4>.
- Honarbakhsh, A., Azma, A., Nikseresht, F., Mousazadeh, M., Eftekhari, M. & Ostovari, Y. 2019 Hydro-chemical assessment and GIS-mapping of groundwater quality parameters in semi-arid regions. *Journal of Water Supply: Research and Technology-Aqua* **68** (7), 509–522. <https://doi.org/10.2166/aqua.2019.009>.
- Hrnjica, B. & Danandeh Mehr, A. 2018 *Optimized Genetic Programming Applications: Emerging Research and Opportunities: Emerging Research and Opportunities*. Available from: <https://www.inso.gov.ir/>, <https://www.wrm.ir/>
- Jafari, R., Torabian, A., Ghorbani, M. A., Mirbagheri, S. A. & Hassani, A. H. 2019 Prediction of groundwater quality parameter in the Tabriz plain, Iran using soft computing methods. *Journal of Water Supply: Research and Technology-AQUA* **68** (7), 573–584. <https://doi.org/10.2166/aqua.2019.062>.
- Jalalkamali, A. & Sheykhbahaei, A. 2022 Modeling of groundwater salinity on the Persian Gulf coastal plain By using linear moments and ANFIS-PSO. *International Journal of Coastal and Offshore Engineering* **7** (3), 43–49. <https://doi.org/10.22034/ijcoe.2022.155143>.
- Kadkhodazadeh, M. & Farzin, S. 2022 A novel hybrid framework based on the ANFIS, discrete wavelet transform, and optimization algorithm for the estimation of water quality parameters. *Journal of Water and Climate Change* **13** (8), 2940–2961.
- Ketchemen-Tandia, B., Boum-Nkot, S. N., Ebondji, S. R., Nlend, B. Y., Emvoutou, H. & Nzegue, O. 2017 Factors influencing the shallow groundwater quality in four districts with different characteristics in urban area (Douala, Cameroon). *Journal of Geoscience and Environment Protection* **5** (08), 99. doi:10.4236/gep.2017.58010.
- Khalaj, M., Kholghi, M., Saghafian, B. & Bazrafshan, J. 2019 Investigation about climate change and human activity effects on groundwater level and groundwater quality in semiarid region. *Iran-Water Resources Research* **15** (2), 278–290. doi: 20.1001.1.17352347.1398.15.2.21.0.
- Khashei-Siuki, A. & Sarbazi, M. 2015 Evaluation of ANFIS, ANN, and geostatistical models to spatial distribution of groundwater quality (case study: Mashhad plain in Iran). *Arabian Journal of Geosciences* **8** (2), 903–912. <https://doi.org/10.1007/s12517-013-1179-8>.
- Komba, E. A., Munubi, R. N. & Chenyambuga, S. W. 2020 Comparative evaluation of water quality parameters and growth performance of sex-reversed Nile tilapia (*Oreochromis niloticus*) raised in two different climatic conditions in Tanzania. Available from: <https://www.suaire.sua.ac.tz/handle/123456789/3703>.
- Li, J., Lu, W., Wang, H., Bai, Y. & Fan, Y. 2020 Groundwater contamination sources identification based on kernel extreme learning machine and its effect due to wavelet denoising technique. *Environmental Science and Pollution Research* **27** (27), 34107–34120. <https://doi.org/10.1007/s11356-020-08996>.
- Lu, H., Tian, P. & He, L. 2019 Evaluating the global potential of aquifer thermal energy storage and determining the potential worldwide hotspots driven by socio-economic, geo-hydrologic and climatic conditions. *Renewable and Sustainable Energy Reviews* **112**, 788–796. <https://doi.org/10.1016/j.rser.2019.06.013>.
- Maghrebi, M., Noori, R., Partani, S., Araghi, A., Barati, R., Farnoush, H. & Torabi Haghighi, A. 2021 Iran's groundwater hydrochemistry. *Earth and Space Science* **8** (8), e2021EA001793. <https://doi.org/10.1029/2021EA001793>.
- Montaseri, M., Zaman Zad Ghavidel, S. & Sanikhani, H. 2018 Water quality variations in different climates of Iran: Toward modeling total dissolved solid using soft computing techniques. *Stochastic Environmental Research and Risk Assessment* **32** (8), 2253–2273. <https://doi.org/10.1007/s00477-018-1554-9>.
- Motevalli, A., Pourghasemi, H. R., Hashemi, H. & Gholami, V. 2019 Assessing the vulnerability of groundwater to salinization using GIS-based data-mining techniques in a coastal aquifer. In: *Spatial Modeling in GIS and R for Earth and Environmental Sciences*. Elsevier, pp. 547–571. <https://doi.org/10.1016/B978-0-12-815226-3.00025-9>.
- Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C. & Gupta, H. V. 2021 What role does hydrological science play in the age of machine learning? *Water Resources Research* **57** (3), e2020WR028091. <https://doi.org/10.1029/2020WR028091>.
- Nordin, N. F. C., Mohd, N. S., Koting, S., Ismail, Z., Sherif, M. & El-Shafie, A. 2021 Groundwater quality forecasting modelling using artificial intelligence: A review. *Groundwater for Sustainable Development* **14**, 100643. <https://doi.org/10.1016/j.gsd.2021.100643>.
- Noto, L. V., Cipolla, G., Pumo, D. & Francipane, A. 2023 Climate change in the Mediterranean basin (Part II): A review of challenges and uncertainties in climate change modeling and impact analyses. *Water Resources Management*. <https://doi.org/10.1007/s11269-023-03444-w>.
- Pannell, D. & Rogers, A. 2022 Agriculture and the environment: Policy approaches in Australia and New Zealand. *Review of Environmental Economics and Policy* **16** (1), 126–145.
- 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 and Environment Journal* **35** (1), 67–83. <https://doi.org/10.1111/wej.12595>.
- Pumo, D. & Noto, L. V. 2023 Exploring the use of multi-gene genetic programming in regional models for the simulation of monthly river runoff series. *Stochastic Environmental Research and Risk Assessment* 1–25. <https://doi.org/10.1007/s00477-022-02373-1>.

- Ren, J., Chen, Q., Ma, D., Xie, R., Zhu, H. & Zang, S. 2021 Study on a fast EC measurement method of soda saline-alkali soil based on wavelet decomposition texture feature. *Catena* **203**, 105272.
- Rhif, M., Ben Abbes, A., Farah, I. R., Martínez, B. & Sang, Y. 2019 Wavelet transform application for/in non-stationary time-series analysis: A review. *Applied Sciences* **9** (7), 1345. <https://doi.org/10.3390/app9071345>.
- Sadeghi, A., Galalizadeh, S., Zehtabian, G. & Khosravi, H. 2021 Assessing the change of groundwater quality compared with land-use change and precipitation rate (Zrebar Lake's Basin). *Applied Water Science* **11** (11), 1–15. <https://doi.org/10.1007/s13201-021-01508-z>.
- Sadeqi, A. & Dinpashoh, Y. 2020 Projection of precipitation and its variability under the climate change conditions in the future periods (Case study: Tabriz). *Environment and Water Engineering* **5** (4), 339–350. <https://doi.org/10.22034/jewe.2020.210941.1339>.
- Shtull-Trauring, E., Cohen, A., Ben-Hur, M., Tanny, J. & Bernstein, N. 2020 Reducing salinity of treated waste water with large scale desalination. *Water Research* **186**, 116322.
- Utomo, Z., Hidayah, I. & Nur Rizal, M. 2021 Comparison of electrical conductivity prediction models using Gaussian process. *IJITEE (International Journal of Information Technology and Electrical Engineering)* **5** (4), 159–165. <http://dx.doi.org/10.22146/ijitee.70684>.
- Yang, Q., Zhang, J., Hou, Z., Lei, X., Tai, W., Chen, W. & Chen, T. 2017 Shallow groundwater quality assessment: Use of the improved Nemerow pollution index, wavelet transform and neural networks. *Journal of Hydroinformatics* **19** (5), 784–794. <https://doi.org/10.2166/hydro.2017.224>.

First received 4 May 2023; accepted in revised form 19 July 2023. Available online 31 July 2023