PREDICTION OF RIVER DISCHARGE USING ARTIFICIAL NEURAL NETWORKS: AN EXAMPLE OF GHARRAF RIVER, SOUTH OF IRAQ

Alaa M. Atiaa
Department of Geology, College of Science, University of Basra. Basra- Iraq

Abstract

The applicability and performance of the artificial neural networks are investigated by predicting river discharge one and two days ahead for Gharraf River, south of Iraq. Gharraf River system is located at southeast Iraq within Mesopotamian Plain. A multilayerd percpetron artificial neural net is selected to achieve experiments which trained by using back-propagation algorithm. Three models are presented firstly to explore the affect of the previous discharge on the specified discharge. The ANN generated results are evaluated using statistical parameters: squared correlation coefficient R^2 and root mean squared error RMSE. The results of this study indicate that ANNs are capable of producing very good results for both one and two days ahead predictions. Correlation between observed and simulated discharge values of both high and low is estimated with a good accuracy.

تخمين التصريف النهري باستخدام الشبكات العصبية الصناعية مثال عن نهر الغراف جنوب العراق

علاء محسن عطية
قسم علوم الارض، كلية العلوم، جامعة البصر. البصرة – العراق.

الخلاصة

تم التحري عن إمكانية تطبيق وأدائية النماذج العصبية الصناعية لغرض التنبؤ القصير المدى بتصريف نهر الغراف في الناصرية جنوب العراق. يمتد نهر الغراف من الكوت وينتهي في اهوار جنوب العراق في الناصرية وفي هور الحمار تحديداً. استخدمت الشبكة العصبية من النوع المتعدد الطبقات لبناء معمارية الشبكة واستخدمت طريقة انسياب الخطأ إلى الخلف لغرض تدريب الشبكة (معايرتها). اختبرت ثلاث نماذج في البداية لغرض معرفة تأثير التصاريف السابقة على معدل التصريف في يوم محدد لاحق، بعدها تم اختيار النموذج الذي يعطي اقل خطأً لغرض إجراء التنبؤ ليوم ويومين لاحقين. بينت نتائج تطبيق الطريقة الإمكانية الكبيرة للشبكات العصبية في إجراء التنبؤ وعزز ذلك من خلال مقدارين إحصائيين هما جذر معدل الخطأ التربيعي ومعامل الارتباط بين التصاريف المقاسة والمحسوبة باستخدام النموذج، كما كان لنموذج الخلية العصبية المصمم القدرة على محاكاة قيم التصاريف الواطئة والعالية على حد سواء.

Introduction

The problem of accurately determining river discharge occupies an important place in water resources management policies. Hydrological phenomena are highly non-linear, spatially and temporally varying, and not easily captured by simple models. Therefore, modeling of such complex systems becomes one of the difficult and challenging tasks for water resources practitioners. Many approaches have been used around the world to model hydrological systems which is in turn needs careful attention both to the reliability of the predictions and safety of the made decisions. Generally, these approaches (models) are grouped into two main categorizes: physically-based and data-driven models. Physically-based models use extensively knowledge of the details about the physical behaviors of the system. They are dependent on mathematical differential equations which describe the system being question and solved analytically or numerically. Data-driven models, on the contrary, do not use any explicit well-defined representation of the physical process and governing equations of the phenomena. They are highly dependent on the quality of the observed data used to train (calibrate) these models. There is another type of model which combines these techniques in one solution and called hyprid models.

Practitioners in water resources have embraced data-driven techniques enthusiastically during the last decades because they are perceived to overcome some of the difficult associated with physically based models.

The ultimate objective of the present study is to predict Gharraf River discharge at Nasiryiah, south of Iraq, one and two days ahead by using artificial neural network (one of the most popular form of data-driven models) and to introduce such models into the branch of water resources of Iraq as alternative to the physically-based models which are time consuming and costly.

Artificial Neural Network

The artificial neural network (ANN) technique is an artificial intelligence (AI) technique that attempt to mimic the human brain's problem solving capabilities. It is an information processing system that is composed of a number of processing elements (or neurons) and interconnection (or weights) between these elements that initiate the synaptic strength in a biological nervous system.

There are many types of ANN, but the most commonly used is the multi-layered percpetron (MPL), it has been applied successfully to solve some difficult and divers problems by training them in a supervised manner with a highly popular algorithm known as the Backpropagation algorithm [1]. It can approximate any arbitrary continuous function, simulate a non-linear system without a prior assumption of process involved, and give a good solution even when input data are incomplete or ambiguous [2] [3].

A (MLP) network consists of at least three layers: an input layer, an output layer, and one or more hidden layer(s) (middle). Each layer is made of neurons which are junctions receiving inputs from neurons in other layers. The input to an individual from neuron consists of signals from other neurons combined using algebraic transfer function and weighting factors associated with each connections. Figure (1) shows a typical MLP with three layers.

Training in ANNs (sometimes called learning) involves feeding samples as input vectors through designed network, calculating the error of the output layer, and then adjusted the weight of the network to minimize error. Training can stop when the network error drops below a specified threshold. Usually training is accomplished using a non-linear optimization algorithm. The Backpropagation algorithm is used in this study. Backpropagation algorithm consists of two pass through the different layers of the network: a forward pass and backward pass. In the forward pass an input vector is presented to the nodes of the network and its effect propagates through the network layer by layer [1]. Finally, a set of output is produced. In the forward pass the synaptic weigh remain fixed through the network. In the backward pass, the weights are all adjusted depending on an error correction rule. The output of the network is subtracted from the target values to produce an error term. This error is then propagated backward through the network. The weights are adjusted to make the calculated values by network move closer to the target values in a statistical term. The root mean square error (RSME) is used to investigate the performance of the network.

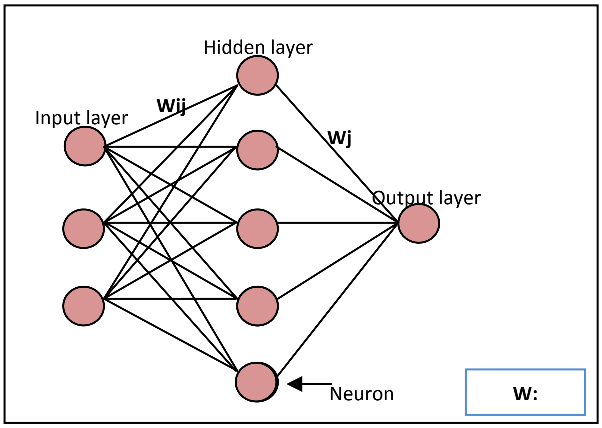

Figure 1: A typical multi-layered percpetron with three layers.

Study area

Gharraf River system is located at southeast of Iraq within Mesopotamian plain, figure.(2), between latitude (32°27′- 31°2′ N) and (46°43′- 45°45′ E). The study area covers an area of about (435052 × 106) m2. The Mesopotamian plain of southeast and central of Iraq is a flat surface bounded by the foothill zone in the east and the Salman zone in the west Figure.(3). Elevation of it varies from 2m above sea level near Basra to 35m near Baghdad and 180m near Baiji. The flood plain of Mesopotamia contains Quaternary fluvial sediments of the interacting Tigris and Euphrates rivers, merging into a marsh land north to Basra and Shatt Al-Arab delta plain between Basra and the Arabian Gulf [4].

Gharraf area was part of the delta of Euphrates River before 100-500 B.C. After that, Gharraf River moved over the original Euphrates deposits and covered these sediments with deposits origin from the Tigris (Iraqi Ministries of Environment, 2006). The river begins in the Kut barrage and runs south between Euphrates and Tigris Rivers, finish at Nasiryiah in Hammer marsh. The length of the river is of about 230 km with hydraulic gradient of about 0.0000475. The course of Shatt Al Gharraf can be subdivided as follows, according to the conditions that governed its development [5]: (1) Hai Delta, which ends at Kalaat Sukkar and in which expansion can take place. (2) Rafai gully extending to about 10 km upstream of Bada'a and in which flow is restricted, no lateral.

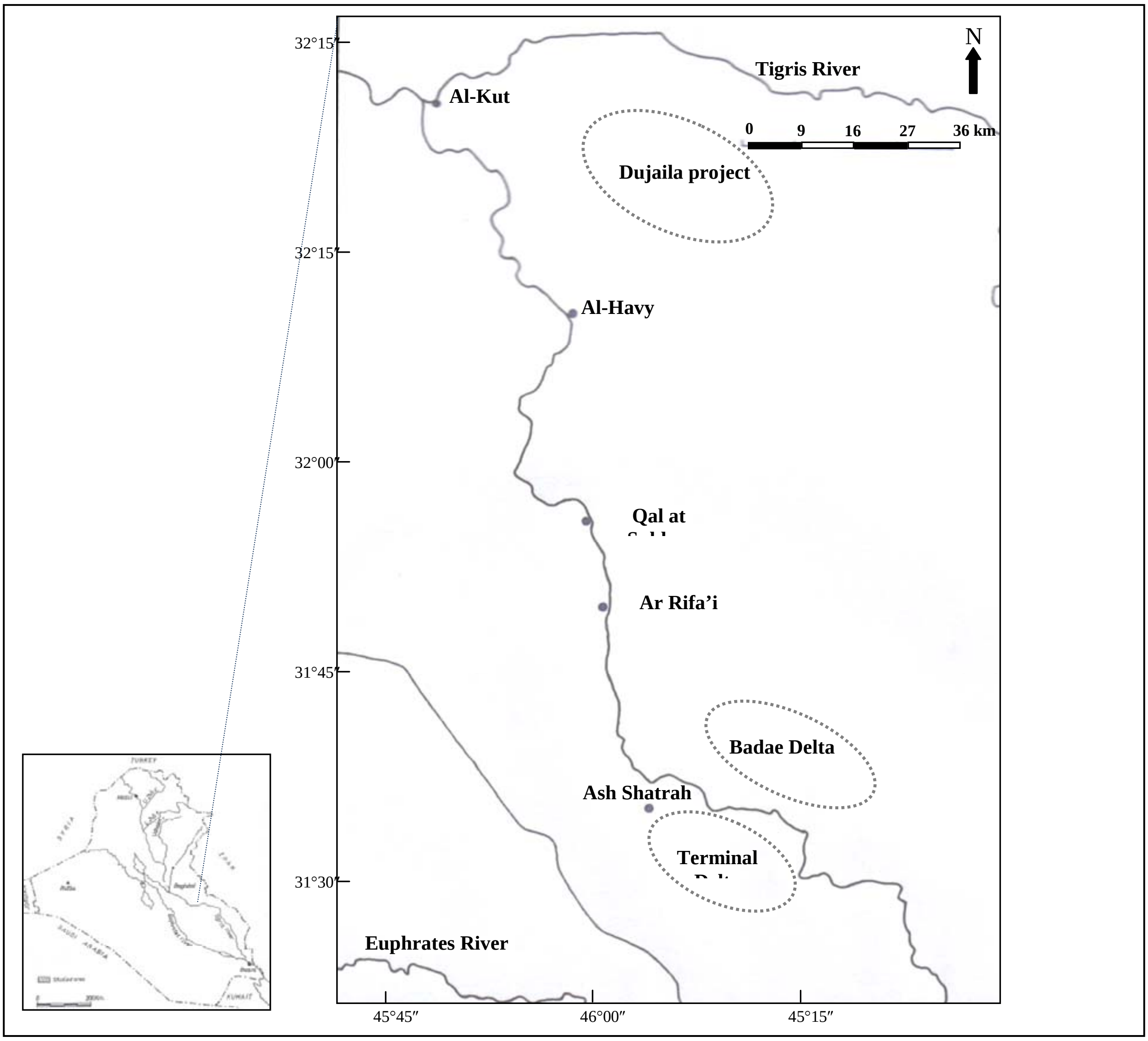

Figure 2: Gharraf River system (Modified after [5])

expansion being possible. (3) The Bada'a Delta that is the most recent region of expansion on the left bank towards the Hor Abu Ijul, Hor H'weynah and Hor Ghamukah depressions. (4) Shattrah and Kasser-Ibrahim Deltas that are regions of expansion at the end of Rafai gully.

Methodology

The main objective of this study is to predict the main daily discharge of Gharraf River regulator at Nassaryia. The available data covers the period from April 2005 to May 2006, hence 333 instances (observations) are available and ready to use. The neural network toolbox in Matlab 2006a is used. Firstly, the data set is scaled using mapminmax function, according this scaling the range of the input lies inside the range $\left(-1 \le x \le 1\right)$. The data set is divided into three parts: 60% for training, 20% for validation, and 20% for testing. The used ANN model is based on its most popular form, i.e., multilayer perceptron with one hidden layer. In training, Backpropagation algorithm is applied with the fast training algorithm known as Levenberg-Marguard technique. The logistic sigmoid transfer function is used in the hidden layer and a linear one in the input layer. The root mean squared error (RMSE) and squared correlation coefficient are used as the performance criteria of the net. The early stopping method is selected to overcome overfitting problem. The number of nodes in hidden layer is estimated using trial and error procedure. The other networks parameters are still as default values.

Results and discussion

The selection of appropriate model inputs is extremely important in any prediction model [6]. Because the available data is river flow only, three models are adapted firstly to explore the effects of previous recorded discharges on the prediction of a specified discharge. These models are described as below:

M1: Qt = f (Qt-1)

M2: Qt = f (Qt-1, Qt-2)

M3: Qt = f (Qt-1, Qt-2, Qt-3)

where Qt is the mean daily discharge at a specified time. Table 1 shows the results of this experiment.

Table 1: the effect of previous recorded discharge on ANNs results

Model	Training period		Testing period		No. of hidden nodes
	R^2	RMSE	R^2	RMSE	
M1	0.84	0.173	0.70	0.293	25
M2	0.85	0.123	0.80	0.187	30
M3	0.87	0.122	0.85	0.159	34

The other causal variables (e.g. rainfall, evaporation, etc.) are of important, but they are not available on the daily basis, only monthly averages are available. Therefore, they are excluded from the study and treat the problem as time series one instead of function approximation.

The performance level of the nets increases with the increase in the number of the included previous discharge in the input data set Table 1. The error statistics, R2 and RMSE confirm this acclaim. According to the results of this experiment the third model (M3) is selected to carry out the second experiment, i.e, prediction.

Table 2 presents the performance statistics of the second experiment and Figure 3 shows the observed versus calculated discharges. The results in Table 2 and Figure 3 ensure that ANNs is capable of prediction of high and low discharge values very well. High values of R2 and low values of RMSE support strongly this statement.

The length period of prediction could be increases if data about the variable under consideration are available and after introduce another causal variables such as rainfall, or uses hyprid models like wevelet-neuro technique [7].

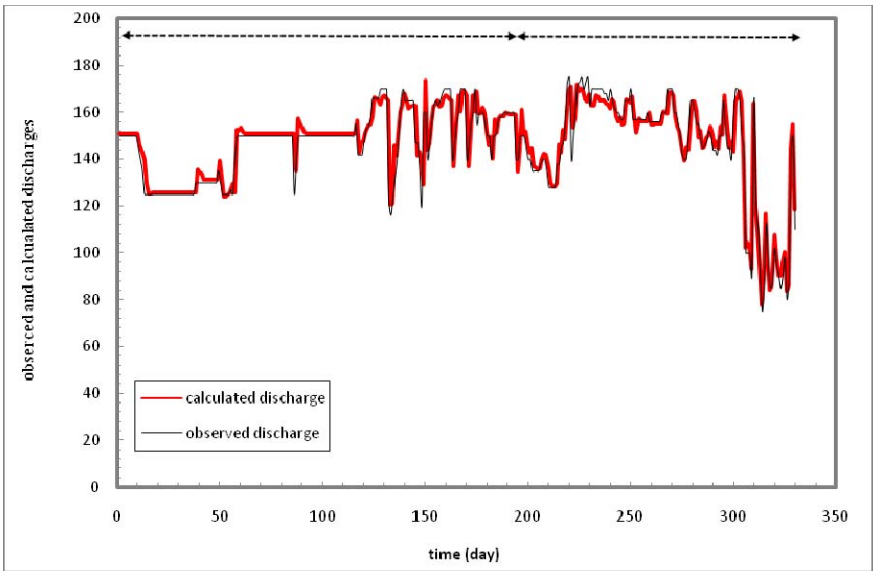

Figure 3: observed versus simulated discharges for one day ahead prediction

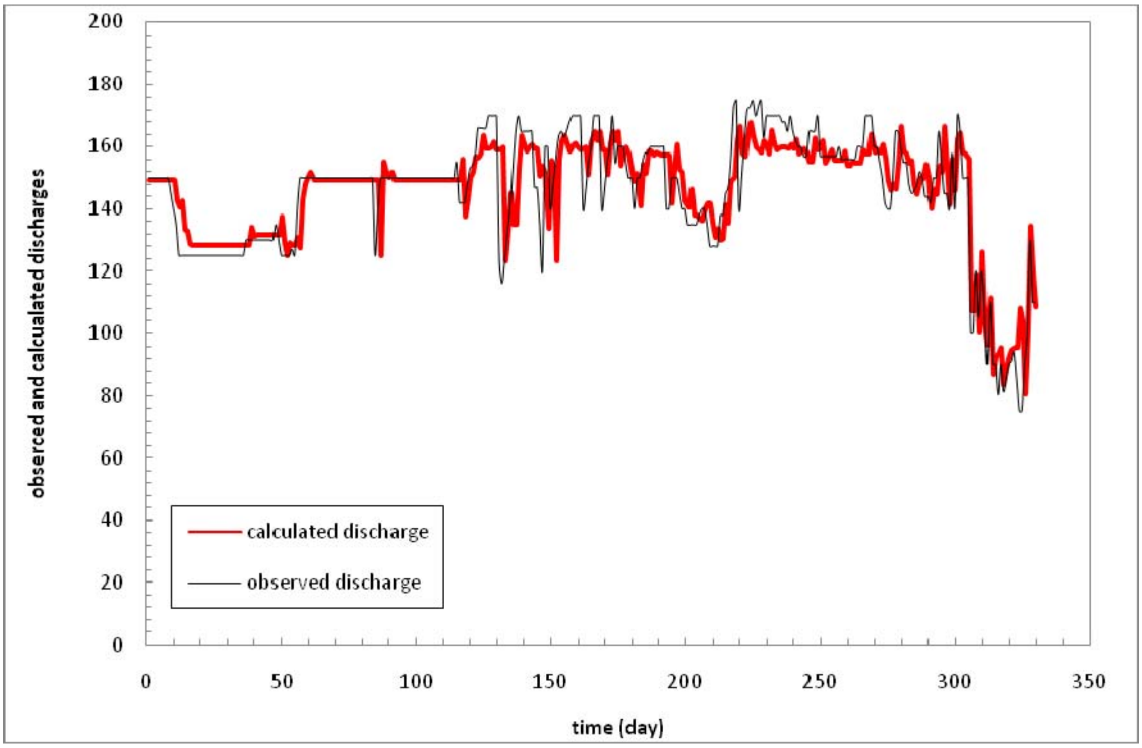

Figure 4: observed versus simulated discharges for two days ahead prediction

Table 2: results of the second experiment

Model	Training period		Testing period	
	R^2	RMSE	R^2	RMSE
Q_{t+1}	0.88	0.139	0.75	0.207
Q_{t+2}	0.83	0.195	0.70	0.262

Conclusions

The possibility of using artificial neural networks to predict Gharraf River discharges at Nasiryiah, south of Iraq is investigated. Two experiments are adopted to perform this task. The first experiment shows the importance of the effect of previous discharges on a specified discharge and on build a suitable ANNs model. The selected suitable model from the first experiment is used to predict river discharge one and two day ahead (second experiment). The error statistics of the results confirm that ANNs capable of prediction of high and low discharge values very well. No trial made to compare results of the study with other useful technical techniques such as ARIM model or physically-base model because researchers around the world study that extensively. The ANNs technique is suitable to predict discharge of natural rivers as well as controlled rivers if the suitable date is available. Physically based models are used to manage water resources of Iraq since eighteenth form the past century, but as remember previously they used a huge data to construct them. Therefore, we recommend to use data driven models in branch of water resources of Iraq due to the fact that they are very useful and easy to construct models, and only use data to perfume specific tasks without need for knowing the behavior of the system under consideration.

Acknowledgment

I would like to offer my special thanks to Dr. Solomatine, Department of Hydroinfo-rmatics and Knowledge, UNESCO-IHE Institute for Water Education, who opens my eyes on the field of data driven models throughout published and freely papers on your web site.

References

1. Haykin, S. **1994**, "*Neural Networks a Comprehensive Foundation*", 1st edition, Macmillan College Publishing Company Inc., New York.
2. ASCE Task Committee, **2000a** "*Artificial neural networks in hydrology I: preliminary concepts*". Journal of Hydrologic Engineering 5(2), 115-123.
3. ASCE Task Committee, **2000b**, "*Artificial neural networks in hydrology II: Hydrologic application*". Journal of Hydrologic Engineering 5(2), 124-137.
4. Saad Z. Jassim and Jeremy C. Goff ,**2006**. "*Geology of Iraq*", Dolin, Prague and Moravian Museum, Brno, 341 p.
5. Iraqi Ministries of Environment, Water resources, Municipalities and Public works, **2006**, "*Overview of present conditions and current use of the water in the water resources marshlands area*", book1, Italy-Iraq, 146p.
6. Solomatine, D. P. **2002**,"*Applications of data-driven modeling and machine learning in control of water resources*". Computational intelligence in control, M. Mohammadian, R. A. Sarkaer, and X. Yao eds., Idea Group Publishing, 197-217.
7. Wang, W. & Ding J., **2003**,"*Wavelet network model and its application to the prediction of hydrology*" Nature and Science, 1(1), 67-71.