Conjunction of SOM-based feature extraction method and hybrid wavelet-ANN approach for rainfall–runoff modeling

Vahid Nourani and Masoumeh Parhizkar

ABSTRACT

In rainfall–runoff modeling, the wavelet-ANN model, which includes a wavelet transform to capture multi-scale features of the process, as well as an artificial neural network (ANN) to predict the runoff discharge, is a beneficial approach. One of the essential steps in any ANN-based development process is determination of dominant input variables. This paper presents a two-stage procedure to model the rainfall–runoff process of the Delaney Creek and Payne Creek Basins, Florida, USA. The two-stage procedure includes data pre-processing and model building stages. In the data pre-processing stage, a wavelet transform is used to decompose the rainfall and runoff time series into several sub-series at different scales. Subsequently, independent sub-series are chosen via a self-organizing map (SOM). In the model building stage, selected sub-series are imposed as input data to a feed-forward neural network (FFNN) to forecast runoff discharge. To make a better interpretation of the model efficiency, the proposed model is compared with the Auto Regressive Integrated Moving Average with eXogenous input (ARIMAX) and with the *ad hoc* FFNN methods, without any data pre-processing. The results proved that the proposed model leads to better outcome especially in term of determination coefficient for detecting peak points (DC_{peak}).

Key words | clustering, feed-forward neural network, rainfall–runoff modeling, self-organizing map, wavelet transform

Vahid Nourani (corresponding author)
Masoumeh Parhizkar
Department of Water Resources Engineering,
Faculty of Civil Engineering,
University of Tabriz,
Iran
E-mail: *vnourani@yahoo.com*;
vnourani@umn.edu

INTRODUCTION

Accurate modeling of hydrological processes such as rainfall–runoff can be helpful in city planning, land uses, water resources management and environmental engineering and is of prime importance for hydrologists and environmental engineers. Therefore, many hydrological models have been developed in order to simulate such a complex process and a comprehensive classification of these models has been presented by Nourani *et al.* (2007).

Conventional time series models such as the Auto Regressive Integrated Moving Average with eXogenous input (ARIMAX) are used widely for hydrological time series forecasting (Salas *et al.* 1980; Nourani *et al.* 2011). These kinds of models, which are basically linear, lose their merit toward modeling hydrological processes that are embedded with high complexity, dynamism and nonlinearity in both spatial and temporal scales – but such models still may be employed for comparison and to evaluate the efficiency of the new developed models.

The artificial neural network (ANN), as a self-learning and self-adaptive approximator, has shown great ability in modeling and forecasting non-linear hydrologic time series. The ability of ANN to relate input and output variables in complex systems without any need for prior knowledge of the physics of the process plus its sufficiency in representing time-scale variability have led to a tremendous surge in use of ANN for rainfall–runoff modeling (see e.g. ASCE 2000b; Maier & Dandy 2000; Dawson & Wilby 2001; Jain & Srinivasulu 2006; Altunkaynak 2007; Iliadis & Maris 2007; Nourani *et al.* 2009a, 2011; Abrahart *et al.* 2012).

Despite the fact that ANN is a flexible tool for modeling hydrological time series, it includes some drawbacks when faced with a high non-stationary signal of a hydrologic

doi: 10.2166/hydro.2013.141

process that involves seasonalities that operate in a large range of scales that vary from 1 day to several decades. Therefore, time and/or space pre-processing of data in such a situation may be an effective approach to overcome the deficiencies. Potency of the wavelet transform in decomposing non-stationary time series into sub-series at different scales (levels) is helpful for better interpretation of the process and it is applied widely to time series analysis of non-stationary signals (Nason & Von Sachs 1999; Adamowski 2008; Kisi 2010; Mirbagheri *et al.* 2010). Therefore, the combination of ANN with wavelet transform as a hybrid wavelet-ANN (WANN) model that can explain simultaneously spectral and temporal information of the signal creates an effective implement for prediction of hydrological processes. The WANN model was firstly proposed by Aussem *et al.* (1998) for financial time series forecasting. Cannas *et al.* (2006) investigated the effect of data pre-processing on the model performance using continuous and discrete wavelet transforms and data partitioning. The results showed that networks trained with pre-processed data perform better than networks trained on un-decomposed, noisy, raw signals. Partal & Cigizoglu (2008) applied WANN to forecast daily suspended sediment load. The combination of ANN with wavelet transform has the potential for forecasting groundwater levels. Adamowski & Chan (2011) proposed a method based on coupling discrete wavelet transform and ANN for groundwater level forecasting. There are several papers that describe the application of WANN model to different fields of hydrology (see e.g. Partal & Kisi 2007; Nourani *et al.* 2009a, b, 2011). In the field of rainfall–runoff modeling, Remesan *et al.* (2009) applied the combination of wavelet transform with ANN for runoff prediction. Nourani *et al.* (2009a) combined wavelet analysis with the ANN concept to model the rainfall-runoff process and investigate the effect of the mother wavelet type on the model performance using four different mother wavelets. The modeling results showed the high merit of the Haar wavelet compared with the others (i.e. sym3, db2 and coif1). Tiwari & Chatterjee (2010) developed a hybrid wavelet–bootstrap–ANN (WBANN) model to explore the potential of wavelet and bootstrapping techniques for the development of an accurate and reliable ANN model for hourly flood forecasting. The results showed that the WBANN forecasting model with confidence intervals can improve the reliability of flood forecasting. Adamowski *et al.* (2012) developed multivariate adaptive regression spline, WANN and regular ANN models for runoff forecasting in a mountainous watershed with limited data. It was determined that the best WANN and multivariate adaptive regression spline models were comparable in terms of forecasting accuracy, and both could provide accurate runoff forecasts with regard to the ANN model.

In any ANN-based modeling, some of the input variables may be correlated, noisy or have no significant relationship with output variables and are not equally informative. Therefore, one of the essential steps in the ANN development process is to determine dominant input variables that are independent, informative and efficiently cover the proposed input domain. Although much information is included in the raw data, data pre-processing magnifies dominant features of the data and consequently the effect of data noise is diminished. More difficult learning, divergence, obscurity and poor model accuracy are some shortcomings that appear with the application of ANN without a proper data pre-processing method. In addition, high dimensionality included in most rainfall-runoff modeling slows the training and simulation of the process via the ANN. Therefore, when prediction of a process includes a long time series, the time required to perform the modeling becomes expensive. As similar as ANN, although WANN method can efficiently model hydrological processes, selection of effective inputs for such model is still a challengeable task, especially when several sub-series at different levels are obtained via the wavelet analysis and should be imposed to the ANN (Maheswaran & Khosa 2012; Nourani *et al.* 2012). Numerous available wavelet-based subsets as input variables forces us to examine several combinations of input variables (i.e. 2^d–1 input combinations can be selected from d input variables) and causes us to look for a robust technique to extract a dominant input combination. The clustering approach as a data mining technique can be utilized for this purpose. In the content of ANN-based hydrological modeling, clustering is usually employed for classification of the data, stations, zones, etc. into some homogeneous classes (see e.g. Wu *et al.* 2009; Chang *et al.* 2010; Nourani & Kalantari 2010; Tsai *et al.* 2012) and optimization of the model structure by

selection of dominant and relevant inputs (Bowden *et al.* 2005). The *K*-means clustering method is a simple clustering algorithm that classifies the data into *K* mutually exclusive clusters and has been applied to different fields of hydrology (see Chang *et al.* 2010; Nourani & Kalantari 2010). The *K*-means algorithm is linear inherent and requires the number of clusters to be specified in advance. These shortcomings reduce the capability of *K*-means in dealing with hydrological processes, which are usually involved with non-linear relationships among the variables. The self-organizing map (SOM) is a kind of unsupervised ANN method that has the authority to classify, cluster, estimate, predict and mine the data (Kalteh *et al.* 2008). It is an effective tool to convert the complex, non-linear, statistical relationship between high-dimensional data items into a simple, geometric relationship on a low-dimensional display in which similar variables are closer to each other in the grid than the more dissimilar variables. It also helps to a better understanding of data relationships due to its capability for visualization of data vectors. A main characteristic of the SOM method is its non-linearity and topologically preservation of the data structure through the algorithm (Corne *et al.* 1999; ASCE 2000a). Typical for a SOM is that the desired solution or targets are not given and the network intelligently learns to cluster the data by recognized different patterns (Kalteh *et al.* 2008). The SOM method was originally proposed by Kohonen *et al.* (1984) in speech recognition. Since Chon *et al.* (1996) firstly applied the SOM to pattern benthic communities in streams; SOM was used in water resources problems. Bowden *et al.* (2002) divided neural network data into training, testing and validation subsets using a SOM. Lin & Chen (2006) used SOM for the identification of homogeneous regions in regional frequency analysis of rainfall data. They compared efficiency of the SOM with regard to *K*-means and Ward's methods in accurately determining the clusters' membership. The results showed that the SOM can identify the homogeneous regions more accurately as compared with the other two clustering methods. Lin & Wu (2007) presented a SOM-based approach to estimate design hyetographs of un-gauged sites. They concluded that the approach performs better than methods based on conventional clustering techniques. Kalteh & Berndtsson (2007) interpolated monthly precipitation values by SOM and multi-layer perceptron (MLP) ANN and showed that the regionalization based on SOM performs better than the MLP. Kalteh *et al.* (2008) reviewed the published applications of SOM to different fields of water resources engineering and Cereghino & Park (2009) provided some additional inputs and further perspective to their review.

As mentioned previously, an important issue regarding the WANN application to hydrological simulation is how to select the dominant wavelet-based sub-series as the model inputs. For this purpose and as a classic methodology, computed linear correlation coefficient between input and output time series is usually employed (see e.g. Partal & Kisi 2007; Maheswaran & Khosa 2012). However, as criticized by Nourani *et al.* (2011, 2012), in spite of a weak linear relationship, a strong non-linear relationship may exist between two time series. In this paper, as an innovation in WANN modeling, a robust-intelligent algorithm is proposed by conjunction of the SOM concept with a WANN model. For this purpose firstly, the wavelet transform is used to decompose the main rainfall–runoff time series into several sub-series. These sub-signals are then clustered via a two-step SOM to choose independent and effective sub-series as input data for the feed-forward neural network (FFNN) model in order to forecast runoff value at 1 day ahead as well as at multi-step ahead. The sensitivity analysis is also applied on wavelet outputs to survey SOM accuracy in selecting dominant sub-series. To evaluate the model performance, the proposed wavelet-SOM-FFNN model (WSNN) is compared with *ad hoc* FFNN approach (without data pre-processing) and classic linear ARIMAX method.

In the next three sections, the concepts of wavelet transform, SOM and FFNN are briefly reviewed, respectively. In the two sections following those, the Efficiency Criteria and the Study Areas are introduced. Next, the models performances are evaluated and discussed. Concluding Remarks will be the final section of the paper.

WAVELET TRANSFORM

The wavelet transform has increased in usage and popularity in recent years. A comprehensive literature survey of wavelet in geosciences can be found in Foufoula-Georgiou & Kumar

(1995), and the most recent contributions are cited by Labat *et al.* (2000).

The time-scale wavelet transform of a continuous-time signal, $x(t)$, is defined as (Addison *et al.* 2001):

$$T(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} g^* \left(\frac{t-b}{a}\right) x(t) dt \quad (1)$$

where * corresponds to the complex conjugate and $g(t)$ is called wavelet function or mother wavelet. The parameter a acts as a dilation factor, while b corresponds to a temporal translation of the function $g(t)$, which allows the study of the signal around b. The main property of wavelet transform is to provide a time-scale localization of process, which derives from the compact support of its basic function. This property is opposed to the classical trigonometric function of Fourier analysis. The wavelet transform searches for correlations between the signal and wavelet function.

For practical applications, the hydrologists do not have at their disposal a continuous-time signal process but rather a discrete time signal. A discrete mother wavelet has the form (Addison *et al.* 2001):

$$g_{m,n}(t) = \frac{1}{\sqrt{a_0^m}} g\left(\frac{t - nb_0 a_0^m}{a_0^m}\right) \quad (2)$$

where m and n are integers that control the wavelet dilation and translation respectively; a_0 is a specified fined dilation step greater than 1; and b_0 is the location parameter and must be greater than zero. The most common and simplest choice for parameters are $a_0 = 2$ and $b_0 = 1$.

This power-of-two logarithmic scaling of the translation and dilation is known as the dyadic grid arrangement. The dyadic wavelet can be written in more compact notation as (Addison *et al.* 2001):

$$g_{m,n} = 2^{-m/2} g(2^{-m} i - n) \quad (3)$$

For a discrete time series, x_i, the dyadic wavelet transform becomes (Nourani 2009):

$$T_{m,n} = 2^{-m/2} \sum_{i=0}^{N-1} g(2^{-m} i - n) x_i \quad (4)$$

where $T_{m,n}$ is wavelet coefficient for the discrete wavelet of scale $a = 2^m$ and location $b = 2^m n$. Equation (4) considers a finite time series, x_i, $I = 0, 1, 2, \ldots, N$–1; and N is an integer power of 2 so that $N = 2^M$. This gives the ranges of m and n as $0 < n < 2^{M-m} - 1$ and $1 < m < M$, respectively.

The inverse discrete transform is given by (Nourani 2009):

$$x_i = \overline{T} + \sum_{m=1}^{M} \sum_{n=0}^{2^{M-m}} T_{m,n} 2^{-m/2} g(2^{-m} i - n) \quad (5)$$

Or in a simple format as (Nourani 2009):

$$x_i = \overline{T} + \sum_{m=1}^{M} W_m(t) \quad (6)$$

which $\overline{T}$ is called approximation sub-series at level M and $W_m(t)$ are details sub-series at levels $m = 1, 2, \ldots, M$.

The wavelet coefficients, $W_m(t)$ ($m = 1, 2, \ldots, M$), provide the detail signals, which can capture small features of interpretational value in the data; the residual term, $T(t)$, represents the background information of data.

SELF-ORGANIZING MAP (SOM)

The SOM is an effective software tool for the visualization of high-dimensional data. It implements an orderly mapping of a high-dimensional distribution onto a regular low-dimensional grid. Thereby, it is able to convert complex, non-linear statistical relationships between high-dimensional data items into simple geometric relationships on a low-dimensional display while preserving the topology structure of the data (Bowden *et al.* 2002). The way SOMs go about reducing dimensions is by producing a map of usually one or two dimensions that plot the similarities of the data by grouping similar data items together. Thus, SOMs accomplish two things; they reduce dimensions and display similarities. The basic SOM network consists of two layers, an input layer and a Kohonen layer, which in most common applications is two dimensional. In the input layer, one neuron is dedicated for each variable. The

Kohonen layer neurons are related to every neuron of the input layer via adjustable weights. A two-level SOM neural network is a better approach to catch a preliminary overview on intricate data set. It augments the conventional SOM network with an additional one-dimensional Kohonen layer in which each neuron is connected to neurons in the previous Kohonen layer. The schematic view of the two-level SOM network is shown in Figure 1.

It is important for every variable to have equal impact as compared with other variables. Therefore, normalization of data is an important step in data clustering. Converting the data to the range of, for example, 0–1 prevents a variable being more important when compared with others. The SOM is trained iteratively. It is recommended that the number of iteration should be at least 500 times the number of neurons in the Kohonen layer (Haykin 1999; Kohonen 2001). Initially the weights are assigned randomly. When the n-dimensional input vector x is sent through the network, the distance between the weight w neurons of SOM and the inputs is computed. The most common criterion to compute the distance is Euclidean distance (Bowden *et al.* 2002):

$$\|X - w\| = \sqrt{\sum_{i=1}^{n} (X_i - w_i)^2} \tag{7}$$

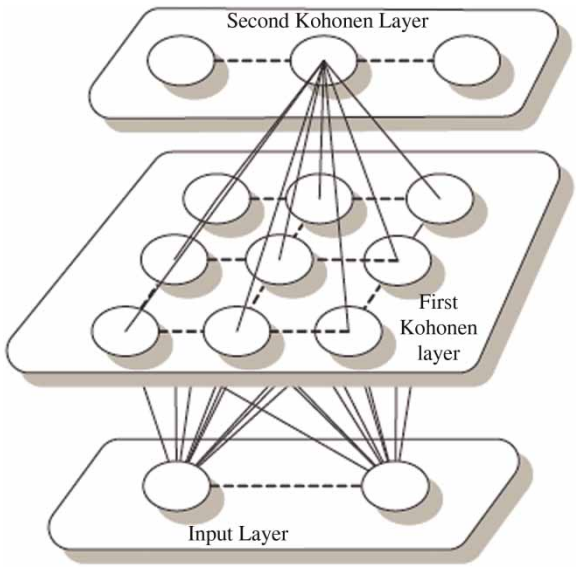

Figure 1 | Architecture of the two-level self-organizing map (SOM) neural network (Hsu & Li 2010).

The weight with the closest match to the presented input pattern is called winner neuron or best matching unit (BMU). The BMU and its neighboring neurons are allowed to learn by changing the weights at each training iteration t, in a manner to further reduce the distance between the weights and the input vector (Bowden *et al.* 2002):

$$w(t+1) = w(t) + \alpha(t) h_{lm}(x - w(t)) \tag{8}$$

where α is the learning rate, ranging in [0 1], l and m are the positions of the winning neuron and its neighboring output nodes and h_{lm} is the neighborhood function. The most commonly used neighborhood function is the Gaussian (Bowden *et al.* 2002):

$$h_{lm} = \exp\left(-\frac{\|1 - m\|^2}{2\sigma(t)^2}\right) \tag{9}$$

where h_{lm} is the neighborhood function of the best matching neuron l at iteration t; and l–m is the distance between neurons l and m on the map grid; and σ is the width of the topological neighborhood. The training steps are repeated until convergence. After the SOM network is constructed, the homogeneous regions, i.e. clusters, are defined on the map.

FEED-FORWARD NEURAL NETWORK (FFNN)

The ANN is widely applied in hydrology and water resource studies as a forecasting tool. In an ANN, feed-forward back-propagation (FFBP) networks are common to engineers. It has been proved that the FFBP network model with three layers is satisfied for the forecasting and simulation of any engineering problem (Hornik 1989; ASCE 2000a; Nourani *et al.* 2008).

As shown in Figure 2, three-layered FFNNs, which have been usually used in forecasting hydrologic time series, provide a general framework for representing non-linear functional mapping between a set of input and output variables. Three-layered FFNNs are based on a linear combination of the input variables, which are transformed by a non-linear activation function.

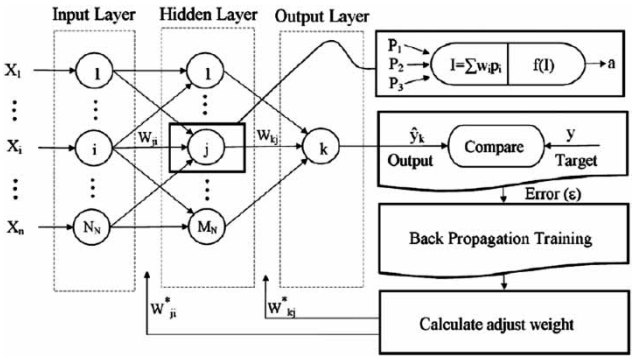

Figure 2 | A three-layered feed-forward neural network with the back propagation (BP) training algorithm.

In Figure 2, i, j and k denote input layer, hidden layer and output layer neurons, respectively and w is the applied weight by the neuron. The term 'feed forward' means that a neuron connection only exists from a neuron in the input layer to other neurons in the hidden layer or from a neuron in the hidden layer to neurons in the output layer and the neurons within a layer are not interconnected to each other. The explicit expression for an output value of a three-layered FFNN is given by (Nourani *et al.* 2011):

$$\hat{y}_k = f_0\left[\sum_{j=1}^{M_N} w_{kj} \cdot f_h\left(\sum_{i=1}^{N_N} w_{ij}x_i + w_{jo}\right) + w_{ko}\right] \tag{10}$$

where w_{ji} is a weight in the hidden layer that connects the ith neuron in the input layer and the jth neuron in the hidden layer, w_{jo} is the bias for the jth hidden neuron, f_h is the activation function of the hidden neuron, w_{kj} is a weight in the output layer that connects the jth neuron in the hidden layer and the kth neuron in the output layer, w_{ko} is the bias for the kth output neuron, f_o is the activation function for the output neuron, x_i is ith input variable for input layer and $\hat{y}_k$, y are computed and observed output variables, respectively. N_N and M_N are the number of the neurons in the input and hidden layers, respectively. The weights are different in the hidden and output layers, and their values can be changed during the process of the network training.

EFFICIENCY CRITERIA

To perform the FFNN model, the rainfall and runoff data sets were split into calibration and verification subsets. To achieve better performance it was tried to put time series extreme values in a calibration data set. For this purpose, the first 25% of total data were used as the verifying set and the remaining 75% were used for training the proposed model. The time series before going through the network were normalized between 0 and 1. A proper model yields comparatively good results in terms of determination coefficient and root mean squared error (RMSE) in training and verification steps. Consequently, to assess model efficiency, the following measurements were used to compare the performance of various models (Nourani 2009):

$$\mathrm{DC} = 1 - \frac{\sum_{i=1}^{N}\left(O_{obs_i} - O_{com_i}\right)^2}{\sum_{i=1}^{N}\left(O_{obs_i} - \overline{O}_{obs_i}\right)^2} \tag{11}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\left(O_{obs_i} - O_{com_i}\right)^2}{N}} \tag{12}$$

where DC, RMSE, N, O_{obs_i}, O_{com_i} and $\overline{O}_{obs}$ are determination coefficient, root mean squared error, number of observations, observed data, computed values and mean value of observed data, respectively. DC evaluates model accuracy by comparing observed data and predicted values (Equation (11)). RMSE is used to measure the discrepancy between observed data on which the model is developed and predicted values that are created via the model (Equation (12)). A high value for DC (up to one) and a small value for RMSE in both training and validation steps indicate high efficiency of the model. Legates & McCabe (1999) showed that a hydrological model can be evaluated sufficiently using these two statistical values.

Also, due to the importance of extreme values of discharge in the rainfall–runoff simulation for water resources management and flood mitigation purposes, Equation (13) can be used to compare the ability of different models to capture the peak values of a runoff time series (Nourani *et al.* 2011):

$$\mathrm{DC}_{\mathrm{peak}} = 1 - \frac{\sum_{i=1}^{n}\left(Q_{PC_i} - Q_{PO_i}\right)^2}{\sum_{i=1}^{n}\left(Q_{PO_i} - \overline{Q}_{PO}\right)^2} \tag{13}$$

where, $\mathrm{DC}_{\mathrm{peak}}$ is the determination coefficient for peak values, n refers to number of peak values

Q_{PO_i}, Q_{PC_i} and $\overline{Q}_{PO}$ are observed data, computed values and mean of observed runoff peak values, respectively.

STUDY AREA AND DATA

In order to evaluate the performance of the proposed model, two study areas (i.e. the Delaney Creek and Payne Creek sub-basins) with various topological characteristics were used in this study. The characteristics of the watersheds are introduced in cases (1) and (2), comprehensively.

Case study (1): Delaney Creek sub-basin

The Delaney Creek sub-basin in Tampa Bay Watershed was selected as the first study area in the current research. The basin is located at Florida State between 27°52′ and 27°56′ north latitude and 82°22′ and 82°24′ west longitude and its drainage area is about 42 km^2 of open water, which drains to Tampa Bay on the Gulf of Mexico, including the area draining from Gasparilla Pass and the watershed of Hillsborough Bay. The city of Tampa and the southern portion of the metropolitan Tampa Bay area are within the watershed (Figure 3).

The Tampa Bay Watershed contains some of the most productive agricultural lands of the state. This watershed is fairly flat and its elevation varies between 10 m above and below sea level. The climate of the region is subtropical and exhibits a transitional pattern from continental to tropical Caribbean. Long, warm and humid summers are typical as well as mild, dry winters. The annual average temperature and the total yearly rainfall are about 23 °C and 1,350 mm, respectively. The observed daily stream flow and rainfall time series of Delaney Creek Station, where the outlet of the Delaney Creek Watershed is, were retrieved via the United States Geological Survey (USGS) website and used in this study. Time series included 6,403 daily data observations from August 1993 to December 2011. The observed data from August 1993 to December 1997 were used as the verification set and the remaining data were employed for training the model.

Case study (2): Payne Creek sub-basin

The Payne Creek sub-basin located in the Peace Tampa Bay watershed in Florida was selected as the second study area in this research (Figure 4). The Peace Tampa Bay watershed connects central Florida to the southwest coast and consists of nine sub-basins. The Payne Creek sub-basin is the second smallest basin in the watershed located at the northwest

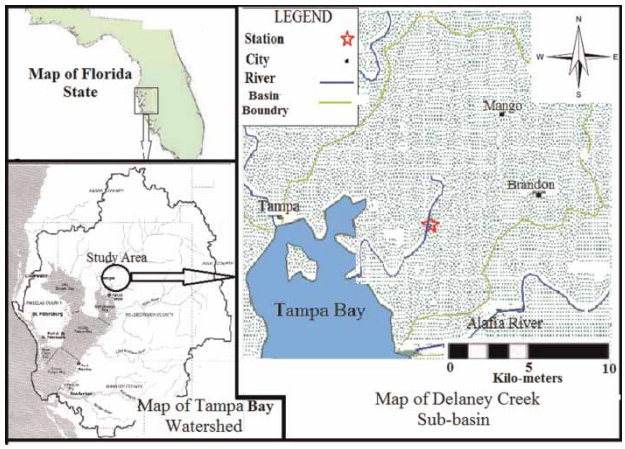

Figure 3 | Study area (Delany Creek sub-basin).

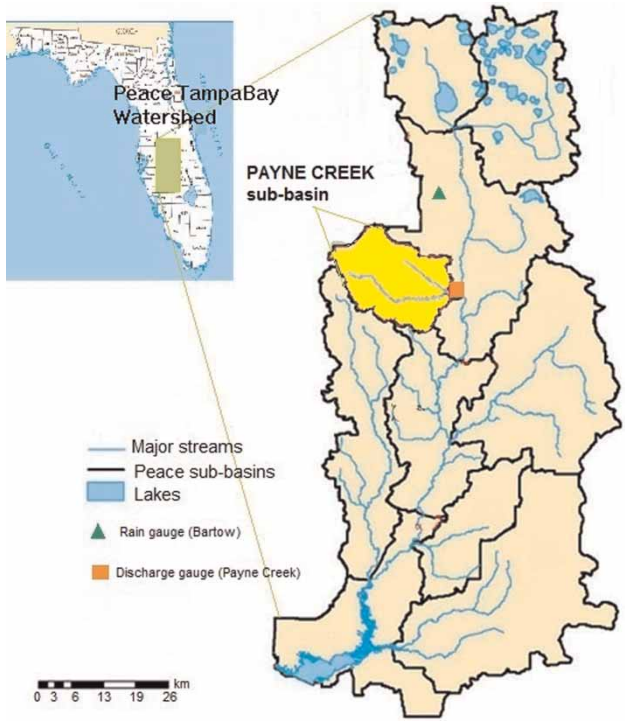

Figure 4 | Study area (Payne Creek sub-basin).

portion of the Peace River watershed. The Payne Creek Basin covers a 322 km^2 area. The Payne Creek River flows through the sub-basin with annual mean flow of 2 cm. The climate of the area is generally subtropical with an annual average temperature of about 23 °C. Annual average rainfall over or around the Payne Creek sub-basin is 1,270–1,420 mm. The developed models require discharge and rainfall time series for training and testing purposes. The observed daily stream flow values of Payne Creek station and rainfall data of Bartow station were used in this study (see Figure 4). Bartow station is located upstream of the sub-basin. The location of discharge gauge is in the middle part of the watershed. Changes in elevation are most conspicuous along the ridges and scarps. The elevation variation between the upstream and middle part of the watershed indicates the 'sloppy' situation (i.e. the watershed includes steep slopes) of the study area.

The rainfall and runoff time series included 5,841 daily data observations from July 1995 to July 2011. The observed data from July 1995 to September 2008 were used for the training set and the remaining data were employed as the verification set. The statistics of the rainfall and runoff time series for both study areas have been tabulated in Table 1.

RESULTS AND DISCUSSION

The increase of wavelet-based sub-signals as WANN inputs may lead to an essential deterioration in the model that is usually reflected in the network overfitting. Hence, data pre-processing (i.e. combination of wavelet and SOM in this research) is a crucial step in the modeling. In order to optimize the input layer and improve the efficiency of the FFNN-based rainfall–runoff model of the watersheds, temporal pre-processing of time series was performed using a combination of wavelet transform and SOM clustering method. Schematic diagram of the proposed model is shown by Figure 5. The details of the proposed methodology and results are presented in the following sub-sections. The methodology was initially explained for the Delany Creek sub-basin. Subsequently, to have a better evaluation of the proposed methodology, WSNN approach was also applied on rainfall and runoff time series for the Payne Creek sub-basin.

Results of proposed WSNN model

Elucidation of time series simultaneously in both spectral and temporal terms helps FFNN for better interpretation of the process. The wavelet transform was used to decompose rainfall and runoff time series at level 5 into six sub-series (one approximation and five details), the optimum decomposition level obtained through trial–error procedure. In order to have a comprehensive overview on decomposition level, initially the following formula (which offers a minimum level of decomposition) was employed (Aussem *et al.* 1998; Nourani 2009):

$$L = \text{int}[\log(N)] \tag{14}$$

where L and N are decomposition level and number of time series data, respectively. For the first case study $N = 6{,}403$, thus $L = 3$. This experimental equation was derived for fully autoregressive signals, and only considering time series length without paying attention to seasonal signatures of a hydrologic process (Nourani *et al.* 2011). Hence, by application of decomposition level 3, only a few seasonalities in the main time series might be taken into account. Decomposition level 3 yields three detailed sub-series

Table 1 | Statistics of rainfall and runoff data for the study areas

		Rainfall time series (mm)				Runoff time series (m^3/s)			
Study area	Data set	Max.	Min.	Mean	Standard deviation	Max.	Min.	Mean	Standard deviation
Delaney Creek	Calibration	154.18	0	3.49	10.78	16.46	0	0.27	0.53
	Verification	81.78	0	2.49	8.90	3.75	0	0.23	0.68
Payne Creek	Calibration	158.8	0	3.60	11.2	77.84	0	3.80	6.06
	Verification	94.00	0	2.80	8.70	15.20	0	2.20	2.70

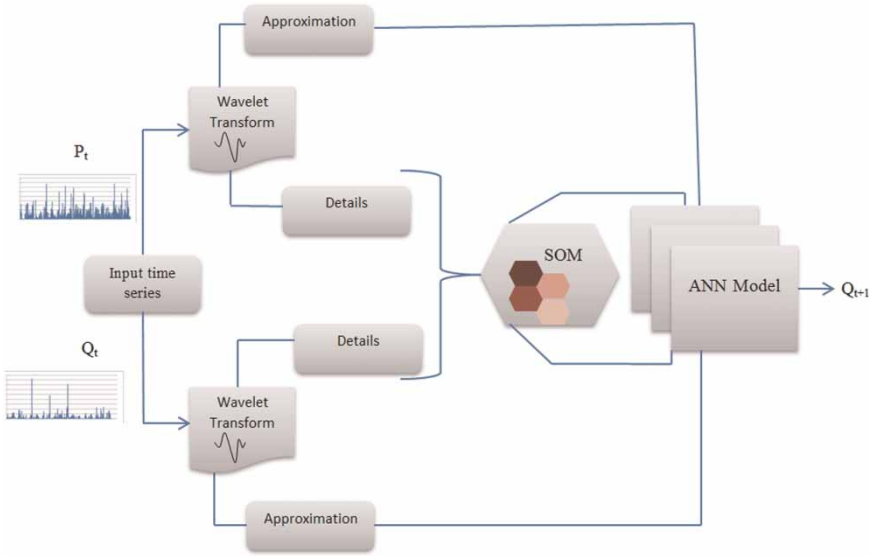

Figure 5 | Schematic diagram of the proposed methodology.

(i.e., 2^1-day mode, 2^2-day mode, 2^3-day mode, which is nearly a weekly mode), but according to the hydrological base of the time series, there might be other dominant seasonalities with longer periods. Therefore, two other decomposition levels (i.e., 5 and 7) were also examined, in which decomposition level 5 led to better results via the presented modeling and was selected as the optimum decomposing level. Level 5 sub-series contains five details as 2^1-day mode, 2^2-day mode, 2^3-day mode, which is nearly weekly mode, 2^4-day mode and 2^5-day mode, which is nearly monthly mode. Therefore, the seasonality of the process up to 1 month could be handled by the model. Due to proportional relationship between amount of rainfall and runoff, these signals were supposed to have the same seasonality level and both time series were decomposed at same level (i.e. level 5). Daubechies-2 (db2), Meyer and coif2 mother wavelets were applied to decompose both rainfall and runoff time series. To investigate the effect of form similarity between mother wavelet and main time series, another selection of mother wavelets were also examined. For this purpose, two other mother wavelets (Haar and Daubechies-4, db4) were chosen according to the formation of main signals. As the Haar wavelet has a pulsed shape, it could properly capture the signal features of rainfall time series and may yield comparatively high efficiency (Nourani 2009). Conversely, there are many jumps in the runoff time series because of sudden start and cessation of rainfall over the related watershed. Therefore, due to the formation of adb4 wavelet that is similar to the runoff signal, it could capture the signal features, especially peak points, efficiently and led to comparatively good results. Thus, Haar and db4 mother wavelets were also used to decompose rainfall and runoff time series, respectively. Mallat (1998) can be referred to for more information on mathematical concepts of the mother wavelets. For instance, Figure 6 shows approximation and details sub-series of rainfall time series decomposed by the Haar mother wavelet at level 5. For application of discrete wavelet transform, a code was developed in the MATLAB environment (MathWorks 2010a). Boundary (edge) effect is one of the deficiencies in the application of wavelet transform that happens due to application of the wavelet to the beginning and end of the time series (signals) at which point there are no data before and after. As a solution, the Zero Padding method was used (Addison *et al.* 2001). Therefore, as the time series were long enough for two case studies, suitable amounts of data were neglected from the beginning and end parts of time series after wavelet application.

The increase of sub-series as inputs of FFNN may lead to network overfitting, divergence, obscurity and poor accuracy. Therefore, to optimize the number of input and improve the model training rate and efficiency, we tried to

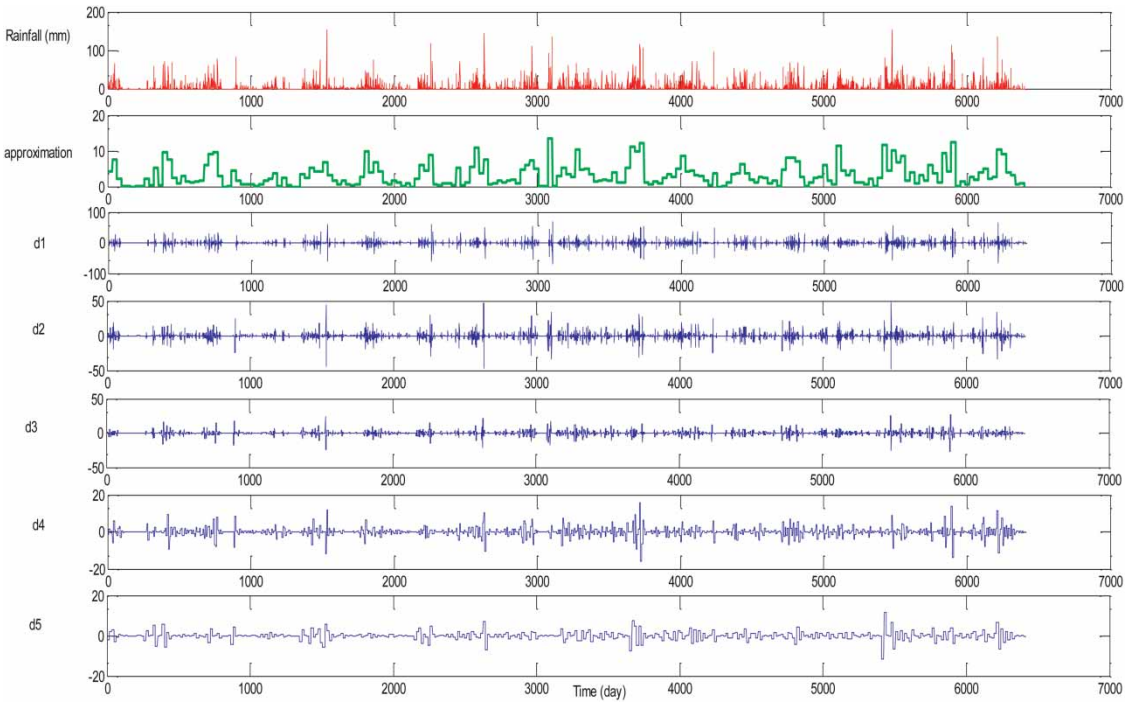

Figure 6 | Approximation and details sub-signals of rainfall time series decomposed at level 5 for the Delaney Creek sub-basin.

choose a dominant sub-series from the decomposed rainfall and runoff time series prior to imposition into the FFNN model. In conventional trial and error methodology, for 10 details as potential inputs, $2^{10}-1$ input combinations (for d inputs, 2^d-1 combinations can be assigned) should be examined. In a complex hydrological system with large value of d, an efficient algorithm, instead of a conventional trial and error method, is needed to select dominant input variables. The problem further intensifies in time series, in which appropriate lags must also be considered. As SOM compresses information while preserving the most important topological and metric relationships of the primary data items on the display, it can be an effective tool to extract dominant features of a process. Therefore, sub-series obtained by wavelet transform were imposed to SOM to be clustered into several groups, as similar sub-series were stood in the same cluster. Centers of clusters that well represented the cluster patterns were imposed to the FFNN as model inputs. For this purpose, a two-step SOM clustering method was employed to select the effective sub-series and reduce the dimensionality of the input space. At the first step, a two-dimensional SOM was applied to have an overview on signals patterns and approximate number of clusters to be assigned, regarding the SOM topology. Subsequently, in the second step, in order to be ensured of the highlighted clusters, a one-dimensional SOM was applied to classify the signals with specific numbers of groups determined at the first step. Afterwards, the Euclidean distance criterion was utilized to select the centroid signal of each cluster that was the best representative of the data pattern within the cluster. For application of a SOM on a data set, the SOM Toolbox from MATLAB was used to cluster the data (MathWorks 2010b). Detailed sub-series were imposed to the SOM in order to extract dominant details that can have a significant role in attaining accurate model results. After decomposition of the Delany Creek time series at level 5 and in order to apply the proposed two-step SOM, the size of Kohonen layer was considered to be 4×4 for the first step. As the number of detailed sub-series was 10 (five rainfall detailed sub-series in addition to five runoff detailed sub-series), the mentioned size was large enough to ensure that a suitable number of clusters are formed from the training data. Figure 7 presents the resulted neighbor weighted distances of the output layer size 4×4 which contains detailed sub-series obtained by decomposing through Haar-db4 mother wavelets. The

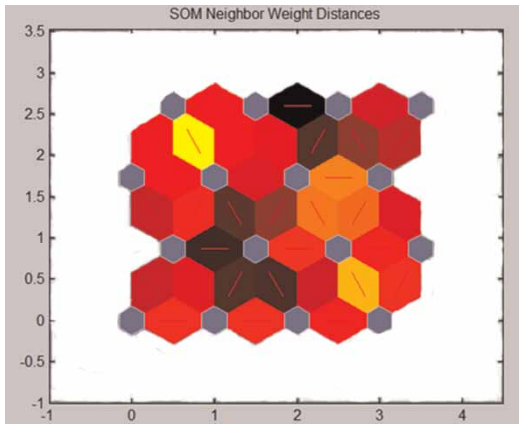

Figure 7 | Neighbor weighted distances obtained by two-dimensional (2D) self-organizing map (SOM) for data from the Delaney Creek sub-basin.

neighbor weight distances' plan presents output neurons and their direct neighbor relationship. The regular hexagonals display the SOM output neurons while the stretched hexagonals indicate the distances between neurons. The darker colors demonstrate larger distances, and lighter colors refer to smaller distances. Figure 8 shows the hits plan of the output layer size of 4×4 that contains detailed sub-series obtained by decomposing through Haar-db4 mother wavelets. The hits plan is an illustration of a SOM output layer, with each neuron showing the number of classified input vectors. The relative number of vectors for each neuron is shown via the size of a colored patch.

According to obtained neighbor weighted distances (Figure 7), darker hexagonals divided the Kohonen layer approximately into four parts, in the two-dimensional SOM. To be assured of the optimized clustering result, the one-dimensional SOM was also examined in the second step (i.e., 1×3, 1×4, 1×5 and 1×6 Kohonen layers). Obtained results via the FFNN models in the next stage proved that output layer of size 1×4 could lead to better results and four clusters was considered as the optimum number of clusters.

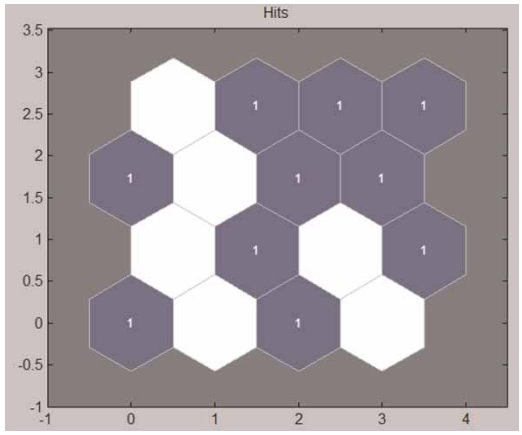

Figure 8 | Hits plan obtained by two-dimensional (2D) self-organizing map (SOM) for data from the Delaney Creek sub-basin.

The centroid of each cluster was selected using the Euclidean distance criterion and assigned as the representative of the cluster. Table 2 presents clustering patterns and selected sub-series for each of applied mother wavelets. Detailed sub-series of rainfall at levels 1 and 2 were usually grouped in a same cluster with runoff detailed sub-series at same seasonalities (i.e., scales of 2^1 days and 2^2 days of rainfall and runoff time series are related). Detailed sub-series of runoff are more effective than rainfall sub-series in runoff prediction, which were truly selected by SOM in this study. In order to highlight the dynamic entities in time series via wavelet continuous wavelet transform (CWT) was applied on a data set using the MATLAB toolbox. The scalogram of runoff for last 600 days, extracted by the db4 mother wavelet, is shown in Figure 9. A scalogram is a visual method to display a wavelet transform. There are three axes: x represents time, y represents seasonalities, and z represents coefficient value. The z axis is often shown by varying the color or brightness. CWT results confirmed the consequences obtained via the proposed methodology. For instance, 16-day mode was interpreted as an initial seasonality in the scalogram, which is in agreement with d_{4q} obtained using a DWT (detailed sub-series) and selected via a SOM (see Table 2).

Consequently, two approximation sub-signals and four details, as the centroids of the classes, obtained by SOM might efficiently constitute dominant inputs of FFNN model for the study case of the Delaney Creek sub-basin.

The proposed hybrid model contains numerous parameters such as number of hidden neurons, training iteration number (epoch), number of SOM clusters, decomposition levels and mother wavelet type; their appropriate determination and selection may lead to improvement of the model efficiency.

Assigned inputs were then imposed to a FFNN model to forecast 1-day-ahead runoff value. Selected input variables

Table 2 | SOM clustering for the Delaney Creek sub-basin

	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
Mother wavelet	**Data**	**Center**	**Data**	**Center**	**Data**	**Center**	**Data**	**Center**
Haar-db4	*d1p, d2p, d3p, d4p, d1q, d2q*	*d2q*	*d3q, d5q*	*d3q*	*d4q*	*d4q*	*d5p*	*d5p*
coif2	*d1p, d2p, d3p, d1q, d2q, d3q*	*d3q*	*d4p, d4q*	*d4q*	*d5q*	*d5q*	*d5p*	*d5p*
db2	*d3p, d4p, d1q, d4q*	*d4q*	*d2q, d3q, d5q*	*d3q*	*d1p, d2p*	*d1p*	*d5p*	*d5p*
Meyer	*d3p, d2q, d3q, d4q*	*d3p*	*d1p, d2p, d1q*	*d1q*	*d5p, d5q*	*d5q*	*d4p*	*d4p*

d_{ip}: detail sub-series of rainfall in level *i*.
d_{iq}: detail sub-series of runoff in level *i*.

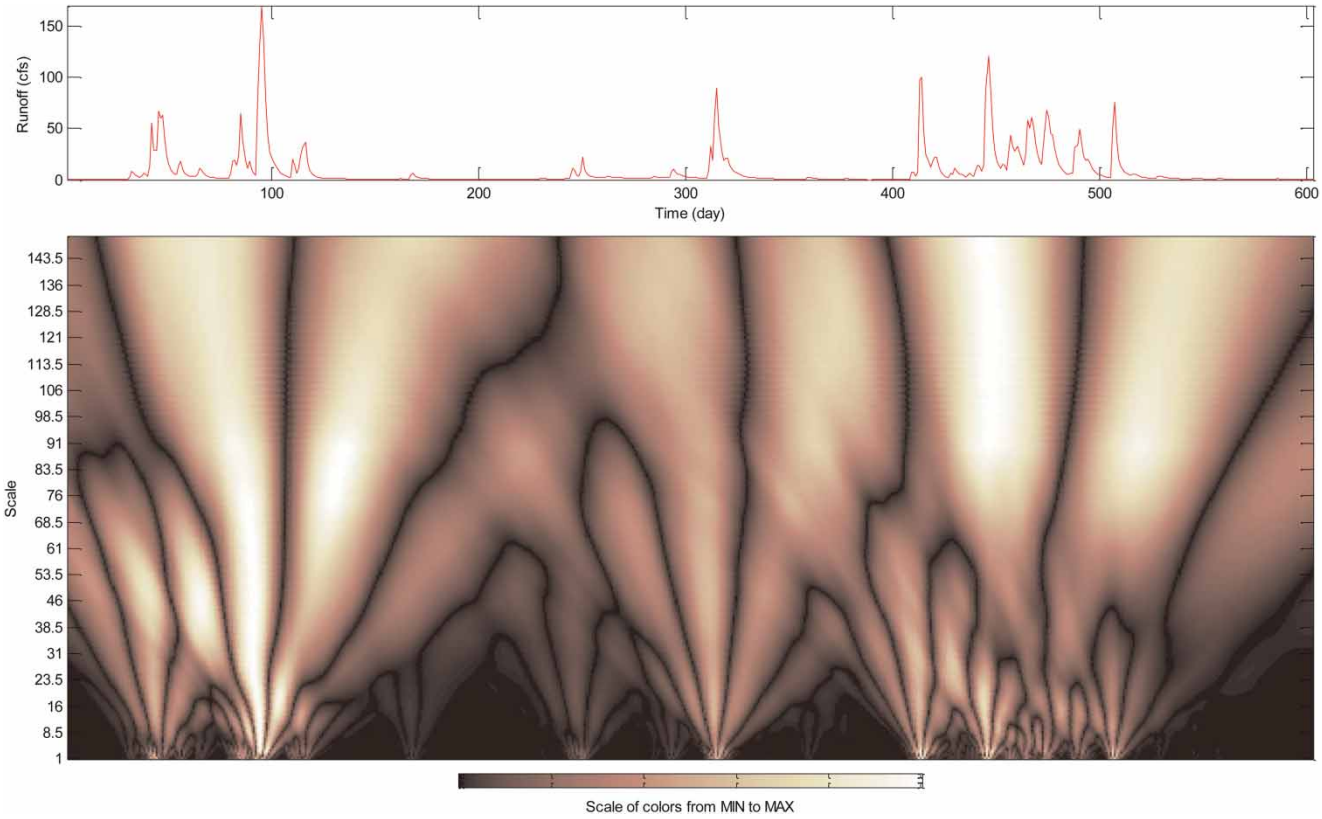

Figure 9 | Scalogram of runoff time series extracted via db4 mother wavelet for the Delaney Creek sub-basin.

via a SOM, best models structures and obtained performance criteria (i.e. DC and RMSE) of the proposed modeling according to the applied mother wavelets are shown in Table 3. Input variables d_1, d_2, d_3, d_4, d_5 included wavelet details of the rainfall and runoff time series that indicate; 2^1-day mode, 2^2-day mode, 2^3-day mode, which is nearly weekly mode, 2^4-day mode and 2^5-day mode, which is nearly monthly mode, respectively. The output variable in Table 3 is 1-day-ahead runoff values. Network structure respectively indicates number of input variables, hidden neurons and output variable of the selected structure. The optimal hidden neuron numbers were obtained through trial and error procedure. In this way, 6–25 hidden neurons were examined in a single hidden layer for each FFNN structure and the optimal number of hidden neurons was determined. Decomposition of the time series by Haar-db4

Table 3 | Results of proposed WSNN model for the Delaney Creek sub-basin

				Calibration		Verification	
Mother wavelet type	**Selected details as input variables[a]**	**FFNN structure[b]**	**Epoch no.**	**DC**	**RMSE (normalized)**	**DC**	**RMSE (normalized)**
Haar-db4	*d2q, d3q, d4q, d5p*	(6-9-1)	300	0.94	0.008	0.86	0.015
coif2	*d3q, d4q, d5q, d5p*	(6-9-1)	180	0.92	0.009	0.82	0.017
db2	*d4q, d3q, d1p, d5p*	(6-9-1)	110	0.86	0.011	0.67	0.023
Meyer	*d3p, d1q, d5q, d4p*	(6-9-1)	110	0.86	0.012	0.70	0.022

[a]Approximation of sub-series of rainfall and runoff were imposed to the FFNN in addition to detailed sub-series.
[b]The results for the best structure have been presented.
DC, determination coefficient; RSME, root mean squared error.

mother wavelets, which matches patterns of main time series, led to a better performance. The results reconfirmed previous recommendations about similarity of mother wavelets formation with main time series (Nourani 2009). The high stochastic feature of precipitation makes its effects unpredictable; therefore rainfall sub-series may not be a good regressor to predict the runoff values, individually. On the other hand, runoff time series, according to its Markovian inherent, shows strong regression with some of its constitutive seasonalities. Hence, runoff sub-series are more effective in runoff prediction than rainfall sub-series. Accordingly, more runoff sub-series were truly selected than rainfall sub-series by the SOM of the applied methodology. In data pre-processed via db2 and Meyer mother wavelets, two detailed sub-series of rainfall time series participated in the modeling which decreased DC to 0.67 and 0.7, whereas in data pre-processed by coif2 and Haar-db4 mother wavelets, only one of the detailed sub-series of rainfall fell separately in the modeling. Therefore, higher DC (i.e. 0.82 and 0.86 for coif2 and Haar-db4 mother wavelets, respectively) was obtained by using the two mentioned mother wavelets. For all mother wavelets, same ANN structure (i.e., six input variables, nine hidden neurons and one output variable) led to better results, which may be due to the same numbers of inputs in all structures. To consider the effect of decomposition level on the model performance, two other decomposition levels of wavelet transformation (i.e. levels 3 and 7) were also examined using Haar-db4 mother wavelets in this research. Decomposition at level 3 yields three detailed sub-series (i.e., 2^1-day mode, 2^2-day mode, 2^3-day mode, which is nearly weekly mode) and decomposition at level 7 that contains four more details (i.e., 2^4-day mode, 2^5-day mode, which is nearly monthly mode, 2^6-day mode and 2^7-day mode). Table 4 includes results obtained by WSNN according to different decomposition levels using Haar-db4 mother wavelets for the Delaney Creek sub-basin. The employed SOM output layer for decomposing at levels 3 and 7 were 1×2 and 1×4, respectively. The level 5 decomposition showed better results among other examined decomposition levels. When using decomposition level 3, few seasonalities from main time series were taken into account. Therefore, an inefficient number of time scales decreased the model performance. Model performance was greatly reduced by

Table 4 | WSNN results in different decomposition levels using haar-db4 mother wavelets for the Delaney Creek sub-basin

		Calibration		Verification	
Decomposition level	**FFNN structure[a]**	**DC**	**RMSE (normalized)**	**DC**	**RMSE (normalized)**
3	(4-7-1)	0.91	0.011	0.85	0.009
5	(6-9-1)	0.94	0.008	0.86	0.015
7	(6-10-1)	0.62	0.020	0.54	0.028

[a]In this table the best result for each model has been presented.
DC, determination coefficient; RSME, root mean squared error.

increasing the decomposition level up to level 7. Level 7 of decomposition included 16 sub-series (i.e. 14 detailed and two approximation sub-series of rainfall and runoff time series) which seemed to be excessive for SOM to handle the clustering well. Figure 10 compares observed and computed data of calibration and verification time series obtained by the proposed methodology through decomposing at level 5, using Haar-db4 mother wavelets and reconstructed via the trained FFNN for the Delaney Creek sub-basin.

One of the important concerns regarding the forecasting models is the capability of the model to provide a useful horizon of forecasts. In this way, a multi-step-ahead approach in FFNN was also applied on decomposed time series obtained by Haar-db4 mother wavelets. In order to explore predictions several time steps ahead, the lead times of runoff time series (i.e. 2, 4, 5, 8 days of lead time) were considered as output neurons and subsequently the relevant FFNN models were trained and verified. The lead time of 5 days led to better performance in term of DC compared with other lead times. Multi-step-ahead forecasting performance poorly compared with 1-day-ahead forecasting. Obtained results have been tabulated in Table 5 for multi-step-ahead forecasting.

Sensitivity analysis

To survey the efficiency of SOM in extracting dominant inputs for the FFNN model, the classic sensitivity analysis was applied on sub-series of the Delaney Creek sub-basin obtained by wavelet transform. To apply sensitivity analysis on sub-series obtained by decomposing the time series at level 5, $2^{12} - 1$ subsets of wavelet outputs should be examined. Table 6 shows samples of analyzed data obtained by decomposing via Haar-db4 mother wavelets. Input combinations of sub-series that included rainfall sub-series performed poorly in the modeling process. It appears that an approximation of sub-series of rainfall did not have a considerable effect on the model performance in the studied watershed. Thus, the supposed hypothesis of the

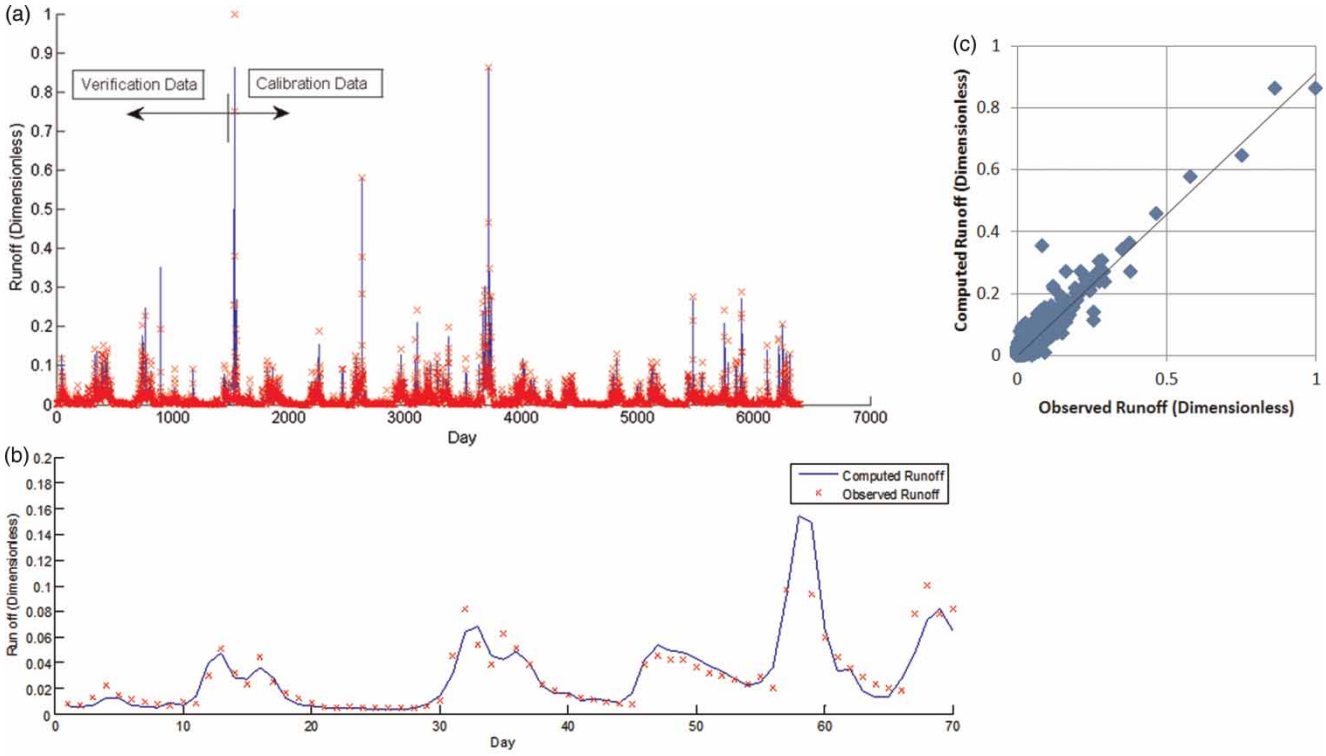

Figure 10 | Results of the wavelet–SOM–FFNN model (WSNN) model for decomposing by Haar-db4. (a) Computed and observed runoff. (b) Detail of a hydrograph. (c) Scatter plot for the Delaney Creek sub-basin.

Table 5 | Results of multi-step ahead forecasting using Haar-db4 mother wavelets for the Delaney Creek sub-basin

				Calibration		Verification	
Input variables[a]	**Output variable**	**FFNN structure[b]**	**Epoch no.**	**DC**	**RMSE (Normalized)**	**DC**	**RMSE (Normalized)**
d2q, d3q, d4q, d5p	Q_{t+2}	(6-9-1)	270	0.87	0.011	0.68	0.023
d2q, d3q, d4q, d5p	Q_{t+4}	(6-8-1)	290	0.81	0.014	0.65	0.024
d2q, d3q, d4q, d5p	Q_{t+5}	(6-8-1)	100	0.81	0.014	0.74	0.022
d2q, d3q, d4q, d5p	Q_{t+8}	(6-7-1)	130	0.66	0.019	0.27	0.035

[a]Approximation sub-series of rainfall and runoff were imposed to FFNN in addition to detail sub-series.
[b]The results for the best structure have been presented.
DC, determination coefficient; RSME, root mean squared error.

Table 6 | Results of sensitivity analysis by FFNN for the Delaney Creek sub-basin

		Determination coefficient (DC)	
ANN structure[a]	**Input variables**	**Calibration**	**Verification**
(1, 3, 1)	ap_q	0.50	0.30
(2, 6, 1)	ap_q, ap_p[b]	0.54	0.36
(2,9,1)	ap_q, d_{1p}	0.70	0.60
(2, 6, 1)	ap_q, d_{2p}	0.64	0.64
(2, 8, 1)	ap_q, d_{3p}	0.7	0.67
(2, 10, 1)	ap_q, d_{4p}	0.57	0.57
(2, 4, 1)	ap_q, d_{5p}	0.57	0.38
(2, 10, 1)	ap_q, d_{1q}	0.79	0.78
(2, 9, 1)	ap_q, d_{2q}	0.73	0.78
(2, 5, 1)	ap_q, d_{3q}	0.71	0.71
(2, 6, 1)	ap_q, d_{4q}	0.45	0.34
(2, 10, 1)	ap_q, d_{5q}	0.74	0.6
(3, 8, 1)	ap_q, d_{q1}, d_{2q}	0.83	0.78
(4, 6, 1)	ap_q, d_{1q}, d_{3q}, d_{4q}	0.89	0.87
(4, 7, 1)	ap_q, d_{2q}, d_{3q}, d_{5q}	0.89	0.85
(4, 5, 1)	ap_q, d_{1q}, d_{2q}, d_{3q}	0.87	0.85
(4, 6, 1)	ap_q, d_{3q}, d_{4q}, d_{5q}	0.91	0.86

[a]The results for the best structure have been presented.
[b]ap_q: approximation sub-series of runoff, ap_p: approximation sub-series of rainfall.

effect of both approximations seems illogical in the case study. This situation may be due to the high stochastic feature, the precipitation event, which makes its effects unpredictable – but runoff time series according to its Markovian inherent is related to some of its constitutive seasonalities. By comparing results with the SOM output, it is seen that the SOM truly neglected the rainfall sub-series and mostly involved the runoff sub-series in the modeling. As there are five detailed sub-series for runoff, 2^5-1 components of runoff details are analyzed to survey the correctness of the components selected by the SOM. Components of the runoff sub-series that performed better are shown in Table 6. The approximation sub-series of runoff, individually, covered 30% of the resultant DC. By comparing the results tabulated in Tables 3 and 6, it is clear that the runoff component selected via SOM (i.e. the component involved detailed the sub-series of levels 2, 3 and 4) is more efficient than the other components. The tabulated results show that although a few of selected inputs by SOM did not have a significant effect on FFNN modeling, the performance of the SOM is still reliable and acceptable. As the SOM clusters data according to its similarity and does not pay attention to importance of data in modeling the process, it is possible for a cluster to contain thoroughly unimportant data. For example in this study, in using Haar-db4 mother wavelets, clusters 3 and 4 contain members that have a low effect on the model efficiency. Thus, the SOM may be considered as a pre-screen tool for the complete sensitivity analysis, and can reduce the trial-error steps from $2^{12}-1$ to 2^5-1 for the first case study. Therefore, not only does the application of sensitivity analysis on SOM outputs reveal important data, but it also eliminates the need to examine all 2^d-1 subsets of the initial sub-series obtained by the wavelet transform.

Comparison of the models

To truly evaluate the efficiency of the proposed hybrid WSNN model, the methodology was similarly applied to

the data of case study 2 (i.e. the Payne Creek sub-basin). The obtained results were also compared with the results of an *ad hoc* FFNN (without data pre-processing) and ARIMAX models.

An *ad hoc* multi-layer FFNN without any data pre-processing was developed as a benchmark to evaluate the proposed model performance for each case study. In this way, four structures were tested for an *ad hoc* FFNN model. First, 3–25 hidden neurons in a single hidden layer were examined for training each FFNN structure using the Levenberg-Marquardt training algorithm (Haykin 1994) to determine the best structure. To ensure that the network did not overfit the training data, the training was terminated when the error in the validation data set began to rise. For instance, the best result for each model has been tabulated in Table 7 for the Delaney Creek sub-basin. The results showed the low performance of the model even when previous days rainfall–runoff data constituted input data. This finding was probably because of significant signal fluctuations around the mean value that reduced the short-term regression between data. By increasing the number of input variables, the number of hidden neurons also was grown. Model efficiency was reduced when rainfall and runoff data from the previous 4 days was involved in the modeling. This situation might be due to the hydrological regime of the study area in that the precipitation over the watershed usually takes up to 3 days and rarely continues to a fourth day. Therefore, the fourth neuron in the input layer usually received a zero value, which acted as the network noise and reduced the model efficiency.

To have a better interpretation of the model performance, the linear ARIMAX model was also employed for both case studies. In this research, the ARIMAX(p,d,q)I(t) model was applied in which p, q and d refer to orders of autoregressive and moving average components and the number of differencing operations, respectively (Salas *et al.* 1980). The best ARIMAX structure for Delaney Creek sub-basin belongs to ARIMAX (3, 0, 1) I(t). For this model, firstly an ARIMAX model was fitted to the calibration data to calibrate the model. The calibrated model was then used to find the values for the verification data, day by day. Table 8 presents the best results obtained by ARIMAX, *ad hoc* FFNN (without data pre-processing) and the proposed WSNN models for both case studies. Tabulated results indicate that the ARIMAX model, due to its linear inherence is unable to completely handle the complex non-linear rainfall–runoff process. Although the *ad hoc* FFNN model is more efficient than the ARIMAX model, it

Table 7 | Results and structure of the *ad hoc* FFNN model for the Delaney Creek sub-basin

		Determination coefficient (DC)	
Network structure[a]	**Input variables[b]**	**Calibration**	**Verification**
(4, 4, 1)	Q_t, Q_{t-1}, I_t, I_{t-1}	0.89	0.73
(4, 9, 1)	Q_t, Q_{t-1}, Q_{t-2}, I_t	0.88	0.85
(5, 14, 1)	Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, I_t	0.90	0.81
(6, 19, 1)	Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, I_t	0.87	0.68

[a]The results for the best structure have been presented.
[b]Q_{t+1}: output variable.

Table 8 | Comparison of results obtained in two study areas

		Calibration		Verification		
Study area	**Model type[a]**	**DC**	**RMSE (normalized)**	**DC**	**RMSE (normalized)**	**DC_{peak} verification**
Delany Creek	ARIMAX	0.78	0.0170	0.65	0.0181	0.82
	WSNN[b]	0.94	0.008	0.86	0.015	0.95
	FFNN	0.90	0.012	0.81	0.016	0.84
Payne Creek	ARIMAX	0.72	0.019	0.64	0.025	0.76
	WSNN	0.87	0.027	0.79	0.016	0.90
	FFNN	0.94	0.015	0.77	0.016	0.81

[a]In this table, the best result for each model has been presented.
[b]Combination of wavelet, SOM and FFNN methods.
DC, determination coefficient; RSME, root mean squared error.

only considered short-term autoregressive features of the process and could not capture long-term seasonality. Therefore, it performed worse compared with when it was linked to wavelet and SOM concepts. Another criterion for comparing performance of different models is the capability of the model to estimate peak values of runoff, a key task in any flood mitigation program (Equation (13)). According to the results presented in Table 8, the proposed hybrid model, which considered seasonal patterns, was more efficient in detecting peak discharges than the two other models. It is evident that extreme or peak values in the rainfall and runoff time series, which occur in a periodic pattern, can be detected accurately by the seasonal models. For example, Figure 11 compares results obtained via FFNN and WSNN models with the observed runoff data for the Delaney Creek sub-basin.

By comparing the data (i.e. maximum and standard deviation values; Table 1) and topologic characteristics (i.e. slope and area, see section on Study Area and Data) of both watersheds, it can be emphasized that the Payne Creek sub-basin, with a vaster area and spatially different topological features, may show more complex dynamic behavior than the Delaney Creek sub-basin. Therefore, in spite of the acceptable performance of the WSNN model for both watersheds, the proposed WSNN model, as well as the *ad hoc* FFNN and ARIMAX models, led to a lower performance for the Payne Creek sub-basin with respect to the Delaney Creek sub-basin (see Table 8).

CONCLUDING REMARKS

The use of wavelet-ANN-based modeling of the rainfall-runoff process that detects non-linear relationships simultaneously in various time scales, has grown greatly in recent years. Although several seasonalities can be found in a process, not all of them are informative or have a considerable effect on the model performance. Application of a model, such as the FFNN, without proper data pre-processing can lead to deterioration in modeling that usually becomes manifested as network overfitting. In this paper, a combination of two methodologies was applied to pre-process available rainfall–runoff time series data, before imposing them to a FFNN. Wavelet transform was used to capture multi-scale features of the signals by decomposing the main time series into several sub-series. A SOM was

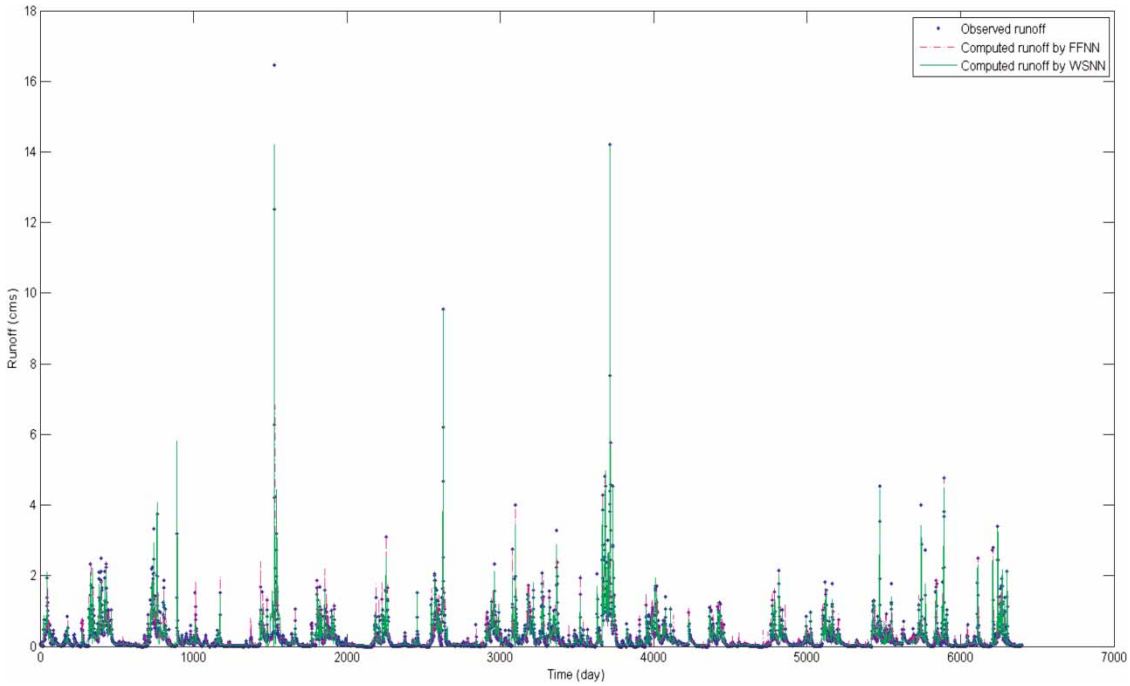

Figure 11 | Comparison of the results of FFNN and wavelet–SOM–FFNN model (WSNN) models for decomposing by Haar-db4 mother wavelets for the Delaney Creek sub-basin.

then applied to classify the decomposed sub-series into several clusters. The center of the clusters as the representative of these clusters were determined and imposed to a FFNN to forecast 1-day-ahead and multi-step-ahead values of runoff discharge. The observed rainfall and runoff data from two study areas with different topological features were used in this study. Three decomposition levels were examined in the modeling (i.e. levels 3, 5 and 7). Decomposing the main time series at level 5, which includes monthly seasonality, led to better performance. Model performance was greatly reduced by increasing the decomposition level up to 7. This change might be due to an increase in the sub-series number that the SOM could not cluster properly. Four types of mother wavelets (i.e. db2, coif2 and Meyer mother wavelets in first three transformations and a combination of Haar and db4 mother wavelets in the fourth transformation for rainfall and runoff time series, respectively) were applied in this research. The results proved that the application of mother wavelets according to similarity of mother wavelets shape with main time series formations may lead to better results. Runoff sub-series, according to its Markovian inherence, had more effect compared with rainfall sub-series for rainfall–runoff modeling of the studied watershed. Multi-step-ahead forecasting showed a poor performance compared with 1-day-ahead forecasting. To evaluate model performance, the proposed methodology was also compared with two conventional models (i.e. the ARIMAX method as a linear model and the *ad hoc* FFNN without any data pre-processing as a non-linear model). The results show that the ARIMAX model due to its linear inherent nature could not detect non-linear relationship between the studied parameters. Although the *ad hoc* FFNN yielded a better performance, it only considered short-term autoregressive features of the process and could not capture long-term seasonalities. The sensitivity analysis showed that as the SOM method classifies sub-series according to similarity and does not pay attention to their importance on the model performance, it is possible for a cluster to contain unimportant sub-series. Therefore, a post-sensitivity analysis of the SOM results seems to be the logical way to ensure the dominance of the sub-series.

To complete the current study, it can be helpful to involve other effective parameters such as evaporation in runoff prediction. If a proper and longer data set is available, we suggest that the phenomenon is also surveyed for longer periods in seasonal and annual scales. The examination of other artificial intelligence approaches to forecast runoff values in the hydrological processes is also suggested. For instance, according to the capability of the Fuzzy Set theory, the conjunction of the Fuzzy and ANN concepts as the Adaptive Neural-Fuzzy Inference System (ANFIS) may led to a promising tool to handle the uncertainty of the hydrological process as well as the hysteretic property.

REFERENCES

Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., Shamseldin, A. Y., Solomatine, D. P., Toth, E. & Wilby, R. L. 2012 Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. *Prog. Phys. Geogr.* **36**, 480–513.

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

Adamowski, J. & Chan, H. F. 2011 A wavelet neural network conjunction model for groundwater level forecasting. *J. Hydrol.* **407**, 28–40.

Adamowski, J., Chan, H. F., Prasher, S. O. & Sharda, V. N. 2012 Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data. *J. Hydroinform.* **14**, 731–744.

Addison, P. S., Murrary, K. B. & Watson, J. N. 2001 Wavelet transform analysis of open channel wake flows. *J. Eng. Mech.* **127**, 58–70.

Altunkaynak, A. 2007 Forecasting surface water level fluctuations of Lake Van by artificial neural networks. *Water Resour. Manage.* **21**, 399–408.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000a Artificial neural networks in hydrology. I: Preliminary concepts. *J. Hydrol. Eng.* **5**, 115–123.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000b Artificial neural networks in hydrology. II: Hydrologic applications. *J. Hydrol. Eng.* **5**, 124–137.

Aussem, A., Campbell, J. & Murtagh, F. 1998 Wavelet-based feature extraction and decomposition strategies for financial forecasting. *J. Comput. Intell. Financ.* **6**, 5–12.

Bowden, G. J., Dandy, G. C. & Maier, H. R. 2005 Input determination for neural network models in water resources applications. *J. Hydrol.* **301**, 75–92.

Bowden, G. J., Maier, H. R. & Dandy, G. C. 2002 Optimal division of data for neural network models in water resources applications. *Water Resour. Res.* **38**, 1010–1011.

Cannas, B., Fanni, A., See, L. & Sias, G. 2006 Data pre-processing for river flow forecasting using neural networks: wavelet transforms and data partitioning. *Phys. Chem. Earth.* **31**, 1164–1171.

Cereghino, R. & Park, Y. S. 2009 Review of the self-organizing map (SOM) approach in water resources: commentary. *Environ. Model. Soft.* **24**, 945–947.

Chang, L. C., Shen, H. Y., Wang, Y. F., Huang, J. Y. & Lin, Y. T. 2010 Clustering-based hybrid inundation model for forecasting flood inundation depths. *J. Hydrol.* **385**, 257–268.

Chon, T. S., Park, Y. S., Moon, K. H. & Cha, E. Y. 1996 Patternizing communities by using an artificial neural network. *Ecol. Model.* **90**, 69–78.

Corne, S., Murray, T., Openshaw, S., See, L. & Turton, I. 1999 Using computational intelligence techniques to model sub-glacial water systems. *J. Geogr. Syst.* **1**, 37–60.

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

Foufoula-Georgiou, E. & Kumar, P. 1995 *Wavelet in Geophysics.* Academic Press, New York.

Haykin, S. 1994 *Neural Networks: A Comprehensive Foundation.* Macmillan, New York.

Haykin, S. 1999 *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Upper Saddle River, New Jersey.

Hornik, K. 1989 Multilayer feed-forward networks are universal approximators. *Neur. Netw.* **2**, 359–366.

Hsu, K. C. & Li, S. T. 2010 Clustering spatial-temporal precipitation data using wavelet transform and self-organizing map neural network. *Adv. Water Resour.* **33**, 190–200.

Iliadis, L. S. & Maris, F. 2007 An artificial neural network model for mountainous water-resources management: the case of Cyprus mountainous watersheds. *Environ. Model. Soft.* **22**, 1066–1072.

Jain, A. & Srinivasulu, S. 2006 Integrated approach to model decomposed flow hydrograph using artificial neural network and conceptual techniques. *J. Hydrol.* **317**, 291–306.

Kalteh, A. M. & Berndtsson, R. 2007 Interpolating monthly precipitation by self-organizing map (SOM) and multilayer perceptron (MLP). *Hydrol. Sci. J.* **52**, 305–317.

Kalteh, A. M., Hjorth, P. & Berndtsson, R. 2008 Review of the self-organizing map (SOM) approach in water resources: analysis, modeling and application. *Environ. Model. Soft.* **23**, 835–845.

Kisi, O. 2010 Daily suspended sediment estimation using neuro-wavelet models. *Int. J. Earth Sci.* **99** (6), 1471–1482.

Kohonen, T. 2001 *Self-Organizing Maps.* Springer-Verlag, Berlin.

Kohonen, T., Makisara, K. & Saramaki, T. 1984 Phonotopic maps insightful representation of phonological features for speech recognition. In: *Proceedings of 7ICPR, International Conference on Pattern Recognition*, CA. IEEE Computer Society Press, Los Alamitos, pp. 182–185.

Labat, D., Ababou, R. & Mangin, A. 2000 Rainfall-runoff relation for karstic spring. Part 2: Continuous wavelet and discrete orthogonal multi resolution analysis. *J. Hydrol.* **238**, 149–178.

Legates, D. R. & McCabe, G. J. 1999 Evaluating the use of goodness of fit measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **35**, 233–241.

Lin, G. F. & Chen, L. H. 2006 Identification of homogeneous regions for regional frequency analysis using the self-organizing map. *J. Hydrol.* **324**, 1–9.

Lin, G. F. & Wu, M. C. 2007 A SOM-based approach to estimate design hyetographs of un-gauged sites. *J. Hydrol.* **339**, 216–226.

Maheswaran, R. & Khosa, R. 2012 Multi-scale nonlinear model for monthly stream flow forecasting: a wavelet-based approach. *J. Hydroinform.* **14**, 424–442.

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

Mallat, S. G. 1998 *A Wavelet Tour of Signal Processing*. Academic Press, San Diego.

MathWorks, Inc. 2010a *MATLAB: User's Guide, Version 7*. The Math Works, Inc., Natick.

MathWorks, Inc. 2010b *MATLAB: SOM Toolbox (NCTOOL) User's Guide, Version 7*. The Math Works, Inc., Natick.

Mirbagheri, S. A., Nourani, V., Rajaee, T. & Alikhani, A. 2010 Neuro-fuzzy models employing wavelet analysis for suspended sediment concentration prediction in rivers. *Hydrol. Sci. J.* **55** (7), 1175–1189.

Nason, G. P. & von Sachs, R. 1999 Wavelets in time series analysis. *Philos. Trans. R. Soc.* **357**, 2511–2526.

Nourani, V. 2009 Reply to comment on an ANN-based model for spatiotemporal groundwater level forecasting. *Hydrol. Process.* **24**, 370–371.

Nourani, V. & Kalantari, O. 2010 Integrated artificial neural network for spatiotemporal modeling of rainfall-runoff-sediment processes. *Environ. Eng. Sci.* **27**, 411–422.

Nourani, V., Monadjemi, P. & Singh, V. P. 2007 Liquid analogue model for laboratory simulation of rainfall–runoff process. *J. Hydrol. Eng.* **12**, 246–255.

Nourani, V., Mogaddam, A. A. & Nadiri, A. O. 2008 An ANN based model for spatiotemporal groundwater level forecasting. *Hydrol. Process.* **22**, 5054–5066.

Nourani, V., Komasi, M. & Mano, A. 2009a A multivariate ANN-Wavelet approach for rainfall-runoff modelling. *Water Resour. Manage.* **23**, 2877–2894.

Nourani, V., Alami, M. T. & Aminfar, M. H. 2009b A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. *Eng. Appl. Artif. Intell.* **22**, 466–472.

Nourani, V., Kisi, O. & Komasi, M. 2011 Two hybrid artificial intelligence approaches for modeling rainfall-runoff process. *J. Hydrol.* **402**, 41–59.

Nourani, V., Komasi, M. & Alami, M. T. 2012 Hybrid wavelet-genetic programming approach to optimize ANN modeling of rainfall-runoff process. *J. Hydrol. Eng.* **16**, 724–741.

Partal, T. & Cigizoglu, H. K. 2008 Estimation and forecasting of daily suspended sediment data using wavelet-neural networks. *J. Hydrol.* **358**, 317–331.

Partal, T. & Kisi, O. 2007 Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. *J. Hydrol.* **342**, 199–212.

Remesan, R., Shamim, M. A., Han, D. & Mathew, J. 2009 Runoff prediction using an integrated hybrid modelling scheme. *J. Hydrol.* **372**, 48–60.

Salas, J. D., Delleur, J. W., Yevjevich, V. & Lane, W. L. 1980 *Applied Modeling of Hydrological Time Series*. Water Resources Publications, Denver.

Tiwari, M. K. & Chatterjee, Ch. 2010 Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. *J. Hydrol.* **394**, 458–470.

Tsai, C. C., Lu, M. C. & Wei, C. C. 2012 Decision tree-based classifier combined with neural-based predictor for water-stage forecasts in a river basin during typhoons: a case study in Taiwan. *Environ. Eng. Sci.* **29**, 108–116.

Wu, C. L., Chau, K. W. & Li, Y. S. 2009 Predicting monthly streamflow using data-driven models coupled with data pre-processing techniques. *Water Resour. Res.* **45**, W08432.

First received 31 August 2012; accepted in revised form 26 November 2012. Available online 11 January 2013