*Indian Journal of Fundamental and Applied Life Sciences ISSN: 2231– 6345 (Online)*
*An Open Access, Online International Journal Available at www.cibtech.org/sp.ed/jls/2014/04/jls.htm*
*2014 Vol. 4 (S4), pp. 630-636/Taalohi and Tabatabaee*

***Research Article***

# PREDICTING BAR DAM WATER QUALITY USING NEURAL-FUZZY INFERENCE SYSTEM

**Mahmoud Taalohi[1] and *Hamid Tabatabaee[2]**
*[1]Department of Computer Science and Technology, Islamic Azad University, Mashhad, Iran*
*[2]Young Researchers and Elite Club, Quchan Branch, Islamic Azad University, Quchan, Iran*
**Author for Correspondence*

## ABSTRACT

Dams and rivers are considered as the major sources of drinking water supply, agriculture and industry. They are also subject to several quality fluctuations because they pass by various beds, regions and they are in direct contact with their environment. Due to the complexity and multiplicity of qualitative processes of surface water resources, using adaptive neural fuzzy inference system that is capable of learning and understanding these kinds processes without the need for governing equations, is a new method presented for predicting the quality of rivers and dams. In this article, the principle of this system is stated along with its application in sampling a set of real-time data during 50 months of data collection period from the dam and Bar River at six stations. This application was shown based on different data sets and ANFIS method .Because of the inference and reasoning capabilities of fuzzy systems and the learning ability of neural networks, ANFIS model can be used as a multivariable model to predict the properties. The process of development and evaluation are based on the learning data and test data, and in this article the prediction of the results was appropriate but due to the unavailability of further statistical samples in some of the sampling stations, the prediction has not shown a good performance. Finally, the updated predicted values were compared with the measured values through the model. Using adaptive neural fuzzy inference system can be applied as a new approach to predict the quality of dams, water transmission lines, treatment plant sand rivers where sufficient data are presented for validation training process.

***Keywords:*** *Water Quality, Dams, Fuzzy Logic, Neural Networks, Water Resources Management*

## INTRODUCTION

Over the past few decades due to the limited fresh water resources, the issue of protecting the quality of water resources has attracted the attention of water resource managers and researchers. Meanwhile, the quality management of surface water resources such as rivers, lakes and estuaries is more important because of the ease of access and the direct exposure to different kinds of contaminants and extensive research has been done in this area. Over the past two decades, due to increasing population and expanding industries and the production of a variety of pollutants in urban, industrial and agricultural areas, disposal of pollutants in surface water resources which are the main suppliers of water needs are endangering the health of communities. To evaluate and estimate the ability of recipient sources in compliance with water source quality standards, there is a need for a qualitative modeling and study of surface sources behaviors toward the drain of different kinds of contaminations. Without a doubt, qualitative modeling cannot be independent of hydraulic and hydrodynamic characteristics of the river and the contaminants profile. With the disposal of pollutants in to rivers, their transmission and diffusion in the downstream is associated with physical, chemical and biological reactions. Because of one-dimensional assumption of flow of rivers due to their limited depth and width in comparison to the length of the river, the study of river quality changes are conducted both one- dimensional and longitudinal manner. Also, longitudinal and transverse diffusion which are in a relatively short length from the contaminated area are disregarded. Proper and efficient use of mathematical models for river quality simulation in order to determine comprehensive and appropriate policies for evaluating the methods of reducing pollution and their management in particular is very significant. From the mid-1970s, with the rapid development of rigorous tools, computational tools, computers, and the development of numerical methods for solving partial differential equations, river qualitative model in ghada considerable growth (Little and Williams, 1975).

The number of simulation parameters, being one or multi- dimensional, methods to solve the governing equations numerically, transmission mechanism and uncertainty analysis are developed and applied in different regions of the world. In all these models, there are limitations due to the calibration coefficient of the reaction, particularly with the increase in number of parameters and their mutual effects on each other, their accuracy in predicting the qualitative behavior of rivers will be reduced .Extensive research has been done in order to calibrate the quality of rivers from which Davidson's work in determine the longitudinal dispersion coefficient (Yih and Davidson, 1975), Mulligan and Brown (1998) in using genetic algorithms for calibrating river quality models quality (Mulligan and Brown, 1998). In analyzing the sensitivity calibration of river quality models in dynamic mode (Van *et al.*, 2002), in using genetic algorithm for calibration and verification of qualitative models of damsare worth to be pointed out.

Because of existence of non-linear equations governing the transport processes of pollution in rivers and the complexity of solving them simultaneously, and the existence of multi ply kinetic constants and coefficients, it is hard to use models and methods with physical basis. Many of these models are applicable only for simple mode and within the conditions that have been calibrated and they are not appropriate for situations outside the desired range of accuracy. Uncertainty and ambiguity in the use of Fuzzy Inference System and in particular its combination with adaptive neural networks should be considered as a new approach. The main objective of this study is to use the capabilities of neural fuzzy inference system in predicting the quality of the river and the reservoir levels was selected as the case study.

***Adaptive Neuro-fuzzy Inference System***

Water quality processes are often nonlinear and complex in the rivers, therefore the use of models with basic physical and mainly linear equations can cause many limitations in modeling and prediction by them. As a result, the use of neural networks, fuzzy inference systems, or a combination of them can be considered as alternative methods. Adaptive Neuro-fuzzy inference system is an approach to solve complex problems for which either there is no specific algorithms for solving, or for which the use of models with basic physical equations requires the application of approximate and severe restrictions on them. The main feature of this system understands the behavior of the desired nonlinear systems (Jang *et al.*, 1993).

Fuzzy system is a system based on logical rules of condition and result, which is not analyzable through classic theories of probability. The purpose of fuzzy logic is to extract accurate results using a series of inaccurate information which is defined through lingual words and sentences. The starting point for creating a fuzzy system is obtaining a set of “if-then” principles. Possessing a method by which one can use the available data in order to make these principles is considered as an efficient tool. On the other hand, because of their learning capabilities using different training patterns, artificial neural networks can make a good relationship between input and output variables. Therefore, combined application of fuzzy inference system and artificial neural networks which are capable of predicting the results using available numerical data is a powerful tool that is called "Adaptive Neuro-Fuzzy Inference System". In this combined application, the fuzzy part makes a relationship between input and output variables (Jang and Gulley, 1995), and the parameters of membership functions in fuzzy part are determined through neural networks (Chang and Ning, 2001).

In adaptive neuro-fuzzy inference system, first the structure of a model with certain parameters is selected which is proportional to inputs, degree of membership, principles and functions of output membership degree. Then, some of available data are chosen as input-output and are used in order to train the system. In training stage, parameters of the model approach to the real values through modifying the parameters of membership degree based on acceptable error rate. After system training and parameter selection, accuracy measurement and validation of the model are required. As the data used in system training are not necessarily complete representatives to a comprehensive training, validation of the model is of great importance. Hence, another part of dataset of input-output which is not used to train the fuzzy inference system is applied as input data of the system in order to make sure about accuracy and validity of the built fuzzy system model to predict corresponding output data. Validity of adaptive neuro-fuzzy inference system is determined thorough another set of data which is known as control dataset. One must be careful about selecting the control data, because they are representatives of all sets of training data for the model to be able to have an appropriate prediction. In practice, half of the data are used in training stage and the remainders of them are applied in accuracy measurement and validation. Often, adaptive neuro-fuzzy inference system is applied through a Sugeno fuzzy system in the form of a progressive network structure which is made of 5 layers (McKone and Deshpande, 2005).

***The 1st Layer*** contains input nodes and each node is a fuzzy set. The output of each node is input variable membership degree in this fuzzy set.

***The 2nd Layer*** contains principle nodes and each node in this layer calculates activity degree of a rule.

***The 3rd Layer*** contains average nodes and calculates the proportion of principle activity to the sum of total principle activities.

***The 4th Layer*** is made of result nodes and the output of each node is calculated by the result parameters.

***The 5th Layer*** is made of output nodes and each node gives final output value.

Methodology of training in this system is back error propagation. Through this method, when moving forward, error rate is propagated towards the inputs using error gradient descent algorithm, and the parameters are corrected. Therefore, outputs of the nodes are calculated up to the 4th layer in each round of training. Then, the result parameters are calculated by error least squares method. After calculation of error when moving backward, ratio of error is propagated towards the condition parameters and their values are corrected through error gradient descent. The general form of the system is shown in Figure.1.

In the mentioned equation, x and y are non-fuzzy inputs of node i; Ai and Bi are fuzzy membership functions. Also, value of membership for each input as the output of the 1st layer is specified as $\mu_{A_i}(X)$ and $\mu_{B_i}(Y)$. Therefore, parameters of membership functions must be specified; they are known as the front part of fuzzy rules and are non-linear parameters (Mamdani *et al.*, 1976).

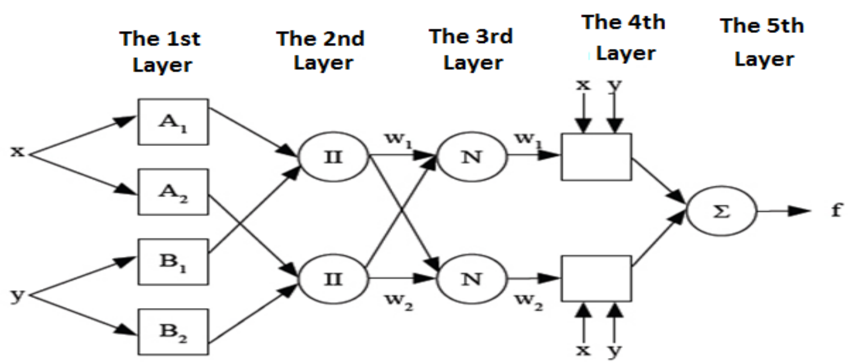


**Figure 1: General Form of Adaptive Neuro-Fuzzy Inference System (ANFIS)**

***Formulas andMathematical Equations***

To evaluate the performance of various criteria including errors (errors), mean square error, (RMSE) Average Error (error_mean) and the standard deviation of error (error_std) between the observed and predicted values, the following equations are used.

1. $errors = \sum_{i=1}^{N} Targets(i) - outputs(i)$

Targets (i) :estimated cases (SAR,TH,Na%)

Outputs (i):a combination of inputs and outputs learning data ANFIS

2. $RMSE = \sqrt{\frac{\left(\sum_{i=1}^{N} Targets\ (i) - outputs\ (i)\right)^2}{N}}$

N: total number of data samples

RMSE: Mean Square Error

3. $error_{mean} = \frac{(\sum_{I=1}^{N} Targets\ (i) - outputs\ (i))}{N}$

4. $error\_std = \sqrt{\frac{1}{N-1} * \sum_{i=1}^{N}((\sum_{I=1}^{N} Targets(i) - outputs(i)) - \overline{(\sum_{I=1}^{N} Targets(i) - outputs(i))})^2}$

Error_std: Standard deviation of error

***Case Study: Bar River and Dam***

Bar dam is built at the 35th kilometer of Firooz-Neishaboorroad, on a river of the same name. The clay structure has a height of 37 meters and a crest length of 1,500 meters, which is capable of storing about 24 million cubic meters of water. An example of the topography of Bar Dam can be seen in Figure 2 (Regional water in Khorasan Razavi, 1983).

Due to its geographical location and the existence of cities, steel plant and scattered industries in its vicinity, this river is recipient of various pollution which is caused by household waste, agricultural waste and industrial pollution which is increasing day by day. This causes a severe reduction of the water quality from Bar Dam site to Neishaboor, so that annually several million cubic meters of the reservoir and the river water is drained by the river by the surrounding industries, municipal wastewater and drainage from agricultural land. Many researches has been done to check the status of Bar River water quality, its simulation and modeling that is mainly based on data collected from several sampling stations (Commission of sampling and test methods for water, 2013).

**Figure 2: The topography of Bar Dam**

***Figures and Charts***

Several quality parameters including pH, dissolved oxygen, conductivity, biochemical oxygen demand, chemical oxygen demand, total dissolved solids, nitrate, sulfate, phosphate, chloride and metals has often been measured on a monthly basis from 2010 till now. In this article, to predict water quality in the reservoir and the river TH, SAR, % Na is considered as the main indicators. Also, 6 sampling stations which are respectively: the river, the one meter, 5 meters, 10 meters, 15 meters and 20 meters from the reservoir level of Bar Dam were considered. Therefore, the prediction of TH, SAR, % Na, have been made in the six stations  using adaptive neuro-fuzzy inference systems for river and dam . It should be noted that this prediction system can be used for all stations and transmission lines but in this paper, this sampling station was skipped.

The effective parameters to calculate the value of TH, SAR, % Na at the station include: electrical conductivity (EC), total dissolved solids (TDS) HCO3, CL, SO4, Ca, Mg, Na. In this study, the four year data which was available the supply and operation center of Khorasan Regional Water was used. The results of implementing 6 Adaptive Neuro-Fuzzy Inference System for 6 existing stations are predicted according to figure 3, in which TH in the 5 meter station and the river, SAR in 10-meter station, and NA % in 1- meter and 15- meter station,  learning data, category partition testing by networking along with the error frequency chart are  shown. It should be noted that in figure 3 some examples of the stations are presented for predicting the estimated parameters. In this study, the test data that is used merely for testing the accuracy of ANFIS system performance, offered acceptable results for this prediction according to learning data. Below in figure 4 TH prediction charts for 4 six stations can be observed but due to the unavailability of data at station 1, 10, 15 the predicted data have not faced a good approximation. In figure 5, prediction charts are presented for all the SAR stations in which the system has an acceptable prediction. In figure 6, all the charts for predicting the indicator NA% are presented which shows a good prediction like SAR charts.

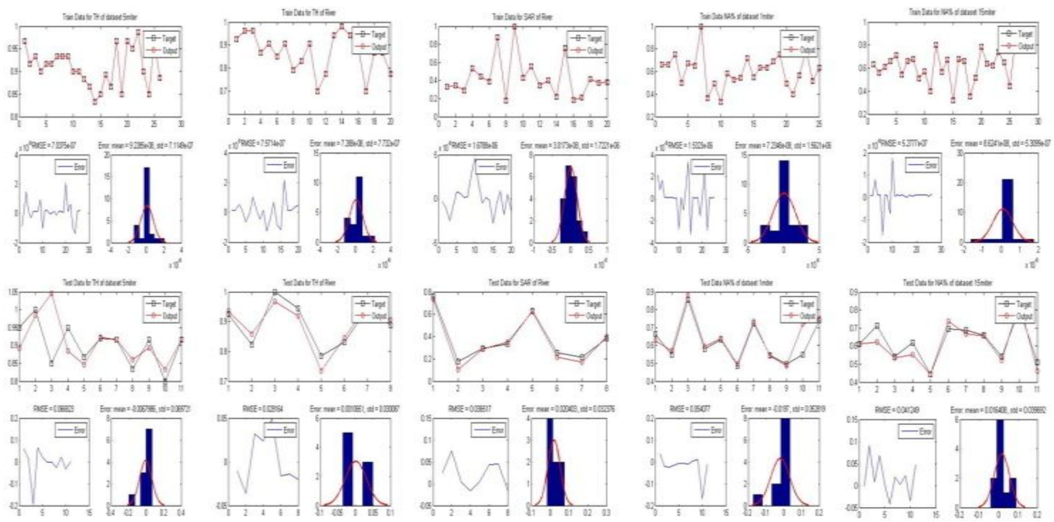

**Figure 3: Presenting charts of Na%, SAT, TH at different stations**

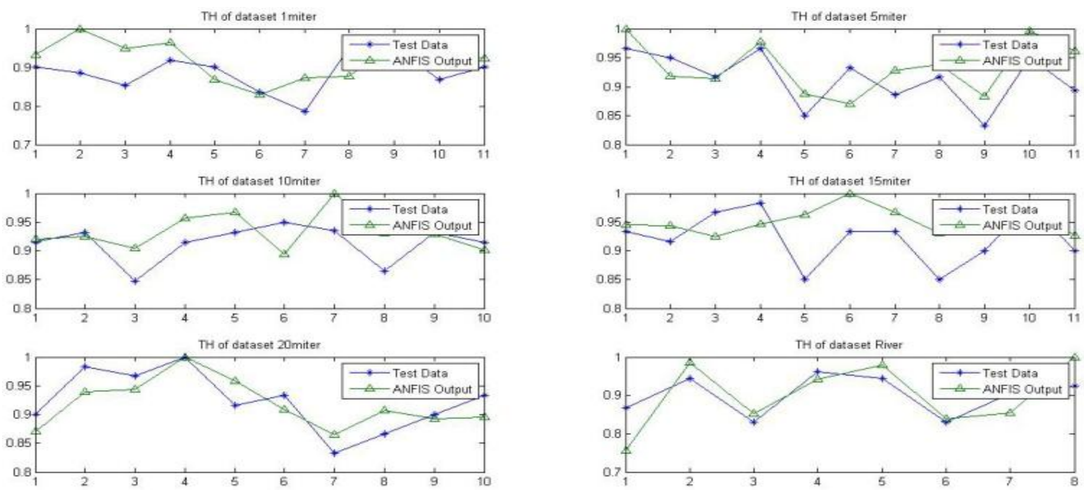


**Figure 4: Presenting the predicted values of TH for 6 stations**

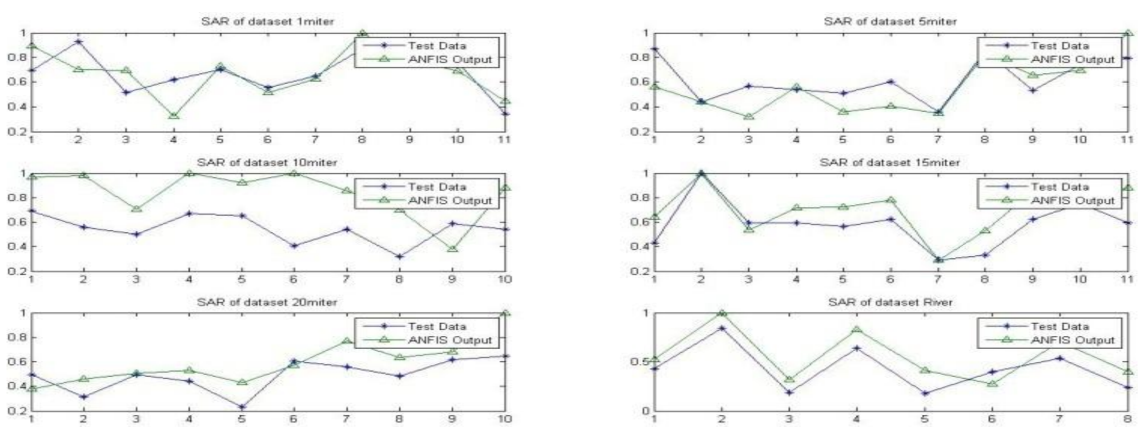


**Figure 5: Presenting the predicted values of SAR for 6 stations**

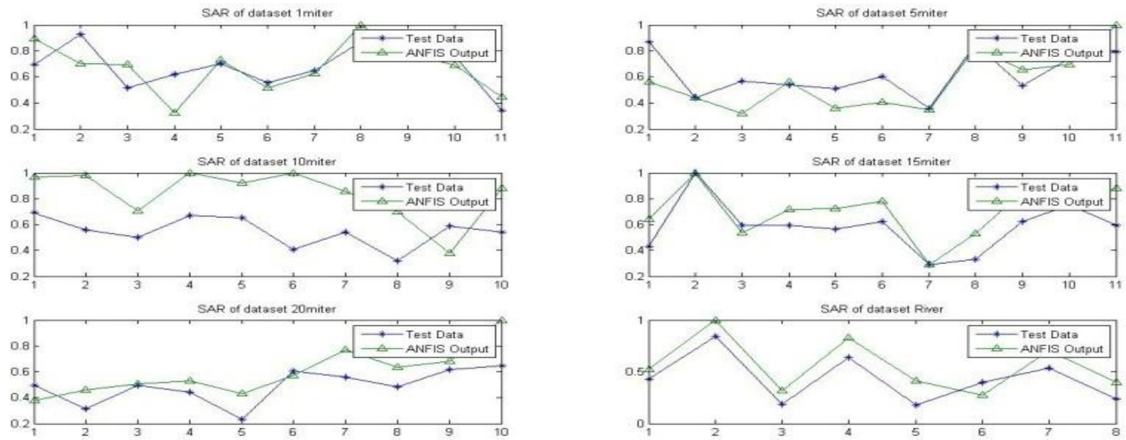


**Figure 6: Presenting the predicted values of Na% for 6 stations**

As it is evident, the performance of all three models in predicting values is acceptable which indicates the ability of ANFIS in predicting based on a training data set (learning). On the other hand, in all three samples the estimate index of the river station is the closest to the training data, which is why it is the right and proper range of samples.

Also, the x-axis in Figures 4, 5, 6 shows the length of a number of the test data (target) and the y-axis in these figures shows ANFIS output.

***Conclusion***

In this paper, we describe the basics of the application of adaptive neuro-fuzzy inference system for quality prediction of surface water sources in rivers that have enough data for both training and testing stages. The accuracy of these models was studied, and it was determined that if the concept of processes and effective input parameters are well defined, even though the set of processes are in form of complex and non-linear analysis and generally there is a need for estimating many of the constants and coefficients, acceptable results can be achieved. As for the case of the river and reservoir levels of Bar dam, in order to predict TH, SAR, NA, the correlation coefficient of 0.953 in calibration stage and 0.931 in the validation stage was obtained by the system. Comparison of the results obtained from the use of three different models showed that the increase in the number of input parameters for the test data necessarily increases the accuracy of the prediction model. So that ANFIS system is suitable for prediction when the right time is not important to them and the deficit lies in predicted shortage of the test data. The comparison of adaptive neuro-fuzzy inference system model results with the measured data showed the high accuracy of the system. However, it should be noted that although the data-driven models do not have the need for the basic physical equations, which are mostly non-linear and complex, understanding the process and the impact of desired factors of the input data on output it is very important and you cannot expect the desired results in case of lacking this understanding.

**REFERENCES**

**Chang NB, Chen HW and Ning SK (2001).** Identification of river water quality using the fuzzy synthetic evaluation approach. *Journal of Environmental Management* **63**(3) 293-305.

**Commission of Sampling and Test Methods for Water (2013).** First edition of IranWater sampling. Standard number, 2347.

**Jang JSR (1993).** ANFIS: Adaptive-network-based fuzzy inference systems, *IEEE Transaction Systems* (Man and Cybernetics) **23** 665-685.

**Jang JSR and Gulley N (1995).** *The Fuzzy Logic Tool Box for Use with MATLAB* (The MathworksInc) Natick, MA.

**Little KW and Williams RE (1975).** Least-squares calibration of QUAL2E, *Water Environment Research* **64**(2) 179-185.

**Mamdani EH (1976).** Advances in the linguistic synthesis of fuzzy controllers. *International Journal of Man-Machine Studies* **8** 669-678.

**McKone TE and Deshpande AW (2005).** Can fuzzy logic bring complex environmental problems into focus? *Environmental Science & Technology* **39**(2) 42A-45A.

**Mulligan AE and Brown LC (1998).** Genetic algorithms for calibrating water quality models, *Journal of Environmental Engineering* **124**(3) 202-211.

**Regional Water in Khorasan Razavi-design and Development (1983).** Institute of Standards and Industrial Research of Iran. *Water Sampling*, Standard number 2347.

**Van Griensven A, Francos A and Bauwens W (2002).** Sensitivity analysis and autocalibration of an integral dynamic model for river water quality, *Water Science and Technology* **43**(7) 321-328.

**Yih SM and Davidson B (1975).** Identification in nonlinear, distributed water quality models, *Water Resources Research* **11**(5) 693-704.