Comparison of Artificial Neural Network and Multivariate Linear Regression Model to Predict Sodium adsorption ratio (SAR) (Case Study: Sistan River, Iran)

Nooshin Sarani, Jaber Soltani, Shirahmad Sarani, Ali Moasheri

***Abstract*--**One of the most important factors in development of any region is availability of appropriate water supplies. Among the most important is water quality parameter, sodium absorption ratio (SAR). Thus, various models used for predicting quality of water. In this study, using the parameters of chlorine (CL), salinity (EC), alkalinity (PH) and total dissolved substances (TDS) measured at station Kohak Dam, located on the river Sistan during 1997-2011, qualitative parameters using SAR artificial neural networks MLP and GFF was predicted. The multivariate regression model to compare the neural network also is used, Finally, considering with all of the models, results showed that use of MLP neural network with one hidden layer is used, with $R^2 = 0.92$ and MSE = 1.05 better than the other models would predict. Results showed that artificial neural network is better performance than the regression model.

***Keywords*--** Neural networks, Multivariate regression, SAR, Sistan River

I. INTRODUCTION

Due to the problems of water quality and quantity; codifying the quality management plans for all water resources, is essential strategy and unavoidable, in order to conservation of water resources. Nowadays, the role of water resources, and particularly river systems, as the vital artery of each region, one of the main bases of stable development are considered [11]. The rivers as the water sources for sectors of agriculture, industry, urban have had a particular importance and because of passing through the platforms and different regions, have qualitative variations. Therefore, the evaluation and prediction of water quality parameters should be considered along a river. Knowledge of condition of surface waters quality provides the possibility to use it in different ways leading into the least damage to the source. Generally, quality of irrigation water by factors such as sodium absorption ratio (SAR), soluble sodium percentage (SSP), the remaining calcium carbonate (RSC) and electrical conductivity (EC) is determined. Among the most important water quality parameters is sodium absorption ratio (SAR), which is used to determine the desirability of irrigation water. Generally, the high SAR causes undesirability [10].

Nooshin Sarani, Jaber Soltani, Shirahmad Sarani and Ali Moasheri are with University of Zabol, College of Agriculture, P.O. Box: 98615-538 R. IRAN, Tel: 0098 (0542) 2240711,Fax: (0542) 2230540, Email: saraninooshin@gmail.com

In recent years different applications of artificial neural network model is presented. Among the hydrological issues, which in recent years have been reviewed by artificial neural networks can be mentioned to predict flooding, modeling rainfall - runoff, water demand, water quality modeling, groundwater management and so on. According to the non-linear nature of hydrological phenomenon, application of artificial neural networks in science is justified [1]. In this study, the performance of multi-layer perceptron artificial neural network algorithm and multiple regressions to predict sodium absorption ratio (SAR) changes were evaluated in the River Sistan. Data about the stations ranged between 1997 to 2011 Kohak coastal dam, the concentration of chlorine (Cl), salinity (EC), alkalinity (PH) and total dissolved substances (TDS) as input parameters and the sodium absorption ratio (SAR) is an output parameter. With the increasing development of computer techniques such as artificial intelligence, artificial neural network model inspired by the structure of the human brain, This model is widely anticipated in the study of various parameters related to water resources, has been used and the accuracy of this method compared with the experimental and regression relationships, have emphasized [6]. Including research on the application of artificial neural networks in determining water quality; can be indicated to Hore *et al* [2] the estimated channel parameters affecting water contamination in India. Also Huiqun and Ling [3] in research in China, the Dong Chang River water quality by fuzzy logic and artificial neural networks evaluated. Kunwar *et al* [5] used in the Guttmann India the perceptron neural networks in modeling DO and BOD water quality parameters. The purpose of this study, the setting of a multi-layer perceptron neural network structure for modeling of water quality (SAR) is a river in Sistan, so in this study, the possibility of training the neural network to model the independent parameter (SAR) using related parameters were evaluated.

II. MATERIALS AND METHODS

Geographical area and data used

Sistan region, northern province of Sistan-Baluchistan, with an area of 15,197 km, equivalent to between 29 to 32 degrees north latitude and between 60 and 64 degrees east along the

prime meridian. Climate of Sistan, due to is located near the Loot desert is hot and dry in summer, and is a kind of dry and warm. The maximum annual temperature is 40 to 45 degrees. The average annual temperature of 21.7 ° C and relative humidity varies between 2 to 52. The River Sistan that in the first of it there are Chah Nimeh reservoirs, is about 70 km wide and over to the plains of Sistan plain Hirmand falls. Kohak, Zahak and Sistan diversion dams on the river, have been constructed to provide water for agricultural lands. In the domestic field, the longest data is related to Kohak dam in Sistan River at the beginning of Iran. For this research, Kohak of hydrometric stations, located along the river in Sistan; the years 1997 to 2011 time interval measured by the Department of Water Affairs Sistan, has been used. Desired parameters include: CL, PH, EC and TDS is to simulate the SAR.

Introduction to Neural Networks

Artificial neural networks, first introduced in 1943 by Mc storm and Pits, But this method was without using for long periods but with development of computers and the algorithms, the networks enter a new phase [8]. The overall structure of artificial neural networks (ANNs) network is biologically inspired by the human brain. Artificial neural network systems that are able to perform arithmetic operations such as the nervous system are normal. Neural network structure is divided to single-layer and multi-layer networks. Multi-layer neural network of three layers including input layer, an intermediate or hidden layer and output layer is composed of [4]. Common types of artificial neural network models of multi-layer perceptron model (MLP), radial basis function models (RBF) network and a public whipping (GFF) [9].

Modeling using artificial neural network

Development of an artificial neural network model, involves the design and technical elements constituting it. To achieve the desired goals from the perceptron neural networks with different structures were tried to the best and most efficient network, with the amount of error and used it. In order to analyze the factors affecting the sensitivity of SAR was used. Finally, select the appropriate model and optimization of the parameters R^2 and MSE were used. Artificial neural network for modeling used the NeuroSolution software. The above model has an input layer, one hidden layer and one output layer. This software has ability to solve optimization problems such as prediction, classification and the estimation of function. Also, in order to increase accuracy and speed of artificial neural networks, data must be normalized in the range of 1 and 0 are used.

Analysis: Since the software NeuroSolution can normalization the data, so this step was performed automatically by the software. Also this software was dedicated the values 60, 15 and 25 percent of the total data, respectively, for education, health surveys and model testing

Pre-processing: Internal display data has been studied and statistical information about each column is visible.

Design: This can be most effective structure of the neural network was designed and Activation.

Education: Graphs can be trained using the training process, the error distribution and the distribution histogram of the input and distribution of weight and was looking for.

Using charts, training, and distribution of the error distribution histogram of the input and weight distribution, the training process can be pursued and investigated. Testing: In this phase, the trained network performance is analyzed and the model output graphs and charts in real output distribution can be examined. Performance of artificial neural network (ANN)

1. Weight is based on the input and output pairs
2. Begun weights with random values
3. This is an example of the learned input
4. All calculations on the input output
5. Modified to reduce the weight of
6. Repeat for all samples tested
7. To end stability with minimal weight change of the error

Transfer functions and output functions are used: Tanh function with domain R (real numbers) and range function in the interval [0, 1] as the Transfer function is expressed:

$$f(x_i, w_i) = \tanh[x_i^{lm}] \quad (1)$$

Linear function with domain R (real numbers) and range function in R as the Transfer function is expressed:

$$f(x_i, w_i) = \beta x_i + w_i \quad (2)$$

Sensitivity analysis: In this study, for the sensitivity analysis model, Statsoft method was used. So that the coefficient sensitivity of input variables, divided by the total error in the absence of a network variable on the total error in the network, all input variables, is obtained. Under this method, if the sensitivity coefficient of a variable is more than one, that variant has a large proportion of explaining dependent variable.

Multivariate regression model: In this model, the change of variables in this equation is a function of changes in other variables, the dependent variable or response variable and other variables called the independent variable. Predict the dependent variables and test research hypotheses, the main purposes of the multivariate regression analysis [12].

The general model is as follows:

$$y=\beta_0+\beta_1\chi_1+\beta_2\chi_2+\ldots.+\beta_{P-1}\chi_{P-1}+\varepsilon \quad (3)$$

Where Y the dependent variable, which is the parameter of quality SAR, β_0 constant model χ_1, χ_2, ..., χ_{P-1} independent variables, β_0, β_1, ..., β_{P-1} regression coefficients and ε represents the error that I can have a zero mean and variance 2σ and will follow the normal distribution [7].

Simulation method:

After training each network configuration, performance evaluation model for the artificial neural networks with different topologies and multiple regression to determine the optimal number of repetitions of the statistics, (the coefficient Correlation Coefficient: r), the mean square error (Mean Square Error : MSE) and mean absolute error (Mean Absolute Error: MAE) was used.

$$R = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2(Y_i - \bar{Y})^2}} \quad (4)$$

$$MSE = \frac{1}{n}\sum[z(X_i) - z(X_i)]^2 \quad (5)$$

$$MAE = \frac{\sum|X_i - Y_i|}{n} \quad (6)$$

These amounts to the following relations are obtained. In the above formula, Xi and Yi, i th real data and the estimates and $\bar{X}$, $\bar{Y}$ and the average Xi and Yi data and $z(X_i)$ value observed in the i th point, $z(X_i)$ the estimated amount of the i th point, n the number of Xi and Yi are examples.

III. Results and Discussion

Artificial Neural Networks

To estimate the SAR quality parameter; using artificial neural network, input factors introduced to the network, including chlorine (CL), PH, TDS, EC and SAR factor is introduced by output.

Functions in artificial neural network designed in this study are described in Table 1:

TABLE 1 - THE ARCHITECTURE OF ARTIFICIAL NEURAL NETWORKS DESIGNED

The output function	The third stimulus	The second stimulus	The first stimulus	The number of hidden layers	Network	Network Index
line araxon	-		tana xon	1	MLP	A
line araxon	-	line araxon	tana xon	2	MLP	B
line araxon	line araxon	line araxon	tana xon	3	MLP	C
araxon	araxon	araxon	xon		LP	
line araxon	-	-	tana xon	1	GFF	D
tana xon	-	-	line araxon	1	GFF	E
tana xon	-	-	line araxon	1	MLP	F
line araxon	-	line araxon	tana xon	2	GFF	G
line araxon	line araxon	line araxon	tana xon	3	GFF	H

Figure 1 and 2 show graphs of SAR measurement and predicted by artificial neural network, with the correlation diagram between the values in Kohak dam station. In these figures, the numbers change during the training phase is shown as point to point. Also compare the actual numbers obtained on the basis of the correlation can be seen in the test phase. Samples gathered on the regression line represents the correlation of this test is appropriate.

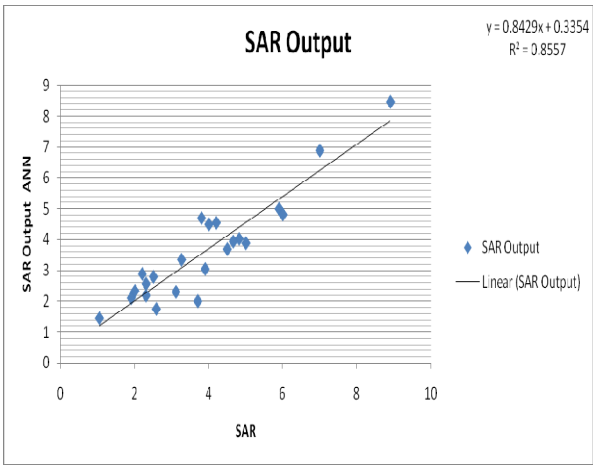

Figure 1 - Comparison of results with the results of measuring SAR and SAR was simulated by artificial neural network.

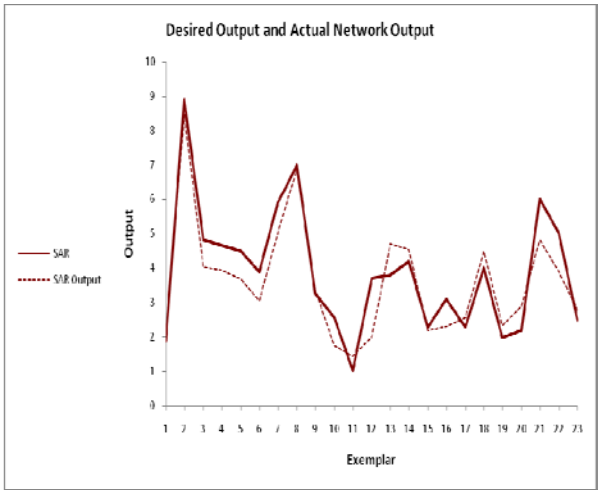

Figure 2 - Results of the neural network compared with the values measured by SAR.

After the implementation of artificial neural networks GFF or MLP with different transfer functions and the functions behind different indicators were R ² and the MSE and MRE were determined, and the training and testing of network software MATLAB and NeuroSolutions done that the results presented in Table 2 is:

TABLE 2 - STATISTICAL PARAMETERS R ² AND MRE MSE AND ARTIFICIAL NEURAL NETWORKS

R^2	MSE	MAE	Network Index
0.92	1.05	0.85	A
0.801	1.23	0.93	B
0.79	2.004	0.97	C
0.83	1.204	0.89	D
0.8003	1.356	0.95	E
0.86	1.08	0.75	F
0.81	1.13	0.91	G
0.793	1.37	0.93	H

According to the above table, the results of performance network with different transfer functions and the number of hidden layers, indicates that MLP network with one hidden layer and with the stimulus tanaxon R^2 =0.92 and MSE = 1.O5 (Network A) has more coefficients solidarity. The correlation coefficient, determines the ability of the network for communication between the input and output corresponding to the network is introduced. Therefore, to determine the sensitivity and effectiveness of each of the input factors can MLP network with a hidden layer and the driving function tanaxon (Network A) can be used. Furthermore, the degree of sensitivity, i.e. the influence of each input factor on the results presented to the network, network output, is expressed in the following table:

TABLE 3 - SENSITIVITY ANALYSIS OF PARAMETERS CL, PH, TDS AND EC ON THE SAR RESULTS

SAR (%)	*Sensitivity*
5.784180829	Cl
0.153107444	PH
28.96265452	TDS
65.10005721	EC

Figure 3 show the influence of various factors on the SAR.

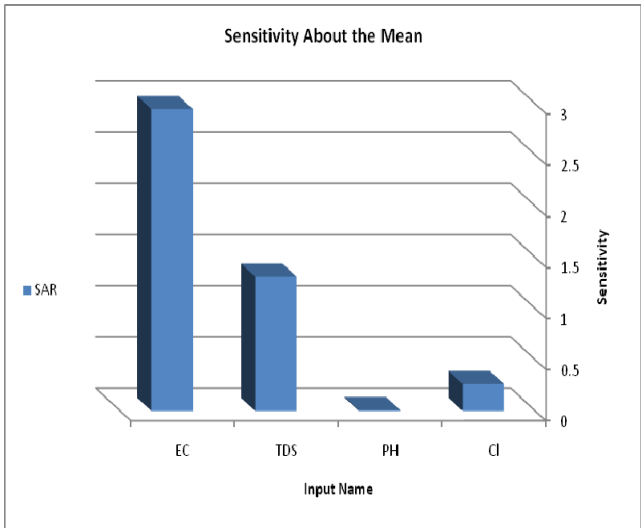

Figure 3- The influence of various factors on the SAR

As the figure 3 shows: EC> TDS> CL> PH. The most influential factor is salinity (EC) and TDS, CL and PH, respectively. The weight of each of the input data, the significant point is that, all these parameters in predicting the output of the network (SAR) are quite effective, and most of the entrances, in the run-owned the belong to EC and TDS and the least affected belongs to PH. Figure 4 represents the influence of input parameters, CL, PH, TDS and EC, the results indicate predictable. The purpose of influencing, in fact, is the positive or negative effects on output values .

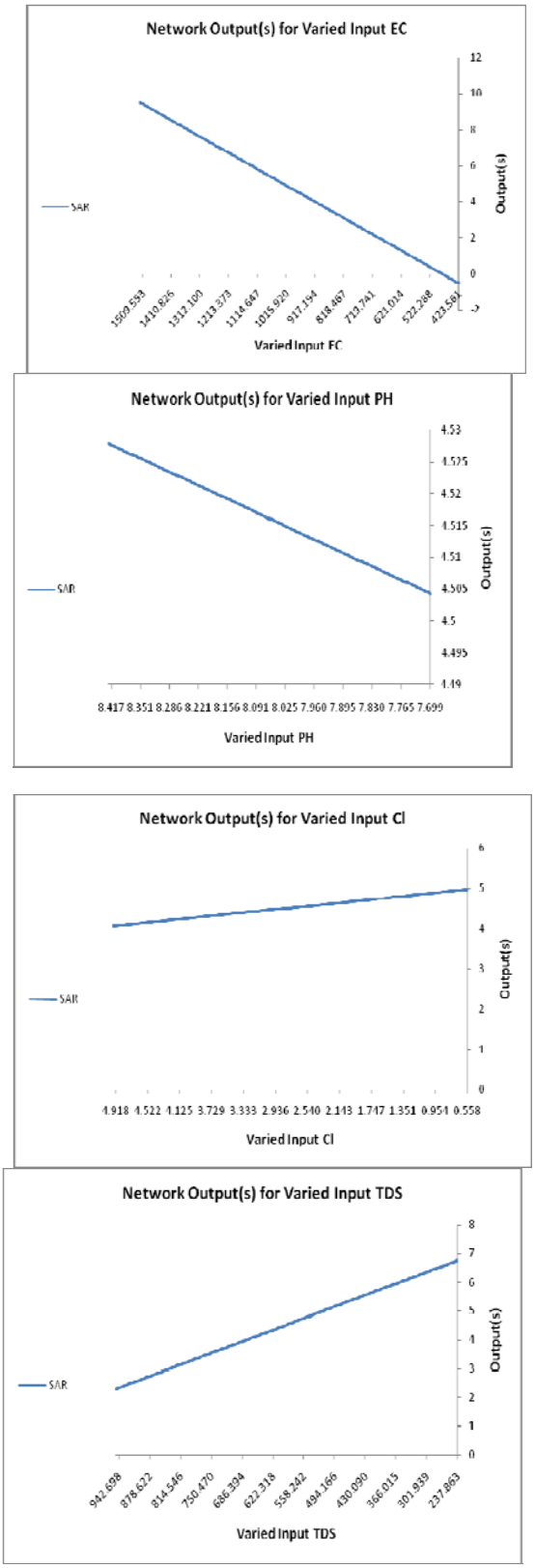

Figure 4– Analysis of the effects of input parameters (EC, PH, Cl, TDS) on SAR

Multivariate regression:

Many parameters on SAR levels are effective.

In this study, the parameters of chlorine (CL), PH, electrical conductivity (EC) and total dissolved substances (TDS) was investigated on the SAR and the regression equation was determined.

Figure 5, shows SAR obtained from the multivariate regression. Correlation coefficient and standard error are 0.74 and 1.35 respectively. Measured and predicted values for regression models, fit relatively well, but at a high SAR value,the amount of differences of the measured values and the model is increased.

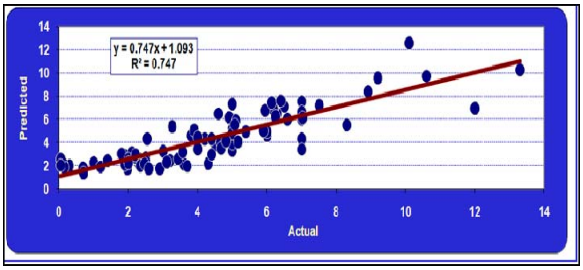

Figure5. Curve fit between predicted and actual valuse

Regression equation fitness the following correlation is equal to 0.747.

$$SAR=(0.46\times Cl)+(0.09\times PH)-(0.01\times TDS)+(0.01\times EC)-(1\times(\pm 1.35)) \quad (7)$$

TABLE4. STATISTICAL INDICATORS R^2, MSE, MRE IN 3RD ORD POLYNOMIAL REGRESSION MODEL

R^2	MSE	MAE	Type of Model
0.747	1.352	0.91	Regression Model

Finally, using comparison between statistical indicators in the artificial neural network and regression model in can be observed superiority of the artificial neural network model in simulated and predicted values of SAR.

REFERENCES

[1]. ASCE Task Committee on Application of Artificial neural networks in Hydrology., (2000), "Artificial neural networks in hydrology (I & II)", Journal of Hydrologic Engineering, 5(2), pp. 115-136.

[2]. Hore A, Dutta S, Datta S and Bhattacharjee C, 2008. Application of an artificial neural network in wastewater quality monitoring: prediction of water quality index. International. Journal of Nuclear Desalination (IJND) 3: 160 - 74.

[3]. Huiqun M and Ling L, 2008. Water quality assessment using artificial neural network. International Conference on Computer Science and Software Engineering. Washington, DC, USA.

[4]. Khazirpoor M, Hamzeh S, Jael A, 2011. Simulated EC values the Karun River using artificial neural networks, 3th National Conference on management of Irrigation and Drainage.

[5]. Kunwar P, Singh AB, Amrita M and Gunja J, 2009. Artificial neural network modeling of the river water quality—A case study. Ecol Model 220:888–95.

[6]. Kuo Y, Liu C and Lin KH, 2004. Evaluation of the ability of an artificial neural network model to assess the variation of groundwater quality in an area of blackfoot disease in Taiwan. Water Res 38: 148–58.

[7]. Lance E. Besaw, Donna M. Rizzo, Paul R. Bierman, William R. Hackett., (2010), "Advances in ungauged streamflow prediction using artificial neural networks", Journal of Hydrology 386, 27–37.

[8]. Mahdizadeh M, 2004. Artificial Neural Networks and its application in civil engineering. Ebadi Publishing, 130 pp.

[9]. Menhaj MB. 2000. Fundamentals of neural networks. Amir kabir University of Technology, Tehran.

[10]. Michael AM, Kherpar SD, Sondhi SD, 2008. Water wells and pumps. McGraw-Hill, New Delhi

[11]. Oliai A, Banejad H, Samadi MH, Rahmani A, Saghi MH, 2011. Performance evaluation of artificial neural network in predicting BOD and DO water quality criteria Darreh Hamadan River. Journal of Science water and soil / Vol 20.1(3).

[12]. Sarabi H, Multiple regression in behavioral research, The center of academic publishing, first published in 1987, Volume I, 441 pp.