ORIGINAL PAPER

The modeling and prediction of the quality of the groundwater resources in Tuyserkan plain using the optimized artificial neural network

Mohamad Parsi Mehr[1], Kamran Shayesteh[1,✉], Kazem Godini[2]

1. Department of Environmental Sciences, Faculty of Natural Resources and Environment, Malayer University, Malayer, Hamedan, Iran
2. Environmental Health Research Center, Research Institute for Health Development, Kurdistan University of Medical Sciences, Sanandaj, Iran

Date of submission: 07 Dec 2019, ***Date of acceptance:*** 25 Jan 2020

ABSTRACT

Tuyserkan plain is an important agricultural plain located in Hamadan province, Iran. Despite the severe decline of the water levels in aquifers, the quality of the plain has not been evaluated in recent years. The present study aimed to analyze the data of 15 wells during 12 years to evaluate the quality of groundwater in this area using the Wilcox diagram for the aquifer. The electrical conductivity (EC) of the plain was interpolated using the Kriging method to evaluate its spatial distribution since this parameter has caused a decline in the quality of the groundwater in the plain. According to the findings, the EC value was higher in the eastern parts of the plain and Tuyserkan city, which was described as the spatial distribution of the parameter. The Pearson's correlation-coefficient was used to assess the correlations between EC and other parameters. To predict and model the EC value, multi-layer perceptron artificial neural network (MLP-ANN) were used. According to the results of the Pearson's correlation-coefficient, the reduced number of the data led to the decreased expenditures of the experiments in obtaining the input data. The third model was finally obtained with the lowest number of the input parameters, low error, and high correlations between the predicted and actual data. In this model, two input parameters and five neurons were obtained in a hidden layer (R: 0.997, mean: 8.634, NRMSE: 0.05) using the momentum and hyperbolic tangent functions, indicating the high potential of MLP-ANNs in the prediction and modeling of groundwater quality.

Keywords: Artificial intelligence, Environmental assessment, Environmental pollution, Groundwater pollution

Introduction

It is a proven fact that water has played a pivotal role in the formation and survival of civilizations throughout history. The shortage of water resources is a major issue in the current era.[1] The concern regarding the quality of water has been severely felt within the past three decades of the 20th century, so that water quality has become as important as its quantity.[2] In some of the countries located in the north of Africa and the Middle East, groundwater is the only source of water for a large number of consumers.[3] The high rate of population growth, economic development, and urbanization have directly or indirectly affected the quality of groundwater in many countries,[4]

✉ *Kamran Shayesteh*
Ka_shayesteh@yahoo.com

Citation: Parsi Mehr M, Shayesteh K, Godini K. The modeling and prediction of the quality of the groundwater resources in Tuyserkan plain using the optimized artificial neural network. J Adv Environ Health Res 2020; 8: 107-118

giving rise to water pollution as a clear consequence of excessive water consumption. The contamination of water and reduction of water quality make it non-consumable, which not only adversely affects human health, but it also threatens socioeconomic development. Therefore, it is paramount to raise awareness of the qualitative condition of groundwater in the face of the challenges associated with the provision of water.[5]

The methods used to assess the quality of groundwater resources and discriminate proper situations to use water are vital issues in the planning of water resources. Computer models are a common approach to the evaluation of the changes in groundwater quality. It is critical to recognize the process of quality changes over time and predict the future status based on the qualitative behaviors in the past.[6] Previously, many efforts were made to predict the qualitative parameters under various scenarios using specific models. However, the statistical accuracy of these models was generally poor since natural systems are extremely complex for deterministic models.

The artificial neural network (ANN) is a fast and flexible tool to develop models for the estimation of water quality. Recently, ANNs have been employed as regression tools, showing exceptional performance, especially when used for the recognition of patterns and performance.[7] Extensive research has been focused on the modeling and prediction of the quality of groundwater resources,[8-12] generally indicating the high potential and accuracy of ANNs in the prediction and modeling of the quantity and quality of groundwater resources in different plains. In addition, ANNs are used as tools for environmental assessments, providing high accuracy in the outcomes and showing potential in the optimization and reduction of experimental expenditures. Tuyserkan plain is a major agricultural plain located in Hamadan province (Iran), where the largest amount of water is provided through the wells drilled in the plain. The plain has a semi-arid climate in one of the plains, which has extensively undergone the consumption of groundwater in recent years, which has led to the reduction of the groundwater levels by 22 meters. Despite the decline in the groundwater resources in Tuyserkan plain, adequate research has not been carried out in this regard.

The present study aimed to investigate the qualitative status and spatial distribution of an effective index in the quality of groundwater for agricultural use and introduce ANN as an optimized tool for the modeling and prediction of the quality of the groundwater resources in Tuyserkan plain.

Materials and Methods

Study area

Tuyserkan plain has a drainage basin of 805 square kilometers and is one of the plains located in the upper basin of Karkheh River in the south of Mount Alvand in Hamadan Province. The plain has 1,200 exploitive wells, 15 observation wells, and three piezometer wells. Most of the consumption on the exploitive wells is assigned to the agriculture. Fig. 1 depicts the location of the study area.

Applied data

In order to evaluate the quality of the groundwater resources in Toyserkan plain for agriculture and the modeling of the water quality data using the systematic network approach, 15 wells were sampled twice a year during 2005-2017. Late spring sampling was due to the end of the spring rainfall, and sampling at the end of summer was due to the end of water harvesting in the aquifer. This type of sampling shows the effects of the quantitative changes in water on the quality changes during one year. Sampling was performed using a standard method by Hamadan Regional Water Company. Table 1 shows the geographical position of the wells in the plain.

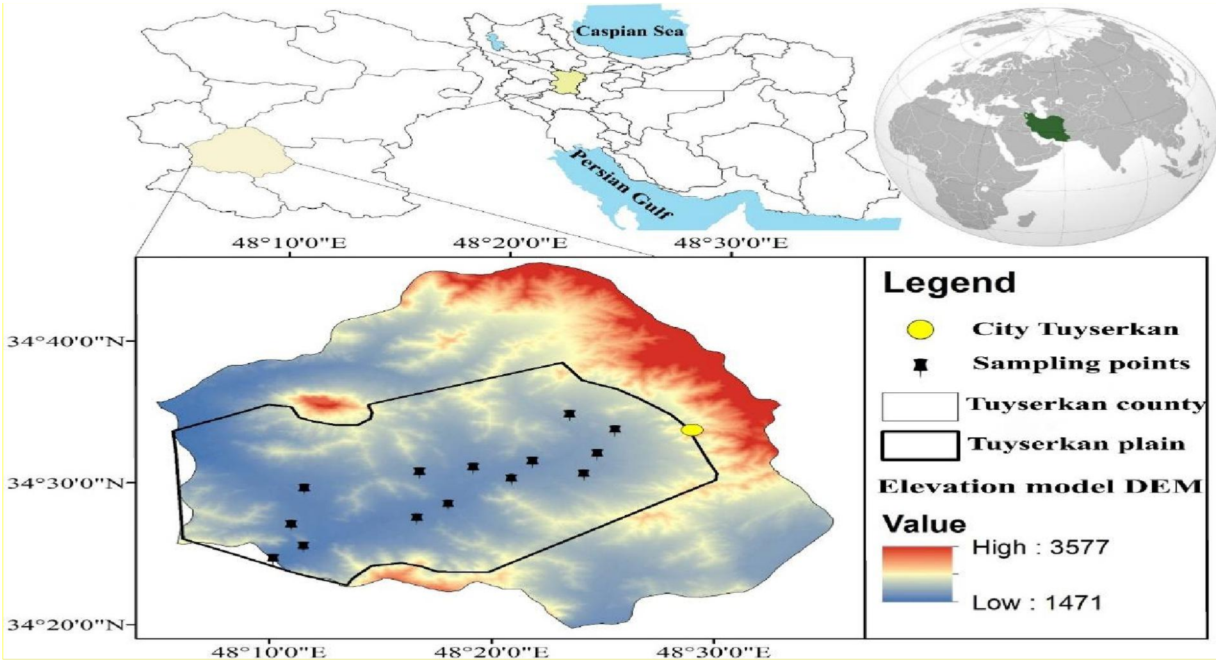

Fig. 1. Study area

Table 1. Geographical characteristics of sampled wells in Tuyserkan plain

Number	Place name	utmy	utmx
1	Mobarak Abad	3828272	263145
2	Korzan	3830256	260023
3	Rud Avar	3825136	261926
4	Noqadeh	3822514	260991
5	Arikan	3824138	257443
6	Mahmudabad	3821900	256000
7	Farasfaj	3822784	249654
8	Amir Abad	3818559	251630
9	Chasht Khoran	3816737	249478
10	Menjan	3811466	239549
11	Karkhaneh	3813085	241641
12	Motasemabad	3815891	240804
13	Deh Musa	3811466	239549
14	Anbar Naft	3823353	253362
15	Qaleh Now	3820630	241708

Groundwater quality

Initially, the groundwater resources of Tuyserkan plain that were used in agriculture were evaluated using the Wilcox diagrams given the importance of this qualitative assessment. The Wilcox method, which is also known as the U.S. Salinity Laboratory Method, is a common approach to the categorization of the agricultural water based on electrical conductivity (EC) and sodium adsorption ratio (SAR). Based on this approach, each parameter was divided into four classes and a total of 16 groups in terms of water quality.[13] The EC parameter was modeled using an ANN with the aim of the monitoring and management of the quality of groundwater resources. In order to decrease the experimental expenditures, the input parameters with the highest accuracy and lowest number were selected.

Statistical analysis

Data analysis was performed in SPSS using the Pearson's correlation-coefficient to assess the correlations between EC and other parameters. The correlations between the parameters indicated the strength of the correlations between the relative movements of the two parameters, which were selected for the modeling and prediction of EC based on the Pearson's correlation-coefficient. The data with the most significant correlations with each parameter were used for the optimized models.

Geostatistical analysis

The ArcGIS v10.3 software was used to obtain the map of spatial distribution. To obtain the spatial distribution map of EC with the optimal results, several techniques were applied, including inverse distance weighting, global/local polynomial interpolation, radial basis functions, and Kriging method. The Kriging method was selected due to lower error compared to the other approaches in this regard. In this method, a semivariogram model is used to determine the weight of function,[14] and the semivariogram is an automated correlation statistic, which is defined as the Eq. 1:[15]

$$\gamma(h) = \frac{1}{2N(h)} = \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (1)$$

In Eq. 1, $\gamma(h)$ is the semivariogram, $N(h)$ is the total number of the paired samples at distance h, $Z(x_i)$ shows the measured sample in point I, and $Z(x_i + h)$ is the measured sample in point $i+h$.

Artificial intelligence

If there was a precise definition of the problem, well-known applicable rules would be useful; however, when the recognition of a phenomenon is difficult, the use of known rules and methods may not be effective. The artificial intelligence system is constructed with the capability of learning, creativity, and flexibility similar to the human brain and neural computing methods. These methods basically rely on an advanced training system.[16] The overall structure of ANNs has three layers with specific tasks, including the input data layer in the network, the middle layer (hidden) for information processing, and the output layer that not only processes the network input, but it also shows the outputs. A neuron is the smallest unit in the ANN processor. Fig. 2 shows the overall structure of ANNs. In the present study, a multilayer neural network (MLP-ANN) with the algorithm of back-propagation was used. More than 90% of the ANN that is used to evaluate water resources involves the back-propagation algorithm.[17]

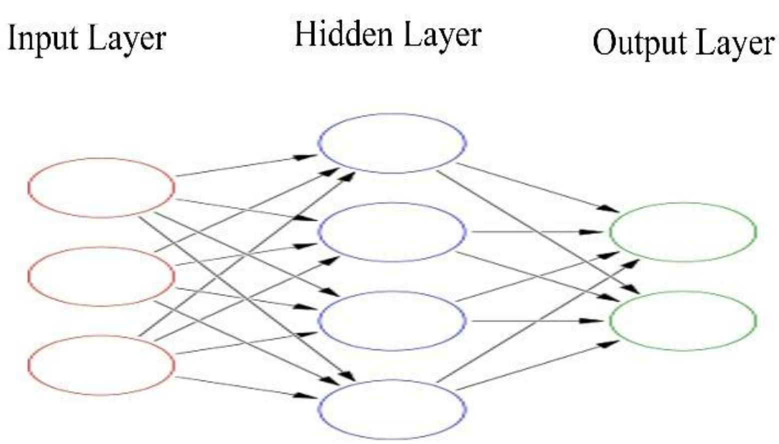

Fig. 1. ANN general structure

In the current research, network design was carried out based on a combination of the information on the parameters affecting groundwater quality in the past and in the form of different structures of data numbers in the input layer. In each structure, the post-processing input information is transmitted through the output way of the neurons in the first layer to those in the next layer and eventually to the output way of the network if it is acceptable. Otherwise, it returns to the previous layers, issuing the computational error and repeated calculations again to obtain an acceptable result.[2] In the present study, normalized data were used as the network input to increase the speed of information processing and prevent the disruption of the network in the local minimums. The implementation of these structures was conducted in the NeuroSolution software under Windows operating system, which is able to normalize data. Another key advantage of the software is the different functions with various algorithms in the software bank.[18]

For ANN modeling, a percentage of the data was initially used for network training. Following that, the remaining percentage of the data was used for validation and network testing. In the current research, 60% of the data were used for training, and 40% were used for validation and network testing. In addition, the predicted data were compared with the actual data, and the error value was calculated as well.

The error values should be at their lowest, which requires the designing of the desired network. Moreover, the practice of training and experimenting should be repeated with different modes to minimize the error. In order to evaluate and compare the results of the correlation-coefficient (R), the mean absolute error (MAE) and relative normal error (NRMSE) were calculated using the following Eqs.

$$MAE = \frac{1}{n}\sum_{i}^{N} = 1\,|xi - yi| \quad (3)$$

In Eq. 3, xi shows the actual data, yi is the estimated data, and n is the total number of the data.

$$R = \frac{\sum_{i=1}^{n}(Y\,act - \hat{Y}\,act)\,(Y\,est - \hat{Y}\,est)}{\sqrt{\sum_{i=1}^{n}\,(Y\,act - \hat{Y}\,act)^2\,\sum_{i=1}^{n}(Y\,est - \hat{Y}\,est)}} \quad (4)$$

In Eq. 4, $Y\ act$ represents the actual values, $\hat{Y}\ act$ is the mean actual values, $Y\ est$ shows the estimated value, and $\hat{Y}\ est$ is the mean estimated values.

$$NRMSE = \frac{RMSE}{EC\ average} \quad (5)$$

In Eq. 5, *RMSE* represents the error, and *EC average* shows the mean EC obtained in the experiments. The correlation-coefficient (R) indicated the amount of the correlation between the predicted results and actual values. In Eq. 5, it is clearly observable that when the value is closer to one, the results are more acceptable, and closer RMSE and MAE to zero indicate lower error.

The order of using the parameters in each model was in accordance with the correlations in Table 5, which were applied in the experiments based on convenience and cost. According to the information in Table 2, the number of the parameters reduced.[19]

Table 2. Neural network structure

	Input parameter	Output parameter
Model 1	SO_4	EC
	Cl	
	HCO_3	
	TH	
	SAR	
	TDS	
	K	
	Na	
	Mg	
	Ca	
Model 2	TDS	EC
	TH	
	Mg	
	Ca	
Model 3	TDS	EC
	TH	

Results and Discussion

Data dispersion indicators

The data used in the present study included EC, total dissolved solids (TDS), SO_4, Cl, HCO_3, TH, SAR, potassium, magnesium, and calcium. Different values such as the maximum, minimum, mean, and standard deviation were also calculated using SPSS software (Table 3).

Table 2. Dispersion of used data

Parameter	Maximum	Minimum	Mean
E.C	1046.00	357.00	548.99
TDS	679.90	228.48	351.22
HCO_3	6.90	2.00	4.27
CL	2.70	0.30	0.78
SO_4	3.40	0.09	0.50
Ca	6.20	1.50	2.98
Mg	4.00	1.00	1.88
Na	2.66	0.21	0.72
K	0.05	0.01	0.02
SAR	1.35	0.15	0.45
TH	455.00	150.00	243.23

Groundwater quality

The results of the Wilcox's diagram (Fig. 3) indicated that the groundwater of Toyserkan plain in Naqadeh and Roodavar regions was saline and consumable for agriculture, while it was relatively saline and proper for agricultural use in the remaining areas of the plain. In general, 86.67% of the groundwater of Toyserkan had the C2S1 symbol (slightly saline), and 13.33% had the C3S1 symbol (saline). The concentration of the SAR parameter was observed to be low, while the concentration of the EC parameter was moderate/high. Therefore, the EC parameter was considered to be the main cause of the reduced quality of the groundwater resources in the plain. Table 4 shows the characteristics and quality of water for agricultural use.

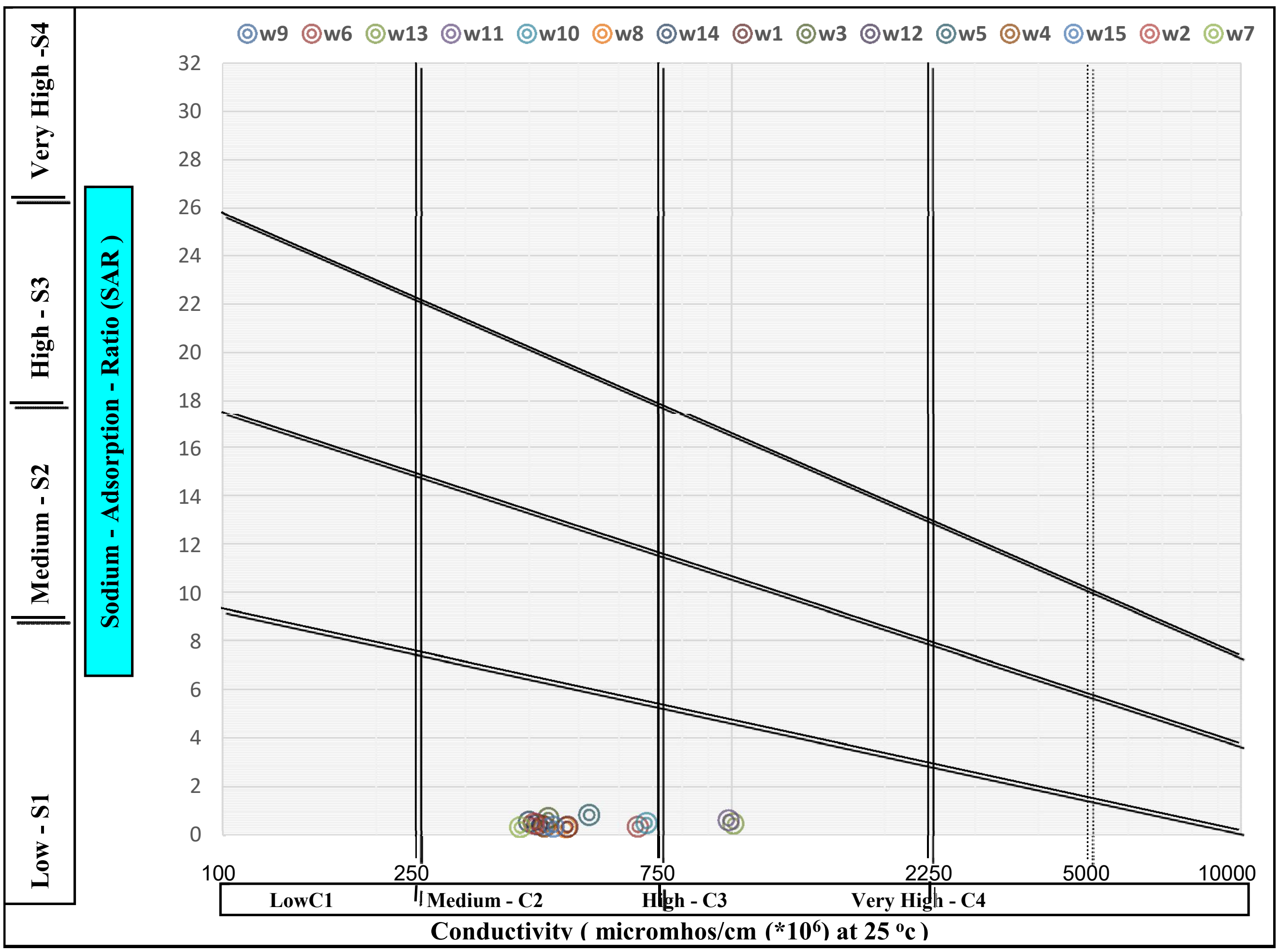

Fig. 2. Wilcox diagram

The assessment of various stations clearly indicated that the closer stations to Tuyserkan city had higher EC values.

Table 3. Results of Wilcox diagram for each point of Tuyserkan plain

Number	Place name	Quality classes	EC	SAR	Symbol	Quality for agricultural use
1	Mobarak Abad	C2-S1	425	0.37	w9	Slightly salty - suitable
2	Korzan	C2-S1	654	0.33	w6	Slightly salty - suitable
3	Rud Avar	C3-S1	1007	0.46	w13	Salted - Can be used
4	Noqadeh	C3-S1	986	0.61	w11	Salted - Can be used
5	Arikan	C2-S1	680	0.49	w10	Slightly salty - suitable
6	Mahmudabad	C2-S1	473	0.3	w8	Slightly salty - suitable
7	Farasfaj	C2-S1	400	0.54	w14	Slightly salty - suitable
8	Amir Abad	C2-S1	476	0.33	w1	Slightly salty - suitable
9	Chasht Khoran	C2-S1	436	0.7	w3	Slightly salty - suitable
10	Menjan	C2-S1	415	0.43	w12	Slightly salty - suitable
11	Karkhaneh	C2-S1	525	0.82	w5	Slightly salty - suitable
12	Motasemabad	C2-S1	429	0.33	w4	Slightly salty - suitable
13	Deh Musa	C2-S1	447	0.34	w15	Slightly salty - suitable
14	Anbar Naft	C2-S1	409	0.46	w2	Slightly salty - suitable
15	Qaleh Now	C2-S1	384	0.31	w7	Slightly salty - suitable

In the current research, the Pearson's correlation-coefficient was used to examine the parameters affecting the EC parameter concentration. The parameters with high correlation-coefficients represented the controller and common origin of the parameters. According to the results of the Pearson's correlation-coefficient (Table 5), all the parameters used in the research had significant correlations with EC, while the parameters of TDS, TH, magnesium, and calcium were most effective with the coefficients of 0.998, 0.977, 0.911, and 0.906, respectively.

Table 4. Results of Pearson's correlation-coefficient

Parameter		E.C	T. D. S	HCO_3	CL	SO_4	Ca	Mg	Na	K	TH	S. A. R
E.C	Pearson Correlation	1	0.998^{**}	0.871^{**}	0.817^{**}	0.856^{**}	0.906^{**}	0.911^{**}	0.746^{**}	0.652^{**}	0.977^{**}	0.559^{**}

Geostatistics

As can be seen in Fig. 4, the results of the interpolation of EC in the studied area using the Kriging method indicated that the highest and lowest values were in the eastern part (near Tuyserkan city) and western part of the plain, respectively. In these regions, the examination of the other parameters showed that the maximum increase and decrease in EC occurred, along with the parameters of TDS, TH, magnesium, and calcium, highlighting the importance of the performed modeling.

MLP-ANN

Regarding the selection of the inputs and outputs of the three models based on the Pearson's correlation-coefficient (Table 5) with different the training and testing, the optimal outcomes were obtained for the number of the hidden layers, training law, and actuator function (Table 6). In addition, the momentum training function and tangent hyperbolic function yielded the best results in the three models. Through fixing the neural network model and reducing the input parameters, the error resulting from the model did not increase significantly, while the results of the model had acceptable accuracy only through changing the number of the mid-layer processors, so that the number of the input parameters would decrease to two in the third model, rendering the results of the model acceptable.

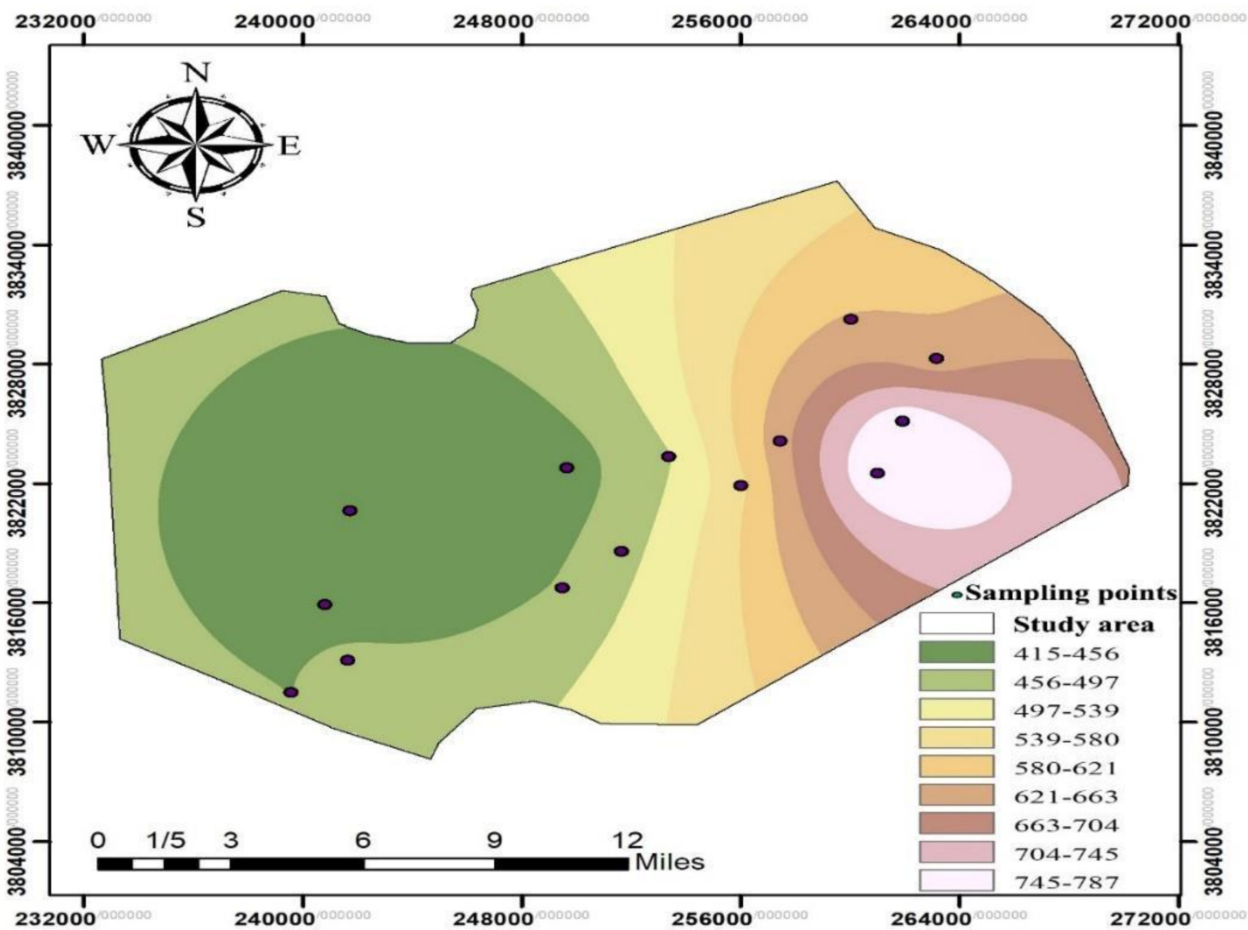

Fig.3. Interpolation of EC values in Tuyserkan plain (micromhos/cm)

Table 6. Results of MLP-ANN models with their structures

	Model		
	1	2	3
Number of input neurons	10	4	2
Teaching law	Momentum	Momentum	Momentum
Actuator function	TanhAxon	TanhAxon	TanhAxon
The number of hidden layers	1	1	1
The number of processors intermediate layer	10	4	5
R	0.998	0.997	0.997
NRMSE	0.004	0.004	0.005
MAE	9.178	8.056	8.634

Fig. 5 shows the dispersion diagrams of the predicted results. In these graphs, the horizontal and vertical axes represent the actual and predicted data, respectively. Moreover, the correlation between the actual data and predicted data in these charts are shown by an equation and standard deviation from the first-half bisector. If the data were closer to the one-

to-one graphs, it would indicate the high performance and accuracy of the constructed models.
s

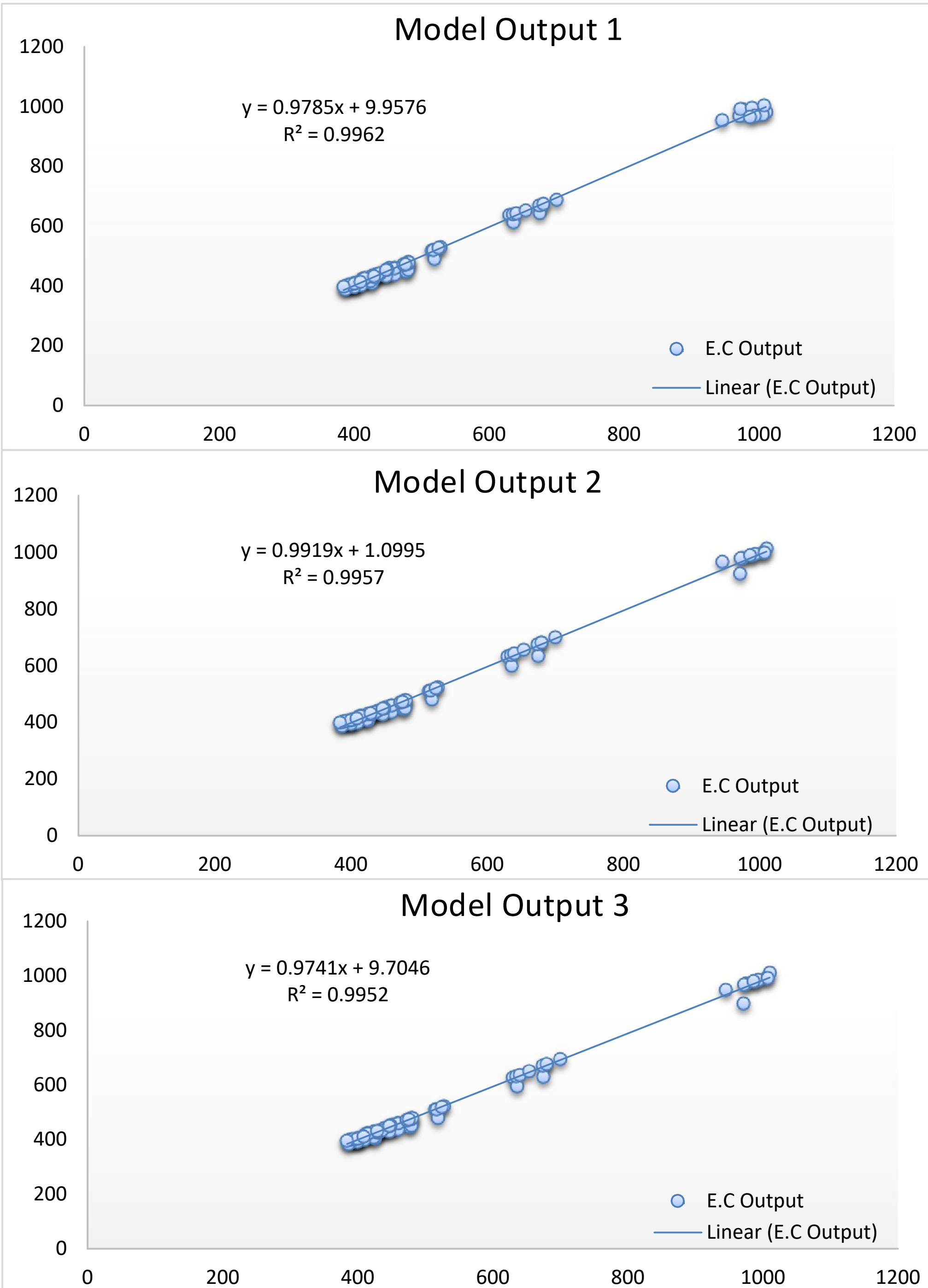

Fig. 4. Scatter plot of predicted and observed EC in models

Fig. 6 depicts the overlap graph between the predicted and actual values. In these diagrams, the horizontal and vertical axes indicated the sample number and EC values, respectively. As is shown in these graphs, the predictive accuracy of EC in the three models was higher than 99%. Therefore, the graphs of the actual values and predicted values completely overlapped. Evidently, the higher overlap between the two charts resulted in the higher accuracy and efficiency of the models.

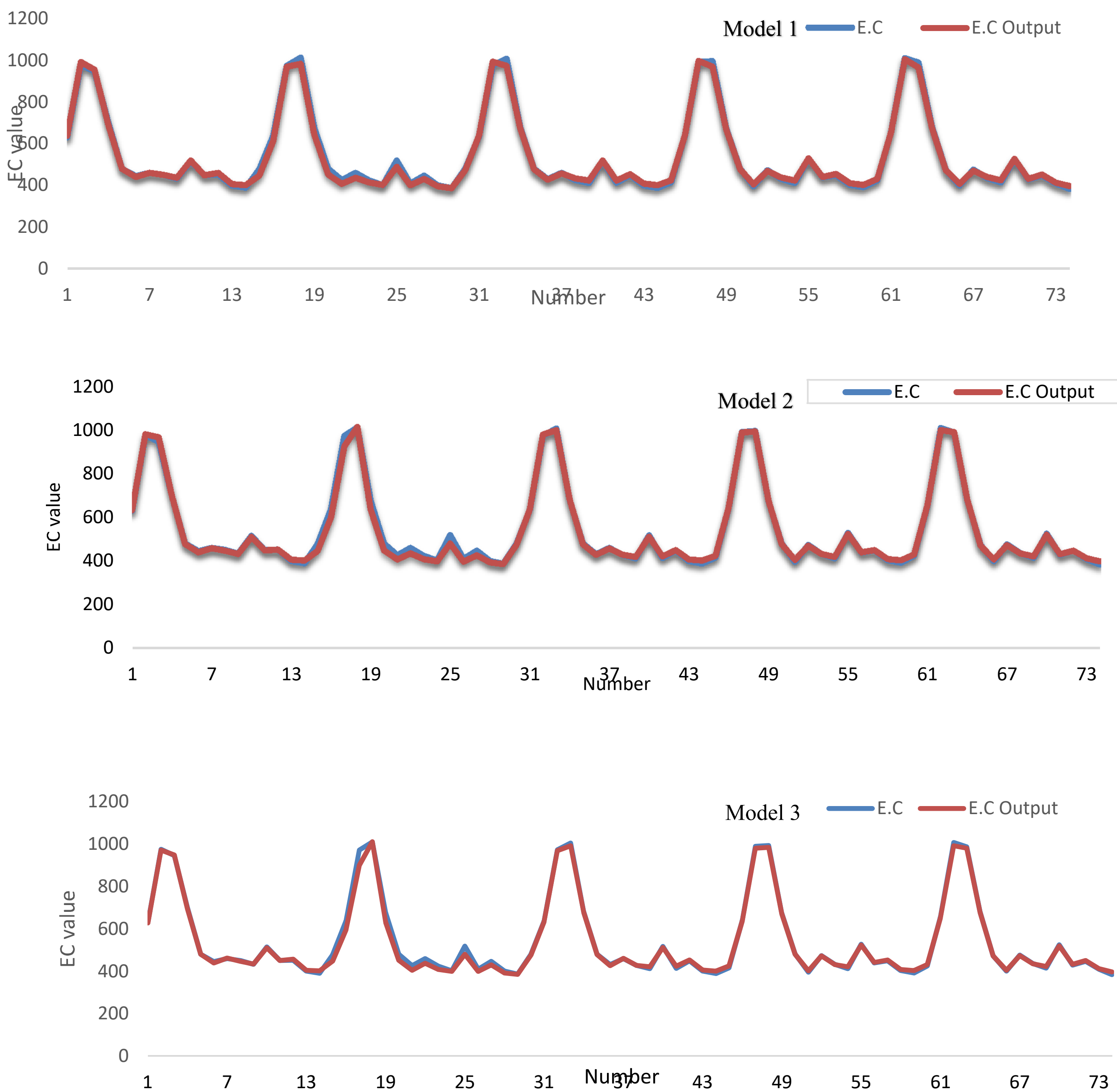

Fig. 5. Overlap graph of predicted values by models with actual magnitudes

Conclusion

According to the results, the high EC parameter caused the loss of water quality in Toyserkan plain for agricultural applications. In terms of the spatial distribution of the EC parameter in the groundwater of the plain, the closer areas to Toyserkan city had higher EC values, which could be due to the urban sewage penetration into the plain aquifer. The reduced quality caused by the high EC and its specific distribution in the aquifer highlighted the importance of the current research. Furthermore, it was observed that the model has the ability and high precision to model and predict the quality of groundwater resources.

The results obtained from various optimized ANN models indicated that all the optimized structures based on the number of the input parameters had very high accuracy for the prediction of the EC parameter. With two input parameters, the third model had the least expenditure despite the high precision and acceptable error in the prediction of EC. This model had the momentum training function and hyperbolic tangent function with two inputs and five neurons in a hidden layer, which resulted in the more significant correlation of the outcomes with the actual real data by more than 99%; this confirmed the high accuracy and significance of the obtained results. Therefore, it could be claimed that MLP-ANN could be used as an efficient tool by the managers and planners of water resources owing to high accuracy and efficiency in the modeling and prediction of water quality parameters.

References

1. Taghizadeh Mehrjardi R, Zareian Jahromi M, Mahmodi Sh, Heidari A. Spatial distribution of groundwater quality with geostatistics (Case study: Yazd-Ardakan Plain). World Appl Sci J 2008; 4 (1): 09-17.
2. Parsimehr M, Shayesteh K, Godini K, Bayat Varkeshi M. Using multilayer perceptron artificial neural network for predicting and modeling the chemical oxygen demand of the Gamasiab River. Avicenna J Environ Health Eng 2018; 5(1): 15–20 .
3. Salman SA, Shahid S, Mohsenipour M, Asgari H. Impact of landuse on groundwater quality of Bangladesh. Sustain Water Resour Manag 2018; 4(4): 1031–6.
4. Lutz A, Thomas JM, Keita M. Effects of population growth and climate variability on sustainable groundwater in Mali, West Africa. Sustainability 2011; 3(1): 21–34.
5. Shah T, Roy AD, Qureshi AS, Wang J. Sustaining Asia’s groundwater boom: An overview of issues and evidence. Natural Resources Forum 2003; 27(2): 130-141.
6. Tahmasebi AR, Zomorrodian SMA. Estimation of soil liquefaction potential using artificial neural network. Second National Student Conference on Water and Soil Resources 2004 . (In Persian)
7. Sarkar A, Pandey P. River water quality modelling using artificial neural network technique. Aquat Procedia 2015; 4: 1070–7.
8. Jalalkamali A, Jalalkamali N. Adaptive network-based fuzzy inference system-genetic algorithm models for prediction groundwater quality indices: A GIS-based analysis. Journal of AI and Data Mining 2018; 6(2):439–45.
9. Alizamir M, Sobhanardakani S. Forecasting of heavy metals concentration in groundwater resources of Asadabad Plain using artificial neural network approach. J Adv Environ Health Res 2016; 4(2):68–77.
10. Azimi S. Azhdary Moghaddam M. Hashemi Monfared SA. Prediction of annual drinking water quality reduction based on groundwater resource index using the artificial neural network and fuzzy clustering. J Contam Hydrol 2019; 220: 6–17 .
11. Soltani Mohammadi A, Sayadi Shahraki A, Naseri AA. Simulation of groundwater quality parameters using ANN and ANN+ PSO Models (Case Study: Ramhormoz Plain). Pollution 2017; 3(2):191–200.
12. Mehrdadi N, Hasanlou H, Jafarzadeh MT, Hasanlou H, Abdolabadi H. Simulation of low TDS and biological units of Fajr industrial wastewater treatment plant using artificial neural network and principal component analysis hybrid method. J Water Resource Prot 2012; 4(6): 370-6.

13. Wilcox Lv. The quality of water for irrigation use. 1948.
14. Webster R, Oliver Ma. Geostatistics for environmental scientists. John Wiley & Sons; 2007.
15. Mabit L, Bernard C. Assessment of spatial distribution of fallout radionuclides through geostatistics concept. J Environ Radioact 2007; 97(2-3):206–19.
16. Ostad-Ali-Askari K, Shayannejad M, Ghorbanizadeh-Kharazi H. Artificial neural network for modeling nitrate pollution of groundwater in marginal area of Zayandeh-Rood River, Isfahan, Iran. Ksce Journal of Civil Engineering 2017; 21(1):134–40.
17. Braddock RD, Kremmer ML, Sanzogni L. Feed-Forward artificial neural network model for forecasting rainfall run-off. Environmetrics1998; 9(4):419–32.
18. Asadpour G, Nasrabadi T. Municipal and medical solid waste management in different districts of Tehran, Iran. Fresenius Environ Bull 2011; 20(12): 3241–5.
19. Zare Abyaneh H, Bayat Varkeshi M, Maroufi S, Amiri Chaijan R. Evaluation of artificial neural network and adaptive neuro fuzzy inference system in decreasing of reference evapotranspiration parameters. Journal of Water and Soil (Agricultural Science and Technology) 2010; 24(2): 297-305. (In Persian)