

Prediction of groundwater quality parameters in Golestan province using response surface method, decision tree and neural network

Omolbani MohammadRezaPour¹, Behrooz Keshtegar^{2*} , Ozgur Kisi³ 

¹ Department of water engineering, Faculty of water and soil, Gorgan University of Agricultural Resources and Natural Resources, Gorgan, Iran.

² Department of civil, Faculty of Engineering, University of Zabol, Zabol, Iran.

³ Department of Civil Engineering, Technical University of Lübeck, Lübeck, Germany

Article Info

Article type:

Research Article

Article history:

Received 05 February 2025

Revised 26 June 2025

Accepted 08 July 2025

Published online 28 August 2025

Keywords:

Groundwater quality prediction; neural network decision tree response surface method

ABSTRACT

Objective: Groundwater quality is a main issue in most of the plains in Iran. Therefore, quality management and monitoring of water resources are of great importance.

Material and Methods: In this study, water quality parameters, including sodium adsorption ratio (SAR), total dissolved solids ratio (TDS), and electrical conductivity (EC), were predicted using an artificial neural network (MLP type), decision tree model (M5Tree), and response surface method (RSM). The quality data acquired from 96 observation wells located in Golestan province were used for model inputs, including sodium, water pH, chloride, sulfate, calcium, and magnesium. Models were evaluated utilizing three criteria: root mean square error (RMSE), determination coefficient (R²), and mean absolute error (MAE). Three different input combinations were considered to predict EC, SAR, and TDS.

Results and Discussion: The results of this study showed that the parameters Na and Cl have the greatest effect on the accuracy of the models. According to the results, the decision tree model (M5Tree) was found to have the highest accuracy in predicting EC (R² =0.997, RMSE=23.351, MAE=13.607) followed by the RSM (R² =0.98, RMSE=56.871, MAE=33.0428) and ANN (R² =0.991, RMSE=37.1073, MAE=17.279) models. However, the RSM model has a higher efficiency than the other models in predicting the SAR and TDS. According to the results obtained, the RSM generally predicts the groundwater quality parameters with relatively better accuracy.

Conclusions: The results obtained from this research showed that the models having all input parameters offered the best accuracy in predicting EC, SAR, and TDS. Also, the results showed that the Na and Cl parameters have the greatest effect on the accuracy of the prediction results for all three methods.

*Corresponding author, Email: Mohammadrezapour@gau.ac.ir

Cite this article: Mohammadezapour, Omolbani., Keshtegar, Behrooz., Kisi, Ozgur. (2025). Prediction of groundwater quality parameters in Golestan province using response surface method, decision tree and neural network.

Journal of New Approaches in Water Engineering and Environment, 4(2), 38-57.

<https://doi.org/10.22034/nawee.2025.504716.1136>



© The Author(s).

Publisher: Gonbad Kavous University.

DOI: <https://doi.org/10.22034/nawee.2025.504716.1136>

1. Introduction

Considering the importance of groundwater studies in the field of drinking, industry, and agriculture, investigating water quality changes is very important for mankind. Due to the undeniable effect of water on human life, such as health, industry, and agriculture, as well as the reduction of atmospheric precipitation and the high volume of groundwater extraction, it is essential to identify the quality characteristics of water and its compatibility with various uses. Iran is located on the dry belt of the world, and this issue has created many problems and dilemmas for managing Iran's water resources, which is considered the most important, such as water supply needed for agriculture, drinking, and industry. Therefore, studying and evaluating the quality of these sources is more important than ever (Mohammadrezapour et al, 2018). Therefore, recognizing the quality and quantity of underground water as one of the most important and vulnerable sources of water supply in recent decades is quite obvious (Rizo and Mouser, 2000). Management and protection of groundwater is important in countries like Iran. So far, many statistical and intelligent methods have been developed for such research, and many studies have been conducted in hydrochemical modeling of underground water using multivariate and intelligent statistical methods (Kumru and Bakac, 2003). The main advantage of these methods is the possibility of using time series learning capabilities and forecasting. One of the important characteristics of artificial intelligence methods is their ability to establish a connection between the input and output of a process without considering the physics of that problem (Vafakhah, 2012). In a research, Yesilnacar et al. (2008) used an artificial neural network to predict the nitrate in the groundwater wells of the Hart plain located in Turkey. They could predict the nitrate concentration with high accuracy. In research, the ability of neural networks to predict the water quality indicators of Johor River in Malaysia and estimate the amount of TDS and EC in this basin were emphasized (Najah et al, 2009). Guler et al (2012) investigated the effect of human activities on the hydrogeochemical parameters of groundwater in the Tartus coastal plain in Turkey with multivariate statistical methods, fuzzy clustering (FC), and GIS technique. The groundwater samples were split into 4 different groups using the FC method. Sivasankar et al (2013) investigated the FC of groundwater quality parameters of Rameswaram in South India. The outcomes revealed that the seasonal changes from summer to winter caused variation in the quality parameters of the underground water samples, resulting from the clusters. Porjebar et al. (2014) searched for the ability of the hierarchical FC method to collect data contaminated with heavy metals produced from a mine in Germany. The outcomes indicated that fuzzy algorithms were better for analyzing geological and hydrogeological data with inherent ambiguity and uncertainty. In a research work, Jalal Kamali and Jalal Kamali (2018) investigated the capabilities of the Adaptive Neural Fuzzy Inference System (ANFIS), a combination of ANFIS with Genetic Algorithm (GA) and Artificial Neural Network (ANN) in predicting groundwater quality in Kerman plain. Different combinations of lagged rainfall and groundwater levels were used as inputs to two different neural-fuzzy models (ANFIS and ANFIS-GA) and ANN. The results showed that the ANFIS-GA method could provide a more efficient model with a lower number of rules used (about 300% reduction) than the standard ANFIS model. This work also showed that fluctuations in groundwater levels and rainfall contribution are two important factors in predicting groundwater quality indicators. Kisi et al. (2019) used four evolutionary-based optimization methods, continuous genetic algorithm (CGA), particle swarm optimization (PSO), ant colony optimization for continuous domains (ACOR) and differential evolution (DE) for tuning neuro-fuzzy models to predict quality variables of groundwater such as electrical conductivity (EC), sodium absorption ratio (SAR) and total hardness (TH). At first, using correlation and sensitivity analysis, the best inputs were

obtained. The results showed that: 1) All the proposed algorithms improve the performance of ANFIS in EC and TH prediction. In SAR prediction, ANFIS-CGA and ANFIS-PSO performed better than the ANFIS method. 2) CGA was found to be the best method for improving the accuracy of ANFIS for predicting the quality variables of groundwater. 3) The results also revealed that correlation analysis was reliable in selecting the best input combination for predicting groundwater quality variables. Adimalal et al. (2022) To assess the quality of groundwater and its suitability for drinking purposes in the semi-arid region of southern India, the parameters include TDS, Mg, Ca, K, HCO₃, SO₄, Cl, Na, SAR, EC, total anion, and cation variables. were analyzed using the standard method of the American Public Health Association. Based on the results, the groundwater was severely affected by fluoride, and about 60% of the groundwater samples were shown to be unsuitable for drinking purposes in the study area. Marashi et al (2023) They studied the trend of groundwater quality changes in 23 selected wells in the Haraz Plain using the nonparametric Mann-Kendall method. For this purpose, the statistics of 20 years (1996-2016), 23 measurement wells in the Haraz plain, located in the north of Iran, were used. The investigated water quality and salinity parameters include TDS, Mg, Ca, K, HCO₃, SO₄, Cl, Na, SAR, EC, total anion, and cation variables. The results showed that the trend of changes in most of the groundwater quality variables in this plain was decreasing. Considering the negative trend of most of the quality variables in the Haraz plain, it can be said that groundwater quality has increased. Heshmatpour and Sajadi (2024) evaluated the agricultural water quality of the Gorganrud River in Iran using a fuzzy inference system (FIS). By integrating sodium adsorption ratio (SAR) and electrical conductivity (EC) data from 1961 to 2014, the fuzzy model enhanced the traditional Wilcox classification. The results indicated that the fuzzy approach provided a more accurate and realistic assessment of the spatial and temporal variations in water quality. Solgi et al. (2025) evaluated the drinking water quality of Malayer City, Iran, using WQI, NPI, and Wilcox indices, as well as the non-carcinogenic risk of nitrate. Sixty tap water samples from 15 urban areas were analyzed for pH, electrical conductivity, and nitrate concentration. The results indicated that water quality was generally excellent, with nitrate and pH within permissible limits, although EC slightly exceeded standards. Non-carcinogenic risk was below one for all age groups, with infants being the most vulnerable, highlighting the need for periodic water monitoring.

Golestan province is one of the northeastern provinces of Iran, and it is considered one of the country's main agricultural areas. It has suitable and wide plains and suitable water sources such as the Gorganrud River. Therefore, the use of chemical fertilizers and poisons in this area is inevitable. Since these fertilizers and chemical poisons, as well as domestic and industrial wastewater, penetrate into the underground aquifers, as a result, the quality of these waters changes over time and space (Mohammadrezapour et al, 2018).

This research uses the qualitative statistics of 96 observation wells located in Golestan province, which has a common statistical year of 19 years (1998-2016), to predict water quality parameters, including sodium absorption ratio (SAR), total dissolved solids (TDS), and electrical conductivity. (EC). Three intelligence methods, including artificial neural network, decision tree, and response surface method, were compared using input combinations of water alkalinity (pH), chloride (Cl), sulfate (SO₄), calcium (Ca), magnesium (Mg), and sodium (Na).

2. Material and Methods

The qualitative parameters of the underground water of Golestan province were predicted by intelligent models using the statistics and information that were available from the regional water company of this province. (pH), chloride (Cl), sulfate (SO_4), calcium (Ca), magnesium (Mg), sodium (Na) data from 96 observation wells covering 19-year period (1998-2016) were utilized to predict sodium absorption ratio (SAR), electrical conductivity (EC) and the ratio of total dissolved solids (TDS). The location of the wells is shown in Figure 1. Artificial intelligence methods, including artificial neural network (ANN-MLP), decision tree model (M5Tree), and response surface method (RSM), were used to predict the qualitative parameters of Golestan province. First, the input data was divided into two groups: calibration (training) and test data. 70% of the data was used for training, and the remaining 30% that did not play a role in training was used as testing data. The training and testing data were randomly selected and used after normalizing. Table 1 enlists the input and output combinations used in the presented study. The input combinations were decided considering correlation analysis, and these three scenarios were employed for three intelligent methods.

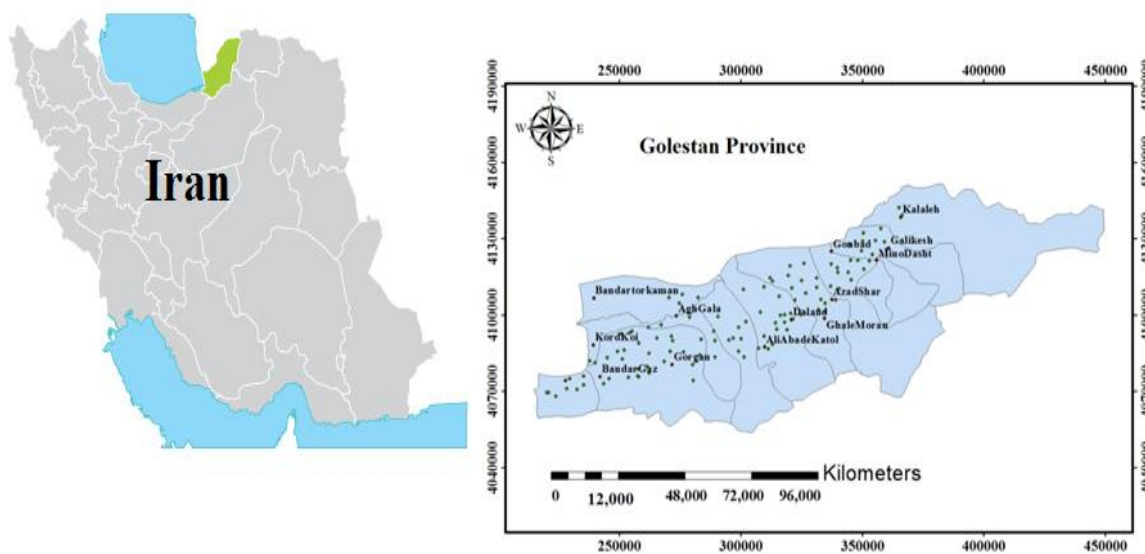


Figure 1- Location of sample wells in the Golestan province aquifer

Table 1- shows different combinations of the studied models

Model	Output	Input variables
1	EC· TDS·SAR	Na,Mg,Ca,SO ₄ ,Cl,pH
2	EC· TDS·SAR	Mg, Ca,SO ₄ , pH
3	EC· TDS·SAR	Na,Cl

2.1. Decision tree model (M5 Tree)

The decision tree model (M5Tree) is the basis of tree models based on decision and overcoming methods. Replacing the line regression equation instead of labels in the nodes is a

method implemented in the M5 model that can predict or estimate continuous numerical variables.

The structure of the decision tree is similar to a tree that consists of 1- root, 2- branches, 3- nodes, and 4- leaves. The construction of the tree model is done in two steps. In the first step, the decision tree is formed by branching the data. The branching criterion in the M5Tree model is to maximize the reduction of the standard deviation of the data in the child node. When it is not possible to reduce the standard deviation of the child node's data, its parent node does not branch and reaches the final node or leaf. The formula used to calculate the standard deviation reduction is (Satari et al. 2013):

$$SDR = sd(T) - \sum_{i=1}^N \frac{T_i}{T} sd(T_i) \quad (1)$$

which in relation (1), T represents a set of samples that enter each node, T_i represents a subset of samples that have been tested for the i -th result, have the potential, and Sd is the standard deviation of the input data and is calculated as follows (Satari et al., 2013).

$$sd(T) = \sqrt{\frac{1}{N} \left(\sum_{i=1}^N y_i^2 - \frac{1}{N} \left(\sum_{i=1}^N y_i \right)^2 \right)} \quad (2)$$

In relation (2), Y_i represents the numerical value of the characteristic of sample i , and N represents the number of data. Since the process of branching (categorization) in the child nodes has less standard deviation than the parent nodes, they have more accurate results and have more homogeneity. After examining all possible categories, the M5 model selects the one that has the minimum expected error. The second stage of designing the tree model is reducing the overgrown tree by pruning the branches and replacing it with linear regression functions.

Response surface method

Response Surface Method (RSM) is a set of mathematical and statistical methods essential for solving problems in which the response parameter is affected by various independent variables. Its goal is optimization of the response variables (Montgomery, 2001).

The RSM is useful for describing quality indicators during food processing (Thomson, 1982). In this procedure, the response parameter is assessed on a fixed scale. In most practical issues, more than one factor is included in the quality and performance of a product, and these factors must be investigated. In response level optimization, the input variables are identified as independent variables, and the influence of these variables on the dependent or output variables is studied. The main advantage of RSM is reducing the number of repetitions of experiments to evaluate multiple parameters and their interrelationships. Basically, the RSM includes the following steps: a) two-factor tests to screen the effective input variables b) regression analysis to estimate the fitting function of the outputs according to the inputs; c) optimization in order to determine the optimal levels of input variables.

RSM is a second-order polynomial function based on a simple and effective approximation approach as follows:

$$f(x) = \beta_0 + \sum_{i=1}^n \beta_i x_i + \sum_{i=1}^n \sum_{j=1}^n \beta_{ij} x_i x_j \quad (3)$$

Where $f(x)$ is the predicted values, n is the number of input variables, and β_i , β_0 , and β_{ij} are unknown coefficients. The ordinary least square estimator is generally used to calibrate the unknown coefficients of equation (3) as follows:

$$a = [P(X)^T P(X)]^{-1} [p(X)^T (SR)] \quad (4)$$

Where SR is the observation vector of the SR polynomial basis function model $P(x)$, which includes the input data as follows.

$$P(x) = \begin{bmatrix} 1 & X_{11} & x_{n1} \\ 1 & & \\ 1 & x_{1N} & x_{nN} \end{bmatrix} \quad (5)$$

2.2. Artificial Neural Network (ANN)

An artificial neural network is a model that processes information like the human brain, and in other words, it is a mathematical simulation of the human nervous system. The network is trained by examples. This model is not planned for doing specific work, but it is such that, over time and more interaction with the work environment, it has become more experienced, can draw logical conclusions, and can learn. The process of simulation and prediction in this model is based on learning the relationships of the observed variables and generalizing these relationships for the future to make new predictions. More details Information about ANNs can be found in, e.g., Haykin (1999).

2.3. Evaluation Criteria:

To evaluate and check the effectiveness of the methods used to predict quality parameters of groundwater in Golestan province, various evaluation criteria were used, including the coefficient of determination (6), the root mean square error (7), and the mean absolute value of the error (8).

$$R^2 = \left(\frac{\sum_{i=1}^N (T_i^O - \bar{T}^O)(T_i^E - \bar{T}^E)}{\sqrt{\sum_{i=1}^N (T_i^O - \bar{T}^O)^2 \sum_{i=1}^N (T_i^E - \bar{T}^E)^2}} \right)^2 \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (T_i^E - T_i^O)^2}{N}} \quad (7)$$

$$MAE = \frac{1}{N} \sum_{i=1}^n |T_i^E - T_i^O| \quad (8)$$

Where T_i^O and T_i^E respectively are the observed and predicted values at the time step i , and \bar{T}^O and \bar{T}^E are the means of the observed and predicted values. N also represents the number of data .

3. Results and Discussion

As stated in the previous part, the purpose of this research is to investigate the performance of three artificial intelligence models, such as artificial neural network (ANN), decision tree (M5Tree), and response surface method (RSM), to predict the quality parameters of groundwater in Golestan province during the years 1998 to 2016.

In this section, the results of the models were first examined, and then the best input model was selected for each artificial intelligence model. The results of these models were compared with each other, and the best model was selected.

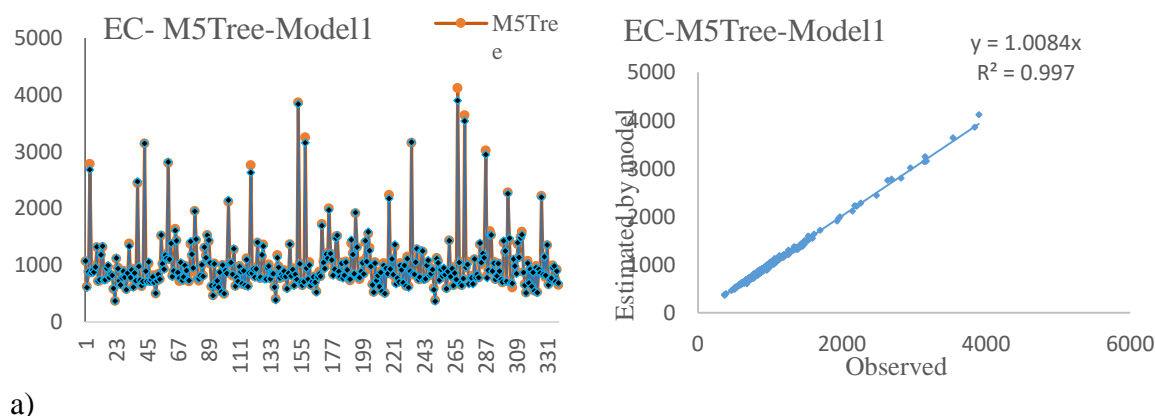
Prediction result of M5Tree

The M5Tree model with different inputs was investigated to predict the quality parameters (SAR, EC, and TDS) of groundwater, and the best models with respect to minimum RMSE, MAE, and maximum R^2 . The prediction results of the M5Tree model are given in Table 2.

Table -2 Prediction results of the M5Tree model in the test stage

Model	Criterion	Model 1	Model 2	Model 3
EC	R ²	0.997	—0.783	0.9193
	RMSE	23.35	349.47	118
	MAE	13.61	159.97	67.13
SAR	R ²	0.995	0.3837	0.957
	RMSE	0.123	1.69	0.379
	MAE	0.0507	0.823	0.146
TDS	R ²	0.894	—0.099	0.745
	RMSE	78.93	271.62	128.7
	MAE	23.94	115.71	44.71

As can be seen from the results, model 1, with the values of the coefficient of determination (0.997), the root mean square error (23.351), and the mean absolute value of the error (13.61), has the best accuracy in predicting the EC parameter. Also, model 3 has better accuracy than model 2 with the values of (0.9193), (118.0), and (67.13) for R², RMSE, and MAE, respectively. In predicting the SAR parameter, model 1 has the best accuracy with the values of the R² (0.995), the RMSE (0.1227), and the MAE (0.0507). After model 1, model 3 has the appropriate accuracy with the values of R² (0.957), RMSE (0.379), and MAE (0.146). In predicting the TDS parameter, model 1 is more accurate than models 3 and 2, with the values of the R² (0.894), the RMSE (78.93) and the MAE (23.94). The results show that removing NA and Cl from the model inputs (model 2) deteriorates the prediction accuracy. It is also understood from the outcomes of model 3 that Na and Cl are highly effective on EC, SAR, and TDS. Figures (2) to (4) show the graph of observed and simulated values of EC, SAR, and TDS parameters using the decision tree model. It is clear from the time variation and scatterplots that the M5Tree model, having all input parameters, has better prediction accuracy than the other two models.



a)

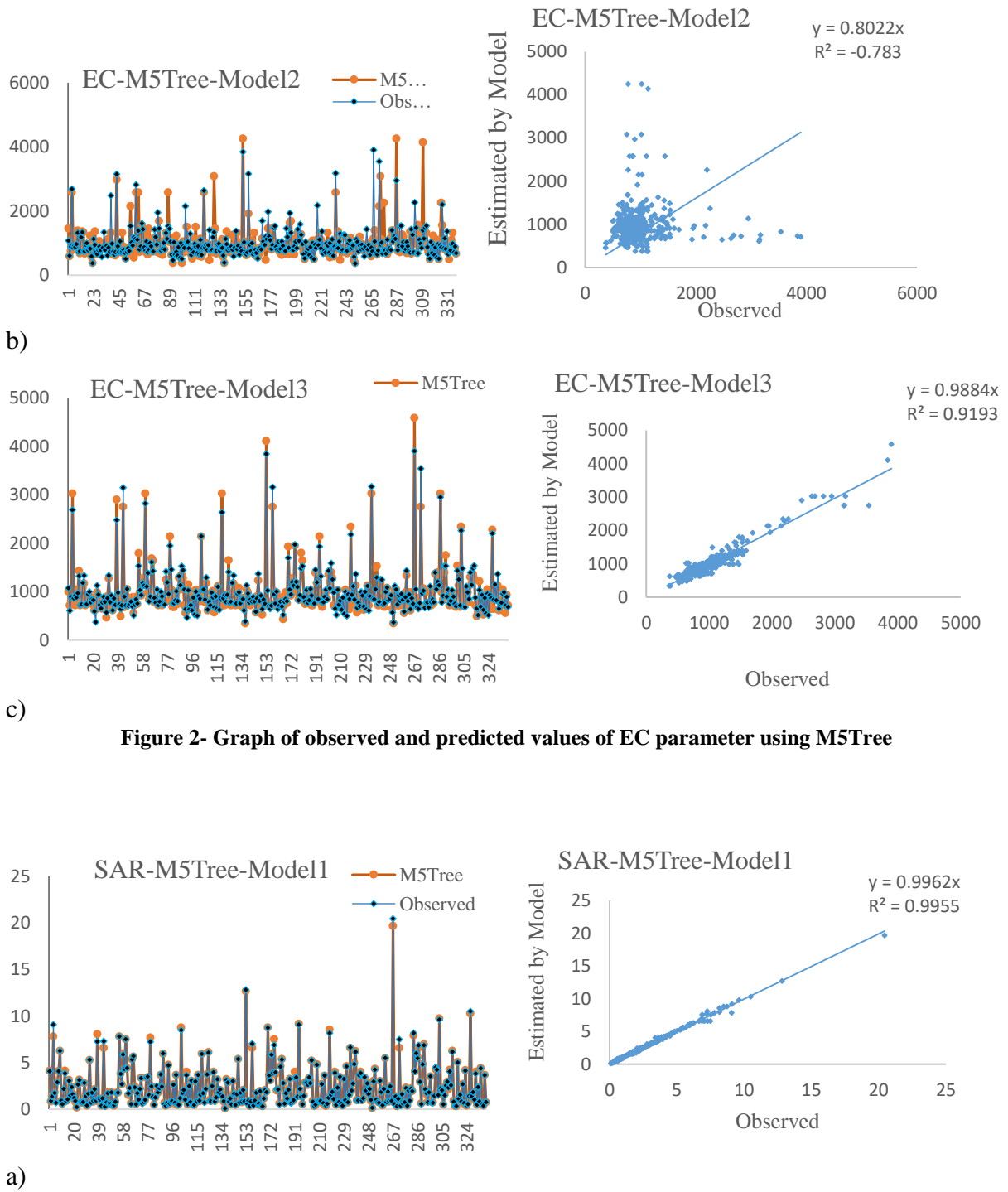


Figure 2- Graph of observed and predicted values of EC parameter using M5Tree

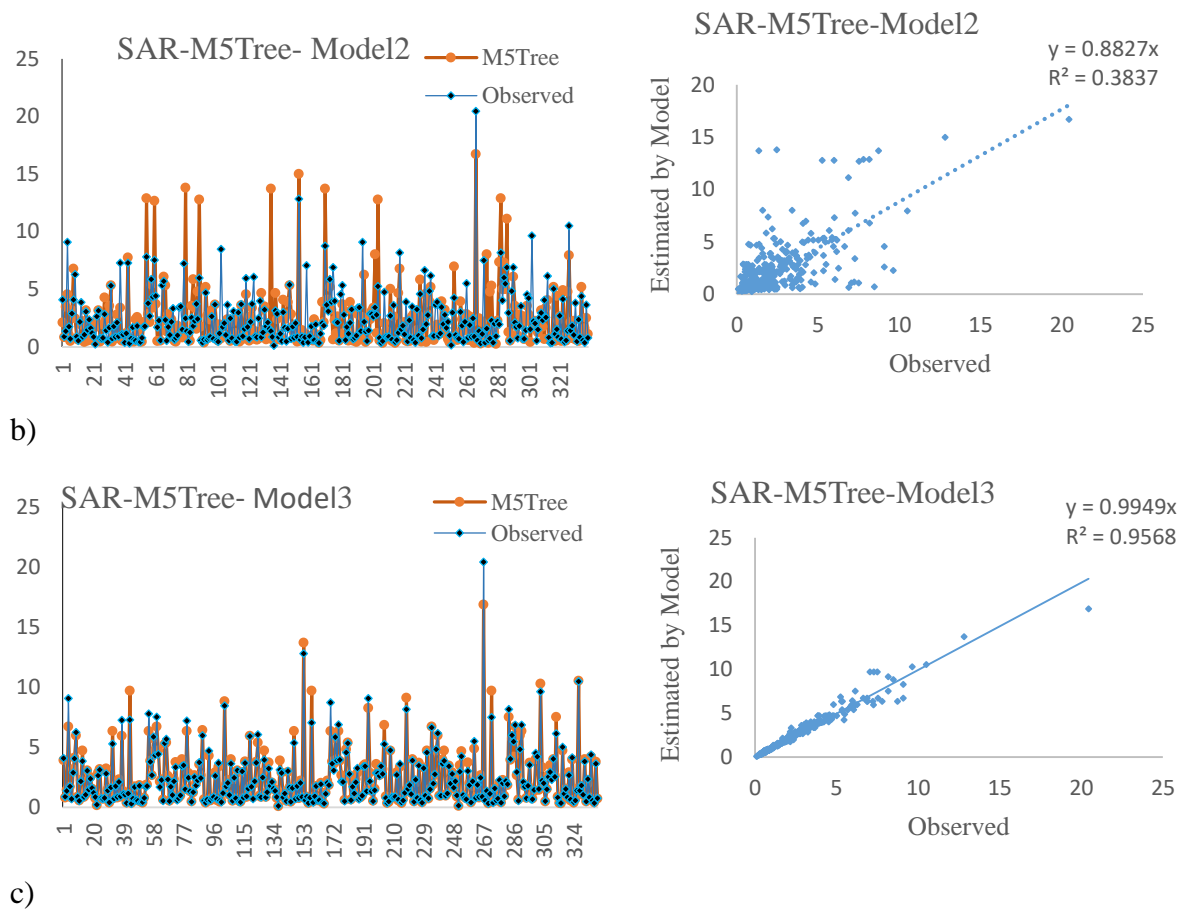
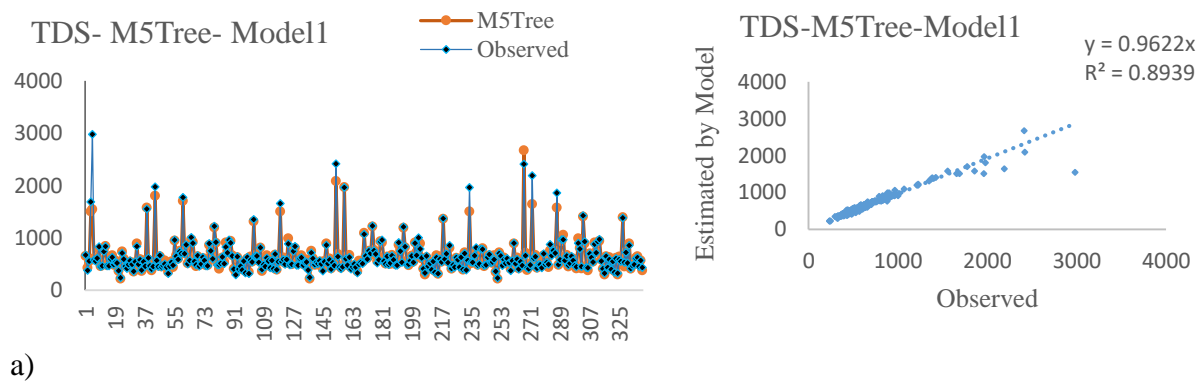


Figure 3- Graph of observed and predicted values of the SAR parameter using M5Tree



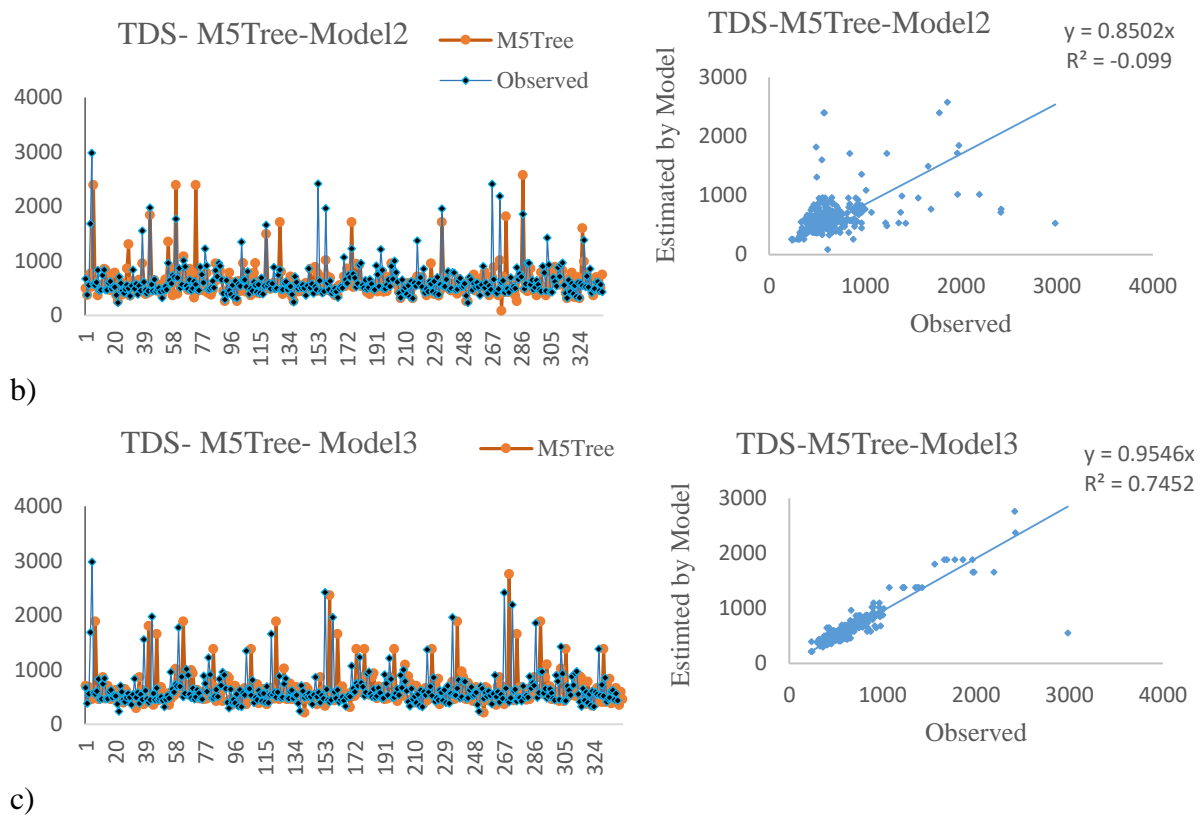


Figure 4-Graph of observed and predicted values of the TDS parameter using M5Tree

Prediction results of RSM

Table 3 shows the results of the evaluation criteria for the test stage of different models for predicting the quality parameters of underground water using the response surface method.

Table 3-Prediction results of the RSM in the test stage

Model	Criterion	model 1	model 2	model 3
EC	R ²	0.9807	0.879	0.9407
	RMSE	56.87	291.47	99.92
	MAE	33.04	143.18	56.63
SAR	R ²	0.9966	-0.045	0.9583
	RMSE	0.111	1.453	0.366
	MAE	0.0494	0.768	0.0333
TDS	R ²	0.932	-1.338	0.775
	RMSE	65.43	169.06	121.04
	MAE	13.95	79.951	39.368

As can be seen from Table 3, model 1 is the most accurate in predicting the EC parameter with the highest coefficient of determination (0.9807), the lowest root mean square error (56.871), and average absolute error (33.04). Also, in predicting the SAR and TDS parameters, model 1 has the highest accuracy, and its accuracy is followed by model 3, as found for the M5Tree model. It can be seen from Table 3 that model 3, including only Cl and Na parameters,

has an acceptable ability to predict SAR, EC, and TDS parameters. It is apparent that removing CL and NA parameters from the model inputs (model 2) significantly decreases its efficiency in predicting EC, SAR, and TDS. This finding supports the M5Tree model (Table 2).

Figures 5 to 7 show the observed and predicted values of EC, SAR, and TDS parameters using the surface response method. It is visible from the graphs that the RSM with full inputs offers better predictions which are closer to the observed ones, compared to the other two models, and the RSM for model 3 can be a good alternative in predicting EC, SAR, and TDS when Mg, Ca, SO₄, and pH are not available.

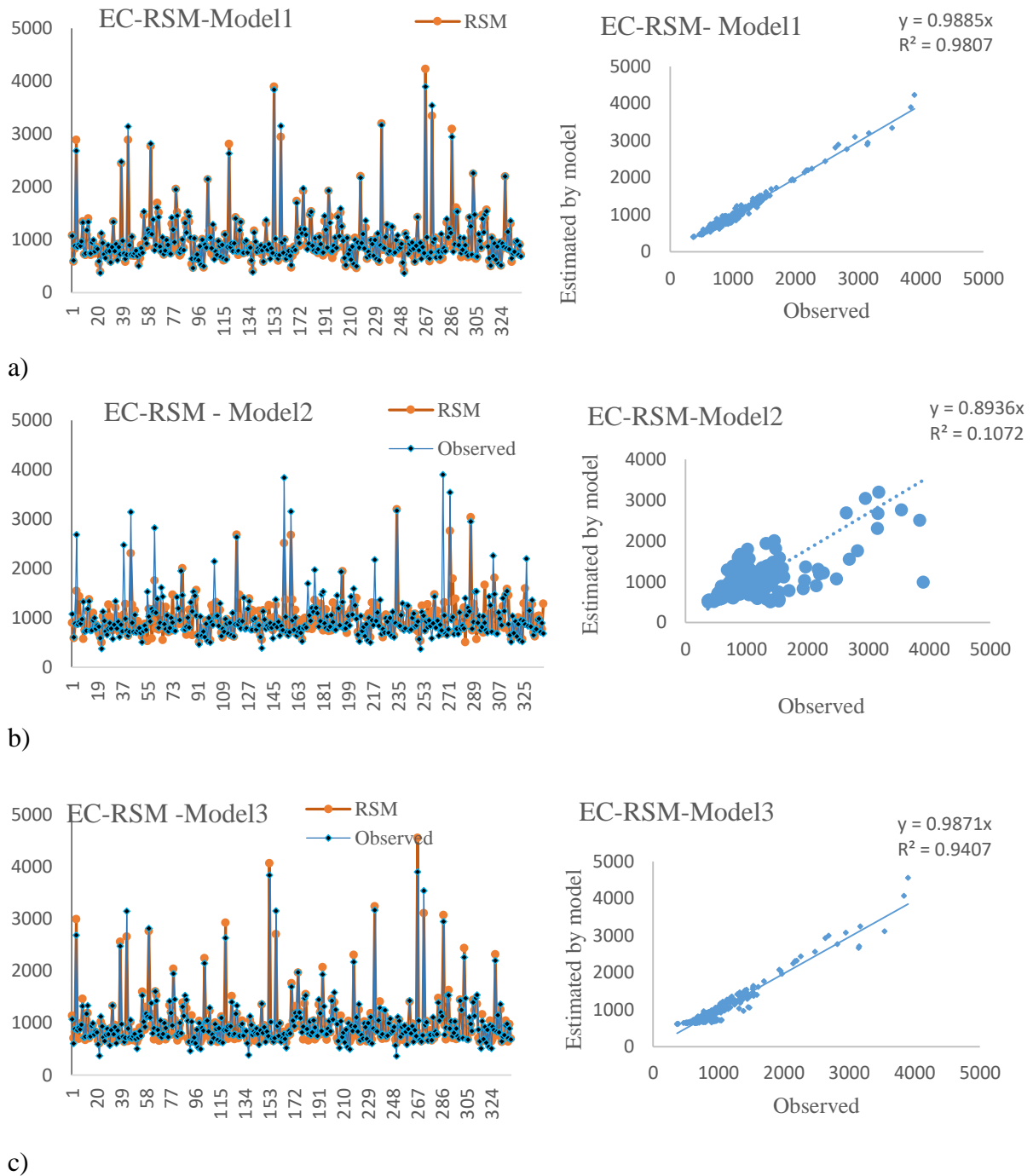


Figure 5- Graph of observed and predicted values of EC parameter using RSM

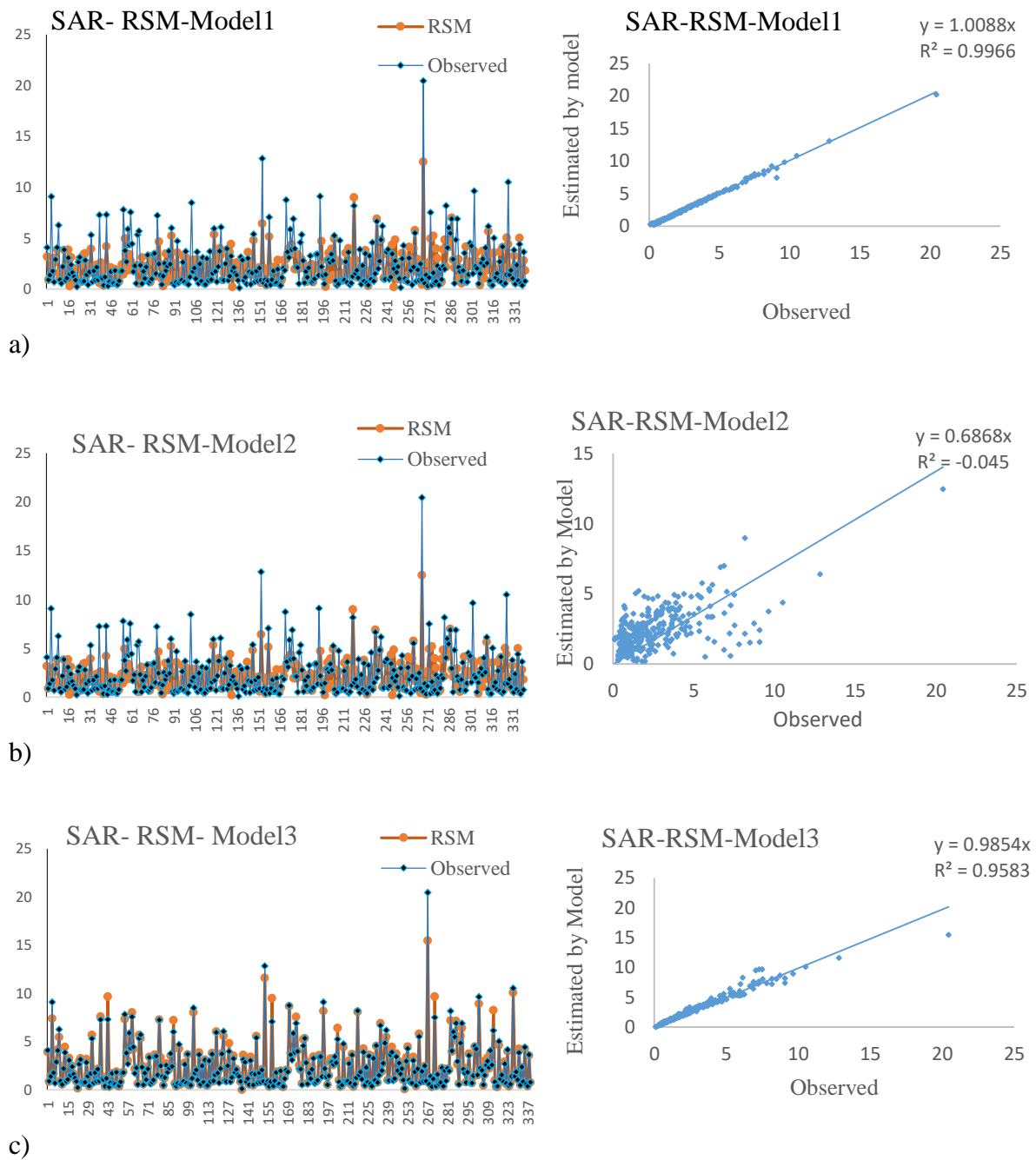


Figure 6- Graph of observed and predicted values of SAR parameter using RSM

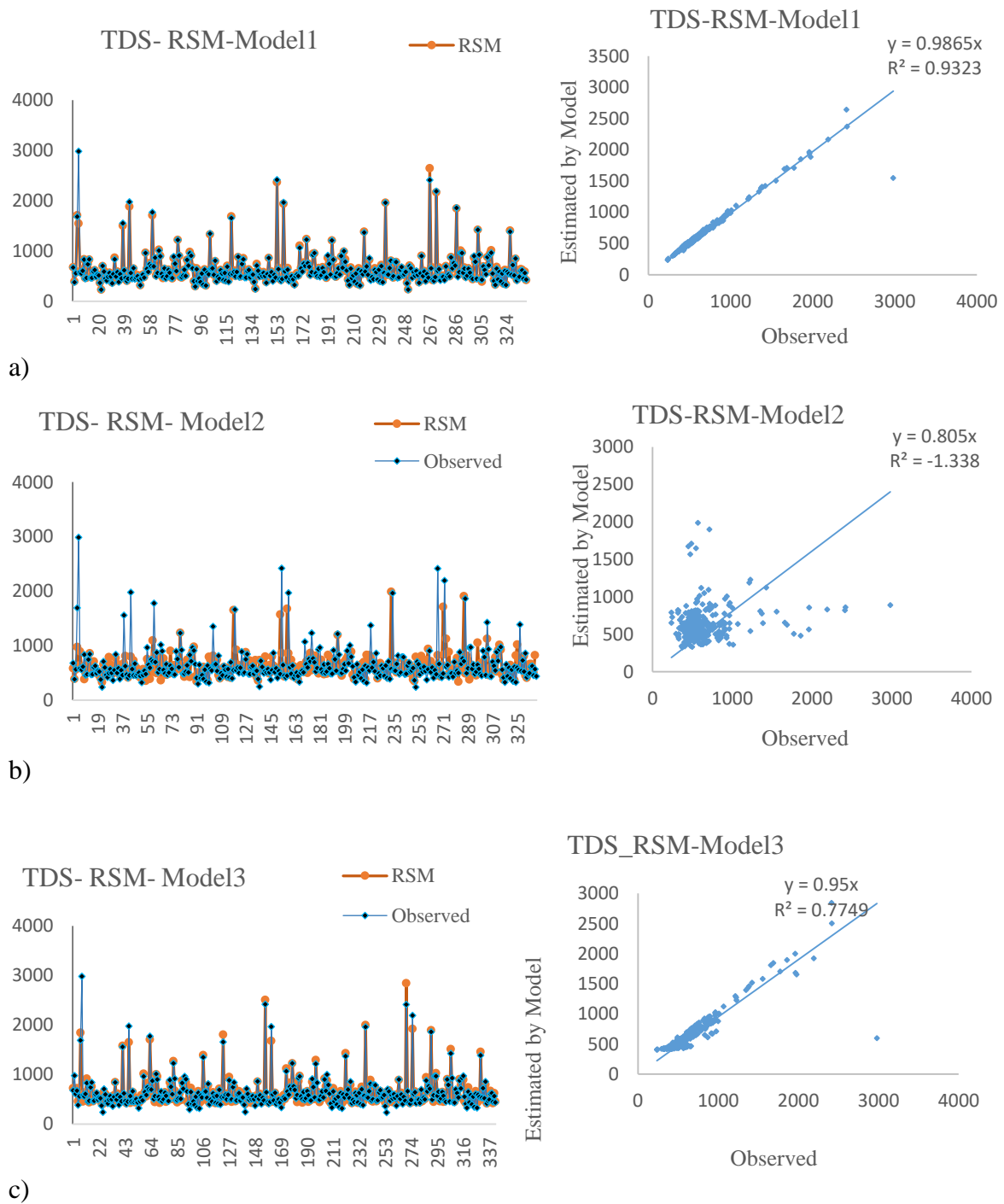


Figure 7- Graph of observed and predicted values of TDS parameter using RSM

Prediction results of ANN

Table 4 shows the results of the evaluation criteria for the testing stage of different models for predicting the quality parameters of underground water using an ANN model. To predict EC, SAR, and TDS parameters, 3 different input combinations were given to the ANN model.

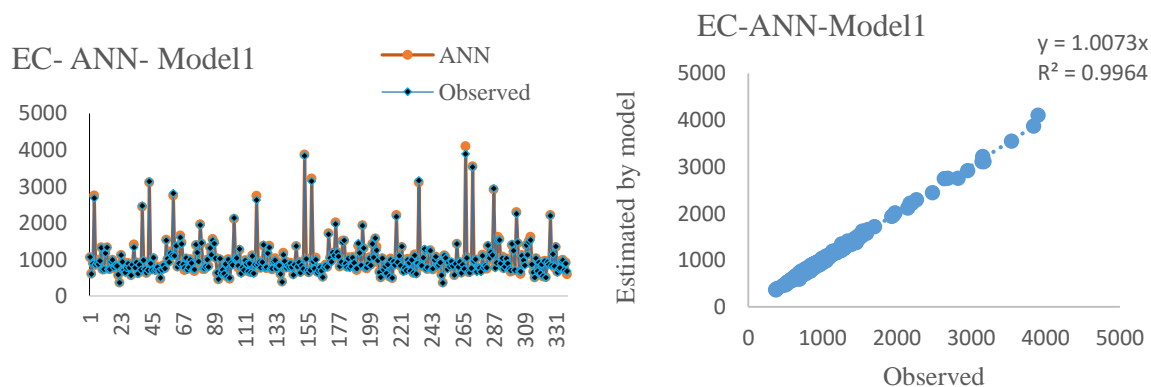
Table 4- Prediction results of ANN in the test stage

Model	Criterion	Model 1	Model 2	Model 3
EC	R ²	0.9964	0.1455	0.9388
	RMSE	25.19	285.84	102.1
	MAE	15.13	144.38	58.04
SAR	R ²	0.9881	0.0473	0.9496
	RMSE	0.2	1.423	0.3907
	MAE	0.1042	0.7693	0.1764
TDS	R ²	0.9282	0.099	0.7796
	RMSE	72.15	211.71	126.53
	MAE	22.148	96.19	42.66

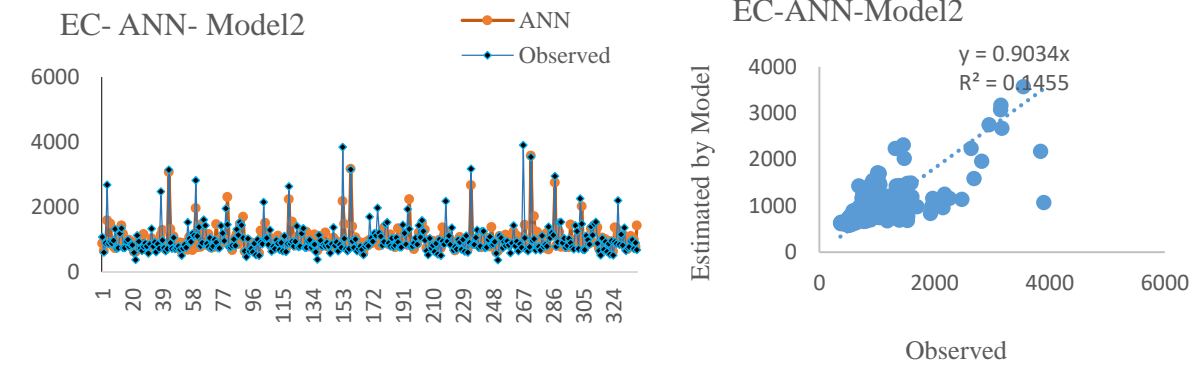
According to the results of Table 4, it can be seen that model 1, including all input parameters, has the highest R² (0.9664) and the lowest RMSE (25.19) and MAE (15.13) in predicting the EC parameter. It can also be seen that the input combination of model 1 in the prediction of SAR and TDS parameters also has the highest R² and the lowest RMSE and MAE values. Here, the model 3 involving only CL and NA inputs provides good accuracy in predicting three groundwater quality parameters. Also, model 2 has the weakest performance among all models for all three parameters predicted in this research, which include Mg, Ca, So₄, and Ph parameters. This reveals the necessity of CL and Na parameters in predicting EC, SAR, and TDS.

Figures (8) to (10) show values and predicted values using the ANN model for EC, SAR, and TDS parameters. Similar to the M5Tree and RSM, model 1 has less scattered predictions and they better follow the corresponding observed values. Model 2 predictions are also acceptable with the high explained variances (see R² values).

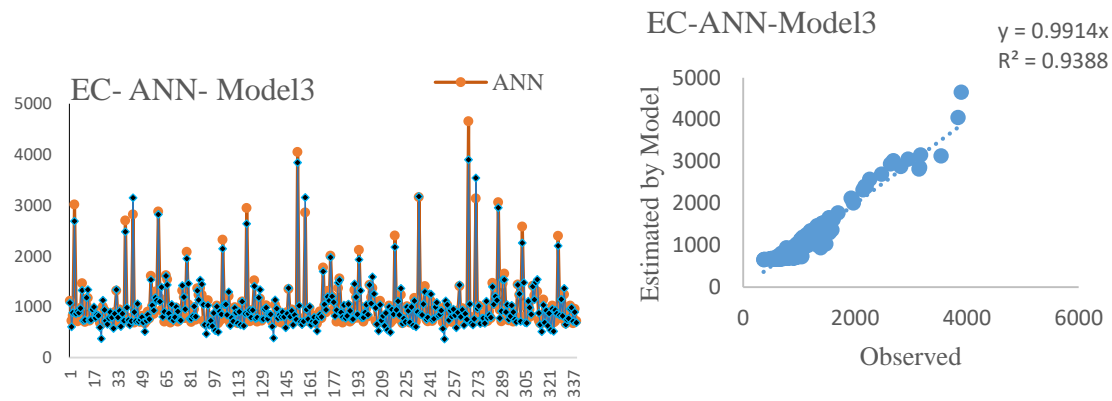
Comparison of the three methods (tables 2-4) reveals that the M5Tree model 1 performs superior to the other models in predicting EC parameter, while the RSM has the best accuracy in predicting SAR and TDS. In addition, in the case of model 3 having CL and Na inputs, the RSM acts better than the M5Tree and ANN in predicting EC, SAR, and TDS.



a)

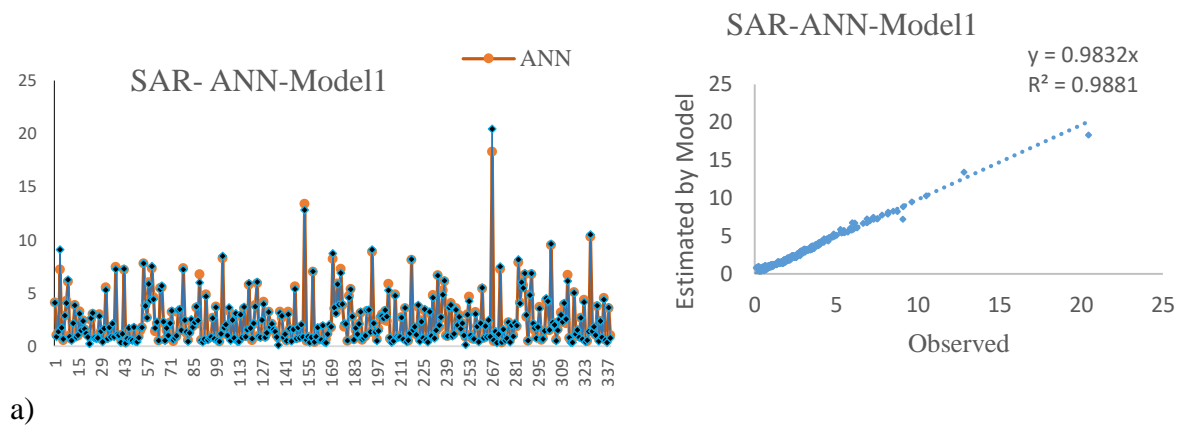


b)

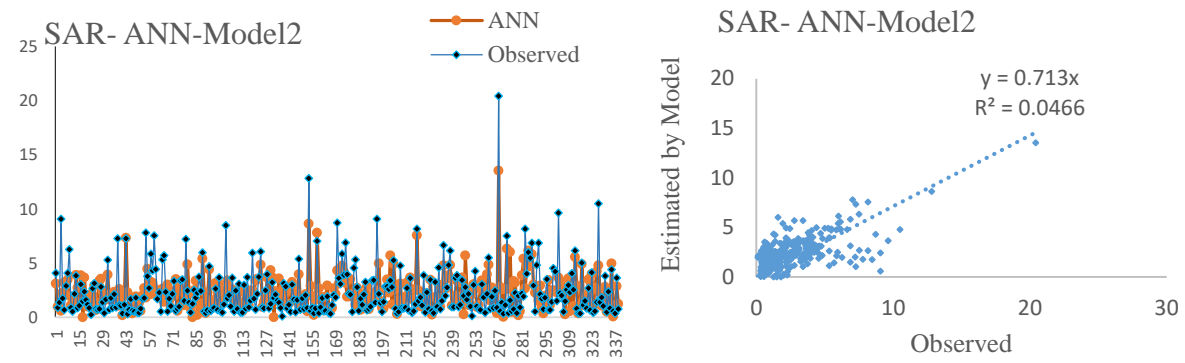


c)

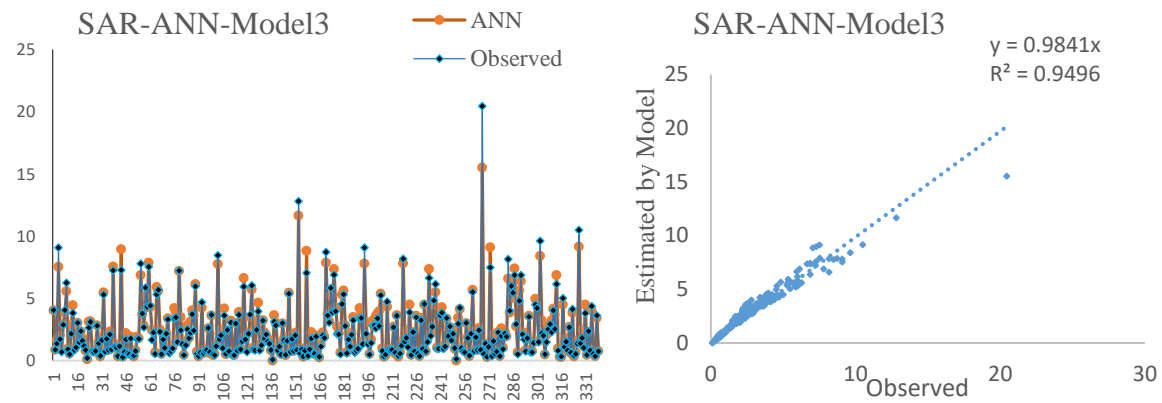
Figure 8-Graph of observed and predicted values of EC parameter using ANN



a)

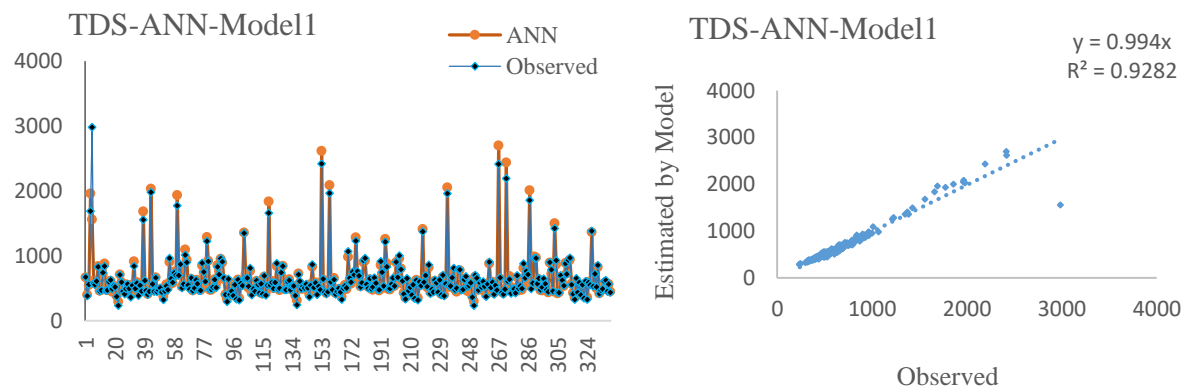


b)



c)

Figure 9- Graph of observed and predicted values of SAR parameter using ANN



a)

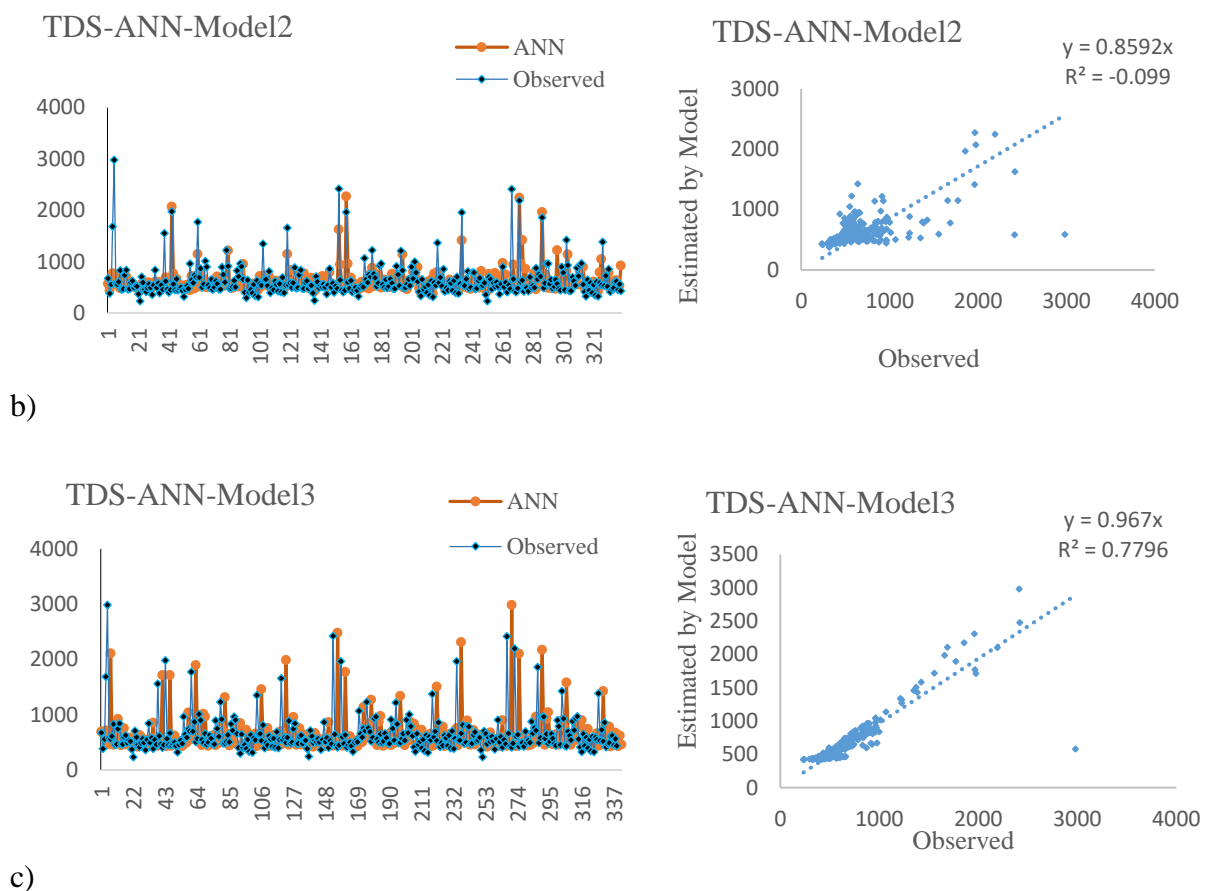


Figure 10-Graph of observed and predicted values of the TDS parameter using ANN

4. Conclusions

This study investigated the ability of artificial neural network (ANN), decision tree model (M5Tree), and response surface method (RSM) in predicting groundwater quality parameters (e.g., SAR, EC, and TDS) of Golestan province using data from 96 observation wells for the period of 19 years (1998-2016). (PH), chloride (CL), sulfate (SO₄), calcium (Ca), magnesium (Mg), and sodium (Na) were utilized as model inputs, and three input scenarios were implemented. The results obtained from this research showed that the models having all input parameters offered the best accuracy in predicting EC, SAR, and TDS. Also, the results showed that the Na and Cl parameters have the greatest effect on the accuracy of the prediction results for all three methods, and the input scenario without Na and Cl parameters provided very weak predictions. According to the full input models (model 1), the RSM was found to be superior to the other methods in predicting SAR and TDS, whereas the M5Tree performed better in predicting EC. On the other hand, the RSM model with limited inputs (Na and Cl) outperformed the other two methods, which had the same inputs in the prediction of all three parameters (EC, SAR, and TDS). The results of this study are consistent with the research of Najah et al (2009) and Yesilnacar et al (2008) which has shown that artificial intelligence models have high ability in estimating water quality parameters.

Author Contributions

All authors contributed equally to the conceptualization of the article and writing of the original and subsequent drafts.

Data Availability Statement

Data available on request from the authors.

If the study did not report any data, you might add “Not applicable” here.

Acknowledgements

The authors would like to thank all participants in the present study.

The authors thank all participants in this study.

Ethical Considerations

The authors avoided data fabrication, falsification, plagiarism, and misconduct.

Funding

The study was funded by the Gorgan University of Agricultural Science and Natural Resources, Iran, and Grant No. 3886. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

“The authors declare no conflict of interest.”

References

- Adimalla, N., Ravi Manne, Y. Z., Panpan, X., Hui, Q. 2022. Evaluation of groundwater quality and its suitability for drinking purposes in semi-arid region of Southern India: An application of GIS. *Geocarto International*, 25: 1–12.
- Guler, C., Kurt, M., Alpasalan, M., Akbulut, C. 2012. Assessment of the impact of anthropogenic activities on the groundwater hydrology and chemistry in Tarsus coastal plain (Mersin, SE Turkey) using fuzzy clustering, multivariate statistics and GIS techniques. *Journal of Hydrology*, 435–451.
- Haykin, S. 1999. *Neural networks: A comprehensive foundation*. Prentice-Hall, Upper Saddle River, NJ.
- Heshmatpour, A., & Sajadi, S. J. 2024. A survey on spatial and temporal variations of agricultural water quality in Gorganrood River using fuzzy rules. *Journal of New Approaches in Water Engineering and Environment*, 3(1): 1–20.
- Jalalkamali, A., Jalalkamali, N. 2018. Adaptive network-based fuzzy inference system-genetic algorithm models for prediction groundwater quality indices: A GIS-based analysis. *Journal of AI and Data Mining*, 6: 439–445.
- Kisi, O., Azad, A., Kashi, H., Saeedian, A., Hashemi, S. A., Ghorbani, S. 2019. Modeling groundwater quality parameters using hybrid neuro-fuzzy methods. *Water Resources Management*, 33: 847–861.
- Kumru, M. N., Bakac, M. K. 2003. R-mode factor analyses applied to the distribution of element in soils from the Aydin basin, Turkey. *Journal of Geological Exploration*, 77: 81–91.
- Marashi, S. E., Solaimani, K., Habibnejad, M., Jahanshahi, A. 2023. Investigating the quality of groundwater using statistical techniques and geographic information system in Haraz Plain. *Journal of Irrigation and Water Engineering*, 13(13): 248–268.

Mohammadrezapour, O., Kisi, O., Purahmadi, F. 2018. Fuzzy c-means and K-means clustering with genetic algorithm for identification of homogeneous regions of groundwater quality. *Neural Computing and Applications*, 32(8): 3763–3775.

Montgomery, D. C. 2001. Design and analysis of experiments. John Wiley and Sons, New York.

Najah, A., Elshafie, A., Karim, O. A., Jaffar, O. 2009. Prediction of Johor River water quality parameters using artificial neural networks. *European Journal of Scientific Research*, 28: 422–435.

Pourjabbar, A., Sarbu, C., Kostarelos, K., Einax, J., Buchel, G. 2014. Fuzzy hierarchical cross-clustering of data from abandoned mine site contaminated with heavy metals. *ScienceDirect*, 122–133.

Rizzo, D. M., Mouser, J. M. 2000. Evaluation of geostatistics for combined hydrochemistry and microbial community fingerprinting at a waste disposal site.

Sattari, M. T., Pal, M., Apaydin, H. 2013. M5 Model Tree application in daily river flow forecasting in Sohu Stream, Turkey. *Water Resources*, 4(3): 233–242.

Sivasankar, V., Kameswari, M., Msagati, T. A. M., Venkatarapathy, M., & Senthil Kumar, M. 2013. Fuzzy set approach—A tool to cluster groundwater quality parameters at Rameswaram, South India. *Journal of Water Resources and Ocean Science*, 33–39.

Solgi, E., Beigmohammadi, F., Ahmadvand, R., Baser, S. 2025. Nitrate risk assessment in drinking water in different areas of Malayer city. *Journal of New Approaches in Water Engineering and Environment*.

Thompson, D. 1982. Response surface experimentation. *Journal of Food Processing and Preservation*, 6: 155–188.

Vafakhah, M. 2012. Application of artificial neural networks and adaptive neuro-fuzzy inference system models to short-term streamflow forecasting. *Canadian Journal of Civil Engineering*, 39(4): 402–414.

Yesilnacar, M. I., Sahinkaya, E., Naz, M., Ozkaya, B. 2008. Neural network prediction of nitrate in groundwater of Harran Plain, Turkey. *Environmental Geology*, 56: 19–25.