Pertanika J. Sci. & Technol. **18 (1)**: 33 – 41 (2010)
ISSN: 0128-7680

Neural Networks for Forecasting Daily Reservoir Inflows

Shahram Karimi-Googhari[1], Huang Yuk Feng[2], Abdul Halim B Ghazali[3] and Lee Teang Shui[4*]

[1]Department of Water Engineering, Shahid-Bahonar University of Kerman, Kerman, I.R. Iran
[2]National Hydraulic Research Institute of Malaysia, Ministry of Natural Resources and Environment, 43300 Seri Kembangan, Selangor, Malaysia
[3]Department of Civil Engineering, Faculty of Engineering,
[4]Department of Agricultural and Biological Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[*]E-mail: tslee@eng.upm.edu.my

ABSTRACT

Proper integrated management of a dam reservoir requires that all components of the water resource system be known. One of these components is the daily reservoir inflow which is the subject matter of this study, i.e. to establish predictions of what is coming in the next rainfall-runoff process over a catchment. The transformation of rainfall into runoff is an extremely complex, dynamic, and more of a non-linear process. The available six-year average daily rainfall data across the Sembrong dam catchment were computed using the well-known Theissen's polygon method. Daily reservoir inflow data were extracted by applying the water balance model to the Sembrong dam reservoir. Modelling of relationship between rainfall and reservoir inflow data was done using feed-forward back-propagation neural networks. The final selected model has one hidden layer with 11 neurons in the hidden layer. The selected model was applied for an independent data series testing. Results in relation to specific climatic and hydrologic properties of a small tropical catchment suggested that the model is suitable to be used in forecasting the next day's reservoir inflow. The efficiencies of the model Abtained indicated the validity of using the neural network for modelling reservoir inflow series.

Keywords: Reservoir inflow, neural network, forecasting, modelling

INTRODUCTION

Dam reservoirs plays a vital function at various times and for different purposes such as supplying water for irrigation, hydropower, mitigating disastrous environmental effects and impacts, as well as ensuring flood mitigation and as an insurance during periods of drought etc. In many instances, these dams have no established ydrometric data collection network. The absence of intense network (for the accuracy of data required for the establishment of a vast network facilities) is usually the norm in view of the high costs involved in setting them up. These conditions can result in a considerable uncertainty in the hydrologic information obtained. The non-linear relationship between input and output variables complicates the effort to forecast reservoir inflow events. Many of the techniques currently used in modelling hydrological time-series consider linear relationships among the variables. The two main technique groups are physically based conceptual models and time-series models. In the first group, the procedure is to mathematically simulate the sub-processes and physical mechanisms that prevail in the hydrological cycle. These models usually

[*]Corresponding Author

combined simplified forms of physical laws and are generally non-linear, time-invariant, and deterministic, with parameters which are representative of watershed characteristics (Hsu *et al*., 1995) but ignore the spatially distributed, time-varying, and stochastic properties evident in the rainfall–runoff process. The implementation and calibration of conceptual models can cause many difficulties which require intricate mathematical tools (Duan *et al*., 1992; Sorooshian and Gupta, 1995), significant amounts of calibration data (Yapo *et al*., 1996), and some degree of expertise and experience with the model (Hsu *et al*., 1995). With the time-series modelling approach in the second group, most of them fall within the framework of multivariate autoregressive moving average (ARMA) models (Raman and Sunilkumar, 1995). In runoff forecasting, the time-series models consider the stochastic structure of the time sequence of the runoff and precipitation values measured over time. They are more practical than the conceptual models in the sense that there is no need to understand the internal structure of the physical processes which are taking place in the system being modelled. The limitation of the univariate time-series methods is that the only information they incorporate is the value of the past flows. Many of the available techniques are deficient in that they are not considered as non-linear dynamics inherent in the transformation of rainfall to runoff.

New computing tools and black-box modelling techniques have been introduced to take into cognizance the above mentioned insufficiency. In the data driven modelling, the input variables connect to the output of a system with only limited knowledge about the physical behaviour of the system. Meanwhile, techniques used for data-driven modelling can be stated as machine learning (decision tree, Bayesian methods, neural networks, reinforcement learning), soft computing (fuzzy inference systems, neuro-fuzzy), data mining (which uses machine learning methods and statistics), non-linear dynamics, and chaos theory. These categories often overlap each other (Solomatine, 2002).

Many researches have applied Artificial Neural Networks (ANNs) to model different complex hydrological processes. Some ANN methods (Rumelhart *et al*., 1986) have been successfully employed to simulate the rainfall-runoff process. In addition, the ANN methods have good generalization efficiency and are commonly used in practical hydrologic projects (Zealand *et al*., 1999). Even when there are missing data values, the ANN methods can be applied to aid in the completion of missing hydrological records (Khalil *et al*., 2001). Some authors have compared Box–Jenkins with the ANN methods (Hsu *et al*., 1995; Abrahart and See, 1998) and confirmed, in most cases, the more accurate performance of the ANNs. Traditionally, it is just a matter of studying the cause-effect relationship with historical data. However, this simple statistics does not take into account other issues which take place in a time series.

This paper presents the results from a study on the application of the feed-forward back-propagation neural networks to forecast the next day's reservoir inflow. The modelling procedure is demonstrated for a dam reservoir in Malaysia, where stream inflows are not measured, catchment overland runoff inflow is virtually not measured and rainfall being measured in a three-station network strategically placed in the catchment.

METHODOLOGY - ANN MODELLING

There is this perpetual growing interest in the modelling of non-linear relationships. The artificial neural networks (ANNs) are essentially semi-parametric regression estimators and therefore suitable for this purpose. A significant advantage of the ANN approach in system modelling is that a well-defined physical relationship for systematically converting an input to an output is not required. What is needed for most networks is a collection of representative examples (input–output pairs) of the desired mapping. Then, the ANN adapts itself to reproduce the desired output

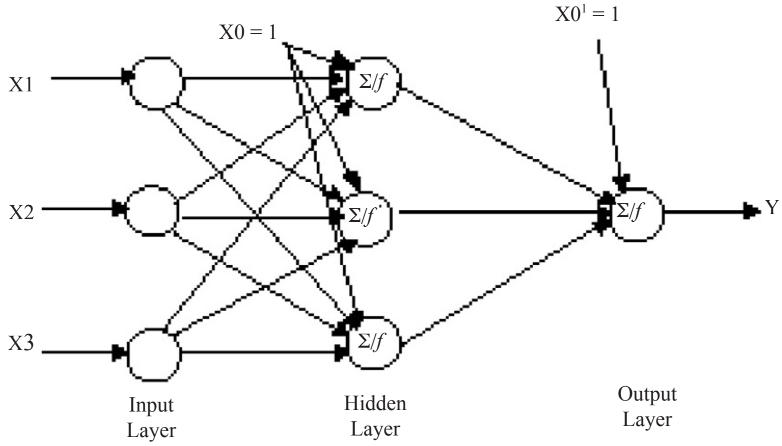

Fig. 1: Structure of a typical MLP

when presented with the training sample inputs. Meanwhile, network architecture determines the number of connection weights (free parameters) and the way information flows through the network. The determination of appropriate network architecture is not only one of the most important, but it is also one of the most difficult tasks in the model building process. The multi-layer perceptron (MLP) is the most popular network architecture in use today, due originally to Rumelhart *et al*. (1986), and it is discussed in most neural network text references. As shown in *Fig. 1*, a typical MLP has neurons which are arranged in a distinct layered topology with its three-layer relationship.

Network geometry determines the number of connection weights and how they are arranged. This is generally done by fixing a number of hidden layers and choosing the number of nodes in each of these. The ANNs with one hidden layer has been shown to be able to approximate any function. The number of nodes in the input layer is fixed by the number of model inputs, whereas the number of nodes in the output layer is equal to the number of the model outputs. In this study, there was only one output, i.e. the reservoir inflow. The selection of the final structure of the ANN model started with a minimum number of nodes in the hidden layer of two since the optimum number of hidden notes cannot be pre-determined, and the network is trained until a minimum mean square error is attained. Then, the number of nodes in a hidden layer is gradually increased until such increase does not significantly improve the performance of the neural network, and thus result in an optimal number of notes in the hidden layer.

The process of optimizing the connection weights is known as 'training' or 'learning.' Here, the Levenberg–Marquardt backpropagation training (LMBP) is used to train a feed-forward neural network. The transfer functions most commonly used are the sigmoidal type functions such as the logistic and hyperbolic tangent functions. In this study, hyperbolic and linear functions were considered for the hidden and output layers, respectively. The performances of the models developed in this study were evaluated using standard statistical performance evaluation based on error measures. More specifically, three different statistical performance indices have been employed, namely normal root mean squared error or NRMSE (which is preferred in many iterative prediction and optimization schemes), Pearson's correlation coefficient or R (which represents the relationship between two parameters giving a scatter plot by a linear relationship), and Nash–Sutcliff efficiency or CE (which includes both observed and predicted values of the same). These are defined as follows:

$$NRMSE = \frac{\left[(1/n)\sum_{t=1}^{n}(Q(t) - Q_o(t))^2\right]^{1/2}}{(1/n)\sum_{t=1}^{n} Q_o(t)} \tag{1}$$

$$R = \frac{\sum_{t=1}^{n}(Q_o(t) - \overline{Q_o})(Q(t) - \overline{Q})}{\sqrt{\sum_{t=1}^{n}(Q_o(t) - \overline{Q_o})^2 (Q(t) - \overline{Q})^2}} \tag{2}$$

$$CE = \frac{E_1 - E_2}{E_1} \tag{3}$$

where

$$E_1 = \sum_{t=1}^{n}\left(Q_o(t) - \overline{Q_o}\right)^2 \tag{4}$$

$$E_2 = \sum_{t=1}^{n}\left(Q(t) - Q_o(t)\right)^2 \tag{5}$$

Where $Q_0(t)$ is the observed inflow at time t, Q(t) is the estimated inflow at time t, n is the total number of inflow data points estimated from the developed ANN model, $\overline{Q_o}$ is the mean observed inflow, Q is the mean estimated inflow. E_i is the respective sum square of differences.

CASE STUDY

The Sembrong dam (*Fig. 2* and *Plate 1*), which is a dual purpose water supply cum flood mitigation dam was considered in this study. The catchment of dam lies to the east of the Kuala Lumpur-Singapore road and to the west of the Kuala Lumpur-Singapore railway in an area bounded by the latitude of $1^0$57'-$2^0$5' N and by the longitude of $103^0$8'-$101^0$17' E in the west part of Peninsular Malaysia, about 10km from Air Hitam town in the state of Johor and on the Air Hitam - Kluang road.

The data which were needed and available were sourced from the water supplies company and the Department of Irrigation and Drainage which is in charge of the reservoir operations whilst reservoir inflow data were extracted using water balance equation for the reservoir. The average rainfall across the catchment was calculated by constructing the well-known Theissen's polygon *(Fig. 3)*. The daily data, which were derived from a period of 6 years, i.e. from 1995-1998 and from 2002-2004 *(Fig. 4)*, were considered as the training-validation sets, whereas the other daily data taken for a period of 10 months, i.e. from March 2005 to the end of December 2005 *(Fig. 5)* were used as the testing set. Model inputs were selected by prior knowledge and research in the study area. Model inputs were reservoir inflow with four lags and average rainfall over the catchment with three lags and the output from the models was just the reservoir inflow for the next day. The evaluation procedure of the number of lags for the reservoir inflow and rainfall is not shown here since it is well-known. After the input (Q(t), Q(t-1), Q(t-2), Q(t-3), Q(t-4), R(t), R(t-1), R(t-2), R(t-3), where Q(t-i) and R(t-i) represent the reservoir inflow and rainfall at the *i*th lag number respectively) and output (Q(t+1) in output layer represented the reservoir inflow at

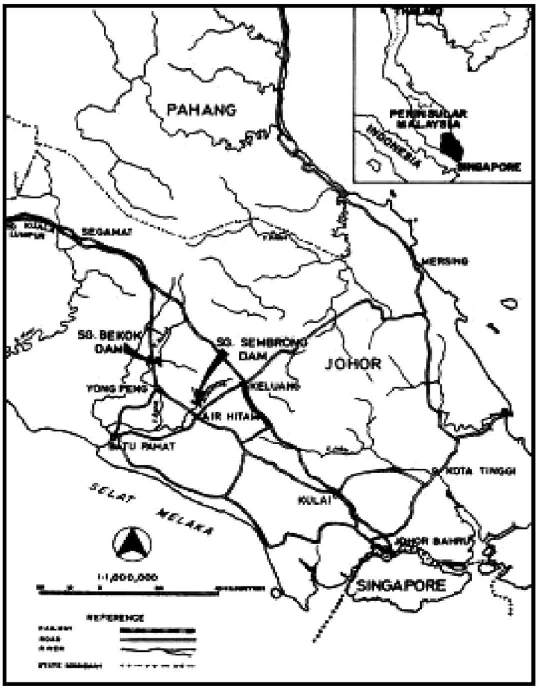

Fig. 2: Location map of the Sembrong Dam in Johor, Malaysia

Fig. 3: Theissen's polygon showing the locations of the three rainfall stations

Plate 1: An aerial view of the Sembrong Dam and reservoir and the river downstream of the dam (Courtesy of Department of Irrigation and Drainage, Kuala Lumpur)

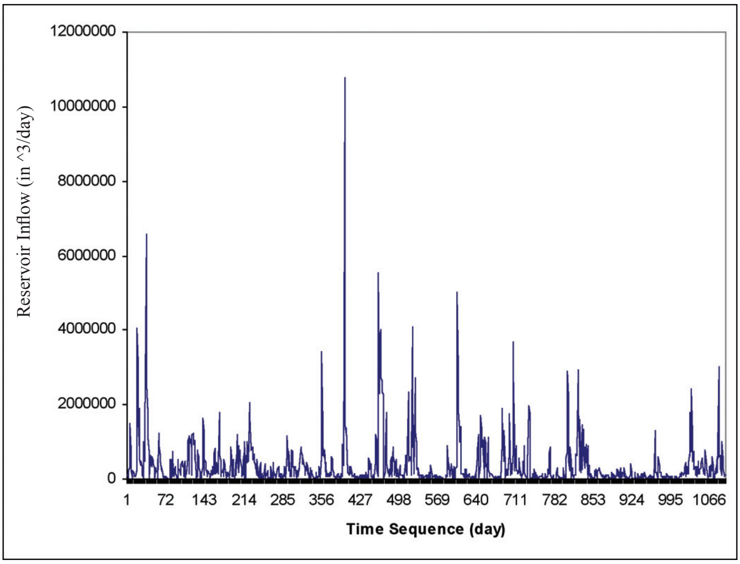

Fig. 4: Daily reservoir inflow data for training and validation during 1995 to1998 and 2002 to 2004

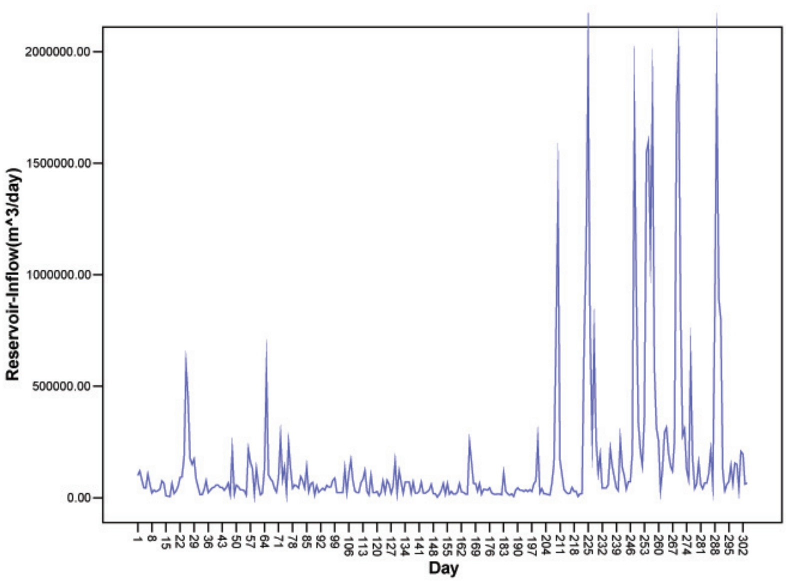

Fig. 5: Daily reservoir inflow data for testing (March – December, 2005)

time t+1) variables were selected, the ANN architecture 9-N-1 was explored for the capturing of the complex, dynamic, and non-linear, rainfall-discharge process. The architecture of the models is a 9-N-1 format in which N represent the number of neuron in hidden layer.

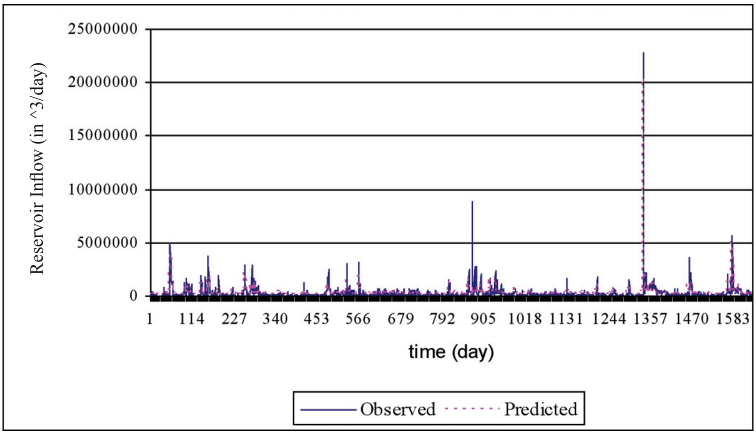

Fig. 6: Observed and predicted of selected AAN model in the training

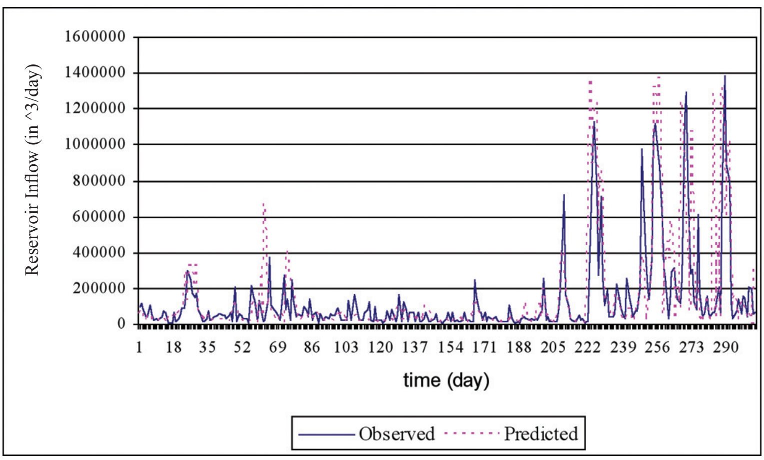

Fig. 7: Observed inflows and selected ANN model predicted inflows during the testing phase

RESULTS AND DISCUSSIONS

Different ANN models were trained with different number of neurons in hidden layer. Each model compromised of nine inputs, one hidden layer and one output. The performance criterion for each model during the training process is presented in Table 1. These results have been obtained using raw reservoir inflow and rainfall data (non-transformed, but scaled). On comparing the results, it was revealed that adding neurons (until 11neurons) to hidden layer increases the R and CE values and decrease NRMSE, which is desirable. Based on table 1, there is no point to add more neurons to the hidden layer thereafter (after 11 neurons) in view of the decreasing performance criteria with increasing neuron numbers. The best architecture obtained was selected with respect to maximum values of R and CE and minimum NRMSE value and this was from the model 9-11-1. This model was further employed to simulate the independent testing data and the results are also presented in Table 1. In the testing stage, although the selected model was found to result in improved NRMSE criterion but however, it simultaneously showed decreased R and CE values.

TABLE 1
Performance criteria of the ANN models with different architecture
during the training and testing phases

During training			
Architecture	Pearson's correlation coefficient (R)	Nash-Sutcliff efficiency coefficient (CE)	Normal root mean square error (NRMSE)
9-2-1	0.758	0.566	173.483
9-3-1	0.841	0.688	147.001
9-4-1	0.830	0.686	147.505
9-5-1	0.822	0.666	152.091
9-6-1	0.823	0.675	150.104
9-7-1	0.831	0.689	146.773
9-8-1	0.838	0.690	146.613
9-9-1	0.812	0.643	157.367
9-10-1	0.844	0.710	141.679
9-11-1	**0.862**	**0.738**	**134.648**
9-12-1	0.829	0.675	150.139
9-13-1	0.840	0.700	144.207
9-14-1	0.821	0.618	162.612
9-15-1	0.847	0.716	140.304
During Testing			

Fig. 6 shows the simulation results of the 9-11-1 model for the training data section. The model has some over prediction in base-flows but nevertheless, it was not able to capture the high peaks. Simulation results for testing period *(Fig. 7)* shows good generalization for base-flows and some over-prediction in peak flows.

CONCLUSIONS

In this study, the potential of a feed-forward multi layer perceptron for forecasting the daily reservoir inflow was investigated for the Sembrong dam catchment in Malaysia. An appropriate architecture of ANN model was found by trial and error. The model was examined for an independent data set. Results showed good generalization for base-flows and medium peaks but over-estimation in high peaks (which tends to err on the conservative side). With respect to specific climatic and hydrologic properties of small tropical catchment, the results of study show the validity of using neural network for modelling reservoir inflow series.

REFERENCES

Abrahart, R.J. and See, L. (1998). Neural networks vs. ARMA modelling: Constructing benchmark case studies of river flow prediction. Retrieved from http://divcom.otago.ac.nz/sirc/GeoComp/GeoComp98/05/gc_05.htm.

Duan, Q., Sorooshian, S. and Gupta, V.K. (19920. Effective and efficient global optimization for conceptual rainfall–runoff models. *Water Resources Research, 28* (4), 1015–1031.

Hsu, K., Gupta, H.V. and Sorooshian, S. (1995). Artificial neural network modeling of the rainfall–runoff process. *Water Resources Research, 31*(10), 2517–2530.

Khalil, M., Panu, U.S. and Lennox, W.C. (2001). Groups and neural networks based stream flow data infilling procedures. *Journal of Hydrology*, *241*, 153–176.

Raman, H. and Sunilkumar, N. (1995). Multivariate modelling of water resources time series using artificial neural networks. *Journal of Hydrology Science*, *40*(2), 145–163.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986). Learning internal representations by error propagation. In D.E. Rumelhart and J.L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (vol. 1) (pp. 318–362). Cambridge, MA : MIT Press.

Solomatine, D.P. (2002). Applications of data-driven modelling and machine learning in control of water resources. In M. Mohammadian, R.A. Sarker and X. Yao (Eds.), *Computational intelligence in control* (pp. 197 – 217). Idea Group Publishing.

Sorooshian, S. and Gupta, V.K. (1995). Model calibration. In V.J. Singh (Ed.), *Computer models in watershed hydrology* (pp. 23 – 68). USA: Water Resources Publications, Co.

Yapo, P., Gupta, V.K. and Sorooshian, S. (1996). Calibration of conceptual rainfall–runoff models: Sensitivity to calibration data. *Journal of Hydrology*, *181*, 23–48.

Zealand, C.M., Burn, D.H. and Simonovic, S.P. (1999). Short-term streamflow forecasting using artificial neural networks. *Journal of Hydrology, 214*, 32–48.