Int. J. Environ. Sci. Technol. (2013) 10:1181–1192
DOI 10.1007/s13762-013-0209-0

ORIGINAL PAPER

Assessment of a conceptual hydrological model and artificial neural networks for daily outflows forecasting

M. Rezaeianzadeh · A. Stein · H. Tabari · H. Abghari · N. Jalalkamali · E. Z. Hosseinipour · V. P. Singh

Received: 16 May 2012 / Revised: 13 October 2012 / Accepted: 13 February 2013 / Published online: 29 March 2013

Abstract Artificial neural networks (ANNs) are used by hydrologists and engineers to forecast flows at the outlet of a watershed. They are employed in particular where hydrological data are limited. Despite these developments, practitioners still prefer conventional hydrological models. This study applied the standard conceptual HEC-HMS's soil moisture accounting (SMA) algorithm and the multi layer perceptron (MLP) for forecasting daily outflows at the outlet of Khosrow Shirin watershed in Iran. The MLP [optimized with the scaled conjugate gradient] used the logistic and tangent sigmoid activation functions resulting into 12 ANNs. The R^2 and RMSE values for the best trained MPLs using the tangent and logistic sigmoid transfer function were 0.87, 1.875 $m^3 s^{-1}$ and 0.81, 2.297 $m^3 s^{-1}$, respectively. The results showed that MLPs optimized with the tangent sigmoid predicted peak flows and annual flood volumes more accurately than the HEC-HMS model with the SMA algorithm, with R^2 and RMSE values equal to 0.87, 0.84 and 1.875 and 2.1 $m^3 s^{-1}$, respectively. Also, an MLP is easier to develop due to using a simple trial and error procedure. Practitioners of hydrologic modeling and flood flow forecasting may consider this study as an example of the capability of the ANN for real world flow forecasting.

Keywords Continuous rainfall–runoff · Multi layer perceptron · HMS SMA model · Activation functions · Khosrow Shirin watershed

Introduction

Hydrological models can be employed to predict flood magnitudes, flooding extent and associated water volumes. Rainfall–runoff (R-R) processes are modeled because hydrological measurements are limited, particularly for ungauged catchments. Hydrological models are often used to study various watershed management alternatives and real time forecasting in order to improve decision-making about the management of a watershed and addressing specific watershed issues. Among several hydrologic modeling approaches, continuous hydrologic simulation is

M. Rezaeianzadeh (✉)
School of Forestry and Wildlife Sciences, Auburn University, 602 Duncan Drive, Auburn, AL 36849, USA
e-mail: mzr0031@auburn.edu

A. Stein
Faculty of Geo-Information Science and Earth Observation (ITC), Twente University, Enschede, The Netherlands

H. Tabari
Department of Water Engineering, Ayatollah Amoli Branch, Islamic Azad University, Amol, Iran

H. Abghari
Faculty of Natural Resources, Urmia University, Urmia, Iran

N. Jalalkamali
Department of Water Engineering, Kerman Branch, Islamic Azad University, Kerman, Iran

E. Z. Hosseinipour
Advanced Planning Section, Ventura County Watershed Protection District, Ventura, CA, USA

V. P. Singh
Department of Biological and Agricultural Engineering, Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX 77843-2117, USA

preferred as it can model under dry and wet weather conditions (Beven 2001). Continuous hydrologic models, unlike event-based models, account for soil moisture balance of a watershed and are suitable for simulating daily, monthly, and seasonal stream flow (Ponce 1989).

The relationship between rainfall and runoff is a complex one due to the spatial and temporal variability of watershed characteristics and precipitation patterns, and the number of variables involved in the modeling of the physical processes (ASCE 2000; Kumar et al. 2005). Since the 1930s, numerous R-R models have been developed to forecast stream flow. Conceptual models provide daily, monthly, or seasonal estimates of stream flow for long-term forecasting on a continuous basis. The physical process in the hydrologic cycle is mathematically formulated in conceptual models that are composed of a large number of parameters. For example, the Stanford Watershed Model is defined by 20–30 parameters. Optimization of model parameters is usually done by a trial and error procedure because of their large number and their complex interaction (Tokar and Johnson 1999; Tokar and Markus 2000).

Artificial neural networks (ANNs) are models for a hydrological system mimicking the biological processes of the human brain. They consist of nodes, arrows and transfer functions and require only limited knowledge of internal functions of a system in order to recognize relationships between input and output (Mutlu et al. 2008). The most commonly applied transfer functions are sigmoidal type functions, such as the logistic and hyperbolic tangent functions (Maier and Dandy 2000).

Kisi (2004) applying ANNs in predicting mean monthly stream flow showed that ANNs provide a superior alternative to autoregressive models for developing input–output simulations and forecasting models. Kumar et al. (2005) showed that determining the input combinations involves finding the lags of rainfall/runoff that have a significant influence on the predicted flow. Kisi (2007) used a correlation analysis for understanding the effects of preceding flows, whereas Sudheer et al. (2002) suggested a statistical procedure, based on cross-correlation, autocorrelation and partial autocorrelation properties of the series for identifying the appropriate input vector for the model. Modarres (2009) showed that global statistics broadly reflect the accuracy of the model but are insufficient indicators for the best ANN. Ahmed and Sarma (2007) showed the optimal performance of an ANN for a synthetic stream flow series of the Pagladia River. Gopakumar et al. (2007) showed that ANNs can be adopted for forecasting river flows in the humid tropical river basins during the monsoonal period. In the other study, river flow forecasting and estimation using different ANN techniques was demonstrated by Kisi (2008).

Recently, a MLP network was optimized with the Levenberg–Marquardt algorithm to predict maximum daily discharge by Rezaeian Zadeh et al. (2009). Results showed that the application of area-weighted precipitation in place of single station precipitation increased forecasting efficiency. The area-weighted precipitation was also used as an input vector for the networks in another study (Rezaeian Zadeh et al. 2010). Despite the accumulated experience, it is still a challenging task to convince practicing engineers that ANNs are reliable and perform better than the traditional hydrological models in many practical problems. To this end there is still a need to carry out more comparative studies between ANNs and conventional hydrological models to build up the ANN portfolio.

This study used the widely used hydrological HEC-HMS (Hydrologic Engineering Center-Hydrologic Modeling System) model as a benchmark to assess the performance of ANNs. The HEC-HMS model, developed by the US Army Corps of Engineers, has a soil moisture accounting (SMA) routine. Rezaeian Zadeh et al. (2012) found that the SCG algorithm is superior to other training algorithms.

In several aspects the current work expands upon Rezaeian Zadeh et al. (2010): (1) The current study applied the scaled conjugate gradient (SCG) algorithm for MLP training, (2) the data were selected at random in the previous study, whereas here we used the data with their own time series and (3) the application of HMS SMA model was evaluated during this new work. The main purpose of this study was to apply artificial intelligence with the conceptual HEC SMA model in forecasting daily outflow. To do so, we explored advantages and disadvantages of black box and conceptual modeling in simulating the behavior of a basin in R-R transformation approach. The study was carried out in the Khosrow Shirin basin located in southwest Iran to simulate the daily R-R relationship. Altogether, 12 ANNs were trained and tested using the data from this watershed.

Materials and methods

An overview of the HMS SMA model

Conceptually, the HMS SMA model divides the potential path of rainfall onto a watershed into five zones (Fig. 1). Twelve parameters are needed to model the hydrologic processes of interception, surface depression storage, infiltration, soil storage, percolation, and groundwater storage. Those are the maximum depth of each storage zone, the percentage that each storage zone is filled at the beginning of a simulation, and the transfer rates, such as the maximum infiltration rate that are required to simulate

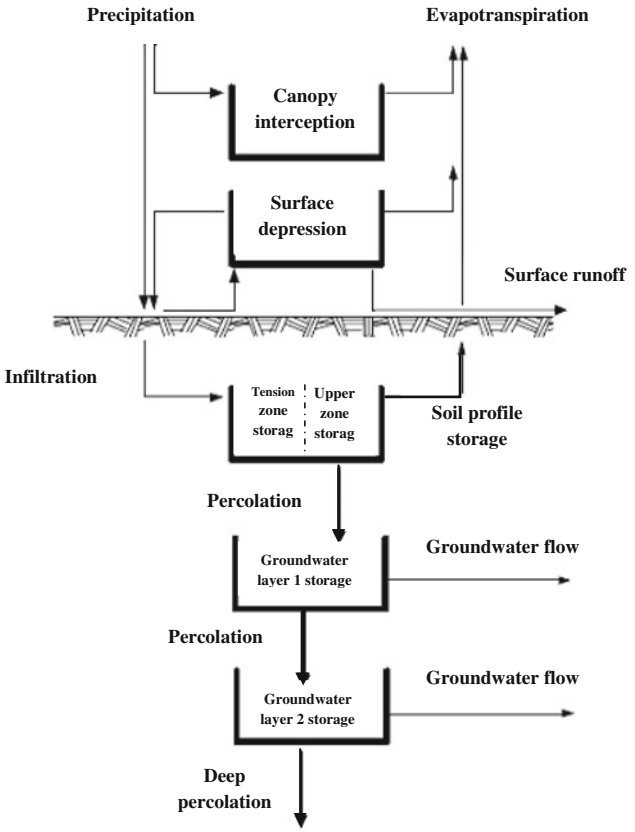

Fig. 1 Conceptual schematic of the continuous soil moisture accounting algorithm (Bennett 1998)

the movement of water through the storage zones. The only other input is the potential evapotranspiration rate. In this study, due to data scarcity for potential evapotranspiration calculations, the monthly values of pan evaporation were used. The pan coefficient was set to 0.7 for all the months. For further details on the HMS SMA model and how to calibrate it, see Fleming and Neary (2004) and HEC (2000).

Multi layer perceptron (MLP)

The MLP is a popular ANN in use today (Dawson and Wilby 1998). MLP is a network formed by simple neurons called perceptrons. A perceptron computes a single output from multiple real-valued inputs by forming a linear combination according to input weights and then possibly putting the output through a nonlinear transfer function. Mathematically this can be represented as:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \quad (1)$$

where, the w_i represent the weights, x_i is an input vector ($i = 1, 2, \ldots, n$), b is the bias, f is the transfer function and y is the output. The transfer function is often chosen to be the logistic sigmoid function defined for any variable S as:

$$f(s) = \frac{1}{(1 + e^{-s})} \quad (2)$$

As an alternative we will also apply the tangent sigmoid function in this study defined for any variable S as:

$$f(s) = \frac{2}{(1 + e^{-2s})} - 1 \quad (3)$$

Commonly applied transfer functions in ANNs are sigmoidal type functions, such as the logistic and hyperbolic tangent functions (Haykin 1999; Maier and Dandy 2000). In a recent study, Rezaeian Zadeh et al. (2010) showed that the tangent sigmoid activation function performed better than the logistic sigmoid activation function in daily outflow prediction when the data was randomly selected. Yonaba et al. (2010) endorsed the priority of tangent sigmoid than others. These two studies indicated the importance of appropriate activation function. In this study, we used time series (not randomized) data to evaluate the priority of these two transfer functions.

Neural network training

Training of a single-layer neural network is relatively simple. Common methods to do so are supervised learning (Sethi 1990), Boltzmann learning (Ackley et al. 1985), counter-propagation (Hecht-Nielsen 1987), and Madaline Rule-II (Widrow et al. 1988). A popular method is supervised learning that provides the network with examples of input–output mapping pairs from which the network learns. The learning is reflected through the modification of connection strengths or weights. This process of learning occurs continuously until the mapping present in the examples is achieved. A key issue in the learning process credit assignment, i.e. identification of what should be the desired output of the neurons in the hidden layers during training. This issue has been addressed by propagating back the error in the output layer to the internal layers, also called the back propagation algorithm (Rumelhart et al. 1986). It minimizes the error at the output layer and is a gradient descent method.

Neural networks training algorithms

The SCG algorithm was carried out by minimizing the global error E defined as:

$$E = \frac{1}{p}\sum_{p=1}^{p} E_p \quad (4)$$

where P is the total number of training patterns and E_P is the error for the training pattern p. E_P is calculated as:

$$E_p = \frac{1}{2}\sum_{k=1}^{n}(o_k - t_k)^2 \quad (5)$$

where n is the total number of output nodes; o_k is the network output at the kth output node; and t_k is the target output at the kth output node. In this study, the global error is reduced by adjusting weights and biases as described by Kisi (2007).

Scaled conjugate gradient algorithm

The basic back propagation algorithm adjusts the weights in the steepest descent direction, i.e. in the negative gradient direction, being the direction in which the performance function decreases most rapidly. It turns out that this direction is not necessarily that of the fastest convergence. In a conjugate gradient algorithm a search is performed along conjugate directions, which generally produces faster convergence than steepest descent directions.

All conjugate gradient algorithms start out by searching in the steepest descent direction (negative of the gradient) on the first iteration:

$$P_0 = -g_0 \quad (6)$$

A line search is then performed to determine the optimal distance to move along the current search direction:

$$X_{k+1} = X_k + \alpha_k \times P_k \quad (7)$$

where X_k is a vector of current weights and biases, g_k is the current gradient, and α_k is the learning rate and $P_1, \ldots, P_k$ a set of non-zero weight vectors.

Then the next search direction is determined so that it is conjugate to previous search directions. The general procedure for determining the new search direction is to combine the new steepest descent direction with the previous search direction:

$$P_k = -g_k + \beta_k \times P_{k-1} \quad (8)$$

The various versions of conjugate gradient are distinguished by the manner in which the constant β_k is computed. As a further refinement, the step size is adjusted at each iteration. A search is made along the conjugate gradient direction to determine the step size, which minimizes the performance function along that line.

In this study we use the SCG algorithm, developed by Moller (1993). It has as an advantage that it avoids the time-consuming line search and that it combines the model-trust region approach, with the conjugate gradient approach.

The target error for the training of networks was equal to 10^{-4}. With the specification of this target error, the number of iterations for all 12 models (different combinations of MLPs based on the correlation analysis obtained from Rezaeian Zadeh et al. 2010) was placed at 1,000. The training of networks was stopped when their performances reached the target error. Hence, the number of iterations for all models was equal to 1,000. Based on Kisi (2007), the same number of iterations can be used for networks training if the same training algorithm is used.

After examination of many (50, 100, 150, 200, …, 2,500) epochs and neuron numbers, the best results were seen at this network architecture (Table 1).

Before applying the ANNs training, all data must be normalized, i.e. they are transformed into the range of (0.05, 0.95) by applying.

$$X_n = 0.05 + 0.9\frac{X_r - X_{\min}}{X_{\max} - X_{\min}} \quad (9)$$

where X_n and X_r are normalized and original input and $X_{\min}$ and $X_{\max}$ are the minimum and maximum of the input range, respectively (Soroosh et al. 2005; Rezaeian Zadeh et al. 2010).

These normalized data were used to train each of the 12 ANN models. Program codes, including Neural Network Toolbox, were written in MATLAB language for the ANN simulations. In this study, the following combination of input data for precipitation and discharge was implemented (Table 2).

The case study and relevant data

The Khosrow Shirin watershed within the Fars Province in Iran, located between (30°37′N, 51°49′E) and (30°59′N, 52°12′E was selected as a case study catchment (Fig. 2). The drainage area at this site is 610 km^2. The average annual precipitation and temperature are equal to 530 mm and 5.5 °C, respectively. Hence, the region belongs to humid climatological class based on the De Martonne (1926) aridity index.

The time series of daily precipitation data from four stations (Khosrow Shirin, Sedeh, Dehkadeh Sefid and Dozd Kord) was used for this study. The time series of the stream flow of Dehkadeh Sefid station (30°39′N, 52°07′E) is presented in Fig. 3.

The average daily precipitation on the watershed area was determined with Thiessen polygon. It calculates the average precipitation over the catchment by weighing each station's rainfall depth in proportion to its area of influence. Hence, the weights belonging to the Khosrow Shirin, Dozd

Table 1 Architecture of network

	Layer		
	Input	Hidden	Output
No. of neuron	5	7	1
Transfer functions	Logsig Tansig	Logsig Tansig	Purelin

Table 2 Different combinations of MLPs based on correlation analysis to examine and find the best inputs to the networks

Model no	Input combinations
MLP1	$P(t)$
MLP2	$P(t)$, $P(t-1)$
MLP3	$P(t)$, $Q(t-1)$
MLP4	$P(t)$, $Q(t-1)$, $Q(t-2)$
MLP5	$P(t)$, $P(t-1)$, $P(t-2)$, $Q(t-1)$
MLP6	$P(t)$, $P(t-1)$, $P(t-2)$, $Q(t-2)$
MLP7	$P(t)$, $P(t-1)$, $Q(t-1)$
MLP8	$P(t)$, $P(t-1)$, $Q(t-1)$, $Q(t-2)$
MLP9	$P(t)$, $P(t-1)$, $Q(t-2)$
MLP10	$P(t)$, $P(t-1)$, $P(t-2)$, $Q(t-1)$, $Q(t-2)$
MLP11	$\mathbf{P(t)}$, $\mathbf{P(t-2)}$, $\mathbf{P(t-3)}$, $\mathbf{Q(t-1)}$
MLP12	$P(t)$, $P(t-2)$, $P(t-3)$, $P(t-4)$, $Q(t-1)$

P precipitation, Q river flow, t time

Kord, Dehkadeh Sefid and Sedeh precipitation stations were 0.68, 0.16, 0.08 and 0.08, respectively.

These data are used as input into the network (Rezaeian Zadeh et al. 2010). Precipitation from 2002 to 2007 water years were used, where the 2002–2006 water years data (in total 1,460 daily values: equal to 80 % of the record) were used for the training of MLPs and for calibration of the HMS SMA model. The 2006–2007 water year data were used for testing (365 daily values: equal to 20 % of data) of MLPs and validation of the calibrated HMS SMA model. It should be noted that in the previous study, data were randomly selected. For comparison of the results of HMS SMA and MLPs the same calibration and validation periods for the two methods were selected. The 1,460 training patterns in this study generate 1 by 1,460 matrices for both the rainfall and the river flow patterns. The rainfall and river flow patterns for the 2002–2003 water year are shown in Fig. 3.

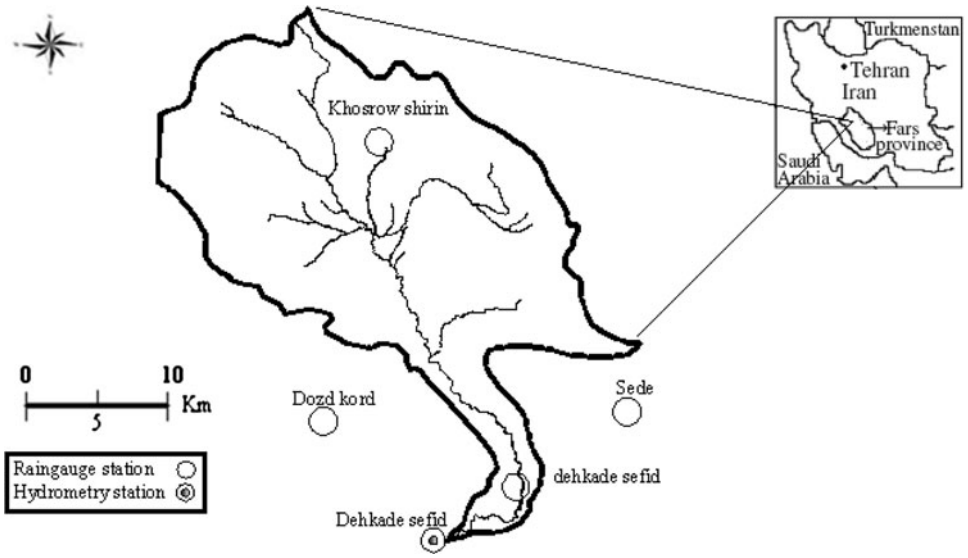

Fig. 2 Map of the location of the study site

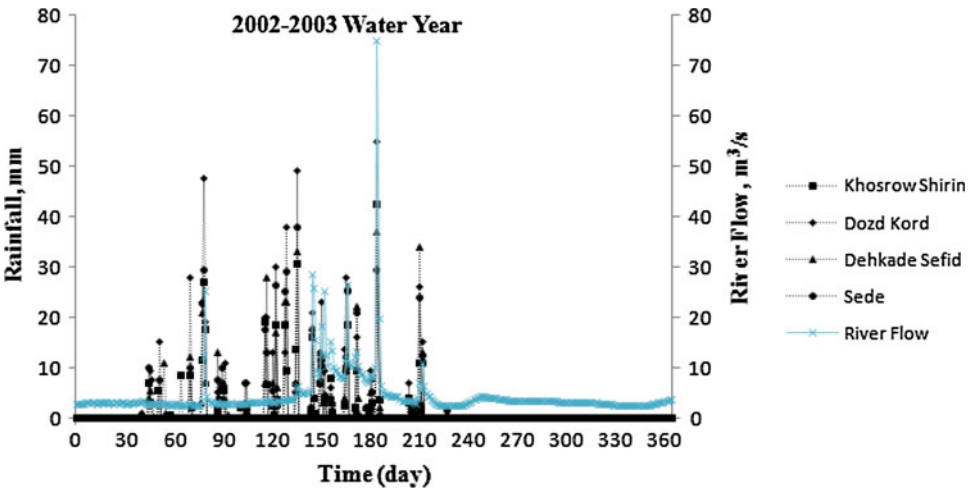

Fig. 3 Rainfall and river flow patterns of the 2002–2003 water year

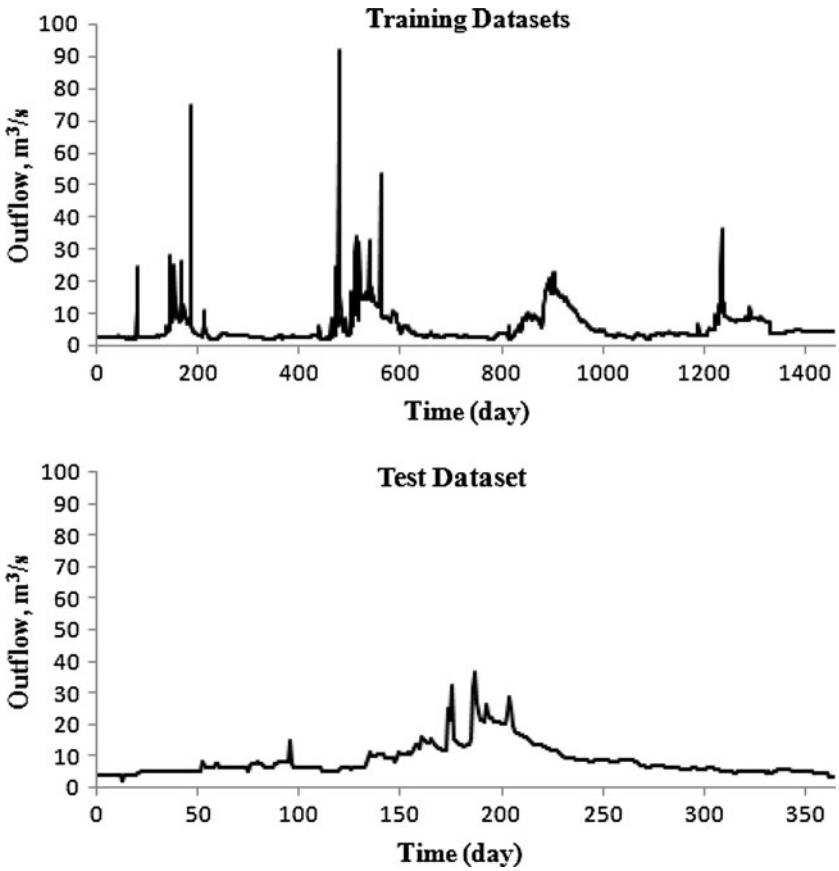

Fig. 4 Training and testing dataset

We let the proportions of the training and test data depend upon the coefficient of variation (CV) of the data. The proportion of training data versus full data could vary between 60 and 80 % and hence from 40 to 20 % for the testing data. Increase in training percentage, however, may result in better forecasts during the testing phase.

To do so, Levene's test and the *t* test were applied for checking the distribution similarity of different flow regimes for both training and test data sets (Rezaeian Zadeh et al. 2010). The equality of two population variances, tested using Levene's test, showed at the 95 % confidence level a *p* value equal to 0.192 and hence the hypothesis of equal variance was not rejected. The *t* test under the assumption of equal variances showed a *p* value equal to 0.001 and thus indicates a significant difference between the training and test datasets. It thus shows that different flow regimes (high, mean and low flows) were not selected both in training and test datasets (Fig. 4) but due to ANN flexibility, the results of forecasting are very satisfactory.

The values of the mean, maximum and minimum annual flow and the annual flow volume of the Dehkadeh Sefid hydrometric station are shown in Table 3. Table 3 shows that the 2006–2007 water year with the mean annual flow equal to 8.45 m^3 s^{-1} has the highest value among these 5-year of data. This water year also has the highest value of the annual flow volume equal to 266.5 MCM.

Table 3 Values of the mean, daily maxima/minima during annual periods of flow and annual flow volume at the Dehkadeh Sefid hydrometric station

Years	Mean annual flow (m^3/s)	Daily maxima during annual periods (m^3/s)	Daily minima during annual periods (m^3/s)	Annual flow volume (MCM)
2002–2003	4.43	74.9	2.24	139.63
2003–2004	6.91	92.1	2.33	217.9
2004–2005	6.66	23.2	2.37	210.0
2005–2006	5.82	36.6	3.17	183.4
2006–2007	8.45	36.5	1.98	266.5

Results and discussion

Model calibration is the process of adjustment of model parameter values until the simulated results of the model match the historical data. In the case of rainfall–runoff models, it is necessary to have observed flow data to be able to develop this process (García et al. 2008). Due to gauge scarcity of Khosrow Shirin watershed, 7 of the 12 parameters needed for the SMA algorithm [canopy interception storage, surface depression storage, maximum infiltration rate, maximum soil storage, tension zone storage, soil zone infiltration rate, and groundwater 1 (GW1) percolation rate] were estimated by means of model

calibration. A sensitivity analysis showed that GW1 percolation rate and Clark storage coefficient had the highest effect on peak flows, and GW2 percolation rate, surface depression storage capacity and impervious area had highest effects on annual flood volume (Table 4).

Four parameters (GW1 and GW2 storage depths and storage coefficients) were estimated by stream flow recession analysis of historic stream flow measurements (Linsley et al. 1958; Burnash et al. 1973; Leavesley et al. 1983).

Table 4 Calibration parameter constraints of HMS SMA

Parameters	Minimum	Maximum
Canopy capacity (mm)	0.01	1,500
Canopy initial storage percentage (%)	0.001	100
Clark storage coefficient (h)	0.01	1,000
Clark time of concentration (h)	0	1,000
GW1 capacity (mm)	0.01	1,500
GW1 initial storage percentage (%)	0.001	100
GW1 percolation rate (mm/h)	0.01	500
GW1 storage coefficient (h)	0.01	10,000
GW2 capacity (mm)	0.01	1,500
GW2 initial storage percentage (%)	0.001	100
GW2 percolation rate (mm/h)	0.01	500
GW2 storage coefficient (hr)	0.01	10,000
Linear reservoir GW1 coefficient (hr)	0	10,000
Linear reservoir GW1 steps ()	1	100
Linear reservoir GW2 coefficient (hr)	0	10,000
Linear reservoir GW2 steps ()	1	100
Soil capacity (mm)	0.01	1,500
Soil infiltration rate (mm/h)	0.01	500
Soil initial storage percentage (%)	0.001	100
Soil percolation rate (mm/h)	0.01	500
Surface capacity (mm)	0.01	1,500
Surface initial storage percentage (%)	0.001	100
Tension zone capacity (mm)	0.01	1,500

The GW2 percolation rate was estimated as well by model calibration. The recession curve is described by Eq. (10),

$$q_t = q_o \times K_r^t \tag{10}$$

where q_t is the average daily flow at a future time with respect to the initial flow q_o, K_r is the recession constant and t is the time (days). There are several recession periods during the surveyed years. Figure 5 shows the daily flows from 2/20/2005 to 8/20/2005 for one of considerable recession periods. The recession constants for the first (left) and second (right) parts of hydrograph and the mean value were found to be equal to 0.983 and 0.993, respectively, with the mean of 0.988.

Based on the field information summarized in Table 5, the slope of the watershed is equal to 13.8 % and the watershed is thus classified as moderately to gently sloping. A slope value of 6.4 was selected as the initial value for estimation of the surface storage parameter. The final calibrated value of this parameter was equal to 10.0.

A correlation analysis of the time series was employed to evaluate the effect of antecedent precipitation and flow by Rezaeian Zadeh et al. (2010). The auto- and partial autocorrelation statistics and the corresponding 95 % confidence bands from lag 0 to lag 10 days were estimated for precipitation and stream flow data.

The results obtained by the partial autocorrelation function indicate a significant correlation up to lag 3 for precipitation and up to lag 1 for the flow series data,

Table 5 Surface depression storage (surface storage values from Bennett 1998)

Description	Slope (%)	Surface storage (mm)
Paved impervious areas	NA	3.2-6.4
Steep, smooth slopes	>30	1.0
Moderate to gentle slopes	5–30	6.4–12.7
Flat, furrowed land	0–5	50.8

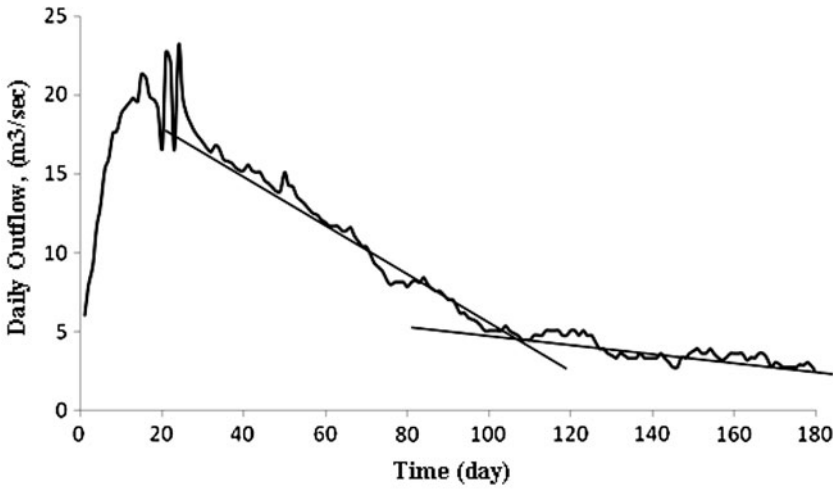

Fig. 5 Daily outflow hydrograph from 2/20/2005 to 8/20/2005

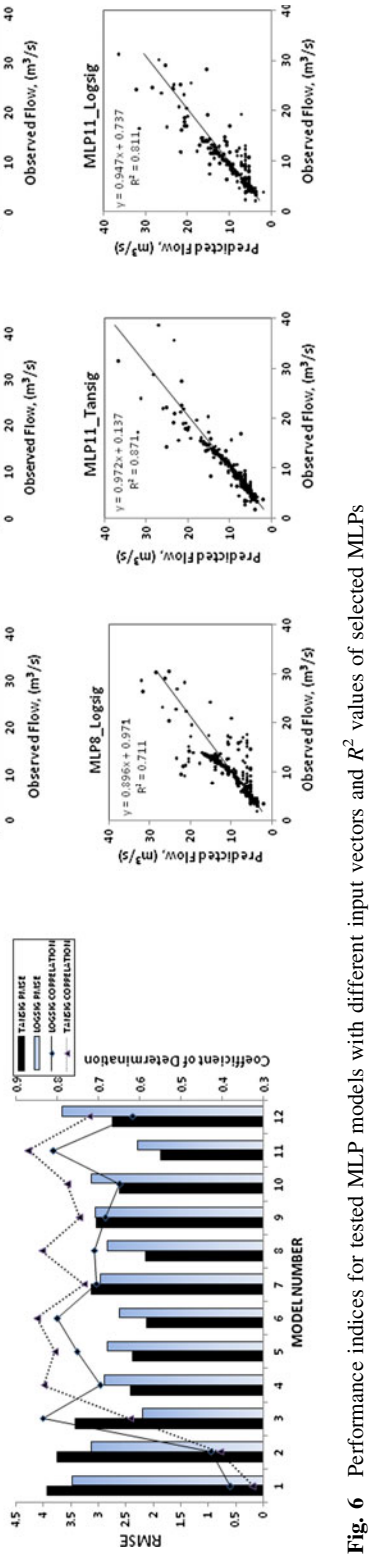

Fig. 6 Performance indices for tested MLP models with different input vectors and R^2 values of selected MLPs

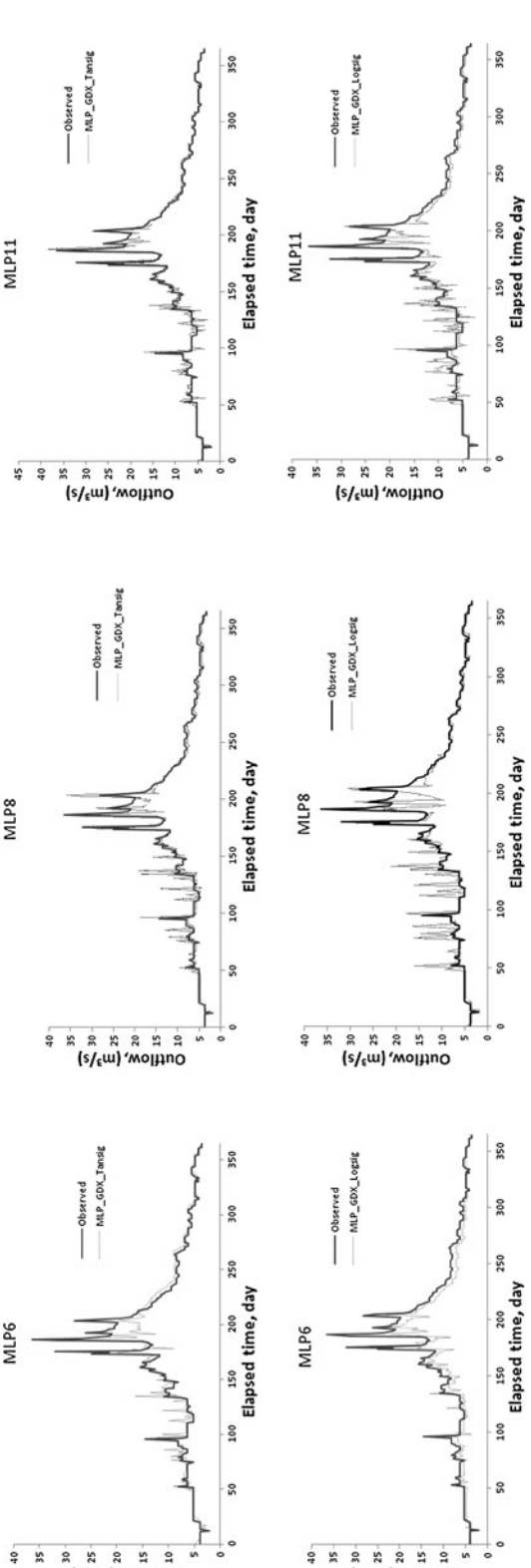

Fig. 7 Observed versus predicted hydrographs of MLP6, MLP8 and MLP11 models with different transfer functions for the validation period

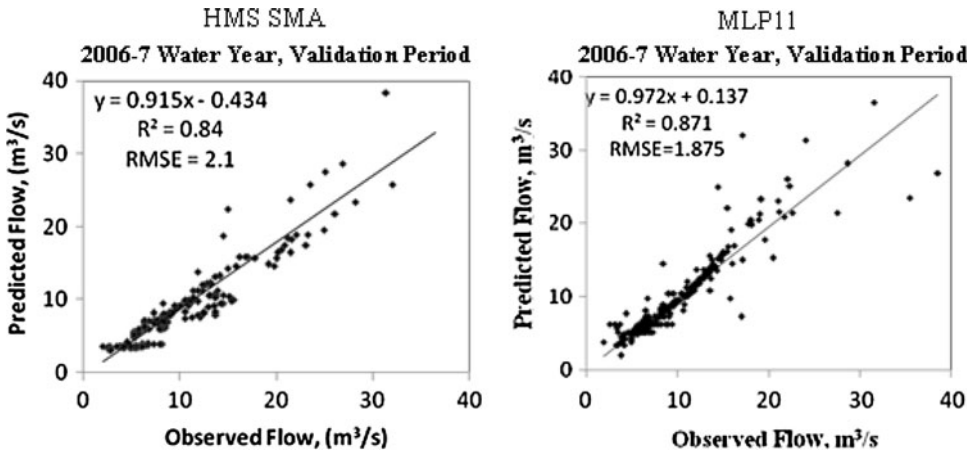

Fig. 8 Scatter plots of HMS SMA and MLP11 optimized with tangent sigmoid transfer function for the validation period

Table 6 Error in peak flow (%) for validation period

Model no.	Tangent sigmoid	Logistic sigmoid
3	6.93	27.71
4	17.37	45.66
5	5.2	3.27
6	25.58	22.5
7	34.7	32.53
8	1.26	16.05
9	46.3	36.37
10	15.53	15.2
11	5.68	14.2
12	27.81	28.17

respectively, that drops subsequently to within the confidence limits. According to the correlation analysis, three and one antecedent precipitation and flow values should be selected for the ANN input, respectively. For more confidence in the selection of all appropriate inputs to ANNs, three antecedent precipitation and flow were selected. Note that in Rezaeian Zadeh et al. (2010), five different input combinations were constructed during model's development whereas in the current study we examined twelve input combinations (12 MLPs).

For all MLPs, three layered ANN models were trained using the scaled conjugate gradient (MLP_SCG) training algorithm. The training of the MLPs was performed by two transfer functions. The ANN results were transformed back to the original domain and the root mean square errors (RMSE) were computed for the test data for each MLP (12 models). The performance of each of these combinations in terms of the RMSE and coefficient of determination (R^2) is shown in Fig. 6. Model performance improves with the addition of more antecedent rainfall and discharge information. Figure 6 shows that models MLP6, MLP8 and MLP11 had the best result for daily flow forecasting.

As can be seen from Fig. 6, the models with tangent sigmoid transfer function have a lower RMSE, except models 1, 2, 3 and 7. The MLP1 gives the highest RMSE. Different numbers of hidden layer neurons were tried for the MLP_SCG in the study. A trial and error procedure showed that the number of neurons in hidden layer that gave the minimum RMSE was 7. The lowest RMSE occurs for the networks with an input vector containing two and three antecedent precipitation and one antecedent flow values (MLP6, MLP11). This result is in direct agreement with the input vector selected based on the correlation analysis. Also, it can be seen that the values of RMSE for about 60 % of the models with tangent sigmoid transfer function are lower than the models with logistic sigmoid transfer function.

Figure 6 shows that the correlation of tangent and logistic sigmoid transfer functions is close to each other for MLP6 and MLP11 but the correlation of MLP8 with these two activation functions is not close to each other. This difference may due to adding one more discharge vector as input combination of MLP8 than the other MLPs. Hence, an increase in input data was not always useful.

Comparison of the three ANN models with two different transfer functions is shown in Figs. 6 and 7.

The R^2 values of MLP6, MLP8, and MLP11 with the tangent and logistic sigmoid activation functions are also shown in Fig. 6. The values of tangent sigmoid transfer function are higher than the related values of logistic sigmoid transfer function. The values of R^2 for MLP6, 8 and 11 for tangent sigmoid transfer function are equal to 0.85, 0.83 and 0.87, respectively. Also, RMSE values for these MLPs with tangent and logistic sigmoid transfer functions are equal to 2.123 and 2.631, 2.148 and 2.850, 1.875 and 2.297 $m^3\ s^{-1}$, respectively. Clearly, the results of MLPs using tangent sigmoid activation function have a higher forecasting efficiency than the MLPs using logistic sigmoid activation function. These metrics are in good agreement with the achieved results by Rezaeian Zadeh et al. (2010).

The correlation between the observed flow and simulated flow for ANNs with tangent sigmoid activation

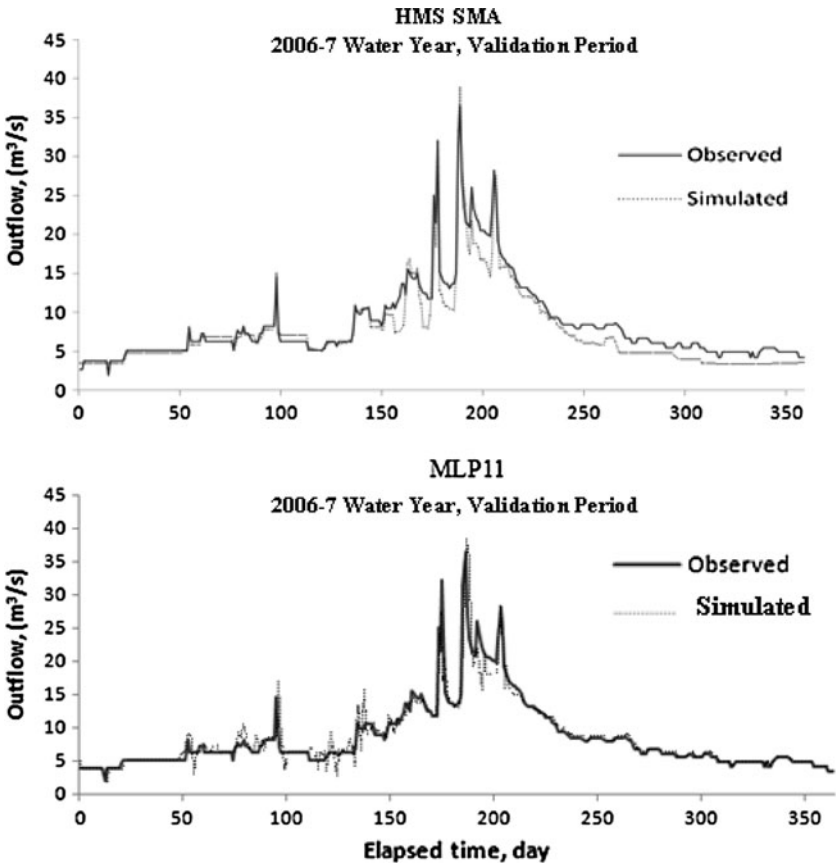

Fig. 9 Observed versus predicted hydrographs of HMS SMA model and MLP11 for the validation period

function is higher than for ANNs with logistic sigmoid activation function. The highest correlation (0.87) is obtained for Model 11. The RMSE value for MLP11 was equal to 1.875 $m^3\ s^{-1}$, so that this value is lower than the RMSE value of the other MLPs. For MLP6, the correlations of two different activation functions are close. Similarly, this proximity is seen at MLP11. This combination of input at MLP11 ($P(t)$, $P(t\text{-}2)$, $P(t\text{-}3)$, $Q(t\text{-}1)$) shows that antecedent precipitation up to 3 days can be very useful for daily flow forecasting. Figure 8 in MLP11 attests to this observation. These findings also validate the results that were reported by Rezaeian Zadeh et al. (2010).

Figure 7 shows that MLP8 with tangent sigmoid activation function predicts peak flows at various times but with higher values than the observed flow. The MLP8 with a logistic sigmoid activation function, however, does not produce a good forecasting. The analysis of the percentage error in peak flow estimation in Table 6 indicates that all models except MLP3, 5, 8 and 11 tend to underestimate the peak. Table 6 shows that MLP8 with tangent sigmoid transfer function has the best forecasting of peak flow as Fig. 7 corroborates this observation.

Figure 7 shows that MLP11 with tangent sigmoid activation function predicts daily flows that are close (error = 5.7 in the peak flow, $R^2 = 0.871$ and $RMSE = 1.875\ m^3\ s^{-1}$) to the observed flows. For all models that have produced better forecasting, antecedent discharge with 1 day time lag was used and it can guide us in building efficient models in the future. Comparing MLP6 and MLP11 performance, it can be seen that increase in input data for a model is not always applicable and useful. In MLP11 with four input data, we gained the best result, whereas MLP6 has five input data series and the forecasting is not as good as MLP11.

The performance metrics of HMS SMA and MLP11 in terms of the root mean square error (*RMSE)* and coefficient of determination (R^2) are shown in Fig. 8.

Figure 8 shows that the values of RMSE for MLP11 are lower than those of the HMS SMA model, whereas the correlation between the observed flow and simulated flow was higher. The correlations for MLP11 and HMS SMA were equal to 0.87 and 0.84, respectively, whereas the RMSE values were 1.875 and 2.1 $m^3\ s^{-1}$, respectively. Clearly MLP11 performed better than HMS SMA and had higher forecasting efficiency. Plots of HMS SMA and MLP11 for flow forecasting for the validation period of 2006–2007 water years are also shown in Fig. 9. Comparison of HMS SMA and MLP11 simulations attests to this conclusion.

The observed and simulated monthly flow volumes for the validation period using HMS SMA and MLP11 are

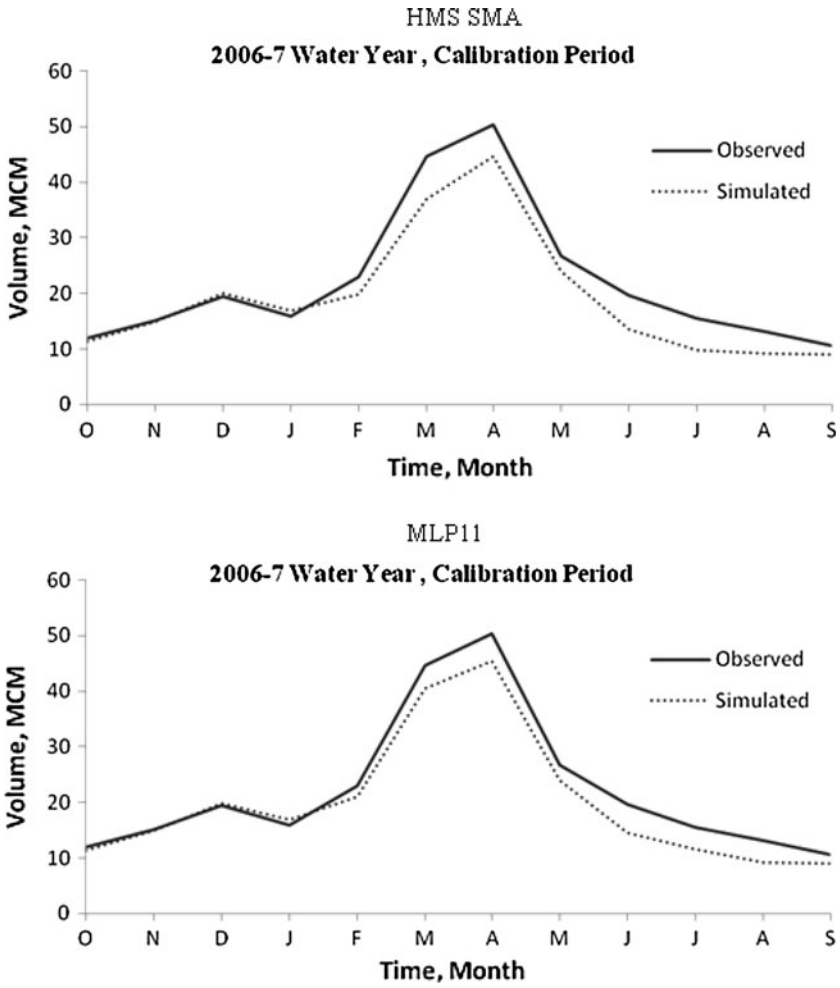

Fig. 10 Observed versus simulated monthly flood volume using HMS SMA and MLP11 for the validation period

shown in Fig. 10. Figure 10 shows that MLP11 can predict monthly and annual flood volumes better than the HMS SMA model. The R^2 values for HMS SMA and MLP11 for annual flood volume forecasting were equal to 0.86 and 0.93, respectively. The RMSE values for HMS SMA model and MLP11 for annual flood volume forecasting were equal to 4.1 and 3.0 $m^3\ s^{-1}$. Overestimation of HMS SMA was less than MLP11. This overestimation may be due to uncertainties about forecasting by MLPs. Overall comparison of performances for these two models indicate that MLP11 is a better tool for continuous hydrologic modeling.

The effective number of parameters for such a MLP network is the sum of the number of feedforward connections (including the bias connections). This may result in a higher number of parameters for MLPs than for the HMS SMA model. But, there are only a few parameters to be assigned a value or function; e.g. number of neurons in each layer and transfer functions. Other parameters are only vaguely assigned a value by using a simple trial and error procedure during MLP training.

Forecasts by the HMS SMA model are reliable as well. One weakness of the HMS SMA is that the additional efforts required for sensitivity analysis of parameters are resulting in several times longer in time and higher in costs to perform the work.

Conclusion

The ability of the HMS SMA model and ANNs with focus on the application of two activation functions was investigated in daily flow modeling. The data of Khosrow Shirin watershed in Iran used as a case study show that there was significant difference between the training and test datasets. This is in contrast to a recent study at the same catchment (Rezaeian Zadeh et al. 2010) where randomized training and testing sets with close similarity were explored. Although the training and test datasets are quite different, due to ANN's flexibility, the results of forecasts are still satisfactory especially for MLP11. As expected, the results showed that the addition of previous daily flows as input vector provides more satisfactory results in daily flow modeling. Finally, the predicted outflow showed that the tangent sigmoid activation function performed better than

did the logistic sigmoid activation function. The results indicate that MLP can predict daily flow with higher forecasting efficiency than the conventional hydrological model such as HMS SMA. Comparison of annual flood volumes predicted by the two models also verifies the superiority of MLP. Application of MLP should be considered when there is not enough time and resources for sensitivity analysis and other procedures that are needed for the HMS SMA applications. Also, MLP is much simpler to apply than the HMS SMA by using a simple trial and error procedure. This study will help the practitioners to gain valuable knowledge about the ANN over the more widely used conceptual hydrological models.

Acknowledgments The data used to carry out this research were provided by the Islamic Republic of Iran Meteorological Office (IRIMO) and Surface Water Office of Fars Regional Water Affair. The first author would like to especially thank Dr. Amir Agha Kouchak and Prof. Dawei Han for their inestimable help.

References

Ackley DH, Hinton GE, Sejnowski TJ (1985) A learning algorithm for Boltzmann machines. Cog Sci 9:147–169

Ahmed JA, Sarma AK (2007) Artificial neural network model for synthetic stream flow generation. Water Resour Manag 21:1015–1029

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000) Artificial neural networks in hydrology. I: preliminary concepts. ASCE J Hydrol Eng 5(2):115–123

Bennett T (1998) Development and application of a continuous soil moisture accounting algorithm for the Hydrologic Engineering Center Hydrologic Modeling System (HEC-HMS). MS thesis, Department of Civil and Environmental Engineering, University of California, Davis

Beven KJ (2001) Rainfall-runoff modelling. The Primer. Wiley, Chichester

Burnash RJC, Ferral RL, McGuire RA (1973) A generalized streamflow simulation system: Conceptual modeling for digital computers. United States Department of Commerce, National Weather Service, and State of California, Department of Water Resources, Sacramento

Dawson CW, Wilby R (1998) An artificial neural network approach to rainfall–runoff modeling. Hydrol Sci J 43(1):47–66

De Martonne E (1926) Une nouvelle fonction climatologique: L'indice d'aridité. *La Météorologie*, p 449–458

Fleming M, Neary M (2004) Continuous hydrologic modeling study with the hydrologic modeling system. ASCE J Hydrol Eng 9(3):175–183

García A, Sainz A, Revillaa JA, Álvareza C, Juanesa JA, Puentea A (2008) Surface water resources assessment in scarcely gauged basins in the north of Spain. J Hydrol 356(3–4):312–326

Gopakumar R, Takara K, James EJ (2007) Hydrologic data exploration and river flow forecasting of a humid tropical river basin using artificial neural networks. Water Resour Manag 21:1915–1940

Haykin S (1999) Neural networks: a comprehensive foundation. Pearson Education Pvt Ltd, Singapore

Hecht-Nielsen R (1987) Counter propagation networks. Applied Optimization 36:4979–4984

Hydrologic Engineering Center (HEC) (2000) Hydrologic modeling system HEC–HMS: Technical reference manual. U.S Army Corps of Engineers, Hydrologic Engineering Center, Davis

Kisi O (2004) River flow modeling using artificial neural network. ASCE J Hydrol Eng 9(1):60–63

Kisi O (2008) River flow forecasting and estimation using different artificial neural network techniques. Hydrol Res 39(1):27–40. doi:10.2166/nh.2008.026

Kişi O (2007) Streamflow forecasting using different artificial neural network algorithms. ASCE J Hydrol Eng 12(5):532–539

Kumar APS, Sudheer KP, Jain SK, Agarwal PK (2005) Rainfall-runoff modeling using artificial neural networks: comparison of network types. Hydrol Process 19:1277–1291

Leavesley GH, Lichty RW, Troutman BM, Saindon LG (1983) Precipitation-runoff modeling system: user's manual. United States Department of the Interior Geological Survey, Denver

Linsley RK, Kohler MA, Paulhus JLH (1958) Hydrology for engineers. McGraw-Hill, New York

Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. Environ Mod Softw 15:101–123

Modarres R (2009) Multi-criteria validation of artificial neural network rainfall-runoff modeling. Hydrol Earth Syst Sci 13:411–421

Moller M (1993) A scaled conjugate gradient algorithm for fast supervised learning. Neural Netw 6(4):525–533

Mutlu E, Chaubey I, Hexmoor H, Bajwa SG (2008) Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed. Hydrol Process 22(26):5097–5106

Ponce VM (1989) Engineering hydrology. Prentice-Hall, Englewood Cliffs

Rezaeian Zadeh M, Abghari H, van de Giesen N, Nikian A, Niknia N (2009) Maximum daily discharge prediction using multi layer perceptron network. Geophys Res 11:3652–3662

Rezaeian Zadeh M, Amin S, Khalili D, Singh VP (2010) Daily outflow prediction by multi layer perceptron with logistic sigmoid and tangent sigmoid activation functions. Water Resour Manag 24(11):2673–2688

Rezaeian Zadeh M, Tabari H, Abghari H (2012) Prediction of monthly discharge volume by different artificial neural network algorithms in semi-arid regions. Arab J Geosci. doi:10.1007/s12517-011-0517-y

Rumelhart DE, McClelland JL, PDP Research Group (1986) Parallel distributed processing. Vol.1: Foundations. The MIT Press, Cambridge, Mass

Sethi IK (1990) Entropy nets: from decision trees to neural networks. Proc IEEE 78(10):1605–1613

Soroosh M, Gity F, Sherafat AR, Farahani KH, Razaghi M (2005) A neural network model for determination of the breakdown voltage for separate absorption and multiplication region avalanche photodiode (SAM-APD). Second IEEE Conference on wireless and optical communication networks, UAE. p 173–177

Sudheer KP, Gosain AK, Ramasastri KS (2002) A data driven algorithm for constructing artificial neural network rainfall-runoff models. Hydrol Process 16(6):1325–1330

Tokar AS, Johnson A (1999) Rainfall-runoff modeling using artificial neural networks. ASCE J Hydrol Eng 4(3):232–239

Tokar AS, Markus M (2000) Precipitation-runoff modeling using artificial neural networks and conceptual models. ASCE J Hydrol Eng 5(2):156–161

Widrow B, Winter RG, Baxter RA (1988) Layered neural nets for pattern recognition. IEEE Trans Acoust Speech Signal Process 36:1109–1118

Yonaba H, Anctil F, Fortin V (2010) Comparing sigmoid transfer functions for neural network multistep ahead streamflow forecasting. ASCE J Hydrol Eng 15(4):275–283