Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Evaluation of total dissolved solids in rivers by improved neuro fuzzy approaches using metaheuristic algorithms

Mahdieh Jannatkhah
University of Tabriz
Rouhollah Davarpanah (davarpanah.rouhollah@gmail.com)
Tarbiat Modares University
Bahman Fakouri
Tarbiat Modares University
Ozgur Kisi
Technical University of Lübeck

Research Article

Keywords: surface water quality, TDS, metaheuristic algorithms, ANFIS, hybrid models, TOPSIS

Posted Date: October 16th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3423568/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations:
No competing interests reported.

Tables 1 to 4 are available in the Supplementary Files section.

Evaluation of total dissolved solids in rivers by improved neuro fuzzy approaches using metaheuristic algorithms

Mahdieh Jannatkhah[1], Rouhollah Davarpanah[*2], Bahman Fakouri[2], Ozgur Kisi[3,4]

[1] Faculty of Civil Engineering., University of Tabriz, Tabriz, Iran

[2] Department of Water Engineering and Management, Tarbiat Modares University, Tehran, Iran

[3] Department of Civil Engineering, Technical University of Lübeck, 23562, Lübeck, Germany

[4] Department of Civil Engineering, Ilia State University, 0162, Tbilisi, Georgia

*** Corresponding Author:** Rouhollah Davarpanah

Department of Water Engineering and Management, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran

Email address: davarpanah.rouhollah@gmail.com

Tel. +989179136613, +987137345680

Abstract

Substantial deterioration of surface water quality, mainly caused by human activities and climate change, makes the assessment of water quality a global priority. Thus, in this study, four metaheuristic algorithms, namely the particle swarm optimization (PSO), differential evolution (DE), ant colony optimization algorithm (ACOR), and genetic algorithm (GA), were employed to improve the performance of the adaptive neuro-fuzzy inference system (ANFIS) in the evaluation of surface water total dissolved solids (TDS). Monthly and annual TDS were considered as target variables in the analysis. In order to evaluate and compare the authenticity of the models, an economic factor (execution time) and statistical indices of the coefficient of determination (R2), Kling Gupta efficiency (KGE), root mean squared error (RMSE), mean absolute error (MAE), and Nash-Sutcliff efficiency (NSE) were utilized. The results revealed that the hybrid methods used in this study could enhance the classical ANFIS performance in the analysis of monthly and annual TDS of both stations. For more clarification, the models were ranked using the TOPSIS approach by simultaneously applying the effects of statistical parameters, temporal and spatial change factors, and execution time. This approach significantly facilitated decision-making in ranking models. The ANFIS-ACOR annual model considering discharge had the best performance in the Vanyar Station; Furthermore, ANFIS-ACOR monthly model ignoring discharge was outstanding in the Gotvand Station. In total, after utilizing two defined and proposed temporal and spatial change factors, ANFIS-ACOR and ANFIS-DE hybrid models had the best and worst performance in TDS prediction, respectively.

Keywords: surface water quality, TDS, metaheuristic algorithms, ANFIS, hybrid models, TOPSIS.

1. Introduction

Rivers are an indispensable natural water resource, fulfilling diverse human requirements including industrial (22%), domestic (8%), and irrigation (70%) demands (Hossain, 2019). Because of the growing demand for fresh water and restricted access to water resources (Khataee et al., 2013)**,** observation and control of river water quality have an incontrovertible role in environmental conservation and sustainable distribution of natural resources (Varol et al., 2022; Zhang et al., 2021; Deng et al., 2021; Jamei et al., 2020).

One of the widely accepted factors that has been effectively used for studying water quality is total dissolved solids (TDS) (Kabolizadeh et al., 2022; Salmani and Salmani Jajaei, 2016; Ghfolamreza et al., 2016;). TDS comprises dissolved organic matters and a variety of inorganic salts e.g., sodium (Na^{+}), magnesium (Mg^{2+}), calcium (Ca^{2+}), and potassium (K^{+}) as cations, as well as chloride (Cl^{-}), sulfate ($SO4^{2}$), nitrates (NO^{-}), and bicarbonates (HCO^{3-}), as anions (Sun et al., 2021). Based on the World Health Organization WHO standards, the acceptable TDS range for agricultural uses is 450–2000 mg/l (Gawande & Sarode, 2021). As high concentrations of this parameter can overshadow human health and the environment, it is important to focus on it in the evaluation of water resources, especially surface waters (Chellaiah et al., 2021). The parametric intricacy brings about a high cost and time for on-site investigations and laboratory examinations of water quality (Asadollah et al., 2021). Numerical methods used for water quality modeling have been shown to be time-consuming and have weak optimization performance. Artificial intelligence-based models eliminate these limitations and also have the benefit of being less sensitive to missing values and being able to perform sophisticated mathematical calculations with a huge quantity of data and nonlinear structures (Najafabadipour et al., 2022). Various studies from different fields have also used artificial intelligence-based techniques like artificial neural networks (ANNs), ANFIS, and gene expression programming (GEP) to estimate and predict water quality (Al-Mukhtar & Al-

74 Yaseen, 2019; Antanasijević et al., 2014; Hu et al., 2020; Khalil et al., 2011; Palani et al., 2008;
75 Yoosefdoost et al., 2022).

76 Among all these methods, ANFIS was introduced as a practical technique for evaluating water
77 quality (Shah et al., 2021). The ANFIS is a neural network based on the Takagi-Sugeno fuzzy
78 inference system (Abd El-Mageed et al., 2022). By taking advantage of the mathematical
79 characteristics of general function estimators, fuzzy systems can effectively process and
80 integrate logical information. This method is transparent to the user and results in fewer
81 memorization errors (Tutmez et al., 2006). In addition, this method combines the advantages
82 of ANN and fuzzy approaches into a single framework by integrating the fuzzy inference
83 system into the adaptive network framework (Ying & Pan, 2008). Therefore, the engineering
84 application of the neural fuzzy method has been widely applied in hydrology and water quality
85 investigations (Gavili et al., 2018).

86 In neuro-fuzzy networks, choosing the proper parameters and determining the appropriate
87 network structure is essential in analysis. The more accurate and efficient the algorithms used
88 for training, the more successful the model will be. Most of the workable variable training
89 algorithms for fuzzy networks are gradient algorithms, especially backpropagation and
90 Levenberg-Marquardt. Despite these algorithms being prevalent in the literature, they have
91 weaknesses that need to be considered. For instance, some of these algorithms are potent for
92 determining only local optimal points (Azad et al., 2018). In addition, some training algorithms,
93 such as Levenberg-Marquardt, are methodologically complicated and need high data
94 processing memory for calculations that are less likely to be optimized. Thus, new approaches
95 are required to avoid the weaknesses of gradient-based algorithms. Not only can the
96 metaheuristic algorithms (EA) discover solutions globally without providing exclusively local
97 optimal points, but also an intricate problem with a high calculation volume can be simply
98 evaluated if a proper algorithm is used.

Many studies have been devoted to estimating surface water quality using modified AI methods. Jalalkamali (2015) applied the ANFIS improved with the ant colony optimization algorithm (ACOR), particle swarm optimization (PSO), differential evolution (DE), and genetic algorithm (GA) to evaluate the groundwater quality of Kerman Province, Iran. ANFIS demonstrated appropriate performance in evaluating all water quality parameters during the training phase. The findings indicated that applied metaheuristic algorithms significantly improve the accuracy of ANFIS in predicting river water quality parameters. Banadkooki et al. (2020) employed the artificial neural networks (ANN), ANFIS, and support vector machine (SVM) models for the estimation of the total dissolved solids of aquifers. The grey wolf optimization (GWO), gravitational search algorithm (GSA), moth flame optimization (MFO), particle swarm optimization (PSO), shark algorithm (SA), and cat swarm optimization (CSO) were used to train the models. The results showed that the ANFIS-CSO and ANFIS-MFO models outperformed the other ones. Aghel et al. (2019) evaluated water quality parameters using a hybrid particle swarm optimization-neural fuzzy PSO-ANFIS approach. The results showed that applying two methods was highly satisfactory for estimating inorganic water quality factors. The flexibility of the ANFIS-PSO method in modeling was better than the ANFIS approach. Azad et al. (2018) predicted the water quality parameters of the Gorganrood River using ANFIS optimized by Genetic Algorithm (GA), Ant Colony Optimization for Continuous Domains (ACOR), and Differential Evolution (DE) algorithms. The results showed that, for predicting electrical conductivity (EC) and total hardness (TH) in the test stage, ANFIS-DE was the most appropriate model. Liu et al. (2022) used support vector machine particle swarm optimization with deep learning to investigate the causes of water pollution and control countermeasures in the Liaohe estuary. The results showed that the SVM-PSO approach has excellent predictive performance.

In recent decades, various pollution loads, including agricultural drainages, and urban and industrial wastewaters have damaged the water quality of many surface water resources (Wang et al., 2020). The Karun River is the largest and only navigable river in southwest Iran, with a contentious water quality (Diagomanolin et al., 2004). An environmental catastrophe has emerged from the building of the massive Upper Gotvand Dam in the Karun River. This is because of the accumulation of 66.5 million metric tonnes of dissolved salt in the reservoir and a dramatic increase in the reservoir's water salinity to 200 g/L (Jalali et al., 2019). The Aji Chay River, in northwest Iran, is also a very important river of the country that is problematic in terms of water quality. This river drains into Lake Urmia (LU), the world's second-largest hyper-saline lake, and its water quality influences LU's water quality (Andaryani et al., 2021). Because of the salt formations that exist before the Vanyar Station of the Aji Chay River basin, the water of this river is salty. These two rivers can serve as excellent representatives in evaluating the capability of optimized artificial intelligence methods in studying surface water quality. Therefore, Vanyar and Gotvand stations were selected as study areas in this research.

As yet, no published research has compared different metaheuristic algorithms for enhancing the precision of the ANFIS method in the simultaneous evaluation of monthly and annual TDS for the Aji Chay and Karun rivers. Thus, in this study, four metaheuristic algorithms including genetic algorithm (GA), ant colony optimization (ACOR), particle swarm optimization (PSO), and differential evolution (DE) were used to optimize the performance of ANFIS model accuracy in the evaluation of monthly and annual TDS of Vanyar Station (the Aji Chay River), and Gotvand station (the Karun River) in Iran. The effect of omitting the discharge parameter from the input dataset was also assessed in all calculations. Eventually, the TOPSIS method was utilized to rank the preferred hybrid models and select the most precise one. Given the expected proximity in performance between these models under specific spatial and temporal conditions, two factors relating to spatial and temporal changes were proposed and employed

148 in the TOPSIS method to facilitate evaluating and decision-making procedures. The outline of
149 this research is also shown in Fig. 1.

150 2. Material and methods

151 2.1. Study area and data description

152 In order to accomplish the objective of this study, two distinct areas in Iran were selected based
153 on their different climatic conditions and exposure to salt domes and water pollution; the Karun
154 River in a hot and dry climate (Eskandari and Mahmoudi Sarab, 2022), and the Aji Chay River
155 in a cold and mountainous climate (Sharafkhani et al., 2018). Fig. 2 shows the maps of the
156 Gotvand and Vanyar Stations.

157 At around 950 kilometers, the Karun River is the longest river situated in the southwest area of
158 Iran and is the only navigable river in the country providing water to several cities and villages
159 (Fakouri et al., 2019; Golshan et al., 2020). It also meets the water demands of various
160 industries, including steel, oil, petrochemical, sugarcane, paper, and cement industries, as well
161 as farmlands along the river. Since agricultural, domestic, and industrial wastewater discharges
162 into the Karun River, its salinity has been increasing (Karamouz et al., 2009). One of the most
163 significant hydrometric stations on the Karun River is Gotvand Station, which is situated at 48°
164 49' east longitude and 32° 14' north latitude. The elevation of this station is 243 meters
165 (Radmanesh et al., 2013). The water quality of the Karun River has been severely impacted by
166 the Gachsaran salt domes in this area. Downstream agricultural fields have also been degraded
167 and lost their productivity as a result of water salinization in the dam reservoir. Saline water
168 destroys soil structure and prevents crops from absorbing irrigated water (Gutiérrez & Lizaga,
169 2016). This research used a 45-year (1971–2015) water quality data of Gotvand Station.

The Aji Chay River, which has an average discharge of 16.4 (m3 s-1) and is recognized as one of the most saline rivers of northwest Iran, was also subject to evaluation in this study. This river originates from the Qusheh-Dagh and Sabalan elevations and eventually ends up in the Urmia Lake Basin, the largest lake in the Middle East. The water quality of the Aji Chay River has a direct impact on the drinking water of West Azerbaijan Province as well as the water quality of Urmia Lake. The salinity of the river water is also significantly affected by the salt domes that lie under the reservoir region behind the Vanyar dam (Hossein and Moghaddam, 2006). Hence, Vanyar Station, one of the major hydrometric stations of the Aji Chay River, was selected as the representative of this river. It is located in the coordination of 38° 07′ 00″ N and 46° 24′ 18″ E, with an elevation of 3882 meters above sea level. A 45-year water quality data (1966-2011) from the Vanyar Station was used in this research.

The monthly WQ data of Gotvand and Vanyar Stations utilized in this research were obtained from Iran water resources management company (wrm.ir). The annual data used in the analysis was derived from the average monthly data of each station. The selection of the studied stations was based on the availability of long-term data, lack of outliers, and water pollution vulnerability.

Among the chemical parameters, the ones that had a high correlation with TDS, including Cl^-, $SO^{2-}4$, Ca2+, $Mg^{2+,}$ and Na^+ were selected as input parameters. In parallel, among the physical parameters, discharge, which has a great impact on the quality of rivers, was taken into account (Gao et al., 2018; Seiler et al., 2020). The maximum, average, and minimum monthly TDS of both stations are shown in Table 1. For the Vanyar station, due to the limited access to the data of recent years, the analysis was performed using the data from 1975 to 2012. Fig. 3 represents the TDS and discharge time series of study areas, and scatter plots of TDS versus discharge of the Gotvand and Vanyar Stations.

The data must be preprocessed in order to prepare them for analysis using ANFIS and metaheuristic approaches. There are various WQ parameters that can be employed in the evaluation of TDS. However, only a limited number of inputs are used for modeling because of constraints like time, memory space, and computational volume. Therefore, candidate parameters need to be evaluated in terms of correlation. For this purpose, in this research, linear regression analysis was performed on various water quality parameters of the Gotvand and Vanyar Stations using the SPSS software). The findings showed that Cl^-, SO^{2-}_4, Ca^{2+}, Mg^{2+}, Na^+, and discharge (Q) had the highest linear regression with TDS (Azad et al., 2019). After choosing the best input parameters, for both study areas, three classification types were used: 30-70 i.e., 30% of the data for testing and 70% for training, 20-80, and 40-60 modes. When compared to the other two dataset classification types, the 30-70 type performed the best (Jannatkhah et al., 2021). In addition, the datasets of both stations were normalized between 0 and 1 using Eq. 1.

$$x_{new} = (x - x_{min})/(x_{max} - x_{min}) \quad (1)$$

where x_{new}, x, x_{min}, and x_{max} are the standardized values of the predicted data, the original data, and the minimum and maximum values in the data set, respectively (Khoi et al., 2022).

2.2. Applied models

2.2.1. Adaptive Neuro Fuzzy Inference System (ANFIS)

The Adaptive Neuro-Fuzzy Inference System ANFIS is an AI-based technique that is derived from Takagi-Sugeno TS fuzzy systems. The ANFIS development procedure consists of distinguishing the most dependent inputs that have a correlation with a targeted output. The optimal rules, types, and the number of the associated membership functions (MFs) should be determined to select the optimum ANFIS structure with minor resulting errors. Two TS fuzzy sets of "if-then" rules in a typical ANFIS structure are shown in Eqs. 2 and 3.

Rule 1: If x_1 is A_1 and x_2 is B_1, then $y_1 = f_1 = p_1x_1 + q_1x_2$ (2)

Rule 2: If x_1 is A_1 and x_2 is B_1, then $y_2 = f_2 = p_2x_1 + q_2x_2$ (3)

Ai and B_j indicate the linguistic degrees, while p_1, q_1, p_2, and q_2 are ANFIS determinants. The ANFIS structure consists of five layers (Ying & Pan, 2008). The following descriptions are concise explanations of the role of these layers (Al-Mukhtar and Al-Yaseen, 2019).

Layer 1, the fuzzification layer, receives the input values and identifies the MFs.

Layer 2, or the rule layer, develops the firing strengths for the rules.

Layer 3, or the standardization layer, normalizes the calculated strengths.

Layer 4 takes the standardized values and then generates parameter sets.

Layer 5, or the defuzzification layer, returns the values to the ultimate result (Shah et al., 2021).

The number of epochs in ANFIS was set at 500 and the results did not represent a noticeable influence on the authenticity of models beyond this epoch value. The best values of step-size-decrement, initial step-size, epochs, number of clusters, and initial-increment were 0.9, 0.01, 500, 10, and 1.1, respectively. The approaches used in this study were all coded in MATLAB R2021b and were executed on a computer equipped with an Intel® CoreTM i7 processor.

2.2.2. Particle swarm algorithm (PSO)

PSO is a metaheuristic algorithm formulated using the stochastic optimization approach. In the PSO method, there is a particular search interval in which the solutions of the algorithm are marked as particles (Banadkooki et al., 2020; Samanataray and Sahoo, 2021). The fitness function assigns a fitness value for every particle. The PSO approach is characterized by exploring the solution domain for particles in better condition. When the position of particles is changed, they try to find the optimum place according to their previous experiences. Particles are cognizant of their location and the position of other particles as well. As an example, in the

problem d-dimensional space, particle i has a location that can be calculated by equation 3 in which t shows the iteration of the particle. In addition, the particle has a velocity V leading the particle to the new location Eq. 4. Particles also have a memory to remember their position and the location of other particles Eq. 5. After every iteration, particles are updated with their two best values including the "best solution" p_{best} and the "global best value" g_{best}. After determination of the p_{best} and g_{best}, the particle velocity is updated based on Eqs. 4 and 5 (Al-qaness et al., 2021).

$$X_i^t = (x_{i1}^t, x_{i2}^t, \dots., x_{id}^t\ ) \quad (4)$$

$$V_i^t = (v_{i1}^t, v_{i2}^t, \dots., v_{id}^t\ ) \quad (5)$$

$$P_i^t = (p_{i1}^t, p_{i2}^t, \dots., p_{id}^t) \quad (6)$$

$$v_i^{t+1} = WV_i^t + C_1^t r_1 (P_i^t - X_i^t) + C_2^t r_2 \left(P_g^t - X_i^t\right) \quad (7)$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} \quad (8)$$

The learning indicators are shown with C_1 and C_2. These two factors qualify the movement of each member after iterations. Generally, $C_1 = C_2$ and have a value equal to 2. r_1 and r_2 are two random numbers with a range between 0 and 1. The inertial weight is shown by W, and its initial range is normally between 0 and 1. In this research, although 200 iterations were used for this procedure, there was no improvement in the outputs as the number of iterations increased over 200. Furthermore, the initial population size was set to 100, the velocity bounds were set to -2 and +2, the inertia weight was selected between 0.5 and 1.5, and the personal and global best learning coefficients were determined between 1 and 5.

2.2.3. Differential evolution (DE)

DE is classified as a stochastic optimization algorithm driven by biological procedures, where the permanence of the fittest is needed for adoption with intrinsic genetic characteristics (Wang

and Zhao, 2013). If the target optimization functions are called with f, in the following equation, R is related to the observed dataset, and D indicates the objective function parameters f (V).

$$f(V): R^D \rightarrow R \tag{9}$$

The purpose of a DE approach is to minimize the objective function by utilizing the values of optimized parameters:

$$V \in R^D \tag{10}$$

The vector is shown by V. It includes the parameters of the objective function of D. The objective function is the mean squared error between the observed and predicted TDS. The objective function parameters are indicated as follows:

$$v_i^L \leq v_i \leq v_i^U \tag{11}$$

The lower and upper boundaries set for the objective function are shown with vi L and vi U, respectively. DE operates not on single solutions but instead on a population PG of candidate solutions; the generation of a population is denoted by G. The population predicted by the DE approach can be shown as follows:

$$P_G = (\vartheta_1, G, \vartheta_2, G, \dots, \vartheta_{1NP}, G) \qquad \mathrm{G} = 0, \dots, \mathrm{G_{max}} \tag{12}$$

In Eq. 12, G_{max} is the uttermost generation that mostly performs as the cessation standard of DE. Each vector includes the accurate parameter of D assumed as a single chromosome.

$$v_{i,G} = \left(v_{1,i,G}, v_{2,i,G}, \dots, v_{D,i,G} \right) \qquad \mathrm{i} = 1,2, \dots, \mathrm{NP} \qquad \mathrm{G} = 0, \dots, \mathrm{G_{max}} \tag{13}$$

An initial population needs to be evolved to assign a leading point for the best search. As a whole, except for the optimum problem variables, it is impossible to find information about the best solutions. Accordingly, PG with the value of 0 is one of the approaches to specifying the initial population. In point of fact, it is the random selection of limitations shown as follows:

$$v_{j,i,0} = \mathrm{rand}_j[0.1](v_j^U - v_j^L) + V_j^L \qquad i = 1,2,\ldots,NP \quad ,j = 1,2.\ldots,D \tag{14}$$

The random value distributed consistently is shown by randj [0, 1] in a range between 0 and 1, which is chosen for every current j. The procedure of DE is different from other metaheuristic approaches. PG is incorporated and sampled haphazardly from the initial output to the regular population of vectors in order to generate candidate vectors for the subsequent generation, P_{G+1}. A candidate population of vectors derived from multiple trials, is determined as follows:

$$u_{I,G+1} = \begin{cases} {}^{V_{j,r3,G+F}V_{j,r1,G-}V_{j,r2,G\text{if } v_j^L \le y_{j,I,,G+1} \le v_j^U}} \\ \text{otherwise and } d_j \qquad [0,1]v_j^U - v_j^L \\ \text{otherwise} \quad v_{j,I,G} \end{cases} \tag{15}$$

In each run, the values of r_1, r_2, and r_3 vary. The value of parameter i should be determined. The precise values of r_1, r_2, and r_3 are selected randomly for each value of i. P_{G+1}, which is selected from the current population P_G, is the afterward generation population, and the children's population follows this equation. (Ebtehaj and Bonakdari, 2017)

In this research, the population size, initial mutation factor set, crossover factor, and maximal metaheuristic iteration number were 80, 0.4, 0.9, and 2000, respectively.

2.2.4. Ant Colony Optimization for Continuous Domains (ACOR)

Studies and observations on ant colonies first made by Marco Dorigo served as the basis for the ACOR model in 1992. The propensity of ants to search for the shortest route between their nest and food is one of the most significant and interesting aspects of their behavior. In the real world, ants first travel randomly to and from food sources. The next step is that they go back to their nest while leaving pheromone trails behind them (Dorigo and Blum, 2005). By using the primary population and correcting answers to earlier questions, the ACOR algorithm is intended to produce the best possible response. The primary population is selected at random. The identified paths are assigned a value in the third step according to the values of target

functions like pheromones, and some of them disappear. The fourth step involves creating new paths using a rotary cycle based on the distance between customers and pheromone following the pheromone setting before proceeding to the third stage (Mullen et al. 2009). Each ant, once stationed at a point, utilizes the following Eq. 16 to decide where to move to next:

$$P_{ij}^{kt} P_{ij}^{kt} = \begin{cases} \frac{\tau_{ij}^{\alpha}(t) \times n_{iu}^{\beta}(t)}{\sum_{u \epsilon N_i^k} \tau_{iu}^{\alpha}(t) \times n_{iu}^{\beta}(t)} \\ 0 \qquad\qquad j \notin N_i^k \end{cases} \tag{16}$$

Where η_{ij} equals to $1/d_{ij}$, and d_{ij} represents the distance between ants from each other. τ_{ij} indicates the pheromone value and N_i^k shows the sum of point i and ant k neighbors. After forming the primary population, pheromone is vaporized, and the algorithm is ultimately evaluated to be continued or terminated:

$$\tau_{ij} \leftarrow (1 - \rho) \times \tau_{ij} \sum_{k=1}^{m} \Delta\tau_{ij}^{k} \tag{17}$$

m is the number of ants, ρ is the coefficient of evaporation, and $\Delta\tau ij$ is the concentration of pheromone between i and j vertices. Using Eq. 18, pheromone route updates are performed (Dai et al., 2019).

$$\tau_{ij} \leftarrow (1 - \rho) \times \tau_{ij} \sum_{k=1}^{m} \Delta\tau_{ij}^{best} \tag{18}$$

where $\Delta\tau^{best}{}_{ij}$ represents the optimal pheromone path between i and j vertices. In this study, ACOR replication number was set to 200, and greater repetition had no discernible effect on the quality of evaluations. Furthermore, the initial population was considered with a value of 100.

2.2.5. Genetic algorithm (GA)

A genetic algorithm GA is a metaheuristic approach that uses the principles of natural selection and genetics (Chau, 2006). Although random in some characteristics, GA is not completely coincidental because it makes use of past data to determine where to explore next. Several fields of knowledge have found success using GAs to solve optimization challenges (Ahmad, 2012). Steps for the modified GA are listed as follows:

1. A random collection of chromosomes can be used to generate the current candidate solutions for the given problems.
2. The fitness of each chromosome is determined. The fitness values are utilized in the genetic algorithm to help the simulations find the best possible design options.
3. Sorting the chromosomes with the lowest fitness yielded the best result.
4. Clone the appropriate chromosomes to improve the positive attributes of the population.
5. In order to perform a crossover operation, new springs are derived from the chromosomes of the parents.
6. If the ratio is low and has been predefined, the chromosomes should be mutated.
7. New fitness values for each chromosome are assessed before the analysis continues.
8. Steps 3 through 7 are performed until the non-minimum error requirement is met.

The evolution of individuals within a population throughout generations typically initiates in a random manner. As a result, during each generation, the appropriateness of each member of the population is analyzed, members of the current population are selected on the basis of their fitness, the population is modified reintegrated, and possibly mutated to form another new population (Fu & Li, 2021; Hassan et al., 2020). In this study, iteration numbers and the population size were 200 and 100, respectively.

2.3. Performance Assessment

2.3.1. Statistical indicators

The execution time of models examined by the applied techniques was one of the performance parameters addressed in this study. In addition, statistical parameters such as the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and Nash-Sutcliffe model efficiency coefficient (NSE) were utilized in this study. Eqs. 19–23 are the respective equations of these criteria:

Where x_i and y_i show the observed and estimated values, x and y represent the mean observed and estimated values obtained from ANFIS and its hybrid models, and n is the number of data points. The R^2 index determines the potential and direction of the linear relationship among variables. The Kling-Gupta efficiency (KGE) more effectively incorporates the three parts of the Nash-Sutcliffe efficiency (NSE) of model errors correlation, bias, ratio of variances, or coefficients of variation. In recent years, this index has been frequently utilized for calibrating and evaluating hydrological models. In this index, CC shows the Pearson coefficient value, rm refers to the average of observed values, cm indicates the average of forecast values, rd is the standard deviation of observation values, and cd is the standard deviation of forecast values. The RMSE index indicates the appropriacy of fit related to high values. The MAE calculates a more balanced aspect of suitability of fit at medium data (Karunanithi et al., 1994; Kisi et al., 2014; Montaseri et al., 2018). NSE is a potentially reliable assessment criterion and is widely used for investigating the suitability of fit of hydrologic models (McCuen et al., 2006).

$$R^2 = \left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}} \right)^2 \quad (19)$$

$$KGE = 1 - \sqrt{((CC-1)^2 + (\{cd/rd - 1)\}^2 + \{cm/rm - 1)\}^2)} \quad (20)$$

$$\mathrm{RMSE} = \sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{n}(x_i - y_i)^2} \quad (21)$$

$$\mathrm{MAE} = \frac{\sum_{t=1}^{n} x_i - y_i}{n} \quad (22)$$

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{n}(x_i - y_i)^2}{\sum_{t=1}^{n}(x_i - \bar{x}_i)^2} \quad (23)$$

2.3.2. Decision making

It is challenging to assess the performance of the models and select the optimal algorithms for diverse circumstances due to the large number of models and variety of analyses. In order to accurately choose the optimal models, multi-criteria decision-making techniques must be used. Therefore, the TOPSIS technique was employed in this study to choose the best models since it is among the most widely used Multiple-criterion decision-making MCDM methods (Mahmood and Ali, 2021; Rezaee et al., 2021). In this method, the best choice is the one that is the closest to the positive ideal solution (PIS) and the farthest from the negative ideal solution (NIS) (Karabašević et al., 2020). The decision matrix is as follows:

$$A = \begin{pmatrix} d_{11} & \cdots & d_{1n} \\ \vdots & \ddots & \vdots \\ d_{m1} & \cdots & d_{mn} \end{pmatrix} \quad (24)$$

Where m, n, and A represent the number of criteria, number of alternatives, and decision matrix, respectively. Then, the decision matrix is normalized as follows:

$$r_{ij} = \frac{d_{mn}}{\sqrt{\sum_{i=1}^{m} d_{ij}^2}} \quad (25)$$

Then, the weighted normalized decision matrix Z is formed:

$$Z_{mn} = \left(W_j\, r_{ij}\right)_{mn} \quad (26)$$

PIS and NIS are computed following the formation of the weighted normalized decision matrix:

393 $$PIS = [a_1^+, a_1^+, \dots, a_1^+] = max\{Z_{ij} \mid i = 1,2, \dots, m\} \quad (27)$$

394 $$IS = [a_1^+, a_1^+, \dots, a_1^+] = min\{Z_{ij} \mid i = 1,2, \dots, m\} \quad (28)$$

395 Four influential criteria were considered in order to rank the models: statistical parameter (R^2,
396 RMSE, MAE, NSE), model run time, spatial change factor (SCF), and temporal change factor
397 (TCF), where SCF shows the effect of changing stations on the analysis, and TCF indicates
398 the impact of changing time scale of the used data. The KGE coefficient was taken into account
399 in TOPSIS for calculating the last two factors. The utilization of SCF and TCF in the TOPSIS
400 method is because of the insufficiency of statistical parameters in providing a satisfactory
401 prioritization of hybrid models and the inability to generalize results to other studies with
402 varying conditions. These two factors are described below:

403 $$\mathrm{SCF} = \frac{|\mathrm{KGE_V} - \mathrm{KGE_G}|}{\mathrm{KGE_{max}}} \quad (29)$$

404 Where $\mathrm{KGE_V}$ and $\mathrm{KGE_G}$ correspond to the Vanyar and Gotvand datasets, respectively. $\mathrm{KGE_{max}}$
405 represents the highest KGE value achieved by each algorithm, taking into account all
406 associated conditions (Considering and ignoring discharge, and monthly and annual time
407 scales) in both hydrometric stations.

408 $$\mathrm{TCF} = \frac{|\mathrm{KGE_m} - \mathrm{KGE_a}|}{\mathrm{KGE_{max}}} \quad (30)$$

409 $\mathrm{KGE_m}$ and $\mathrm{KGE_a}$ are utilized to assess the monthly and annual datasets, respectively. $\mathrm{KGE_{max}}$
410 denotes the utmost KGE value attained by each algorithm during the analysis. This
411 comprehensive analysis encompasses all relevant conditions.

412 **3. Results and discussion**

413 Based on accuracy-error and economic criteria time, the standalone ANFIS approach and its
414 hybrid models were assessed in estimating monthly and annual TDS for both Gotvand and

Vanyar stations. The findings of the analysis are presented in Tables 2 and 3 for the Gotvand and Vanyar Stations, respectively. Furthermore, Figures 4 and 5 show the scatter plots of the observed versus simulated data for the ANFIS and the strongest hybrid method of each station.

3.1. Analysis of monthly TDS

3.1.1. Considering discharge

Based on the statistical assessment criteria of the Gotvand Station given in Table 2, it was determined that the ANFIS approach had the weakest performance compared to the other approaches (train: R^2 = 0.93, KGE = 0.92, RMSE = 0.041, MAE = 0.016, NSE = 0.92) and test(R^2 = 0.76, KGE = 0.77, RMSE = 0.077, MAE = 0.04, NSE = 0.7). This might be because it failed to find the global optimum due to getting confined to local optima (Azad et al., 2019). By applying the meta-heuristic algorithms, the output of ANFIS was boosted to reasonable values. The ANFIS-PSO method had superior results compared to other metaheuristic methods (R^2 = 0.98, KGE = 0.99, RMSE = 0.019, MAE = 0.0035, and NS = 0.98) were derived for train, and (R^2 = 0.97, KGE = 0.98, RMSE = 0.022, MAE = 0.0034, and NS = 0.98) for test. In terms of execution time, the ANFIS technique was the quickest method with 350 seconds, whereas the ANFIS-PSO took 3300 seconds.

In respect of the statistical assessment criteria reported in Table 3, the analysis conducted on the Vanyar Station dataset using the standalone ANFIS method had poor results because of the significant difference between the test and train outputs: in train (R^2 = 0.92, KGE= 0.99, RMSE = 0.044, MAE = 0.013, NSE = 0.917) and in test (R^2 =0.61, KGE= 0.66, RMSE = 0.13, MAE = 0.106, NSE= 0.38). The results of all optimized models were quite close to one another, with only minor differences in performance criteria values. By combining ANFIS with meta-heuristic methods, it was found that all the applied algorithms improved the performance of ANFIS in the evaluation of the monthly TDS dataset of the Vanyar station with applying

discharge. Analysis revealed that ANFIS-PSO has the highest potential: in train (R^2 = 0.98, KGE = 0.98, RMSE = 0.018, MAE =0.001, NSE = 0.983) and in test (R^2 = 0.97, KGE = 0.98, RMSE = 0.023, MAE = 0.006, NSE = 0.96). An investigation of the running times of the models revealed that the ANFIS method was the fastest, taking 250 seconds, while the ANFIS-DE was the slowest, with 3100 seconds of execution time.

From the results obtained for the methods used in the assessment of the Gotvand Station, it can be inferred that when there were sufficient and appropriate amounts of data and parameters for training the ANFIS-PSO was superior. Previous investigations have also underscored the crucial role played by both the quantity and quality of data employed in the modeling process (e.g., Alwosheel et al., 2018). This was also true for the analysis of Vanyar data. Compared to the Vanyar Station analysis, however, the models of the Gotvand Station required more execution time, but they delivered more precise results. This is probably due to the nature of the data. One explanation for this can be that dataset of the Gotvand Station is larger than the Vanyar Station. Additionally, it is possible that the monthly data from Vanyar are of lower quality than the data from the Gotvand Station as a result of challenges like discharge measurement error (Potash and Steinschneider, 2022) or a stronger influence of TDS from inputs other than discharge and ions employed in this part of the calculations. Moreover, the low accuracy of the Vanyar Station calculation findings compared to the Gotvand Station may be because of its higher TDS fluctuations and a broader spectrum of maximum and minimum TDS values (Kouadri et al., 2021). The longer execution time at the Gotvand Station might be attributed to its larger period of data series.

3.1.2. Ignoring discharge

In the Gotvand Station, eliminating the discharge factor improved the modeling results in both ANFIS and hybrid methods. The results of the ANFIS analysis were (R^2 = 0.98, KGE = 0.98, RMSE = 0.0198, MAE = 0.0017, and NS = 0.97 in train), and (R^2 = 0.82, KGE = 0.87, RMSE

464 = 0.053, MAE = 0.003, and NSE = 0.786 in test). The ANFIS model had the worst performance
465 among all the approaches that were investigated during the analysis. In this Station also,
466 ANFIS-GA was the most appropriate approach among the others, after eliminating the
467 discharge factor from the input parameters. The result of ANFIS-GA was as follows: in train
468 (R^2 = 0.98, KGE = 0.99, RMSE = 0.0163, MAE = 0.001, NSE = 0.98) and test (R^2 = 0.98,
469 KGE = 0.99, RMSE = 0.016, MAE = 0.0002, NSE = 0.98). It is noteworthy that the ANFIS
470 technique required approximately 220 seconds to execute, whereas the ANFIS-DE strategy
471 lasted 2800 seconds. In the analysis of Gotvand Station monthly TDS when ignoring discharge,
472 more accurate outputs and fewer errors were obtained as compared to considering discharge as
473 an input parameter. The poor quality of the discharge data may be the reason for this result.
474 Generally, the measurement of the discharge resulting from non-point pollution sources in Iran
475 is accompanied by numerous challenges, (Loucks and van Beek, 2017; Pan et al., 2016; Parra
476 et al., 2016) and this may have affected the quality of discharge data. The high variability of
477 discharge - TDS ratio compared to variations in the other utilized parameters in relation to TDS
478 might be the reason for the superior result of analysis without discharge than when using
479 discharge as an input (Kouadri et al., 2021).

480 In terms of the Vanyar Station, It is clear from Table 3 that the inclusion of discharge lowers
481 the accuracy of models, which improved after this factor was removed. The statistical
482 parameters for ANFIS analysis were (R^2 = 0.99, KGE = 0.99, RMSE = 0.007, MAE = 0.0015,
483 NSE = 0.99 for train) and (R^2 = 0.69, KGE= 0.7, RMSE = 0.091, MAE = 0.052, NSE = 0.38
484 for test). These values were significantly lower than the ones obtained by hybrid models. The
485 obtained information from optimizations indicates that ANFIS-PSO has the most acceptable
486 performance in train (R^2 = 0.984, KGE = 0.99, RMSE = 0.0181, MAE = 0.002, NSE = 0.98)
487 and test (R^2 = 0.98, KGE=0.94, RMSE = 0.021, MAE = 0.0053, NSE = 0.98) among the
488 implemented models. When comparing how long each model took to run, it was clear that the

ANFIS approach was the quickest, with 200 seconds. The ANFIS-DE approach required the most time to run, with 2600 seconds.

The analysis represented that the assessment of Vanyar Station's monthly TDS when omitting discharge required a shorter running time and had more reliable results than the monthly models including discharge. With disregarding the discharge as input (Similar to the condition where the discharge was factored into the computations), the applied methodologies performed better in the calculations of monthly data of the Gotvand Station in comparison to the Vanyar Station (the possible reasons are mentioned in section 3.1.1). Furthermore, the analysis of the monthly data of the Gotvand Station in the condition of ignoring discharge required more execution time compared to the monthly data of the Vanyar Station regardless of discharge.

Based on the results, the ANFIS-PSO method performed better than other methods in all cases of monthly data analysis (except in the monthly models of Gotvand Station disregarding discharge in which the ANFIS-GA model was the most appropriate approach). This output yields a conclusion that the performance of employed models varies temporally and spatially; and can also be influenced by a variety of factors, including the complexity of the problem, the size of the search space, the number of variables and constraints, and the quality and quantity of the data available (Djebedjian et al., 2021; Kim et al., 2016).

3.2. Analysis of annual TDS

3.2.1. Considering discharge

The annual analysis of the Gotvand Station on the Karun River revealed that ANFIS has poor performance in predicting total dissolved solids in train (R^2 = 0.99, KGE=0.99, RMSE = 0.0127, MAE = 0.025, NSE = 0.99) and test (R^2 = 0.69, KGE= 0.77, RMSE = 0.115, MAE = 0.048, NSE = 0.689). The ANFIS-ACOR method demonstrated perfect results in train (R^2 = 0.97, KGE= 0.961, RMSE = 0.027, MAE = 0.023, NSE = 0.97) and test (R^2 = 0.94, KGE= 0.9,

RMSE = 0.0503, MAE = 0.026, NSE = 0.93). Based on a comparison of execution durations, the ANFIS method took over 130 seconds.

The results indicated that, while the annual models of the Gotvand Station required less time to run, they were less accurate than its monthly models. The findings of the annual data analysis at the Gotvand Station were superior to those at Vanyar Station in both the case with and without discharge. Researchers often use average monthly data to estimate annual data when annual data is not available (e.g., Yu et al., 2014). Thus, this study utilized average monthly water quality data to create annual data for both of the studied rivers. The aim was to conduct the analysis in cases that the comprehensive data indicating temporal changes are not available. The results indicated that the models utilized for processing the annual data demonstrated satisfactory performance, though not as well as the monthly data.

Through an annual dataset analysis of the Vanyar Station using the discharge variable, it was discovered that ANFIS was the least effective when compared to the other models. There was a significant gap between the test and train results of this approach: (R^2 = 0.99, KGE= 0.99, RMSE = 0.0035, MAE = 0.0078, and NSE = 0.99 for train), and (R^2 = 0.66, KGE= 0.77, MSE = 0.128, MAE = 0.09, and NSE = 0.569 for test). After applying meta-heuristic algorithms, the statistical factors yielded by ANFIS evolved to appropriate values. ANFIS-ACOR Method (R^2 = 0.97, KGE=0.98, RMSE = 0.02, MAE = 0.007, NSE = 0.97 in train) and (R^2 = 0.94, KGE = 0.9, RMSE = 0.035, MAE = 0.08, NSE = 0.93 for test), shows the best results among the other three approaches. ANFIS-DE, ANFIS-PSO, and ANFIS-GA ranked as the second, third, and fourth best, respectively. A comparison of the execution times of the models demonstrated that the ANFIS method was the fastest with 95 seconds whereas the ANFIS-DE method demanded the most amount of time 3000 seconds to tune the ANFIS parameters.

The annual models of the Gotvand Station considering discharge were more reliable than the annual models of the Vanyar Station considering discharge., but required more time to execute.

Moreover, The results of calculations of the monthly dataset of Vanyar encompassing the discharge were more accurate and had less error compared to annual models of Vanyar considering discharge. In our study, except for the following cases, the modeling of monthly data yielded better results: Gotvand Station without discharge modeled with ANFIS-ACOR, Vanyar Station considering discharge analyzed with ANFS, Vanyar Station considering discharge assessed with ANFIS-ACOR, and Vanyar Station ignoring discharge evaluated with ANFIS-ACOR. However, there have been some studies showing that the outputs obtained from modeling monthly data do not significantly differ from those obtained from modeling annual data (Yu et al., 2014).

3.2.2. Ignoring discharge

Similar to the analysis of the monthly dataset of Gotvand, in the evaluation of annual data of the Vanyar Station in this part of the calculations also the discharge factor's removal led to better results for both the ANFIS and hybrid methods (Table 2). The ANFIS findings were (R^2=0.97, KGE = 0.985, RMSE=0.033, MAE= 0.0003 and NSE=0.976 for train) and (R^2= 0.8, KGE = 0.81, RMSE= 0.082, MAE= 0.001, NS= 0.783 for test). Among the algorithms, the ANFIS-ACOR method with (R^2=0.97, KGE= 0.96, RMSE= 0.002, MAE=0.022, NSE= 0.97 for train) and (R^2= 0.97, KGE = 0.92, RMSE= 0.032, MAE=0.005 and NS=0.971 for test) was the best method. It took 95 seconds for the ANFIS method to execute.

Annual models of Gotvand without discharge were less accurate and quicker than the monthly models. According to the results, the annual models of the Gotvand Station ignoring discharge models were faster, but their accuracy was lower than this station’s modeling of monthly data without discharge. In addition, Gotvand’s annual models ignoring discharge models were less precise and required more time than Vanyar’s annual analysis ignoring discharge.

In the next step, The accuracy of assessments was raised by leaving the discharge factor out of the input parameters for the Aji Chay River's annual water quality. ANFIS model yielded statistical parameters with values of R^2=0.99, KGE= 0.99, RMSE = 0.0005, MAE = 0.0001, NS = 0.9 in train and R^2 = 0.8, KGE = 0.793, RMSE = 0.086, MAE = 0.044, NS = 0.677 in test. When compared to other models, the optimizations revealed that ANFIS-ACOR has the highest performance in train R^2 = 0.973, KGE= 0.96 RMSE = 0.04, MAE = 0.0059, NSE = 0.971, and test R^2 = 0.96, KGE= 0.91, RMSE = 0.04, MAE = 0.007, NS = 0.962.

According to the outputs of the total analyses, in the assessment of monthly data, ANFIS-GA and ANFIS-PSO approaches were more accurate, but with annual data evaluations, the ANFIS-ACOR resulted in higher performance. In the case of omitting the discharge, the assessment of annual data of Vanyar Station yielded weaker results than the annual data modeling of the Gotvand Station; also the annual analysis of Vanyar resulted in poor outputs than the monthly assessment of this station. Furthermore, in the circumstance including the discharge, the annual models of Vanyar were weaker than this station's annual model without discharge.

The box plot of errors for the approach used in this study is shown in Fig. 6. It is evident that hybrid models had a narrower error change range than ANFIS, and all models outperformed ANFIS. The ANFIS-ACOR represented the most superior performance, while the ANFIS-DE demonstrated the least efficiency. The reason for the disparity between the information of this chart with the findings of TOPSIS is that TOPSIS took into account the spatial-temporal sensitivity and execution time factors.

3.3. Decision making

It is challenging to make decisions and choose between poor and superior models due to the abundance of available models and the variety of modeling situations. As a result, this research

584 employed the TOPSIS technique for multi-criteria decision-making. In the previous studies,
585 the ranking of models was mostly done focusing on the statistical indicators (R^2, RMSE, etc.)
586 (e.g., Kadkhodazadeh and Farzin, 2022). Our study used additional factors (spatial and
587 temporal change factors) to make the results more realistic and informative to address practical
588 water management needs, as opposed to solely statistical considerations. This strengthens the
589 practical usefulness and actionable value of our proposed monitoring station selection process.

590 The matrix of experts employed in TOPSIS includes an impact coefficient of 0.2 assigned for
591 the matrix elements. It should be emphasized that Kling Gupta's coefficient served as the
592 ranking criterion for assessing the effects of applying and removing discharge, the type of data
593 interval for the station, and the effects of river stations.

594 An analysis of the results of different methods showed that the ANFIS-ACOR algorithm has
595 superior performance in predicting water quality. Likewise, Azad et al. (2019) demonstrated
596 that predicting precipitation, as a crucial parameter influencing the water quantity, had the best
597 performance using the same model (Kitan & Nang, 2020; Lintern et al., 2018). Thus, this model
598 can be considered reliable for both quantitative and qualitative evaluations. On the other hand,
599 the TOPSIS showed that the ANFIS-DE was the weakest model in this study, considering the
600 whole criteria (spatial and temporal variation factors, statistical factors, and model execution
601 time). However, a closer examination of the algorithms used in the quantitative and qualitative
602 research on water resources reveals that, in most cases, the performance of hybrid models in
603 predicting water quality has been suitable. This makes selecting the most effective model for
604 problems that use hybrid models challenging (Vazquezl et al., 2021).

605 **3.4. Suggestions for future studies**

606 The methods used in this research can contribute to a proper management of ecological
607 consequences, reduce the high costs of water quality data collection, and save time in water

quality assessment (Zhu et al., 2022). However, future studies can focus on the following recommendations:

- Explore the performance of ANFIS and its hybrid models (ANFIS-PSO, ANFIS-DE, ANFIS-ACOR, and ANFIS-GA) models in modeling TDS in other rivers and hydrometric stations to assess the generalizability of the results.
- Investigate the use of alternative metaheuristic algorithms to further enhance the performance of ANFIS models.
- Consider the impact of other factors, such as temperature, rainfall, and land use, on TDS modeling to provide a more comprehensive understanding of the factors that affect water quality.
- Evaluate the effectiveness of other evaluation factors to rank the models when using the TOPSIS method.
- Investigate the effect of different temporal and spatial scales on the performance of ANFIS models in modeling TDS.
- Compare the performance of ANFIS hybrid models with the hybrid of other machine learning techniques, such as artificial neural networks and support vector machines.

4. Conclusions

The performance of the ANFIS approach in the modeling of TDS was assessed and improved by the utilization of four different metaheuristic algorithms, including PSO, DE, ACOR, and GA. This study focused on two key hydrometric stations, Vanyar in the Aji Chay River and Gotvand on the Karun River. The linear regression analysis was used for parameter selection. In addition, results revealed that metaheuristic techniques employed in this study enhanced ANFIS performance in all models. The running time of models assessed using ANFIS was

significantly less than that required for hybrid models. The monthly data calculations produced better findings than the annual data in the majority of the models. However, the results of the annual data were likewise reliable, and it appears that they can be utilized in analyses when monthly data is unavailable. In terms of running time, the annual models were faster than the monthly models. Investigation of the effect of discharge on modeling showed that removing the discharge from analysis improved the results in both accuracy and running time. Although the Gotvand Station dataset required more time to process than the Vanyar Station dataset, More precise outcomes were obtained for the Gotvand Station than Vanyar. In the final step, the ranks of models were obtained using the TOPSIS approach. In order to increase the reliability of TOPSIS performance, the temporal and special change factors were used in the computations. The comprehensive findings of the study revealed the ANFIS-ACOR model's superiority, while concurrently exposing the ANFIS-DE model's weak. ness in total. Consequently, the findings of this research support the hypothesis that hybrid ANFIS models can be used to evaluate the TDS with high accuracy, which is of vital importance in water quality management**.**

5. References

Abd El-Mageed, A. M. G., Enany, T. A., Goher, M. E., Hassouna, M. E. M. 2022. Forecasting water quality parameters in Wadi El Rayan Upper Lake, Fayoum, Egypt using adaptive neuro-fuzzy inference system. Egyptian Journal of Aquatic Research, 481. https://doi.org/10.1016/j.ejar.2021.10.001

Aghel, B., Rezaei, A., Mohadesi, M. 2019. Modeling and prediction of water quality parameters using a hybrid particle swarm optimization–neural fuzzy approach. International Journal of Environmental Science and Technology, 168. https://doi.org/10.1007/s13762-018-1896-3

Ahmad, H. A. 2012. The best candidates method for solving optimization problems. Journal of Computer Science, 85. https://doi.org/10.3844/jcssp.2012.711.715

Al-Mukhtar, M., Al-Yaseen, F. 2019. Modeling water quality parameters using data-driven models, a case study Abu-Ziriq marsh in south of Iraq. Hydrology, 61. https://doi.org/10.3390/hydrology6010021

Al-qaness, M. A. A., Fan, H., Ewees, A. A., Yousri, D., Abd Elaziz, M. 2021. Improved ANFIS model for forecasting Wuhan City Air Quality and analysis COVID-19 lockdown impacts on air quality. Environmental Research, 194October 2020, 110607. https://doi.org/10.1016/j.envres.2020.110607

Alwosheel, A., van Cranenburgh, S., & Chorus, C. G. (2018). Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. Journal of Choice Modeling, 28, 167–182. https://doi.org/10.1016/J.JOCM.2018.07.002

Andaryani, S., Nourani, V., Haghighi, A. T., Keesstra, S. 2021. Integration of hard and soft supervised machine learning for flood susceptibility mapping. Journal of Environmental Management, 291. https://doi.org/10.1016/j.jenvman.2021.112731

Antanasijević, D., Pocajt, V., Perić-Grujić, A., Ristić, M. 2014. Modeling of dissolved oxygen in the danube river using artificial neural networks and Monte carlo simulation uncertainty analysis. Journal of Hydrology, 519PB. https://doi.org/10.1016/j.jhydrol.2014.10.009

Asadollah, S. B. H. S., Sharafati, A., Motta, D., & Yaseen, Z. M. (2021). River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. Journal of Environmental Chemical Engineering, 9(1). https://doi.org/10.1016/j.jece.2020.104599

677 Azad, A., Karami, H., Farzin, S., Mousavi, S. F., Kisi, O. 2019. Modeling river water quality
678 parameters using modified adaptive neuro fuzzy inference system. Water Science and
679 Engineering, 121. https://doi.org/10.1016/j.wse.2018.11.001

680 Azad, A., Karami, H., Farzin, S., Saeedian, A., Kashi, H., Sayyahi, F. 2018. Prediction of water
681 quality parameters using ANFIS optimized by intelligence algorithms case study: Gorganrood
682 river. KSCE Journal of Civil Engineering, 227, 2206–2213. https://doi.org/10.1007/s12205-
683 017-1703-6

684 Banadkooki, F. B., Ehteram, M., Panahi, F., Sh. Sammen, S., Othman, F. B., & EL-Shafie, A.
685 2020. Estimation of total dissolved solids TDS using new hybrid machine learning models.
686 Journal of Hydrology, 587. https://doi.org/10.1016/j.jhydrol.2020.124989

687 Deng, C., Liu, L., Li, H., Peng, D., Wu, Y., Xia, H., Zhang, Z., Zhu, Q. 2021. A data-driven
688 framework for spatiotemporal characteristics, complexity dynamics, and environmental risk
689 evaluation of river water quality. Science of the Total Environment, 785.
690 https://doi.org/10.1016/j.scitotenv.2021.147134

691 Chau, K. 2006. A review on integration of artificial intelligence into water quality modeling.
692 A Review on Integration of Artificial Intelligence into Water Quality Modeling,
693 52doi:10.1016/j.envsoft.2007.06.008, 726–733.
694 https://doi.org/10.1016/j.marpolbul.2006.04.003

695 Chellaiah, E. R., Ravi, P., Uthandakalaipandian, R. 2021. Isolation and identification of high
696 fluoride resistant bacteria from water samples of Dindigul district, Tamil Nadu, South India.
697 Current Research in Microbial Sciences, 2, 100038.
698 https://doi.org/10.1016/J.CRMICR.2021.100038

699 Dai, X., Long, S., Zhang, Z., Gong, D. 2019. Mobile robot path planning based on ant colony
700 algorithm with a∗ heuristic method. Frontiers in Neurorobotics, 13.
701 https://doi.org/10.3389/fnbot.2019.00015

702 Diagomanolin, V., Farhang, M., Ghazi-Khansari, M., & Jafarzadeh, N. (2004). Heavy metals
703 (Ni, Cr, Cu) in the Karoon waterway river, Iran. Toxicology Letters, 151(1).
704 https://doi.org/10.1016/j.toxlet.2004.02.018

705

706 Djebedjian, B., Abdel-Gawad, H. A. A., & Ezzeldin, R. M. (2021). Global performance of
707 metaheuristic optimization tools for water distribution networks. Ain Shams Engineering
708 Journal, 12(1), 223–239. https://doi.org/10.1016/J.ASEJ.2020.07.012

709 Dorigo, M., Blum, C. 2005. Ant colony optimization theory: A survey. Theoretical Computer
710 Science, 3442–3. https://doi.org/10.1016/j.tcs.2005.05.020

711 Ebtehaj, I., & Bonakdari, H. 2017. Design of a fuzzy differential evolution algorithm to predict
712 non-deposition sediment transport. Applied Water Science, 78.
713 https://doi.org/10.1007/s13201-017-0562-0

714 Fakouri, B., Mazaheri, M., Samani, J. M. 2019. Management scenarios methodology for
715 salinity control in rivers case study: karoon river, Iran. Journal of Water Supply: Research and
716 Technology - AQUA, 681. https://doi.org/10.2166/aqua.2018.056

717 Fu, H., Li, H. 2021. Research on water resources dispatch model based on improved genetic
718 algorithm – water resources dispatch model. Water Science and Technology: Water Supply,
719 213. https://doi.org/10.2166/ws.2020.344

720 Gao, H., Li, C., Sun, B. 2018. The impact of changed river discharge on water quality
721 deterioration in a prairie lake revealed by the sedimentary evidence. Water Science and
722 Technology: Water Supply, 181. https://doi.org/10.2166/ws.2017.120

723 Gavili, S., Sanikhani, H., Kisi, O., Mahmoudi, M. H. 2018. Evaluation of several soft
724 computing methods in monthly evapotranspiration modeling. Meteorological Applications,
725 251. https://doi.org/10.1002/met.1676

726 Gawande, S. M., Sarode, D. D. 2021. Water Pollution and Its Prevention Through Development
727 of Low Cost Wastewater Treatment System. RILEM Bookseries, 29.
728 https://doi.org/10.1007/978-3-030-51485-3_35

729 Ghfolamreza, A., Afshin, M. D., Shiva, H. A., & Nasrin, R. (2016). Application of artificial
730 neural networks to predict total dissolved solids in the river Zayanderud, Iran. Environmental
731 Engineering Research, 21(4), 333–340. https://doi.org/10.4491/eer.2015.096

732 Golshan, M., Dastoorpour, M., Birgani, Y. T. 2020. Fuzzy environmental monitoring for the
733 quality assessment: Detailed feasibility study for the Karun River basin, Iran. Groundwater for
734 Sustainable Development, 10, 100324. https://doi.org/10.1016/J.GSD.2019.100324

735 Gutiérrez, F., Lizaga, I. 2016. Sinkholes, collapse structures and large landslides in an active
736 salt dome submerged by a reservoir: The unique case of the Ambal ridge in the Karun River,
737 Zagros Mountains, Iran. Geomorphology, 254.
738 https://doi.org/10.1016/j.geomorph.2015.11.020

739 Hassan, O. F., Jamal, A., Abdel-Khalek, S. 2020. Genetic algorithm and numerical methods
740 for solving linear and nonlinear system of equations: A comparative study. Journal of
741 Intelligent and Fuzzy Systems, 383. https://doi.org/10.3233/JIFS-179572

742 Hossain, Md. F. 2019. Water. Sustainable Design and Build, 301–418.
743 https://doi.org/10.1016/B978-0-12-816722-9.00006-9

744 Hossein, M., Moghaddam, R. 2006. Geomorphologic hazards for Vanyar Dam with emphasis
745 on the reactivation of Tabriz fault , northwest Iran. 339, 1–5.

746 Hu, J. H., Tsai, W. P., Cheng, S. T., Chang, F. J. 2020. Explore the relationship between fish
747 community and environmental factors by machine learning techniques. Environmental
748 Research, 184. https://doi.org/10.1016/j.envres.2020.109262

749 Jalali, L., Zarei, M., & Gutiérrez, F. (2019). Salinization of reservoirs in regions with exposed
750 evaporites. The unique case of Upper Gotvand Dam, Iran. Water Research, 157, 587–599.
751 https://doi.org/10.1016/J.WATRES.2019.04.015

752 Jalalkamali, A. 2015. Using of hybrid fuzzy models to predict spatiotemporal groundwater
753 quality parameters. Earth Science Informatics, 84. https://doi.org/10.1007/s12145-015-0222-6

754 Jamei, M., Ahmadianfar, I., Chu, X., Yaseen, Z. M. 2020. Prediction of surface water total
755 dissolved solids using hybridized wavelet-multigene genetic programming: New approach.
756 Journal of Hydrology, 589. https://doi.org/10.1016/j.jhydrol.2020.125335

757 Jannatkhah, M., Akbari, A., Bagheri Basmanji, A., Rahmani, E., Peter Cox, J. 2021. Estimation
758 of Monthly Total Dissolved Solids Using ANN and LS-SVM Techniques in the Aji Chay
759 River, Iran. Journal of Civil Engineering and Construction, 101.
760 https://doi.org/10.32732/jcec.2021.10.1.1

761 Kabolizadeh, M., Rangzan, K., Zareie, S., Rashidian, M., Delfan, H. 2022. Evaluating quality
762 of surface water resources by ANN and ANFIS networks using Sentinel-2 satellite data. Earth
763 Science Informatics, 151. https://doi.org/10.1007/s12145-021-00741-z

764 Kadkhodazadeh, M., & Farzin, S. (2022). Introducing a Novel Hybrid Machine Learning
765 Model and Developing its Performance in Estimating Water Quality Parameters. Water
766 Resources Management, 36(10), 3901–3927. https://doi.org/10.1007/s11269-022-03238-6

767 Karabašević, D., Stanujkić, D., Zavadskas, E. K., Stanimirović, P., Popović, G., Predić, B.,
768 Ulutaş, A. 2020. A novel extension of the TOPSIS method adapted for the use of single-valued

neutrosophic sets and hamming distance for e-commerce development strategies selection. Symmetry, 128. https://doi.org/10.3390/SYM12081263

Karamouz, M., Kerachian, R., Akhbari, M., Hafez, B. 2009. Design of river water quality monitoring networks: A case study. Environmental Modeling and Assessment, 146. https://doi.org/10.1007/s10666-008-9172-4

Karunanithi, N., Grenney, W. J., Whitley, D., Bovee, K. 1994. Neural Networks for River Flow Prediction. Journal of Computing in Civil Engineering, 82. https://doi.org/10.1061/asce0887-380119948:2201

Khalil, B., Ouarda, T. B. M. J., St-Hilaire, A. 2011. Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. Journal of Hydrology, 4053–4. https://doi.org/10.1016/j.jhydrol.2011.05.024

Khataee, A. R., Vafaei, F., Jannatkhah, M. 2013. Biosorption of three textile dyes from contaminated water by filamentous green algal Spirogyra sp.: Kinetic, isotherm and thermodynamic studies. International Biodeterioration and Biodegradation, 83. https://doi.org/10.1016/j.ibiod.2013.04.004

Khoi, D.N., Quan, N.T., Linh, D.Q., Nhi, P.T.T., Thuy, N.T.D. 2022. Using Machine Learning Models for Predicting the Water Quality Index in the La Buong River, Vietnam. Water, 141552.https://doi.org/10.3390/w14101552

Kim, J. H., Lee, H. M., Jung, D., & Sadollah, A. (2016). Performance measures of metaheuristic algorithms. Advances in Intelligent Systems and Computing, 382. https://doi.org/10.1007/978-3-662-47926-1_2

Kisi, O., Latifoğlu, L., Latifoğlu, F. 2014. Investigation of empirical mode decomposition in forecasting of hydrological time series. Water Resources Management, 2812. https://doi.org/10.1007/s11269-014-0726-8

793 Kitan, Y. A., & Nang, S. C. S. (2020). Influence of seasonal rainfall to the water quality of slim
794 river lake in Perak, Malaysia. *Plant Archives*, *20*(1).

795 Kouadri, S., Elbeltagi, A., Islam, A. R. M. T., & Kateb, S. (2021). Performance of machine
796 learning methods in predicting water quality index based on irregular data set: application on
797 Illizi region (Algerian southeast). Applied Water Science, 11(12).
798 https://doi.org/10.1007/s13201-021-01528-9

799 Lintern, A., Webb, J. A., Ryu, D., Liu, S., Bende-Michl, U., Waters, D., Leahy, P., Wilson, P.,
800 & Western, A. W. (2018). Key factors influencing differences in stream water quality across
801 space. *Wiley Interdisciplinary Reviews: Water*, *5*(1). https://doi.org/10.1002/WAT2.1260

802 Liu, G., Ye, J., Chen, Y., Yang, X., Gu, Y. 2022. Analysis of water pollution causes and control
803 countermeasures in Liaohe estuary via support vector machine particle swarm optimization
804 under deep learning. CMES - Computer Modeling in Engineering and Sciences, 1301.
805 https://doi.org/10.32604/cmes.2022.016224

806 Loucks, D. P., & van Beek, E. (2017). Water resource systems planning and management: An
807 introduction to methods, models, and applications. In Water Resource Systems Planning and
808 Management: An Introduction to Methods, Models, and Applications.
809 https://doi.org/10.1007/978-3-319-44234-1

810 Varol, M., Karakaya, G., & Alpaslan, K. (2022). Water quality assessment of the Karasu River
811 (Turkey) using various indices, multivariate statistics and APCS-MLR model. Chemosphere,
812 308(136415).

813 Mahmood, T., Ali, Z. 2021. Entropy measure and TOPSIS method based on correlation
814 coefficient using complex q-rung orthopair fuzzy information and its application to multi-
815 attribute decision making. Soft Computing, 252. https://doi.org/10.1007/s00500-020-05218-7

McCuen, R. H., Knight, Z., Cutter, A. G. 2006. Evaluation of the Nash–Sutcliffe Efficiency Index. Journal of Hydrologic Engineering, 116. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:6(59

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, 328. https://doi.org/10.1007/s00477-018-1554-9

Mullen, R. J., Monekosso, D., Barman, S., Remagnino, P. 2009. A review of ant algorithms. In Expert Systems with Applications Vol. 36, Issue 6. https://doi.org/10.1016/j.eswa.2009.01.020

Najafabadipour, A., Kamali, G., Nezamabadi-Pour, H. 2022. Application of Artificial Intelligence Techniques for the Determination of Groundwater Level Using Spatio-Temporal Parameters. ACS Omega, 712. https://doi.org/10.1021/acsomega.2c00536

Palani, S., Liong, S. Y., Tkalich, P. 2008. An ANN application for water quality forecasting. Marine Pollution Bulletin, 569. https://doi.org/10.1016/j.marpolbul.2008.05.021

Pan, F., Wang, C., Xi, X. 2016. Constructing river stage-discharge rating curves using remotely sensed river cross-sectional inundation areas and river bathymetry. Journal of Hydrology, 540. https://doi.org/10.1016/j.jhydrol.2016.06.024

Parra, B. G., Rojas, L. E. P., Barrios, M., Estrada, J. C. M. 2016. Uncertainty of discharge estimation in high-grade Andean streams. Flow Measurement and Instrumentation, 48, 42–50. https://doi.org/10.1016/J.FLOWMEASINST.2016.02.005

Potash, E., & Steinschneider, S. (2022). A Bayesian Approach to Recreational Water Quality Model Validation and Comparison in the Presence of Measurement Error. Water Resources Research, 58(1). https://doi.org/10.1029/2021WR031115

840 Radmanesh, F., Zarei, H., Salari, M. 2013. Water Quality Index and Suitability of Water of
841 Gotvand Basin at District Khuzestan, Iran. International Journal of Agronomy and Plant
842 Production, 44.

843 Rezaee, A., Bozorg-Haddad, O., Chu, X. 2021. Reallocation of water resources according to
844 social, economic, and environmental parameters. In Scientific Reports Vol. 11, Issue 1.
845 https://doi.org/10.1038/s41598-021-96680-2

846 Eskandari, S., Ali Mahmoudi Sarab, S. 2022. Mapping land cover and forest density in Zagros
847 forests of Khuzestan province in Iran: A study based on Sentinel-2, Google Earth and field
848 data. Ecological Informatics, 70. https://doi.org/10.1016/j.ecoinf.2022.101727

849 Salmani, M. H., Salmani Jajaei, E. 2016. Forecasting models for flow and total dissolved solids
850 in Karoun river-Iran. Journal of Hydrology, 535. https://doi.org/10.1016/j.jhydrol.2016.01.085

851 Samanataray, S., Sahoo, A. 2021. A Comparative Study on Prediction of Monthly Streamflow
852 Using Hybrid ANFIS-PSO Approaches. KSCE Journal of Civil Engineering, 2510, 4032–
853 4043. https://doi.org/10.1007/S12205-021-2223-Y

854 Seiler, L. M. N., Fernandes, E. H. L., & Siegle, E. 2020. Effect of wind and river discharge on
855 water quality indicators of a coastal lagoon. Regional Studies in Marine Science, 40, 101513.
856 https://doi.org/10.1016/J.RSMA.2020.101513

857 Shah, M. I., Javed, M. F., Alqahtani, A., Aldrees, A. 2021. Environmental assessment-based
858 surface water quality prediction using hyper-parameter optimized machine learning models
859 based on consistent big data. Process Safety and Environmental Protection, 151.
860 https://doi.org/10.1016/j.psep.2021.05.026

861 Sharafkhani, R., Khanjani, N., Bakhtiari, B., Jahani, Y., Sadegh Tabrizi, J. 2018. Physiological
862 Equivalent Temperature Index and mortality in Tabriz The northwest of Iran. Journal of
863 Thermal Biology, 71. https://doi.org/10.1016/j.jtherbio.2017.11.012

Sun, K., Rajabtabar, M., Samadi, S. Z., Rezaie-Balf, M., Ghaemi, A., Band, S. S., Mosavi, A. 2021. An integrated machine learning, noise suppression, and population-based algorithm to improve total dissolved solids prediction. Engineering Applications of Computational Fluid Mechanics, 151. https://doi.org/10.1080/19942060.2020.1861987

Tutmez, B., Hatipoglu, Z., Kaymak, U. 2006. Modeling electrical conductivity of groundwater using an adaptive neuro-fuzzy inference system. Computers and Geosciences, 324. https://doi.org/10.1016/j.cageo.2005.07.003

Vazquezl, M. Y. L., Peñafiel, L. A. B., Muñoz, S. X. S., & Martinez, M. A. Q. (2021). A Framework for Selecting Machine Learning Models Using TOPSIS. Advances in Intelligent Systems and Computing, 1213 AISC. https://doi.org/10.1007/978-3-030-51328-3_18

Wang, X., Zhao, S. 2013. wan, 2013. https://doi.org/10.1155/2013/419372

, M. Y. L., Peñafiel, L. A. B., Muñoz, S. X. S., & Martinez, M. A. Q. (2021). A Framework for Selecting Machine Learning Models Using TOPSIS. *Advances in Intelligent Systems and Computing*, *1213 AISC*. https://doi.org/10.1007/978-3-030-51328-3_18

Ying, L. C., Pan, M. C. 2008. Using adaptive network based fuzzy inference system to forecast regional electricity loads. Energy Conversion and Management, 492. https://doi.org/10.1016/j.enconman.2007.06.015

Yoosefdoost, I., Khashei-Siuki, A., Tabari, H., Mohammadrezapour, O. 2022. Runoff Simulation Under Future Climate Change Conditions: Performance Comparison of Data-Mining Algorithms and Conceptual Models. Water Resources Management, 36(4). https://doi.org/10.1007/s11269-022-03068-

Yu, J., Qin, X., Larsen, O., & Chua, L. H. C. (2014). Comparison between Response Surface Models and Artificial Neural Networks in Hydrologic Forecasting. Journal of Hydrologic Engineering, 19(3). https://doi.org/10.1061/(asce)he.1943-5584.0000827

Zhang, P., Cao, C., Wang, Y. H., Yu, K., Liu, C., He, C., Shi, Q., Wang, J. J. 2021. Chemodiversity of water-extractable organic matter in sediment columns of a polluted urban river in South China. Science of the Total Environment, 777. https://doi.org/10.1016/j.scitotenv.2021.146127

Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Wu, B., & Ye, L. (2022). A review of the application of machine learning in water quality evaluation. Eco-Environment & Health, 1(2), 107–116. https://doi.org/10.1016/J.EEHL.2022.06.001

6. Statements and Declarations

6.1. Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

6.2. Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

6.3. Author Contributions

Mahdieh Jannatkhah: Conceptualization, Methodology, Data Collection, Data Analysis, Writing, Visualization, Validation.

Rouhollah Davarpanah: Methodology, Data Analysis, Writing - Review & Editing, Supervision, Visualization, Validation, Project Administration.

Bahman Fakouri: Methodology, Data Analysis, Writing - Review & Editing, Validation

Ozgur Kisi: Writing - Review & Editing, Supervision.

6.4. Data Availability statement

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Figures

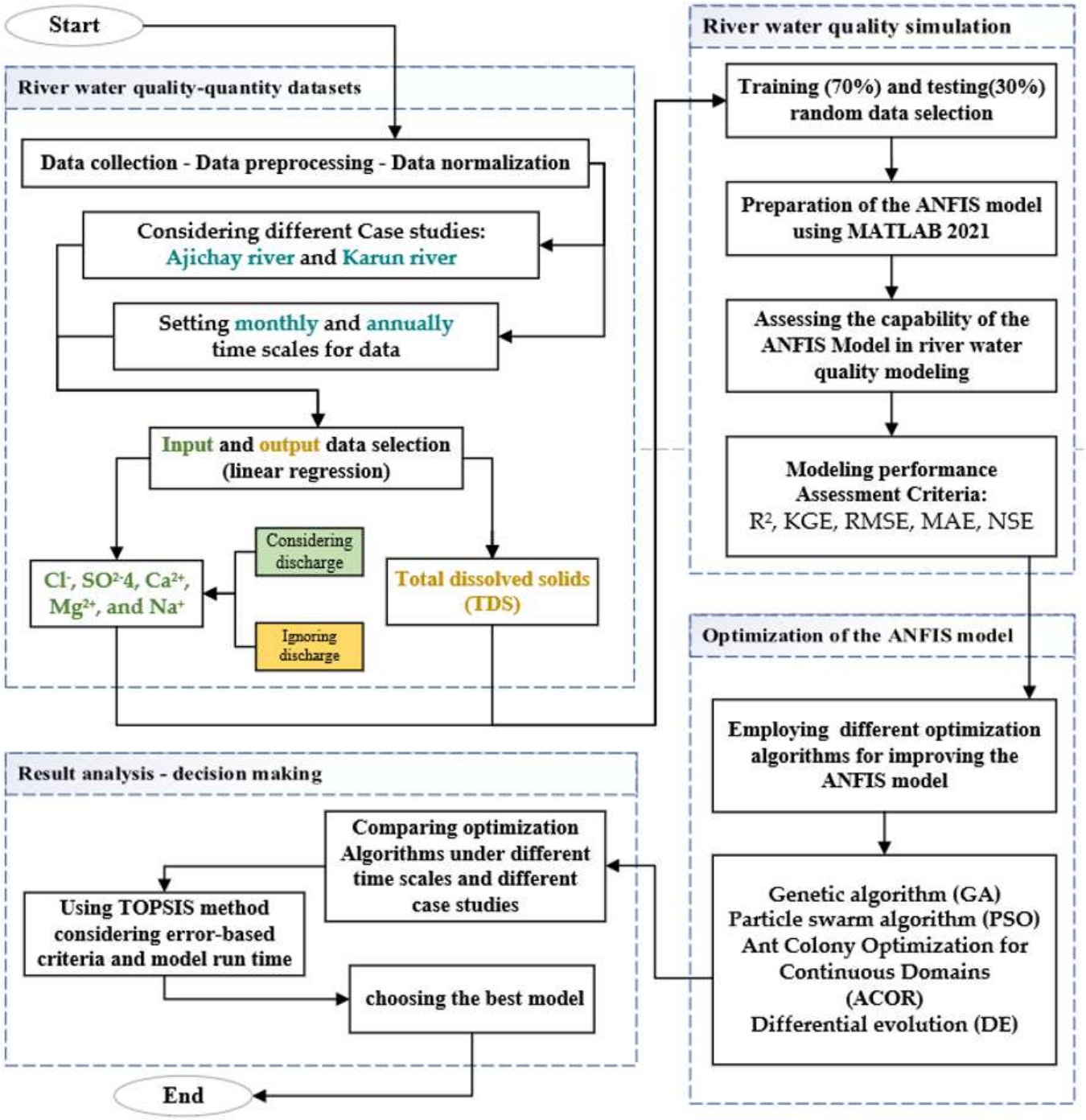

Figure 1

The research process followed in this study.

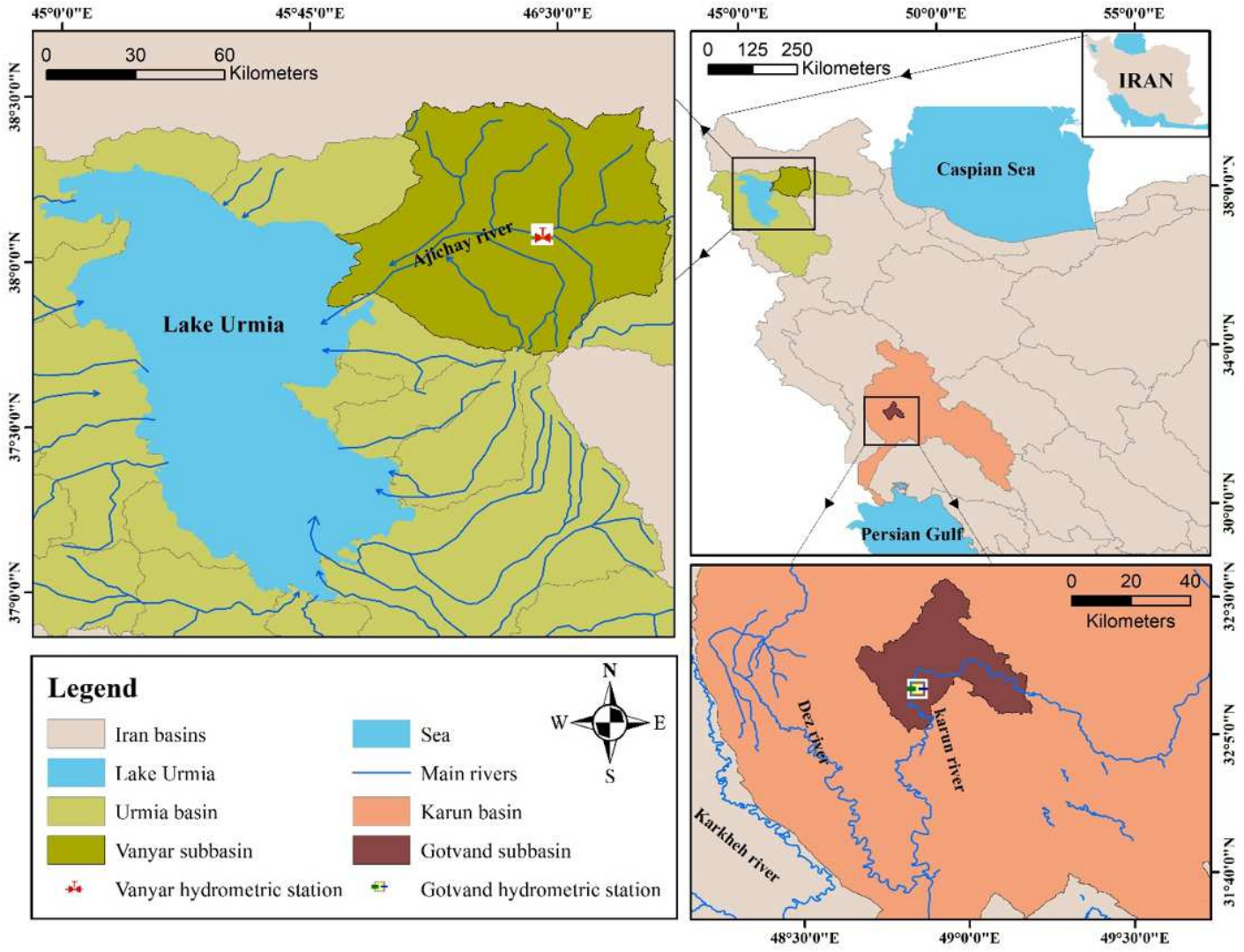

Figure 2

Study area and hydrometric stations.

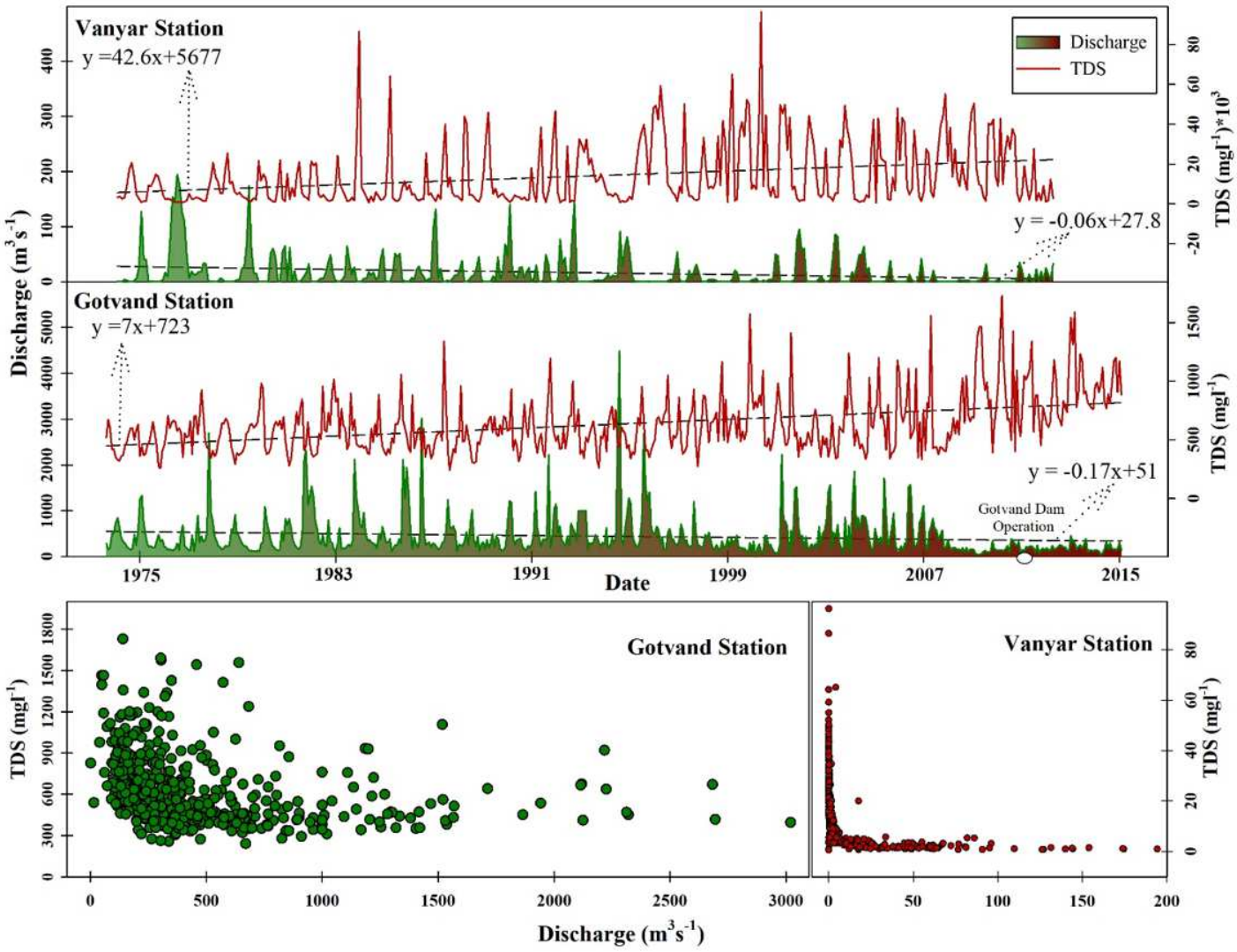

Figure 3

Time series of TDS (mg l^{-1}) and discharge (m^3 s^{-1}) in Gotvand and Vanyar hydrometric stations (the Karun and Aji chay rivers respectively); and Scatter plots of TDS versus discharge of both stations.

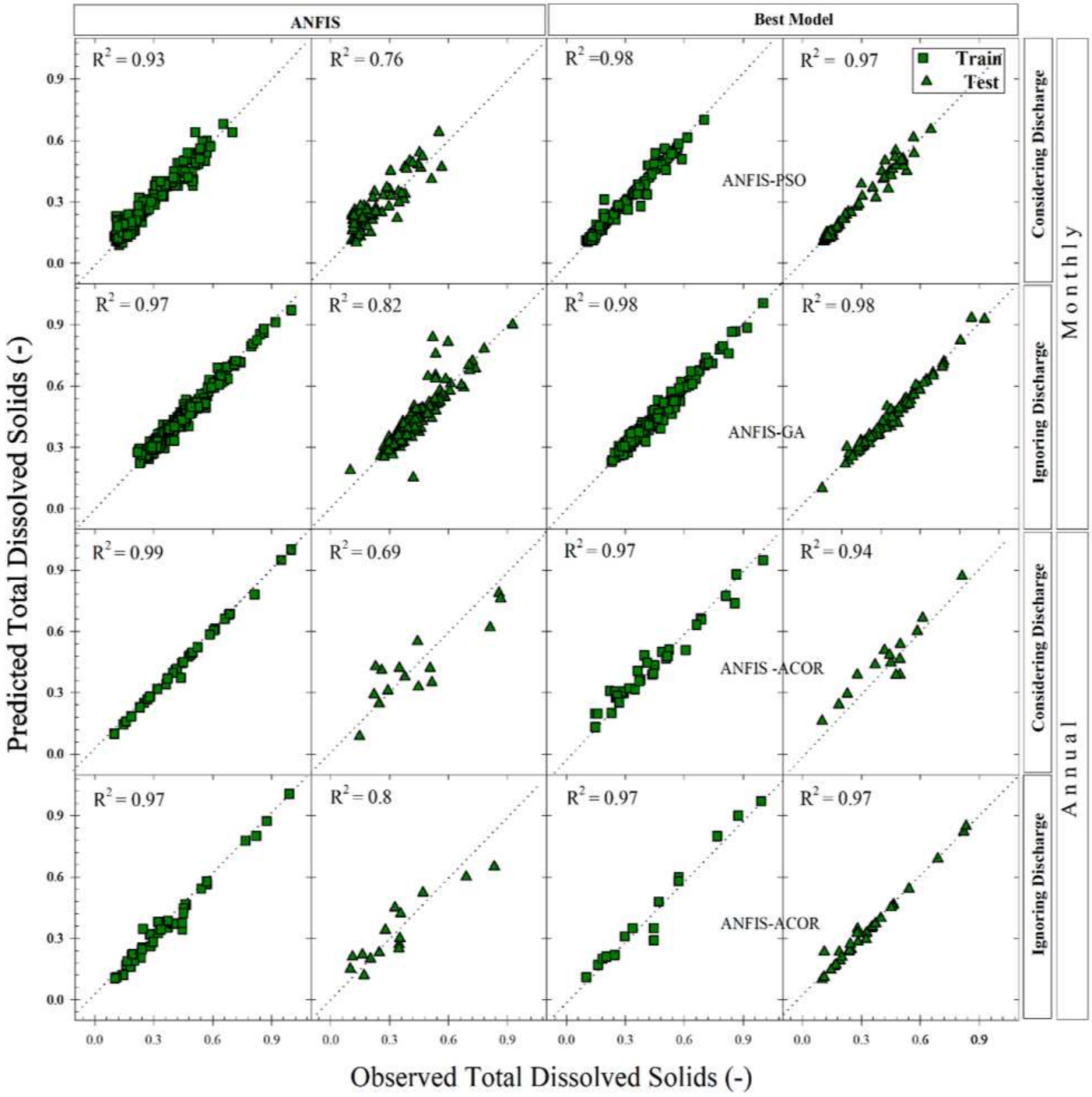

Figure 4

Scatter plots of normalized observed and predicted TDS of the Karun River water for ANFIS and the best hybrid models.

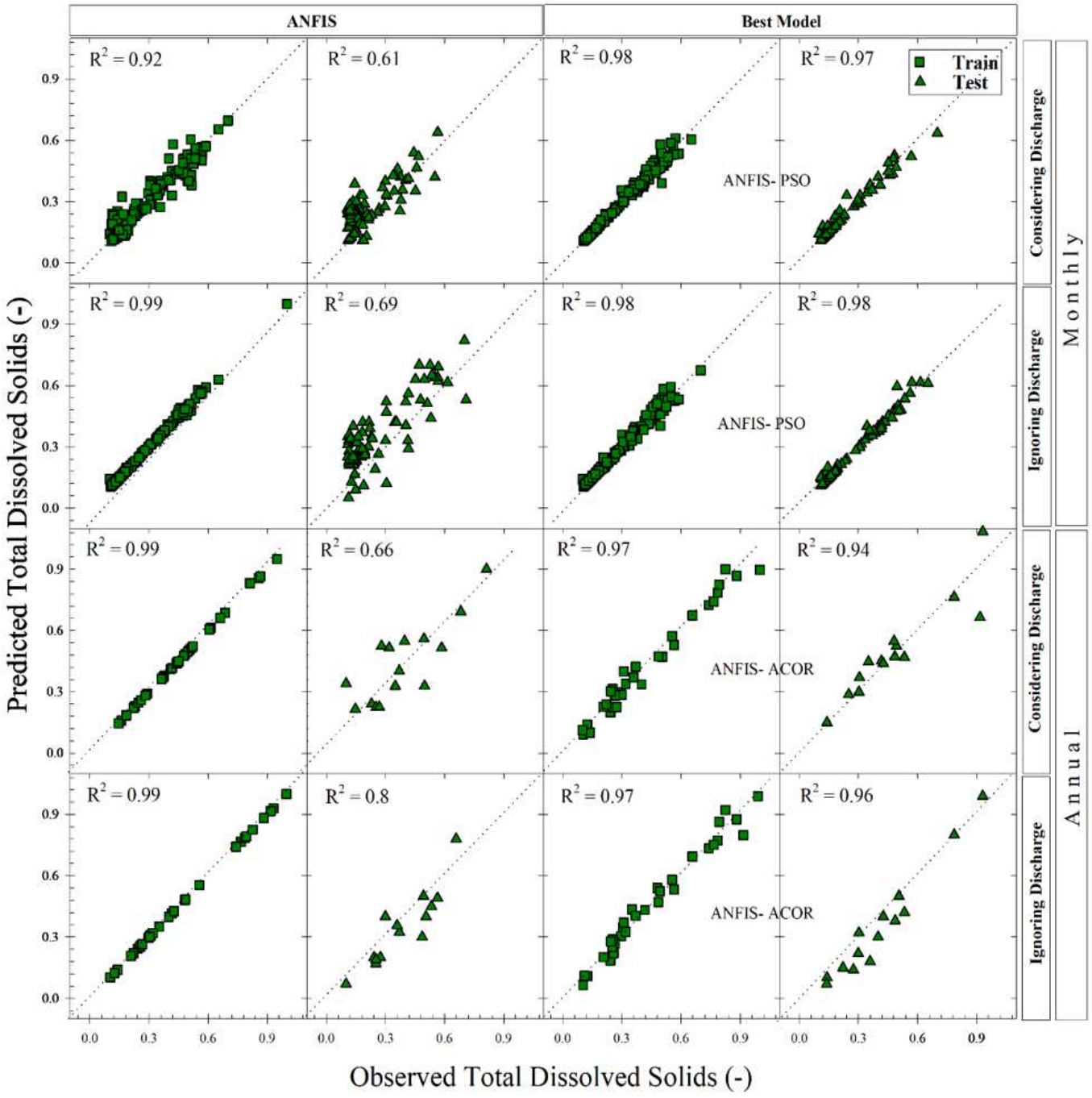

Figure 5

Scatter plots of normalized observed and predicted TDS of the Aji Chay river water for ANFIS and the best hybrid models.

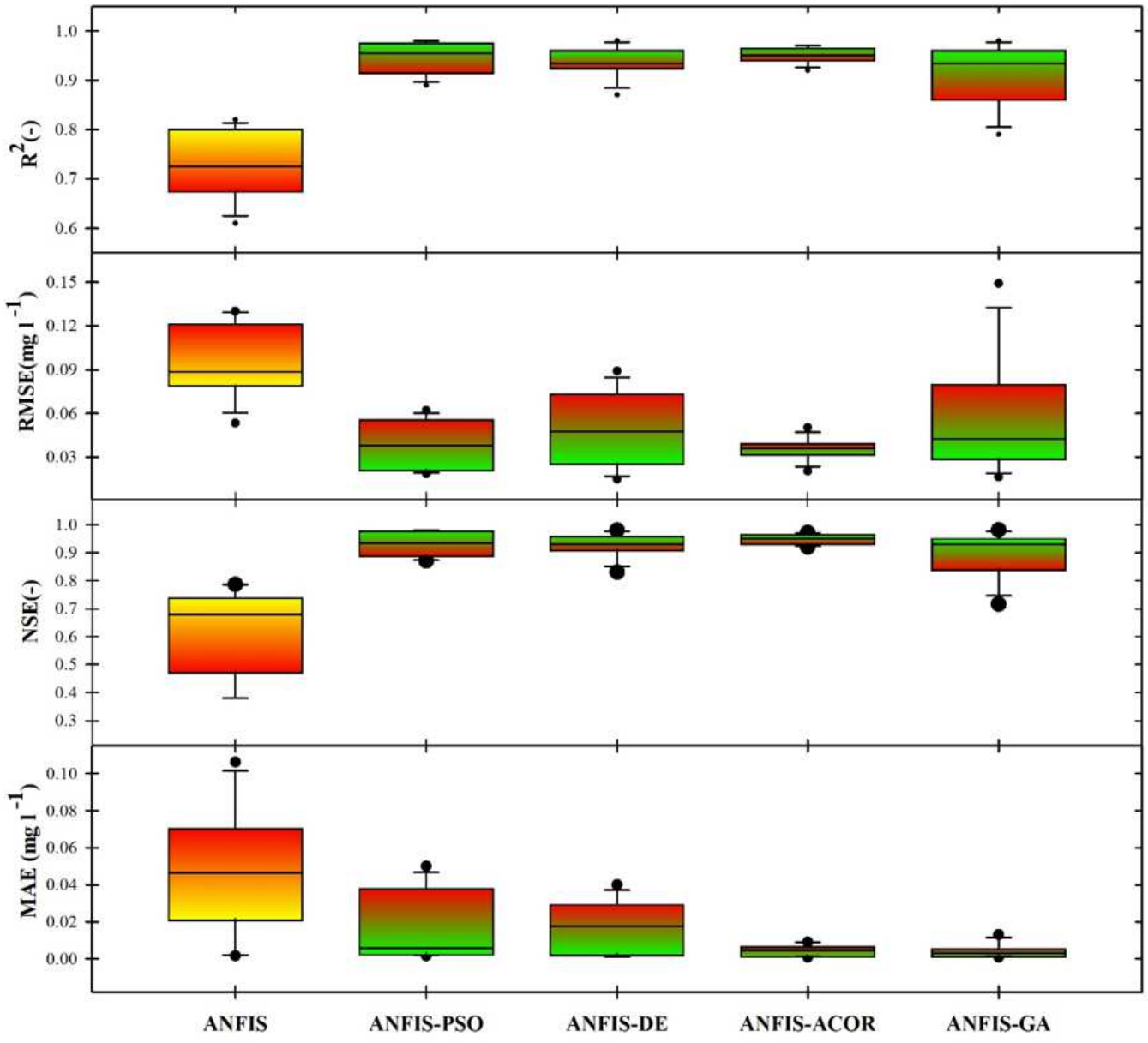

Figure 6

Box plots of Statistical parameters for ANFIS and hybrid approaches in this study.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- GraphicalAbstract.tiff
- Highlights.docx
- Tables.docx