Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data

Jan Adamowski, Hiu Fung Chan, Shiv O. Prasher and Vishwa Nath Sharda

ABSTRACT

Himalayan watersheds are characterized by mountainous topography and a lack of available data. Due to the complexity of rainfall–runoff relationships in mountainous watersheds and the lack of hydrological data in many of these watersheds, process-based models have limited applicability for runoff forecasting in these areas. In light of this, accurate forecasting methods that do not necessitate extensive data sets are required for runoff forecasting in mountainous watersheds. In this study, multivariate adaptive regression spline (MARS), wavelet transform artificial neural network (WA-ANN), and regular artificial neural network (ANN) models were developed and compared for runoff forecasting applications in the mountainous watershed of Sainji in the Himalayas, an area with limited data for runoff forecasting. To develop and test the models, three micro-watersheds were gauged in the Sainji watershed in Uttaranchal State in India and data were recorded from July 1 2001 to June 30 2003. It was determined that the best WA-ANN and MARS models were comparable in terms of forecasting accuracy, with both providing very accurate runoff forecasts compared to the best ANN model. The results indicate that the WA-ANN and MARS methods are promising new methods of short-term runoff forecasting in mountainous watersheds with limited data, and warrant additional study.

Key words | artificial neural network, Himalayan watersheds, multivariate adaptive regression splines, rainfall–runoff modeling, time series forecasting, wavelet

Jan Adamowski (corresponding author)
Hiu Fung Chan
Shiv O. Prasher
Department of Bioresource Engineering,
McGill University,
21 111 Lakeshore Road,
Ste. Anne de Bellevue, QC,
Canada H9X 3V9
E-mail: *jan.adamowski@mcgill.ca*

Vishwa Nath Sharda
Central Soil and Water Conservation Research and
    Training Institute,
218 Kaulagarh Road,
Dehradun 248 195,
India

NOTATION

CWT	continuous wavelet transform
g	activation function
I_i	input value to node i of the input layer
L	number of wavelet decomposition
N	number of data points used
O_k	output at node k of the output layer
R^2	coefficient of determination
RMSE	root mean square error
SEE	sum of squared errors
V_j	hidden value to node j of the hidden layer
s	scale parameter
$x(t)$	signal
$\bar{y}_i$	mean value taken over N
y_i	observed peak weekly water demand
$\hat{y}_i$	forecasted peak weekly water demand
τ	translation parameter
*	complex conjugate
$\psi(t)$	mother wavelet

INTRODUCTION

In the Himalayan region of India, conversion of forest areas into agricultural areas and other land use changes have

altered existing water runoff patterns, resulting in increased surface runoff and reduced ground water recharge (Sharda *et al*. 2006). In an area where agriculture is the major economic activity (Sharda *et al*. 2006), it is important for future water use planning to analyze the effect of environmental and morphological parameters on the hydrological cycle of micro-watersheds in this area, and to be able to accurately forecast runoff in a watershed.

Process-based models can be used in runoff forecasting, and examples of such models include SWAT (Soil Water Assessment Tool) (Arnold & Fohre 2005), ANSWERS (Areal Nonpoint Source Watershed Environment Response Simulation) (Dillaha *et al*. 1998), and AnnAGNPS (Annualized Agricultural Non-Point Source) (Bosch *et al*. 1998). However, process-based models usually necessitate large numbers of input parameters that are not easily obtained in mountainous regions such as the Himalayas. In addition, temporal and spatial unpredictability in watershed characteristics can increase the complexity of rainfall–runoff relationships, resulting in difficulties in developing accurate process-based models.

'Data based models' are an alternate method used for runoff forecasting, and are particularly useful in areas such as the Himalayas where there is a lack of data. Multiple linear regression (MLR) and autoregressive moving average (ARMA) models are probably the most common data based methods currently employed for forecasting runoff (e.g., Raman & Sunilkumar 1995; Young 1999; Adamowski 2008a, c). Data based artificial intelligence models can also be developed for areas with limited data and without *a priori* knowledge of the mathematical relationships interlinking inputs with outputs. Artificial intelligence models often require few (and generally more easily available and measurable) input parameters. These models are generally still able to model complex phenomena satisfactorily. Recently, artificial neural networks (ANNs) have been introduced for runoff forecasting applications (e.g., Tawfik 2003; Kisi 2004; Corani & Guaris 2005; Adamowski 2008a, c; Akhtar *et al*. 2009; Nourani *et al*. 2009a; Piotrowski & Napiorkowski 2011). An advantage of ANNs is that they are often effective with non-linear data. However, ANNs and other linear and non-linear data based methods often have restrictions with time series data that is non-stationary (Cannas *et al*. 2006). Methods such as ANNs may not be

able to deal with non-stationary data if input data pre-processing is not done.

It has been noted that techniques for handling non-stationary data are not as advanced as techniques for stationary data (Cannas *et al*. 2006). Two recent publications (Solomatine & Ostfeld 2008; Maier *et al*. 2010) on future directions in the use of data based modeling in hydrological forecasting noted two very important issues that need to be investigated in more depth: (i) the development and testing of hybrid model architectures that build on the strengths of different modeling methods; and (ii) the development and testing of robust modeling techniques that are able to handle 'noisy' data. This research project was focused on addressing these two issues in the context of runoff forecasting in mountainous watersheds with limited data.

One method that deals effectively with multi-scale and non-stationary behavior is wavelet analysis, which can be used to detect and extract signal variance both in time and scale simultaneously. It does not require any assumptions of stationarity. Wavelet transforms are able to deal with non-stationary time series in forecasting because they can automatically localize and filter the non-stationary component of a signal, instead of attempting to de-trend or suppress quasi-periodic smooth components as, for example, in the non-stationary autoregressive integrated moving average approach. Over the course of the last 10 years, wavelet analysis has begun to be explored in the hydrology and water resources literature.

Several studies have recently been published that investigate the use of hybrid wavelet transform and ANN (WA-ANN) models for hydrological forecasting applications. A WA-ANN model for monthly runoff forecasting was developed by Cannas *et al*. (2006) for a watershed on the Italian island of Sardinia. Kisi (2008) and Partal (2009) developed a hybrid model for monthly runoff forecasting in Turkey. Kisi (2009) explored the use of WA-ANN models for daily runoff forecasting of intermittent rivers. Wu *et al*. (2009a) developed WA-ANN models for 1, 2, and 3 days ahead forecasting. Nourani *et al*. (2009b) proposed a WA-ANN model that can predict both short- and long-term runoff discharges. Adamowski & Sun (2010) developed a coupled wavelet transform and neural network method for runoff

forecasting of non-perennial rivers in semi-arid regions. These studies found that the WA-ANN models outperformed ANN models for runoff forecasting.

Regarding other new methods, Wu *et al.* (2009b) proposed a crisp distributed support vector regression (CDSVR) model for monthly streamflow forecasting and compared it with four other methods: autoregressive moving average (ARMA), K-nearest neighbors (KNN), ANNs and crisp distributed artificial neural networks (CDANN). To improve model performance, the data pre-processing techniques of singular spectrum analysis (SSA) and moving average (MA) were coupled with all five models. They found that models fed by pre-processed data performed better than models fed by original data, and that the CDSVR method outperformed other models. Wang *et al.* (2009) compared ARMA, ANN, adaptive neural-based fuzzy inference systems (ANFIS), genetic programming (GP) and support vector machine (SVM) models for monthly river flow forecasting, and found that the best models were the ANFIS and SVM models. Nourani *et al.* (2011) proposed a wavelet-ANFIS method for watershed rainfall–runoff modeling. They found that the wavelet-ANFIS method provided accurate forecasts because it used multi-scale time series of rainfall and runoff data in the ANFIS input layer.

Another new artificial intelligence method, multivariate adaptive regression splines (MARS), was first introduced by Friedman (1991). MARS has been found to be a rapid, flexible and accurate method for forecasting continuous and binary output variables (Salford Systems 2001). MARS models use a nonparametric modeling approach without identifying the functional relationship between the input and output variables (Friedman 1991). Instead, MARS models construct this functional relation from a set of coefficients and basis functions from the regression data. The main advantage of MARS models is that the relationship of the MARS models is additive and interactive, which involves fewer variable interactions (Lee *et al.* 2006). In recent years, various environmental and hydrological research studies have been conducted using the MARS method. Yang *et al.* (2003) simulated pesticide transport trends in soils with the MARS method. Leathwick *et al.* (2005) used the MARS method to predict the distributions of New Zealand's freshwater diadromous fish. Bera *et al.*

(2006) simulated pesticide concentrations in soil with the MARS method. Leathwick *et al.* (2006) used the MARS method to study the relationships between the distributions of 15 freshwater fish species and their environment. Sharda *et al.* (2006, 2008) utilized the MARS method for the prediction of runoff from the Sainji watershed in India, and concluded that MARS models have the potential to accurately forecast total runoff for hilly watersheds. Balshi *et al.* (2009) analyzed the response of burned areas due to climate change in western boreal North America using the MARS method. And finally, Latinez-Sotomayor (2010) concluded that the performance of the MARS method is more accurate than the ANN method for the forecasting of rain and temperature in the Mantaro River basin in Peru. Based on the results of the above studies, it can be seen that the MARS method has the potential to produce comparable or even better results than the ANN method, which has been widely used in runoff forecasting in the last decade. To the best knowledge of the authors: (i) the WA-ANN method has not been tested for runoff forecasting in mountainous watersheds with limited data; and (ii) the MARS and WA-ANN methods have not been compared to date for runoff forecasting.

To address the above described issues, in this research WA-ANN, ANN and MARS models were developed and compared for runoff forecasting in three micro-watersheds with limited data in the mountainous watershed of Sainji in the Himalayan region of India. The MARS models were developed by one of the authors of this paper in a previous study (Sharda *et al.* 2006) for the same data set from the three Himalayan micro-watersheds located in Uttaranchal State, India.

METHODS

MARS

MARS models are developed to forecast continuous numeric outcomes. The MARS algorithm consists of a forward and a backward stepwise procedure. In the forward stepwise procedure, it can be viewed as a selection of a set of appropriated input variables. However, after a number of splits, this excessive forward stepwise selection procedure

could generate a complex and over-fitted model (Andres *et al.* 2010). Such a model will have poor forecasting performance. To improve forecasting accuracy, the backward stepwise procedure eliminates the unnecessary variables among the previously selected set. This function projects variable X to a new variable Y by using either of the following two basis functions, using a knot or value of a variable that defines an inflection point along the range of inputs (Sharda *et al.* 2006):

$$Y = \max(0, X - c) \tag{1}$$

$$Y = \max(0, c - X) \tag{2}$$

where c is some chosen threshold value. Two adjacent splines will intersect at a knot to maintain the continuity of the basis functions. The function is applied in a forward–backward stepwise approach to each input variable to recognize the location of knots where the function value changes. For more detailed information on the development of the MARS models used in this research the reader can refer to Sharda *et al.* (2006).

Artificial neural networks

An artificial neural network is composed of many artificial neurons that are linked together according to a specific network architecture. The objective of the neural network is to transform the inputs into meaningful outputs. A neural network can be used to predict future values of possibly noisy multivariate time series based on past histories. In the last decade, ANNs have become popular for hydrological forecasting such as runoff modeling, ground water and precipitation forecasting, and water quality forecasting (e.g., Shrestha *et al.* 2005; Han *et al.* 2006; Sahoo & Ray 2006; Han *et al.* 2007; Adamowski 2008b; Banerjee *et al.* 2009, Pramanik & Panda 2009; Wu & Chau 2011). The most popular ANN model used in these applications is the multilayer perceptron (MLP). The structure of the MLP has nodes organized in layers, with each node only connected with the nodes in adjoining layers. An overall relationship is formed through weighted linear functions at each of the nodes, with the output of each node forming the input of the nodes in the following layers. By having many layers and many functions, a non-linear function is created.

Coupled wavelet and artificial neural networks (WA-ANN)

Wavelets are mathematical functions that use time-scale representations to analyze time series that may contain non-stationarities. There are two types of wavelet transforms: the continuous wavelet transform (CWT) and the discrete wavelet transform (DWT).

The CWT is shown by Cannas *et al.* (2006):

$$\mathrm{CWT}_x^\psi(\tau, s) = |s|^{1/2} \psi^* \int_{-\infty}^{+\infty} \left(\frac{t - \tau}{s}\right) \mathrm{d}t \tag{3}$$

while the DWT is also shown by Cannas *et al.* (2006):

$$\psi_{j,k}(t) = s_0^{-j/2} \psi\left(\frac{t - k\tau_0 s_0^j}{s_0^j}\right) \tag{4}$$

In these equations, s is the scale parameter, τ is the translation parameter, $\psi(t)$ is the mother wavelet and transforming function, '$*$' denotes the complex conjugate, j and k are integers and $s_0 > 1$ is a fixed dilation step. The DWT is obtained by modifying the mother wavelet function $\psi(t)$ to the form shown in Equation (4).

The DWT is the more commonly used of the two transforms because by rendering the continuous function discrete it requires less computational time and resources to implement (Cannas *et al.* 2006). Another advantage of using the DWT is that it allows for the implementation of digital filtering effects. High and low pass filters can be easily implemented, producing detailed coefficients and approximation series (Cannas *et al.* 2006). This allows for the analysis of trends in specific frequency bands that may not otherwise be apparent, and the separation of these trends for further analysis. In this study, the DWT was used. In forecasting models that use wavelet transforms, attention must be given to the boundaries of the signal to ensure that future information is not included (e.g., Kim & Valdes 2003; Murtagh *et al.* 2004, Renaud *et al.* 2005). As such, the 'a trous' DWT algorithm used in this study was modified to address this issue.

The coupled wavelet and neural network models are ANN models which use wavelet decomposed sub-series components as inputs. The development of the WA-ANN models is described in detail in the section 'Coupled wavelet and artificial neural network models'.

Model performance comparison

The performance of different models may be assessed in terms of goodness of fit. For this research two commonly used performance indices were used to evaluate the accuracy of the models: the coefficient of determination (R^2) and the root mean squared error (RMSE).

R^2 shows the discrepancy between the observed and forecasted data and indicates how close the points are to the bisector in the scatter plot of two variables. R^2 is calculated via the following formula:

$$R^2 = 1 - \left(\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (\hat{y}_i - \bar{y}_i)^2} \right) \qquad (5)$$

where N, $\hat{y}_i$, y_i, $\bar{y}_i$ are the number of observations, observed data, predicted values and mean of observed data, respectively. A perfect fit between observed and forecasted values is described by an R^2 of 1.

The RMSE evaluates the variance of errors independently of the sample size via the following formula:

$$RMSE = \sqrt{\frac{SEE}{N}} \qquad (6)$$

where

$$SEE = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \qquad (7)$$

and N is the number of data points used. SEE is known as the sum of squared errors, with the variables as defined above. RSME values range from 0 to infinity, with a perfect fit between observed and forecasted values described by an RMSE of 0.

Study areas and data

Study watershed

The data used for the development of the models in this research were obtained from the Sainji watershed in the Uttaranchal State of India. The main watershed (WS1) consists of two sub-watersheds (WS2 and WS3), having third order and second order streams, respectively. WS1, WS2 and WS3 are shown in Figure 1. The areas of WS1, WS2 and WS3 are 255, 52 and 163 ha, respectively. WS1 is composed mainly of mixed forest scrub, WS2 is composed mainly of agriculture and scrub forest, and WS3 is composed mainly of mixed forest. The average slopes of the watersheds are between 62 and 66%, and the drainage densities are between 2.2 and 3.83 km/km^2. For further information on the watersheds the reader is referred to Table 1.

Data

The morphological variables used in this study were: length of the watershed (m); area of the watershed (ha); relief (m); circulatory ratio (CR); compactness coefficient (CC); elongation ratio (ER); drainage density (DD); time of concentration (TOC, min); stream length (m); main channel length (m); and the percentage of land under agriculture, forest and scrub. The contour map of the main watershed was digitized and the morphometric parameters extracted with ArcInfo 8.0 GS at 20 m vertical intervals (Sharda *et al.* 2006). The following variables were also used in this study: antecedent precipitation index (API5), rainfall, day of the year, and runoff number. Runoff was estimated using the SCS curve number.

Due to the absence of available data, all three micro-watersheds (WS1, WS2, and WS3) were gauged specifically for this research project to monitor daily rainfall and runoff at their respective outlets. This was done by constructing broad crested weir-type structures and equipping them with automatic stage level recorders and recording and non-recording rain gauges. Data from all three sub-watersheds were recorded continuously on a daily basis for a 2-year period from July 1, 2001 to June 30, 2003.

All of the above data was then used together (i.e., data from all the three watersheds was compiled together) to develop the MARS, ANN and WA-ANN models of this study.
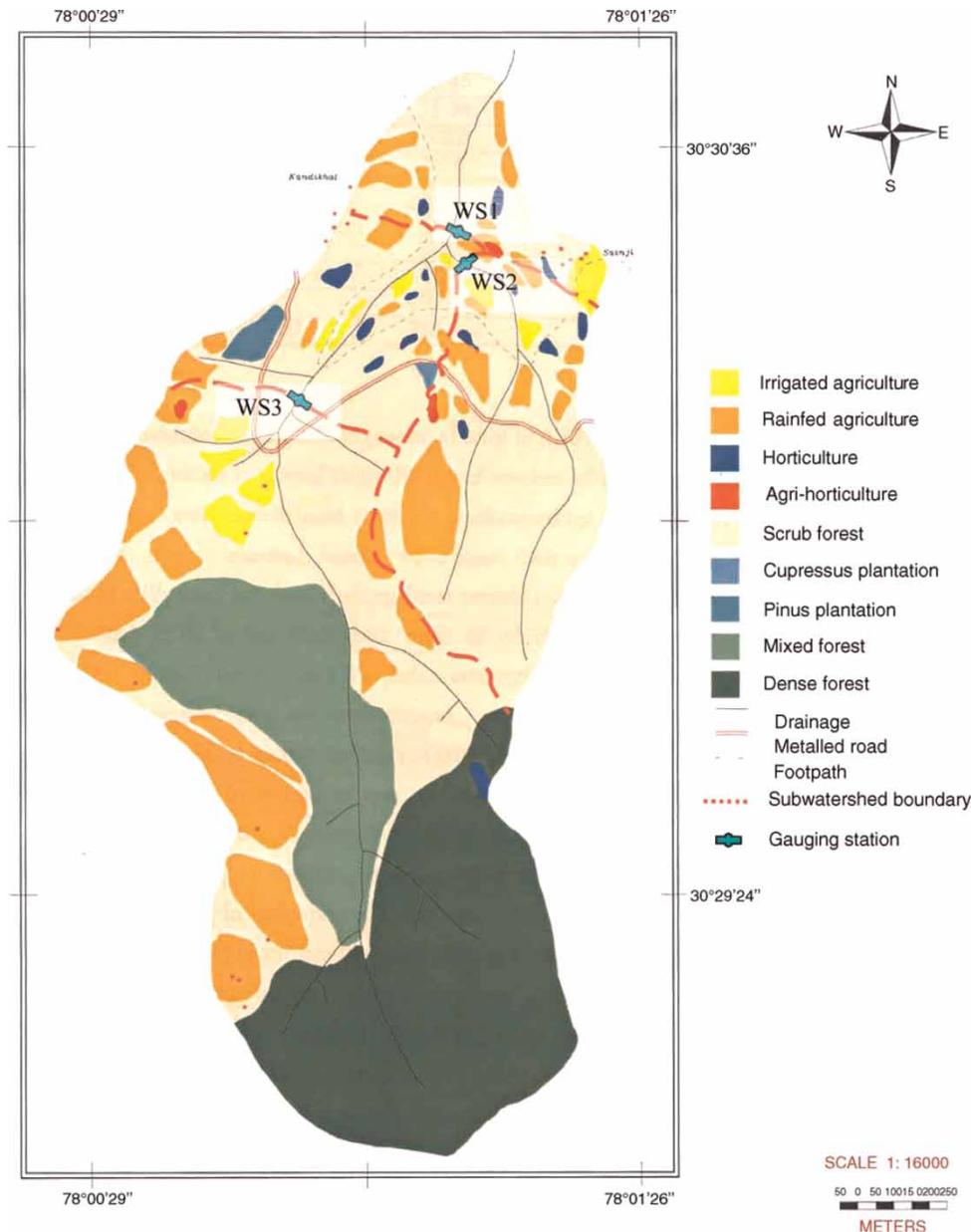
Figure 1 | Land use map of the Sainji watershed and location of gauging stations in the three sub-watersheds (Sharda *et al.* 2006).

Model development

MARS models

The MARS™ software (version 2.0) (Salford Systems 2001) was used to build the MARS models (Sharda *et al.* 2006). The MARS™ software first maximizes the number of knots and the corresponding basis functions by a trial and error approach. It then prunes the unwanted knots and basis functions to create a simplified model, and also assesses the relative importance of each input variable in the development of the model by sequentially eliminating input parameters while assessing the corresponding change in the goodness of fit. Different combinations of the data described in the previous section were tested as inputs to determine the best MARS model.

In order to develop MARS models with smaller data sets (e.g., the 2-year data sets from all three watersheds used in

Table 1 | Morphological characteristics of the three Sainji sub-watersheds (Sharda et al. 2006)

Category	Watershed characteristics	WS1	WS2	WS3
General features	Area (ha)	255	52	163
	Length (m)	2,950	1,360	2,100
	Relief (m)	1,020	635	870
Shape indicators	Circulatory ratio (dimensionless)	0.553	0.704	0.705
	Compactness coefficient (dimensionless)	134	1.19	1.18
	Elongation ratio	0.610	0.598	0.686
Drainage pattern	Drainage density (km/km^2)	2.76	3.83	2.2
	Time of concentration (minutes)	14	6.76	9.86
	Length of streams (m)	7,050	2,010	3,595
	Main channel length (m)	2,950	1,360	2,100
Land use pattern	Agriculture (%)	16.55	22.94	14.87
	Forest (%)	36.53	0.64	54.01
	Scrubs (%)	46.92	76.42	29.12
Hydrologic soil cover complex	Weighted curve number	64.99	69.57	62.57

this study), a 10-fold cross-validation procedure has been recommended to assess the generalization ability of the model (Weiss & Kulikowski 1991). As such, all the models in this study were developed with a 10-fold cross-validation procedure to verify the generalization ability of the model. In this process, the data are randomized and then divided into 10 equal parts. The first nine parts of the data are used to train the model, and the final 10th part of the data is used to test the model (Sharda et al. 2006). Repeating this procedure for all 10 possible combinations allows for the development of a robust model, with each 'fold' being tested on 'unseen' data. The MARS models that were developed were then compared using statistical measures of goodness of fit (R^2 and RMSE).

Artificial neural network models

The primary objective of ANN modeling is to optimize the architecture of the ANN that captures the relationship between the input and output variables. The regular ANN models (i.e., those not using wavelet decomposed input data) consisted of an input layer, one single hidden layer, and one output layer consisting of one node denoting the targeted daily total runoff. The number of nodes in the input and output layers directly corresponds to the number of input and output parameters being modeled. The optimal structure of the hidden layer is determined through a trial and error procedure (Jain et al. 2001). Each ANN model was tested for one to 10 hidden neurons to determine the optimum number of neurons in the one hidden layer (found to be seven).

The Levenberg-Marquardt (LM) algorithm was utilized to train the ANN models in MATLAB because it has been found to be accurate and reliable (e.g., Adamowski & Karapataki 2010). It was determined by the MARS modeling process that rainfall, antecedent precipitation index (API5), season (day of the year), and the runoff number are the most important input parameters (Sharda et al. 2006). Therefore, in order to determine the best ANN model, various combinations of these variables were tested as input nodes in the ANN models.

Individual ANN models for total runoff were developed with a 10-fold cross-validation procedure to verify the generalization ability of the model (Weiss & Kulikowski 1991). To do this the data were randomized, and then divided into 10 equal parts. The model was created using nine parts of the data (the training data set) and the remaining 'unseen' 10th part was used to test the model (Sharda et al. 2006). This procedure was repeated for all 10 possible combinations. Thus, each time, a model was constructed and

tested with an 'unseen' data set. The ANN models that were developed were then compared using statistical measures of goodness of fit (R^2 and RMSE).

Coupled wavelet and artificial neural network models

As mentioned above, it was determined by the MARS models that rainfall, antecedent precipitation index (API5), season (day of the year), and the runoff number are the most important input parameters (Sharda et al. 2006). Therefore, in order to determine the best WA-ANN model, various combinations of these variables were tested as input nodes in the WA-ANN models. The original data (rainfall, antecedent precipitation index (API5), season (day of the year), and the runoff number data series) were decomposed using a modified version of the 'a trous' wavelet algorithm (to ensure that future data are not used). The original time series for each variable was first decomposed into an approximation and accompanying detail signal. The decomposition process was then iterated with successive approximation signals being decomposed in turn, and in this way the original time series was broken down into lower resolution components.

The following formula was used to determine the number of decomposition levels (Nourani et al. 2009a):

$$L = \text{int}[\log(M)] \tag{8}$$

where L and M are the decomposition level and number of time series data, respectively. For this study $M = 2,190$ for training, which results in $L = 4$ approximately. Therefore, four wavelet decomposition levels were selected. A new series was obtained for each variable by adding the details and approximate series for a specific variable, and these series were then used as inputs to the ANN models.

The ANN networks developed for the WA-ANN models (which also used the Levenberg-Marquardt training algorithm) consisted of an input layer, a single hidden layer, and one output layer. Each WA-ANN model was tested for one to 10 hidden neurons to determine the optimum number of neurons in the hidden layer (found to be seven).

Individual WA-ANN models for total runoff were developed with a 10-fold cross-validation procedure to verify the generalization ability of the model (Weiss &

Kulikowski 1991). To do this, the data were randomized, and then divided into 10 equal parts. Models were created using nine parts of the data (the training data set) and the remaining 'unseen' 10th part was used to test a model (Sharda et al. 2006). This was repeated for all 10 possible combinations. The WA-ANN models that were developed were then compared using statistical measures of goodness of fit (R^2 and RMSE).

RESULTS AND DISCUSSION

MARS models

The results from the Sharda et al. (2006) 10-fold cross-validation MARS modeling to forecast the total runoff are presented in Table 2. The best MARS model had an R^2 of 0.939 for the testing data set. The MARS results indicate that the models were not only able to learn the relationship between the inputs and the total runoff, but were also able to apply it successfully to unseen data sets. The low testing RMSE value (0.292) is a further indication of the good performance of the MARS model. It can be seen that MARS modeling has the potential to forecast total runoff effectively for mountainous watersheds with limited data. Figure 2 illustrates the observed total runoff and the forecasted total runoff using the best MARS model.

Table 2 | Statistical comparison of the best MARS, WA-ANN and ANN models for total runoff predicted from 10-fold cross validation

Model	MARS statistical results	WA-ANN statistical results	ANN statistical results
Best network structure	N/A	(7-7-1)	(6-7-1)
Training			
Maximum R^2	0.970	0.856	0.813
Minimum RMSE (mm)	0.250	0.750	0.550
Testing			
Maximum R^2	0.939	0.907	0.724
Minimum RMSE (mm)	0.292	0.423	0.624

Note: R^2: coefficient of determination, RMSE: root mean square error.
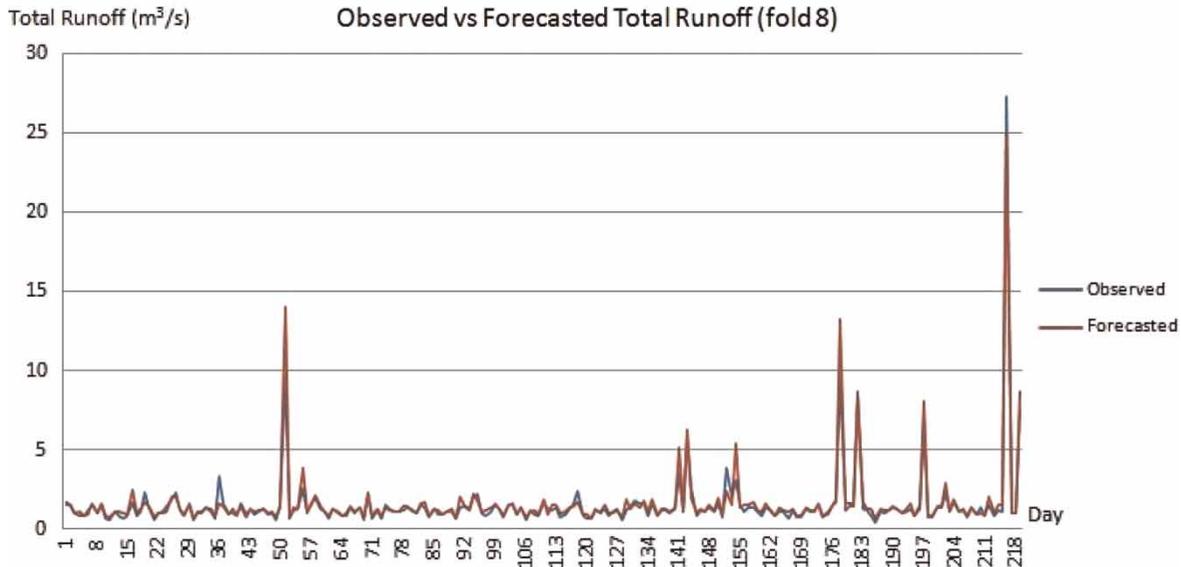
Figure 2 | Comparison of forecasted versus observed total runoff using the best testing MARS model (Fold 8).

ANN models

The results of the 10-fold cross-validation ANN modeling to forecast the total runoff using data from the three micro-watersheds are shown in Table 2. The best ANN model is a function of the rainfall from the previous day, the runoff number from the current day and the previous day, and the antecedent precipitation index from the current day, previous day and 2 days ago. This ANN model had seven neurons in the hidden layer. The best ANN model had a testing correlation coefficient (R^2) of 0.724. The minimum RMSE value of this ANN model for the testing period was 0.624. Figure 3 illustrates the observed total runoff and the forecasted total runoff using the best ANN model.
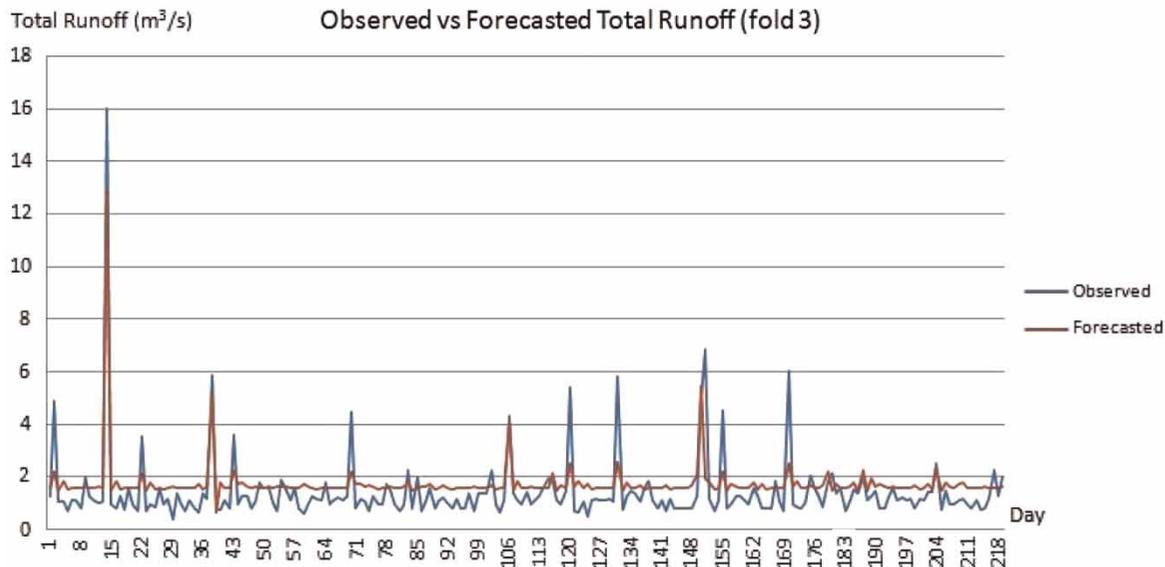

Figure 3 | Comparison of forecasted versus observed total runoff using the best testing ANN model (Fold 3).

WA-ANN models

The results of the 10-fold cross-validation WA-ANN model-ing to forecast the total runoff using data from the three watersheds are presented in Table 2. The best WA-ANN model is a function of the rainfall from the current day and the previous day, the runoff number from the previous day, and the antecedent precipitation from the current day, previous day, 2 days ago and 3 days ago. This WA-ANN model had seven neurons in the hidden layer. This WA-ANN model had a testing correlation coefficient (R^2) of 0.907. The minimum RMSE value of this WA-ANN model for the testing period was 0.423, which is low. Figure 4 illus-trates the observed total runoff and the forecasted total runoff using the best WA-ANN model.

Comparison of MARS, ANN and WA-ANN models

Table 2 shows that the best WA-ANN model had a similar level of accuracy to the best MARS model. The MARS model had a slightly better R^2 (0.939) than the WA-ANN model (0.907), and both the WA-ANN and MARS models had a significantly better R^2 than the best ANN model ($R^2 = 0.724$). The best MARS model had a testing RSME of 0.292, compared to the best

WA-ANN model (RMSE = 0.423) and the best ANN model (RMSE = 0.624). The lower RMSE value indicates that the best MARS model had smaller differences between the total forecasted runoff and the total observed runoff. Overall, it can be seen that the best WA-ANN and MARS models were comparable in terms of forecasting accuracy, with both model types providing very accurate runoff forecasts compared to the best regu-lar ANN model.

Figures 2–4 compare the observed and forecasted total runoff from the best MARS, ANN and WA-ANN models, respectively. It can be seen that the MARS model slightly under forecasts total runoff peaks, as does the WA-ANN model, but to a lesser degree. The ANN model tends to over-forecast average and low runoff, and under-forecasts runoff peaks. Overall, the MARS model provides closer esti-mates to the corresponding observed total runoff compared to the WA-ANN and ANN models.

Figures 5–7 are scatterplots that compare the observed and forecasted runoff using the best MARS model, the best ANN model and the best WA-ANN model during the test-ing period. It can be seen from these plots that the values for the MARS and WA-ANN models are less scattered and more centered around the 1:1 line than for the ANN model.
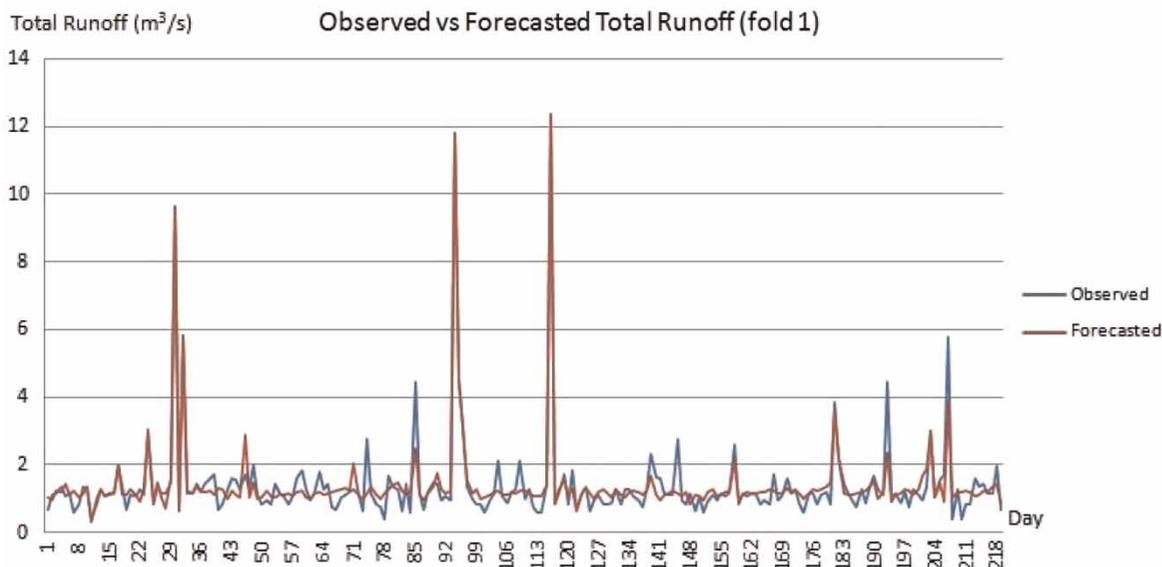

Figure 4 │ Comparison of forecasted versus observed total runoff using the best testing WA-ANN model (Fold 1).
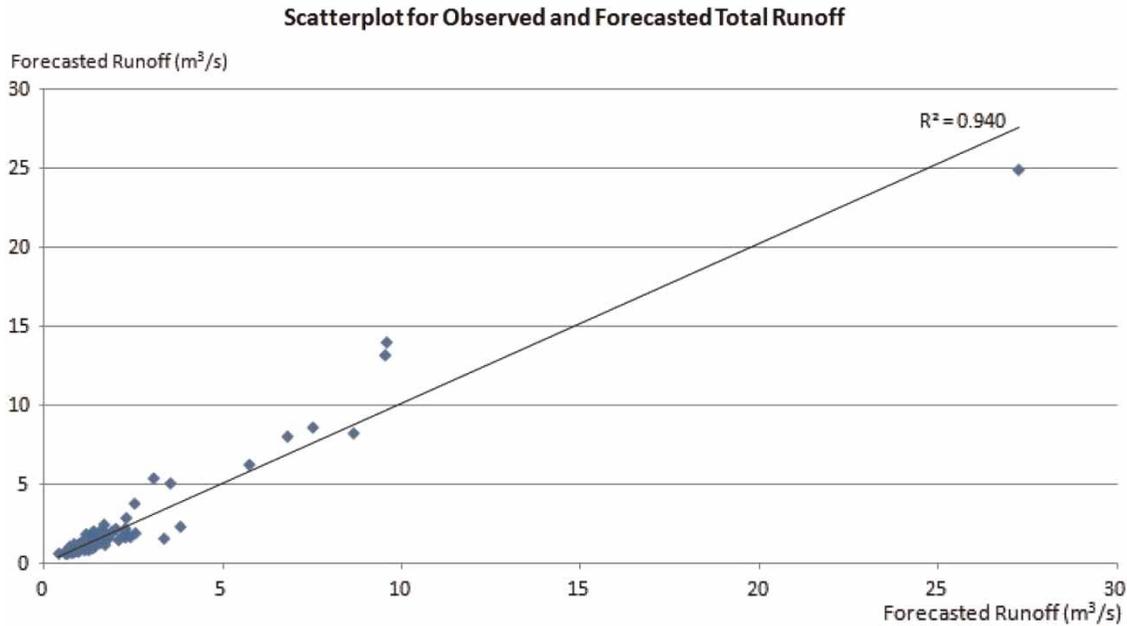
Figure 5 | Scatterplot comparing observed and forecasted total runoff using the best testing MARS model (Fold 8).
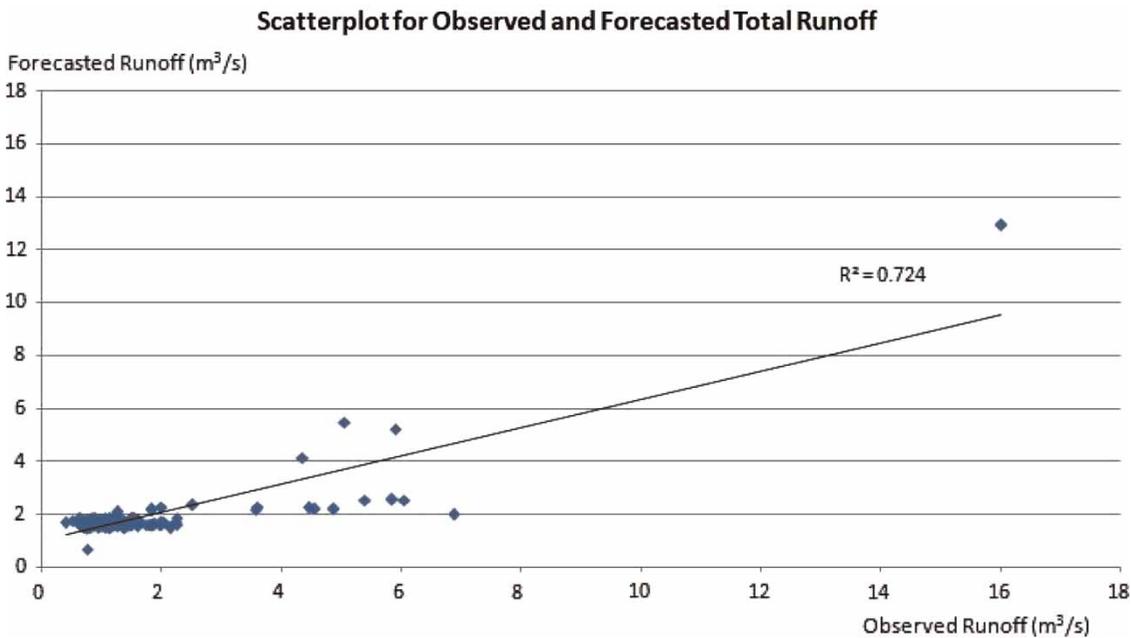

Figure 6 | Scatterplot comparing observed and forecasted total runoff using the best testing ANN model (Fold 3).

CONCLUSIONS

This research involved the development and testing of a novel hybrid model architecture (WA-ANN) for runoff forecasting in mountainous watersheds with limited data that attempted to draw on the strengths of two different modeling methods (wavelet analysis and artificial neural networks). The wavelet-neural network models were
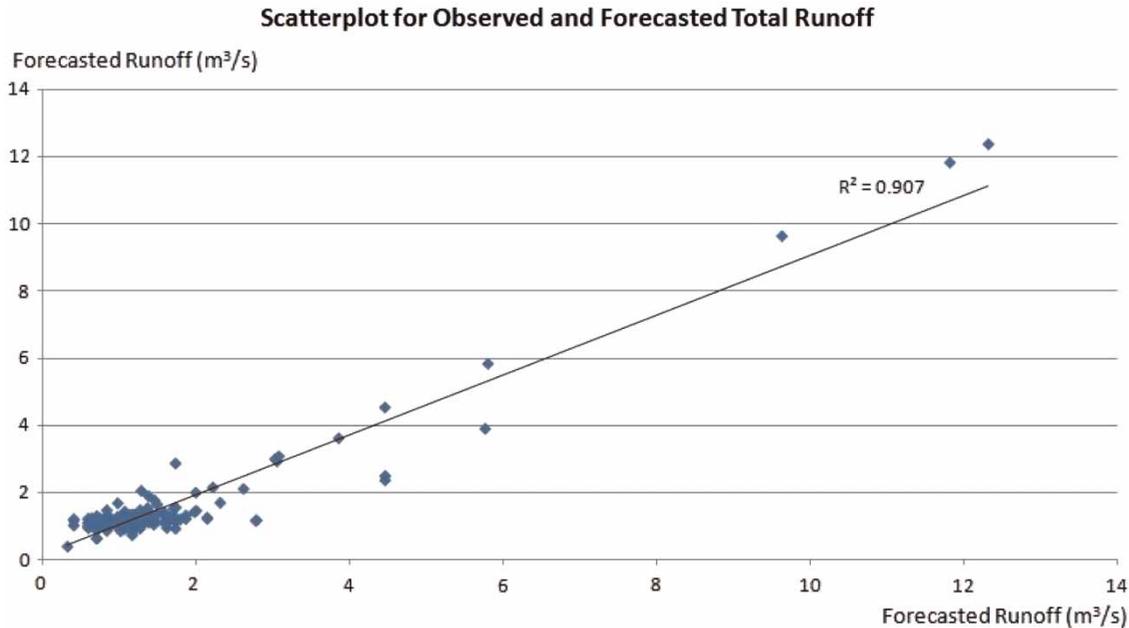
Figure 7 │ Scatterplot comparing observed and forecasted total runoff using the best testing WA-ANN model (Fold 1).

developed by combining discrete wavelet transforms and artificial neural networks. The wavelet-neural network models were compared with ANN models, as well as with MARS models developed by Sharda *et al.* (2006) for total runoff forecasting.

It was found that the best WA-ANN and MARS models were comparable in terms of forecasting accuracy, with both providing very accurate runoff forecasts compared to the best ANN model. This research found that the use of carefully selected sub-series, obtained via wavelet analysis, as inputs for ANN models results in very accurate runoff forecasts in mountainous watersheds with limited data. The results of this research study confirm the initial findings of several authors (Cannas *et al.* 2006; Kisi 2008, 2009; Partal 2009; Wu *et al.* 2009a; Adamowski & Sun 2010) who also found that wavelet-neural network models appear to be a promising new method of short-term runoff forecasting. It is hypothesized that the WA-ANN models are more accurate than the regular ANN models since wavelet transforms allow for the useful decomposition of original time series data, and the decomposed data can then be selectively used to develop artificial neural network forecasting models. This process allows some of the 'noisy' data to be removed.

It was found that both the MARS and WA-ANN methods do not require extensive input data sets for highly accurate runoff forecasting in mountainous watersheds, and as such both methods warrant additional research in other mountainous watersheds with limited data. Future suggested studies stemming from this research include: testing the application of WA-ANN and MARS models in a wide variety of different mountainous watersheds with limited data; testing the application of WA-ANN and MARS models using more recent data of a longer duration; investigating different modified mother wavelets for use in the coupled wavelet-neural network models; comparing the WA-ANN and MARS methods with other runoff forecasting methods such as support vector regression models; and exploring how to assess the uncertainty of WA-ANN and MARS forecasts.

ACKNOWLEDGEMENTS

assistance of Dr Patel and Mr Vaughn Grey of McGill is appreciated.

REFERENCES

Adamowski, J. 2008a Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. *Journal of Hydrology* **353**, 247–266.

Adamowski, J. 2008b Peak daily water demand forecast modeling using artificial neural networks. *Journal of Water Resources Planning and Management* **134** (2), 119–128.

Adamowski, J. 2008c River flow forecasting using wavelet and cross-wavelet transform models. *Journal of Hydrological Processes* **22**, 4877–4891.

Adamowski, J. & Karapataki, C. 2010 Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: Evaluation of different ANN learning algorithms. *Journal of Hydrologic Engineering* **15** (10), 729–743.

Adamowski, J. & Sun, K. 2010 Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology* **390**, 85–90.

Akhtar, M. K., Corzo, G. A., Van Andel, S. J. & Jonoski, A. 2009 River flow forecasting with artificial neural networks using satellite observed precipitation pre-processed with flow length and travel time information: case study of the Ganges river basin. *Hydrology Earth System Science* **13**, 1607–1618.

Andres, J. D., Lorca, P., de Cos Juez, F. J. & Sánchez-Lasheras, F. 2010 Bankruptcy forecasting: a hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS). *Expert Systems with Applications* **38**, 1866–1875.

Arnold, J. G. & Fohrer, N. 2005 SWAT 2000: current capabilities and research opportunities in applied watershed modeling. *Hydrological Processes* **19**, 563–572.

Balshi, M. S., McGuire, A. D., Duffy, P., Flannigan, M., Walsh, J. & Melillo, J. 2009 Assessing the response of area burned to changing climate in western boreal North America using a Multivariate Adaptive Regression Splines (MARS) approach. *Global Change Biology* **15**, 578–600.

Banerjee, P., Prasad, R. K. & Singh, V. S. 2009 Forecasting of groundwater level in hard rock region using artificial neural network. *Environmental Geology* **58** (6), 1239–1246.

Bera, P., Prasher, S. O., Patel, R. M., Madani, A., Lacroix, R., Gaynor, J. D., Tan, S. C. & Kim, S. H. 2006 Application of MARS in simulating pesticide concentrations in soil. *Transactions of the American Society of Agricultural Engineers* **49** (1), 297–307.

Bosch, D., Theurer, F., Binger, R., Felton, G. & Chaubey, I. 1998 Evaluation of the AnnAGNPS water quality model. *ASAE* Paper No. 98-2195.

Cannas, B., Fanni, A., See, L. & Sias, G. 2006 Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. *Physics and Chemistry of the Earth* **31**, 1164–1171.

Corani, G. & Guariso, G. 2005 Coupling fuzzy modelling and neural networks for river flood prediction. *IEEE Transactions on Man, Systems and Cybernetics* **35** (3), 382–391.

Dillaha, T., Wolfe, M., Shirmohammadi, A. & Byne, F. 1998 ANSWERS 2000. *The ASAE Annual Meeting*, Paper No. 98–2199.

Friedman, J. H. 1991 Multivariate adaptive regression splines (with Discussion). *Annals of Statistics* **19** (1), 1–141.

Han, D., Kwong, T. & Li, S. 2006 Uncertainties in real-time flood forecasting with neural networks. *Hydrological Processes* **21** (2), 223–228.

Han, D., Chan, L. & Zhu, N. 2007 Flood forecasting using support vector machines. *Journal of Hydroinformatics* **9** (4), 267–276.

Jain, A., Varshney, A. & Joshi, U. 2001 Short-term water demand forecast modeling at IIT Kanpar using artificial neural networks. *Water Resources Management* **15** (5), 299–321.

Kim, T. & Valdes, J. 2003 Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. *Journal of Hydrologic Engineering* **8** (6), 319–328.

Kisi, O. 2004 River flow modeling using artificial neural networks. *Journal of Hydrologic Engineering* **9** (1), 60–63.

Kisi, O. 2008 Stream flow forecasting using neuro-wavelet technique. *Hydrological Processes* **22**, 4142–4152.

Kisi, O. 2009 Neural networks and wavelet conjunction model for intermittent streamflow forecasting. *Journal of Hydrologic Engineering* **14** (8), 773–782.

Latinez-Sotomayor, K. A. 2010 Comparison of adaptive methods using multivariate regression splines (MARS) and artificial neural networks backpropagation (ANNB) for the forecast of rain and temperatures in the Mantaro river basin. *Hydrology Days* pp. 58–68.

Leathwick, J. R., Rowe, D., Richardson, J., Elith, J. & Hasti, T. 2005 Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshwater Biology* **50**, 2034–2052.

Leathwick, J. R., Elith, J. & Hasti, T. 2006 Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling* **199**, 188–196.

Lee, T. S., Chiu, C. C., Chou, Y. C. & Lu, C. J. 2006 Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics and Data Analysis* **50**, 1113–1130.

Maier, H. R., Jain, A., Dandy, G. C. & Sudheer, K. P. 2010 Methods used for the development of neural networks for the prediction of water resources variables: current status and future directions. *Environmental Modelling & Software* **25**, 891–909.

Murtagh, F., Starck, J. & Renaud, O. 2004 On neuro-wavelet modeling. *Decision Support Systems* **37**, 475–484.

Nourani, V., Alami, M. T. & Aminfar, M. H. 2009a A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. *Engineering Applications of Artificial Intelligence* **22** (3), 466–472.

Nourani, V., Komasi, M. & Mano, A. 2009b A multivariate ANN-wavelet approach for rainfall–runoff modeling. *Water Resource Management* **23** (14), 2877–2894.

Nourani, V., Kisi, O. & Komasi, M. 2011 Two hybrid Artificial Intelligence approaches for modeling rainfall–runoff process. *Journal of Hydrology* **402** (1–2), 45–59.

Partal, T. 2009 River flow forecasting using different artificial neural network algorithms and wavelet transform. *Canadian Journal of Civil Engineering* **36** (1), 26–38.

Piotrowski, A. P. & Napiorkowski, J. J. 2011 Optimizing neural networks for river flow forecasting – Evolutionary Computation methods versus the Levenberg–Marquardt approach. *Journal of Hydrology* **407** (1–4), 12–27.

Pramanik, N. & Panda, R. K. 2009 Application of neural network and adaptive neuro fuzzy inference systems for stream flow prediction. *Hydrological Sciences Journal* **54** (2), 247–260.

Raman, H. & Sunilkumar, N. 1995 Multivariate modelling of water resources time series using artificial neural networks. *Hydrology Science Journal* **40**, 145–163.

Renaud, O., Starck, J. & Murtagh, F. 2005 Wavelet-based combined signal filtering and prediction. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **35** (6), 1241–1251.

Sahoo, G. B. & Ray, C. 2006 Flow forecasting for a Hawaii stream using rating curves and neural networks. *Journal of Hydrology* **317**, 63–80.

Salford Systems 2001 *MARSTM User Guide*. Salford Systems, San Diego, CA.

Sharda, V., Prasher, S. O., Patel, R. M., Ojavasi, P. R. & Prakash, C. 2006 Modeling runoff from middle Himalayan watersheds employing artificial intelligence techniques. *Agriculture Water Management* **83**, 233–242.

Sharda, V. N., Prasher, S. O., Patel, R. M., Ojasvi, P. R. & Prakash, C. 2008 Performance of multivariate adaptive regression splines (MARS) in predicting runoff in mid-Himalayan micro-watersheds with limited data. *Hydrological Sciences Journal* **53** (6), 1165–1175.

Shrestha, R. R., Theobald, S. & Nestmann, F. 2005 Simulation of flood flow in a streamsystem using artificial neural networks. *Hydrology and Earth System Sciences* **9** (4), 313–321.

Solomatine, D. P. & Ostfeld, A. 2008 Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics* **10** (1), 3–22.

Tawfik, M. 2003 Linearity versus non-linearity in forecasting Nile River flows. *Advances in Engineering Software* **34** (8), 515–524.

Wang, W., Chau, K., Cheng, C. & Qui, L. 2009 A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology* **374**, 294–306.

Weiss, S. M. & Kulikowski, C. A. 1991 *Computer Systems That Learn*. Morgan Kaufmann Publishers, Inc., San Mateo, CA.

Wu, C. L. & Chau, K. 2011 Rainfall–runoff modeling using artificial neural networks coupled with singular spectrum analysis. *Journal of Hydrology* **399**, 394–409.

Wu, C. L., Chau, K. & Li, Y. S. 2009a Methods to improve neural network performance in daily flows prediction. *Journal of Hydrology* **372** (1–4), 80–93.

Wu, C. L., Chau, K. & Li, Y. S. 2009b Predicting monthly streamflow using data-driven models coupled with data pre-processing techniques. *Water Resources Research* **45** (8), WO8432–12.

Yang, C. C., Prasher, S. O., Lacroix, R. & Kim, S. H. 2003 A multivariate adaptive regression splines model for simulation of pesticide transport in soils. *Biosystems Engineering* **86** (1), 9–15.

Young, P. C. 1999 Nonstationary time series analysis and forecasting. *Progress in Environmental Science* **1**, 3–48.