Development of Water Rating Curve in Shatt Al-Arab River

Zuhal A. Al-Hadi
University of Basra / College of Engineering / Civil Engineering
Noor Ali @yahoo.com

Abstract

This study investigates the abilities of artificial neural networks (ANN) to improve the accuracy of stream flow-water rating curve in Shatt Al-Arab River. Development of stage-discharge relationships for the daily stream flow to Shatt Al-Arab is a challenging task.

In this study, the hydrological data was used as a tool for the identification of critical (information) segments in a river, to covers all study area in Shatt Al-Arab over a period of seven years started from January /2009 to January/2015from water resources office in Basra Province.

Data from different gauging sites were used to compare the performance of ANN trained on the whole data set. The neural network toolbox available in MATLAB was used to develop several ANN models. Five layers feed- forward network with Log-sigmoid transfer function was used. The networks were trained using Levenberg-Marquradt (LM) back-propagation

The Levenberg-Marquradt (LM) back-propagation was found to be the best ANN model with minimum Mean Squared Error (MSE) and maximum correlation coefficient (R) 0.9, and MSE $1.05*10^{-7}$, respectively. The optimum neuron number in the two hidden layers of (LM) was 8 neurons with R greater than 0.9, and MSE $1.05*10^{-7}$, respectively.

Key ward: rating curve, ANN,water.

الملخص

تهدف هذه الدراسة الى تحقيق قابلية الشبكات العصبية الاصطناعية في تطوير دقة جدولة جريان المياه في منطقة الدراسة(شط العرب). ان تطوير العلاقة بين المناسيب–التصاريف لنهر شط العرب اليومية تعتبر مهمة جديرة بالتحدي. في هذه الدراسة المعلومات المناخية المستعملة كانت اداة للتعرف على المقاطع الحرجة لتغطي عدة مقاطع من نهر شط العرب ولسبع سنين اعتبارا من كانون الثاني\ 2009 ولغاية كانون الثاني\2015 والتي تم الحصول ليها من دائرة الموارد المائية في محافظة البصرة.

تم استخدام الأدوات المتاحة للشبكة العصبية في برنامج MATLAB عدة نماذج للشبكات العصبية الاصطناعية. بنيت هذه النماذج باستخدام خمس طبقات تغذية أمامية مع دالة تحويل من نوع الاس السيني. تم تدريب الشبكات باستخدام خوارزمية التقدم العكسي من نوع Levenberg–Marquradt . وبناء" على نتائج هذه الدراسة تم إيجاد الأتي:

خوارزمية التقدم العكسي Levenberg–Marquradt (LM) أفضل تركيب نموذج شبكة عصبية اصطناعية تم الحصول علية لغرض التنبوء في منطقة الدراسة حيث أوضحت قيم معاملات المعايرة وهي معامل الارتباط ومعدل مربع الخطأ (0.9 و 1.05×10^{-7}) على التوالي. كانت الشبكة تحتوي على طبقتين مخفيتين تحوي كل منهما ثمان عقد.

الكلمات المفتاحية: تصنيف منحنى،IN ، الماء.

Introduction

Water is one of our most precious resources important to all ecological and socioeconomic activities, including food and energy production, transportation, waste disposal, industrial development, and human health. In water resource planning, design and infrastructure development reliable quantitative estimation of discharge is crucial. The Shatt Al-Arab River is an important water resource for Basra city, Iraq's second largest city, with many household, agricultural and industrial activities reliant on its waters.

Although it is possible to identify sophisticated models taking into consideration the hydrological and hydro meteorological variables, it is economically preferable to have a model that can simulate the water variations on the basis of flow discharge and depth records.

In Iraqi rivers is generally classified as wet, average or dry.

Accordingly, water management plans and the seasonal operation of Iraq's hydraulic infrastructure, including storing or releasing, are based mainly on these three scenarios.

Discharge can be measured reasonably well, however due to the spatial and temporal variability of rainfall; the forcing function that causes the discharges is not easy to characterize [Wagener *et. al*.,2008], thus this has necessitated researchers to use different methodologies in modeling stream flow.The variation in the hydrological cycle manifested in the magnitude and sequence of floods or droughts is very critical.

Shatt Al-Arab is the only water source of Basra city for different purposes like drinking and irrigation forms the outlet of the two main rivers of Iraq, the Tigris and Euphrates. Shatt Al-Arab flows along a wide channel in a south-easterly direction and downstream of Al-Fao discharges into the Arabian Gulf [Saad,2007].Nowadays, the flow of Shatt Al-Arab is much smaller than before because of massive dam projects in Turkey [Abdullah *et. al*.,2001].

Previous studies have been supportive of the latter especially artificial neural networks (ANN) for flow prediction. ANNs are effective in pattern recognition and function approximation [Lingireddy *et. al*.,2005] which are the main characteristics of water resources problems. The advantage of ANN is that no prior knowledge of the catchment characteristics is required, because even if the exact relationship between the input and output is unknown but is acknowledged the network can be trained to learn the relationship [Minns and Hall,1996]. ANNs can be taken as black box models since they neither learn based on assumptions relating to neither the input-output transfer function nor the physical interaction of the parameters. ANNs have found increasing applications in water resources and environmental systems for instance in rainfall runoff modeling [Dawson]. Stage-discharge (rating curve) modeling [Hertz *et. al*.,1991]. Most of the conventional pre-processing techniques, such as transformation and/or normalization of data, do not perform well, because of the large variation in magnitude and scale, as well as the presence of many zero values in data series [Jothiprakash and Kote,2011]. Data from the real world are never perfect; it can be an incomplete record with missing information, occurrence of zeros, improper types, erroneous records,etc.; hence, data pre-processing can be an iterative and tedious task.

The main purpose of this study is to develpe and discusses the performances of ANN technique in prediction of water rating curve. The ANN is applied to stream flow data of different stations on Shatt Al-Arab and it was a good tool comparing with previous studies.

Climate Properties

The climate of the area is categorized as a dried desert climate with winter rains and normally coded as (Bwhs), according to [Koeppe and De Long, 1961]. It is a continental climate with an extreme temperature and a long time period of hot season (summer), which extends to 230 days while the winter days are 75.No distinguishing of Spring and Autumn happened. The natural elements (temp., wind, rain …etc.) have direct effects on the explanations of environment phenomena. Meteorological data were collected from four stations (center of Shatt Al-Arab, Gurna- Madaina, and Sehan)[دائرة الموارد المائية] near the study area.

The information contained in any hydrological data is not homogeneous [Wagener *et. al*.,003]. The data that contain lots of hydrological variability may be the best choice for training, because they contain most of the information for parameter (weights) identification [Gupta and Sorooshian,1985]. In this study, the data with the most hydrological variability in a data series is termed as input data.

Artificial Neural Networks

A neural network (NN) consists of a large number of simple processing elements called neurons (units, cells, or nodes) which are considered as information-processing units. Each neuron is connected to other neurons by means of directed communication links, each with an associated weight. weights represent information being used by the network to solve a problem.It has an internal state, called its activation or activity level, which is a function of the received inputs. Typically, a neuron sends its activation as a signal to several other neurons. It is important to note that a neuron can send only one signal at a time, although that signal is broadcast to several other neurons. As in nature, the network function is determined largely by the connections between elements.A neural network can be trained to perform a particular function by adjusting the values of the connections (weights) between elements [Flood and Kartam,1994].

Neurons are divided to input layer, hidden layer and output layer. Figure (1) shows a schematization of artificial neural networks. The activity of the input units represents the raw information that is fed into the network. In the input layer there is one neuron for each model input. There are one or more hidden layers in NN. The hidden layers perform a weighted sum of inputs from each neuron of the previous layer, transform the sum according to some activation function and distribute the result to each neuron of the next layer. Subsequently, the output layer produces the final output. Each neuron takes many input signals and based on an internal weighting system, produces a single output signal, which is typically sent as input to another neuron.

In this paper, Feed-forward back propagation Neural Network (FNN). was used due to simplicity compared to other networks and its ability to learn the implicit governing relation between the inputs and outputs if sufficient training data is supplied. The FNN typically consist of two or three layers including hidden layer and output layer. Herein, the input layer processes no signal and this layer is not considered to be a neuron layer [Pastravan and De Keyser,2002].

Optimal number of neurons has to be determined in order to obtain a good identified network.The selected ANN inputs include set of elevation, rainfall, temperature, humidity and wind variables at sampling time (k-1), (k-2),(k-3)(k-4), and (k-5).While, the outputs (targets) has been selected to be the quantity of the flowrate (k).

Study Area

The length of Shatt Al-Arab river is about 200 Km, its width about 2 Km. The depth average is ranging between 8m to 15m. The mean hydraulic slope is about 1 to 1.5 cm/km [Hsian *et. al*.,1991].

Increasing and decreasing discharges of the river between periods was noted frequently which is consider a matter of course. This is may be due to environmental changes upstream the river which is made by human activity.It is clear that the decline changes in freshwater input has made. The discharges increases in the month April, May, and June, and decreasing in months, September, October, and November in Shatt Al-Arab.

This research covers the study and analysis of daily water data from four sampling stations of Shatt Al-Arab River in Basra region in Iraq, these stations are located at center of Shatt Al-Arab, Gurna, Madaina, and Sehan. Data for Shatt Al-Arab River in Basra region in these stations were collected form the period 2009 to 2015.

Data Used in the Study

The records of the stage can be transferred into records of discharge using a rating curve. Normally, a rating curve has the form: (دائرة الموارد المائية)

$$Q = a \times H^{b} \quad \text{... (1)}$$

Where, Q is discharge (m^3/s), H is river stage (m) and a, b are constant. The establishment of a rating curve is a non-linear problem. The ANN can represent the stage and discharge relation better than the conventional way, which uses Equation (1) [Jain and Chaligaonkar,2000].The inputs to the ANN were river stages at the current and previous times. The other inputs were water discharge at previous times. The input to the ANN model was standardized before applying ANN. The input was normalized by dividing the value with the maximum value to fall in the range [0,1].

Since the aim of this study is to test the feasibility of training the ANN model, we used the same setup of network as given by [16]. Hence, an integrated three-layer ANN, as described [Jain and Chaligaonkar,2000],was trained using the training period data pertaining to river stage, discharge. The number of nodes in the hidden layer was determined based on the best correlation coefficient (R) and the least root mean square error (RMSE) or mean square error (MSE). Using the weights obtained in the training phase for each case, the performance of the ANN was checked by using the testing period data. After examining the data and noting the periods in which there were gaps in the data variables. Programs were developed in MATLAB 10 software using the neural network toolbox to pre-process the data, train the ANN and test it. The weights were obtained by the Levenberg–Marquardt algorithm, which is computationally efficient

Results and Discussion

Stage, discharge rating relation was determined for different sites using the ANN. The site river stage and discharge relationship was fitted using Equation (1); Matlab 7.1 is used to develop ANN technology. The characteristics of applied ANN for predicting stage, discharge rating in Shatt Al-Arab River is, the number of hidden neurons is estimated according to trial and error procedure based on Mean Square Error (MSE) and Correlation Coefficient (R). MSE and R are calculated [Hagan *et. al.*,1996].

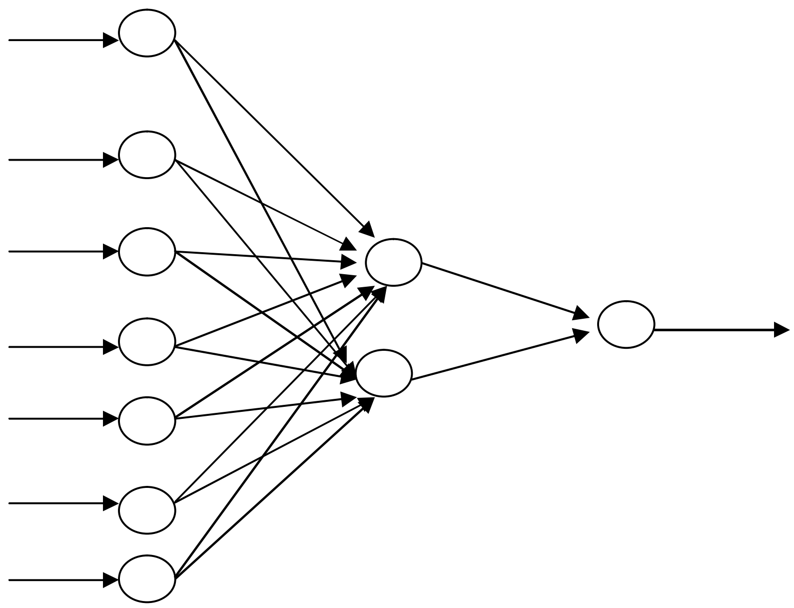
$$MSE = \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n} \quad \text{... (2)}$$

$$R = \sqrt{1 - \frac{\sum(x_i - y_i)^2}{\sum x_i^2 - \frac{\sum y_i^2}{n}}} \quad \text{...(3)}$$

Where x_i is actual data and y_i is calculated data by network, and n is the number of data. Zero is the best condition for MSE and one is the most desirable condition for R.

To satisfy the goal of training NN,the optimum number of neurons was determined by being based on the minimum value of MSE of training and the prediction sets. These was done by plotting the learning curve and observing the MSE, and try to change number of hidden layers, number of neurons in the hidden layers, type of activation function and type of Back propagation algorithm, until to arrive to the lowest possible values of MSE.

A set of 962 sampling points were used for training and test NN in order to carry out the learning procedure good, where 80 % of the learning samples were used for training and 20 % used for test.



Input layer Hidden layer Output layer

Fig. 1 Schematic representation of a typical ANN

Selection of Hidden Layers Numbers and Number of Neurons in Each Hidden Layer

The number of hidden layers and the number of neurons in one hidden layer are not straightforward to ascertain, there is no theoretical limit on the number of hidden layers but typically there are just one or two. A dilemma arises when determining the number of hidden-layer neurons, since there are no rules available to determine the exact number of neurons in the hidden layer the appropriate number of hidden layer (s) and number of neurons in one hidden layer is to be selected by trial and error. In this study, two types of multilayer models were developed, firstly single layer NN model consist of only one hidden layer. Secondly Multilayer NN models consist of two hidden layers. The optimal topology was determined by using one hidden layer with activation function as hyperbolic tangent (tansig) function in hidden layer and linear (purelin) function in output layer,NN was trained with the fast Backpropagation algorithm (Quasi-Newton Algorithm (BFGS)),and by taken different numbers of neurons from (1 to 20) as shown in Fig. (2). From this figure the network with 12 neurons in the hidden layer give best performance for both training and testing equal to (MSE= 1.2×10^{-4}).

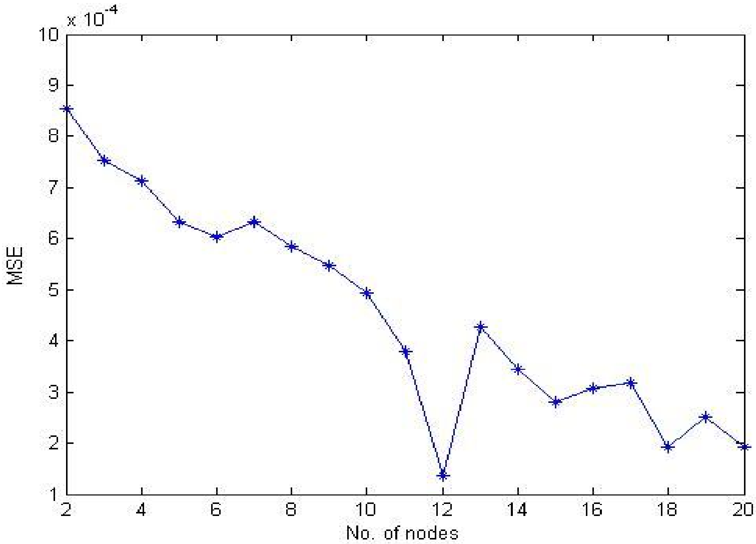


Fig.(2) Effect of the neuron number in one hidden layer on MSE

Then, two hidden layer was used with activation function as hyperbolic tangent (tansig) function in both, and linear (purelin) function in the output layer was used, NN was trained with the fast Backpropagation algorithm (Quasi-Newton Algorithm (BFGS)). Different numbers of neurons in each hidden layer from (1 to 20) neurons were used. The performance of these topologies of network for both training and testing is shown in Fig.(3). It can be shown that the error (MSE) for both training and test phase decreases until the number of neurons reach (8,8), Beyond these values, it observe a stagnation of the errors indicating that a greater number of neuron is useless and does not lead to an improvement of the results.

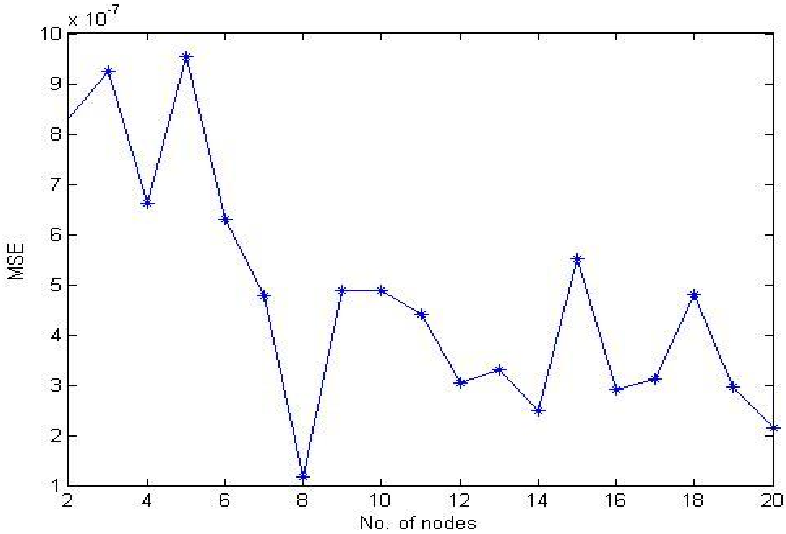


Fig.(3) Effect of the neuron number in two hidden layer on MSE

From above figs. it can be concluded that:-

Two hidden layers were chosen to satisfy the convergence because it's better than one hidden layer, were with optimal number of neurons in the hidden layer, MSE value was 1.2×10^{-4} for one hidden layer NN and 1.05×10^{-7} for two hidden layer NN.

Training of any neural network is considered to be successful if the trained network works well on the testing data set. The analysis of results and the discussion presented above clearly show that the ANN trained on critical events has performed equally well for the tested data set.

Generalization of the Neural Network

The performance of a trained network can be measured to some extent by the error on the training, and test sets, but it is often useful to investigate the network response in more detail. One option is to perform a regression analysis between the network response and the corresponding targets. The routine 'postreg' in **MATLAB** program is designed to perform this analysis.

Fig. (4) shows the regression analysis between the output of the NN and the corresponding targets for the best chosen topology. The format of this routine is [m,b, r], where m and b correspond to the slope and the intercept of the best linear regression that relates the targets to the network outputs. If the fit is perfect (outputs exactly equal to targets), the slope would be 1, and the intercept with the y-axis would be 0. The third variable, r, is the correlation coefficient between the outputs and targets. It is a measure of how well the variation in the output is explained by the targets. If this number is equal to 1, then there is perfect correlation between targets and outputs.

Fig.(4) shows comparison results between NN results and target results for training data testing data for center of Shatt Al-Arab,the outputs are plotted versus the targets as open circles. A dashed line indicates the best linear fit and the solid line indicates the perfect fit (output equal to targets). The values of the slope are 0.88 for each, respectively. The interceptions with the y-axis are 0.029, and .039, respectively. The correlation coefficients are 0.95196, and 0.9021, respectively.

The mapping of the NN for the training patterns was very good; also the generalization of the NN for the test patterns was very good.

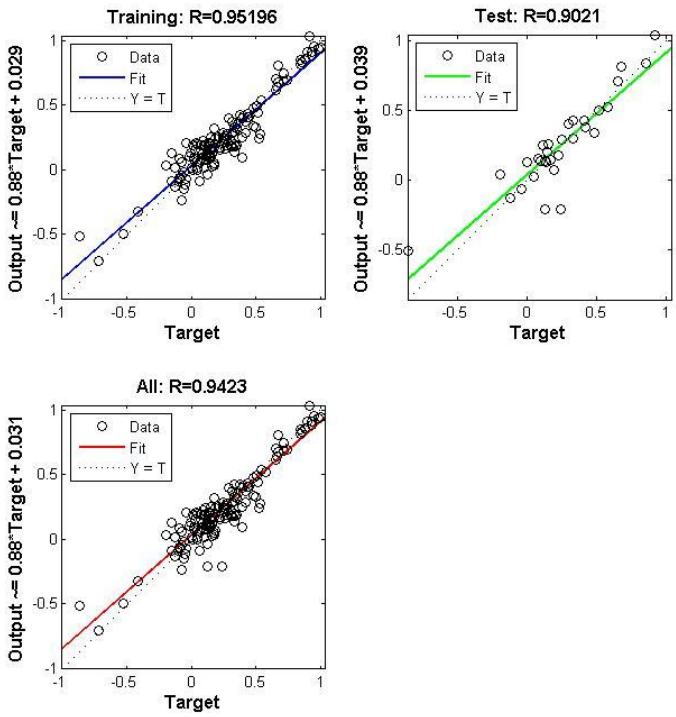


Fig.(4) R-values for NN Results for the center of Shatt Al-Arab

In Figs (5), (6), and (7) show comparison results between NN results and target results for training data testing data for other stations take in this study

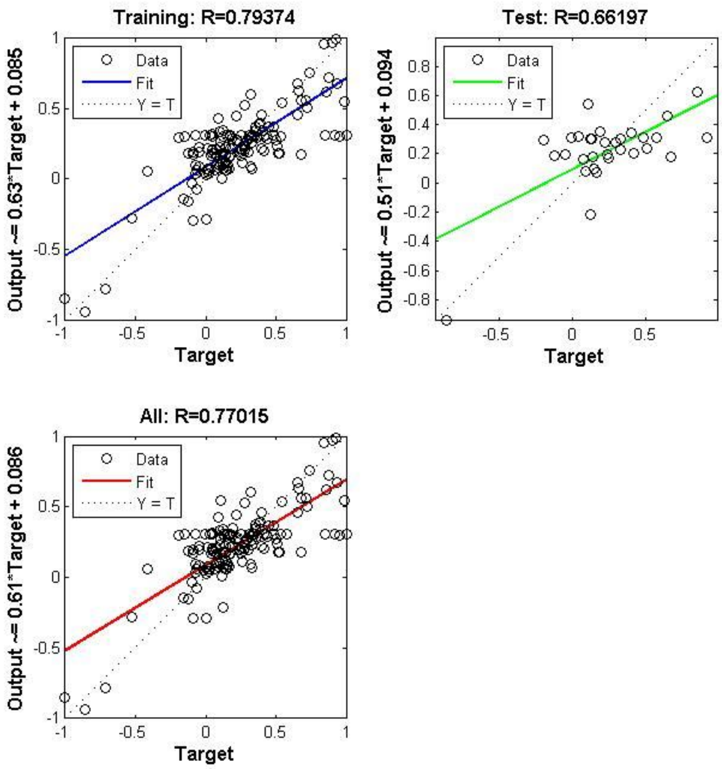


Fig.(5) R-values for NN Results for Gurna of Shatt Al-Arab

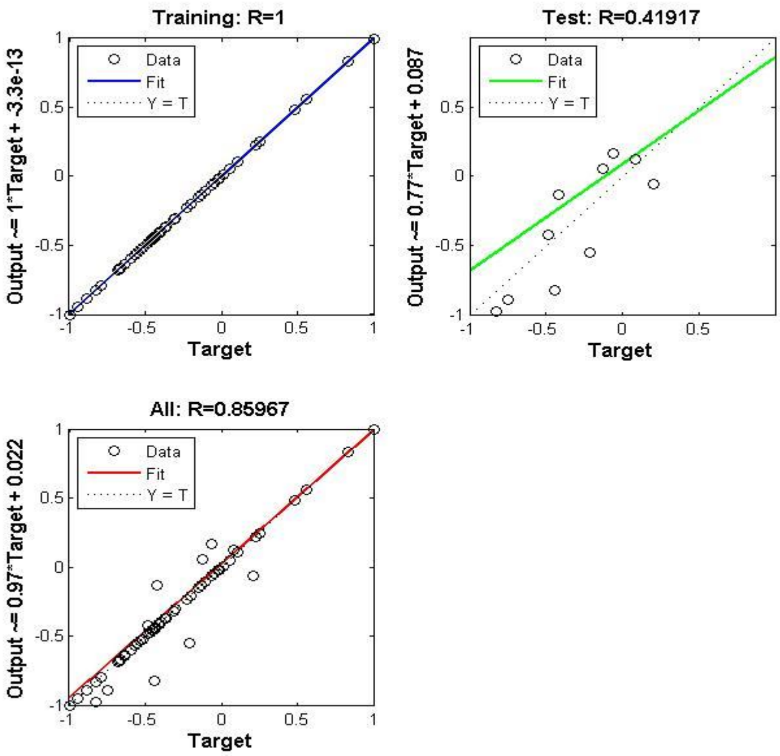


Fig.(6) R-values for NN Results for the Madaina of Shatt Al-Arab

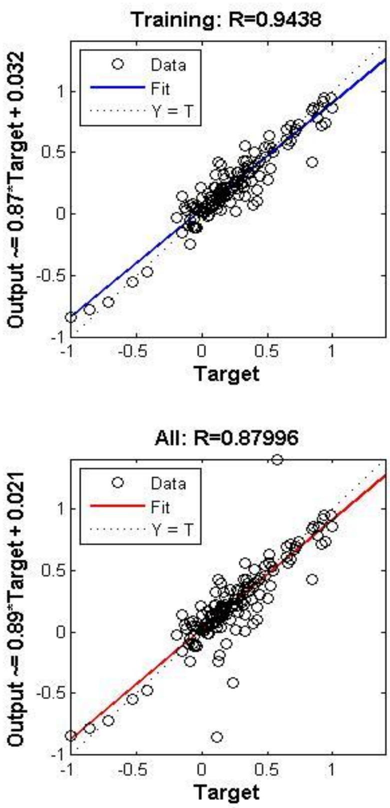


Fig.(7) R-values for NN Results for Sehan of Shatt Al-Arab

Figs.(8) to (10) present the best training performance for the relations between MSE and epoch (predicted water levels in the study area) using the best structure (two hidden layers, 8,8 nodes) of ANN.

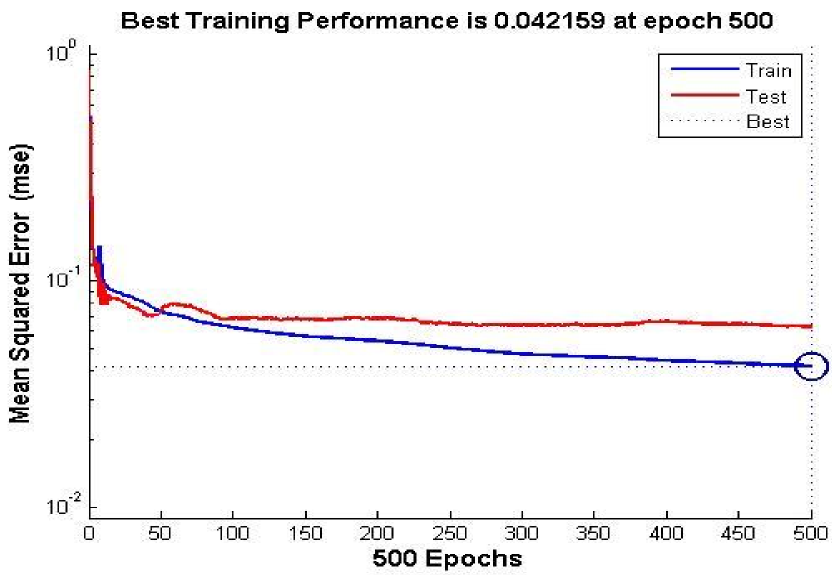


Fig (8) MSE versus Epochs for ANN with two hidden layer for Gurna

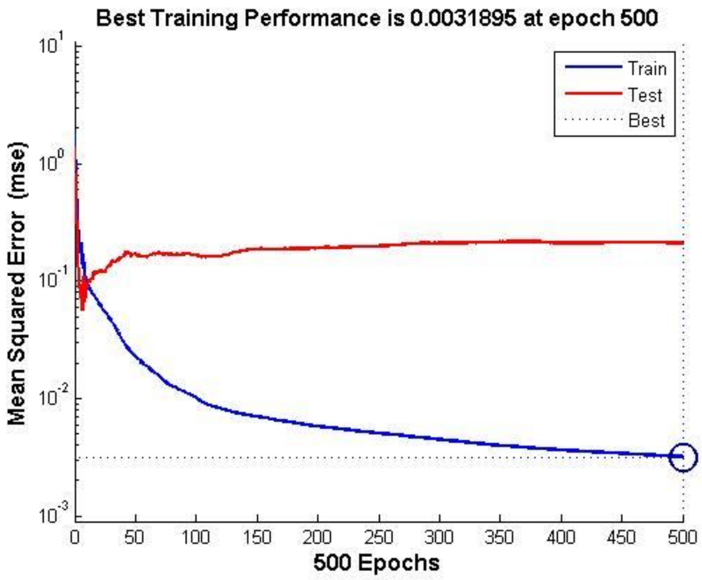


Fig.(9) MSE versus Epochs for ANN with two hidden layer for Madaina

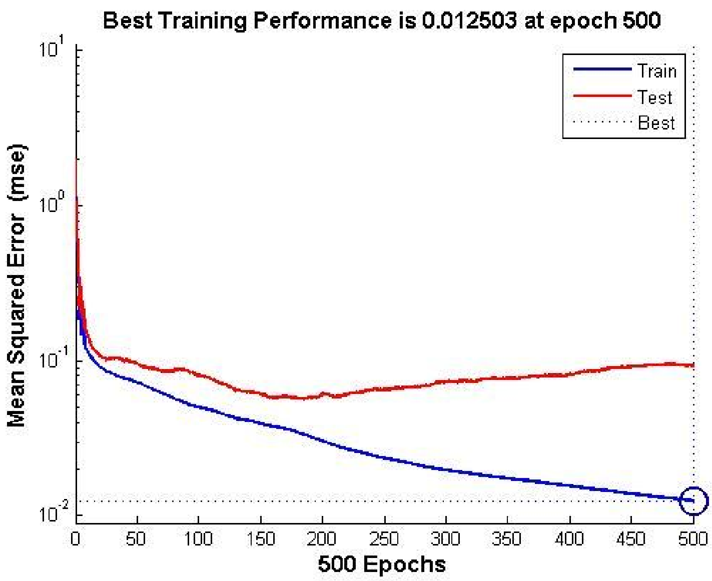


Fig. (10) MSE versus Epochs for ANN with two hidden layer for Sehan

Conclusions

The main aim of this study is to show the ability of an artificial neural network in predicting stream flow. The application of this technique to the Shatt Al-Arab River at different stations in Basra has shown the possibility of using available data in a given catchment to predict stream flow with the following points:

1. This strategy can result in substantial savings in time and effort in the training of models based on data-driven approaches, such as ANNs. There is always great effort and expertise needed to select the proper data set for the training.
2. The methodology can be used for other kinds of ANNs, such as radial basis function neural networks or support vector machines.
3. The concept of the paper can be used to predict the performance of the model a priori by looking at the geometry of the training and testing data. A possible criticism of the use of information-rich data could be that it may not result in

substantial savings of time, since some time and effort will be spent in the identification of critical events

4. This paper shows that using two hidden layers in the NN, rather than single hidden layer, significantly improve performance of network. Number of nodes (12),(8,8) in the first and second hidden layer showed to give good performance to NN by reducing MSE.
5. The ANN model was able to reproduce the complex nonlinear Q-H relationship and its multiple loops that are observed in Shatt Al-Arab.

References

Abdullah, AA, Issa., M, Saleh., AQJ, Qassim., A,2001,"**Physical and chemical properties of water for the northern part of the Shatt Al-Arab River**", *Basra Agriculture Sciences*, vol. 14, no. 3, pp. 123-42.

Dawson .C. W. and Wilby .R., 1998,"**An artificial neural network approach to rainfall runoff modelling,**" Hydrological Sciences, vol. 43, no. 1, pp. 47-66,.

Flood, I. and Kartam, N.,1994,"**Neural Networks in Civil Engineering I: Principles and Understanding**" Journal of Computing in Civil Engineering, Vol.8, No.2, pp.131-148,

Gupta, V.K.; Sorooshian,S.,1985,"**The relationship between data and the precision of parameter estimates of hydrologic models". J**. Hydrol. 81, 57–77.

Hagan, M. , Demuth, H., and Beale, M., 1996,"**Users Guide for the Neural Network Toolbox for MATLAB"**, Ver. 6, MA, The Mathworks Inc., 901 pp.

Hertz, J., A. Krogh, and R.G. Palmer, 1991,"**Introduction to the Theory of Neural Computation**", Redwood city, California, USA. Addison Wesley.

Husian, N.A., Al-Najar, H.H.; Al-Saad, H.T and Al-Sabonchy, A.A,1991," **Shatt Al-Arab Basic Scientific Studies**" University of Basra, Marines Science Center,.

Jain,S.K.;Chalisgaonkar,D.,2000,"**Setting up stage-discharge relations using ANN**". J. Hydrol. Eng. 5, 428–433,

Jothiprakash,V.;Kote,A.S.,2011,"**Improving the performance of data-driven techniques through data pre-processing for modelling daily reservoir inflow**". Hydrol. Sci. J., 56, 168–186..

Koeppe,E. C. and De Long,G.C.,1961,"**Weather and climate**",Mc Graw-Hill, NewYork,.

Lingireddy,S.,D.Ramalingam, and J.Pavoni,2005,"**ANNs as function approximation tools- a case study**," in Artificial Neural Networks in Water Supply Engineering, B. G. Lingireddy, Virginia: American Society of Civil Engineers, pp. 160-170.

Minns. A. W. and Hall .M.J.,1996,"**Artificial neural networks as rainfall runoff models**," Hydrological Sciences Journal, vol. 3, pp. 399-417, June.

Pastravanu, O., and De Keyser, R.,2002,"**Neural Network and Hybrid Modelling" ,** Diploma thesis, University of Gent , Department of Control Engineering and Automation.

Saad, M. A. H.,2007,"**Review of limnological and oceanographic data base in Iraq during 1973-1977"** Int. j. ocean. Oceanorg. 2(1):1-16.

Wagener, T.; McIntyre, N.; Lees, M.J.; Wheater, H.S.; Gupta, H.V., 2003,**"Towards reduced uncertainty in conceptual rainfall-runoff modeling":**Dynamic identifiability analysis. Hydrol. Process. 17, 455–476..

Wagener .T., Wheater .H. S., and H. V.Gupta, 2008,"**Rainfall-Runoff Modelling in Gauged and Ungauged Catchment"|**, London: Imperial College Press.

– دائرة الموارد المائية، محافظة البصرة، البيانات.