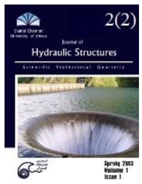
Journal of Hydraulic Structures
***J. Hydraul. Struct.*, 2016; 2(2):62-73**
DOI: 10.22055/jhs.2016.12853



Flood Forecasting Using Artificial Neural Networks: an Application of Multi-Model Data Fusion Technique

Yaser Tahmasebi Biragani[1,2]
Farhad Yazdandoost[3]
Hossein Ghalkhani[4]

Abstract

Floods are among the natural disasters that cause human hardship and economic loss. Establishing a viable flood forecasting and warning system for communities at risk can mitigate these adverse effects. However, establishing an accurate flood forecasting system is still challenging due to the lack of knowledge about the effective variables in forecasting. The present study has indicated that the use of artificial intelligence, especially neural networks is suitable for flood forecasting systems (FFSs). In this research, mathematical modeling of flood forecasting with the application of Artificial Neural Networks (ANN) and data fusion technique were used in estimating the flood discharge. Sensitivity analysis was performed to investigate the significance of each model input and the best MLP ANN architecture. The data used in developing the model comprise discharge at different time steps, precipitation and antecedent precipitation index for a major river basin. Application of model on a case study (Karun River in Iran) indicated that rainfall-runoff process using data fusion approach produces results with higher degrees of precision.

Keywords: Flood Forecasting, Neural Networks, Data Fusion, Sensitivity Analysis, Karun River

Received: 05 June 2016; Accepted: 23 October 2016

1. Introduction

Flood protection and awareness have continued to rise on the political agenda over the last decade accompanied by a drive to ‘improve’ flood forecasts (Demeritt, et al. [5]; van Berkom, [12]). Without a doubt, flood flow information is very important for early warning. Flood flows in downstream areas are strongly influenced by upstream conditions. The artificial neural network (ANN) model has shown to be appropriate for the above-mentioned problem (Chen et

[1] Department of Environmental Health Engineering, School of public Health, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran, (**Corresponding author**)
[2] Environmental Technologies Research Center (ETRC), Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran
[3] Department of Civil Engineering, K. N. Toosi University of Technology, Tehran, Iran
[4] Department Water Resources Researches, Water Research Institute, Tehran, Iran

al. [3]). Neural network models possess a distributive processing system and are able to store inherent characteristics of data in the form of a large storage for later generalization (French et al. [7]). The ANN model has various mathematical compositions capable of modeling extremely complex physical systems. It has the potential to be more flexible and less assumption dependent approach to act as a simulation model for the hydrologic systems (Sudheer and Jain [11]). In this paper, a model has been presented for flood forecasting using data fusion technique and artificial neural networks (ANNs). Data fusion is an emerging area of research that covers a broad spectrum of application areas ranging from ocean surveillance to strategic warning, and medical diagnosis (Hall [8]). The principal objective of data fusion, which is process of combining or amalgamating information from multiple sensors and/or data sources, is to provide a solution that is either more accurate according to some measure of evaluation, or allows one to make additional inferences above and beyond those that could be achieved through the use of single source data alone (Dasarathy [4]). Data fusion researches are divided into two board groups. The first takes the view that data fusion is the amalgamation of raw information to produce an output, while the second advocate a more generalized view of data fusion in which both raw and processed information can be fused into useful outputs including higher level decision. See and Abrahart used data fusion approach for continuous river level forecasting where data was the amalgamation of information from multiple sensors and different data sources (See and Abrahart [9]). Abrahart and See evaluated six data fusion strategies and found that data fusion by an (ANN) model provided the best solution (Abart and See [1]). Shu and Burn applied (ANN) ensembles in pooled flood frequency analysis for estimating the index flood and the 10-year flood quintiles (Shu and Burn [10]). Araghinejad et al. [2] applied the combination of data fusion technique and probabilistic method to forecast peak discharge of Red River in Canada and seasonal stream flow in Zayandeh-rud River in Iran. Despite remarkable studies in this field, further investigations need to be conducted to find the effects of various inputs in forecasting models. This study considers several combination of the different-input models. These inputs range from discharge at different time steps to precipitation and antecedent precipitation index (API).

The remainder of the paper is organized in the following sections: section 2 is associated with the methodology of the study. The methodology section explains data fusion method, typical ANN model as well as different-input ANN models. In section 3, proposed methodology is applied on Karun River in Iran as a case study. Combination of the different ANN model with various inputs are tested on this case study. In the last section, this paper finally gives a conclusion according detailed results.

2. Methodology

2.1. Data Fusion Methods

The general equation of a hydrological event forecasting model is

$$y_{i=}f(X_i) + \varepsilon_i \qquad i = 1,2, \dots n \tag{1}$$

where X= vector of predictors, y= forecast variable, $\varepsilon =$model error and n= number of observation data. In the case of using multiple models to forecast y, and considering similar predictors, Eq. (1) is changed to the following matrix from

$$[Y_i] = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{im} \end{bmatrix} = \begin{bmatrix} f_1(X_i) \\ f_2(X_i) \\ \vdots \\ f_m(X_i) \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{im} \end{bmatrix} \tag{2}$$

where m= number of forecast models used to estimate y., $[Y]$= matrix of estimations of y provided by different individual models. Using the data fusion approach, $[Y]$ is the sum up to a unique estimation of y.

2.2. Artificial Neural Networks (ANNs) Model

Empirical models, particularly (ANNs) are known as powerful tools for function mapping. See and Abrahart [1] have suggested the use of ANNs as a data fusion method. The general form of this method is

$$y_i = g([Y_i]) \qquad [Y] = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{im} \end{bmatrix}; i = 1, \dots, n \tag{3}$$

where g is a non-linear function which maps outputs of different individual forecast models to a single output of y_i using ANN model. ANN models generally are a mathematical model that is devised by analogy with biological brain cells. It is a highly interconnected structure consisting of an input layer, a hidden layer (or hidden layers), and an output layer. Fig. 1 shows a general structure of the neural network with one hidden layer. The nodes (called neurons) receive and process input signals and send an output signal to other nodes in the network. The output of each node is defined by

$$net_j = \sum_{i=1}^{n} w_{ij} . o_i \tag{4}$$

where net_j is the net input information arriving at node j, w_{ij} is the connection strength or weight between nodes i and j and $o_j$is the activation at node i. The level of activation is then updated using the following sigmoid function

$$o_j = \frac{1}{1+\exp(net_j)} \tag{5}$$

where $o_j$is the output at node j.

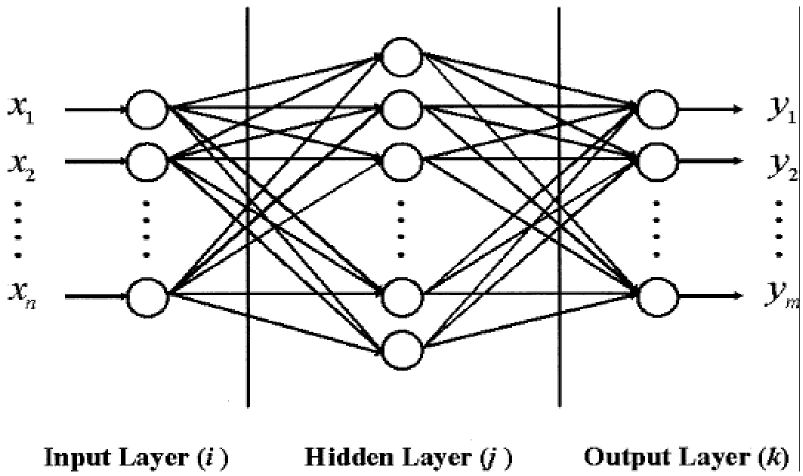


Fig. 1 A General Architecture of an Artificial Neural Network.

In the error back propagation algorithm, errors are estimated to observe the performance of model and the values of weight are recalculated accordingly. This is referred to as a learning process. The error, E, is defined by

$$E=\frac{1}{2}\sum_{p=1}^{P}\sum_{k=1}^{K}(d_{pk}-o_{pk})^2 \quad (6)$$

where P is the total number of patterns in the training set, K is the total number of output nodes, and d_{pk} and o_{pk} are the desired and calculated outputs at node j for pattern p, respectively.

A Multi-Layer Perceptron (MLP) method with back propagation is used here which has input, output, and hidden middle layers.

2.2.1. Model Development

One of the main aims in design of warning systems is to increase the pre-warning time based on accurate simulation of flood inundation levels. Two models were used for this purpose

$$Q(t)=f[Q(t-2),Q(t-1),P_i(t),P_i(t-1),API(X),T(X),T(t-1)] \quad (7)$$

$$Q(t+1)=f[Q(t-1),\ Q(t),\ P_i(t),P_i(t-1),API(X),T(t),T(t-1)] \quad (8)$$

where Q is the discharge, P is the precipitation, T is the temperature and $API(x)$ is the previous precipitation index at previous x time steps and t is the computational time step. The previous precipitation index may be obtained from:

$$API(X)=\sum_{j=1}^{i}P_{X-j}.K^{-j} \quad (9)$$

$$K=EXP(\frac{-ET}{Wm}) \quad (10)$$

where i is the number of passed days, K is the precipitation constant varying between 0.85 and 0.95, P_{x-j} is the precipitation at day $x-j$, ET is the evapotranspiration and Wm is the maximum available soil moisture for evaporation.

2.2.2. Data preprocessing

Since input variables appear in different scale, therefore the values of inputs should be changed in the same scale in a specified interval. Following equations can be used to normalize input values between (0.05-0.95):

$$X_n=0.05+0.9\frac{X-X_{min}}{X_{max}-X_{min}} \quad (11)$$

where, X_n is the normalized value of input X, X_{min} is the minimum of the input X and X_{max} is the maximum of the input X.

2.2.3. Evaluating the performance of the model

In order to evaluate the performance of ANN models, Nash-sutcliffe, Mean Relative Error and correlation coefficient parameters are used as following equations:

$$E=1-\frac{\sum_{i=1}^{n}(Q_{obsi}-Q_{simi})^2}{\sum_{i=1}^{n}(Q_{obsi}-\bar{Q}_{obs})^2} \quad (12)$$

$$MRE=\frac{\sum_{i=1}^{n}\left|\frac{Q_{obsi}-Q_{simi}}{Q_{obsi}}\right|}{n} \quad (13)$$

$$R=\frac{n\sum_{i=1}^{n}(Q_{obsi}\cdot Q_{simi})-(\sum_{i=1}^{n}Q_{obsi})\cdot(\sum_{i=1}^{n}Q_{simi})}{\sqrt{[n.(\sum_{i=1}^{n}Q_{obsi}{}^2)-(\sum_{i=1}^{n}Q_{obsi})^2].[n.(\sum_{i=1}^{n}Q_{simi}{}^2)-(\sum_{i=1}^{n}Q_{simi})^2]}} \quad (14)$$

where, E is the Nash-Sutcliffe parameter, MRE is the mean relative error, R is the correlation coefficient, Q_{obsi} is the ith observed discharge , Q_{simi} is the ith simulated discharge and n is the number of data.

3. Case study

The case under examination was that of Karun River in Khuzestan province, southwest of Iran (Fig.. This is the longest river of Iran (about 950Km) stretched from longitude $50^{\circ}18'28''E$ and latitude $31^{\circ}35'1''N$to longitude $48^{\circ}9'54''E$ and latitude $39^{\circ}25'40''N$. Karun supplies water demand of the important cities located along the river such as Ahvaz and Shushtar. Karun river collects the runoff of a 62570 km^2-basin and conveys to the Persian Gulf (Emamgholizade et al [6]). Therefore, there is a potential of flood risk in cities in the vicinity of Karun. Therefore flood management options such as flood warning systems should be established for Karun River. In this paper, data were gathered from hydrometric stations along the Karun (See Fig. 2 for their locations). The input data are those that had been mentioned in section 2.2.1. Out of the total observation data, some 5844 days had daily corresponding values which in turn provided 5844 patterns. 4197 first patterns were chosen for training and the remaining 1647 patterns for testing the neural network.

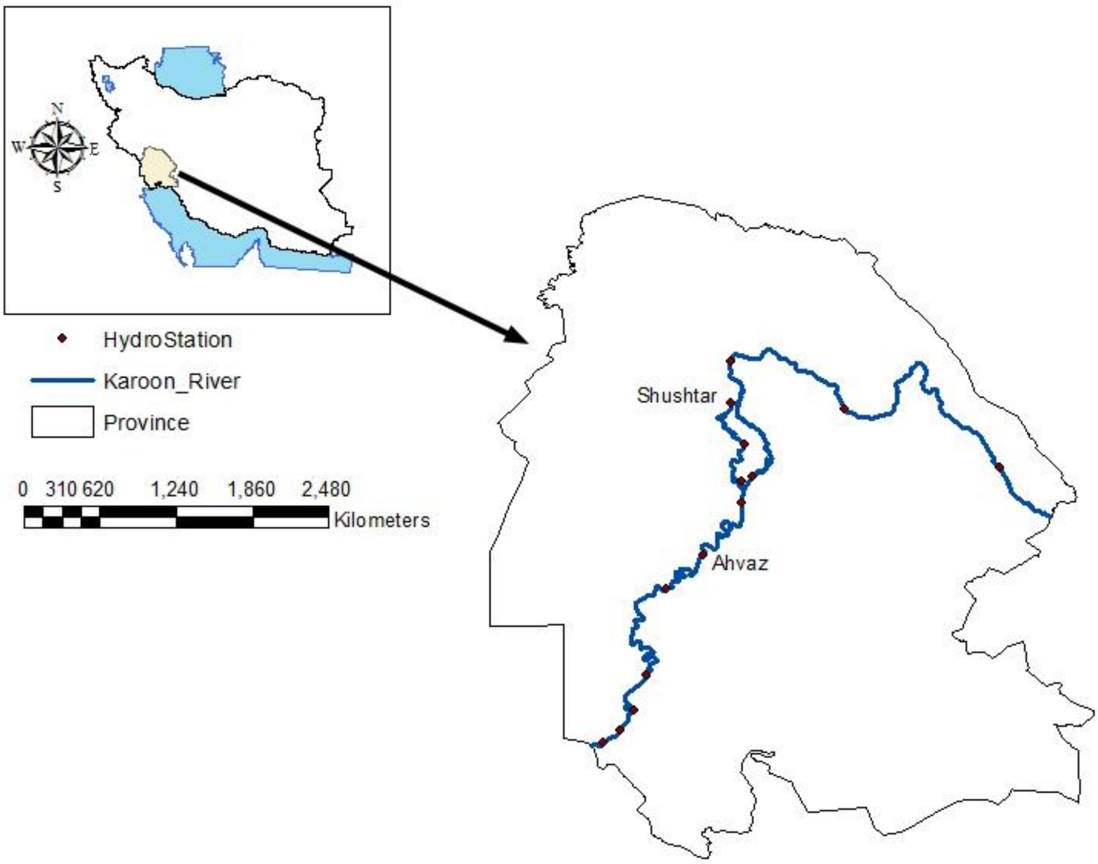


Fig. 2 location of the study area

4. Application of Models and Test of their Reliability

4.1. forecasting discharge in time step t

4.1.1. Model 1

In this model an MLP network with two inputs of discharge at time steps (t-1) and (t-2) were used for training the discharge at time step t. Upon a trial and error procedure the network was designed as (2-2-1) i.e. 2 input neurons, 2 neurons at hidden layer and 1 output neuron. Values of MRE, E and R^2 were obtained to be equal to 10.53, 0.815 and 0.797 respectively. Results are presented in Figure 2. Scattering of the data around the bisector line as depicted in Fig.3 shows that this model cannot efficiently predict flood flow except for a few small floods. In fact, applying this model can increase the risk of inaccurate flood forecasting. Fig. 3 also implies that considering only discharge at two previous time steps cannot give an accurate prediction of floods.

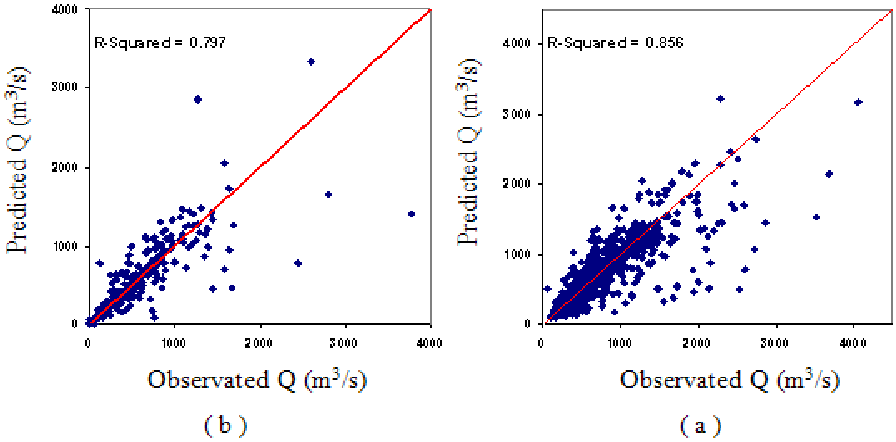


Fig. 3 Observed discharge versus predicted discharge for model 1: a) training data, b) testing data

4.1.2. Model 2

The second model uses precipitation data from 24 stations along the discharge from the previous time step of the output station to simulate the river flow. An MLP network with 25 inputs was trained to obtain the discharge at time step t. The network design was initially set at (25-5-1). Upon sensitivity analysis and optimization, the architecture of the network was designed at (14-6-1). Values of MRE, E and R^2 were obtained to be equal to 12.83, 0.889 and 0.883 respectively. As it may be seen from Fig. 3, this model performs better as it can predict low flow floods with more accuracy. However, large flood flows are not yet predictable (Fig. 4).

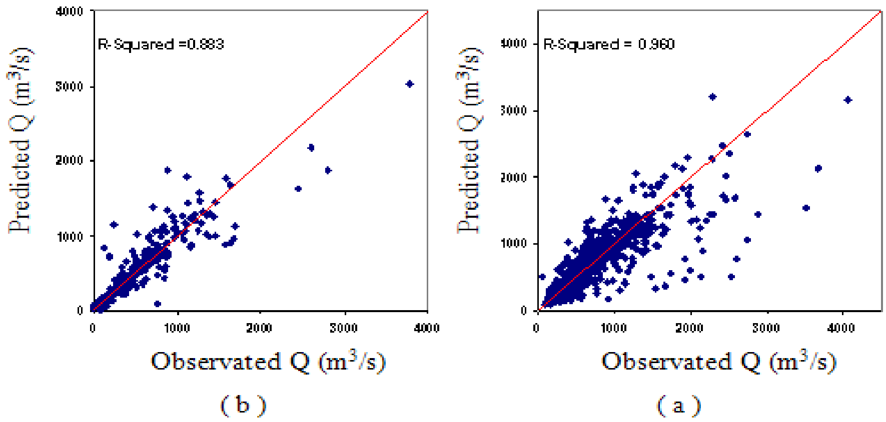


Fig. 4 Observed discharge versus predicted discharge for model 2: a) training data, b) testing data

4.1.3. Model 3

The second model uses precipitation data from 24 stations along the discharge from the previous time step of the output station as well as the precipitation index values for the last 4 to 8 days. Neural network with architecture 29-8-1 was trained using several trials and errors. Based on sensitivity analysis, 15 inputs were found as effective inputs for ANN model. Sensitivity analysis was accomplished by applying change in one normalized input while the other inputs remain constant. As an example, the result of the sensitivity analysis has been shown in Fig. 5. Accordingly, an initial MLP network at (29-8-1) was replaced by a network at (15-5-1). Values of MRE, E and R^2 were obtained to be equal to 13.07, 0.967 and 0.905 respectively. Fig.6 shows even better results than that obtained in the second model. However, none of these models have performed reasonably for predicting large flood flows. The reason may be attributed to the fact that these models use time step patterns for both rainy and no rainy days. These were separated in the next two models.

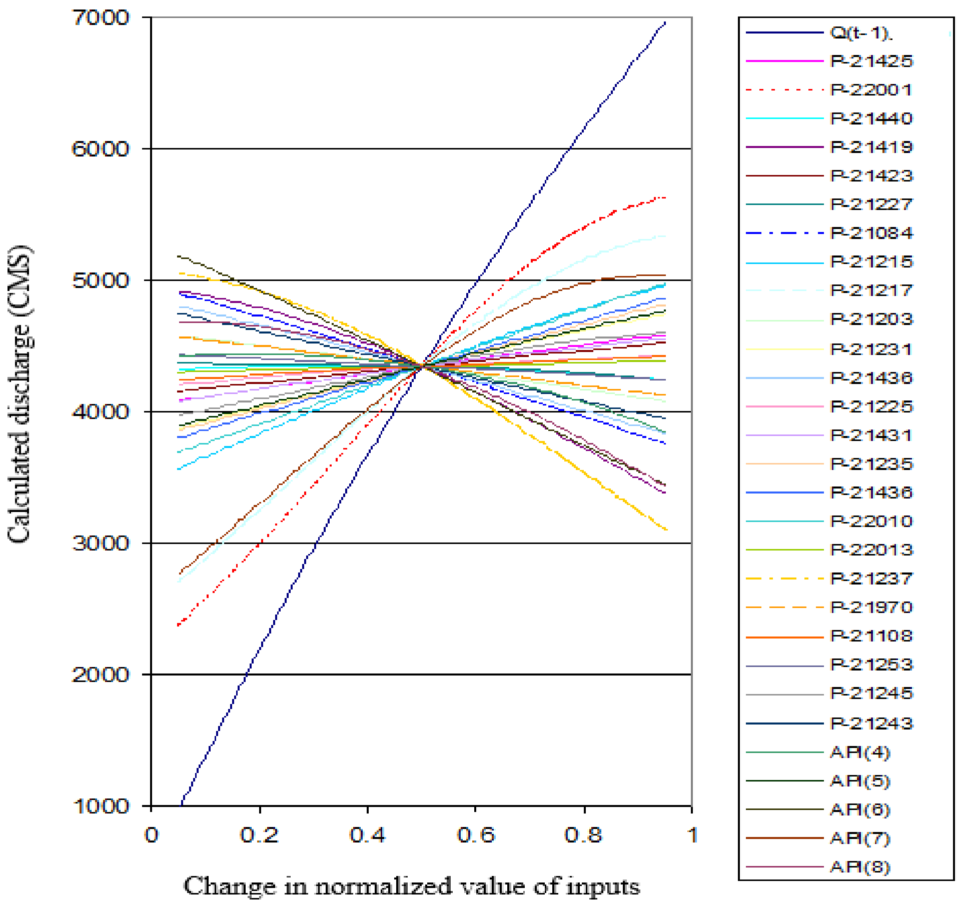


Fig. 5 Sensitivity analysis for Model 5 with initial 29 input variables

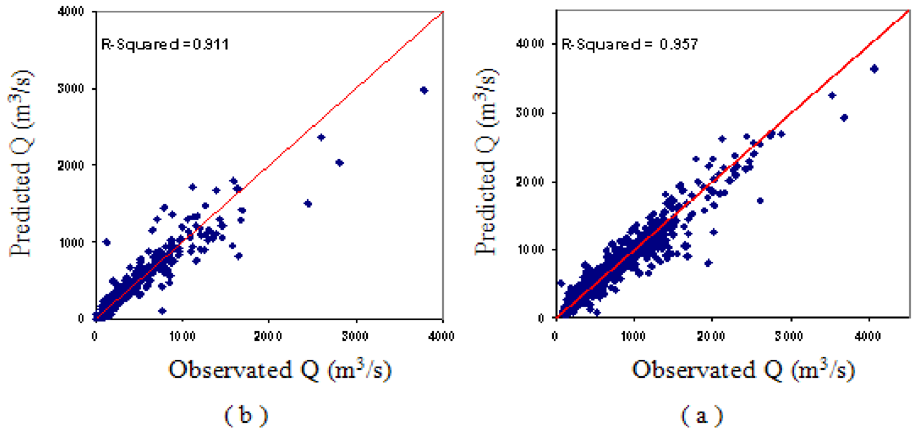


Fig 6. Observed discharge versus predicted discharge for model 3: a) training data, b) testing data

4.1.4. Model 4

Input data for flow on no rainy days include: $Q(t-1)$, $Q(t-2)$ and $API(7)$. Therefore in this model a MLP network with 3 inputs was used for training the discharge at time t. The architecture was based on (3-2-1). Values of MRE, E and R^2 were obtained to be equal to 5.78, 0.977 and 0.978 respectively (Fig.7). As shown in this figure, flood forecasting with model 4 is much more accurate than that with model 2. This implies that the large number of inputs does not necessarily improve forecasting.

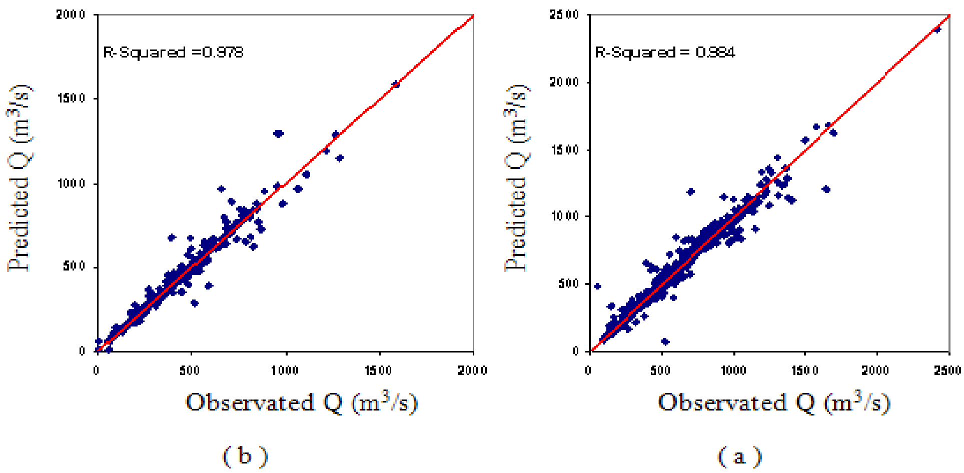


Fig. 7 Observed discharge versus predicted discharge for model 4: a) training data, b) testing data

4.1.5. Model 5

Inputs for the fifth model for rainy days were based on data from 13 stations, the outflow station discharge at last time step and the precipitation index for the last seven days. The network design was set at (15-9-1). Values of MRE, E and R^2 were obtained to be equal to 19.10, 0.957 and 0.859 respectively.

Figure 8 shows that omission of data with high correlation has reduced the correlation factor in this case. However, comparison of observed values of peak flood flows indicates higher accuracies than those obtained in model 3.

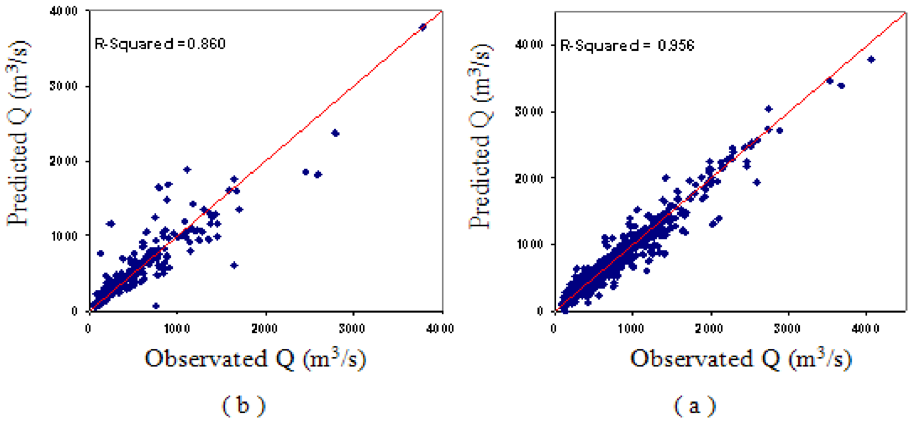


Fig. 8 Observed discharge versus predicted discharge for model 5: a) training data, b) testing data

4.1.6. Combination of Model 4 and Model 5

The combined model may be readily compared with any individual model. Values of MRE, E and R^2 were obtained to be equal to 9.10, 0.913 and 0.911 respectively. As may be seen in Fig. 9, this combined model can predict the flood hydrographs more accurately than the first three models. Flood predictions are made more accurately whilst both rainy and no rainy days have been considered. For more simple comparison, errors and R^2 have been listed in table 1.

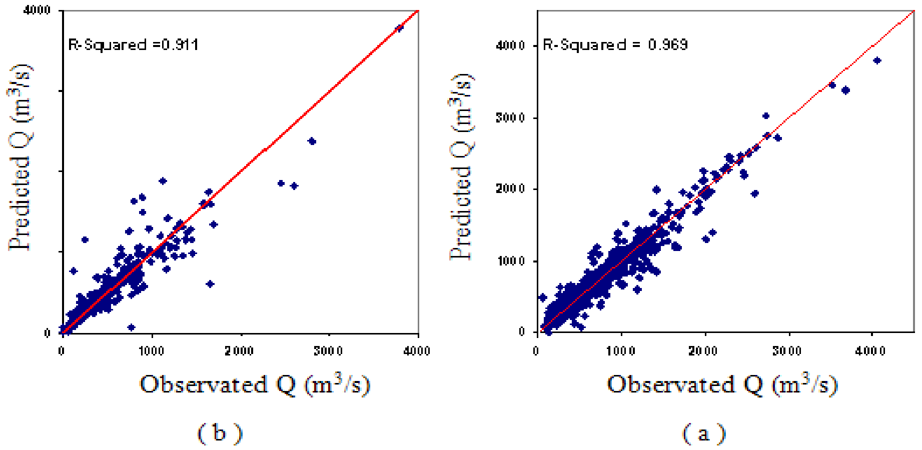


Fig. 9 Observed discharge versus predicted discharge for combined model: a) training data, b) testing data

Table 1. The value of errors for models

Model	R^2		E		MRE	
	Training	Test	Training	Test	Training	Test
1st	0.856	0.897	0.862	0.815	7.87	10.53
2nd	0.96	0.883	0.959	0.889	6.69	12.83
3rd	0.946	0.905	0.948	0.967	8.61	13.07
4th	0.984	0.978	0.999	0.977	2.92	5.78
5th	0.957	0.859	0.999	0.957	10.96	19.10
Combined model	0.969	0.911	0.971	0.913	5.76	9.10

4.2. forecasting discharge in time step *t+1*

Input variable considered for this ANN model, are discharges in rainy days, precipitation data form 24 stations, discharge in previous time step as well as API data. After several trials and errors, ANN model was set with 26 inputs with architecture 26-8-1 to forecast $Q(t+1)$. Sensitivity analysis showed $Q(t-1)$ was the most effective input parameters in developed ANN model. Results of the sensitivity analysis has been shown in Fig. 10. Accordingly, ANN model with architecture 15-6-1 was developed using several trial-and-error attempts. Values of MAE, E and R^2 were obtained to be equal to 20.83, 0.865 and 0.82 respectively. For graphical assessment, Fig. 11 shows predicted discharge by ANN model against observed discharge. The results show the appropriate performance of ANN model to predict large flood flows.

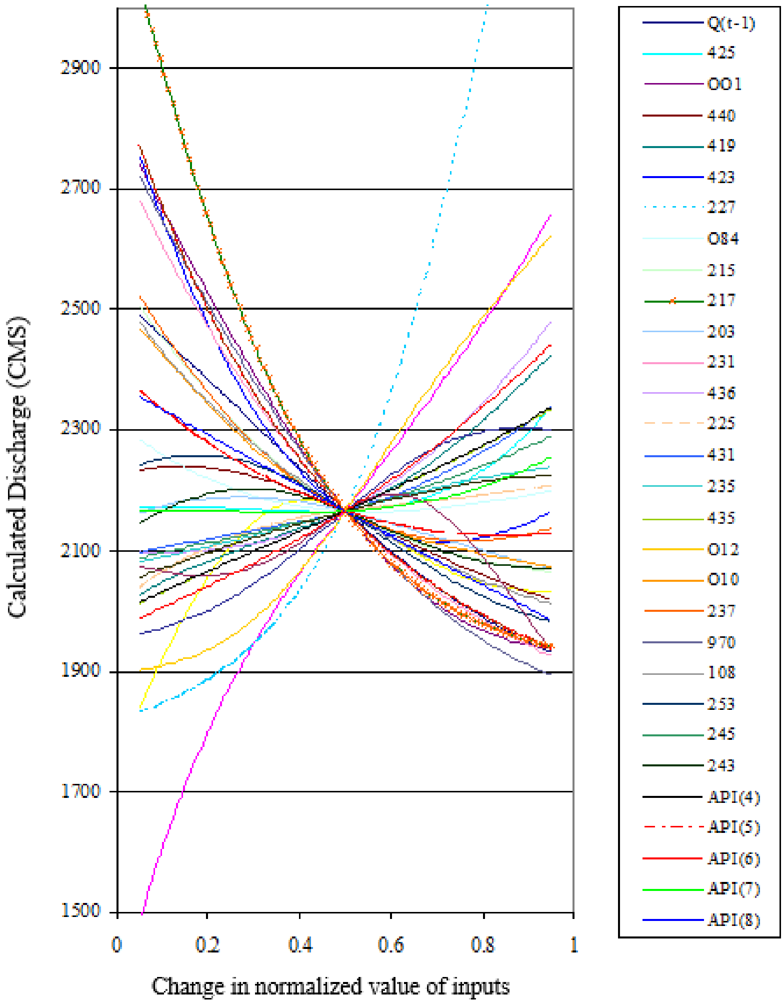


Fig. 10. Sensitivity analysis for ANN model with 26 input variables to forecast $Q(t+1)$

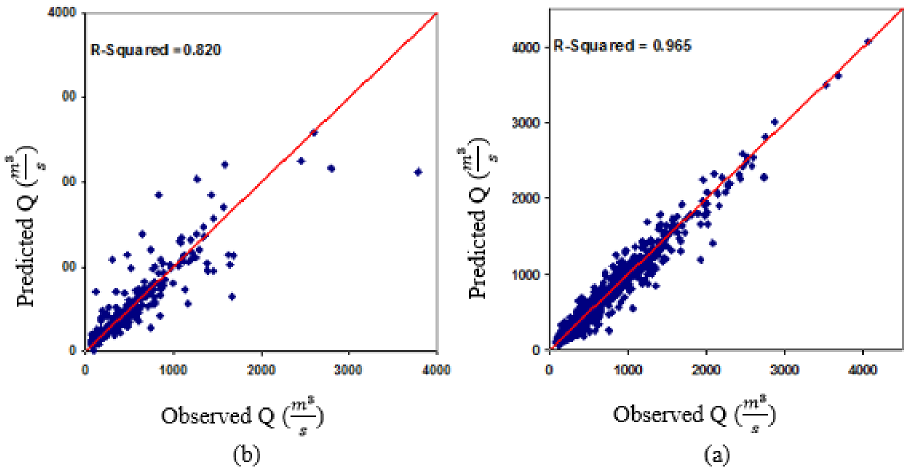


Fig. 11. Observed discharge versus predicted discharge for forecasting $Q(t+1)$: a) training data, b) testing data

4. Conclusion

Establishing an accurate flood warning system is vital option for timely flood management. Artificial neural network are typically considered as a suitable method to model physical phenomena such as riverine flood forecasting. However, determination of the effective inputs for ANN modelling is significant. In this research a neural network model and a data fusion technique were used to predict flood flows in the Karun River. Five ANN models were developed with different inputs combinations. Data for the fourth and fifth model were combined based on input variables for discharge, precipitation and the precipitation index for the previous seven days. In spite of variety of input data, the resulting model showed greater accuracy in predicting floods compared to models with less variety of input variables. Sensitivity tests were utilized to omit inputs with less effects on prediction of floods resulting in even greater accuracy.

Notations

X= vector of predictors
Y= forecast variable
ε =model error
n= number of observation data
$[Y]$= matrix of estimations of y provided by different individual models
net_j = the net input information arriving at node j
w_{ij} = the connection strength or weight between nodes i and j
o_i= the activation at node i
$o_{j=}$ the output at node j
P= the total number of patterns in the training set
K= the total number of output nodes,
d_{pk} and o_{pk} = the desired and calculated outputs at node j for pattern p
X_n = the normalized value of input X
X_{min}= the minimum of the input X
X_{max} = the maximum of the input X.
E = Nash-Sutcliffe parameter

MRE = mean relative error
R= correlation coefficient
Q_{obsi} = ith observed discharge
Q_{simi} = ith simulated discharge
n = the number of data.
Q = the discharge
P= the precipitation
T= the temperature
$API(x)$= the previous precipitation index at previous x time steps and
t = the computational time step
K = the precipitation constant varying between
P_{x-j} = the precipitation at day $x - j$
ET = the evapotranspiration
Wm= the maximum available soil moisture for evaporation.

References

1. Abart, R., SEE, L. (2002) Multi-model data fusion for river flow forecasting: an valuation of six alternative methods based on two contrasting catchment. Hydrology and Earth System Sciences 6(4): 655-670.
2. Araghinejad,S., Azmi, M., Kholghi, M., (2011) Application of artificial neural network ensembles in probabilistic hydrological forecasting. Journal of Hydrology 407(1-4): 94-104.
3. Chen, C., Chen, B., Chou, F., Yang, C., (2010) Development and application of a decision group Back-Propagation Neural Network for flood forecasting. Journal of Hydrology 385: 173–182.
4. Dasarathy, B. V. (1997) Sensor fusion potential explotation-innovative architectures and illustrative applications. Proceeding of the IEEE 85: 24-38.
5. Demeritt, D. et al., (2007) Ensemble predictions and perceptions of risk, uncertainty and error in flood forecasting. Environmental Hazards 7(2): 115.-127.
6. Emamgholizade, S., Kashi, H., Maroofpoor, I., Zalghi, E. (2014) Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. International Journal of Environmental Science and Technology 11(3): 645-656.
7. French, M. N., Krajewshi, W. F., and Cuendall, P. P., (1992) Rainfall Forecasting in Space and Time using Artificial Neural Networks, Journal of Hydrology 137: 1-31.
8. Hall, D. L. (1992) Mathematical technique in multi-sensor data fusion. Artech House, Boston,MA.
9. See, L., Abrahart, R. J. (2001) Multi-model data fusion for hydrological forecasting, Comput.. Geosci. 27: 987-994.
10. Shu, C., Burn, D. H., (2004) Artificial neural network ensembles and their application in pooled flood frequency analysis. Water Resour. Res. 40(9): W09301.
11. Sudheer, K.P., Jain, A., (2004) Explaining the internal behavior of artificial neural network river flow models. Hydrological Processes 18 (4): 833–844.
12. Van Berkom, F., Van de Watering, C., de Gooijer, K., Neher, A., (2007) Inventory of flood information systems in Europe – a study of available systems in Western-, Central- and Eastern Europe, INTERREG IIIC Network Flood Awareness and Prevention Policy in Border Areas' (FLAPP), The Netherlands.