Advances in Environmental Technology 2 (2017) 109-117

Advances in Environmental Technology

journal homepage: http://aet.irost.ir

Artificial neural network modeling for predicting of some ion concentrations in the Karaj River

Kamyar Movagharnejad*, Alireza Tahavvori, Forogh Moghaddam Ali
Babol Noushiravani University of Technology, Babol, Iran

A R T I C L E I N F O

Article history:
Received 30 November 2016
Received in revised form
15 August 2017
Accepted 25 August 2017

Keywords:
Ca concentration
Karaj river
Artificial neural network
Prediction

A B S T R A C T

The water quality of the Karaj River was studied through collecting 2137 experimental data set gained by 20 sampling stations. The data included different parameters such as T (temperature), pH, NTU (turbidity), hardness, TDS (total dissolved solids), EC (electrical conductivity) and basic anion, cation concentrations. In this study a multi-layer perceptron artificial neural network model was designed to predict the calcium, sodium, chloride and sulfate ion concentrations of the Karaj River. 1495 data set were used for training, 321 data set were used for test and 321 data set were used for validation. The optimum model holds sigmoid tangent transfer function in the middle layer and three different forms of the training function. The root mean square error (RMSE), mean relative error (MRE) and regression coefficient (R) between experimental data and model's outputs were measured for training, validation and testing data sets. The results indicate that the ANN model was successfully applied for prediction of calcium ion concentration.

1. Introduction

Freshwater sources such as lakes, rivers and streams are parts of a complex interconnected system which is of vital importance for the earth's ecosystems and human societies. Growing populations and the limitations of natural fresh water sources in the planet have necessitated the better management of these limited resources. Pure natural water is very rare. Even the rainwater contains some impurities such as dissolved gases and traces of mineral and organic compounds. Natural waters may also contain mineral and organic matter from the soil [1]. Mineral materials may include calcium, sodium, chloride and sulfate ions. Calcium, in the form of the Ca^{2+} ion, is one of the major inorganic cations existing in natural waters, originating from streams flowing over limestone, $CaCO_3$, gypsum, $CaSO_4 \bullet 2H_2O$, and other calcium-containing rocks [2]. Most natural waters contain less than 20 mg of sodium per liter. In addition, water-treatment chemicals may rise the sodium levels up to 30 mg/liter. Chloride may originate from different sources, such de-icing salts, inorganic fertilizers, animal feeds, industrial effluents, and so on. It is reported that the typical sulfate levels in fresh water are near 20 mg/liter in average and may range from 0 to 630 mg/liter in rivers, 2 to 250 mg/liter in lakes, and 0 to 230 mg/liter in the underground water [3]. The collection of the relevant information about the natural water supplies is called the water quality monitoring [4]. The qualification and characterization of the natural waters may be considered as an essential step before any realistic planning, process design and optimization for the limited freshwater resources, but as the experimentation is not always possible, the predictive methods are widely used. The relative importance of water quality has acted as a motivation for several other researchers to propose mathematical models for prediction of this property. McCleskey et al. developed a model in which predicted the natural water's electrical conductivity with a mean relative error of about 7% [5]. Marandi et al. predicted the EC of the ground water with similar accuracy [6]. Artificial neural

*Corresponding author. Tel: +981132334204
E-mail address: movagharnejad@yahoo.com
DOI: 10.22104/AET.2017.1802.1084

network models (ANN) have been widely used to describe the versatile phenomena such as drying of agricultural products [7], prediction of thermodynamic properties of materials [8] and forecasting of the oil prices [9]. This mathematical tool has been also used to predict water quality. This technique is used in modeling the different parameters of natural waters. Zare Abyaneh et al. predicted the nitrate parameter using an artificial neural network [10]. They also used the same technique to forecast the underground water level in Malayer Plain [11]. Mehrdadi et al. developed an ANN model to predict the TDS of the waste waters from Fajr refinery located in Boushehr, south of Iran [12]. Moghaddam Ali and Movagharnejad introduced an ANN model to predict the EC of the Jajrud River [13]. They also claimed that this model proved to be more successful than other conventional mathematical methods. The objective of this research is to develop a new Artificial Neural Network model to predict some of the mineral ion concentrations for this important natural water resource, locating near the capital city of Tehran.

The geology of the Karaj water basin

This water basin which is placed in the southern slopes of the Central Alborz Mountains between Bilghan to Dizin, is located between 51^0 to 51^0 35' East longitudes and 35^0 5' to 36^0 11' North latitudes. The average height of this basin is about 1600 m from the free Sea level. 61% of the area of this basin is located at the heights higher than 2500 meters. The minimum height of the Karaj river basin was about 1320 meters and the maximum height being equal to 4000 m in the north. The Karaj River with a length of 75 km, width of 8 to 15 m and depth of 1-2 m, is considered as one of the most important rivers flowing in the South Alborz Mountains. The average flow of the river is about $17 m^3/s$. The water flow is higher in the winter and early spring and lower in the summer and autumn. The river finally empties in the Salt Lake of Ghom in the Central Iran.

Karaj River and Karaj dam have always been considered as one of the main sources of drinking and farming water in the Tehran province. Therefore, the Tehran water and sewage company has established several sampling stations alongside the river route from the springs up to the Karaj dam. The samples are gathered monthly by the well-equipped groups, and then the picked samples were transferred to the laboratory to be examined according to the technical rules and standard procedures. The results then were recorded and archived in three, six and twelve (annual) month formats. 2137 measured data of 20 stations on Karaj River were collected from the official reports of the Tehran Water and Sewage Company. These data which belonged to the different seasons of the year are briefly described in the Table 1.

Table 1. Description of the sampling data

Parameter	Max	Min
Temperature (℃)	24	0
pH	8.84	7.38
NTU	562	0.3
Total Hardness	328	88
TDS (mg/L)	664.95	124
Ca^{2+} (mg/L)	86.4	22.4
Na^+ (mg/L)	40	5
Cl^- (mg/L)	110	2
So_4^{2-} (mg/L)	145	15
NO_2^- (mg/L)	0.34	0
EC (μS/cm)	729.6	212

The locations of the sampling stations are also displayed in the Figure 1.

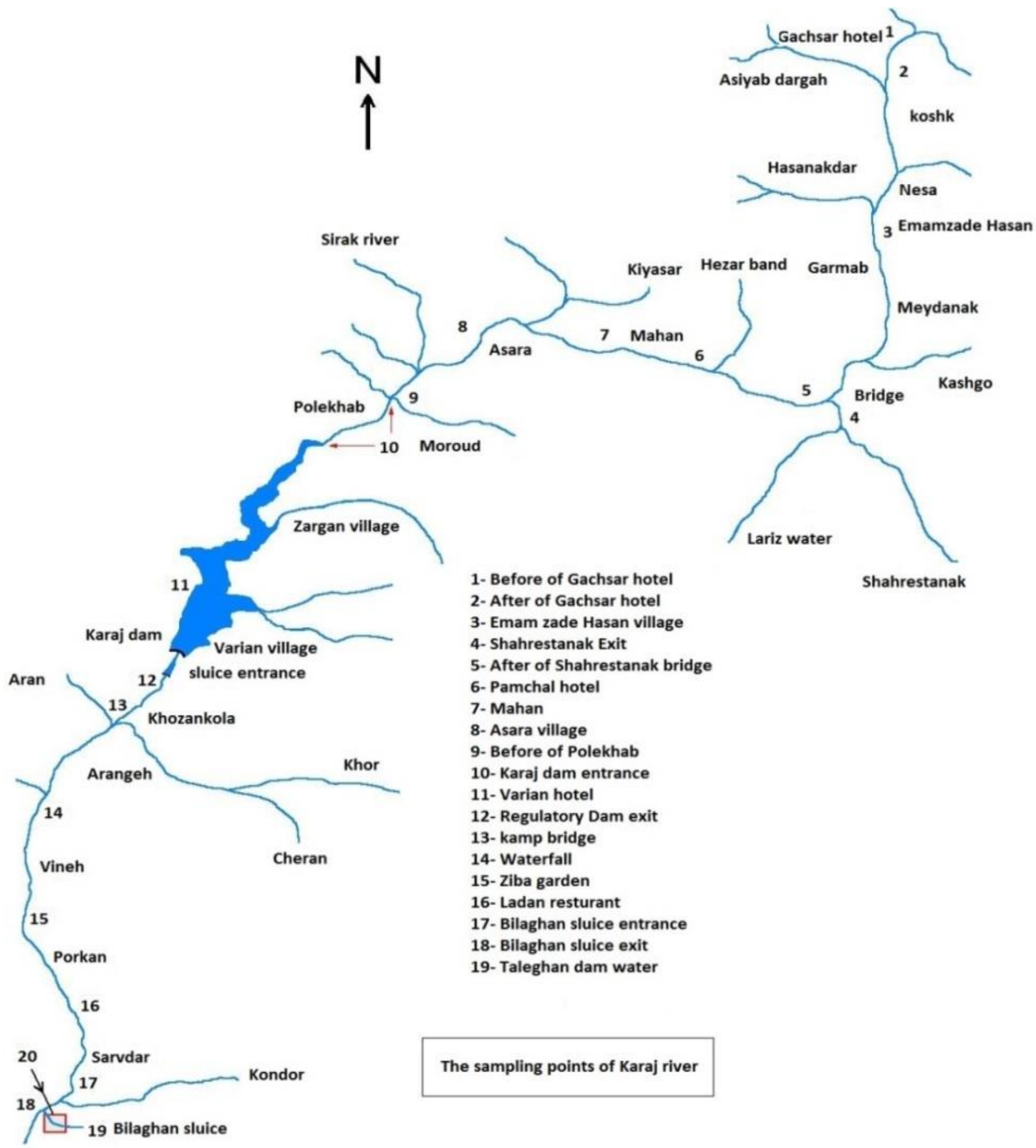


Fig. 1. Location of sampling stations on the Karaj River

The ANN modeling

The collected data bank of the Karaj River data was modeled by the MATLAB's Artificial Neural Network toolbox. The ANN architecture of the Multilayer perceptron (MLP) with Back propagation (BP) algorithm was used in this study. The link between input elements of the first layer and those of the last layer is shown in the Figure.2.

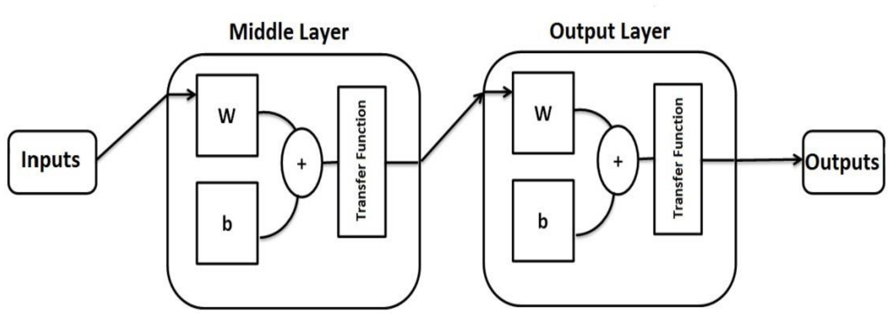


Fig. 2. Artificial neural network model used in the research

The input data are processed and transferred by the first input layer of neurons to the next layers and finally to the net outlet. The computed error would be used to train the neurons and finding better results. This process will continue until the ultimate results are found [14]. The collected data were randomly divided into three distinct groups: 70% for training, 15% for testing and 15% for validating. The tangent transfer function of the sigmoid was used in the middle layer and the linear transfer function in the outer layer. The training functions of Levenberg-Marquardt and conjugate gradients were used to train the ANNs. 7 independent ANN models with the different input and output parameters and different training functions were used in this work which are briefly described in the Table 2.

Table 2. Parameters investigated in 6 models.

Model	Input parameters	Hidden layer Neurons	Output parameter	Training Function
A1	Temperature, NTU, pH, Hardness, TDS, EC	8	Ca^{2+} concentration	**trainscg**
A2	Temperature, NTU, pH, Hardness, TDS, EC	16	Ca^{2+} concentration	**traincgb**
A3	Temperature, NTU, pH, Hardness, TDS, EC	12	Ca^{2+} concentration	**trainlm**
A4	Hardness, TDS, EC	10	Ca^{2+} concentration	**trainlm**
A5	Temperature, NTU, pH, Hardness, TDS, EC	18	Na^{+} concentration	**trainlm**
A6	Temperature, NTU, pH, Hardness, TDS, EC	10	Chloride concentration	**trainlm**
A7	Temperature, NTU, pH, Hardness, TDS, EC	10	Sulfate concentration	**trainlm**

The transfer function of all of the models was selected to be the "tansig" and the optimized number of neurons in the middle layer was varied from 8 neurons to 18 neurons for different ANN models.Various criteria are selected for the acceptance of the predicted results in each individual ANN model. Root mean square error (RMSE), regression max (R) and Mean relative error (MRE) which are shown in the equations 1 and 2 and were used in several previous works [15,16] are selected as the criteria in this work.

$$\mathrm{RMSE} = \sqrt{\frac{\sum(\mathrm{N\ actual} - \mathrm{N\ forecast})^2}{\mathrm{n}}} \quad (1)$$

$$\mathrm{MRE} = \frac{\mathrm{N\ actual} - \mathrm{N\ forecast}}{\mathrm{N\ actual}} \quad (2)$$

Which N_{actual} is the experimental ion concentrations value, $N_{forecast}$ is the predicted ion concentrations value and n is the number of experimental data.

2. Materials and methods

In this study, 7 different ANN models were developed. At first, 6 parameters of EC, TDS, temperature, pH, hardness and turbidity were selected as the input variables and 6 ANN model were developed with the change of outputs and training functions. For A_1, A_2 and A_3 models, the output variable is Calcium concentration and the training functions are Scaled Conjugate Gradient (scg), Conjugate Gradient Backpropagation (cgb) and Levenberg- Marquardt (lm), respectively. And for A_5, A_6, and A_7 models the training function is constant (Levenberg- Marquardt) and the output variables are sodium, chloride and sulfate, respectively. To reduce the number of input variables, we have designed the A_4 model with Levenberg- Marquardt training function and three input variables of EC, TDS and hardness with the calcium concentration as the output of the model. The number neurons in the hidden layer of each ANN model were optimized by varying the number of neurons for each model from 1 to 25 and selecting the number of neurons with the best statistical measures listed in the equations (1) and (2).

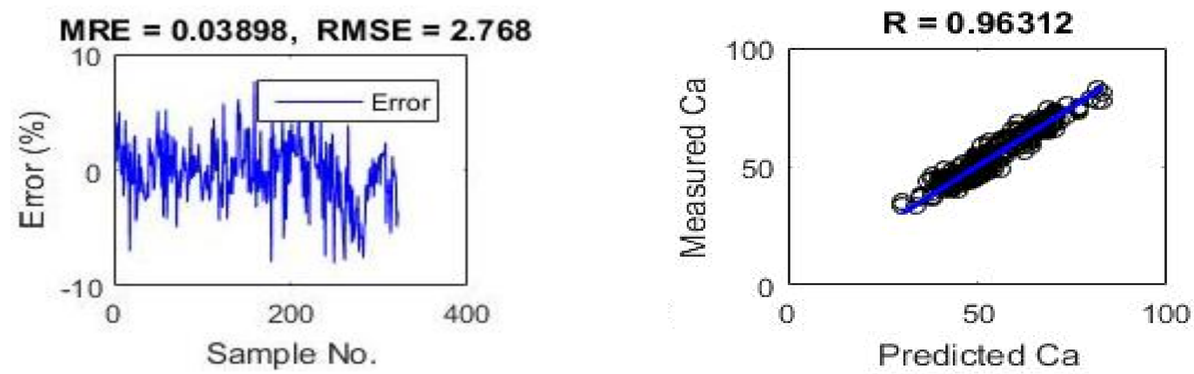
3. Results and discussion

The results of 7 different ANN models for training, test and validation data have been compared with the experimental data and the results have been summarized in Table 3.

It is clear that the models with the Levenberg- Marquardt training function (trainlm) are more accurate than those with scaled conjugate gradient (trainscg) and conjugate gradient back propagation (traincgb) training functions. The models with Ca concentration as output give the best results and after that the models with sulfate concentration, sodium concentration and chloride concentration as the output, respectively. The results show that the models that predicts the calcium concentration are quite successful. There is a logical rule in science that if you have two models with identical accuracy, then the simpler model with fewer inputs may be selected. For Ca concentration, the models A3 and A4 are shown to be more accurate, but the model A4 is much simpler and does not require 3 inputs of temperature, NTU and pH. Consequently, the model A4 with the minimum number of input variables, highest R and lowest RMSE and MRE would be selected as the best model for prediction of Ca concentration of the Karaj River. The results of the 7 proposed models have been shown in figures 3 through 9. Part A of each figure shows the amount of errors for each model and the overlap of experiments and the model outputs are displayed in part B. Figures 10 through 16 show the result of changing the neuron numbers on the performance of different ANN models. In these figures, the horizontal axis shows the number neurons in the hidden layer and the vertical axis represents the MSE.

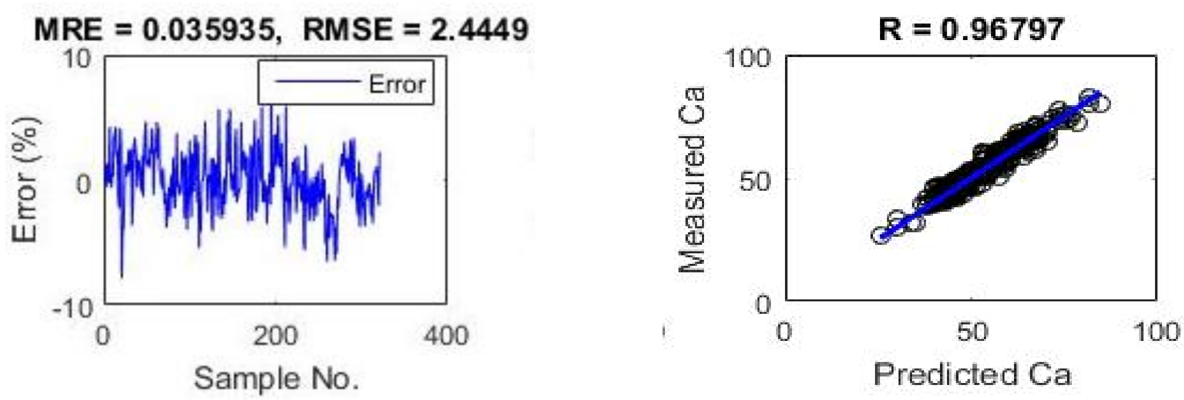
Table 3. Results of neural network models and structures associated with each model.

	Training			**Validation**			**Test**		
Mode.	RMSE	MRE	R	RMSE	MRE	R	RMSE	MRE	R
A1	2.875	0.0395	0.96	2.559	0.0356	0.967	2.768	0.0389	0.96
A2	2.705	0.0377	0.96	2.447	0.0349	0.97	2.444	0.0359	0.96
A3	2.544	0.0351	0.969	2.52	0.037	0.967	2.445	0.0352	0.9742
A4	2.541	0.0358	0.97	2.86	0.0384	0.96	2.505	0.0359	0.9703
A5	1.981	0.0956	0.88	2.182	0.1034	0.88	1.998	0.0990	0.89
A6	2.076	0.1248	0.91	2.492	0.169	0.81	2.420	0.1552	0.87
A7	6.672	0.0837	0.95	7.022	0.0876	0.94	6.222	0.0775	0.96



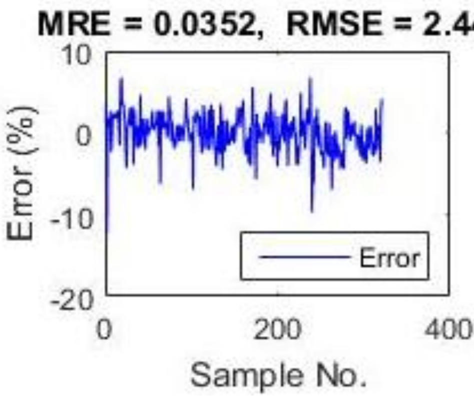
Part A **Part B**

Fig. 3. The relative error and overlap graphs for A1 model.

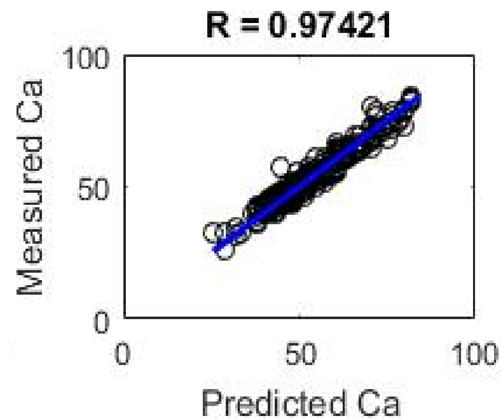


Part A **Part B**

Fig. 4. The relative error and overlap graphs for mode A2

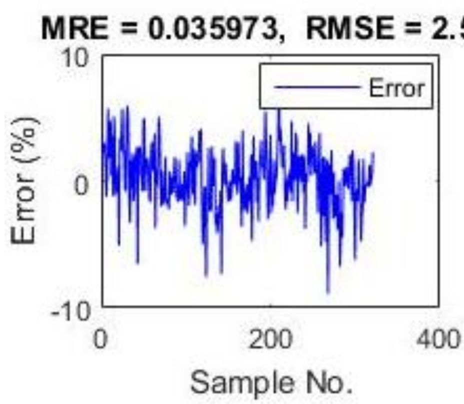


Part A

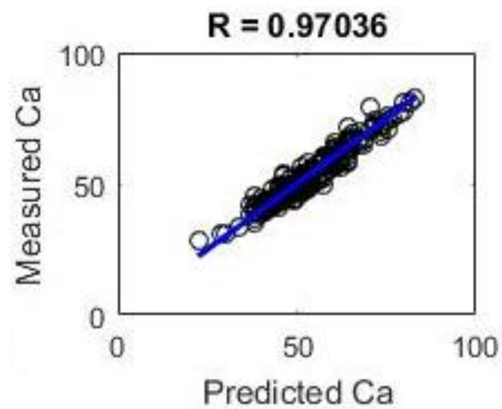


Part B

Fig. 5. The relative error and overlap graphs for mode A3.

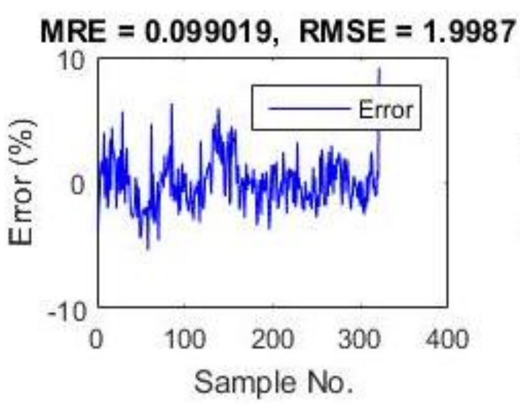


Part A

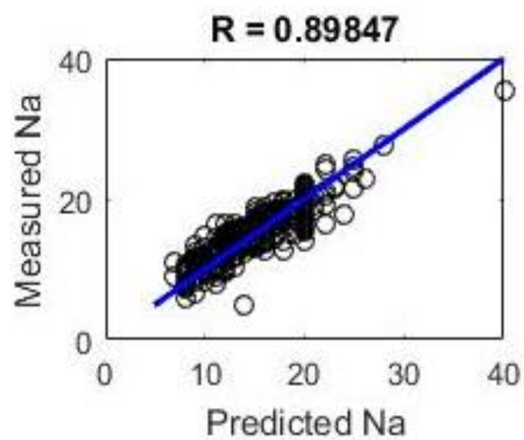


Part B

Fig.6. The relative error and overlap graphs for mode A4.

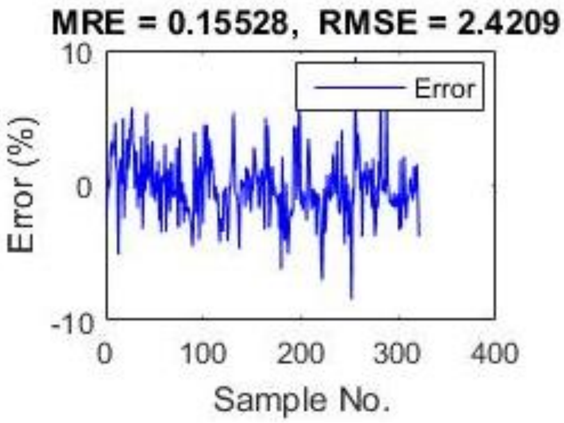


Part A

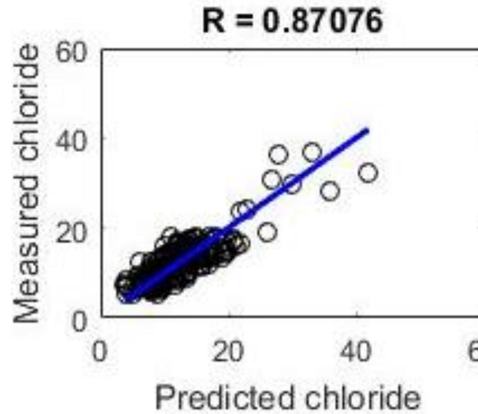


Part B

Fig.7. The relative error and overlap graphs for mode A5.

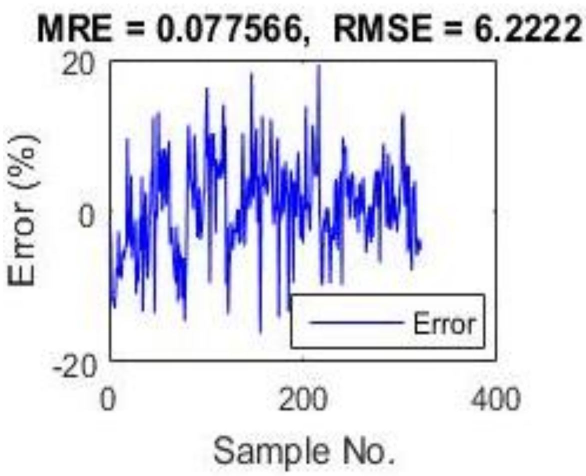


Part A

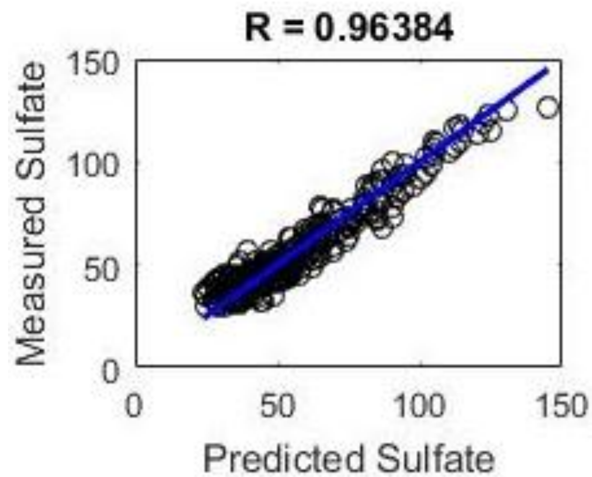


Part B

Fig.8. The relative error and overlap graphs for mode A6.



Part A



Part B

Fig. 9. The relative error and overlap graphs for mode A7.

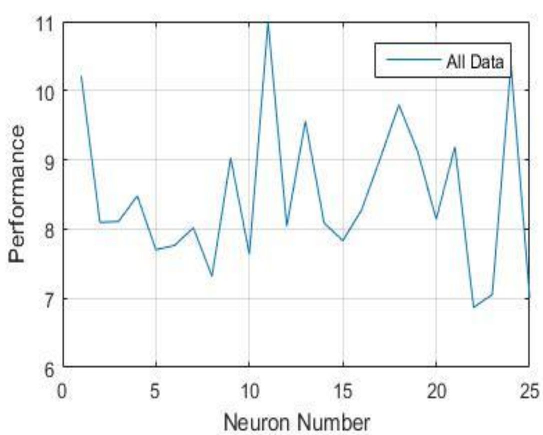


Fig. 10. The variation of performance of the A1 model with the number of neurons in the middle layer

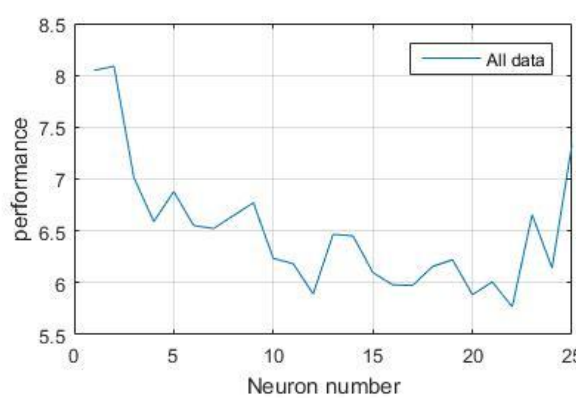


Fig. 11. The variation of performance of the A2 model with the number of neurons in the middle layer.

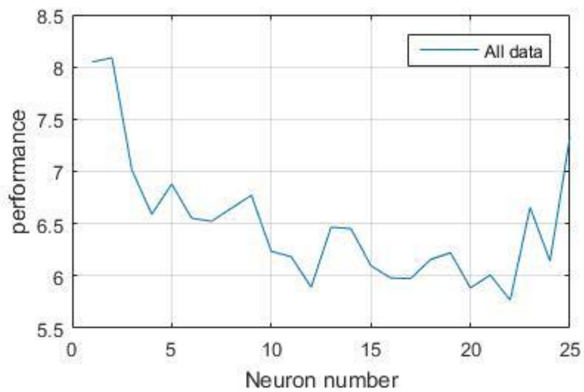


Fig. 12. The variation of performance of the A3 model with the number of neurons in the middle layer.

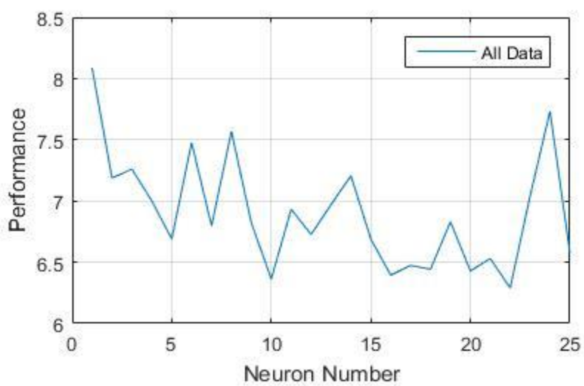


Fig. 13. The variation of performance of the A4 model with the number of neurons in the middle layer.

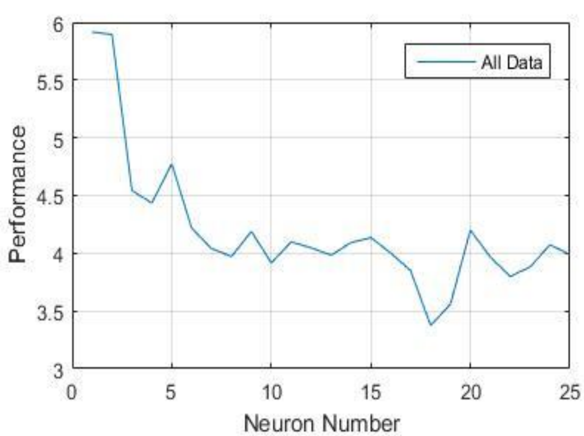


Fig.14. The variation of performance of the A5 model with the number of neurons in the middle layer.

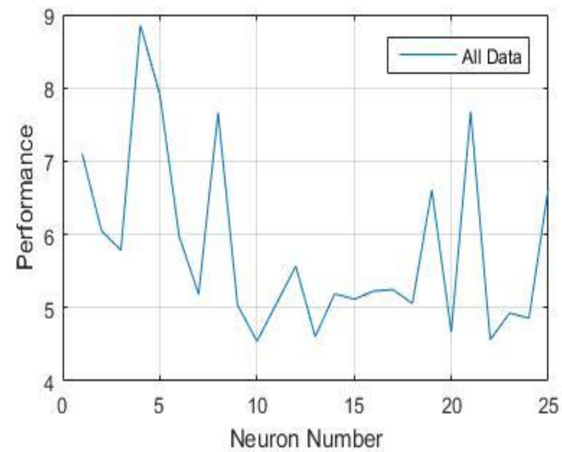


Fig. 15. The variation of performance of the A6 model with the number of neurons in the middle layer.

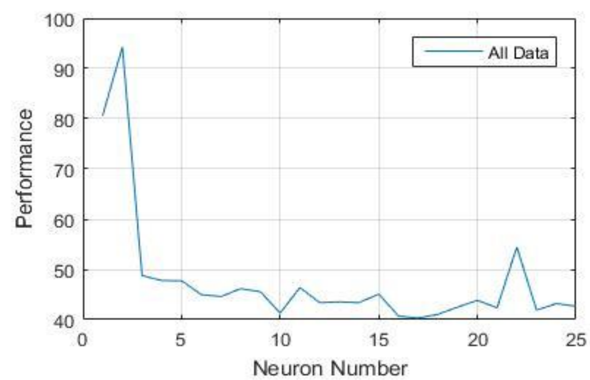


Fig. 16. The variation of performance of the A7 model with the number of neurons in the middle layer

6. Conclusions

The comparison of the seven ANN models with the experimental data shows that all of these models agree well with the real field data. But, the A4 model for Ca concentrations seems to be more accurate with fewer input values, so the use of this model may lead to the decrease of experimental expenses. Although the A3 model is as accurate as the A4 model, it needs additional input values which are not always available and also needs more neurons in the hidden layer which makes the model more complicated. So the A4 model with the Levenberg-Marquardt training function, the tangent-sigmoid transfer function and 10 neurons in the middle layer is selected as the optimum model with the correlated coefficient of 0.97, root mean square error of 2.3562 and mean relative error of 0.034558. This research indicate that the prediction of Ca concentration is more accurate than other concentrations such as Na, Sulfate and chloride in the Karaj River. It can be also concluded that certain types of ANN models may be used to predict the Ca or other ions concentration in natural waters with a reasonable accuracy and limited input parameters.

References

[1] Boyd, C. E. (2000). Water Quality: An introduction kluwer academic publishers. *Norwell, Massachusetts*, *2061*.

[2] Chen, L. (2017). *A Case study of dissolved oxygen characteristics in a wind-induced flow dominated shallow stormwater pond subject to hydrogen sulfide production* (Doctoral dissertation, université d'Ottawa/University of Ottawa).

[3] World Health Organization. (2004). *Guidelines for drinking-water quality* (Vol. 1). World Health Organization.

[4] Chapman, D. V., World Health Organization. (1996). Water quality assessments: a guide to the use of biota, sediments and water in environmental monitoring.

[5] McCleskey, R. B. (2011). Electrical conductivity of electrolytes found in natural waters from (5 to 90) C. *Journal of chemical and engineering data*, *56*(2), 317-327.

[6] Marandi, A., Polikarpus, M., Jõeleht, A. (2013). A new approach for describing the relationship between electrical conductivity and major anion concentration in natural waters. *Applied geochemistry*, *38*, 103-109.

[7] Aghbashlo, M., Hosseinpour, S., Mujumdar, A. S. (2015). Application of artificial neural networks (ANNs) in drying technology: a comprehensive review. *Drying technology*, *33*(12), 1397-1462.

[8] Mirarab, M., Sharifi, M., Ghayyem, M. A., Mirarab, F. (2014). Prediction of solubility of CO_2 in ethanol–[EMIM][Tf_2N] ionic liquid mixtures using artificial neural networks based on genetic algorithm. *Fluid phase equilibria*, *371*, 6-14.

[9] Movagharnejad, K., Mehdizadeh, B., Banihashemi, M., Kordkheili, M. S. (2011). Forecasting the differences between various commercial oil prices in the Persian Gulf region by neural network. *Energy*, *36*(7), 3979-3984.

[10] Zare, A. H., Bayat, V. M., Daneshkare, A. P. (2011). Forecasting nitrate concentration in groundwater using artificial neural network and linear regression models. *International agrophysics*, *25*(2), 187-192.

[11] Zare, A. H., Yazdani, V., Azhdari, K. H. (2009). Comparative study of four meteorological drought index based on relative yield of rain fed wheat in Hamedan province. *Physical geography research quarterly,* 69, 35-49.

[12] Mehrdadi, N., Hasanlou, H., Jafarzadeh, M. T., Hasanlou, H., Abdolabadi, H. (2012). Simulation of low TDS and biological units of Fajr industrial wastewater treatment plant using artificial neural network and principal component analysis hybrid method. *Journal of water resource and protection*, *4*(6), 370.

[13] Moghaddamali, F., Movagharnejad, K. (2014). Predicting electrical conductivity in Jajrud river by an artificial neural network. *Caspian journal of applied sciences research*, *3*(11), 21-29.

[14] Demuth, H., Beale, M., Hagan, M. Neural network toolbox: for use with MATLAB2000. *The mathworks*.

[15] Coppola Jr, E., Szidarovszky, F., Poulton, M., Charles, E. (2003). Artificial neural network approach for predicting transient water levels in a multilayered groundwater system under variable state, pumping, and climate conditions. *Journal of hydrologic engineering*, *8*(6), 348-360.

[16] Coulibaly, P., Anctil, F., Aravena, R., Bobée, B. (2001). Artificial neural network modeling of water table depth fluctuations. *Water resources research*, *37*(4), 885-896.