
Journal of Hydroinformatics

River instantaneous peak flow estimation using daily flow data and machine-learning-based models

Mohammad T. Dastorani, Jamile Salimi Koochi, Hamed Sharifi Darani, Ali Talebi and M. H. Rahimian

ABSTRACT

Estimation of the design flood flow for hydraulic structures is often performed by adjusting probabilistic models to daily mean flow series. In most cases, this may cause under design of the structure capacity with possible risks of failure because instantaneous peak flows may be considerably larger than the daily averages. As there is often a lack of instantaneous flow data at a given site of interest, the peak flow has to be estimated. This paper develops new machine-learning-based methods to estimate the instantaneous peak flow from mean daily flow data where long daily data series exist but the instantaneous peak data series are short. However, the presented methods cannot be used where only daily flow data are available. Developed methodologies have been successfully applied to series of flow information from different gauging stations in Iran, with important improvements compared to traditional empirical methods available in the literature. Reliable results produced by the machine-learning-based models compared to the traditional methods show the superior ability of these techniques to solve the problem of inadequate measured peak flow data periods, especially in developing countries where it is difficult to find sufficiently long instantaneous peak flow data series.

Key words | daily flow, estimation, flood flow, hydraulic design, hydrological data, peak flow

Mohammad T. Dastorani (corresponding author)
Faculty of Natural Resources,
Yazd University (temporarily in Ferdowsi University of Mashhad),
Iran
E-mail: *mdastorani@yazduni.ac.ir*

Jamile Salimi Koochi
Hamed Sharifi Darani
Ali Talebi
Faculty of Natural Resources,
Yazd University,
Yazd,
Iran

M. H. Rahimian
National Research Centre for Salinity,
Yazd,
Iran

INTRODUCTION

The first step in designing a culvert or bridge or any hydraulic structure for a particular location is determination of design flood. This is the flood that the culvert or bridge must be able to carry safely. This involves first choosing the suitable return period for the design flood, which involves considerations of cost, risk, consequences of failure, how to deal with uncertainty, etc. In fact, design flood is an instantaneous peak discharge (IPF) with a specific return period. In particular, estimation of the design floods of hydraulic structures requires the determination of instantaneous peak flows because there may be significant stream flow fluctuations within hours or even minutes, especially in the case of small basins (drainage area up to 1,000 km^2). Another interesting aspect of this issue is related to the process by which gauge operating agencies evaluate and maintain hydrological data. In most cases, these agencies publish only the mean daily flow (MDF) data, and the use of these data in flood studies may cause underestimation of the design flood with possible risk of failure. In many hydrologic studies, particularly flood routing in reservoirs or channels, it is necessary to use the complete flood hydrograph. An important input to estimate this hydrograph is the instantaneous peak flow. There are several ways of estimating peak flow, depending on the area and the type of data available. Methods for estimating the instantaneous peak flow based on mean daily data have been studied by hydrologists for almost a century. In general, to tackle this problem, two different approaches have been presented. The first approach includes methods that seek a relationship between the so-called peak flow coefficient, defined as the ratio of instantaneous peak and the corresponding MDF, with physiographic characteristics of the basin (Fuller 1914;

doi: 10.2166/hydro.2013.245

Silva 1997; Silva & Tucci 1998). Taguas et al. (2008) established possible linear relationships between annual IPF and the corresponding MDF stream for southeastern Spain, and a regional equation to estimate IPF from MDF was developed. This equation was applied to a series of flows of nine stations in the southeast basin of Spain, and a significant improvement was achieved when applying this formula in comparison to the traditional method of Fuller. This study indicates possible restrictions to take into account when the traditional hydrological models are applied in semi-arid areas. The second approach includes methods that use the sequence of MDF data to estimate the peak flow (Jarvis 1936; Langbein 1944; Linsley et al. 1949; Sangal 1983).

Using data-driven techniques such as artificial neural networks (ANNs) and adaptive neuro-fuzzy inference systems (ANFISs) is another possibility for solving this problem. In the last few decades, many types of data-driven model have been developed to simulate hydrologic processes, and this has led to an increasing interest in ANN and ANFIS techniques that consider non-linearity in the hydrological process. Karunanithi et al. (1994) predicted river flow using adaptive ANNs; Hsu et al. (1995) evaluated the application of ANNs for rainfall-runoff processes; Minns & Hall (1996) and Tokar & Johnson (1999) employed this method as a tool for rainfall-runoff modeling; Dawson & Wilby (1998a) applied ANNs for rainfall-runoff modeling; Dawson & Wilby (1998b) also compared the application of different types of ANN for river flow forecasting; Luk et al. (1998) tried to forecast rainfall events using ANNs; Bhattacharya & Solomatine (2000) used ANN techniques to evaluate stage discharge relationships; Dastorani & Wright (2003) completed a research project on flow estimation for ungauged catchments using a neural network method; and Dastorani & Wright (2002) used ANNs for real-time river flow prediction in a multi-station catchment. Dastorani & Wright (2004) employed ANNs to optimize the results of a hydrodynamic approach used for river flow prediction; Jacquin & Shamseldin (2009), after a general overview about the use of fuzzy inference systems in river flow forecasting, stated that fuzzy inference systems can be used as effective tools for river flow forecasting, even though their application is rather limited in comparison to the popularity of neural network models. Kisi (2009) used a neurowavelet conjunction model (a combination of a discrete wavelet transform and an ANN) for intermittent stream flow forecasting, and compared the results to those of the single ANN model. Comparison of the results showed higher accuracy of the combined model over the single ANN model. Dastorani et al. (2010a) evaluated the application of ANN and ANFIS models for reconstruction of missing flow data, and reported their superior abilities compared to the related traditional methods. Dastorani et al. (2010b) used ANNs for estimation of instantaneous peak flow using daily flow data, and compared the results to existing empirical methods. Shamseldin (2010) used ANNs with different input patterns for real-time flow forecasting in Sudan, and mentioned the considerable potential of ANNs for river flow forecasting. Seckin (2011) used ANN and ANFIS models to predict flood discharge at ungauged catchments in Turkey. The results were compared to those presented by regression techniques, showing higher accuracy of the outputs produced by ANN and ANFIS models over the regression techniques. Kisi & Partal (2011) used a combined model called a wavelet-neuro-fuzzy model (a combination of wavelet transform and the neuro-fuzzy techniques) to forecast monthly stream flow, and compared the results to those presented by the classic neuro-fuzzy method. Comparison of results indicated higher accuracy of the predictions presented by the combined model over the classical neuro-fuzzy model.

In this research, after evaluation of some existing methods, several ANN- as well as ANFIS-based models have been developed and used to estimate instantaneous peak flow using daily flow data. However, it must be mentioned that the proposed method is applicable for the places where long daily flow data series are available but the instantaneous peak flow data series are short. This method cannot be used where only daily flow data are available. The results produced by these models have been compared to those presented by the most famous existing methods, and related strengths and limitations have been discussed.

MATERIALS AND METHODS

The methods used in this research are briefly presented in the following subsections.

The Fuller method

Fuller (1914) used data from 24 river basins with drainage areas varying from 3.06 to 151,592 km^2 and proposed the relationship:

$$Q_{\max} = Q(1 + 2.66A^{0.3}) \quad (1)$$

where $Q_{\max}$ is the predicted peak flow (m^3/s), Q is the maximum MDF (m^3/s) and A is the drainage area (km^2).

Following Fuller's method, many other authors have presented relationships between the ratio of peak flow and the MDF as a function of the drainage area for different regions of the world. Table 1 summarizes several formulae proposed in the literature.

The Sangal method

A more recent and well-known technique proposed by Sangal (1983) is based on the assumption of a triangular hydrograph:

$$Q_{\max} = \frac{(4Q_2 - Q_1 - Q_3)}{2} \quad (2)$$

where $Q_{\max}$ is the predicted instantaneous peak flow (m^3/s), Q_2 is the MDF of the day that contains the peak (m^3/s), and Q_1 and Q_3 are the MDFs for the posterior and anterior day, respectively (m^3/s).

Table 1 | Relationship between the ratio of peak flow and mean daily flow as a function of drainage area in the literature (Fill & Steiner 2003)

Equation	Region of study	Author
$Q_{\max}/Q_d = 3.9A^{*-0.22}$	Rocky Mountains	Gray (1973)
$Q_{\max}/Q_d = 10A^{*-0.46}$	Cypress Hills	Gray (1973)
$Q_{\max}/Q_d = 11A^{*-0.26}$	Central Plains	Gray (1973)
$Q_{\max}/Q_d = 0.37A^{*-0.38}$	Manitoba Encarp	Gray (1973)
$Q_{\max}/Q_d = 1 + 1.2A^{-0.036}$	Portugal	Correia (1983)
$Q_{\max}/Q_m = 1 + 68A^{-0.5}$	Italy	Tonini (1939)[a]
$Q_{\max}/Q_m = 32A^{-0.313}$	Italy ($A < 120$ km^2)	Cottechina (1965)[a]
$Q_{\max}/Q_m = 16A^{-0.19}$	Italy ($A > 120$ km^2)	Cottechina (1965)[a]
$Q_{\max}/Q_m = 2.39A^{-0.112}$	Italy	Tonini (1969)[a]
$Q_{\max}/Q_d = 1 + 15.03A^{-0.59}$	Brazil	Tucci (1991)

Note: A = drainage area (km^2); A^* = drainage area (mi^2); $Q_{\max}$ = peak flow; Q_d = highest observed flow; Q_m = maximum mean daily flow. [a]Reported in Canuti & Moisello (1981).

Sangal (1983) used the MDF data of 3 consecutive days. The method was tested with streams in Ontario, Canada, using 3,946 station-years of flow data collected from 387 stations. The method leads to results with reasonable accuracy, but it is downward biased for small basins.

Despite the fact that almost half of the data used in Sangal's study are from snowmelt floods, his method has been widely used in practice for flood routing through reservoirs for feasibility studies of hydroelectric plants in different parts of the world. For watersheds with drainage areas greater than 1,000 km^2, results based on Sangal's method have indicated that calculated peak flow values are significantly higher (about 50%) than the observed values (Sangal 1983). This trend of overestimating theoretical peak flows has been the motivation to review Sangal's methodology.

The Fill and Steiner method

Fill & Steiner (2003) presented a formula almost similar to the Sangal (1983) method to estimate the instantaneous peak flow from the MDF of 3 consecutive days including the peak day. The 3 days are the day with the maximum MDF and the adjacent days. Similar to the Sangal (1983) formula, it is assumed that the peak flow could be estimated by a linear combination of the MDFs of these days:

$$Q_{\max} = 0.8Q_2 + 0.25(Q_1 + Q_3) \quad (3)$$

where $Q_{\max}$ is the predicted instantaneous peak flow (m^3/s), Q_2 is the MDF of the peak day (m^3/s), and Q_1 and Q_3 are the MDFs of the day preceding and succeeding the peak day (m^3/s), respectively.

As can be seen, the equations presented by Sangal (1983) and Fill & Steiner (2003) are mostly similar. In both methods, peak flow is estimated using daily flow data of 3 consecutive days, with higher effects of the peak day (Q_2). However, in analyzing the results computed by Equation (3), Fill & Steiner (2003) evaluated the correlation between the ratio of the estimated and the observed peak flow value and the ratio of the average of the MDFs of the

1st and 3rd days and the MDF of the peak day. Next, it was verified that if a correction factor 'k' is applied to Equation (3), it would be possible to obtain better agreement between estimated and observed data. Hence:

$$Q_{\max} = [0.8Q_2 + 0.25(Q_1 + Q_3)]/k \tag{4}$$

The factor 'k' may be obtained from linear regression between the hydrograph shape factor $x = (Q_1 + Q_3)/2Q_2$ and k for the observed cases.

The ANN-based method

In this research, in addition to the three above-mentioned methods, the ANN technique was also used to estimate instantaneous peak flow data using daily measured flow data. The ANN architecture used was a three-layer feed-forward network (Figure 1).

This type of network is normally trained with the back-propagation algorithm. The backpropagation rule propagates the errors through the network and allows adoption of hidden processing elements. One hidden layer with a tangent sigmoid transfer function was used, while the output layer function was a logistic one. As the figure shows in this study, the number of processing elements (nodes) in the hidden layer was two, and the backpropagation Levenberg–Marquardt algorithm was used as the training algorithm for updating the network weights. The number of epochs was 1,000, with an acceptable error of 0.00001. Annual MDF data were used as the input, and instantaneous peak flow data of the same year were used as the network output. Matlab software was used for construction and implementation of the ANN model. Several ANN structures, including radial-basis, Hopfield, generalized regression and cascade-forward networks, were also used and the presented results were evaluated. However, feed-forward networks presented the most reliable outputs.

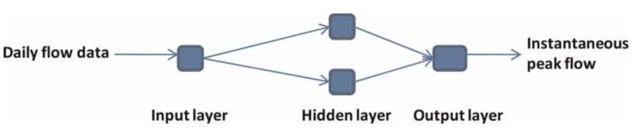

Figure 1 | A typical three-layer feedforward neural network used in this research.

The ANFIS-based method

ANFIS is another model used to estimate instantaneous peak flow data series using daily measured flow data in this research. ANFIS is an improved tool and a data-driven modeling approach for determining the behavior of imprecisely defined complex dynamical systems (Kim & Kasabov 1999). Kurian *et al.* (2006) state that ANFIS model has human-like expertise within a specific domain and learns to perform better in changing environments. A neuro-fuzzy system is defined as a combination of the ANN and the fuzzy inference system. Neuro-fuzzy systems have attracted growing interest in various scientific and engineering areas due to the increasing need of intelligent systems (Abraham & Nath 2001). Figure 2 represents a typical ANFIS architecture that is based on the work of Jang (1993).

***Layer 1*:** Every node in this layer is an adaptive node with a node function that may be a generalized bell membership function (MF) (Equation (5)) or a Gaussian MF (Equation (6)):

$$\mu_{A_i}(x) = \frac{1}{1 + \left|\dfrac{x - c_i}{a_i}\right|^{2bi}} \tag{5}$$

$$\mu_{A_i}(x) = \exp\left[-\left(\frac{x - c_i}{a_i}\right)^2\right] \tag{6}$$

where a_i, b_i and c_i are parameters determining the shape of the MFs. In addition, x is the input to node i, and A_i is the linguistic label (for example, low and high) or simply fuzzy numbers associated with this node function.

***Layer 2*:** Every node in this layer is a fixed node labeled $\prod$, representing the firing strength of each rule, and is calculated by the fuzzy AND connective of 'product' of the

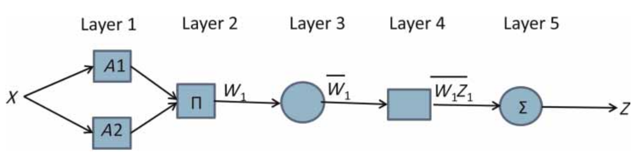

Figure 2 | A typical ANFIS architecture based on Jang (1993) used in this research.

incoming signals using:

$$w_i = \mu_{A_i}(x) \times \mu_{B_i}(x) \quad i = 1, 2 \tag{7}$$

where $\mu_{A_i}(x)$ and $\mu_{B_i}(x)$ are membership grades of fuzzy sets A and B, and w_i is the firing strength of each rule.

Layer 3: Every node in this layer is a fixed node labeled N, representing the normalized firing strength of each rule. The ith node calculates the ratio of the ith rule's firing strength to the sum of two rules' firing strengths using:

$$\overline{w_i} = \frac{w_i}{w_1 + w_2} \quad i = 1, 2 \tag{8}$$

where $\overline{w_i}$ is the normalized firing strength that is the ratio of the ith rule's firing strength (w_i) to the sum of the first and second rules' firing strengths (w_1, w_2).

Layer 4: Every node in this layer is an adaptive node with a node function (Equation (9)), indicating the contribution of ith rule toward the overall output:

$$w_i z_i = \overline{w}_i(p_i x + q_i y + r_i) \tag{9}$$

where z_i is equal to $(p_i x + q_i y + r_i)$, and p_i, q_i and r_i are consequent parameters.

Layer 5: The single node in this layer is a fixed node labeled $\sum$, indicating the overall output as the summation of all incoming signals calculated by:

$$Z = \sum_i \overline{w}_i z_i = \frac{\sum_i w_i z_i}{\sum_i w_i} \tag{10}$$

where Z is the summation of all incoming signals.

When inspecting the above layers, the three different types of components that can be adapted as follows are important (Lughofer 2003).

1. Premise parameters as nonlinear parameters that appear in the input MFs.
2. Consequent parameters as linear parameters that appear in the rule consequents (output weights).
3. Rule structure that needs to be optimized to achieve a better performance.

In this study, different MFs were used including Gaussian, triangle, trapezoidal, sigmoidal and generalized bell MFs, and the results were compared. Finally, the model with the generalized bell MF presented the more accurate results in the testing phase. There is a wide variety of algorithms available for training a network and adjusting its weights. In this study, after consideration of the available algorithms, the hybrid algorithm (a combination of gradient descent and least squares) was used for training. Matlab software was used for construction and implementation of the ANFIS model. Different numbers of epochs, including 100, 300, 500, 1,000, 1,500, 2,000 and 2,500, were used and evaluated, but 500 was the appropriate number of epochs to obtain more accurate results in the testing phase.

Study sites and data

Records of MDF and peak flow data from eight stations located in different places in Iran were applied in this study. The drainage areas varied from 418 to 7,432 km^2. Table 2 shows the gauging station code, climate condition, drainage area and location of the sites. The length of data series used in this research was 30 years (1975–2005), which was divided into two parts. The first part (24 years) was for training of the machine-learning models, and the remaining 6 years for testing purposes. Generally, this is not enough data for data-driven models, but this is what

Table 2 | The gauging stations used in this research

Station code	Climate condition	Latitude	Longitude	Elevation (m above sea level)	Drainage area (km^2)
31-011	Semi-arid	37°30′	45°01′	1,390	418
14-101	Sub-humid	36°02′	51°09′	1,790	725
41-009	Semi-arid	32°21′	50°00′	2,000	817
42-003	Semi-arid	32°39′	50°28′	2,100	1,440
12-005	Semi-arid	37°28′	55°29′	132	1,524
47-015	Arid	35°18′	52°25′	1,000	3,209
12-023	Semi-arid	37°13′	55°00′	30	6,560
31-015	Semi-arid	38°07′	46°24′	1,450	7,432

Table 3 | Data used for training and testing of ANN and ANFIS models (station code 12-005)

Data used for training						Data used for testing		
Year	Daily flow (m^3/s)	Peak flow (m^3/s)	Year	Daily flow (m^3/s)	Peak flow (m^3/s)	Year	Daily flow (m^3/s)	Peak flow (m^3/s)
1977	9.23	10.8	1989	32.8	42	2001	86.4	95.50
1978	3.04	3.67	1990	81.2	162	2002	74.48	85.40
1979	14.5	18.07	1991	95	95	2003	53.32	138.00
1980	137	214	1992	8.95	13.4	2004	15.7	48.50
1981	9.83	15.5	1993	74.1	124	2005	73.8	127.00
1982	32.4	38.1	1994	23.9	66.4	2006	225	783.00
1983	4.56	4.9	1995	5.6	7.7	–	–	–
1984	3.03	3.23	1996	10.1	85	–	–	–
1985	9.1	13.2	1997	8.19	42	–	–	–
1986	8.7	11.57	1998	61.4	197.39	–	–	–
1987	45.9	168	1999	80.2	257	–	–	–
1988	16.8	25.2	2000	90.4	120.3	–	–	–

was available on the studied stations. Therefore, 80% (24 data) was used for training and the remaining 20% (six data) was used for testing. It is clear that when few data are available, it is not possible to use many parameters in ANNs and ANFISs to train. Table 3 shows data used for training and testing of ANN and ANFIS models for one of the stations (station code 12-005) as an example.

In this study, for comparing the results, two performance criteria were used: root mean square error (RMSE) and coefficient of determination (R^2), which are calculated, respectively, by Equations (11) and (12):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(O_i - P_i\right)^2} \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \tag{12}$$

where n is the number of data points, O_i is the observed value, $\bar{O}$ is the mean of the observed values, and P_i is the predicted value. Smaller values of RMSE show satisfactory results. R^2 assesses the goodness of fit by indicating the deviation of the estimates values from the line of the best fit or the regression line. The R^2 value can vary between min infinity and 1. A value close to 1 indicates a satisfactory result, while a low value or close to 0 implies inadequate quality of the results.

RESULTS AND DISCUSSION

For comparison of the results of the proposed ANN- and ANFIS-based methods and those of Fuller (1914), Sangal (1983) and Fill & Steiner (2003), the values of R^2 and RMSE were computed for each site. The RMSE values are shown in Table 4. Figures 3 and 4 show the quality of

Table 4 | Values of RMSE calculated for the results of different methods (testing phase)

	RMSE (m^3/s)				
Station code	Fuller	Sangal	Fill & Steiner	ANN	ANFIS
31-011	19.01	21.42	13.67	7.45	4.81
14-101	15.63	20.77	9.44	13.94	5.58
41-009	4.15	14.05	9.03	4.03	1.42
42-003	60.03	85.9	70.45	118.41	40.47
12-005	203.55	150.85	189.51	131.04	32.30
47-015	37.7	36.01	28.80	6.71	5.84
12-023	18.02	22.8	30.30	9.66	6.63
31-015	30.45	68.66	59.61	22.05	14.52

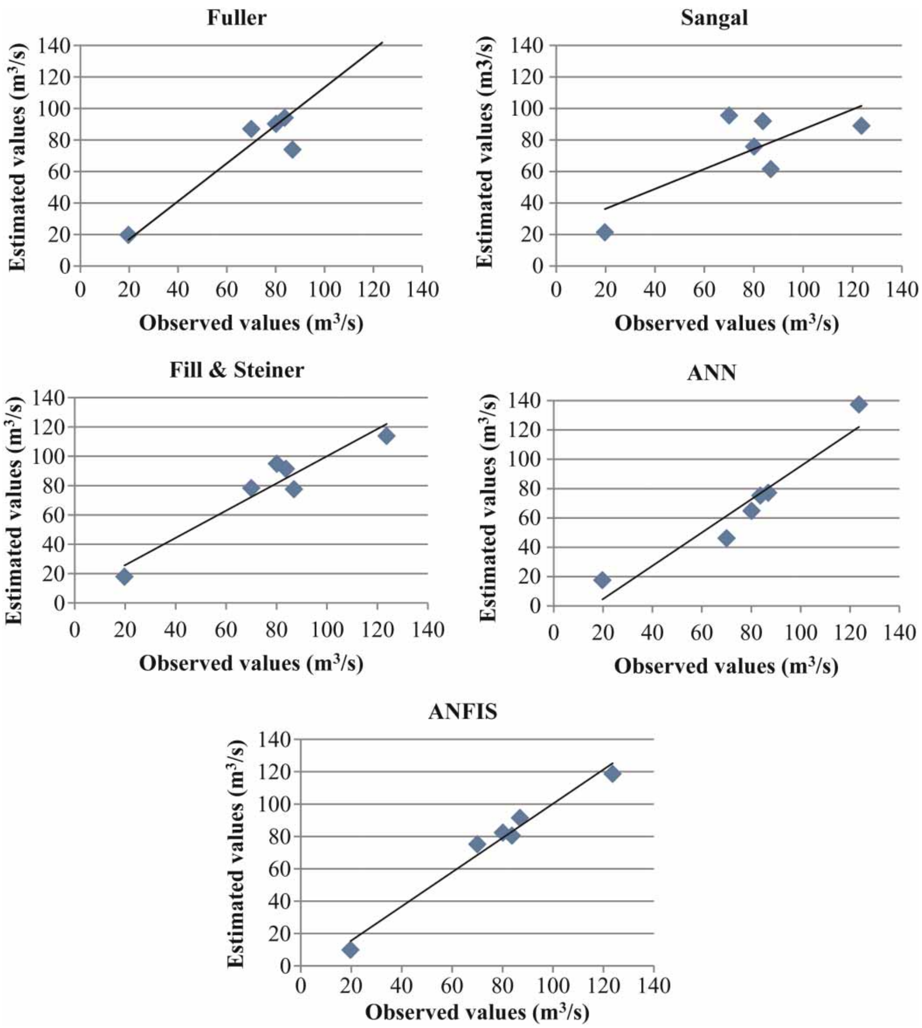

Figure 3 | Instantaneous peak flow estimated by different methods versus the observed values in gauging station code 14-101.

estimations made by different methods against the measured values for gauging stations codes 14-101 and 12-005, respectively, in the testing phase as samples. Results taken in the training phase in station code 12-005 are shown in Figure 5.

The quality of the results produced by these methods varies, and as Table 4 shows, in all cases ANFIS presented the most accurate results in comparison to other applied methods. It would be interesting to compare the data-driven models with traditional methods. According to the results, the ANFIS model shows superiority in the accuracy of estimations. The results produced by ANN also show relatively good levels of accuracy in some cases, but not all. In all eight stations, the accuracy of the ANN outputs is lower than the ANFIS outputs. Comparing the outputs of ANN models to those of empirical methods (Fuller 1914; Sangal 1983; Fill & Steiner 2003) indicates that in six stations out of eight, the accuracy of the ANN-based model is better. However, among the remaining two stations, the outputs of the Fill & Steiner method shows higher accuracy in station code 14-101, whereas in station code 42-003, the accuracy of the results presented by empirical methods is higher than those of the ANN models. In general, the results presented by the Sangal method are the poorest outputs in this research. The present study confirms the very high potential of the ANFIS model to be used for estimation of instantaneous peak flow from the MDF data. It seems that this model can perform quite well, and can be used as a powerful tool over existing methods for generation of required data.

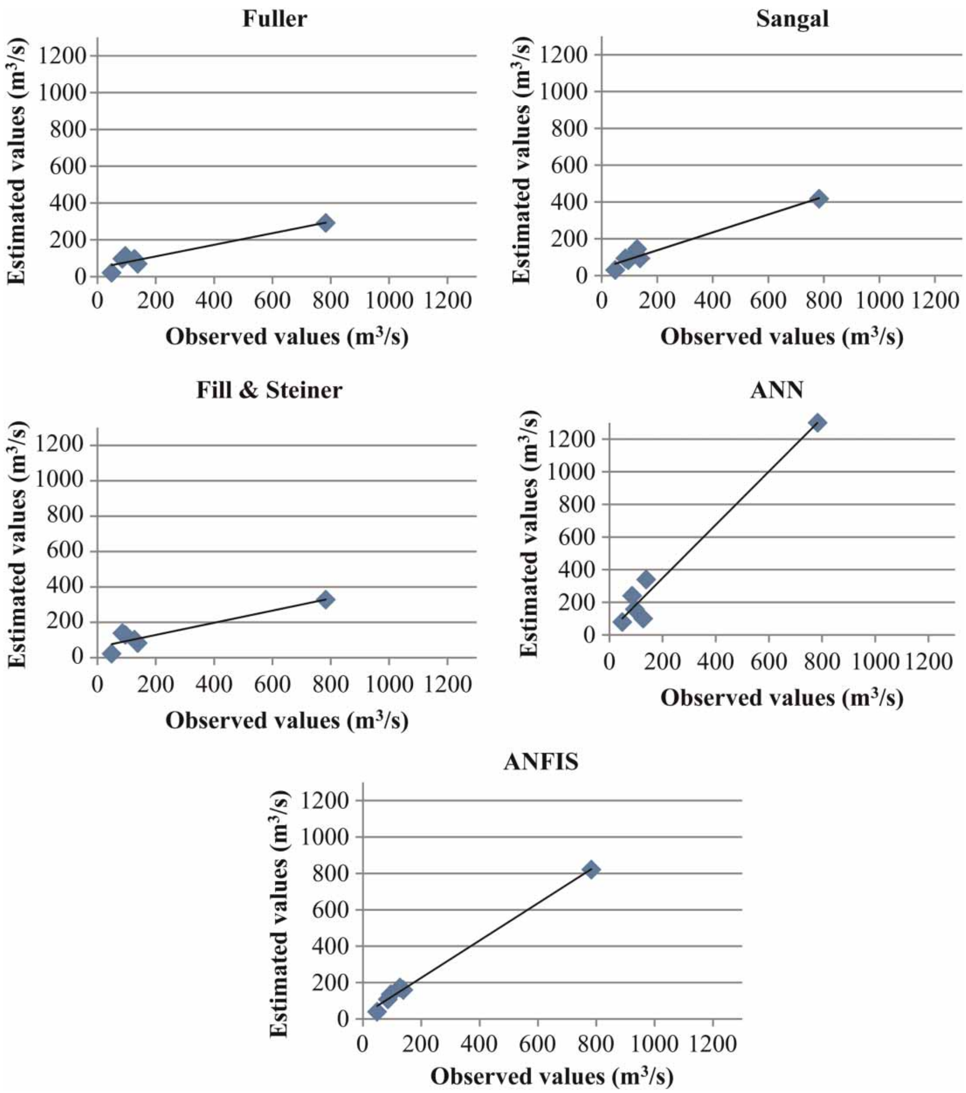

Figure 4 | Instantaneous peak flow estimated by different methods versus the observed values in gauging station code 12-005.

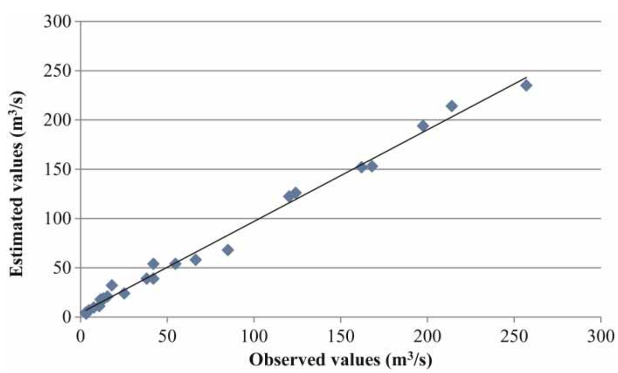

Figure 5 | Instantaneous peak flow estimated by different methods versus observed values in gauging station code 12-005 (training phase).

Due to lack of appropriate performance of traditional and statistical formulae used in hydrology, the interest of applying data-driven models such as ANNs and ANFISs for hydrological simulations is further increased. However, one of the major limitations of ANNs is their lack of explanatory power, also referred to as their 'black box problem'. Neuro-fuzzy techniques remove some of the shortcomings of ANNs. They merge neural networks and fuzzy logic into an integrated system, with relatively higher abilities.

The obtained results confirmed the main hypothesis of the research (preference of the data-driven techniques over traditional methods on peak flow data estimation using daily records). However, it can be seen that although a large number of studies has been carried out and reported on, applications of ANN and ANFIS in hydrology, none of them is related to estimation of instantaneous peak flow using daily data series. Therefore, more investigations need to be completed on application of the mentioned techniques in this specific field.

The five methods mentioned earlier have been applied to the same problem under similar conditions and compared with the same performance indices. Referring to the relative performance of the five methods (Table 4), it can be observed that the lowest RMSE between observed and simulated results in all of the stations belonged to the ANFIS method. In fact, these results indicate that ANFIS is superior for estimation of instantaneous peak flow. Such superiority may be problem related, and may need extensive applications on various data sets to be generalized. However, one can say that the superiority of this technique might be attributed to its ability to capture the nonlinear dynamics of the data. Table 4 and Figures 3 and 4 (as samples) clearly show that the behavior of the used models varies when dealing with data in different gauging stations.

It should be noted that some recent studies have indicated that the noise that exists in hydrologic data may limit the performance of many modeling techniques. Some methods have been proposed to reduce the level of noise in the data set, which may lead to improvement in the accuracy of the estimation of instantaneous peak flow data. It seems suitable that machine-learning techniques such as ANFIS have better abilities to deal with the problem of noise in data.

Extreme hydrologic events have great importance in most countries, including Iran, regarding the economic damage they imply. The commonly used design parameter for hydraulic structures is the maximum annual instantaneous stream flow recorded in conventional gauging stations. However, most available data in different parts of Iran are mean daily stream flows. In most of the stations, for example, daily mean flow data are available for 40 years, whereas the related instantaneous peak flow data can be found for only 30 years. In such a case, the peak flow data for the remaining 10 years can be estimated with high accuracy using the proposed models to prolong the peak data time series, which is important in water-related projects. In this condition, models are trained using the mentioned 30 year data (as both mean daily and peak flow data are available). Inadequate peak data are possibly the problem for most developing countries, not just Iran, and finding a sophisticated approach to use the mean daily data to predict peak values (with acceptable accuracy) is an important step ahead in optimization of water-related projects.

CONCLUSIONS

This research was designed to evaluate the applicability of machine-learning techniques, including ANNs and ANFISs for estimation of instantaneous peak flow from MDF data in different climatic regions. It can be observed from the results of this research that the proposed ANFIS-based method displays an average relative RMSE considerably lower and an R^2 higher than other methods used. Comparing the results taken from this research shows the ability of ANFIS over the existing mentioned methods for this specific application. Important improvements of the results produced by machine-learning methods compared to traditional methods shows the superior abilities of machine-learning techniques to solve the problem of inadequate measured instantaneous peak flow data, and optimize water-related designs. In fact, the proposed method considerably improves the accuracy and precision of other traditional methods. This suggests that it is an appropriate method for estimating missing instantaneous peak flows for flood studies and the design of hydraulic structures when MDFs have been measured for long enough, but the instantaneous peak flows data series are short. A regional analysis is possible using different regressions for each basin or group of similar basins if sufficient numbers of data are available. This approach may result in further improvement of the accuracy. Of course, the methodology proposed in this study is not the last step for the estimation of instantaneous peak flows. More research in this subject is recommended in order to obtain better estimators, including methods using basin physiographic characteristics.

REFERENCES

Abraham, A. & Nath, B. 2001 A neuro-fuzzy approach for modeling electricity demand in Victoria. *Applied Soft Computing* **1**, 127–138.

Bhattacharya, B. & Solomatine, D. P. 2000 Application of artificial neural network in stage-discharge relationship. In: *Proc. 4th International Conference on Hydroinformatics*, Iowa City, USA.
Canuti, P. & Moisello, U. 1981 Relazionetra I massiniannualidelleportate al como e mediegiornaliere per alcunibacinidellatoscana. *L'Energia Elettrica* **48**, 11. In Italian.
Correia, F. N. 1983 Métodos de Análise e Determinação de Caudais de Cheia. Laboratório Nacional de Engenharia Civil, Tese, Investigador, Auxiliar, Lisboa, Portugal. In Portuguese.
Dastorani, M. T. & Wright, N. G. 2002 Artificial neural network-based real-time river flow prediction. In: *5th International Conference of Hydroinformatics*, Cardiff, UK.
Dastorani, M. T. & Wright, N. G. 2003 Flow estimation for ungauged catchments using a neural network method. In: *6th International River Engineering Conference*, Ahwaz, Iran.
Dastorani, M. T. & Wright, N. G. 2004 A hydrodynamic/neural network approach for enhanced river flow prediction. *International Journal of Civil Engineering* **2** (3), 141–148.
Dastorani, M. T., Moghadamnia, A., Piri, J. & Rico-Ramirez, M. 2010a Application of ANN and ANFIS models for reconstructing missing flow data. *Journal of Environmental Monitoring and Assessment* **166**, 421–434.
Dastorani, M. T., Salimi, J., Talebi, A. & Abghari, H. 2010b River instantaneous peak flow estimation using mean daily flow data. In: *9th International Conference on Hydroinformatics (HIC 2010)*, Tianjin, China.
Dawson, C. W. & Wilby, R. 1998a An artificial neural network approach to rainfall-runoff modelling. *Journal of Hydrological Sciences* **43** (1), 47–66.
Dawson, C. W. & Wilby, R. 1998b A comparison of artificial neural network used for river flow forecasting. *Journal of Hydrology and Earth System Sciences* **3** (4), 529–540.
Fill, H. D. & Steiner, A. A. 2003 Estimating instantaneous peak flow from mean daily flow data. *Journal of Hydrologic Engineering* **8** (6), 365–369.
Fuller, W. E. 1914 Flood flows. *Transactions of American Society of Civil Engineering* **77**, 564–617.
Gray, D. M. 1973 *Handbook of the Principles of Hydrology*. Water Information Center, Huntington, NY.
Hsu, K., Gupta, H. V. & Sorooshian, S. 1995 Artificial neural network modeling of the rainfall-runoff process. *Journal of Water Resources Research* **31** (10), 2517–2530.
Jacquin, A. P. & Shamseldin, A. Y. 2009 Review of the application of fuzzy inference systems in river flow forecasting. *Journal of Hydroinformatics* **11** (3), 202–210.
Jang, J. R. 1993 Adaptive network based fuzzy inference systems. *IEEE Transactions on Systems, Man and Cybernetics* **23**, 665–685.
Jarvis, C. S. 1936 *Floods in United States, Water Supply Paper*. US Geological Survey, Reston, VA.
Karunanithi, N., Grenney, W. J., Whitley, D. & Bovee, K. 1994 Neural networks for flow prediction. *Journal of Computing in Civil Engineering* **8** (2), 201–220.
Kim, J. & Kasabov, N. 1999 HyFIS: adaptive neuro-fuzzy inference systems and their application to nonlinear dynamical systems. *Neural Networks* **12**, 1301–1319.
Kisi, O. 2009 Neural networks and wavelet conjunction model for intermittent streamflow forecasting. *Journal of Hydrologic Engineering* **14** (8), 773–782.
Kisi, O. & Partal, T. 2011 Wavelet and neuro-fuzzy conjunction model for streamflow forecasting. *Hydrology Research* **42** (6), 447–456.
Kurian, C. P., George, V. I., Bhat, J. & Aithal, R. S. 2006 ANFIS model for the time series prediction of interior daylight illuminance. *International Journal on Artificial Intelligence and Machine Learning* **6** (3), 35–40.
Langbein, W. B. 1944 Peak discharge from daily records. *US Geological Survey Bulletin* **August**, 145.
Linsley Jr., R. K., Kohler, M. A. & Paulhus, J. L. H. 1949 *Applied Hydrology*. McGraw-Hill, New York, pp. 223–224.
Lughofer, E. 2003 Online Adaptation of Takagi-Sugeno Fuzzy Inference Systems. Technical Report, Fuzzy Logic Laboratorium, Linz-Hagenberg.
Luk, K. C., Ball, J. E. & Sharma, A. 1998 Rainfall forecasting through artificial neural networks. In: *Hydroinformatics 98* (V. Babovic & L. C. Larsen, eds). Balkema, Rotterdam.
Minns, A. W. & Hall, M. J. 1996 Artificial neural networks as rainfall-runoff models. *Journal of Hydrological Sciences* **41** (3), 399–417.
Sangal, B. P. 1983 Practical method of estimating peak flow. *Journal of Hydraulic Engineering* **109** (4), 549–563.
Seckin, N. 2011 Modeling flood discharge at ungauged sites across Turkey using neuro-fuzzy and neural networks. *Journal of Hydroinformatics* **13** (4), 842–849.
Shamseldin, A. Y. 2010 Artificial neural network model for river flow forecasting in a developing country. *Journal of Hydroinformatics* **12** (1), 22–35.
Silva, E. A. 1997 Estimativa regional da vazãomáximainstantâneaemalgumasbaciasbrasileiras. MS Thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil. In Portuguese.
Silva, E. A. & Tucci, C. E. M. 1998 Relação entre as vazõesmáximasdiárias e instantâneas. *Revista Brasileira de Recursos Hýdricos* **3** (1), 133–151. In Portuguese.
Taguas, E. V., Ayuso, J. L., Pena, A., Yuan, Y., Sanchez, M. C., Giraldez, J. V. & Pérez, R. 2008 Testing the relationship between instantaneous peak flow and mean daily flow in a Mediterranean area Southeast Spain. *Catena* **75**, 129–137.
Tokar, S. A. & Johnson, P. A. 1999 Rainfall-runoff modeling using artificial neural networks. *Journal of Hydrologic Engineering* **4** (3), 232–239.
Tucci, C. E. 1991 *Regionalização de vazões no Rio Grande do Sul*. IPH/UFRGS, Porto Alegre, Brazil. In Portuguese.

First received 26 October 2011; accepted in revised form 7 December 2012. Available online 14 February 2013