Modeling flood discharge at ungauged sites across Turkey using neuro-fuzzy and neural networks

Neslihan Seckin

ABSTRACT

One of the most important problems in hydrology is the reliable forecasting of maximum discharge at an ungauged site of interest. Statistical techniques are commonly used for finding the maximum discharge and return period relationship. However, these techniques are generally considered to be inadequate because of the complexity of the problem. Hence, neural network techniques are preferred. In this study, two different neural network models developed based on the following techniques – a multi-layer perceptron neural network with Levenberg–Marquardt algorithm and a radial basis neural network behind an adaptive neuro-fuzzy inference system – are employed in order to capture the nonlinear relationship between discharge and five independent variables – drainage area (km^2), elevation (m), latitude, longitude, return period (year) and maximum discharge (m^3/s). For a modeling study, watershed data from 543 catchments across Turkey were used. Statistical models with regression techniques were also applied to the same data, providing a wider comparison. The results of the models were then compared and assessed with respect to mean square errors, mean absolute error, mean absolute relative error and determination coefficient. Based on these results, it was found that the neural network techniques demonstrated better performance in predicting the maximum discharge based on five independent variables than the regression techniques, and were comparable to the adaptive neuro-fuzzy inference system.

Key words | adaptive neuro-fuzzy inference system, neural networks, regional flood frequency, regression, ungauged site

Neslihan Seckin (corresponding author)
Faculty of Engineering, Civil Engineering Department,
Cukurova University,
Balcali-Adana 01330,
Turkey
E-mail: *nseckin@cu.edu.tr*

INTRODUCTION

In hydrological and hydraulic engineering applications, maximum discharge and its frequency are required for flood risk assessment projects, management of water resources, design of hydraulic structures such as dams, spillways, road and railway bridges and culverts, design of urban drainage systems and flood plain zoning as well as economic evaluation of flood protection projects. Whenever rainfall or river flow records are not available at or near the site of interest, it is difficult for hydrologists to derive reliable flood estimates directly.

Estimation must be fairly accurate to avoid excessive costs in the case of overestimation of the flood magnitude, or excessive damage and even loss of human lives on underestimating the flood potential. Regional flood frequency analysis (RFFA) usually involves the identification of homogeneous regions, selection of suitable regional frequency distributions and estimation of flood quantiles at sites of interest (Dalrymple 1960; Hosking & Wallis 1997). It does so by using data from several sites, which are judged to obey the same probability distribution peculiar

doi: 10.2166/hydro.2010.046

to the site of interest. One of the most important procedures in RFFA is the delineation of the homogeneous regions. Many methods have been developed for determining homogeneous regions (e.g. Stedinger & Tasker 1985; Wiltshire 1986; Acreman & Wiltshire 1989; Burn 1990; Reed & Robson 1999; Ouarda *et al.* 2000, 2001, 2006; Chokmani & Ouarda 2004; Shu & Burn 2004a). RFFA models were commonly used in the literature with complex computational efforts and a large requirement of input data, in comparison with less detailed models such as regression methods, adaptive neuro-fuzzy inference system (ANFIS) and neural network approaches (ANNs).

Regression methods are frequently used to predict flood quantiles as a function of site physiographical and other characteristics, as pointed out by Shu & Ouarda (2008). These methods have been applied when no historical flood data are available (Zrinji & Burn 1994; Hosking & Wallis 1997; Pandey & Nguyen 1999; Ouarda *et al.* 2006).

In the hydrological forecasting context, recent studies have shown that both ANNs and fuzzy logic may offer a promising alternative for rainfall–runoff modelling (Smith & Eli 1995; Tokar & Johnson 1999; Lin & Chen 2004), streamflow prediction (Chibanga *et al.* 2003; Cigizoglu 2003; Kisi 2004a; Cigizoglu & Kisi 2005), suspended sediment modeling (Tayfur 2002; Kisi 2004b, 2005; Cobaner *et al.* 2009), reservoir inflow forecasting (Coulibaly *et al.* 2005; Jain *et al.* 1999), prediction of hydropower energy (Cobaner *et al.* 2008) and flood frequency modeling (e.g. Muttiah *et al.* 1997; Jingyi & Hall 2004; Shu & Burn 2004a,b; Dawson *et al.* 2006; Shu & Ouarda 2008).

Muttiah *et al.* (1997) used the cascade correlation (CC) neural network model with the Quickprop algorithm (Fahlman 1988) to predict the two-year peak discharge for major regional river basins of the continental United States (US). They concluded that the CC neural network provided reasonable estimates of peak discharges and it has simpler variable input requirements, which consisted of drainage area, average annual precipitation and mean basin elevation, than existing models.

Jingyi & Hall (2004) applied a geographical approach to Ward's cluster method, the fuzzy c-means method and a Kohonen neural network for 86 sites from the Gan and Ming river basins in China to delineate homogeneous regions based on site characteristics. They showed that lower standard errors of estimates can be produced using an artificial neural network (ANN).

Shu & Burn (2004a) developed a few methods to derive an objective similarity measure between catchments using fuzzy expert systems (FES) with a genetic algorithm and applied these methods to flood data from Great Britain. They concluded that the proposed FES methods have the capability of utilizing inputs from different categories of catchment characteristics and producing similarity measures that have a high reliability. A full detailed description of the datasets used by Shu & Burn (2004a) are given by Cunderlik & Burn (2002).

Shu & Burn (2004b) also developed six approaches for creating ANN ensembles and applied these approaches in pooled flood frequency analysis for estimating the index flood and the 10-year flood quantile. The results showed that ANN ensembles generate improved flood estimates and were less sensitive to the choice of initial parameters when compared with an ANN, namely the multi-layer perceptron (MLP) neural network with Levenberg–Marquardt (LM) algorithm (Hagan & Menhaj 1994). On the other hand, Shu & Burn (2004b) pointed out that ensemble approaches resulted in a much heavier computational load than for an ANN.

Dawson *et al.* (2006) applied the MLP neural network model to data from the Centre for Ecology and Hydrology's Flood Estimation Handbook (FEH) to predict *T*-year flood events and the index flood for 850 catchments across the UK. They showed that the MLP approach trained by a standard error back-propagation algorithm was more reliable than multiple regression models.

Shu & Ouarda (2008) applied ANFIS to 151 catchments in the province of Quebec, Canada for flood quantile estimation at ungauged sites and compared ANFIS results with those of the MLP and nonlinear regression approaches. They used the LM algorithm for MLP training. Shu & Ouarda (2008) identified the structure of the ANFIS using the subtractive clustering algorithm. They used a hybrid learning algorithm consisting of back-propagation and least-squares estimation for ANFIS system training. They concluded that the ANFIS approach has a much better generalization capability than nonlinear regression techniques and is comparable to the ANN approach.

The purpose of this study was to estimate the peak flood discharge at ungauged sites in river basins across Turkey by utilizing the spatial location parameters and observed maximum discharge values of sites while implementing the available developments in ANN models such as namely an MLP neural network with LM algorithm and a radial basis neural network (RBNN) behind an ANFIS. Multiple-linear (MLR) and multiple-nonlinear regression (MNLR) models were also used in the analysis for wider comparison. This paper differs from other studies cited in the literature since these modeling techniques as well as RBNN have been applied in all river basins in Turkey for the first time.

STUDY AREA

The annual instantaneous flood peaks were determined for 543 runoff gauging stations operated by the Electrical Power Resources Survey and Development Administration and the State Hydraulic Works in all River Basins for a total of 26 across Turkey (SHW 1994). The peak flood discharges were the highest recorded event for each gauging station with varying record periods between 15 and 57 years. Figure 1 shows the boundaries of the basin areas with the locations of the gauging stations.

MULTI-LAYER PERCEPTRON (MLP) NEURAL NETWORKS

Among the many ANN paradigms, the multi-layer back-propagation network (MLP) is by far the most popular (Lippman 1987). The network consists of layers of parallel processing elements, called neurons, with each layer being fully connected to the preceding layer by interconnection strengths fully connected to the preceding layer by interconnection strengths, or weights, W. Initial estimated weight values are progressively corrected during a training process that compares predicted outputs to known outputs, and back-propagates any errors to determine the appropriate weight adjustments necessary to minimize the errors. The Levenberg–Marquardt algorithm was used here for adjusting the weights. Throughout all MLP simulations the adaptive learning rates were used for the purposes of faster training speed and solving the local minima problem. For each epoch, if performance decreases toward the goal, then the learning rate is increased by the factor learning increment. If performance increases, the learning rate is adjusted by the factor learning decrement. A common strategy for finding the optimum number of hidden layer neurons starts with a few numbers of neurons and increasing the number of neurons while monitoring the performance criteria, until no significant improvement is observed. Accordingly, the numbers of hidden layer neurons were found using a simple trial-and-error method in all the applications herein.

RADIAL BASIS FUNCTION-BASED NEURAL NETWORKS (RBNN)

RBNN were introduced into the neural network literature by Broomhead & Lowe (1988). The RBNN consists of two layers whose output nodes form a linear combination of basis

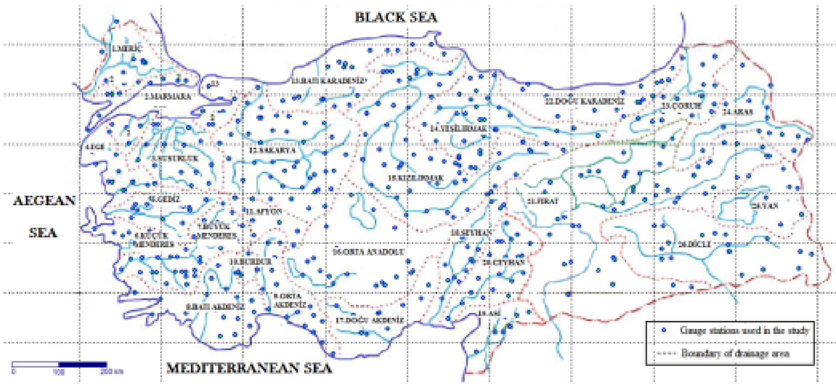

Figure 1 | River basin stations with boundaries of drainage areas in the study area.

functions. The basis functions in the hidden layer produce a significant non-zero response to input stimulus only when the input falls within a small localized region of the input space. Hence, this paradigm is also known as a localized receptive field network (Lee & Chang 2003). Transformation of the inputs is essential for fighting the curse of dimensionality in empirical modeling. The type of input transformation of the RBNN is the local nonlinear projection using a radial fixed-shape basis function. After nonlinearly squashing the multi-dimensional inputs without considering the output space, the radial basis functions play a role as regressors. Since the output layer implements a linear regressor the only adjustable parameters are the weights of this regressor. These parameters can therefore be determined using the linear least-squares method, which gives an important advantage for convergence.

ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM (ANFIS)

ANFIS was first introduced by Jang (1993). By using a hybrid learning procedure, the ANFIS can construct an input–output mapping based on both human knowledge and stipulated input–output data pairs. The ANFIS uses back-propagation learning to determine premise parameters (to learn the parameters related to membership functions) and uses least-mean-squares estimation to determine the consequent parameters. A step in the learning procedure has got two parts: In the first part the input patterns are propagated and the optimal consequent parameters are estimated by an iterative least-mean-squares procedure, while the premise parameters are assumed to be fixed for the current cycle through the training set. In the second part the patterns are propagated again, and in this epoch, back-propagation is used to modify the premise parameters while the consequent parameters remain fixed. This procedure is then iterated.

APPLICATION AND RESULTS

The data for predicting the peak flood discharge through modeling used in this study were obtained from the *Handbook of Flood Frequency Analysis for Peak Discharges Observed through River Basins in Turkey* (State Hydraulic Works (SHW) 1994). Watersheds were selected according to the following criteria: minimum record lengths for the watersheds should be a minimum of 15 years, as it is often of concern from a design perspective as suggested by SHW (1994). Variables used to estimate the maximum discharge were selected according to three criteria as discussed by Shu & Burn (2004b): (1) a relatively small set of variables must be used in the analysis, (2) variables must provide a good statistical fit and (3) variables must represent a hydrological model. In the light of these statements, the maximum discharge values at ungauged sites have been estimated by means of five independent variables, i.e. drainage area (*DA*), elevation above sea level (*EASL*), longitude (*LO*) and latitude (*LA*) of the gauge sites and return period (*T*). Herein, *T* was computed as follows:

$$T = \frac{1}{1-F} \tag{1}$$

where F is the plotting probability of the ith-order statistic and is given as follows by Hosking *et al.* (1985):

$$F = \frac{i-0.35}{n} \tag{2}$$

in which n is the sample size.

The cross-validation provides a rigorous testing of ANN skill (Dawson & Wilby 2001). It involves dividing available data into three sets, namely a training set, a validation set and a testing set. The training set is used to fit ANN model weights, the validation set is used to select the model variant and the testing set is used to evaluate the chosen model against unseen data. On the other hand, Maier & Dandy (2000) stated that: "It is common practice to split the available data into two sub-sets; a training set and an independent validation set." The following references can be given as examples of the application of the splitting process: Jain *et al.* (1999), Kisi (2005) and Alp & Cigizoglu (2007). Similar to those references, before the application of MLR, MNLR, MLP, ANFIS and RBNN methods, in this study, the dataset was first split into two groups as a training set (380 gauge stations) and testing set (163 gauge stations). The minimum and maximum values of model variables are summarized in Table 1. Next, both MLR and MNLR techniques were applied to the training dataset. The following formulae, using MLR

Table 1 | Minimum and maximum values of the input and output parameters

	Training dataset		Testing dataset	
Parameters	**Min**	**Max**	**Min**	**Max**
Drainage area (*DA*) (km^2)	9.9	75120.8	18.8	52531.6
Elevation above sea level (*EASL*) (m)	2.0	2188.0	2.0	2075.0
Latitude (*LA*) (deg)	36.1	42.0	36.2	41.7
Longitude (*LO*) (deg)	26.3	48.1	26.7	44.1
Return period (*T*) (years)	25.7	174.3	28.6	160.0
ln(discharge) (ln(*Q*)) (m^3/s)	1.4	9.0	2.2	7.9

and MNLR techniques, were found to offer the best statistical fit for the training dataset, respectively:

$$\ln(Q) = 3.987\,667 + 4.12\mathrm{E}{-}05DA - 0.001\,34EASL + 0.078LO - 0.0414LA + 0.0139T \tag{3}$$

$$\ln(Q) = \frac{1333\,DA^{0.099}LO^{0.242}\,T^{0.088}}{EASL^{0.068}LA^{0.0405}}. \tag{4}$$

Different ANN, ANFIS and RBNN architectures were tested thanks to the Matlab package with many codes and neural network toolboxes and, hence, the appropriate model structures were determined for the input parameters.

A difficult task with ANNs is to choose the required parameters such as the number of hidden nodes, the learning rate and the initial weights. Determining an appropriate architecture of a neural network for a particular problem is an important issue, since the network topology directly affects its computational complexity and its generalization capability. The optimum network geometry is obtained utilizing a trial-and-error approach in which ANNs are trained with one hidden layer. It should be noted that one hidden layer could approximate any continuous function, provided that sufficient connection weights are used (Hornik *et al.* 1989). Here, the hidden layer node numbers of each model were determined after trying various network structures since there is no theory yet to tell how many hidden units are needed to approximate any given function. In the training stage, the adaptive learning rates and the same initial weights were used for each ANN network. The sigmoid and linear activation functions were used for the hidden and output nodes, respectively. The best MLP model results were obtained from the ANN (4, 8, 1) model using the logarithmic sigmoid activation functions for both hidden and output layer neurons. The iteration number of MLP was 43.

For the RBNN applications, different numbers of hidden layer neurons and spread constants were examined in the study. The number of hidden layer neurons that provided the minimum mean square errors (*MSE*) was found to be 19. The spread is a constant which is selected before RBNN simulation. The larger that spread is, the smoother the function approximation will be. Too large a spread means a lot of neurons will be required to fit a fast-changing function. Too small a spread means many neurons will be required to fit a smooth function and the network may not generalize well. The spread that gives the minimum *MSE* is 0.18. These were found with a simple trial-error method adding some loops to the program codes.

Before applying the MLP and RBNN to the data, the training input and output values were normalized using the equation

$$a\frac{x_i - x_{\min}}{x_{\max} - x_{\min}} + b \tag{5}$$

where $x_{\min}$ and $x_{\max}$ denote the minimum and maximum of the stage and discharge data. Different values can be assigned for the scaling factors, *a* and *b*. There are no fixed rules as to which standardization approach should be used in particular circumstances (Dawson & Wilby 1998). The values of *a* and *b* were taken herein as 0.6 and 0.2, respectively.

For the applications using ANFIS, the hybrid learning algorithm, which combines back-propagation and the least-squares method, was used to rapidly train and adapt the fuzzy inference system. The triangle and Gaussian membership functions were tried for different neuro-fuzzy models. The best results with the lowest *MSE* values were obtained from the ANFIS (2, trimf) model with five triangle membership functions. The iteration number of ANFIS was 5.

The mean square errors (MSE), the mean absolute errors (MAE), the mean absolute relative errors (MARE) and the correlation coefficient (*R*) statistics were used in the applications. The MSE, MAE and MARE statistics are defined as

Table 2 | Performances of MLR, MNLR, MLP, RBNN and ANFIS modeling techniques in training and testing phases

	Training phase				Testing phase			
Method	**MARE (%)**	**MAE (m^3/s)**	**MSE $(m^3/s)^2$**	***R***	**MARE (%)**	**MAE (m^3/s)**	**MSE $(m^3/s)^2$**	***R***
MLR	18.3	0.840	1.031	0.683	17.3	0.781	0.971	0.675
MNLR	15.7	0.745	0.842	0.760	17.5	0.824	1.030	0.685
MLP	12.7	0.566	0.531	0.851	14.5	0.649	0.656	0.797
RBNN	17.6	0.785	0.948	0.713	16.1	0.729	0.821	0.736
ANFIS	15.1	0.718	0.801	0.765	15.8	0.721	0.928	0.700

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(Y_i\ \text{observed} - Y_i\ \text{predicted})^2 \qquad (6)$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|Y_i\ \text{observed} - Y_i\ \text{predicted}| \qquad (7)$$

$$\mathrm{MARE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Y_i\ \text{observed} - Y_i\ \text{predicted}}{Y_i\ \text{observed}}\right| \times 100 \qquad (8)$$

in which N is the number of data and Y_i is the maximum discharge.

The MSE, MAE, MARE and R values of each method in the training and testing phases are summarized in Table 2. As seen in Table 2, the MLP model provided the lowest MSE (0.531), MAE (0.566) and MARE (%12.7) and the highest R (0.851) in the training phase. The MLP model also gave the smallest MARE (%14.5), MSE (0.656) and MAE (0.649) and the highest R (0.797) in the testing phase. According to the testing results, it was found that the MLP estimations were still better than those of the MLR, MNLR, ANFIS, and RBNN. Although the MLR gave similar statistical results in comparison with the MNLR in the testing phase, the MNLR model demonstrated better accuracy than the MLR in the training phase. It can be seen in Table 2 that the ANN models performed better than the MLR and MNLR models in the testing periods.

The ANFIS model, on the other hand, ranked second based on the MARE value of 15.1% in the training phase after the MLP model. The model also provided the second-best R with 0.765 after the MLP model with 0.851. The RBNN was the second-best model based on R and MSE values during the testing phase after MLP. During testing, the ANFIS model provided a MARE of 15.8%, which was the second-best, and better than all models except the MLP model.

Figure 2 plots the observed and predicted maximum discharge values for all methods during the training (left column) and testing (right column) phases. When the data points on and around the lines of equality were assessed, it is apparent that the data points fell on the line better for the MLP and ANFIS models during both training and testing phases, while the other models suggest a nonlinear relationship between the observed and predicted data.

CONCLUSION

The models tried in this study were the MLR, MNLR, MLP, RBNN and ANFIS for the determination of maximum flood discharges. The results showed that MLP gave better results although the others, especially ANFIS and RBNN, produced reasonable and comparable estimations. In the current study, five input parameters (drainage area, elevation above sea level, longitude and latitude of the gauge sites, and return period) were used for the prediction of maximum flood discharges during modeling. Instead of these parameters for the hydrological modeling studies, it is, of course, extremely probable that the predictions would present more accuracy provided that more available independent variables such as physiographic, soil and land use properties, and climate were included in the modeling set-up. Thus, further studies for the same region are recommended including more available independent variables in the modeling set-up.

ACKNOWLEDGEMENTS

The author would like to extend her thanks to Prof. Tefaruk Haktanir, who is an expert on flood frequency analysis, and

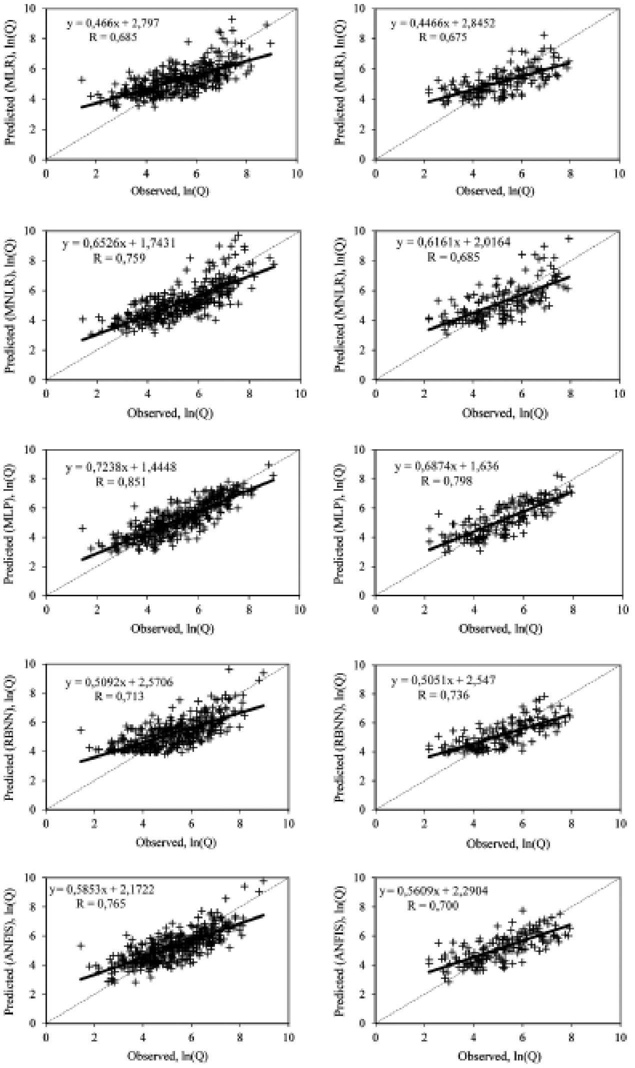

Figure 2 | Comparison of the observed and predicted maximum discharge values using MLR, MNLR, MLP, RBNN and ANFIS models during training (left column) and testing (right column) phases.

Associate Prof. Ozgur Kisi, who is an expert on artificial neural network modeling, for their valuable input as well as to the three anonymous reviewers for their constructive comments in improving this paper.

REFERENCES

Acreman, M. C. & Wiltshire, S. E. 1989 The regions are dead; long live the regions. Methods of identifying and dispensing with regions for flood frequency analysis. *IAHS Publ.* **187**, 175–188.

Alp, M. & Cigizoglu, H. K. 2007 Suspended sediment estimation by feed forward back propagation method using hydro meteorological data. *Environ. Modell. Software* **22**(1), 2–13.

Broomhead, D. & Lowe, D. 1988 Multivariable functional interpolation and adaptive networks. *Complex Syst.* **2**, 321–355.

Burn, D. H. 1990 An appraisal of the 'region of influence' approach to flood frequency analysis. *Hydrol. Sci. J.* **35**(2), 149–165.

Chibanga, R., Berlamont, J. & Vandewalle, J. 2003 Modelling and forecasting of hydrological variables using artificial neural networks: the Kafue River sub-basin. *Hydrol. Sci. J.* **48**(3), 363–379.

Chokmani, K. & Ouarda, T. B. M. J. 2004 Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. *Wat. Res. Res.* **40**(12), W12514.

Cigizoglu, H. K. 2003 Estimation, forecasting and extrapolation of river flows by artificial neural networks. *Hydrol. Sci. J.* **48**(3), 349–361.

Cigizoglu, H. K. & Kisi, O. 2005 Flow prediction by three back propagation techniques using k-fold partitioning of neural network training data. *Nordic Hydrol.* **36**(1), 49–64.

Cobaner, M., Haktanir, T. & Kisi, O. 2008 Prediction of hydropower energy using ANN for the feasibility of hydropower plant installation to an existing irrigation dam. *Wat. Res. Mngmnt.* **22**, 757–774.

Cobaner, M., Unal, B. & Kisi, O. 2009 Suspended sediment concentration estimation by an adaptive neuro-fuzzy and neural network approaches using hydro-meteorological data. *J. Hydrol.* **367**, 52–61.

Coulibaly, P., Hache, M., Fortin, V. & Bobée, B. 2005 Improving daily reservoir inflow forecasting with model combination. *J. Hydrol. Engng.* **10**(2), 91–99.

Cunderlik, J. M. & Burn, D. H. 2002 The use of flood regime information in regional flood frequency analysis. *Hydrol. Sci. J.* **47**, 77–92.

Dalrymple, T. 1960 *Flood–Frequency Analyses*. US Geological Survey Water-Supply Paper 1543-A. Reston, VA, USA.

Dawson, C. W., Abrahart, R. J., Shamseldin, A. Y. & Wilby, R. L. 2006 Flood estimation at ungauged sites using artificial neural networks. *J. Hydrol.* **319**, 391–409.

Dawson, C. W. & Wilby, R. L. 1998 An artificial neural network approach to rainfall-runoff modeling. *Hydrol. Sci. J.* **43**(1), 47–66.

Dawson, C. W. & Wilby, R. L. 2001 Hydrological modeling using artificial neural networks. *Progr. Phys. Geogr.* **25**(1), 80–108.

Fahlman, S. E. 1988 *An Empirical Study of Learning Speed in Backpropagation Networks*. CMU-CS-88-162, School of Computer Science, Carnegie Mellon University, Pittsburgh.

Hagan, M. T. & Menhaj, M. B. 1994 Training feed forward networks with the Marquaradt algorithm. *IEEE Trans. Neural Networks* **6**, 861–867.

Hornik, K., Stinchcombe, M. & White, H. 1989 Multilayer feedforward networks are universal approximators. *Neural Networks* **2**, 359–366.

Hosking, J. R. M. & Wallis, J. R. 1997 *Regional Frequency Analysis: An Approach Based On L-Moments*. Cambridge University Press, New York.

Hosking, J. R. M., Wallis, J. R. & Wood, E. F. 1985 Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27**, 251–261.

Jain, S. K., Das, D. & Srivastava, D. K. 1999 Application of ANN for reservoir inflow prediction and operation. *J. Wat. Res. Plann. Mngmnt.* **125**(5), 263–271.

Jang, J. S. R. 1993 ANFIS: adaptive-network based fuzzy inference systems. *IEEE Trans. Syst. Man Cybern.* **23**(3), 665–685.

Jingyi, Z. & Hall, M. J. 2004 Regional flood frequency analysis for the Gan-Ming River basin in China. *J. Hydrol.* **296**, 98–117.

Kisi, O. 2004a River flow modelling using artificial neural networks. *J. Hydrol. Engng.* **9**(1), 60–63.

Kisi, O. 2004b Multi-layer perceptrons with Levenberg–Marquardt optimization algorithm for suspended sediment concentration prediction and estimation. *Hydrol. Sci. J.* **49**(6), 1025–1040.

Kisi, O. 2005 Suspended sediment estimation using neuro-fuzzy and neural network approaches. *Hydrol. Sci. J.* **50**(4), 683–696.

Lee, G. C. & Chang, S. H. 2003 Radial basis function networks applied to DNBR calculation in digital core protection systems. *Annal. Nucl. Energy* **30**, 1561–1572.

Lin, G. F. & Chen, L. H. 2004 A non-linear rainfall-runoff model using radial basis function network. *J. Hydrol.* **289**, 1–8.

Lippman, R. 1987 An introduction to computing with neural nets. *IEEE ASSP Mag.* **4**, 4–22.

Maier, H. R. & Dandy, G. C. 2000 Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. *Environ. Modell. Software* **15**, 101–124.

Muttiah, R. S., Srinivasan, R. & Allen, P. M. 1997 Prediction of two year peak stream discharges using neural networks. *J. AWRA* **33**(3), 625–630.

Ouarda, T. B. M. J., Cunderlik, J. M., St-Hilaire, A., Barbet, M., Bruneau, P. & Bobée, B. 2006 Data-based comparison of seasonality-based regional flood frequency methods. *J. Hydrol.* **330**, 329–339.

Ouarda, T. B. M. J., Girard, C., Cavadias, G. S. & Bobée, B. 2001 Regional flood frequency estimation with canonical correlation analysis. *J. Hydrol.* **254**, 157–173.

Ouarda, T. B. M. J., Haché, M., Bruneau, P. & Bobée, B. 2000 Regional flood peak and volume estimation in a northern Canadian basin. *J. Cold Region Engng.* **14**, 176–191.

Pandey, G. R. & Nguyen, V.-T.-V. 1999 A comparative study of regression based methods in regional flood frequency analysis. *J. Hydrol.* **225**, 92–101.

Reed, D. W. & Robson, A. J. 1999 *Flood Estimation Handbook* vol 3. Institute of Hydrology, Wallingford.

Shu, C. & Burn, D. H. 2004a Homogeneous pooling group delineation for flood frequency analysis using a fuzzy expert system with genetic enhancement. *J. Hydrol.* **291**, 132–149.

Shu, C. & Burn, D. H. 2004b Artificial neural network ensembles and their application in pooled flood frequency analysis. *Wat. Res. Res.* **40**(9), W09301.

Shu, C. & Ouarda, T. B. M. J. 2008 Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. *J. Hydrol.* **349**, 31–43.

Smith, J. & Eli, R. N. 1995 Neural network models of rainfall–runoff process. *J. Wat. Res. Plann. Mngmnt. ASCE* **121**(6), 499–508.

State Hydraulic Works (SHW), General Directorate 1994 *Handbook of Flood Frequency Analysis for Peak Discharges Observed through River Basins in Turkey*. SHW, Ankara (in Turkish).

Stedinger, J. R. & Tasker, G. D. 1985 Regional hydrologic analysis 1. *Wat. Res. Res.* **21**(9), 1421–1432.

Tayfur, G. 2002 Artificial neural networks for sheet sediment transport. *Hydrol. Sci. J.* **47**(6), 879–892.

Tokar, A. S. & Johnson, P. A. 1999 Rainfall–runoff modelling using artificial neural networks. *J. Hydrol. Engng.* **4**(3), 232–239.

Wiltshire, S. E. 1986 Regional flood frequency analysis I: Homogeneity statistics. *Hydrol. Sci. J.* **31**(3), 321–333.

Zrinji, Z. & Burn, D. H. 1994 Flood frequency analysis for unguaged sites using a region of influence approach. *J. Hydrol.* **153**, 1–21.

First received 16 March 2010; accepted in revised form 22 May 2010. Available online 22 November 2010