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

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
\( \tilde{y}_i \) | mean value taken over \( N \)
\( y_i \) | observed peak weekly water demand
\( \hat{y}_i \) | forecasted peak weekly water demand
\( \tau \) | translation parameter
\( \dagger \) | 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 2005; Kisi 2004; Corani & Guaris 2005; Adamowski 2008a, c; Akhtar et al. 2009; Nourani et al. 2009a; Piotrowski & Napiorowski 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 preprocessing 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 Uttarakhand 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):

\[ Y = \max (0, X - c) \]  

(1)

\[ Y = \max (0, c - X) \]  

(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):

\[ \text{CWT}_j^y(\tau, s) = |s|^{1/2} \phi^* \left( \int_{-\infty}^{+\infty} \left( \frac{t - \tau}{s} \right) dt \right) \]  

(3)

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

\[ \psi_{j,k}(t) = s_j^{-1/2} \phi \left( \frac{t - k \tau s_j}{s_j} \right) \]  

(4)

In these equations, \( s \) is the scale parameter, \( \tau \) is the translation parameter, \( \psi(t) \) is the mother wavelet and transforming function, \( e^{i\omega} \) denotes the complex conjugate, \( j \) and \( k \) are integers and \( s_0 \geq 1 \) 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 2005; 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:

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

where $N$, $y_i$, $\hat{y}_i$, $\bar{y}$ 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:

$$\text{RMSE} = \sqrt{\frac{\text{SEE}}{N}}$$

where

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

and $N$ is the number of data points used. SEE is known as the sum of squared errors, with the variables as defined above. RMSE 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/km². 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.
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
this study), a 10-fold cross-validation procedure has been recommended to assess the generalization ability of the model (Weiss & Kulikowski 1994). 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

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

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

$$L = \text{int} \{ \log (M) \}$$  \hspace{1cm} (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>
<thead>
<tr>
<th>Model</th>
<th>MARS statistical results</th>
<th>WA-ANN statistical results</th>
<th>ANN statistical results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best network structure</td>
<td>N/A</td>
<td>(7-7-1)</td>
<td>(6-7-1)</td>
</tr>
<tr>
<td>Training Maximum $R^2$</td>
<td>0.970</td>
<td>0.856</td>
<td>0.813</td>
</tr>
<tr>
<td>Minimum RMSE (mm)</td>
<td>0.250</td>
<td>0.750</td>
<td>0.550</td>
</tr>
<tr>
<td>Testing Maximum $R^2$</td>
<td>0.939</td>
<td>0.907</td>
<td>0.724</td>
</tr>
<tr>
<td>Minimum RMSE (mm)</td>
<td>0.292</td>
<td>0.423</td>
<td>0.624</td>
</tr>
</tbody>
</table>

*Note: $R^2$: coefficient of determination, RMSE: root mean square error.*
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.
WA-ANN models

The results of the 10-fold cross-validation WA-ANN modeling 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 illustrates 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 \( \text{RMSE} = 0.423 \) and the best ANN model \( \text{RMSE} = 0.624 \). 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 regular 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 estimates 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 testing 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.
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
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**

This research was partially funded by an NSERC Discovery Grant (RGPIN 382650-10) held by Jan Adamowski. The
assistance of Dr Patel and Mr Vaughn Grey of McGill is appreciated.

REFERENCES


First received 8 April 2011; accepted in revised form 16 September 2011. Available online 11 October 2011.