Turkish J. Eng. Env. Sci.
34 (2010) , 57 – 65.

doi:10.3906/muh-0908-7

Flood forecasting in similar catchments using neural networks

G. R. RAKHSHANDEHROO[1], M. VAGHEFI[2], M. M. SHAFIEE[3]
[1]*Associate Professor, Civil Engineering Department, Shiraz University, Shiraz-IRAN*
e-mail: rakhshan@shirazu.ac.ir
[2]*Assistant Professor (Hydraulic Structures), Civil Engineering Department,*
Persian Gulf University, Bushehr-IRAN
e-mail: vaghefi@pgu.ac.ir
[3]*Master of Science, Civil Engineering Department, Shiraz University, Shiraz-IRAN*
e-mail: ehsan_shafi@yahoo.com

Received 14.08.2009

Abstract

Flood forecasting is an essential requirement for solving a wide range of scientific and/or management problems. Neural networks have become attractive models in many practical applications, including flood forecasting. In this paper, 4 similar catchments in Iran, with high quality rainfall-runoff records, were studied. An artificial neural network (ANN) was built and trained as an event-based modeling tool utilizing data from only 2 of the catchments. The flood forecasting ability of the model was then evaluated for all catchments. Results showed that during the training, the model simulated observed peak flows very closely. Based on a good simulation of the response from unseen but similar catchments, it was concluded that the model may be utilized for flood forecasting in catchments that lie on the same cluster. However, as such catchments become less similar to the cluster; the simulation error would increase accordingly.

Key Words: Neural network, flood forecasting, similar catchments, ANN

Introduction

Flood forecasting is an essential requirement for solving a wide range of scientific and/or management problems related to the design and operation of river systems (Abrahart and See, 2000; Damangir, 2001; Akhtar et al., 2009;. There is no doubt that this forecasting can have a great impact on any human activity linked to a river. Flood forecasting is not only important in anticipating the hydraulic behavior of a river system, but also in predicting its environmental impacts, such as the transport and concentration of pollutants. In fact, forecasting river flow is one of the most important parameters in the environmental impact assessment (EIA) of different projects related to river systems. Furthermore, peak runoff is an essential parameter that affects the design of spillways for dams. The spillway must bear the huge amount of runoff that passes it; a magnitude that is usually determined by statistical analysis of peak river flows.

Peak runoff may be forecasted utilizing rainfall records. However, the rainfall-runoff process is believed to be highly nonlinear, time-varying, spatially distributed, and not easily described by simple models (Srinivasulu and

Jain, 2006). In general, river flow may be considered as the response of catchments to rainfall, snow melt, and other relevant hydrological parameters under certain meteorological conditions. Hydrological parameters that are known to influence the response of catchments to rainfall include temperature, humidity, wind, land cover, land use, and soil type. The more accurately we characterize the influence of such parameters in rainfall-runoff modeling of a catchment; the more precise is the flood forecasting in the catchment (Damangir, 2001).

Different types of models have been offered to characterize the rainfall-runoff process in a catchment. Two major approaches have been undertaken in these models; conceptual and black-box (Sorooshian et al., 1995). While conceptual models offer a powerful forecasting method in prediction of river flows, they are often considered to be too complex or too demanding. These models are designed to estimate internal subprocesses involved in the prototype, and physical mechanisms that govern the process. On the other hand, black-box models are more practical and less demanding, and at the same time perform well, and sometimes superiorly, when compared to conceptual models. Such models are becoming attractive alternatives in many practical applications, especially when data collection is limited (Nilsson et al., 2006).

Neural networks have been utilized by many researchers for flood forecasting and/or rainfall-runoff modeling in single catchments (Karunanithi et al., 1994; Tokar and Johnson, 1999; Abrahart and See, 2000; Dibike and Solomatine, 2002;; Tayfur et al., 2007). The networks have been described by researchers as having shown great promise as an effective tool for modeling, performing well in comparison with conventional methods, and probably the most successful machine learning technique with a flexible mathematical structure that is capable of identifying a complex nonlinear relationship between input and output data without attempting to reach an understanding as to the nature of phenomena (Dibike and Solomatine, 2002). Recorded rainfall-runoff events in single catchments were used by the mentioned researchers in order to develop appropriate networks applicable to the catchments under consideration. Floods in unseen but similar catchments, however, may be forecasted by a network trained in more or less similar catchments. Smith and Eli built an artificial neural network (ANN) model for the rainfall-runoff process and compared it to conceptual runoff modeling (Smith and Eli, 1995). They concluded that although both have capabilities to model the process successfully, the ANN model yielded improved results. Srinivasulu and Jain (2006) investigated many methods for modeling the rainfall-runoff process and used self-organizing maps (SOM) as a means for data clustering.

The use of flow length and travel time is a preprocessing step for incorporating spatial precipitation information into ANN models used for river flow forecasting. Spatially distributed precipitation is commonly required when modeling large basins, and it is usually incorporated in distributed, physically based hydrological modeling approaches. The incorporation of remote sensing data of spatially distributed precipitation information as a preprocessing step was shown to be a promising alternative for the setting up of ANN models for river flow forecasting (Akhtar et al., 2009).

Cévenol flash floods are famous in the field of hydrology because they are archetypical of flash floods that occur in populated areas, thereby causing heavy damages and casualties. As a consequence, their prediction has become a stimulating challenge to designers of mathematical models, whether based on physics or machine learning. Because current, state-of-the-art hydrological models have difficulty performing forecasts in the absence of rainfall previsions, new approaches are necessary. The results show that an appropriate model selection methodology, applied to neural network models, provides reliable flood forecasts with 2 h of warning (Toukourou et al., 2010).

In this paper, 4 similar catchments in Iran, with high quality rainfall-runoff records, were studied. They were relatively small catchments with similar hydrologic properties. Their similarity was assessed automatically by SOM and adaptive resonance theory (ART) using simple physical properties of the catchments. An ANN was built and trained as an event-based modeling tool utilizing data from only 2 of the catchments. The flood forecasting ability of the model was then evaluated for all catchments.

Theory

Many parameters influence the response of a catchment to rainfall (Sorooshian et al., 1995). Major parameters may be grouped as follows:

1. Catchment physical properties, such as area, perimeter, shape, location, river length, or slope;
2. Catchment soil properties, such as hydrologic soil type, soil structure, erosion ability, or soil composition;
3. Land cover and land use in the catchment;
4. Meteorological conditions of the catchment;
5. Sediment load in rivers; and
6. Human activities in the catchment.

It is believed that, by far, a catchment's physical properties play the most important role in its hydrologic behavior (Sorooshian et al., 1995).

Neural networks offer an important alternative to traditional methods of data analysis and modeling (Flood and Kartam, 1994). While different types of neural networks exist, the one that is of interest at the moment is the feed-forward multilayer perceptron (MLP). The basic structure of an MLP is not complicated. It consists of a number of simple processing elements (neurons) arranged in different layers, joined together to form a network (Fausett, 1994). The processing elements sum their inputs, effect a nonlinear data squashing process, and then transmit a single output to all processing elements in the next layer via connections (Fausett, 1994).

Not all recorded rainfall or runoff data are good or reliable. Input data screening makes an extreme improvement on the quality of output from an ANN model (Sorooshian et al., 1995; El-Din and Smith, 2002;. In fact, it is not important to have many sets of rainfall-runoff records, but it is important to have a criterion for choosing reliable data from all recorded data. There are many criteria for determining data quality in rainfall-runoff records. As input dimensionality increases, the computational complexity and memory requirements of the model increase, and learning becomes more difficult with irrelevant inputs (Sorooshian et al., 1995). Researchers have employed 5 methods for data screening to improve performance of rainfall-runoff modeling with ANN. Some methods utilize linear cross-correlation while others rely on the use of a priori knowledge of the system being modeled. The cross-correlation method has been the most popular analytical technique for selecting appropriate inputs (Sorooshian et al., 1995).

Data used

Fifteen years of high quality rainfall and runoff records for Amameh and Kasilian catchments in northern, Fakhrabad in central, and Kardeh in northeastern Iran were gathered (Figure 1). These catchments are called determined catchments and their data have been collected by the responsible governmental organizations under the Ministry of Energy since 1969. Physical properties of the catchments were utilized in order to assess their similarity (Table 1). As shown in Table 1, the catchments have different physical properties, with Amameh and Kardeh having the smallest and the largest areas, respectively.

Table 1. Physical properties of determined catchments utilized for similarity assessment.

Catchment			Kardeh	Amameh	Kasilian	Fakhrabad
Area	A	km^2	431.4	37.2	68.4	206
Length	L	km	38	13	18.7	15.25
Shape index	FF		0.30	0.22	0.20	0.89
Slope	S	%	2.6	15.1	10.7	9.8
Lag time	$\underline{t}$	h	4.40	0.98	1.48	1.31

Rainfall-runoff records of the catchments included 15 min rainfall hyetographs and 1 h runoff hydrographs. A screening procedure was performed and over 100 sets of reliable and consistent data were extracted for model training and evaluation. The screening applied the cross-correlation method to exclude inconsistent data. Average rainfall intensity and peak flow (Q_{max}) were chosen as rainfall and runoff indices for cross-correlation, respectively. Figures 2 and 3 depict the data before and after the screening, respectively. A much more consistent scatter of data on Figure 3 reflects the improvement made by the screening procedure.

Figure 1. Location of catchments.

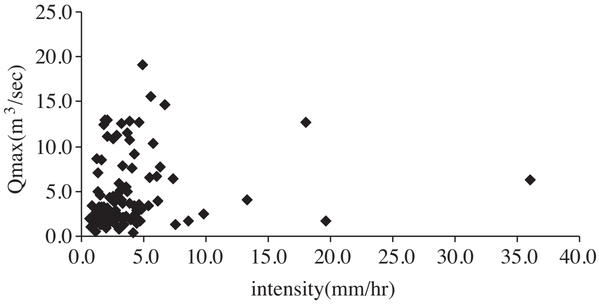

Figure 2. Rainfall and runoff indices before the screening.

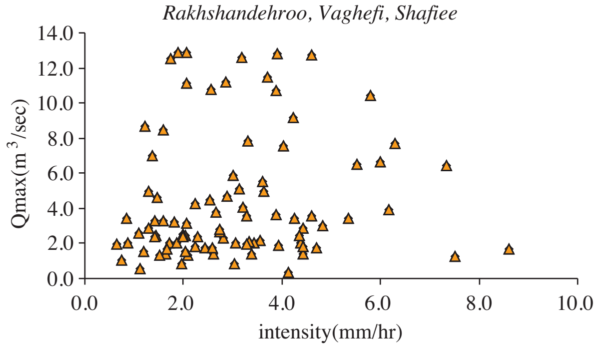

Figure 3. Rainfall and runoff indices after the screening.

Catchment clustering

SOM and ART were employed to cluster the catchments. SOM placed the Kasilian and Amameh catchments in a common cluster, whereas ART verified the same result with maximum similarity for the 2 catchments (57.5%). Fakhrabad and Kardeh had less than 29% and 13% similarity with the cluster, respectively. Hence, Fakhrabad was considered closer to the cluster when compared with the Kardeh catchment. Percent similarity for each pair of catchments (or the vigilance parameter according to ART) is shown in Table 2.

Table 2. Percent similarity for pairs of catchments according to ART.

Catchments	Kardeh	Amameh	Kasilian	Fakhrabad
Kardeh	100	0.5	12.5	34
Amameh	—-	100	57.5	28.8
Kasilian	—-	—-	100	20.7
Fakhrabad	—-	—-	—-	100

Method of study

Observed sets of rainfall and runoff for the Amameh and Kasilian catchments (the most similar catchments) were collected from 1983 to 1998. Rainfall durations and average intensities of over 100 events were extracted as the base input data for the model. Runoff peak flow of the same events was the base output data. Both sets of data were normalized according to the following equation, so that all data lay between 0 and +1:

$$x_{normalized} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

In this normalization,X was the observed data (rainfall intensity, duration, or runoff peak flow) and X_{min} and X_{max} were the minimum and maximum values of the same data, respectively.

It was found that the response of catchments to a certain rainfall strongly depended on the season. It was postulated that different hydrological conditions in different seasons (such as land cover, land use, weather and soil conditions, or humidity) had an influence on the runoff peak flow. Its inclusion in the model increased the accuracy of results by approximately 6%. Thus, an indicator was employed to reflect the seasonal influence in the model. For this purpose, observed runoff peak flows in each season were averaged for the period of study

(15 years). Season indices were calculated by normalizing this data for the entire year (Table 3). The season index was used as additional input data in the model.

Table 3. Season indices for the cluster of Amameh and Kasilian catchments.

Season	Season index
Fall	0.53
Winter	1.0
Spring	0.37
Summer	0.0

An MLP network was employed to model rainfall intensity, duration, and season index as input and runoff peak flow (Q_{max}) as output data. In order to achieve the best performance for the network, the authors examined many MLP networks with different numbers of hidden layers and neurons. The best network was selected based on the least mean square error (MSE). This network comprised 1 hidden layer and 10 neurons.

The model was employed to perform 3 tests. The first test was on the learning data; the data used for model training. This data consisted of approximately 90 sets of data for Amameh and Kasilian; the 2 catchments considered as seen. The model performance was verified by the second test, which was performed on 20 unseen sets of data for the same catchments. The flood forecasting capability of the model for the unseen but more or less similar catchments was tested by the third test, in which data from the Fakhrabad and Kardeh catchments were employed. In this test, runoff peak flows for 15 observed rainfalls at the Fakhrabad and Kardeh catchments were predicted. In all tests, the observed and calculated runoffs were compared and the root mean square error (RMSE), sum of squared errors (SSE), and average percent error (APE) were examined.

Results and Discussion

Figure 4 shows the observed versus calculated peak flows for the training data of Amameh and Kasilian catchments (the first test). As seen in the figure, both sets of values are so close to each other that the 2 graphs can hardly be distinguished from one another. Table 4 quantifies the error for this figure. The small values for RMSE, SSE, and APE shown in the table reflect the fact that during the training, the model simulated observed peak flows very closely.

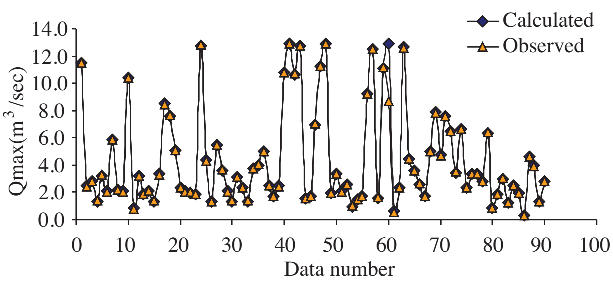

Figure 4. Calculated versus observed peak flows for the training data (the first test).

Table 4. Error for the training data (the first test).

RMSE (m^3/s)	SSE (m^6/s^2)	APE (%)	
0.0377	0.1281	1.79	First test

Figure 5 depicts the calculated versus observed peak flows for unseen data from the same cluster of catchments (Amameh and Kasilian). This figure represents the second test for the model. As shown in the figure, the model again simulated observed peak flows very closely. Small values of 7.29%, 1.95%, and 0.305% were obtained for the APE, SSE, and RMSE, respectively. It was concluded that the model may be utilized for flood forecasting in catchments that lie on the same cluster as Amameh and Kasilian with a small error.

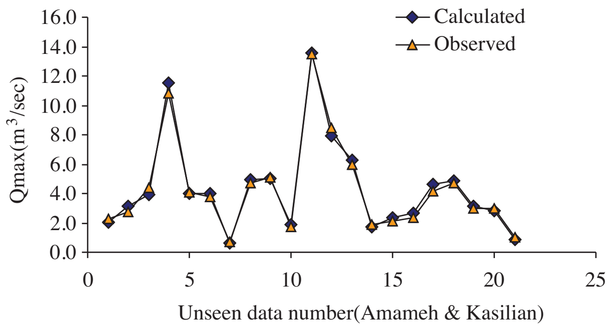

Figure 5. Calculated versus observed peak flows for unseen data from the Amameh and Kasilian catchments (the second test).

Figure 6 shows the calculated versus observed peak flows for Fakhrabad and Kardeh catchments. As mentioned earlier, according to ART, these catchments were less than 34% similar to the cluster under study. Considering the fact that the model was only trained for the Amameh and Kasilian catchments, Figure 6 reflects a good match with an acceptable error. The APE for Fakhrabad and Kardeh catchments were 27.1% and 43.2%, respectively. The SSE and RMSE for this test were 37.16 and 1.69, respectively. In fact, the model yielded a better match for the more similar catchment to the cluster (Fakhrabad). It was concluded that the model may be utilized for flood forecasting in unseen but similar catchments. However, as such catchments grow more similar to the cluster; one may expect the simulation error to decrease.

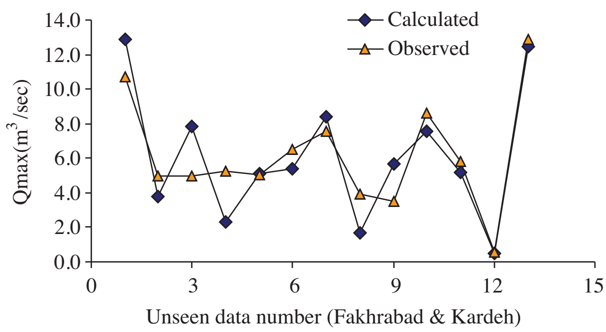

Figure 6. Calculated versus observed peak flows for unseen data from the Kardeh and Fakhrabad catchments (the third test).

Conclusion

The capability of event-based modeling using ANN to forecast floods in similar catchments was explored. A multilayer perceptron (MLP) network was employed to model rainfall intensity, duration, and season index as input and runoff peak flow (Q_{max}) as output data for 4 similar catchments in Iran. The network was trained for 2 catchments that lay on the same cluster according to ART. Results showed that during the training, the model simulated observed peak flows very closely. Based on a good simulation of the unseen data, it was concluded that the model may be utilized for flood forecasting in catchments that lie on the same cluster. It was shown that the model may also be utilized for flood forecasting in unseen but similar catchments. However, as such catchments grow less similar to the cluster, the simulation error would increase accordingly.

Notation

A	Area; km^2)
L	Length;km)
FF	Shape Index
S	Slope;%)
t	Lag Time;h)
$Q_{\max}$	Peak Flow;m^3/s)
X	Observed Data (e.g. m^3/s)
$X_{\min}$	Minimum Values of the Same Data
$X_{\max}$	Maximum Values of the Same Data
$RMSE$	Root Mean Square Error;m^3/s)
SSE	Sum of Squared Errors;m^6/s^2)
APE	Average Percent Error;%)

References

Abrahart, R. J. and See, L., "Neural Network vs. ARMA Modeling: Constructing Benchmark Case Studies of River Flow Prediction", Hydrological Processes, 14, 2157-2172, 2000.

Akhtar, M.K., Corzo, G.A., Van Andel, S.J. and Jonoski, A., "River Flow Forecasting with Artificial Neural Networks Using Satellite Observed Precipitation Pre-Processed With Flow Length and Travel Time Information: Case Study Of The Ganges River Basin", Hydrology and Earth System Science, 13, 1607-1618, 2009.

Damangir, H., "Dynamic Training of ANN for Its Application in Real-Time Flood Forecasting", M.S. Thesis, Shiraz University, Shiraz, Iran, 2001.

Dibike, Y. B. and Solomatine, D. P., "River Flow Forecasting Using Neural Networks", Hydro-information Section, International Institute for Infrastructural Hydraulic and Environmental Eng., Delft; The Netherlands, 2002.

El-Din, A. and Smith, D., "A Neural Network Model to Predict the Wastewater Inflow Incorporating Rainfall Events", Water Resources Research, 36, 1115-1126, 2002.

Fausett, L. V., "Fundamentals of Neural Networks", Prentice Hall International, 527, 1994.

Flood, I. and Kartam, N., "Neural Networks in Civil Engineering 1: Principles and Understanding", Journal of Computing in Civil Eng., 8, 131-148, 1994.

Gaven, J.B., Greame, C.D. and Holger, R.M., "Input Determination for Neural Network Models in Water Resources Applications. Part 1-Background and Methodology", Journal of Hydrology, 254, 175-182, 2005.

Karunanithi, N., Grenny, W.J., Whitley, D. and Bovee, K., "Neural Network for River Flow Prediction", Journal of Computing in Civil Eng., 8, 201-220, 1994.

Nilsson, P., Uvo, C.B., and Berndtsson, R., "Monthly Runoff Simulation: Comparing and Combining Conceptual and Neural Networks Models", Journal of Hydrology, 321, 344-363, 2006.

Smith, J. and Eli, R., “Neural Network Models of Rainfall-Runoff Process”, Journal of Water Resources Planning and Management, 121, 499-508, 1995.

Sorooshian, S., Hsu, K. and Gupta, H.V., “Artificial Neural Network Modeling of the Rainfall-Runoff Process”, Water Resources Research, 31, 2517-2530, 1995.

Srinivasulu, S. and Jain, A., “A Comparative Analysis of Training Methods for Artificial Neural Network Rainfall-Runoff Models”, Applied Soft Computing, 6, 295-306, 2006.

Tayfour, G., Moramarco, T. and Singh, V.P., “Predicting and Forecasting Flow Discharge at Sites Receiving Significant Lateral Inflow”, Hydrological Processes, 21, 1848-1859, 2007.

Tokar, A.S. and Johnson, P., “Rainfall-Runoff Modeling Using Artificial Neural Networks”, Journal of Hydrological Engineering, 4, 232-239, 1999.

Toukourou, M., Johannet, A., Dreyfus, G. and Ayral, P., “Rainfall-Runoff Modeling of Flash Floods in the Absence of Rainfall Forecasts: The Case of ‘Cévenol Flash floods”’, Applied Intelligence, published online, 20 February 2010.