Assessment of dissolved oxygen in Shatt Al-Arab River by other quality parameters of water using Artificial Neural Networks

Zaynab A.Khudhur[a], Saad A.Arab[b] and Ammar S.Dawood[c]

[a,b,c] Civil Engineering Department, Engineering College, University of Basrah , Iraq.

ABSTRACT
The Major sources of water are surface and subsurface. Surface water includes Rivers, Reservoirs, Creek, Streams, etc. This paper deals with using– a neural network model to recognize dissolved oxygen in Shatt Al-Arab. Within the present study, Shatt Al-Arab River (Basrah-Iraq) is considered as the study area with monthly observed data from 2009-2014. Artificial Neural Network (ANN) has been applied to pattern the relations among eight (8) water quality parameters which are devoted for predicting one parameter (1) so that to decrease the load of long experimental procedure. Physical and chemical parameters that are inserted in the model are: pH, total dissolved solids, electrical conductivity, sulphate, phosphate, calcium, magnesium and nitrate. Dissolved oxygen (DO) is included in the output models. The three layered feed-forward model with back-propagation multi-layer perception (MLP) models architecture of 8-8-1 for DO. The artificial neural network has got training successfully and has been tested with 70% and 30% of the data groups. Statistical criteria of correlation coefficient (R^2) and mean square error (MSE) are used to evaluate performance of the models. The correlation coefficients of the artificial neural network model for predicting DO have been 0.99354 and 0.98237, and mean square error for the model are 0.007698 and 0.00122 respectively. It can be concluding that these techniques provide similar accuracy in estimating DO concentration and predicting the dissolved oxygen (DO) in Shatt Al-Arab

Introduction

Water is an essential natural resource for sustaining the environment and life, which is always thought to be available in abundance and the gift of Allah. Rivers, lakes, and springs are the most significant source of fresh water and they represent dynamic system and their nature may be changed several times during their course due to changes in physical circumstances such as slope and bedrock geology and chemical properties due to the dissolved gases, organic and non-organic materials. They hold a horizontal and continuous one-way flow of a significant load of the matter dissolved and particulate matters from both natural and anthropogenic sources [1].

The quality of water could be illustrated as the chemical, physical and biological features of water, often in relation to its suitability for the designated use. Another general perception of water quality is that of the simple property that tells whether the water is polluted or not. As it is known, water is used numerously, for example: in recreation, beverages, fishing, agriculture, and industrial products. Such uses require different chemical, physical and biological qualities which are required to uphold these uses. e.g. High-qualities are expected for water that is specified for drinking and swimming in comparison with the one which is used in agricultural needs, and some industries. The physical and chemical properties of water play a direct role in the distribution and behavior of organisms and have a significant impact on the suitability of this water for human consumption.

water quality standards are defining the water quality that depends on the intended use of the water which may by either human consumption or it may be for industries, irrigation, recreation, etc. The standard will provide benefits in the provision of quality water by eliminating pollutants, which are harmful to human health, or by lowering their concentrations to a minimum [2].

One of the most commonly used indicators of overall ecological health of a body of water is dissolved oxygen (DO). It is used as a signal for the health of a water body, wherein high levels of productivity and little pollution will result higher dissolved oxygen concentrations. It is an important parameter when assessing the quality of water due to the influence on the organisms living in that particular body of water. In the study of lakes - limnology, the dissolved oxygen is an extremely important factor. A dissolved oxygen average which is either too high or low could affect water systems and harm aquatic life [3].

Biodegradable organic material is one of the most widespread types of river pollutants. When a raw sewage, which is high-energy organic material, is discharged into a stream, many changes show up downstream from the discharge point . Pollutants in Shatt Al-Arab have increased fast in the aquatic

environment of Iraq in the past two decades, which are due to:

• Inefficiency and lack of sanitation projects of health

• Disqualification units of sewage treatment prior to release to the river,

• Failure to comply with treatment of water used in industrial projects, and emptied directly to the river without treatment.

• Illegal sewerage connection by industrial projects for the disposal of their polluted water without treatment

• Throwing household waste in the river and on its bank.

. Artificial neural network is a smart technique and a typical black box model [4]. Pollution indicators are related to the quantity and quality of pollutants in water bodies [5] which are estimated by using the relationships between dependent and independent variables. The present study focuses on the use of artificial neural network models on Shatt Al-Arab so that to explore the most suitable simulated neural network architecture for the process of recognizing the pollution of water and evaluating the action of the processes of artificial neural network models after eliminating the input parameters that have less significance by the stepwise regression analysis.

Recently, many applications in many engineering fields as well as in many other scientific fields have successfully applied neural networks [6]. Feed-forward neural networks are commonly adopted in these applications. The standard error back-propagation algorithm is used to train the networks. Nevertheless, a fundamental limitation of such training algorithm is that the architecture of the network must be specified in advance. This indicates that a suitable architecture has to be designed, by the end user, by a costly trial-and-error approach. As a result, neural networks are considered good for regression and classification tasks in practical cases. This makes ANN a very efficient tool to solve and deal with many engineering problems, particularly the problems that have complex or insufficient data [7].

This paper focuses on assessing the water quality of Shatt Al-Arab River due to this river subjected to varying degree of pollution caused by numerous untreated and or/ or partially treated. Therefore, the water quality modeling is estimating the (2009-2014) measured values of Shatt Al-Arab River as sample data to conduct training and testing in order to verity the viability of the model. In addition, objective of this study is to investigate whether it is possible to predict the values of water quality parameters measured by a water quality monitoring.

Material and Methods

1. **General description of Shatt Al-Arab River**

The location of Shatt Al Arab is in the lower Mesopotamian delta in Basra/Iraq. The confluence of Tigris and Euphrates at Al-Qurna city, Basra (31° 00'17"N and 47° 26'29"E) have formed it . Then, it flows from north to south of Basra to the Arabian Gulf [8]. Recently, the discharge of Tigris and the tides from the Arabian Gulf affected the hydrological regime of Shatt Al Arab [9]. Shatt Al Arab River has a 200 km length, a (0.25-2) km width range and a depth varies on the range (5-12) m [10].
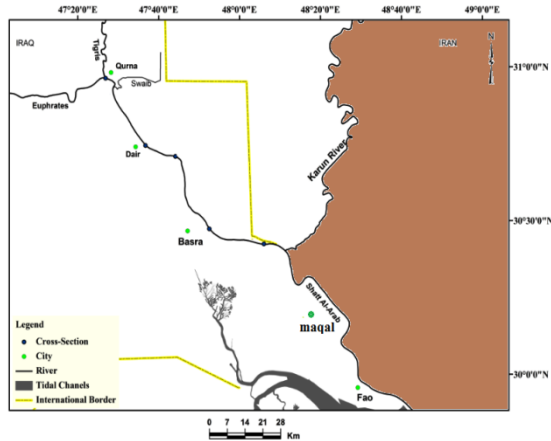

Figure 1:Map of Shatt Al-Arab River

The tides in the Gulf are nearly semidiurnal, and their influence is spread towards Al-Qurna city in the north of Basrah [11].There has been increasing in the salinity of the water of Shatt Al Arab recently and this is related to both the quantity and quality deterioration of the incoming fresh water and the seawater interference. In general, many factors can affect the quality of water of the Shatt Al-Arab [12]. Shatt Al-Arab is the main fresh water source in the south of Iraq (Figure1). It has a remarkable role in the development of the industrial and agricultural sectors in the country and has additional values related to transportation, irrigation and the provision of drinking water. The physical, chemical and hydrological regime aspects of the river's water have been affected by the quality of water coming from its tributaries, which in turn, depend on the amount of rainfall. The effects of groundwater and the water storage circumstances in the upstream areas are highly significant. At its mouth, the river is also affected by the saline water tides which come from the Arabian Gulf [13].

2. Water quality data

The data that are used in this study were provided by Ministry of Water Resources in Iraq-Basrah for the year 2009-2014 which represented the monthly measurements at rate a twice measured values for fourteen water parameters that are used to model water quality for Shatt Al-Arab River. 538 data were chosen for the study to have a complete idea of the overall water quality of the river in its area and for the

purpose of estimating the water quality. The physic-chemical of water quality parameters are: pH value, Total hardness (TH), Chlorides (Cl) , total dissolved solids (TDS), electrical conductivity (EC), Calcium (Ca), Sulfate (SO4), Nitrate (NO3), Alkalinity (ALK), Magnesium (Mg), sodium (Na), potassium (k) ,dissolved oxygen (DO) and PO4" as shown in Table(1).

Table 1: Physic-Chemical parameters of water quality of Shatt Al-Arab River

Parameters	units	Experimental value	
		Min	Max
Dissolved oxygen	mg/l	2.19	13.3
pH value	-	6.7	9.2
Electrical conductivity	μS/cm	1303	52200
Total dissolved solids	mg/l	96	39462
Calcium	mg/l	35	1008
Sulfate	mg/l	100	3000
Magnesium	mg/l	27	1728
Nitrate	mg/l	3.212	19.64
PO4	mg/l	0.03	2.56
Chlorides	mg/l	142	19550
Alkalinity	mg/l	100	1905
Total hardness	mg/l	8.2	8600

Input selection technique

The inputs of a strong correlation with dissolved oxygen were chosen by an input selection method. A basic assumption for this paper is the idea that the input has to be removed when it has a low correlation coefficient. The reason is that having a smaller number than this one means weak correlation exists.

Using back-propagation learning algorithm for feed-forward neural networks, the research purpose is evaluating the intelligent model in determining the quality of water by considering the values of DO as output of the model. Since measuring DO is basic , such values are proposed as outputs of the FFNN based model that considers the power of "hydrogen (pH), Phosphate(PO4), Magnesium (Mg), Calcium (Ca), Sulphates (SO4),electrical conductivity (EC) and total dissolved solid (TDS) as inputs of the model".

Artificial neural networks

The source of the core of ANNs theory is the need for creating a processing system that functions like the human brain which is highly complex as it contains billions of neurons and inter-neuronal connections. The mechanism of this is not recognized fully. However, it is suggested that it has a complex network of neurons which function in a parallel manner to analyze information. The Artificial neural networks are designed to simulate such structure and processing philosophy within a computer environment. Implementing these networks, parallel computation and high computing performance are

necessary to do specific learning tasks. ANNs, which are based on a recent mathematical model, have highly influenced scientists. The interest in these networks has been increased enormously in several scientific fields recently as part of the need to solve some complex issues as they have the ability of learning form any internal representation of data in whatsoever form.

The most important benefit of ANN techniques might be their nonparametric nature. i.e. the underlying idea is that the frequency distribution of the data does not exist. The features of the training data (or their internal structure) are acquired iteratively, thus, they might be considered as data-dependent models. ANNs are supposed to achieve much accurate results out of small training input sets in comparison with conventional statistical classifiers [14].

The mathematical model

A network is formed when artificial neurons are connected together. The processing node is the fundamental element of ANNs (Figure
2) that theoretically matches the neuron of the human brain. The processing nodes acquires and sums sets of input values, and pass these sums by an activation operation that provides the output values of the nodes. These, then, form inputs to a processing node in the following layer of ANN.

The mathematical symbol, x(n) represents various inputs to the network. Each of these inputs is multiplied by a connection weight. W(n) which refers to the weights. For simple cases, such products are summed and lead to a

transfer function (activation function) to produce a result which represent the output. This could happen with other network structures that have other summing and transfer functions.
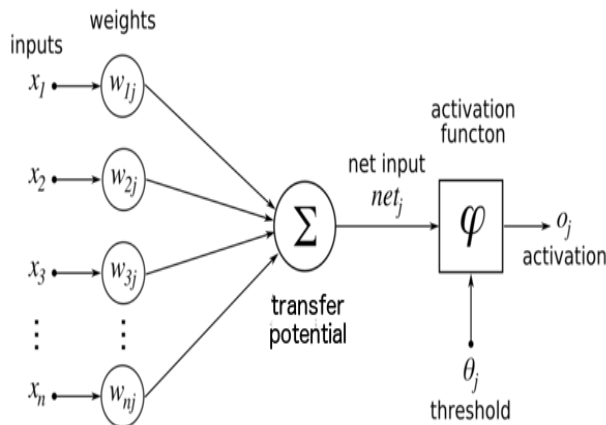

Figure 2: Model of artificial neurons

Architectural of ANN model

Artificial neural networks (ANN) are used to assessment and forecasting of dissolved oxygen in shat Al-Arab River by other water quality parameters. The neural network toolbox, in MATLAB (R2013a), is used for implementing the neural network in this study. This program has many advantages such as containing several types of networks and implementing many different training algorithms, such as the back-propagation algorithm.

In this study, two case studies were studied in:
• Evaluation of the overall quality of Shatt Al-Arab water during the period (2009-2014) of previously recorded data
• Forecast the dissolved oxygen of the Shatt al-Arab River.

Table 2: Specifications of the Proposed Network

Items	Descriptions
Nodes in the input layer	8
Hidden layers	[1 – 2]
Nodes in the hidden layers	[1-25]
Type of activation function	1st : tansig 2nd : purelin 3rd : purelin
Training function	Different types of functions
Nodes in the output layer	1

Back-Propagation training functions

In applying the proposed model (ANN), the Levenberg-Marquardt back-propagation algorithm (trainlm) was first used to train the network of the proposed model (ANN). This network is trained with different back-propagation algorithms to choose the algorithm that gives the best results.

The data set of training is devoted for training the network. After that, the network is tested through the test data set. The optimal size of the network is selected from the one that we acquire as a result of the highest coefficient of correlation (R) and least root mean square error (MSE) from the data sets of both training and testing. The whole data sets of water quality of the river is classified to two sub-sets to identify ANN. These data subsets of training and testing (calibration and validation) (or testing) comprise (80% and 20%), respectively or (70% and 30%), respectively.

Table 3:Variables Training , Testing and Range

Item	Variables	Range of Data	
		Training Min- Max	Testing Min- Max
Input variables	PH -	6.7 – 9.2	7 -8.6
	PO4(mg/l)	0.03- 2.56	0.119- 1.19
	SO4(mg/l)	220 -3000	100 - 220
	NO3(mg/l)	6.98 - 19.64	3.212- 6.97
	Mg (mg/l)	57- 1728	27- 57
	Ca (mg/l)	120 - 1008	35 - 120
	TDS (mg/l)	39462-1260	96 - 1258
	Ec (µS/cm)	1905 - 52200	1303 -1897
Output variables	DO(mg/l)	2.4 -13.30	2.19 – 11.2

Table 4: ANN operation having one and two hidden Layers for the average of training and testing data (80:20)

Training function	One hidden layer				Two hidden layer			
	Nodes NO	mse (test)	R (test)	Epoch	Nodes NO	mse (test)	R (test)	Epoch
trainlm	19	0.166	0.9899	500	[8,12]	0.00123	0.9786	100
trainrp	8	0.00298	0.9616	1000	[10,12]	0.00511	0.9734	1000
traingda	10	0.00211	0.9736	5000	[10,8]	0.000112	0.9726	5000
traingdx	10	0.00256	0.9733	5000	[8,8]	0.00113	0.9703	5000
traincgf	10	0.00212	0.9765	500	[9,8]	0.00188	0.9536	300
Traincgp	10	0.00998	0.971	500	[10,5]	0.00155	0.9562	300
traincgb	8	0.00144	0.9677	500	[12,5]	0.000998	0.9735	343
trainscg	8	0.00178	0.96701	500	[12,7]	0.0144	0.9845	300
trainbfg	12	0.00198	0.9801	300	[9,12]	0.0112	0.9824	1000
trainoss	8	0.00476	0.9733	300	[10,10]	0.0132	0.9838	1000
traingdm								

Table 5: ANN operation with One and Two Hidden Layers for ratio of training and testing data (70:30)

13

Training function	One hidden layer				Two hidden layers			
	Nodes NO	mse (test)	R (test)	Epoch	Nodes NO	mse (test)	R (test)	Epoch
trainlm	19	0.02322	0.64156	500	[8,12]	0.0312	0.82341	100
trainrp	8	0.0212	0.95207	1000	[10,12]	0.00199	0.98455	1000
traingda	10	0.000711	0.98438	5000	[10,8]	0.00521	0.98372	5000
traingdx	10	0.000721	0.98482	5000	[8,8]	0.00545	0.9803	5000
traincgf	10	0.00187	0.97753	500	[9,8]	0.00199	0.97843	300
Traincgp	10	0.00178	0.98333	500	[10,5]	0.00223	0.97258	300
traincgb	8	0.00544	0.97996	500	[12,5]	0.00298	0.93638	343
trainscg	8	0.00122	0.98237	500	[12,7]	0.0132	0.991	300
trainbfg	12	0.0211	0.9483	300	[9,12]	0.00689	0.97733	1000
trainoss	8	0.00321	0.98407	300	[10,10]	0.00187	0.99067	1000
traingdm	8	0.0114	0.98628	100000	[9,10]	0.00355	0.97642	5000

Figure 3:Regression in one hidden layer and two hidden layer of the suggested Network

Comparatively study

The proposed neural model (ANN) is used to predict the dissolved oxygen of the water quality for Shatt Al-Arab River using some selected testing set and training set of data collection. A comparison between the two models (ANN)and for almost specimens the proposed model has given more accurate results.

1.   The first ratio of training to test data records employed in the experiment is 80:20. The Levenberg-Marquardt algorithm and different types of training functions that are used by algorithms for training the neural network and the hidden units which influence the network operation directly. Thus, several investigations have been conducted. The optimal result is provided by specifying the number of hidden nodes. As shown in Table 4, (trainlm) training functions provide the best testing performances having one layer and (traincgb) training functions provide the best testing performances having two hidden layers.
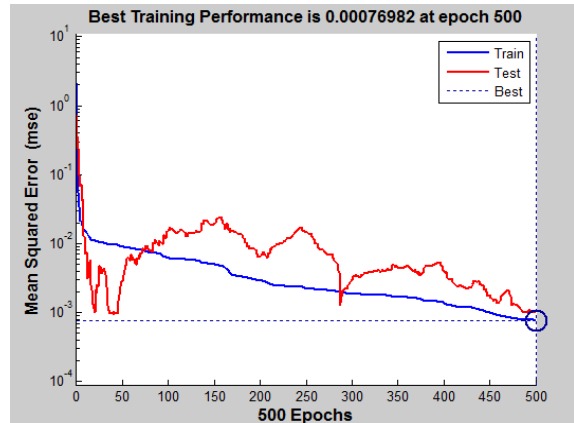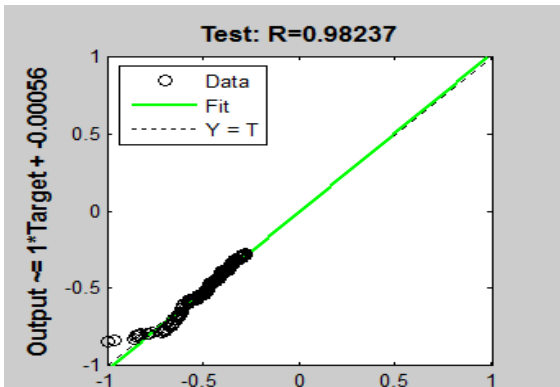
The architecture of the network, as a result, is 8- 19 -1. The hidden nodes are set to 19. No significant difference between the results is available having two hidden layers but provided the most suitable performance having mse = 0.000998 as well as R = 0.9735.

2.   70:30 is the second average of training to test data sets that are used in the present experiment. The neural networks are trained by using different types of training functions algorithms and the process of the network is directly influenced by the number of hidden units. As it is shown in Table 5, the weak testing performance was provided with one and two hidden layers by (trainlm) training function, but the most suitable testing operation which was given to one and two hidden layers by other types of training function.

. Figures 4 illustrates the regression while 5, shows mse of the suggested network of the second ratio.
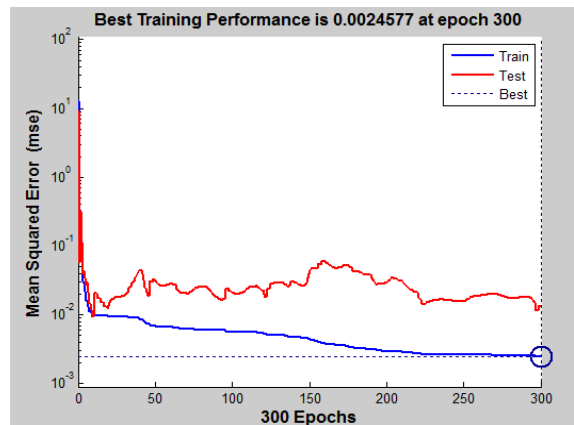
Results AND Discussion

Choosing an optimal ANN design, not only the type of network, but also the design of that network gets its performance limited in terms of quality and speed. One of the significant matters of ANN design is to find a good ANN architecture. A good ANN architecture might be regarded as one yielding good performance in terms of error minimization, while a simple and compact structure is kept. The number of input units and output neurons are problem dependent, but the hardness lies when the optimal structure of a network is determined in terms of hidden neurons and layers, which might be chosen freely. Unfortunately, the design of such an architecture has no rule.

Generally, an appropriate and parsimonious architecture for a problem can be recognized by applying a trial-and-error procedure. Other possibilities, besides trial and error, involve the use of algorithms by which the combination of training and ANN architecture optimization is featured.

For neural network model construction with monthly data randomly partitioned into (training 80%, testing 20% and taining70%, testing30%), the cascade correlation algorithm chooses the training of neural network which is feed-forward and supervised algorithm. The least value of MSE of the training and testing groups specified the best number of neurons. High homogeneity is available among the tested data which is due to high correlation. Eight to twenty neurons every time have been applied individually in the network so that the most suitable number of neurons in the hidden layers is found.

In the present study, it was noticed that the prediction of DO concentrations is affected by ANN model with ratio (80:20) trained with algorithm (trainlm). The selection structure had the highest correlation value ($r = 0.9899$) and the least error (MSE = 0.0166 for Test data) after epoch 500 for one hidden layer, the structure of the network has been 8- 19 -1. The input nodes have been 8 which represent the parameters of water quality which influence the dissolved oxygen, the hidden nodes have been 19, and the output nodes are 1.

In addition to the fact that the prediction of DO concentrations is affected by train with algorithm (traincgb). The choice structure had the highest correlation value ($r = 0.9845$) and the least error (RMSE = 0.00144 for Test data) after epoch 300 for two hidden layer, the structure of the network is 8-[12,5] -1. The input nodes are 8 which represent the parameters of water quality by which dissolved oxygen is affected. The hidden nodes are [12,5], and the output nodes are 1. The optimal average of hidden neurons is nineteen as this could give the maximum correlation coefficient and minimum errors.

The second ration for the same model which found the training algorithm (trainscg) obtained best performance indicators for both hidden layers:

(R=0.98237 MSR=0.0007698) for one hidden layer and (R=0.991 MSR=0.0024577)for two hidden layer. Artificial neural network is assumed to have significant capability of learning and making predictions. The results has shown that DO could be predicted if unknown data is given through the neural networks. Thus, The developed form of ANN model could be used effectively in predicating the water quality. According to the results of these models, there is a good correlation between dissolved oxygen and other variables, which is (pH, TDS, EC, SO_4, PO_4, Ca, Mg and NO_3). These models can be used to predict dissolved oxygen in the future and can be used safely, which gives a great explanation of the quality of the river's water.

These variables (pH, TDS, EC, SO_4, PO_4, Ca, Mg and NO_3) have a great effect on dissolved oxygen, by increasing and decreasing it, and therefore it will affect the water quality.

Conclusions

Many hydrological studies have used the ANNs and this was an influencing factor for their applications to the present study. RMSE was used to test the performance of ANN and R. Coefficient of correlation (R) and root mean square error statistics (RMSE) were used to evaluate performance of the models. The test results referred that the ANN model with minimum input parameters, pH, total dissolved solid (TDS), EC, Ca, Mg, SO_4, PO_4 and NO_3. The results show that the correlation coefficients and mean square error between the same model for both ratio, which obtained second ratio 70:30 (70% training and 30% testing) better than first ratio due to have results high correlation and low error. The model has shown ability of recognizing the pattern of the water quality parameters to give significant prediction concerning changes of water quality of Shatt Al-Arab every month. The optimal networks are indicated to be able to record long-termed data spotted for the tedious water quality changes. Therefore, Artificial Neural Network is highly efficient in computing the water quality and can be applied on other water bodies to raise the awareness of water pollution trends.

Reference

[1] Bakan, Gülfem ; Hülya Böke Özkoc ;Sevtap Tülek ; Hüseyin Cüce (2010)."Integrated Environmental Quality Assessment of Kizilirmak River and its Coastal Environment" . Turk. J. Fish. Aquat. Sci., 10(4): 453-462

[2] F. Baltaci, , 2003 "Determination of the Quality of Water Resources and Related Standards", pollution effects and solutions in Turkey's inland water resources, water 73-86 p.

[3] Kemker, C. Fondriest Environmental (2014). "Conductivity, Salinity and Total Dissolved Solids." Fundamentals of Environmental Measurements.".

[4]   Abdulkadir, T. S., Sule, B. F. and Salami, A. W. 2012. Application of Artificial Neural Network Model to the Management of Hydropower Reservoirs along River Niger, Nigeria. Annals of Faculty Engineering, Hunedoara–International Journal of Engineering, Tome X –FASCICULE 1(ISSN 1584-2673), 419-424

[5]   Waziri, M. and Ogugbuaja, V. O. 2010. Interrelationships between physicochemical water pollution indicators: A case study of River Yobe-Nigeria; Am. J. Sci. Ind. Res., 2010, 1, 1, 76-80.

[6]   Cheu, R. L., Ritchie, S. G., Recker, W. W., and Bavarian, B. (1991). \Investigation of a Neural Network Model for Freeway Incident Detection". Proc. 2nd Int. Conf. on the Applicability of AI to Civil and Structural Eng.. B. H. V. Topping, ed., Vol. 1, AI and Civil Eng., Oxford, England. Civil-Comp Press, pp. 267-274.

[7]   Weigend, A. S., Huberman, B. A., and Rumelhart, D. E. (1990). \Predicting the Future:A Connectionist Approach". Int. Journal of Neural Systems., 1, 193-209.

[8]   MED INGEGNERIA S.r.l, Studio Galli Ingegneria S.p.A and EL CONCORDE LLC, 2011, "Main Report MED517-Technical Report-Volume I_Rev01".

[9]   Moyal, M., S., 2014, "Assessment of Water Quality of the Shatt Al-Arab River Using Multivariate Statistical Technique" Mesopotamia Environmental Journal (MEJ), Vol.1, No.1:39-46.

[10]   Abdulla, S. S., 1990 "An Investigation to River Load of Shatt Al-Arab in Basrah" Unpublished M.Sc. Thesis, Marine Science Centre, Basrah Univ., pp.98.

[11]   Al-Mahdi, A. A., Abdullah, S. S., and Husain, N. A. 2009,"Some Features of The Physical Oceanography in Iraqi Marine Waters." Mesopotamian J. Mar. Sci, 24: 13-24.

[12]   Consulting Engineering Bureau, 2011,"Shatt Al Arab, Final Study", College of Engineering University of Baghdad,.

[13]   Garcia, A., Revilla, J.A, Medina, R., Alvarez, C. & Juanes, J. A., 2002,"A model for predicting the tempoeral evolution of dissolved oxygen concentration in shallow estuaries", Hydrobiologia, pp 205-211.

[14]   Foody, G. M. (1995) Using prior knowledge in artificial neural network classification with a minimal training set. International Journal of Remote Sensing, 16, 301-312.