Prediction of Ryznar Stability Index for Treated Water of WTPs Located on Al-Karakh Side of Baghdad City using Artificial Neural Network (ANN) Technique

Asst. Prof. Dr. Awatif Soaded Alsaqqar
Department of Civil Engineering
College of Engineering
University of Baghdad
Email: d.alsaqqar@yahoo.com

Asst. Prof. Dr. Basim Hussein Khudair
Department of Civil Engineering
College of Engineering
University of Baghdad
Email: basim22003@yahoo.com

Lect. Dr. Sura Kareem Ali
Department of Civil Engineering
College of Engineering
University of Baghdad
Email: srak39@yahoo.com

ABSTRACT

In this research an Artificial Neural Network (ANN) technique was applied for the prediction of Ryznar Index (RI) of the flowing water from WTPs in Al-Karakh side (left side) in Baghdad city for year 2013. Three models (ANN1, ANN2 and ANN3) have been developed and tested using data from Baghdad Mayoralty (Amanat Baghdad) including drinking water quality for the period 2004 to 2013. The results indicate that it is quite possible to use an artificial neural networks in predicting the stability index (RI) with a good degree of accuracy. Where ANN 2 model could be used to predict RI for the effluents from Al-Karakh, Al-Qadisiya and Al-Karama WTPs as the highest correlation coefficient were obtained 92.4, 82.9 and 79.1% respectively. For Al-Dora WTP, ANN 3 model could be used as R was 92.8%.

Key words: artificial neural network; Reynar index; water stability; water treatment plants; correlation coefficient.

التنبؤ بمؤشر الاستقرار(RI) للمياه المعالجة من محطات تصفية الماء على جانب الكرخ من مدينة بغداد باستخدام تقنية الشبكات العصبية الاصطناعية (ANN)

أ.م. د. عواطف سؤدد عبدالحميد
قسم الهندسة المدنية
كلية الهندسة/جامعة بغداد

أ.م.د. باسم حسين خضير
قسم الهندسة المدنية
كلية الهندسة/جامعة بغداد

م.د. سرى كريم علي
قسم الهندسة المدنية
كلية الهندسة/جامعة بغداد

الخلاصة

في هذا البحث تم تطبيق تقنية الشبكات العصبية الاصطناعية (ANN) للتنبؤ بمؤشر الاستقرار (RI) من المياه المتدفقة من محطات تصفية الماء في الجانب الكرخ (الجانب الأيسر) في مدينة بغداد للعام 2013. قد طورت وفحصت ثلاثة نماذج (ANN1, ANN2, and ANN3) باستخدام البيانات من أمانة بغداد بما في ذلك نوعية مياه الشرب للفترة من 2004 الى 2013. وتشير النتائج إلى أنه من الممكن جدا استخدام الشبكات العصبية الاصطناعية في التنبؤ بمؤشر الاستقرار القياسي (RI) مع درجة جيدة من الدقة. حيث يمكن استخدام النموذج (ANN2) لتنبؤ (RI) لمياه المنتجة من محطات التصفية الكرخ، القادسية والكرامة بأعلى معامل الارتباط الذي تم الحصول عليه 92.4 ، 82.9 و 79.1% على التوالي . بينما يمكن استخدام النموذج (ANN3) لمحطة تصفية الدورة حيث كان معامل الارتباط 92.8%.

الكلمات الرئيسية: الشبكات العصبية الاصطناعية، مؤشر الاستقرارية، استقرارية الماء، محطات تصفية الماء، معامل الارتباط.

1. INTRODUCTION

Water quality measurements include a variety of physical, chemical and biological parameters. Basic problem in the case water quality monitoring is the complexity associated with the analyzing the large number of variables. Different multivariate statistical techniques, such as cluster analysis, principal component analysis and factor analysis are used for the interpretation of complex data, **Vesna et al., 2010**. Water quality modeling using mathematical simulation techniques in fact classical process based modeling approach could provide good prediction for different water quality parameters. However these models rely on lengthy data and require a large number of input data which may be not available or unknown. Artificial Intelligence techniques (Artificial Neural Network, ANN) have proven their ability and applicability for simulating and modeling various physical phenomena in the water engineering field. In addition this technique (ANN) captures the embedded spatial and unsteady behavior in the investigated problem using its architecture and nonlinearity nature compared with other modeling techniques, **Najah et al. 2009**. Recently applications of ANNs in water engineering, ecological science and environmental engineering, have been reported and used intensively.

In **2005 Diamantopoulou et al.**, used ANN to drive and develop models to predict the monthly of same water parameters at the Axioupolis station of Axios River, Greece. These parameters included, DO, conductivity, NO_3, Na, Ca and Mg. The monthly values of these parameters and six other water parameters with the discharge at this station for the period 1980 to 1994 were selected for this analysis. A feed forward and supervised ANN was achieved with kalman's learning rule to modify the ANN weights. The results indicated that the ANN models can be used for the prediction of these parameters and allowed the filling of missing values of time series of water quality parameters which are very serious problem in most of the Greek monitoring stations.

Aoyama et al., 2007, developed a model for purification mechanisms in Tamagawa River, Tokyo, Japan. The model was proposed to express changes in BOD, COD, Total Nitrogen and Total Phosphorus concentrations as the combination of inflows, streams and weirs for data in 2002. The ANN model used in this study constructed of functions based on observations and then used the derivatives to evaluate the cause and effect of pollution in the river. The model suggested that the cause of pollution in streams from inflows of sewage, the Tamagawa River has purification functions for COD and Total Phosphorus, but little ability for Total Nitrogen.

Mozejko and **Gniot, 2008,** applied ANNs for timeseries modeling of Total Phosphorus (P) concentrations in the Odra River Szczecin, Poland. Different types of ANN models were employed in this study, where the optimal model which gave the minimum error and best correlation between the predicated and observed data was the Generalized Regression Neural Network (GRNN) with two hiddenlayers. Data of the year's 1991 to2004 were used for the development of the model. The model performed satisfactory over the range of the data used for calibration with mean absolute error (MAE) of 0.032 mg P/dm^3 and correlation coefficient (R) 0.931. As for the prediction of P concentration in year 2005 the model gave MAE of 0.024 and R 0.865.

Predication of fecal coliform concentration in the Achencovil River, India using an ANN model was developed by **Swapna** and **Vijayan in 2009**. Water quality parameters used in this study were, DO, pH, temperature and turbidity in the river for the period of 1996 to 2000. The best ANN model achieved the highest correlation coefficient (R^2) was 0.911 using eight neurons in the hidden layer. Using the same input parameters, a statistical model was developed using SPSS which gave R^2 0.874. Hence it can be inferred that the ANN model slightly outperforms the statistical model and thus can be used for predicating coliform concentrations with better accuracy.

In **2010, Vesna et al.,** developed a feed forward neural network (FNN) model to predict DO concentration in the Gruza reservoir, Serbia. Monthly sampling of water quality was carried out during the period of 2000 to 2003from three sites in the reservoir. Water parameters included pH, NO_3, NO_2, NH_4, Cl, Fe, Mn, P, temperature and conductivity. From the sensitivity analysis, the most effective input parameters were pH and temperature. The Levenberg Marquardt algorithm was used to train the FNN model. The results obtained that the best FNN model was having 15 hidden neurons with the highest correlation coefficient (R^2) of 0.974 for training and 0.8738 for testing. The respective values of mean absolute error (MAE) and mean square error (MSE) for the three sets were 0.4693 and 0.667 for training, 1.179 and 2.7585 for testing as for training + testing the model determined 0.5797 and 0.9923 respectively.

2. WATER TREATMENT PLANTS UNDER STUDY

The sampling sites were chosen to be the effluents from the water treatment plants on the Tigris River in Baghdad City. In Baghdad City there are eight water treatment plants located on the banks of the Tigris River along a distance of 50–60 km. These plants are Al-Karakh, Al-Karama, Al-Qadisiya and Al-Dora in Al-Karakh side (left) of the city. Where on Al-Rasafa side (right) are East Tigris, Al-Wathba, Al-Wahda and Al-Rashed WTPs. The water quality of the treated water from these plants was taken as the necessary water parameters for the determination of the water stability index (Ryznar) in this study. The different water parameters required for these calculations were provided from Baghdad Mayoralty (Amanat Baghdad) for the period from January 2004 to December 2013 for the recorded of these WTPs, which included: pH value, Alkalinity, Total Dissolved Solids, Calcium concentration and Temperature.

3. WATER STABILITY

Corrosive water can dissolve minerals and other types can deposit minerals known as scaling water, this behavior of water is known as stability. Corrosive or scaling water can be harmful to the distribution systems as it can dissolve minerals that detriment water quality or dissolve harmful metals such as lead and copper. Scaling water deposits a film of minerals that may reduce the carrying capacity of the pipes (but it may be a protective layer to prevent pipe corrosion). Also, excessive scaling may damage water heaters (boilers) and increase the friction coefficient in the pipes. Therefore the most desirable water is of stability in the range of slight scaling, **Qasim et al., 2000**. In general there are several ways to calculate the water stability such as Langelier and Ryznar index.

4. COMMON METHODS USED TO MEASURE WATER STABILITY

The US Environmental Protection Agency (USEPA) has recommended the use of Langelier (LSI) and Ryznar (RSI) Stability Indices to monitor the corrosion potential of water, **Degremont, 1991, Kawamura, 2000. Qasim, et al., 2000** and **MWH, 2005**. In this study, Ryznar index is used. This index is a quantitative index of the amount of calcium carbonate scale that would be formed and to predict the corrosiveness of waters that are not scale forming. The equation for the determination of RI is:

RI = 2pH saturation - pH actual (1)

Where:

pH actual = measured pH of water.

pH saturation = (pk`2 – pk`s) + pCa+2 + pAlk + S (2)

(pk`2 – pk`s) = dissociation constant based on temperature and total dissolved solids or ionic strength.

pk`2 = acidity constant for the dissociation of bicarbonate.

pk`s = mixed solubility constant for CaCO3pCa+2 = - log (calcium ion in moles / liter).

pAlk = - log (total alkalinity in equivalent of CaCO3/ liter).

S = salinity correction term = 2.5 μ ½ / (1+5.3 μ ½ + 5.5 μ) (3)

Where μ = ionic strength.

For total dissolved solids content less than 500 mg/L, the ionic strength may be estimated by 2.5 x 10^{-5} x TDS. An alternate approximation of ionic strength can be made using the total hardness and total alkalinity, **Millete et al., 1980**. **Table 1** lists the scale formation or corrosive tendencies of waters with various Ryznar index values, **Qasim et al., 2000**.

5. ARTIFICAL NEURAL NETWORK (ANN)

Forecasting models can be divided into statistical and physically based approaches. Statistical approaches determine relationships between historical data sets, whereas physical based approaches models the underlying process directly. Multilayer Perception (MLP) networks are one type of Artificial Neural Networks (ANN) suited for forecasting applications and are closely related to statistical models, (the modeling philosophy for ANN is similar to that used in traditional statistical approaches), **Najah et al., 2009**.

ANN models are specified by network topology, node characteristics and training or learning rules. It is an interconnection set of weights that contains the knowledge generated by the model, **Hafizan et al., 2004**. Different types of ANNs exist; the most common types are the feed forward network and backward network multilayer perceptron. In these networks, the artificial neurons or processing units are arranged in a layered configuration as:

Input layer - connecting the input information to the network.

Hidden layer (one or more) – acting as the intermediate computational layer.

Output layer – producing the desired output.

Units in the input layer introduce normalized of filtered values of each input into the network. Units in the hidden and output layers are connected to all of the units in the preceding layer.

Each connection carries a weighting factor. The weighted sum of all inputs to a processing unit is calculated and compared to a threshold value. An activation signal then is passed through a mathematical transfer function to create an output signal that is sent to processing units in the next layer. Training an ANN is a mathematical exercise that optimizes all of the network weights and threshold values, using some fraction of the available data. ANN learns as long as the input data set contains a wide range of patterns that the network can predict. The final model is likely to find those patterns and successfully use them in its prediction **,Stewart, 2002**.

Several neural network softwares are available; Neuframe 4 has been used in this study. Three ANN models were constructed for the prediction of the Ryznar index (RI) for the four water treatment plants on the Al-Karakh Side(Al-Karakh, Al-Karama, Al-Qadisiya And Al-Dora WTPs) for the year 2013 which was considered the target year. The first step for the determination of the ANN model is the selection of the data to be the input variables.The inputs chosen for each model are listed in **Table 2**. The effect of the different combinations of these parameters has a great influence on the model performance.

The data have to be divided into three sets, training, testing and validation. This step is achieved by trial and error to select the best division with respect to the lowest testing error followed by training error and high correlation coefficient of the validation set. The general strategy adopted for finding the optimal network architecture and internal parameters that control the training process is by trial and error using the default parameters of the software. In this step, first the nodes of the hidden layer are increased until no significant improvement is gained in the model performance. Then the model is tested by changing the default parameters of the software, the momentum term which is 0.8 and the learning rate 0.2. Finally the transfer functions of the input and hidden layers are tested where the default functions of the software are, linear in the input layer and sigmoid in the hidden layer. The default alternatives of the software are to test the following functions: linear, sigmoid and hyperbolic tangent (tanh). The effect of the different combinations of these parameters will be discussed in the following section.

6. RESULTS AND DISCUSSION

1-Ryznar Stability Index (RI) was calculated using Eq.(1) for the data supplied from Baghdad Mayoralty (Amanat Baghdad) for the period from January 2004 to December 2013 for the eight WTPs in Baghdad City, which included: pH value, Alkalinity, Total Dissolved Solids, Calcium concentration and Temperature. The treated water is within the drinking water standards but the Ryznar Index of this water shows that it is corrosive to very corrosive water (RI more than 6.8).

2-ANN models that were the result of applying Neuframe 4 software and the effect of the different combinations of the input parameters are summarized in **Table 3** which shows the best model performance according to the lowest testing error and the highest correlation coefficient (R^2).All models have three layers (input layer with 5 inputs, one hidden layer with one node and one output layer). Almost all models worked best with the default parameters of the software, momentum rate 0.8 and learning rate 0.2. Finally the transfer functions of the input, hidden and output layers where also the default functions of the software which is, linear in the input layer and sigmoid in the hidden and output layers.

Model ANN 3 gave the highest R^2 (97.4%) but not the less testing error (4.0274%) where model ANN 1 for Al-Karakh WTP had the less testing model (3.7049%) and R^2 (94.3%). These three models were tested to predict RI in the year 2013 for the four WTPs on Al-Karakh side of Baghdad city. **Table 4** shows correlation coefficient (R^2%) in each plant between the predicated and observed RI values using the suggested models in **Table 3**. From this table it is clear that RI could be predicted by ANN 2 model in Al-Karakh, Al-Karama and Al-Qadisiya WTPs as the highest correlation coefficient were obtained 92.4, 82.9 and 79.1% respectively. For Al-Dora WTP ANN3 model could be used as R^2 was 92.8%. **Fig. 1 to 4** show the variation of RI in year 2013 for each plant according to the best model represented in **Table 4**.

REFERENCES

- Aoyama Tomoo, Junko Kambe, Aiko Yamauchi and Umpei Nagashima, 2007, *Construction of a Model for Water Purification Mechanisms in a River by Using a Neural Network Approach*, J. Comput. Chem. JPN., Vol. 6, No.2, PP. 135-144.

- Degremont, 1991, *Water Treatment Handbook*, 6th edition, Lavoisier Publishing, France.

- Diamantopoulou, M. J., V. Z. Antonopoulos and D. M. Papamichail, 2005, *The Use of a Neural Network Technique for the Prediction of Water Quality Parameters of Axios River in Northern Greece*, EWRA European Water 11/12, 55-62.

- Hafizan Juahir, Sharifuddin M. Zain, Mohd Ekhwan, Mazlin Mokhtar and Hasfalina Man, 2004, *Application of Artificial Neural Network Models for Predicting Water Quality Index*, Journal Kejuruteraan Awam Vol. 16, No. 2, PP. 42-55.

- Kawamura, S., 2000, *Integrated Design and Operation of Water Treatment Facilities,* 2nd edition, John Wiley & Sons, Inc. New York.

- MWH, 2005, *Water Treatment Principles and Design*, 2nd edition, .John Wiley & Sons, Inc. Hoboken N.J.

- Mozejko J. and R. Gniot, 2008, *Application of Neural Networks for the Prediction of Total Phosphorus Concentration in Surface Waters*, Polish J. of Environ. Stud. Vol. 17, No. 3, PP. 363-368.

- Millette, J. R., Arthur F. Hammonds, Michael F. Pansing, Edward C. Hansen, and Patrick J. Clark, 1980, *Aggressive Water: Assessing the Extent of the Problem*, AWWA, Vol. 72, No. 5, PP. 262-266.

- Najah Ali, Ahmed Elshafie, Othman A. Karim and Othman Jaffer, 2009, *Prediction of Jobor River Water Quality Parameters Using Artificial Neural Network*, European Journal of Scientific Research,Vol. 28, No.3, PP. 422-435.

- Qasim S. R., E. M. Motley and G. Zhu, 2000, *Water Works Engineering Planning, Design and Operation*, Prentice Hall PTR. USA.

- Stewart A. Rounds, 2002, *Development of a Neural Network Model for Dissolved Oxygen in the Tualatin River, Oregon*, Proceedings of the 2nd Federal Interagency Hydrologic Modeling Conference, Las Vegus, Nevada July 29- August 1.

- Swapna Varma and N. Vijayan, 2009, *Prediction of Fecal Coliform Concentration in Surface Water Using Artificial Neural Network*, 10th National Conference on Technology Trends (NCTT09) 6-7 Nov.

- Vesna Rankovic, Jasna Radulovic, Ivana Radojevic, Aleksandar Ostojic and Ljiljana Comic, 2010, *Neural Network Modeling of Dissolved Oxygen in the Gruza Reservoir, Serbia*, J. Ecological Modelling, 221, PP. 1239-1244, Elsevier.

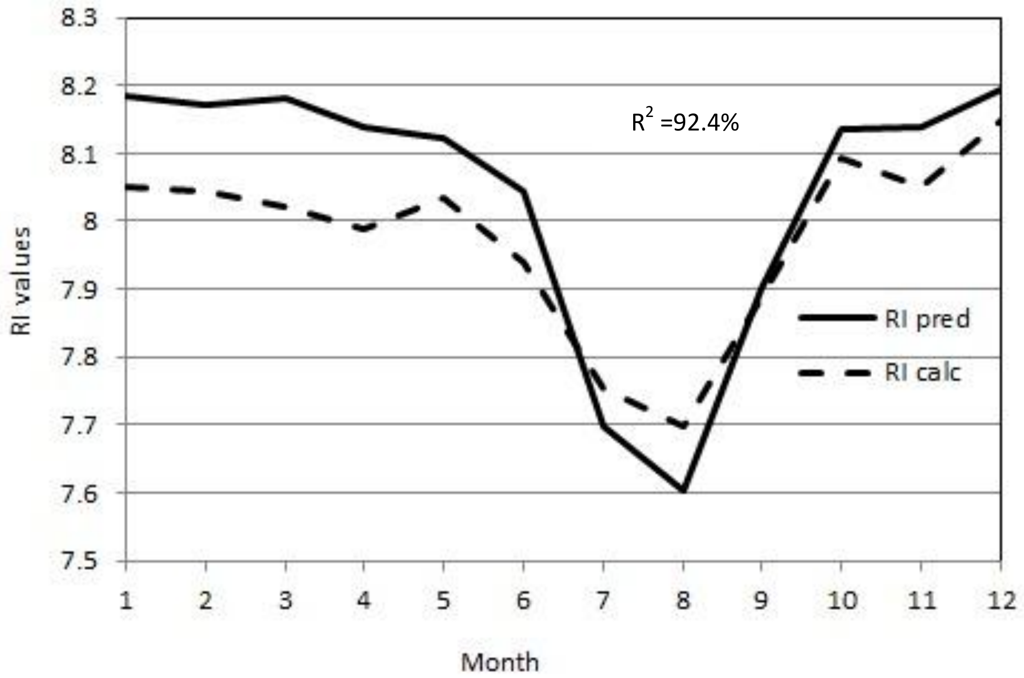

Figure 1. Calculated and predicted RI for Al-Karakh WTP during 2013 by ANN2.

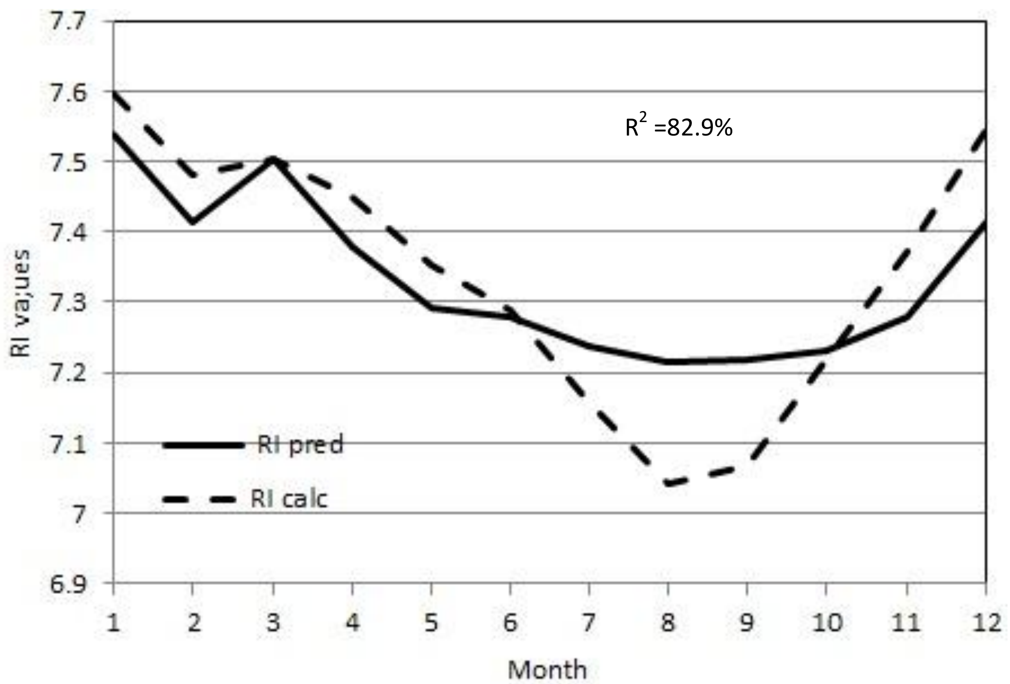

Figure 2. Calculated and predicted RI for Al-Karama WTP during 2013 by ANN2.

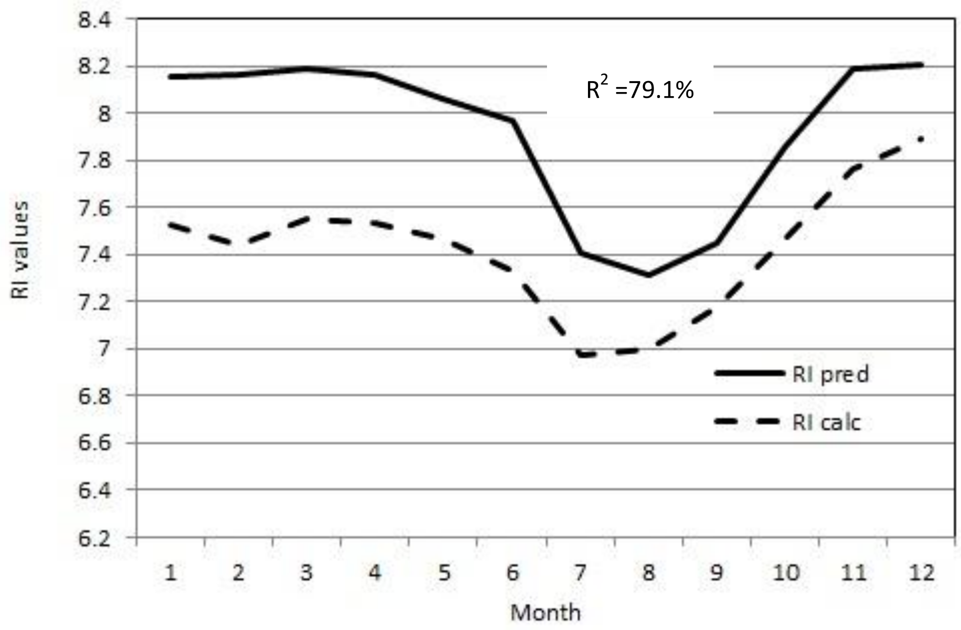

Figure 3. Calculated and predicted RI for Al-Qadisiyia WTP during 2013 by ANN2.

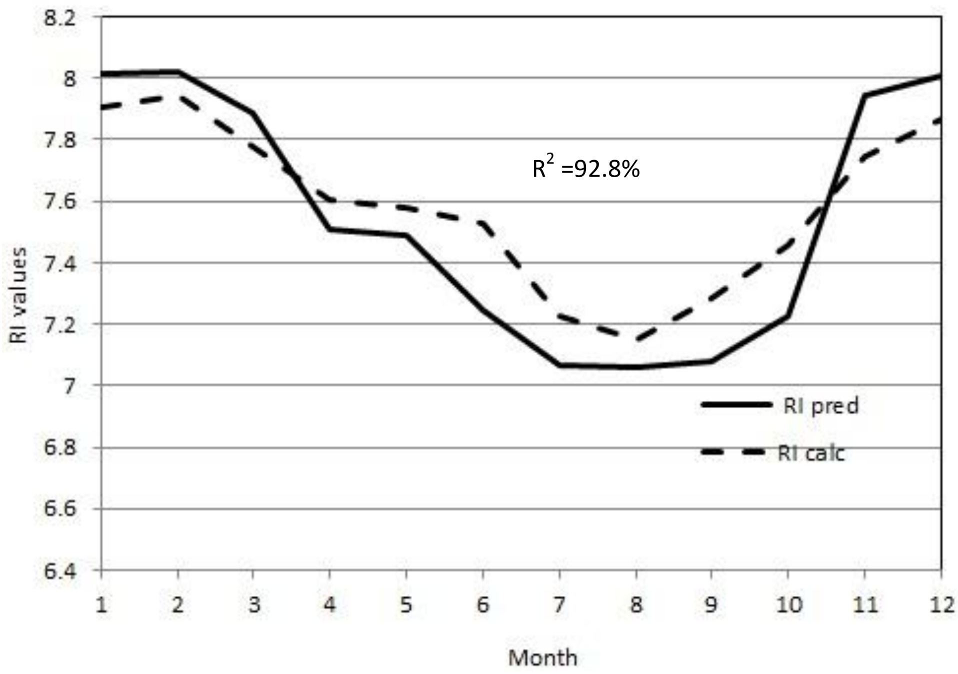

Figure 4. Calculated and predicted RI for Al-Dora WTP during 2013 by ANN3.

Table 1. Scale and corrosion tendencies of water with various Ryznar index (RI) values **(Qasim et al., 2000).**

RI Range	Indication
Less than 5.5	Heavy scale formation
5.5 to 6.2	Some scale will form
6.2 to 6.8	Non-scaling or corrosive
6.8 to 8.5	Corrosive water
More than 8.5	Very corrosive water

Table 2. Data used in ANN models.

Model	Input	Output
ANN 1	Water quality from each plant for 2004 to 2012	RI for each plant in year 2013
ANN 2	Water quality from the 8 plants in Baghdad for 2004 only	RI for each plant in year 2013
ANN 3	Water quality from 4 plants on Al-Karakh side from 2004 to 2012	RI for each plant in year 2013

Table 3. ANN Models, optimization and stopping criteria.

ANN model		Momentum rate	Learning rate	Testing error (%)	Training error (%)	Correlation coefficient (R^2%)
ANN 1	Al-Karakh WTP	0.8	0.2	3.7049	4.9451	94.3
	Al-Karama WTP	0.75	0.2	4.0542	5.1127	94.9
	Al-Qadisiya WTP	0.78	0.2	6.4884	4.9715	90.9
	Al-Dora WTP	0.8	0.2	6.1195	4.9809	90.2
ANN 2		0.79	0.18	4.9415	4.9286	96.5
ANN 3		0.8	0.2	4.0274	5.3908	97.4

Table 4. Correlation coefficient (R^2%) in each plant.

ANN Model	WTP			
	Al-Karakh	Al-Karama	Al-Qadisiya	Al-Dora
ANN 1	79.6	72.8	77.2	90.4
ANN 2	92.4	82.9	79.1	83.3
ANN 3	87.2	56.8	61	92.8