Journal homepage www.jzs.univsul.edu.iq
Journal of Zankoy Sulaimani
Part-A- (Pure and Applied Sciences)

Model of Prediction of Trihalomethanes (THMs) formation in chlorinated water in water treatment plant using Artificial Neural Network

Khairi Ali Omar[1]

1Department of Water Resources Engineering, College of Engineering, Duhok-Kurdistan Region / Iraq
E-mail: Khairi.ali@uod.ac

Article info

Original: 29 June 2016
Revised: 30 August 2016
Accepted: 16 October 2016
Published online: 20 March 2017

Key Words:
Disinfection-By Product, Trihalomethane, Artificial Neural Network, Graphical User Interface, Modeling, Water Treatment.

Abstract

This study focuses on validity of modeling of trihalomethane formation using artificial neural network with a feed-forward back-propagation neural network approach and graphical user interface function of MATLAB at a water treatment plant. Regularly measured parameters, which are conducted by Uyak et al., 2005, consisting of pH, total organic carbon, temperature and applied chlorine dose were utilized to implement the model for forecasting of trihalomethane formation. Attempting to investigate and compare with the use of traditional multiple linear regressions. The default Levenberg-Marquardt algorithm was used for training the network structures. It is noted that the best validation performance depending on the mean square error was 62.8539 at epoch 49. The simulated result tracked the measured data through a regression plot with a correlation coefficient of 0.95842, 0.94717 and 0.93369 for training, validation and test respectively. With respect to simulated testing the model demonstrates great performance in predicting the trihalomethane formation in chlorinated water. It is concluded that the artificial neural network generally is better than the multiple linear regression for forecasting trihalomethane formation.

Introduction

The most significant process in the treatment of drinking water supplies is disinfection as it eliminates or deactivates pathogenic microorganisms that are responsible for producing waterborne diseases for instance typhoid, dysentery and cholera (White, 1992). It generally agrees that chlorination is a widely used technique for disinfection because of its very efficient and cost-effective properties (Abdulla, 2003). Nearly all water treatment plant operators in Iraq likewise use chlorine for water disinfection purposes. However, it has been discovered that using of chlorine as a disinfectant caused potential health risks due to the formation of halogenated organic compounds which causes carcinogenic known as disinfection by products (DBPs) (Rook, 1974) (Bellar, 1974). Among disinfection by product that was found in chlorinated water, trihalomethanes (THMs) have been the concentration of certain attention because they are considered potentially carcinogenic for the bladder (McGeehin, 1993).

The hazardous classes for different disinfection by products among THMs, HAAs and inorganic DBPs; $CHCI_3$, $CHBr_2C1$ and $CHBr_3$ defined by The United States Environmental Protection Agency were classified as possible carcinogens to human beings (EPA, 2002) (Cantor, 1987). Concerning about health risks linked with trihalomethane have imposed many developed countries to set a maximum acceptable level

for trihalomethane concentration in drinking water supply (Rodriguez, 2003) (Milot, 2000). The Rule of Disinfectant/Disinfection by Product (D/DBP) developed by the US EPA in 1998 that is setting a maximum contaminant level of 80 µg/L for trihalomethane in drinking water supply (EPA, 1998). Furthermore, the most of the European countries controlled trihalomethane in their drinking water at the maximum water level of 100 µg/L (EECD, 1997).

It is highly required that to monitor the formation of trihalomethane during water treatment processes. The modeling of trihalomethane involves forming an empirical or mechanistic relationship between trihalomethane level in treated water and the quality of water and operational control parameters of water treatment such as; temperature, pH and applied of chlorine dose. Raw water chlorination was used for development of model equations not chlorination of treated water because it is considered most appropriate (Chowdhury, 1999). Numerous models were developed in literature to improve predictive equations of trihalomethane based on reaction kinetics of disinfection by product formation. Progressed development of new models by (Elshorbagy, 2000), (Golfinopoulos, 1998), (Gallard, 2002), (Abdulla, 2003),(Sohn, 2001),(Golfinopoulos,2002), and (Milot,2000) have participated in analyzing the effect of water characteristics on trihalomethane formation meanwhile optimizing the process of coagulation that is vital for treatment of water.

It has been pointed out by (Milot, 2000) that the most of the models published are empirical and are depended on statistical regression equations predicting the level of trihalomethane from the number of operational control and water quality characteristics. Even though regression based models have revealed an acceptable predictive amount of trihalomethane formation, it is still not free from lack of consideration of non-linear and very complex interactions between trihalomethane and operational parameters in one hand and trihalomethane and water quality parameters on the other hand. Saeed and George (2012) investigate the Feed-Forward Artificial Neural Networks for crack identification and estimates and turbine operating conditions in Francis Turbine type. Al-Suhili and Karim (2015) used artificial neural networks to forecast daily inflow for Dukan reservoir. Maier and Dandy (2000) highlighted the capability of artificial neural network in several modeling fields such as solving classification problems and evaluating raw water quality. Heller and Singh (1994) used artificial neural network for predicting the demand of urban water. Joo et al. (2000) stressed the ability of artificial neural network in establishing coagulation dosage. Hashem and Hassan (2005) concluded that the artificial neural network well predicts the residual chlorine decay in water distribution system.

In recent decades the modeling of trihalomethanes in drinking water supplies has gained much interest. The capability of artificial neural network in modeling both complex and non-linear problems lead the researchers to utilize this approach in predicting trihalomethane formation in water chlorination. The aim of this paper is to weigh the validity of a current approach of non-linear modeling, known as Artificial Neural Networks, to predict the concentrations of trihalomethane formation under conditions of controlled chlorination. Artificial Neural Networks will be evaluated via comparing with the results of trihalomethane predicted by multiple linear regression technique. It is worth noted that the developed model depends on the database published by Uyak et al. (2005).

Materials and Methods

A. Process of Simulation

To evaluate the utility of artificial neural network (ANN) for modeling of trihalomethane formation, both ANN and multiple linear regression models are required. The development of this model is based on the database of independent experimentation developed by Uyak et al. (2005). The statistical distribution of raw water quality parameters and processed water utilized for model development purposes are presented in Table (1).

Table (1). Statistical distribution of raw water quality parameters and processed water by (Uyak, 2005)

Variables	Sample #	Minimum	Maximum	Mean	Standard deviation
TOC, mg/L	74	4.20	6.20	4.86	0.66
pH	120	7.10	7.90	7.45	0.31
Applied Cl_2, mg/L	112	2.82	6.75	4.51	1.39
Temperature, ^oC	124	7.20	22.70	15.23	5.68
THMs of processed water, µg/L	96	48.0	102.0	68.7	18.0

B. Description of ANNs and MLRs

A Multiple Linear Regression Analysis is a well-recognized technique of modeling methodology used in various fields of interested researches which is establishing the power of a linear correlation between a dependent variable and a set of independent variables (Menard, 1995). Rodriguez, et al. (2003) described an equation illustrating the relationship between variables as shown in the following from:

$$Y = \sum \beta_0 + \sum_{i=1}^{m} \beta_i X_i \quad (1)$$

Where Y represents the dependent variable, Xi is the independent variables with m indicating the number of independent variables considered, β_0 representing the intercept and β_i is the coefficient of partial slope which provides a partial prediction to the corresponding value of Y. The ordinary least square method is generally used to estimate the parameters of multiple linear regression model which is resulting in a line that the sum of squared vertical distances were minimized from the predicted data to the line (Neter, 1990).

Artificial neural networks are a modeling strategy stimulated by training system of brain's nervous. The capability of the model to learn from the example of provided data that illustrate either a physical phenomenon or a decision process (Rumelhart, 1994). An Artificial neural network model delivers specific theoretical advantages over traditional techniques like a multiple linear regression, comprising its great capability for generalization and its improved tolerance to very noisy data (Hammerstrom, 1993). An artificial neural network comprises of neurons which categorized into firstly, an input layer which is receiving the input data; secondly, one or more hidden layers which is processing data; and finally, an output layer which is producing the output of the network. Many structures of artificial neural network have been suggested and discovered since the 1960s. Among them the multi-layer feed-forward networks with back-propagation training algorithms is the most researched and widely used structures in both hydrology and water resources phenomena (Govindaraju, 2000). The typical topology that is considered in this study is as shown in Figure1.

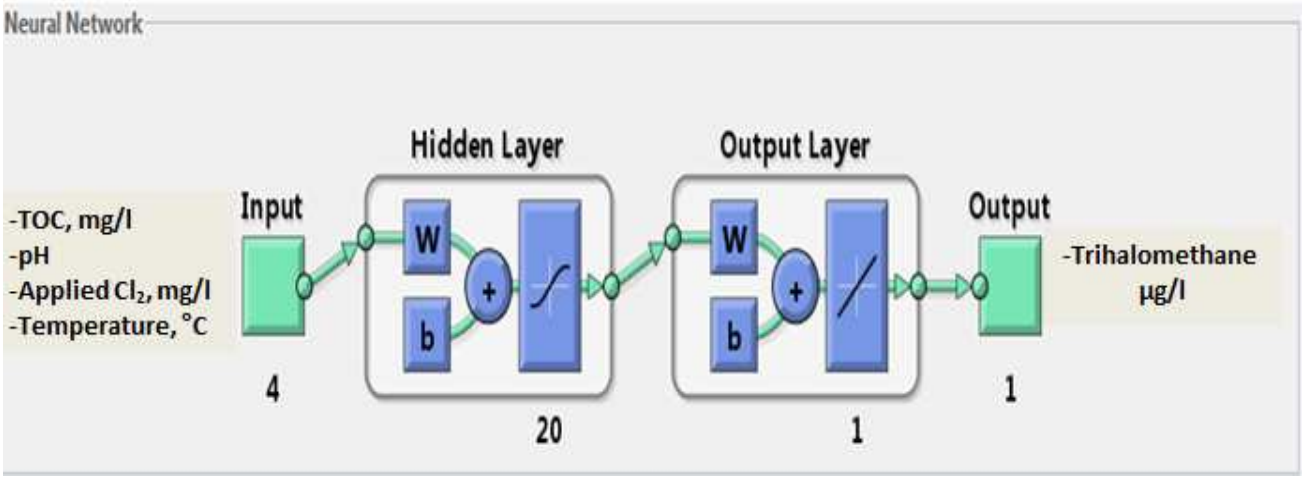

Figure 1: Schematic diagram of Artificial Neural Networks

"The nodes of one layer are connected to the nodes of another layer with connection weight, but they are not connected to nodes of the same layer. Thus, each node in a layer receives signals from nodes of the previous layer with connection weights, adds the weighted inputs of all nodes, converts the weighted sum into an output signal, and transmits the output signal to the nodes of the following layer" (Hashem, 2007). The known input and target values are used to optimize the connection weights between nodes via an iterative process and error-minimization approach, in order that the network generates outputs very close or equal to the well-known target values. This process is so-called network training. The training network associated with an enhanced set of connection weights is then applied to the data set for validation purposes which is estimating the output.

The network layout where data flow is in single direction is called the feed-forward network. Conversely, where the estimated error between the target values and artificial neural network predicted values is weight is named the feed-forward with back- propagation network. The *nntool* graphical user interface obtained in the Neural Network Tool box of MATLAB Software was utilized to generate artificial neural network model (MathWorks, Inc. 2010a). The model of artificial neural network is best implemented compromising of four inputs such as total organic carbon (TOC), pH, temperature and applied chlorine dose, single output which is trihalomethane through one hidden layer consisting of 20 hidden elements which required a learning rate of 0.1 with momentum of 0.001, and epochs 1000 with maximum fail 50.The predicted value of a neuron can be expressed as:

$$predict = f(n) \tag{2}$$

Where $n = \sum_{j=1}^{k} w_j\, x_j + b$ (3)

Where: $x_1, x_2, \ldots\ldots, x_k$ are the input signals; $w_1, w_2, \ldots\ldots, w_k$ are neuron weights, b is value of bias, and $f(*)$ is an activation function.

Both linear and sigmoid activation functions are most commonly used in the construction of artificial neural network.

Linear activation function has form:

$$f(n) = n \tag{4}$$

An example of the sigmoid is the hyperbolic tangent function (Haykin, 1999):

$$f(n) = \frac{1-e^{-n}}{1+e^{-n}} \tag{5}$$

C. Creating trihalomethane predict model with MATLAB

Via typing in *nntool* in the command window of MATLAB, the user will enter the main page of the neural network Graphical User Interface GUI (Figure 2), the Network/Data Manager.

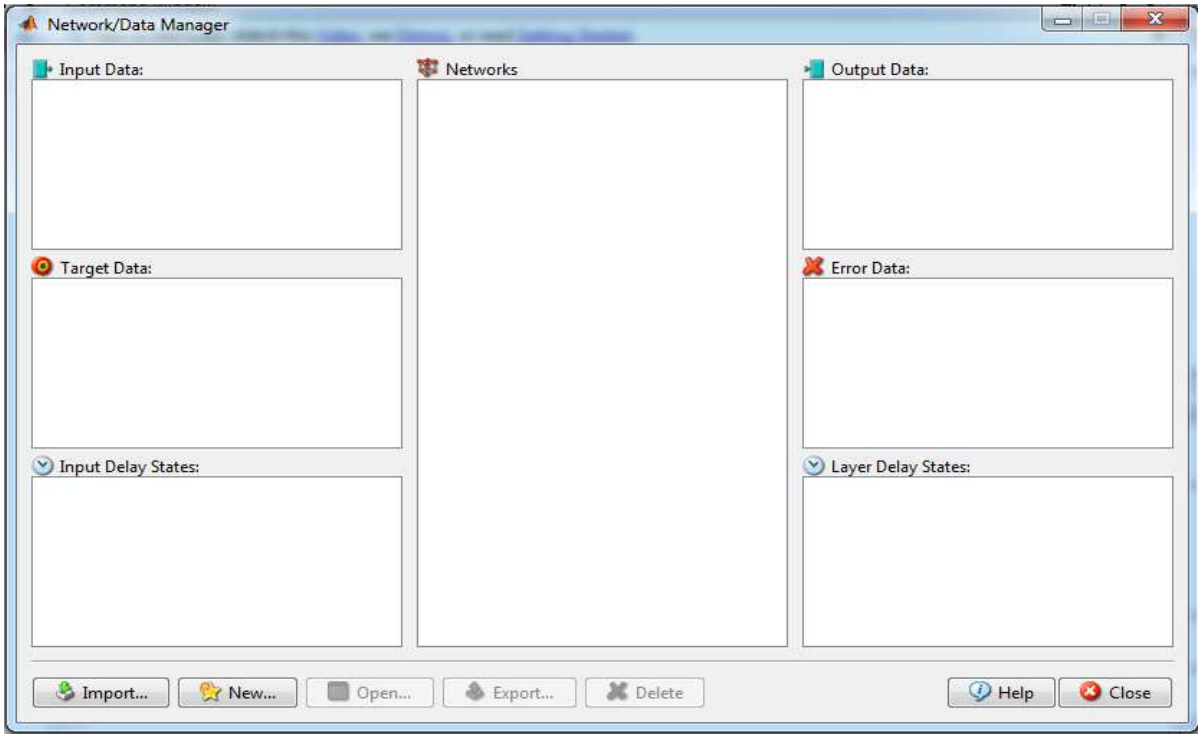

Figure 2: Main Page of Graphical User Interface (GUI)

To start with, both input and target (output) data are imported in the main page of Graphical User Interface. The input variables are set at 4 and output variables 1. Then, click New icon from main page of GUI to construct a new network model as shown in Figure 3. Feed-forward backdrop was selected as the Network Type. The number of neurons in the hidden layer was set to 20, and TANSIG as the neuron transfer function of the hidden layer. It is worth noted that Purelin is chosen as the transfer function for the neurons of the output layer. Figure 4 present an example of constructed model which will be trained and tested with sample data.

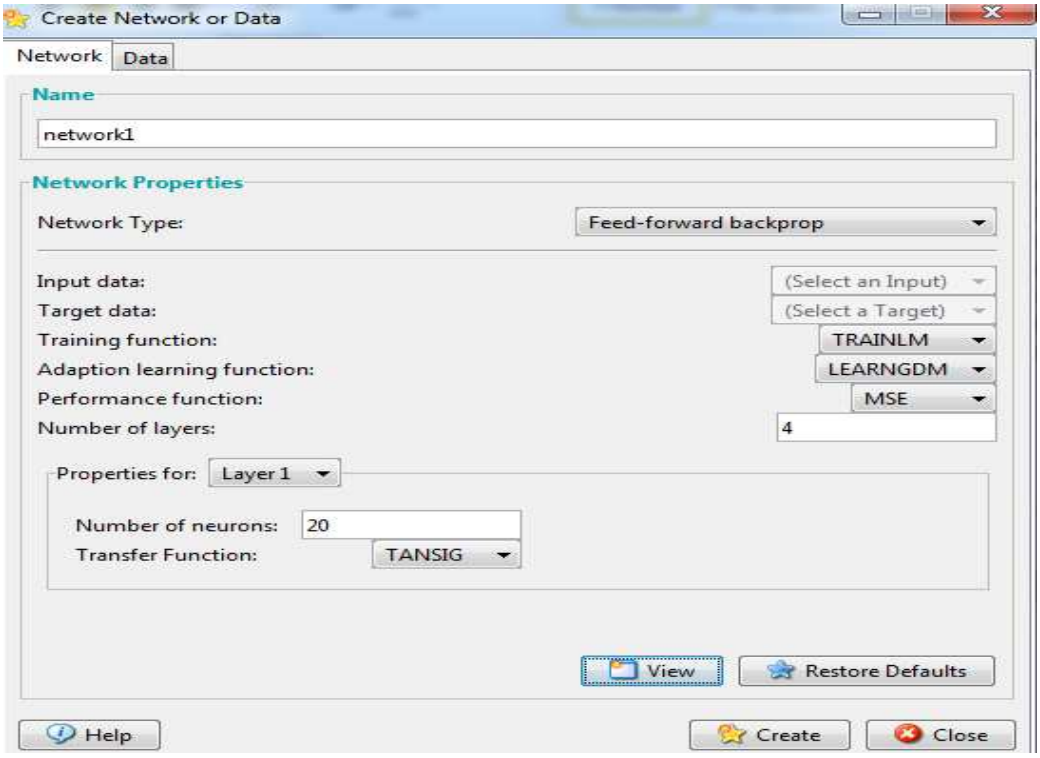

Figure 3: Generating new network

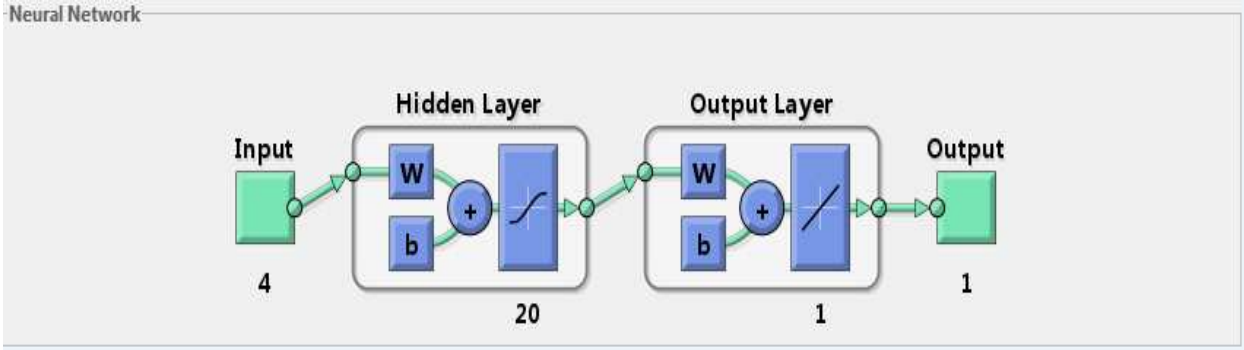

Figure 4: structure of created network

D. Application of Artificial Neural Network to predict THMs

Feed-forward back propagation training algorithm was used throughout this study to predict trihalomethane formation in a water treatment plant. This is done via correlating the relation between input data which are covered of TOC, pH, applied Cl_2 and temperature and output data that is consisted of measured trihalomethane. The structure of artificial neural network (ANN) was identified based on a prior research done by ZHAO and et al (2007). In addition that several trials were attempted till achieving the best regression outcomes with free from over-fitting see Figure 4. The network properties were selected as follows:

- Network input: TOC, pH, Applied Cl_2 and Temperature.
- Network output: Trihalomethane (THM) concentrations.
- Network type: Feed-forward back-propagation.
- Training function: Levenberg–Marquardt algorithm (TRAINLM).
- Adaptation learning function: Gradient descent with momentum weight/bias learning function (LEARNGDM).
- Performance function: Mean square error (MSE).
- Number of layers (input layer): 4
- Number of hidden layer: 20 neurons.
- Transfer function; hidden layers: TANSIG.
- Transfer function; output layer: PURELIN.
- Data were randomly divided into three subsets such that training: 75%, validation: 15% and test: 15%.

Results and Discussion

It is seen from figure 5 that the values of the gradient as well as the validation checks number were utilized to stop training of the network. The magnitude of gradient was equal to 4.3634 when the number of epoch reached 55 iterations, meaning that the training will stop as gradient reached below 10e-10, the validation checks number was equal to 55; which is believed an appropriate value for stopping a trained network.

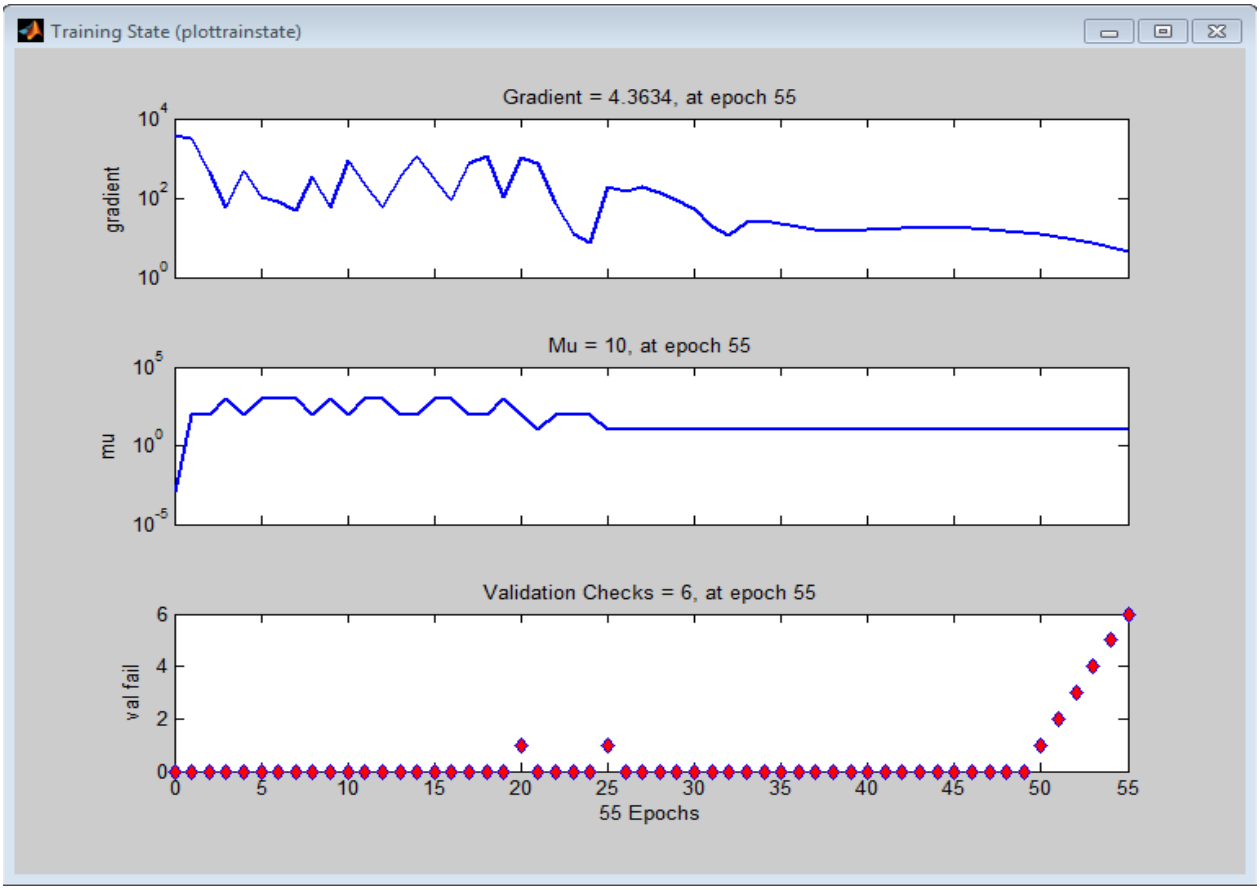

Figure 5: Training State plot of predicted ANN model

As it is presented in figure 6 the performance plot demonstrates the magnitude of the function with respect to training, validation and testing behaviors against the epoch which is known as iteration numbers. It shows that the best validation performance depending on the mean square error was 62.8539 at epoch 49. No major problems and over-fitting were occurred since the validation and test curves are very similar.

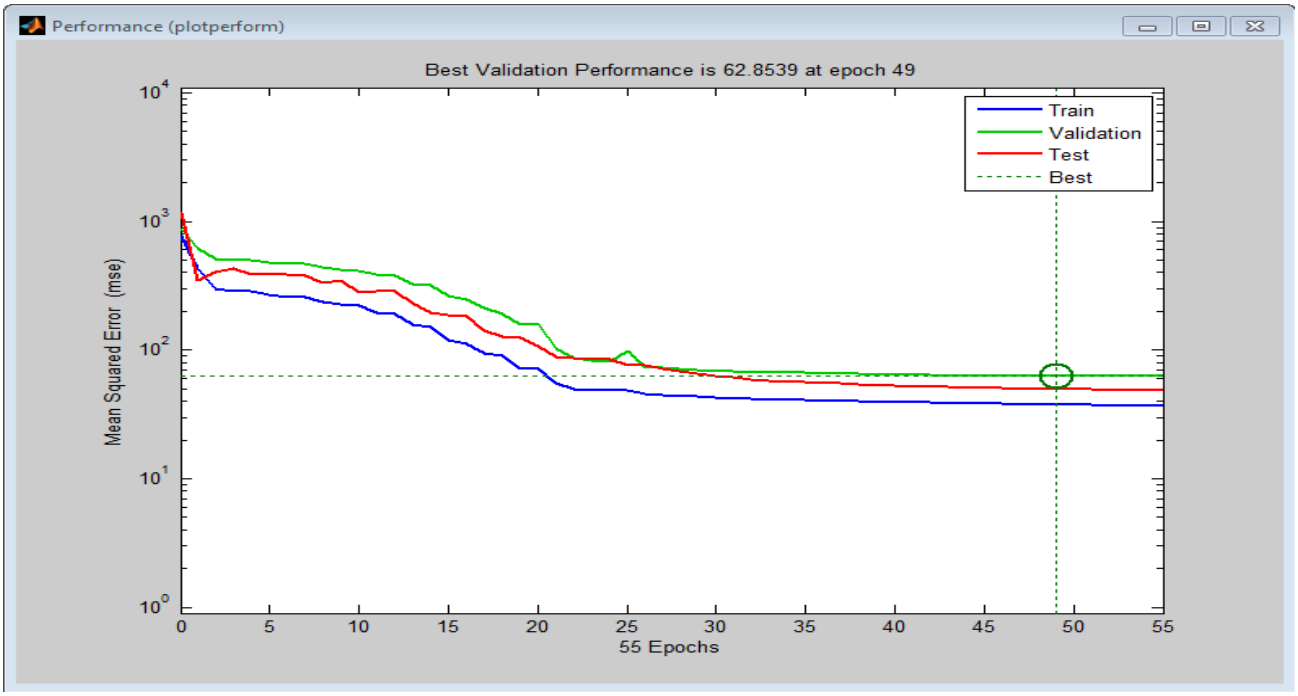

Figure 6: Performance plot of predicted ANN model

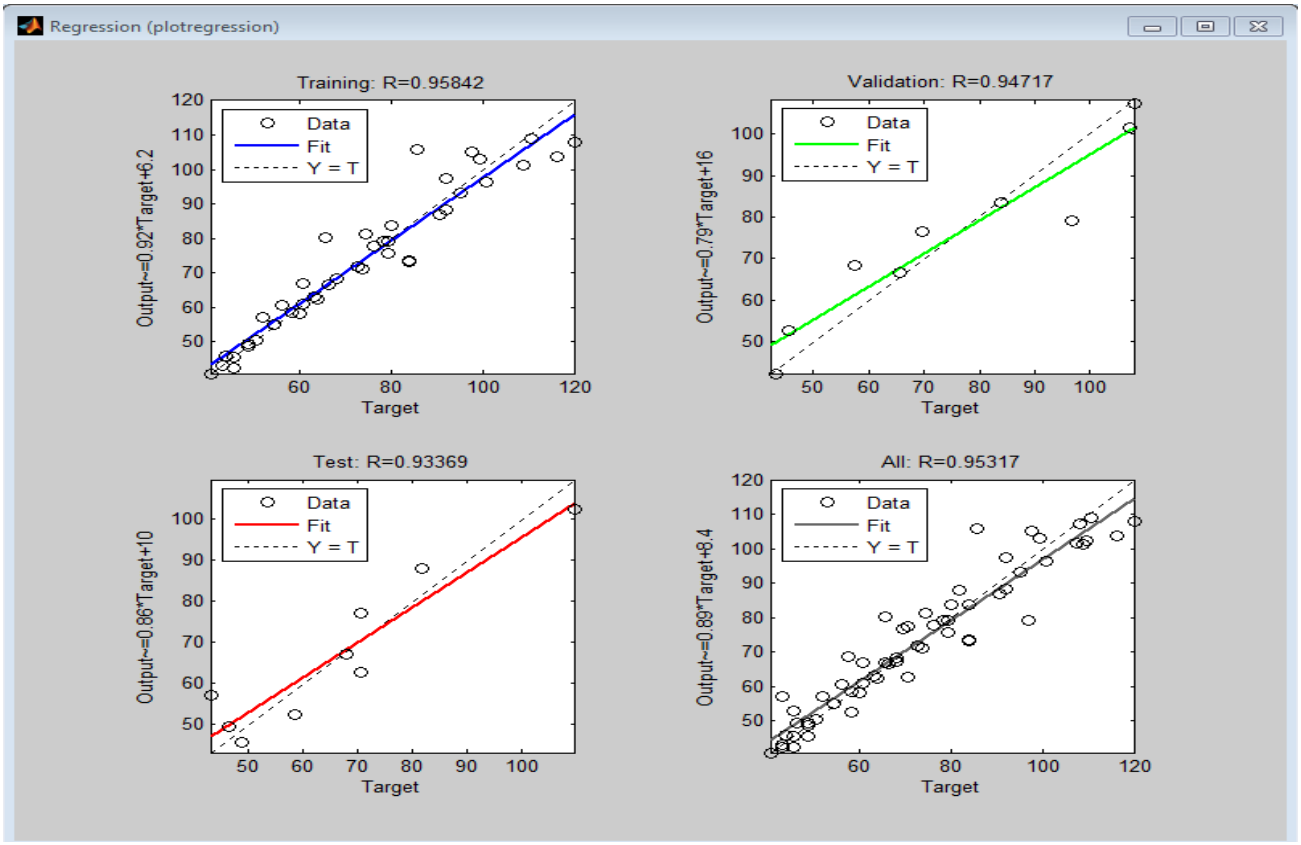

Figure 7: Regression plot of Training, Validation and Test for THMs using ANN

Figure 7 shows the regression plot of training, validation and test behaviors. In order to determine the relationship between the outputs of the network and its target an analysis of linear regression is required to be conducted for training, validation and testing. The dashed line that is shown in each plot represents the exact outcome, meaning that output is equal to target, meanwhile the solid line matches the best fit of linear regression. An exact linear relation-ship is reached as R-value which is a correlation coefficient approaches to one. The result of regression plots (R-Value) were 0.95842, 0.94717 and 0.93369 for training, validation and test respectively. Hence; all those results were matching to a whole response of 0.95317. Then the generated model can be used to predict the parameter of trihalomethane through new inputs of TOC, pH, T °C and applied Cl_2. Hashem and Karkory (2007) conducted a similar research that is used artificial neural network to predict trihalomethane formation in chlorinated water of Libya. The study found that R^2 for whole results are 0.983 between the predicted and measured date of trihalomethane using input parameters: Applied chorine dose (Cl_2), Total organic carbon(TOC), water pH, Ultra-violet (UV), water temperature (T), Bromide concentrations (Br), and reaction time of chlorine residual in water (t). The current study demonstrated that it is possible to forecast trihalomethane formation with access to input parameters such as TOC, pH, applied Cl_2 and T with a correlation coefficient of 0.9532. It was seen that the measured data of trihalomethane was very much close to the predicted data that calculated from the model configuration, hence; the validity of this model was remarkably confirmed. Figure 8 presents the results of measured trihalomethane and predicted trihalomethane using multiple linear regressions. Uyak et al. (2005) pointed out that the predicted trihalomethane curve in most cases overlapped the measured trihalomethane. The resultant of linear regression model is as follow:

$$THMs = 7.07x10^{-2}(TOC + 3.2)^{1.314}(pH4.0)^{1.496}(dose - 2.5)^{-0.197}(temp. + 10)^{0.724} \quad (6)$$

Where TOC is total organic carbon in mg/l, the dose is the chlorine in mg/l, temp. is temperature in °C.

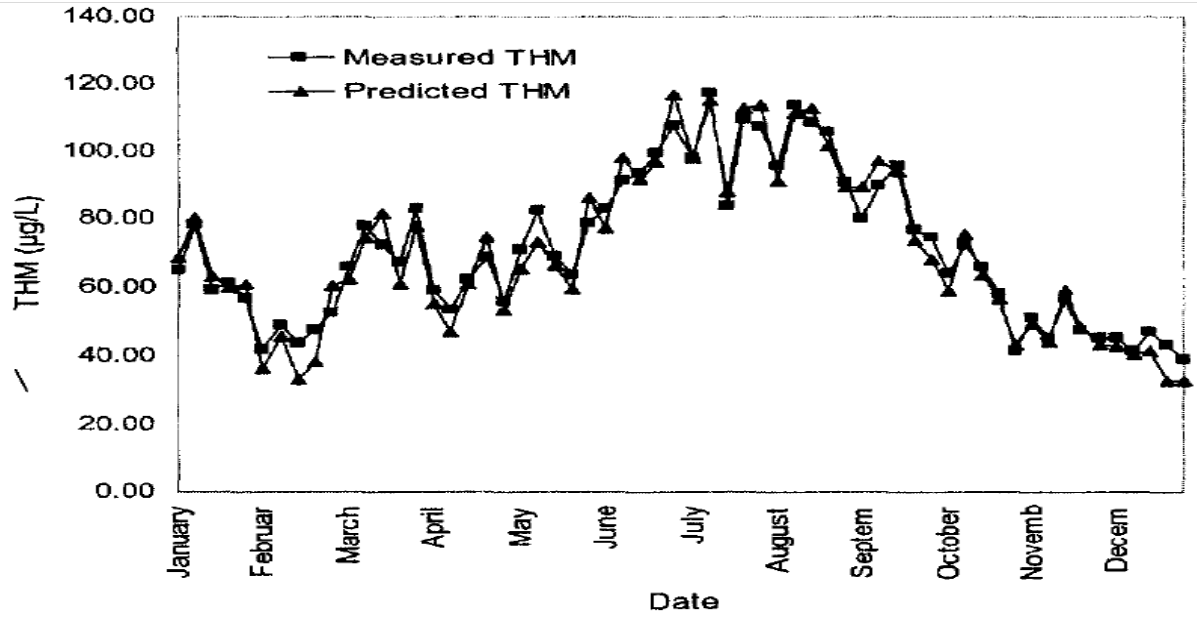

Figure 8: Measured and Predicted magnitudes of THMs by multiple linear regression (Uyak, 2005)

Figure 9 compared the results of measured trihalomethane with the predicted outputs of trihalomethane using both multiple linear regression that were acquired from the database of Uyak and others,2005; and artificial neural network. It is apparent that estimation of trihalomethane using ANN modeling is considerably more accurate than predicted trihalomethane which is estimated using multiple linear regressions.

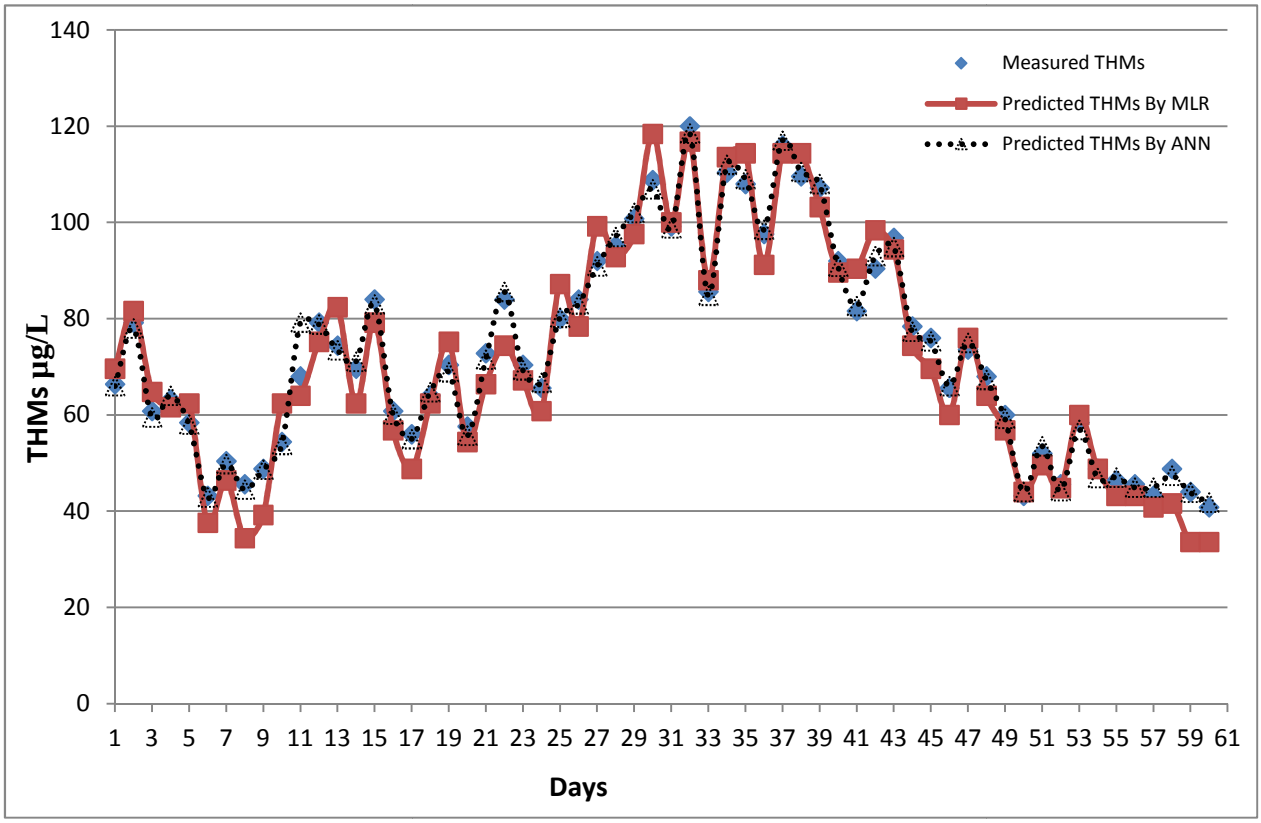

Figure 9: Measured and Predicted values of THMs using MLR and ANNs

Conclusions

This study implemented the back-propagation neural network technology and graphical user interface function of MATLAB to obtain an easier and quicker predicted of trihalomethane of chlorinated water.

Varying in TOC, pH, Temperature and applied chlorine dose have directly influenced concentration of trihalomethane formation in processed water. With the use of artificial neural network approach, it is possible developing a model that simulates and predicts the trihalomethane during water treatment plant. In this research ANN model for predicting of trihalomethane formation and comparing results with MLR was conducted. It is expected that the majority of trihalomethane formation models published in literature are based on multiple linear regressions modeling technique, therefore; the performance of such model can be considerably enhanced with the use of artificial neural network as substituted method for modeling trihalomethane. The result of regression plots (R-Value) were 0.95842, 0.94717 and 0.93369 for training, validation and test respectively. Hence; the developed artificial neural network demonstrated a well-fitted function. The outputs of this study confirmed that ANN is a suitable alternated approach over MLR for modeling THM formation in a water treatment plant. It is worth noted that this results is compatible with Hashem and Karkory (2007) research paper.

References

[1] Abdulla, M. Y. *"Formation, modeling and validation of trihalomethanes (THM) in Malaysian drinking water: a case study in the districts of Tampin,* Negeri Sembilan and Sabak Bernam, Selangor, Malaysia. Water Reaserch, Vol. 37, 4637-4644. (2003).

[2] Agency, U. S. *"Fedrelal Register. 67. Al-Suhili, R.H. and Karim, R.A. (2015) 'Daily inflow forecasting for Dukan reservoir in Iraq using artificial neural networks*", Int. J. Water, Vol. 9, No. 2, pp.194–208. (2002).

[3] Bellar, T. L. *"The Occurence of Organohalides in Chlorinated Drinking Water*". American Water Works Associations, Vol. 66, pp. 703-706. (1974).

[4] Cantor, K. H. *"Bladder cancer, drinking water source, and tap water consumption: a case-control study*". Journal of National Cancer Institiute, Vol. 79, pp. 1269-1279. (1987).

[5] Chowdhury, Z. A. *"Modelling disinfection by-product formation, in: Formation and Control of Disinfection By-products in Drinking Water*". American Water Work Associations. (1999).

[6] EECD. *"Amended proposal for a Council Directive concerning the quality of water intended for human consumption common position.* In: Proc. Council of the European Union, Directive 80/778/EEC, Com (97) 228, Final 95/0010 SYN, Brussels. (1997).

[7] Elshorbagy, W. Q. *"Simulation of THM species in water distribution systems"*. Water Reaserch, Vol. 34, pp. 3431-3439. (2000).

[8] EPA, US. *"National Primary Drinking Water Regulations: Disinfectants and Disinfection Byproducts*", Final Rule,. 40 CGR Part 9, pp. 141-142. (1998).

[9] EPA, US.. Federal Register. 67. (2002)

[10] Gallard, H. A. *"Chlorination of natural organic matter: kinetics of chlorination and of THM formation"*. Water Reaserch, Vol. 36, pp. 65-74. (2002).

[11] Golfinopoulos, S. *"Use of a multiple regression model for predicting trihalomethane formation"*. Water Reaserch, Vol. 32, pp. 2821-2829. (1998).

[12] Golfinopoulos, S. A. *"Quantitative assessment of trihalomethane formation using simulations of reaction kinetics*". Water Reaserch, Vol. 36, pp. 2856-2868. (2002).

[13] Govindaraju, R. A. *"Artificial Neural Networks in Hydrology*". Boston, USA: Kluwer Academic Publishers. (2000).

[14] Hammerstrom, D. *"Neural networks at work*". IEEE Spectrum, pp. 26-32. (1993).

[15] Hashem, M. A. *"Using artificial neural network model as a new technique for simulating residual chlorine*". Journal of Engineering Sciences, Vol. 33, pp. 735-743. (2005).

[16] Hashem, M. A. *"Artificial Neural Networks as Alternative Approach for Perdicting Trihalomethane Formation in Chlorinated Waters"*. Eleventh International Water Technology Conference, IWTC11 2007 Sharm El-Sheikh, (pp. 703-710). Egypt. (2007).

[17] Heller, M. A. *"Forecasting with cascade correlation: an application to potable water demand*" . ANNIE, pp 1150-1160. (1994).

[18] Haykin, S. *"Neural Networks: A Comprehensive Foundation*". Prentice-Hall, New Jersey, 842 pp. (1999).

[19] Joo, D. C. *"Determination of optimal coagulant dosing rate using an artificial neural network*". Wat. Suppl.: Res. & Technol., Vol. 49, pp. 49-55. (2000).

[20] Maier, H. A. *"Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and application*". Environmetal Modelling & Software, Vol. 15, pp. 101-124. (2000).

[21] MathWorks, Inc. 2010a. *"MATLAB"*. 3 Apple Hill Drive, Natick, MA, USA.

[22] McGeehin, M. *"Case-control study of bladder cancer and water disinfection methods in Colorado*". American Journal of Epidemiology, Vol. 138, pp. 492-501. (1993).

[23] Menard, S. *"Applied logistic regression analysis*". Quantitative Appl. Social Sci., 7-106. (1995).

[24] Milot, J. R. *"Modelling the susceptibility of drinking water utilities to form high concentrations of trihalomethanes*". Environmental Management, Vol. 60, pp. 155-171. (2000).

[25] Neter, J. W. *"Applied linear statistical models*". 3rd edition. Irwin: Homewood, IL. (1990).

[26] Rodriguez, M. *"Trihalomethanes in drinking water of greater Quebec region (Canada): occurrence, variations and modelling*". Environmental Monitoring Assessment, Vol. 89, pp. 69-93. (2003).

[27] Rodriguez, M. M. *"Predicting trihalomethane formation in chlorinated waters using multivariate regression and neural networks"*. Wat. Suppl.: Res. & Technol., Vol. 52, pp. 199-215. (2003).

[28] Rook, J. *"Formation of haloforms during chlorination of natural waters*". Water Treatment Exam, Vol. 23, pp. 234-243. (1974).

[29] Rumelhart, D. W. *"The basic idea of neural networks*". Communications of the ACM, Vol. 37, pp. 87-92. (1994).

[30] Saeed, R. A. and George, L. E. *"Apply Pruning Algorithm for Optimizing Feed Forward Neural Networks for Crack Identifications in Francis Turbine Runner"*. International Journal of Soft Computing and Engineering (IJSCE), Vol. 2, No. 4. (2012).

[31] Sohn, J. *"Monitoring and modeling of disinfection by-products (DBPs)*". Environmetal Monitoring Assessment, Vol. 70, pp. 211-222. (2001).

[32] Uyak, V. T. *"Monitoring and modeling of trihalomethanes (THMs) for a water treatment plant in Istanbul*". Desalination, Vol. 176 , pp. 91-101. (2005).

[33] White, G. C. *"Handbook of Chlorination and Alternative Disinfections*", 3rd ed. New York: Van Nostrand Reinhold, (1992).

[3'] ZHAO, Y and et al *"Water quality forecast through application of BP neural network at Yuqiao reservoir*". Journal of Zhejiang University SCIENCE A, Vol. 8, No. 9, pp. 1482-1487. (2007).