ISEIS

Journal of Environmental Informatics 26(1) 52-70 (2015)

Journal of Environmental Informatics

www.iseis.org/jei

Application of Entropy Concept for Input Selection of Wavelet-ANN Based Rainfall-Runoff Modeling

V. Nourani*, T. R. Khanghah, and A. H. Baghanam

Department of Water Resources Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz 51666-16471, Iran

Received 14 November 2013; revised 28 February 2015; accepted 18 March 2015; published online 23 July 2015

ABSTRACT. This paper presents a Wavelet-based Artificial Neural Network (WANN) approach to model rainfall-runoff process of the Delaney Creek and Payne Creek watersheds with distinct hydro-geomorphological characteristics, located in Florida. Wavelet is utilized to handle the multi-frequency characteristics of the process in daily and monthly time scales. Thus, rainfall and runoff time series were decomposed into several sub-series by various mother wavelets. Due to multiple components obtained through wavelet decomposition, input sets to the Feed Forward Neural Network (FFNN) were enhanced. The application of two information content based criteria (i.e., entropy, *H*, and mutual information, *MI*) to select more reliable input sets (among all potential input sets) and to have better insight into the physics of process is considered as the basic innovation of the study which led to a more accurate and compact model. The increase in the number of input of the FFNN might lead to a complex structure and low performance. The results demonstrated that *MI* as a supervised feature extraction criterion could lead to more reliable outcomes due to its non-linear nature. Furthermore, results indicated the superiority of proposed entropy-based WANN model (EWANN) in comparison to simple FFNN. Moreover, multi-step-ahead FFNN, conventional WANN and classic Auto Regressive Integrated Moving Average with eXogenous inputs (ARIMAX) models could not reveal appropriate forecasting results with regard to EWANN model. The superiority of the EWANN over the WANN and FFNN models is not only in terms of efficiency criteria, but also due to its appropriate ability to provide information about the physics of the process. The consequences of EWANN for rainfall-runoff modeling of two watersheds revealed that the proposed EWANN could simulate the process of a small and flat sub-basin slightly reliable than a sloppy and wild watershed. The poor outcome of monthly modeling in regard to daily modeling might be due to involvement of more uncertainty in the monthly data.

Keywords: rainfall-runoff modeling, Feed Forward Neural Network, wavelet transform, feature extraction, Shannon entropy (information content), multi-step-ahead forecasting

1. Introduction

Precise modeling of hydrological processes such as rainfall-runoff can give effective information for city and environment planning and water resources management. Due to the complexity of designating a reasonable physical relationship between the inputs and output in the rainfall-runoff modeling, application of black box models such as Auto Regressive Integrated Moving Average with eXogenous input (ARIMAX) and Artificial Neural Network (ANN) has been recommended by hydrologists (e.g., Nourani et al., 2011). The ARIMAX model is basically linear and fails in coping with hydrologic processes that are embedded with high complexity, non-stationary and non-linearity. On the other hand, ANN as a self-learning and self-adaptive approximator has shown great authority in non-linear hydrological modeling. A review of ANN applications in hydrology in general and in rainfall-runoff modeling in particular have been presented by American Society of Civil Engineering (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000) and Abrahart et al. (2012), respectively.

Despite the fact that ANN is a reliable approach to model hydrologic processes, there are some disadvantages in dealing with high non-stationary signals such as hydrological time series. The concealed frequencies in hydrological time series may lead to inability of ANN in coping with non-stationary data if proper pre-processing of data is not performed (Cannas et al., 2006). The wavelet transform as such a pre-processing technique, has been widely utilized in hydrology. Several studies addressed the efficiency of wavelet technique for modeling hydrological processes (e.g., Adamowski, 2008; Rajaee et al., 2011; Maheswaran and Khosa, 2012a; Sang, 2013).

To take the advantages of both wavelet and ANN concepts, the hybrid wavelet-ANN (WANN) model which benefits the wavelet-based decomposed sub-series as inputs of the ANN can be an effective tool and may yield more reliable forecasting results than a classic ANN model. WANN model was first presented by Aussem et al. (1998) for financial time series forecasting and nowadays, this methodology has become one of the common approaches in rainfall-runoff modeling (e.g., Can-

* Corresponding author. Tel.: +98 411 3392409; fax: +98 413 3344287.
E-mail address: vnourani@yahoo.com (V. Nourani).

ISSN: 1726-2135 print/1684-8799 online
 doi:10.3808/jei.201500309

nas et al., 2006; Nourani et al., 2009, 2013; Adamowski et al., 2012).

In spite of appropriate flexibility of WANN model for simulation of rainfall-runoff process, a main shortcoming associated is the large number of inputs. As several time scales can be conceived in the rainfall and runoff time series, number of such sub-series obtained via wavelet transform (input neurons) are drastically increased which in turn provide a complicate structure to the WANN model. The WANN model as an interconnected group of artificial neurons processes information based on iterative approach. Each iteration (i.e., epoch) performs training processes by adjusting the weighted connections found between neurons and biases in the network. Since time series data as an input data involve noises, WANN outputs might be affected by those noises. Apparently, the increase in numbers of input neurons magnifies the noise effect and might yield to divergence while adjusting values of weights and bias in a network. Therefore as well as any ANN-based modeling, determination of the most effective sub-series as inputs of WANN model that are independent, informative and efficiently cover the proposed input domain is essential to overcome the mentioned disadvantages (Maheswaran and Khosa, 2012b). Therefore robust algorithms are required to extract dominant features (input sub-sets) for a WANN model. For this purpose, computed linear correlation coefficient (*CC*) between input and output time series, which only detects linear relationships embedded in data is usually employed as a conventional feature extraction technique (e.g., Rajaee et al., 2011; Maheswaran and Khosa, 2012b). However as criticized by Nourani et al. (2011), in spite of a weak linear relationship, a strong nonlinear relationship may exist between the time series (or sub-series) of two parameters.

In this paper, Shannon entropy (information content) as a statistical non-linear measure is applied to extract dominant features of the process as inputs of the WANN model for rainfall-runoff modeling. Thereafter in the current paper, the entropy-based WANN model is abbreviated as EWANN.

Since the entropy measures dependencies between random variables, it is suitable to be applied to complex classification tasks, where methods based on linear relationships may lead to unacceptable outcomes. Entropy can be regarded as a statistical measure of information, disorder, chaos and uncertainty (Shannon, 1948). As a mathematical study, Ebrahimi et al. (1999) discussed the role of variance and entropy in ordering distributions and concluded that unlike variance which measures concentration only around the mean, entropy measures diffuseness of the probability density function (PDF) irrespective of the location of concentration. Amorocho and Espildora (1973), Caselton and Husain (1980) and Krstanovic and Singh (1992) were the noteworthy pioneers who utilized the information content at different fields of hydrology. Harmancioglu and Singh (1998) and also Singh (1997, 2011), surveyed applications of entropy to environmental and water resources problems. Although the most applications of entropy theory in hydrology have been proposed to handle uncertainties of hydrologic quantities and to model several water resources systems (e.g., Mishra et al., 2009), May et al. (2008) focused on an input selection algorithm for ANN models and demonstrated the superior performance of this non-linear measure.

The utilization of entropy as a feature extraction method for the WANN models used in hydrological simulations is quite novel methodology presented in the current paper. Thus, a robust intelligent algorithm is proposed by conjunction of ANN and wavelet concepts to the entropy theory for modeling rainfall-runoff process. For this purpose firstly, the wavelet transform is used to decompose the main rainfall and runoff time series into several sub-series. Then, Shannon entropy is used to select independent and more effective sub-series, resulted in wavelet-based decomposition, as input data to the ANN model. The purpose of this application is not only to increase the accuracy of the modeling by reducing the dimension of input data set, but also to have a physical insight into the process by extracting dominant frequencies (features) of the process. Furthermore, to investigate the effect of physical features of the watershed on the performance of the proposed methodology, two watersheds in Florida (i.e., the Delaney Creek and Payne Creek Sub-basins) with almost same climatological conditions but distinct topographic characteristics were considered in the current study.

2. Material and Methods

2.1. Wavelet Transform

The wavelet transform has increased in usage and popularity in recent years. The hydrological contributions of wavelet transform have been cited by Labat (2005).

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

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

where g is called wavelet function or mother wavelet, m and n are integers that control the wavelet dilation and translation, respectively and $T_{m,n}$ is wavelet coefficient for the discrete wavelet of scale 2^m and location $2^m n$. Equation 1 considers a finite time series, x_i, $i = 0, 1, 2, \ldots, N-1$; and N is an integer power of 2: $N = 2^M$. This gives the ranges of m and n as, respectively, $0 < n < 2^{M-m} - 1$ and $1 < m < M$, where M is the decomposition level. At the largest wavelet scale (i.e., 2^m where $m = M$) only one wavelet is required to cover the time interval, and only one coefficient is produced. At the next scale ($2^m - 1$), two wavelets cover the time interval, hence two coefficients are produced, and so on down to $m = 1$. At $m = 1$, the scale is 2^1, i.e., 2^{M-1} or $N/2$ coefficients are required to describe the signal at this scale. The total number of wavelet coefficients for a discrete time series of length $N = 2^M$ is then $1 + 2 + 4 + 8 + \ldots + 2^{M-1} = N - 1$.

Mother wavelet determination is one of the important issues at the wavelet transform application. The similarity between the shape of the mother wavelet and the shape of raw time-series is the most exclusive guideline to choose the reliable mother wavelet (Nourani et al., 2009). Generally, the mother

wavelets which have a compact support form (e.g., Daubechies-1 or Haar and Daubechies-4 or db4) have effective time localization characteristics and appropriate for time series which have a short memory with short duration transient features. In contrast, mother wavelets with wide support form (e.g., Daubechies-2 or db2) yield reliable forecasting efficiency for time series which have long term features (Maheswaran and Khosa, 2012a).

The inverse discrete transform is calculated by (Nourani et al., 2009):

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

or in a simple format as (Nourani et al., 2009):

$$x_i = \overline{T}(t) + \sum_{m=1}^{M} W_m(t) \quad (3)$$

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

The wavelet coefficients, $W_m(t)$, provide the detailed series, which can capture small features of interpretational value in the data; the residual term, $\overline{T}(t)$, represents the background information of data. Also, the normalized wavelet energy (E) is defined as (Morchen, 2003):

$$E = \frac{|W_m(t)|^2}{\sum_{i=1}^{M} |W_i(t)|^2} \quad (4)$$

2.2. Shannon Entropy (Information Content)

Shannon (1948) mathematically defined entropy in terms of its probability distribution and presented entropy as a statistical measure of the randomness or uncertainty. For a discrete random variable X, which takes values x_1, x_2, …, x_N with probabilities p_1, p_2, …, p_N, respectively, entropy is defined as (Shannon, 1948):

$$H(X) = -\sum_{i=1}^{N_s} p(x_i) \log(p(x_i)) \quad (5)$$

where $H(X)$ is entropy of X (also referred to entropy function) and N_s is the number of intervals or bins to form the histogram and thereinafter PDF.

If X is a deterministic variable, $H(X)$ will be zero, which can be considered as the lower limit of the entropy and this corresponds to the case of absolute certainty. On the other hand, when the variable X is uniformly distributed ($p(x_i) = 1/N_s$ and i= 1, 2, …, N_s), thereafter, Equation 5 yields $H(X) = H_{max}(X) = \log N_s$.

Although entropy can statistically illustrate the rate of uniformity for a single variable, it cannot detect the uniformity and relations between two variables (for instance between an input and output variables of the FFNN model). To overcome the drawback, joint entropy between two variables; X and Y, is also defined as (Gao et al., 2008):

$$H(X,Y) = -\sum_{i=1}^{N_s} \sum_{j=1}^{M_s} p(x_i, y_j) \log(p(x_i, y_j)) \quad (6)$$

where $p(x_i, y_i)$ is the joint probability of X and Y with number of bins N_s and M_s, respectively. In order to calculate the probability and joint probability of X and Y (applying Equations 5 and 6) the univariate and bivariate PDFs should be computed, respectively. Histogram and kernel function methods are commonly used methods to create the PDF of variables. To avoid the risk of assuming a wrong probability distribution for the variables in kernel function method, it is suggested to utilize the histogram method (Yang et al., 2000). The histogram method is not only a simple way to generate PDF of variables, but also because of its data classification inherent is the best suited technique to entropy-based methods (Yang et al., 2000). The proper values of N_s and M_s in Equations 5 and 6 were assumed according to the sample size to perform the PDFs.

To compute the values of H (for a single variable and for joint entropy utilizing Equations 5 and 6) for each sub-series, the data were normalized by scaling between zero and one to eliminate the dimensions of the variables using following equation:

$$x_{Normal} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (7)$$

where x_{Normal}, x_i, x_{min} and x_{max} are normalized value of variable, variable, minimum and maximum values in data set, respectively.

As another entropy-based criterion, the mutual information (MI) measures the dependency between two random variables (Yang et al., 2000). MI can measure the statistical non-linear dependency between two random variables and it is zero when the two random variables are independent (Cover and Thomas, 1991). MI between two random variables X and Y can be computed as (Yang et al., 2000):

$$MI(X,Y) = H(X) + H(Y) - H(X,Y) \quad (8)$$

As mentioned, MI and H are employed in this study as feature extraction criteria and the results are compared with E and CC based methods. E, H and MI are calculated using Equations 4, 5 and 8, respectively. On the other hand linear CC value between two random variables (e.g., X and Y) is compu-

ted as:

$$CC = (\rho_{X,Y})^2 \tag{9}$$

in which,

$$(\rho_{X,Y}) = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2 \sum_{i=1}^{N}(y_i - \overline{y})^2}} \tag{10}$$

where $\overline{x}$ and $\overline{y}$ are the mean values of X and Y, respectively.

It should be noted, *MI* and *CC*, as supervised measures, are calculated on the basis of relationship between input and target values (Equations 8 and 9) while *E* and *H* are unsupervised criteria and measure the structural characteristics of data (Equations 4 and 5). On the other hand, *H* and *MI* are non-linear feature extraction criteria and may be more suitable to be used in a non-linear model such as FFNN.

2.3. Artificial Neural Network and Efficiency Criteria

ANN offers an effective approach for handling large amounts of dynamic, non-linear and noisy data, especially when the underlying physical relationships are not fully understood. This makes ANNs well suited to time series modeling of a data-driven nature.

ANN is composed of a number of interconnected simple processing elements called neurons with the attractive attribute of information processing characteristics such as non-linearity, parallelism, noise tolerance, and learning and generalization capability. Among the applied ANNs, the Feed Forward Neural Network (FFNN) with back propagation (BP) training algorithm and multilayer perceptron structure is the most commonly used method in solving various engineering problems (Hornik et al., 1989).

It has been approved that networks with a single hidden layer can approximate any function to a desired accuracy and is enough for most forecasting problems (Hornik et al., 1989).

The Levenberg-Marquardt (LM) algorithm is the widely used optimization algorithm for ANN training. It is an iterative technique that locates the minimum of a multivariate function that is expressed as the sum of squares of non-linear real-valued functions (Levenberg, 1994). The mathematical notation for training FFNNs with the LM algorithm is fully described by (Hagan and Menhaj, 1994). The LM algorithm has become a standard technique for nonlinear least-squares problems and could be thought of as a combination of the steepest descent and the Gauss-Newton method. The LM algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares (as is typical in training feed-forward networks), then the Hessian matrix can be approximated as (Hagan and Menhaj, 1994):

$$H = J^T J \tag{11}$$

and the gradient can be computed as:

$$g = J^T e \tag{12}$$

where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases and e is a vector of the network errors. The Jacobian matrix can be computed through a standard BP technique, which is much less complex than computing the Hessian matrix. The LM algorithm uses this approximation to the Hessian matrix in the following Newton-like update (Hagan and Menhaj, 1994):

$$x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T e \tag{13}$$

when the scalar μ is zero, this is simply Newton's method using the approximate Hessian matrix. When μ is large, this becomes the gradient descent method with a small step size. Newton's method is faster and more accurate near an error minimum, thus the aim is to shift toward Newton's method as quickly as possible. Therefore, μ isdecreased after each successful step (reduction in performance function) and is increased only when a tentative step would increase the performance function. Accordingly, the performance function is always reduced at each iteration of the algorithm.

The network architecture that yields the most reliable result in terms of root mean square error (*RMSE*) and determination coefficient (*DC*) on the calibration, test and validation steps is determined through trial-error process. The time series data before going through the network are usually normalized between 0 and 1 (using Equation 7). Legates and McCabe (1999) reported that a hydrological model can be sufficiently evaluated by *DC* and *RMSE* as (Nourani et al., 2009). *DC* is also known as Nash-Sutcliffe efficiency criterion (Nourani, 2010):

$$DC = 1 - \frac{\sum_{i=1}^{N}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{N}(\hat{Q}_i - \overline{Q})^2} \tag{14}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Q_i - \hat{Q}_i)^2}{N}} \tag{15}$$

where Q_i, $\hat{Q}_i$ and $\overline{Q}$ are respectively observed data, predicted values and mean of N observed values. The *RMSE* is used to measure the accuracy of forecasted values, which produces a positive value by squaring the error. The *RMSE* increases from zero for perfect forecasts through large positive values as the discrepancies between forecasts and observations become increasingly large. Obviously high value for *DC* (up to one) and

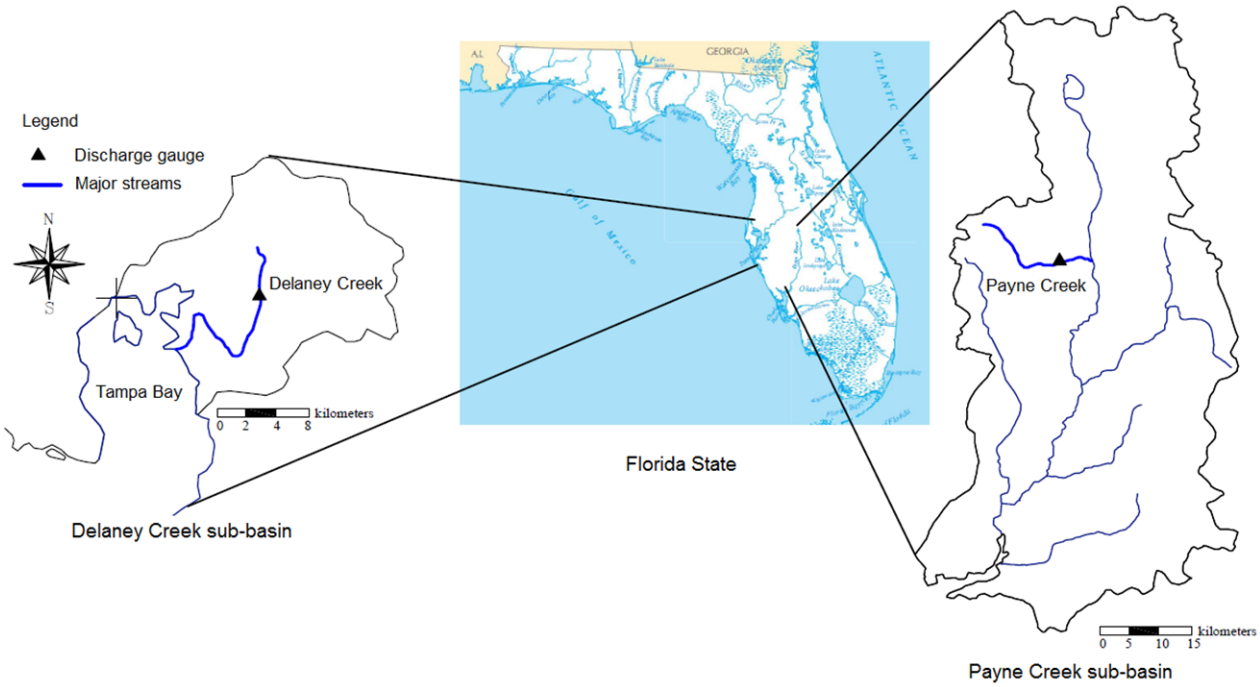

Figure 1. Location map of the Delaney Creek and Payne Creek Sub-basins, Florida.

Table 1. Statistics of Daily Rainfall and Runoff Data for the Study Areas

Sub-basin	Data set	Rainfall time series (mm)				Runoff time series (m^3/s)			
		Max	Min	Mean	Standard deviation	Max	Min	Mean	Standard deviation
Delaney Creek	Calibration	154.18	0	3.53	10.68	16.46	0	0.33	0.58
	Test	81.78	0	2.72	9.15	5.12	0	0.25	0.61
	Validation	80.03	0	2.42	9.01	3.75	0	0.22	0.71
Payne Creek	Calibration	158.80	0	3.56	11.50	77.84	0	3.83	6.26
	Test	94.00	0	3.21	9.18	13.40	0	2.71	4.38
	Validation	91.78	0	2.88	8.62	15.20	0	2.37	3.02

small value for *RMSE* indicate high efficiency of the model.

Firstly, in training step the model is calibrated for each epoch by the calibration data set. The descending rate of calibration diagram by the increase on epoch numbers illustrates the decrease on RMSE (i.e., the evaluation criterion in current study). Secondly in order to avoid the over-fitting problem, the test step is proceeded by the test data for the networks obtained in the calibration step. In order to avoid trapping in local minima and over-fitting problem, the testing step continues until an epoch where in spite of decrease in RMSE of calibration, the RMSE of test begins to increase. At such epoch, the training process is terminated and the relevant epoch to the minimum point of the testing diagram is selected as the optimum epoch. Finally, the optimal model which is selected based on calibration and testing procedures is verified using validation data set.

2.4. Study Area and Data

To investigate the effect of physical features of the watershed on the performance of the proposed methodology, two watersheds in Florida (i.e., the Delaney Creek and Payne Creek Sub-basins) with almost same climatological conditions but distinct topographic characteristics were considered in the current study. The watersheds are introduced as cases (1) and (2), in below.

2.4.1. Case study (1): The Delaney Creek Sub-basin

The Delaney Creek Sub-basin of Tampa Bay Watershed at Florida was selected as the first study area. The watershed located 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 the Tampa Bay on the Gulf of Mexico (Figure 1). This watershed is fairly flat and its elevation varies between 10 meters above and below sea level. The climate of this region is subtropical, exhibiting 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, the total yearly rainfall

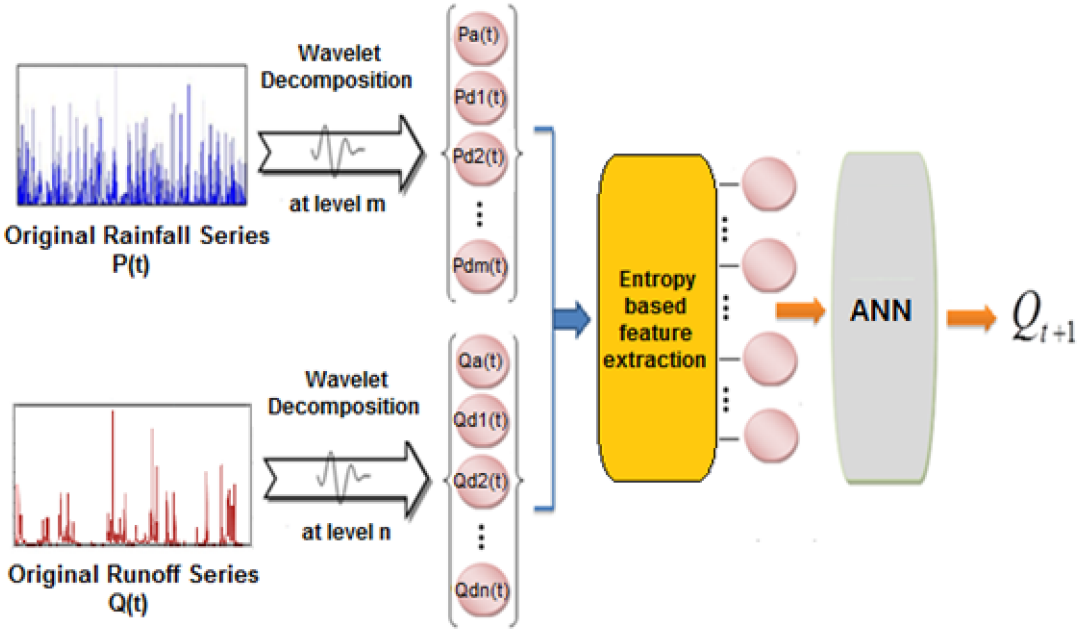

Figure 2. Schematic of the proposed EWANN model.

and the time of concentration for the watershed are about 23 °C, 1,350 mm and 230 minutes, respectively.

The observed daily runoff (Q) and rainfall (P) time series of the Delaney Creek Station, where is the outlet of the Delaney Creek Sub-basin, were retrieved from the United States Geological Survey (USGS) website (http://waterdata.usgs.gov/usa/nwis/uv?site-no=02301750). Time series are included 6,403 daily and 210 monthly data observed from August 1993 to December 2011.

2.4.2. Case study (2): The Payne Creek Sub-basin

The Payne Creek Sub-basin located in the Peace Tampa Bay Watershed in Florida State was selected as the second study area in this research. 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 North West portion of the Peace River Watershed. The Payne Creek Basin covers 322 km^2 areas (Figure 1). The Payne Creek River flows through the sub-basin with annual mean flow of 2 cms. The climate of the area is generally subtropical with an annual average temperature of about 23 °C. Annual average rainfalls in or near the Payne Creek Sub-basin is 1,270 to 1,420 mm and the time of concentration for the watershed is about 250 minutes. Daily stream flow values of the Payne Creek station downloaded from the USGS website (water.usgs.gov/cgi-bin/realsta.pl?station=02295420) and rainfall data of Bartow station were used in this study. The Bartow Station is located in upstream part of the sub-basin. Rainfall and runoff time series are included 5,841 daily and 192 monthly data observed from July 1995 to July 2011.

In order to perform adequate demonstration about ANN-based generalization, data set division into three sub-sets of calibration, test and validation have been suggested by several studies in context of ANN-based modeling of rainfall-runoff process (Abrahart et al., 2012). Therefore for both study areas, data set was divided into three sub-sets: the first 60% of total data were used as calibration set (from August 1993 to August 2004 for Delaney Creek and from July 1995 to February 2005 for Payne Creek), the second 20% for testing (from September 2004 to April 2008 for Delaney Creek and from March 2005 to April 2008 for Payne Creek) and the rest 20% were employed for validation (from May 2008 to December 2011 for Delaney Creek and from May 2008 to July 2011 for Payne Creek) of the models. Statistics of the data sub-sets for both study areas are presented in Table 1. As tabulated in Table 1, the large values of maximum and standard deviation of time series have been appeared in calibration data set. Such data division scheme helps FFNN, as a data interpolator, to learn the pattern of process much effective and leads to more accurate predictions in the testing and validation steps.

Although both study areas are climatically identical, the elevation variation between the upstream and middle parts as well as high value of standard deviation for runoff data (according to Table 1) in the Payne Creek Watershed indicate the sloppy topography and wild situation of this area with regard to the Delaney Creek Watershed.

3. Proposed Entropy-Based Wavelet–ANN (EWANN) Model

The proposed EWANN model consists of a three-layer FFNN with the input neurons of dominant wavelet-based sub-

series extracted by entropy concept. The schematic of the developed model is presented in Figure 2. Basically, the proposed EWANN modeling includes three steps as:

Rainfall and runoff time series are decomposed into several sub-series at different time scales (levels) using discrete wavelet transform (decomposing step). For example, decomposition level M leads to an approximation or trend sub-series and M detailed sub-series (frequencies). Values in approximation sub-series present general view about data. Values in each detailed sub-series describe frequency in data (for instance in daily data, discrete dyadic wavelet decomposition can capture $2^1 = 2$, $2^2 = 4$, $2^3 = 8$, …, and 2^M days frequencies). Indeed, decomposing step is assumed as a data pre-processing step and can efficiently elucidate the non-stationary involved in data. Type of the mother wavelet and also the level of decomposition are the basic parameters of the EWANN model in this step.

Dominant sub-series are extracted via Shannon entropy (screening step). Reasonable modeling is trying not only to increase the accuracy of the modeling, but also to optimize the structure (epochs and number of input and hidden neurons) of model. Although classic WANN model could reach desirable performance in terms of *DC* and *RMSE* with regard to the FFNN, it should be fed by all of the sub-series resulted in the decomposing step. WANN specialists usually present a large number of inputs to the model and rely on the network to recognize dominant inputs (sub-series). There are $d = 2(M + 1)$ sub-series obtained via rainfall and runoff decomposition at level M. Usually, not all of the input variables (sub-series) will be equally informative since some may be correlated, noisy or have no significant relationship with the output variable. Although a large number of information is included in raw data, screening step extracts dominant features (sub-series) of the data and consequently the effect of data noise is diminished. For d potential input sub-series, there are $2^d - 1$ input sub-sets and it is too difficult to test all sub-set combinations for large values of d. Screening step of the EWANN employs an entropy- based non-linear technique to designate dominant sub-series. Since histogram method (to form PDF of variables) assumes data in each interval or bin have same value (to have better compliance between entropy-based criteria and disorder involved in data), the selection of bin size is the main limitation of the EWANN in this step.

The extracted dominant sub-series are imposed into a FFNN to forecast runoff value (simulating step). Due to non-linear structure of FFNN, it is selected as simulation tool which can efficiently model non-linear rainfall-runoff process. The limitation associated with EWANN in this step is the number of epochs and the hidden neurons which should be determined via a trial-error process.

Generally, the first step of the proposed EWANN model handles the non-stationary associated with rainfall and runoff time series and the second and third steps cope with the non-linearity involved in the rainfall-runoff process.

The novelty of EWANN in comparison with classic WANN method is related to its screening step where optimal inputs of FFNN (simulating step) are determined. In WANN model, the FFNN allocates a weight parameter for each input neuron. Values of such weights determine the importance of sub-series. Because of structural noise and redundant information involved in data, the weighting procedure is prone to mistake when huge number of data are considered as inputs. Thereinafter the lack of screening step yields to drastically complex structure (epochs and number of input and hidden neurons) of WANN model. In other words, screening step provided the proposed EWANN model more efficient via selecting limited proper sub-series instead of all sub-series as inputs of model. Basically, classic WANN and FFNN models need three separate trial-error processes to find: i) optimum sub-set of sub-series as inputs, ii) number of epochs, and iii) number of hidden neurons, to arrange the structure of FFNN. Since the proposed model has no necessity to trial-error procedure to find such set of sub-series, the EWANN model can be considered as a modified and robust version of the conventional WANN model. However, the number of epochs and hidden neurons should be yet determined via a trial-error process.

4. Results and Discussion

Prior to apply the proposed EWANN model, FFNN model (without any data pre-processing) was utilized to model rainfall- runoff process of the watersheds. To have reliable interpretation about the performance of the EWANN model in both daily and monthly time scales, the results were also compared with ARIMAX (Auto Regressive Integrated Moving Average with eXogenous inputs, See Box and Jenkins, 1976) and conventional WANN models. For all models, the input and output variables were normalized by scaling between zero and one (using Equation 8) to eliminate the dimensions of the variables and to be ensured that all variables receive equal attention during the calibrating phase.

4.1. Results of FFNN Model

At first, a three-layer single-step-ahead FFNN model without any data pre-processing was used to model the Delaney Creek and Payne Creek Sub-basins rainfall–runoff process. This kind of ANN model accompanied by BP training algorithm has been widely used in hydrological modeling (ASCE, Task Committee on Application of Artificial Neural Networks in Hydrology, 2000).

Four structures were examined for the single-step-ahead FFNN modeling of the watersheds (it should be noted that input layers consisted of rainfall data in current time step and runoff time series up to three days lag in daily scale and a single month lag in monthly scale simulations). Because the values of precipitation at previous time steps have been implicitly affecting the values of runoff in the lead times, the rainfall value at current step accompanied by runoff values at previous time steps could be considered as inputs. The results have been tabulated in Table 2. Each FFNN was calibrated using 2 to 18 hidden neurons in a single hidden layer. Among the training algorithms, the LM algorithm was selected due to its fast convergence ability (Hagan and Menhaj, 1994). Some other studies which have compared different training algorithm of ANNs

Table 2. Results of single-step-ahead FFNN model for the Delaney Creek and Payne Creek Sub-basins

Sub-basin	Modeling time scale	Input variables*	Epoch No.	Structure**	DC			RMSE (normalized)		
					Cal.***	Test	Val.****	Cal.	Test	Val.
Delaney Creek	Daily	Q_{t+1}, Q_t	250	2-4-1	0.88	0.72	0.74	0.0175	0.0183	0.0188
		Q_{t-2}, Q_{t-1}, Q_t	280	3-8-1	0.89	0.80	0.82	0.0173	0.0180	0.0178
		Q_{t-3}, Q_{t-2}, Q_{t-1}, Q_t	280	4-15-1	0.85	0.79	0.77	0.0176	0.0181	0.0184
	Monthly	Q_t	170	1-4-1	0.81	0.73	0.73	0.0176	0.0191	0.0192
Payne Creek	Daily	Q_{t-1}, Q_t	200	2-5-1	0.77	0.65	0.64	0.0188	0.0199	0.0202
		Q_{t-2}, Q_{t-1}, Q_t	320	3-9-1	0.77	0.73	0.74	0.0186	0.0193	0.0192
		Q_{t-3}, Q_{t-2}, Q_{t-1}, Q_t	330	4-10-1	0.69	0.69	0.68	0.0195	0.0196	0.0195
	Monthly	Q_t	190	1-3-1	0.70	0.63	0.64	0.0186	0.0199	0.0192

* Q_{t+1} is output variable (runoff at time $t + 1$).
** Best result has been presented among several structures.
*** Calibration step.
**** Validation step.

Table 3. Results of Daily Multi-step-ahead FFNN Modeling for the Delaney Creek and Payne Creek Sub-basins

Sub-basin	Input variables	Output variable	Structure*	DC			RMSE (normalized)		
				Cal.	Test	Val.	Cal.	Test	Val.
Delaney Creek	Q_{t-4}, Q_{t-2}, Q_t, P_t	Q_{t+2}	3-7-1	0.65	0.60	0.61	0.0198	0.0205	0.0204
	Q_{t-6}, Q_{t-3}, Q_t, P_t	Q_{t+3}	3-7-1	0.65	0.61	0.61	0.0199	0.0207	0.0206
	Q_{t-8}, Q_{t-4}, Q_t, P_t	Q_{t+4}	3-8-1	0.82	0.75	0.76	0.0162	0.0172	0.0171
	Q_{t-12}, Q_{t-6}, Q_t, P_t	Q_{t+6}	3-6-1	0.64	0.59	0.57	0.0212	0.0215	0.0216
	Q_{t-16}, Q_{t-8}, Q_t, P_t	Q_{t+8}	3-6-1	0.76	0.72	0.72	0.0178	0.0183	0.0184
	Q_{t-24}, Q_{t-12}, Q_t, P_t	Q_{t+12}	3-5-1	0.62	0.62	0.61	0.0203	0.0204	0.0205
	Q_{t-32}, Q_{t-16}, Q_t, P_t	Q_{t+16}	3-8-1	0.55	0.51	0.48	0.0232	0.0236	0.0240
Payne Creek	Q_{t-4}, Q_{t-2}, Q_t, P_t	Q_{t+2}	3-5-1	0.62	0.57	0.56	0.0210	0.0218	0.0223
	Q_{t-6}, Q_{t-3}, Q_t, P_t	Q_{t+3}	3-7-1	0.57	0.51	0.52	0.0223	0.0230	0.0229
	Q_{t-8}, Q_{t-4}, Q_t, P_t	Q_{t+4}	3-5-1	0.73	0.64	0.66	0.0184	0.0194	0.0192
	Q_{t-12}, Q_{t-6}, Q_t, P_t	Q_{t+6}	3-7-1	0.60	0.52	0.51	0.0219	0.0225	0.0227
	Q_{t-16}, Q_{t-8}, Q_t, P_t	Q_{t+8}	3-7-1	0.64	0.60	0.61	0.0201	0.0199	0.0199
	Q_{t-24}, Q_{t-12}, Q_t, P_t	Q_{t+12}	3-6-1	0.53	0.51	0.51	0.0191	0.0220	0.0222
	Q_{t-32}, Q_{t-16}, Q_t, P_t	Q_{t+16}	3-6-1	0.50	0.48	0.48	0.0244	0.0245	0.0246

* The result has been presented for the best structure.

have reported the superiority of this algorithm over other algorithms (e.g. see, Nourani et al., 2012).

The numbers of hidden neurons and training epoch were determined using trial-error process. As tabulated in Table 2, comparison of the obtained *DC* values for different structures in validation step revealed that consideration of two days lag in runoff time series in daily modeling was sufficient to predict runoff value one day ahead. The comparison of daily and monthly modeling results confirmed the weak Markovian (auto regressive) property of the process in monthly time scale. According to Table 2 because of sloppy nature of the Payne Creek Watershed, single-step-ahead FFNN reveals low performance with regard to the results of Delaney Creek Watershed.

One of the vital concerns associated with runoff predicting models when designing a flood alert system is the capability of the model to provide reliable forecasts for future with a lead time larger than one time step (Chang et al., 2007). The multi-step-ahead estimations can probably provide long-term predicting horizon comparing with the single-step-ahead modeling in daily time scale (Nourani et al., 2013). The multi-step-ahead approach in the current study comprised of rainfall data in current step and the runoff antecedents in daily time scale as input neurons of model, which could represent the frequencies involved in the rainfall-runoff process. Since monthly runoff exhibited poor auto regressive property (Table 2) and subsequently values of runoff in antecedent months were drastically independent; the multi-step-ahead simulation in monthly time scale was ignored.

In this paper various lead time steps were examined for the runoff time-series as output of multi-step-ahead FFNN model to determine the accuracy of the model in detecting the frequencies involved in rainfall-runoff process (Safehian et al., 2012). The results and details of the input and output neurons of daily multi-step-ahead models for both study areas are presented in Table 3 in which designation of lag sequences for the runoff time series as input was established on the base of the periodic nature of the process. The results presented in Tables 2 and 3 reveal that although generally by increasing the prediction horizon, the model performance was decreased, in both study areas (because of climatic similarity) the lead times of 4 and 8 days led to reliable performance compared to other lead times which implicitly denoted to dominant frequencies of the process.

4.2. Results of EWANN Model

To perform the first step (decomposing step) of EWANN model, both daily and monthly rainfall and runoff time series relevant to the study areas were decomposed into several sub-series via discrete wavelet transform. The decomposition step could accurately elucidate the non-stationary characteristics of the rainfall-runoff process. Haar, db2 and db4 mother wavelets (Mallat, 1998) were used in this study to compare the capability of different mother wavelets to decompose the original daily and monthly data series (see 2.1. Wavelet Transform). Aussem et al. (1998), Singh (2012), Maheswaran and Khosa (2012a) and Shoaib et al. (2014) can be referred for more information about mathematical and practical concepts of the mother wavelets.

In addition to the wavelet type, determination of the appropriate decomposition level (scale) is another important issue in the wavelet-based modeling of hydrological processes. In this study, due to proportional relationship between amount of rainfall and runoff, they were supposed to have the same frequency levels and both time series were decomposed at the same level. Nourani et al. (2011) concluded decomposition level 7 could capture frequencies of time series from 2 days to approximately a season. Decomposition level 7 for both rainfall and runoff time series in daily time scale provided an approximation sub-signal and 7 detailed sub-series (i.e., 2^1-*day* mode, 2^2-*day* mode, 2^3-*day* mode which is nearly weekly mode, 2^4-*day* mode, 2^5-*day* mode which is nearly monthly mode, 2^6-*day* mode and 2^7-*day* mode). In current study decomposition levels 3 and 7 were selected for monthly and daily data, respectively. Although there is not any deterministic method to find the exact decomposition level of hydrological time series and yet trial-error methods remain as the reliable solution, application of Equation 16 can play a guide role in order to select the sufficient level of decomposition (Aussem et al., 1998; Nourani et al., 2011):

$$L = \text{int}[\log(N)] \tag{16}$$

in which L and N indicate the decomposition level and number of time series data, respectively. This experimental equation was derived for fully autoregressive signals with low/without any seasonal or periodic patterns in contrast to hydrologic time series. Comparison of daily and monthly data reveals that the autoregressive property of monthly data is more dominant than daily. Thus, the decomposition level number 3 obtained by Equation 16 could be an appropriate guide for monthly data. On the other hand, the maximum periodicity of monthly data, 12-month mode, approximately coincides 2^3-month mode. In contrast to monthly rainfall data that Equation 16 can be guide for decomposition level number, in daily rainfall data it might not because it is needed to monitor large periodicities such as 365-days (i.e., approximately 2^8-day mode). Selection of decomposition level 8 for daily rainfall data which is nearly an annual mode and can include dominant seasonal periodicity in a rainfall time series might be considered in such studies. Here in this study, sensitivity analysis on levels 7 and 9 beside level 8 demonstrated that there is slight difference on the results of modeling based on these three levels, but it is noticeable that the high decomposition level leads to a large number of parameters in the complex non-linear relationship of the FFNN model. Consequently, although this relationship may monitor and fit the calibration data appropriately, each parameter produces an error in the testing data and net errors decrease the model's efficiency at the verification stage. Also the large amount of input data requires the amount of runtimes in the training network. Thus, decomposition level 7 was selected for daily rainfall data.

As a classic methodology, the computed linear *CC* between the potential inputs and the output of the model has been already applied to select proper inputs (sub-sets) for WANN models (e.g., Rajaee et al., 2011; Maheswaran and Khosa, 2012b). However in a non-linear complex hydrological process, in spite of a weak linear relationship, strong nonlinear relationship may exist between input and output variables. Therefore in such cases where the methods on the basis of linear relationship may lead to unacceptable result, the entropy theory can be a suitable choice to recognize features in complex classification tasks (such as input selecting for FFNN models). To perform screening step of proposed EWANN model, *H* and *MI* (as non-linear Shannon entropy-based criteria) were used in this study for wavelet-based feature extraction of rainfall-runoff process.

To evaluate the efficiency of the entropy-based feature extraction criteria, performances of *H* and *MI* are illustrated and compared with *CC* and *E*-based criteria. The values of *H, MI, E* and *CC* were calculated using decomposed data of the Delaney Creek and Payne Creek Sub-basins. In this way, the sub-series derived from wavelet decomposition procedure were ranked according to the values of *CC, E, H* and *MI* for daily and monthly data. As an instance, normalized values and general ranking of sub-series for daily data at level 7 and monthly data at level 3 of the Delaney Creek Sub-basin data using various mother wavelets are shown in Table 4 for the used feature extraction criteria (*CC, E, H* and *MI*). In Table 4, *Qa* and *Pa* present approximation sub-series of decomposed runoff and rainfall time series, respectively. *Qd1*, …, *Qd7* refer to runoff detailed sub-series and *Pd1*, …, *Pd7* refer to rainfall detailed sub-series at levels 1 to 7. To rank the sub-series obtained via the wavelet transform, values of *E, H, MI* and *CC* were calculated for all sub-series using Equations 4, 5, 8 and 9. Subsequently rank of each sub-series was determined so that the first rank was belonged to the maximum value (more effective input) and the last one to the minimum value.

Several results could be inferred according to the ranking results presented in Table 4 that are briefly listed as below:

- The type of mother wavelet had no major effect on sub-series ranking, based on *H* and *E* (unsupervised criteria).
- For the used mother wavelets, both CC and E based sub-series. ranking were almost identical (probably because of linear inherent of the criteria). However *E*, as an unsupervised criterion, was calculated without any attention to the target and consequently the probable noise and error, contained in the target, might have no direct effect on the ranking.
- In contrast to *H*-based sub-series ranking, most of the prime

Table 4. The Ranking of Sub-series Using Different Feature Extraction Criteria and Different Mother Wavelets for Decomposition at Level 7 in Daily and Level 3 in Monthly Time Scales for the Delaney Creek Sub-basin

Time scale	Rank	db4				db2				Haar			
		H	MI	CC	E	H	MI	CC	E	H	MI	CC	E
Daily	1	*Pd7(1)*	*Qd4(1)*	*Qd3(1)*	*Qa(1)*	*Pa(1)*	*Qa(1)*	*Qa(1)*	*Qa(1)*	*Pa(1)*	*Qd2(1)*	*Qd3(1)*	*Qa(1)*
	2	*Pa(0.94)*	*Qd2(0.96)*	*Qd4(0.96)*	*Qd3(0.98)*	*Pd7(0.93)*	*Qd4(0.98)*	*Qd4(0.98)*	*Qd4(0.96)*	*Pd7(0.91)*	*Qd3(0.97)*	*Qa(0.96)*	*Qd3(0.97)*
	3	*Qa(0.91)*	*Qd3(0.93)*	*Qd5(0.96)*	*Qd4(0.75)*	*Qa(0.92)*	*Qd3(0.95)*	*Qd3(0.92)*	*Qd3(0.94)*	*Qa(0.69)*	*Qd4(0.92)*	*Qd5(0.9)*	*Qd5(0.96)*
	4	*Pd6(0.88)*	*Qd5(0.9)*	*Qa(0.71)*	*Qd5(0.7)*	*Pd6(0.92)*	*Pa(0.73)*	*Pa(0.75)*	*Qd2(0.90)*	*Pd6(0.69)*	*Qa(0.91)*	*Qd4(0.9)*	*Qd4(0.87)*
	5	*Qd7(0.84)*	*Qd6(0.71)*	*Qd7(0.71)*	*Qd7(0.69)*	*Qd7(0.91)*	*Qd5(0.71)*	*Qd2(0.74)*	*Qd7(0.8)*	*Qd7(0.67)*	*Pa(0.78)*	*Qd7(0.83)*	*Qd2(0.74)*
	6	*Pd5(0.60)*	*Qa(0.69)*	*Qd6(0.61)*	*Qd6(0.63)*	*Pd5(0.90)*	*Qd2(0.63)*	*Qd7(0.65)*	*Qd5(0.76)*	*Qd6(0.53)*	*Qd1(0.77)*	*Qd2(0.62)*	*Qd7(0.66)*
	7	*Pd4(0.51)*	*Pd7(0.67)*	*Qd1(0.60)*	*Qd1(0.61)*	*Qd6(0.59)*	*Qd7(0.62)*	*Qd5(0.65)*	*Qd1(0.75)*	*Pd5(0.52)*	*Qd5(0.76)*	*Pa(0.50)*	*Qd6(0.62)*
	8	*Qd6(0.49)*	*Qd7(0.64)*	*Qd2(0.59)*	*Qd2(0.59)*	*Pd4(0.59)*	*Qd1(0.61)*	*Qd1(0.63)*	*Qd6(0.69)*	*Pd4(0.51)*	*Qd6(0.70)*	*Qd6(0.48)*	*Qd1(0.60)*
	9	*Pd3(0.48)*	*Pd4(0.63)*	*Pa(0.53)*	*Pd1(0.55)*	*Qd5(0.58)*	*Pd1(0.52)*	*Qd6(0.55)*	*Pd1(0.66)*	*Pd3(0.44)*	*Pd3(0.64)*	*Pd7(0.48)*	*Pd1(0.59)*
	10	*Qd5(0.47)*	*Pa(0.54)*	*Pd7(0.49)*	*Pd2(0.51)*	*Pd3(0.48)*	*Qd6(0.51)*	*Pd4(0.53)*	*Pd2(0.58)*	*Pd2(0.43)*	*Pd2(0.59)*	*Qd1(0.46)*	*Pd2(0.59)*
	11	*Pd2(0.47)*	*Pd3(0.48)*	*Pd4(0.48)*	*Pd3(0.49)*	*Pd1(0.44)*	*Pd3(0.50)*	*Pd3(0.51)*	*Pa(0.56)*	*Qd5(0.42)*	*Qd7(0.57)*	*Pd4(0.46)*	*Pd3(0.51)*
	12	*Qd4(0.38)*	*Pd5(0.47)*	*Pd5(0.45)*	*Pa(0.38)*	*Pd2(0.42)*	*Pd5(0.48)*	*Pd7(0.44)*	*Pd3(0.52)*	*Qd4(0.39)*	*Pd4(0.53)*	*Pd5(0.40)*	*Pa(0.49)*
	13	*Pd1(0.33)*	*Pd6(0.46)*	*Pd6(0.43)*	*Pd4(0.37)*	*Qd4(0.39)*	*Pd4(0.45)*	*Pd5(0.42)*	*Pd4(0.51)*	*Pd1(0.32)*	*Pd7(0.52)*	*Pd6(0.32)*	*Pd4(0.48)*
	14	*Qd3(0.32)*	*Pd2(0.41)*	*Pd3(0.39)*	*Pd5(0.36)*	*Qd3(0.38)*	*Pd7(0.42)*	*Pd6(0.39)*	*Pd5(0.48)*	*Qd2(0.31)*	*Pd1(0.49)*	*Pd3(0.31)*	*Pd5(0.47)*
	15	*Qd2(0.28)*	*Qd1(0.39)*	*Pd2(0.38)*	*Pd6(0.20)*	*Qd2(0.31)*	*Pd6(0.38)*	*Pd2(0.38)*	*Pd6(0.45)*	*Qd3(0.28)*	*Pd6(0.43)*	*Pd2(0.30)*	*Pd6(0.33)*
	16	*Qd1(0.21)*	*Pd1(0.35)*	*Pd1(0.31)*	*Pd7(0.19)*	*Qd1(0.29)*	*Pd2(0.3)*	*Pd1(0.33)*	*Pd7(0.39)*	*Qd1(0.23)*	*Pd5(0.42)*	*Pd1(0.20)*	*Pd7(0.29)*
Monthly	1	*Pd3(1)*	*Qd2(1)*	*Qd2(1)*	*Qa(1)*	*Pa(1)*	*Qd3(1)*	*Qd3(1)*	*Qa(1)*	*Pd3(1)*	*Qa(1)*	*Qa(1)*	*Qa(1)*
	2	*Pd2(0.96)*	*Qd3(0.97)*	*Pd2(0.94)*	*Pd2(0.98)*	*Pd3(0.96)*	*Qa(0.96)*	*Pd2(0.98)*	*Pd2(0.89)*	*Pa(0.93)*	*Qd3(0.95)*	*Pd2(0.94)*	*Pd2(0.91)*
	3	*Qa(0.94)*	*Qd1(0.93)*	*Qd3(0.89)*	*Qd3(0.63)*	*Pd2(0.89)*	*Qd1(0.92)*	*Qa(0.97)*	*Qd3(0.86)*	*Qa(0.88)*	*Qd1(0.91)*	*Qd2(0.88)*	*Qd3(0.81)*
	4	*Qd3(0.6)*	*Qa(0.61)*	*Qd1(0.55)*	*Qd2(0.57)*	*Pd1(0.63)*	*Pd2(0.58)*	*Qd1(0.95)*	*Qd2(0.63)*	*Pd1(0.84)*	*Pd2(0.89)*	*Qd1(0.77)*	*Qd2(0.72)*
	5	*Pd1(0.58)*	*Pd1(0.57)*	*Qa(0.54)*	*Qd1(0.54)*	*Qd3(0.62)*	*Pd1(0.53)*	*Qd2(0.59)*	*Qd1(0.57)*	*Qd1(0.64)*	*Pd1(0.52)*	*Qd3(0.65)*	*Qd1(0.44)*
	6	*Qd2(0.47)*	*Pd3(0.48)*	*Pd3(0.53)*	*Pd3(0.43)*	*Qd2(0.51)*	*Pd3(0.48)*	*Pd1(0.48)*	*Pd3(0.38)*	*Qd2(0.55)*	*Pd3(0.43)*	*Pd1(0.50)*	*Pd3(0.36)*
	7	*Pa(0.45)*	*Pd2(0.46)*	*Pd1(0.32)*	*Pd1(0.40)*	*Qa(0.32)*	*Qd2(0.32)*	*Pd3(0.36)*	*Pd1(0.34)*	*Pd2(0.33)*	*Qd2(0.4)*	*Pd3(0.24)*	*Pd1(0.22)*
	8	*Qd1(0.28)*	*Pa(0.31)*	*Pa(0.26)*	*Pa(0.30)*	*Qd1(0.31)*	*Pa(0.28)*	*Pa(0.28)*	*Pa(0.29)*	*Qd3(0.22)*	*Pa(0.28)*	*Pa(0.23)*	*Pa(0.22)*

* Numbers in brackets show the normalized values of each criterion.

Table 5. Recognizing Best Feature Extraction Criterion (H, MI, E or CC) for Daily Data, Decomposed at Level 7

Mother wavelet	Ranking base	Delaney Creek				Payne Creek			
		Input variables*	DC			Input variables	DC		
			Cal.	Test	Val.		Cal.	Test	Val.
db4	H	*Pd7,Pa*	0.23	0.04	0.02	*Pd7,Qd2*	0.23	0.11	0.07
	MI	*Qd4,Qd2*	0.79	0.67	0.68	*Qd3,Qd2*	0.69	0.66	0.66
	CC	*Qd3,Qd4*	0.73	0.67	0.66	*Qd3,Qd4*	0.67	0.62	0.61
	E	*Qa,Qd3*	0.70	0.61	0.62	*Qa,Qd3*	0.62	0.58	0.58
	H	*Pd7,Pa,Qa*	0.24	0.15	0.16	*Pd7,Qd2,Qa*	0.24	0.25	0.22
	MI	*Qd4,Qd2,Qd3*	0.83	0.75	0.79	*Qd3,Qd2,Qd4*	0.80	0.75	0.77
	CC	*Qd3,Qd4,Qd5*	0.75	0.66	0.68	*Qd3,Qd4,Qd5*	0.71	0.63	0.66
	E	*Qa,Qd3,Qd4*	0.76	0.66	0.69	*Qa,Qd3,Qd5*	0.70	0.64	0.64
db2	H	*Pa,Pd7*	0.25	0.10	0.08	*Pa,Pd7*	0.22	0.17	0.14
	MI	*Qa,Qd3*	0.66	0.47	0.46	*Qa,Qd3*	0.53	0.38	0.40
	CC	*Qa,Qd4*	0.57	0.44	0.45	*Qa,Qd5*	0.43	0.40	0.38
	E	*Qa,Qd4*	0.57	0.44	0.45	*Qd3,Qa*	0.53	0.38	0.40
	H	*Pa,Pd7,Qa*	0.36	0.14	0.10	*Pa,Pd7,Qa*	0.32	0.15	0.17
	MI	*Qa,Qd3,Qd4*	0.73	0.62	0.60	*Qa,Qd3,Qd5*	0.65	0.55	0.59
	CC	*Qa,Qd4,Qd3*	0.73	0.62	0.60	*Qa,Qd5,Qd3*	0.65	0.55	0.59
	E	*Qa,Qd4,Qd3*	0.73	0.62	0.60	*Qd3,Qa,Qd4*	0.59	0.53	0.52
Haar	H	*Pa,Pd7*	0.30	0.16	0.13	*Pd6,Pa*	0.30	0.22	0.21
	MI	*Qd2,Qd3*	0.62	0.57	0.56	*Qd2,Qd3*	0.64	0.55	0.58
	CC	*Qd3,Qa*	0.58	0.47	0.48	*Qa,Qd3*	0.59	0.50	0.52
	E	*Qa,Qd3*	0.58	0.47	0.48	*Qa,Qd5*	0.59	0.51	0.48
	H	*Pa,Pd7,Qa*	0.29	0.24	0.22	*Pd6,Pa,Qa*	0.39	0.32	0.28
	MI	*Qd2,Qd3,Qd4*	0.75	0.66	0.68	*Qd2,Qd3,Qd4*	0.72	0.66	0.67
	CC	*Qd3,Qa,Qd5*	0.69	0.61	0.60	*Qa,Qd3,Qd4*	0.73	0.61	0.62
	E	*Qa,Qd3,Qd5*	0.69	0.61	0.60	*Qa,Qd5,Pa*	0.60	0.55	0.58

*Q_{t+1} is output variable.

ranks belonged to *Q* components in the ranking procedure based on *CC*, *MI* and *E*.

- For all mother wavelets in *E*-based ranking, first rank belonged to *Qa* sub-series. It is remarkable that the energy of signal is mostly distributed in the approximation sub-series rather than details.

To develop the EWANN model, the FFNN model was fed by extracted dominant sub-series (simulating step) in order to predict runoff values one day ahead for daily and one month ahead for monthly data sets. Therefore for both daily and monthly data of two case studies (the Delaney Creek and Payn selected sub-series via four feature extraction criteria (i.e., *H, MI, E* and *CC*). In order to survey the performance of *H*, *CC*, *E* and *MI* feature extraction criteria, for instance in daily data, the sub-series of the first, second and third ranks (according to the Table 4) were primarily considered as the inputs of FFNN; thereinafter, the sufficient number of inputs was consequently determined using the maximum reduction rate method proposed by Nourani and Zanardo (2014). Table 5 presents the results of developed FFNN models using the selected sub-series decomposed by Haar, db2 and db4 mother wavelets for daily data.

High differences between the most efficient and poor validation *DC*s tabulated in Table 5, approved the sensitivity of the model to the utilized feature extraction criterion. Presented results in Table 5 indicated that almost all models with 2 and 3 input neurons for daily data could lead to acceptable outcomes except when *H* was used as the extraction criterion. According to Table 4, the first and second ranks in H-based ranking for both daily and monthly scales belonged to precipitation components and consequently H-based models were yielded unacceptable results in terms of efficiency criteria (according to Table 5). Such results may be due to the structure of entropy. Although *H* measures information content of the sub-series, it is remarkable that high value of the criterion may be arisen due to noise or redundant information (the information is implicitly presented via other sub-series particularly runoff sub-series in the current study) involved in the sub-series. This fact reconfirmed the sufficiency of the input layer of ad hoc FFNN models presented in Table 2. Furthermore, the results inferred from Table 5 confirmed the superiority of db4 mother wavelet in comparison to db2 and Haar for decomposition of the main time series. The results confirmed that the selected dominant inputs by *MI* criterion tremendously led to more accurate results. Therefore in the rest of paper, only the results of models, developed on the basis of *MI* criterion are presented and the procedure of EWANN is explained for db4-based decomposition (for other mother wavelets, the EWANN models were developed by the same procedure).

Reasonable values for *DC* and *RMSE* guarantee the accu-

Table 6. Results of MI-based EWANN Models for the Study Areas Using db4 Mother Wavelet, Decomposed at Level 7 in Daily and Level 3 in Monthly Time Scales

Sub-basin	Modeling time scale	Input variables*	Structure**	DC		
				Cal.	Test	Val.
Delaney Creek	Daily	*Qd4,Qd2,Qt,*	(3,6,1)	0.94	0.72	0.74
		Qd4,Qd2,Qd3,Qt,	(4,7,1)	0.94	0.80	0.82
		Qd4,Qd2,Qd3,Qd5,Qt,	(5,9,1)	0.95	0.92	0.94
	Monthly	*Qd2,Qd3*	(2,5,1)	0.90	0.86	0.86
		Qd2,Qd3,Qd1	(3,5,1)	0.95	0.90	0.89
Payne Creek	Daily	*Qd3,Qd2,Qt,*	(3,5,1)	0.90	0.73	0.72
		Qd3,Qd2,Qd4,Qt,	(4,6,1)	0.92	0.81	0.82
		Qd3,Qd2,Qd4,Qa,Qt,	(5,7,1)	0.92	0.89	0.91
	Monthly	*Qd2,Qd3*	(2,4,1)	0.84	0.81	0.82
		Qd2,Qd3,Qd1	(3,5,1)	0.91	0.85	0.86

*Q_{t+1} is output variable.
** Best result has been presented among several structures.

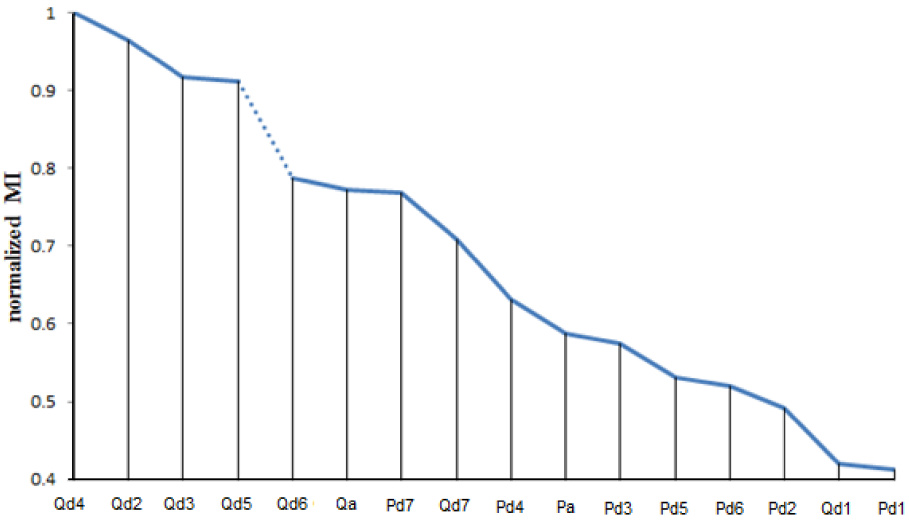

Figure 3. Descending order of the daily sub-series derived by db4 mother wavelet in level 7 for the Delaney Creek Sub-basin (normalized MI).

racy of the model with an optimum structure. The increase of input neurons leads to a complex structure for the FFNN and carries out some difficulties in the calibration step. Therewith, the input time series structurally contain some noise, thus, the increase of inputs magnifies error and may lead to an undesirable result in the test and validation steps. In this study, in order to arrange the structure of FFNN, sub-series with acceptable values of normalized *MI* were selected and imposed to FFNN as the inputs of the proposed EWANN model. In order to calculate the normalized *MI* for the sub-series inferred from decomposition step, value of *MI* for each sub-series was divided by maximum value of *MI*. Thus, normalized *MI* for the first rank (as presented at Table 4) was equal to one and for the others were between one and zero. For instance, Figure 3 presents descending orders of the sub-series derived from db4 mother wavelet for daily data at level 7 for the Delaney Creek Sub-basin. The maximum reduction rate of normalized *MI* separated dominant inputs from the other sub-series. Such reduction in *MI* value was interpreted as maximum decrease in non-linear dependency between two random variables.

In the current study for db4-based decomposition, maximum reduction rate method presented *Qd4*, *Qd2*, *Qd3 and Qd5* as dominant inputs of the model for daily data and *Qd2*, *Qd3 and Qd1* for monthly data. *Qd3*, *Qd2*, *Qd4 and Qa* and *Qd2*, *Qd3 and Qd1* were also determined as dominant sub-series for daily and monthly data of the Payne Creek Sub-basin, respectively. The determined dominant sub-series for two study areas were almost identical which revealed the almost similar climatic condition of the watersheds (see section 2.4).

A shortcoming associated with maximum reduction rate method is probably when no significant reduction is taken place in the normalized *MI* values. In such situation, it is suggested to change the decomposition level.

According to the dyadic representation of the discrete wavelet transform, the finest scale (first level) of the daily decomposition is 2^1 = 2-day mode. However due to the Markovian (auto regressive) property of the runoff process in daily time

Table 7. Mother Wavelet Impact on the Results of Proposed EWANN Model for both Study Areas, in Daily and Monthly Time Scales

Sub-basin	Modeling time scale	Mother wavelet	Input variables*	Structure**	DC			RMSE (normalized)		
					Cal.	Test	Val.	Cal.	Test	Val.
Delaney Creek	Daily	Haar	*Qd2,Qd3,Qd4,Qa,Qt,*	(5,9,1)	0.90	0.82	0.84	0.090	0.014	0.015
		db2	*Qa,Qd4,Qd3,Qt,*	(4,8,1)	0.86	0.83	0.84	0.015	0.017	0.017
		db4	*Qd4,Qd2,Qd3,Qd5,Qt,*	(5,9,1)	**0.95**	**0.92**	**0.94**	**0.008**	**0.010**	**0.009**
	Monthly	Haar	*Qa,Qd3,Qd1,Pd2*	(4,4,1)	0.92	0.86	0.87	0.010	0.014	0.013
		db2	*Qd3,Qa,Qd1*	(3,5,1)	0.93	0.89	0.88	0.008	0.013	0.012
		db4	*Qd2,Qd3,Qd1*	(3,5,1)	**0.95**	**0.90**	**0.89**	**0.008**	**0.011**	**0.011**
Payne Creek	Daily	Haar	*Qd2,Qd3,Qd4,Qa,Qt,*	(5,11,1)	0.87	0.82	0.83	0.014	0.017	0.017
		db2	*Qa,Qd3,Qd5,Pd2*	(4,8,1)	0.85	0.82	0.82	0.016	0.018	0.019
		db4	*Qd3,Qd2,Qd4,Qa,Qt,*	(5,7,1)	**0.92**	**0.89**	**0.91**	**0.010**	**0.011**	**0.010**
	Monthly	Haar	*Qd3,Qd2,Pd2*	(3,4,1)	0.87	0.81	0.83	0.013	0.014	0.013
		db2	*Qa,Qd2,Qd1*	(3,6,1)	0.91	0.86	0.85	0.011	0.013	0.014
		db4	*Qd2,Qd3,Qd1*	(3,5,1)	**0.91**	**0.85**	**0.86**	**0.011**	**0.015**	**0.015**

*Q_{t+1} is output variable.
** Best result has been presented among several structures.

Table 8. Results of Multi-step-ahead EWANN Modeling for the Delaney Creek and Payne Creek Sub-basins in Daily Time Scale

Sub-basin	Output variable*	Structure**	DC			RMSE (normalized)		
			Cal.	Test	Val.	Cal.	Test	Val.
Delaney Creek	Q_{t+2}	4-7-1	0.69	0.65	0.66	0.0191	0.0203	0.0201
	Q_{t+3}	4-5-1	0.65	0.60	0.62	0.0201	0.0211	0.0208
	Q_{t+4}	4-7-1	**0.81**	**0.80**	**0.79**	0.0182	0.0184	0.0186
	Q_{t+6}	4-6-1	0.64	0.61	0.59	0.0197	0.0209	0.0212
	Q_{t+8}	4-5-1	**0.81**	**0.75**	**0.76**	0.0185	0.0194	0.0193
	Q_{t+12}	4-6-1	0.61	0.58	0.58	0.0200	0.0209	0.0208
	Q_{t+16}	4-8-1	**0.79**	**0.76**	**0.78**	0.0188	0.0193	0.0191
Payne Creek	Q_{t+2}	4-4-1	0.68	0.63	0.65	0.0193	0.0207	0.0204
	Q_{t+3}	4-7-1	0.61	0.59	0.60	0.0202	0.0205	0.0204
	Q_{t+4}	4-4-1	**0.82**	**0.76**	**0.79**	0.0181	0.0191	0.0188
	Q_{t+6}	4-5-1	0.63	0.60	0.58	0.0207	0.0209	0.0211
	Q_{t+8}	4-8-1	**0.81**	**0.77**	**0.78**	0.0189	0.0194	0.0191
	Q_{t+12}	4-4-1	0.63	0.61	0.59	0.0205	0.0206	0.0207
	Q_{t+16}	4-6-1	**0.79**	**0.75**	**0.74**	0.0187	0.0190	0.0192

* The input variables for all of the models are identical (i.e., *Qa, Qd2, Qd3* and *Qd4).*
**The result has been presented for the best structure.

scale, a strong relationship is expected between runoff values at two antecedent time steps. Thus, adding the general runoff time series with one day lag time (Q_t) as an input neuron in daily modeling, in addition to the extracted sub-series might lead to more accurate results. For example, daily db4-based decomposition using the Delaney Creek Sub-basin data led to impose Q_t, *Qd4*, *Qd2*, *Qd3 and Qd5* time series into the FFNN to predict Q_{t+1}. The results of sensitivity analysis on the extracted sub-series for the Delaney Creek and Payne Creek Sub-basins are shown in Table 6, using both daily and monthly data. As presented in Table 6, the structure (the neurons of input and hidden layers of FFNN) in monthly modeling was simpler than daily case.

Table 7 demonstrats the most effective structures of EWANN models using different mother wavelets (Haar, db2 and db4) for both daily and monthly data of the Delaney Creek and Payne Creek Sub-basins. Although the tabulated results in Table 7 illustrate that models on the basis of db4 wavelet decomposition were more accurate than Haar and db2 based models, it was notable that when db2 was applied, the structure of network (the neurons of input and hidden layers), was simpler than other cases. There were several jumps in the runoff time series because of sudden start and cessation of rainfall over the related watershed. Therefore due to the shape of db4 mother wavelet, which is similar to the runoff signal, it could capture the signal features, especially peak points, well and could lead to reliable outcomes. Consequently, it is strongly recommended that the used mother wavelet type is selected according to the formation of the main time series. As tabulated in Table 7, since majority of accurate models included *Qd3*,

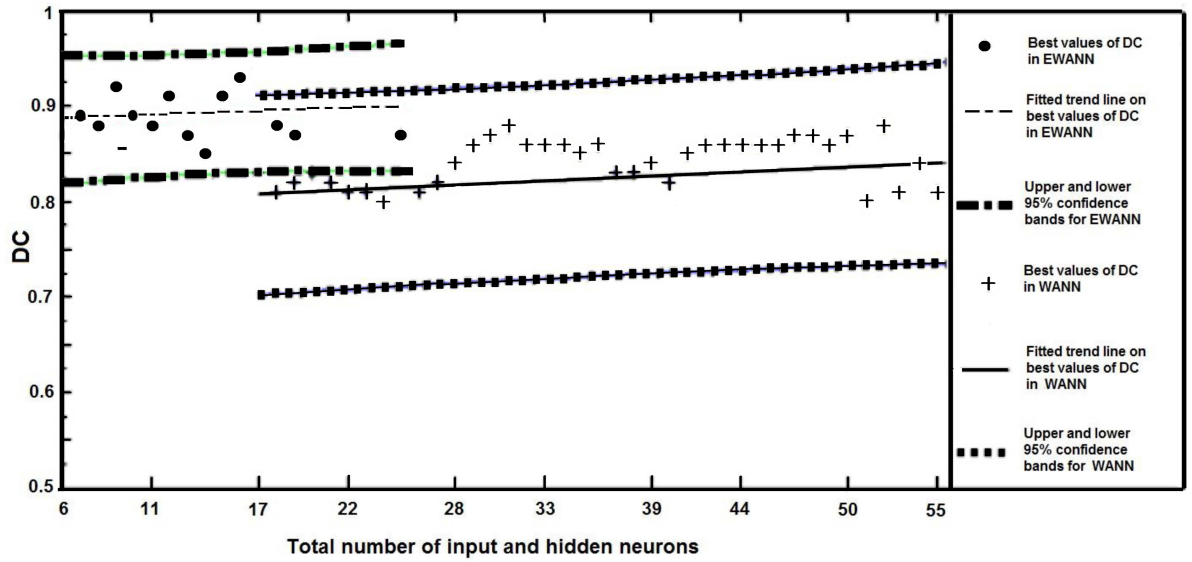

Figure 4. Uncertainty analysis to compare the structural features of seasonal models.

Table 9. Comparison of Best Results Obtained by Different Models for Daily and Monthly Modeling of Delaney Creek and Payne Creek Sub-basins

Sub-basin	Modeling time scale	Name and type of mode*		Epoch No.	Structure**	DC			RMSE (normalized)		
						Cal.	Test	Val.	Cal.	Test	Val.
Delaney Creek	Daily	ARIMAX	Linear	--	(3, 0, 1)P(t)	0.76	0.68	0.69	0.021	0.022	0.022
		FFNN	Non-linear	280	(3,8,1)	0.89	0.80	0.82	0.017	0.018	0.018
		WANN	Seasonal non-linear	770	(16,35,1)	0.93	0.83	0.88	0.008	0.017	0.015
		EWANN	Optimum seasonal non-linear	220	(5,9,1)	**0.95**	**0.92**	**0.94**	**0.008**	**0.010**	**0.009**
	Monthly	ARIMAX	Linear	--	(3, 1, 2)P(t)	0.78	0.66	0.65	0.017	0.022	0.023
		FFNN	Non-linear	170	(1,4,1)	0.81	0.73	0.73	0.018	0.019	0.019
		WANN	Seasonal non-linear	470	(8,18,1)	0.83	0.80	0.84	0.014	0.019	0.018
		EWANN	Optimum seasonal non-linear	210	(3,5,1)	**0.95**	**0.90**	**0.89**	**0.008**	**0.011**	**0.011**
Payne Creek	Daily	ARIMAX	Linear	--	(3, 0, 1)P(t)	0.72	0.64	0.63	0.022	0.023	0.025
		FFNN	Non-linear	320	(3,9,1)	0.77	0.73	0.74	0.019	0.019	0.019
		WANN	Seasonal non-linear	730	(16,38,1)	0.88	0.84	0.86	0.014	0.018	0.017
		EWANN	Optimum seasonal non-linear	230	(5,7,1)	**0.92**	**0.89**	**0.91**	**0.010**	**0.011**	**0.010**
	Monthly	ARIMAX	Linear	--	(2, 1, 1)P(t)	0.63	0.55	0.54	0.021	0.023	0.023
		FFNN	Non-linear	190	(1,3,1)	0.70	0.63	0.64	0.019	0.020	0.019
		WANN	Seasonal non-linear	380	(8,14,1)	0.87	0.81	0.83	0.014	0.018	0.017
		EWANN	Optimum seasonal non-linear	190	(3,5,1)	**0.91**	**0.85**	**0.86**	**0.011**	**0.015**	**0.015**

*Q_{t+1} is output variable.
** Best result has been presented among several structures.

Qd2 and *Qa* in their input layer, they were the most dominant sub-series to develop daily model of the process. In other words, *Qa* (trend or approximation sub-series), *Qd2* ($2^2 = 4$ days) and *Qd3* ($2^3 = 8$ days, approximately weekly) modes as dominant

daily frequencies of the process, were more suitable than other sub-series to be used as input variables. It was remarkable that, *Qd1* (2^1 = 2 months) and *Qd3* (2^3 = 8 months) modes were recognized as dominant monthly frequencies of the process. It is also inferred from Table 7 that almost all of dominant inputs are belonged to the runoff sub-series.

In order to evaluate the ability of EWANN method in long-term prediction horizon, the multi-step-ahead approach was also attached to EWANN model for both study areas. According to the climatic similarity of the study areas, the maximum reduction rate method determined *Qa*, *Qd2* (2^2 = 4 days), *Qd3* (2^3 = 8 days) and *Qd4* (2^4 = 16 days) modes sub-series as dominant inputs of EWANN model among all potential sub-series resulted in wavelet-based decomposition utilizing Haar mother wavelet and decomposition level 7 (for instance) for daily data. To evaluate the authority of *MI*-based feature extraction method in long term prediction, the determined dominant sub-series were imposed as inputs of multi-step-ahead EWANN model; and to survey the ability of the approach for detecting the frequencies involved in rainfall-runoff process, various lead time steps in runoff values were assumed as outputs of the model. The values of efficiency criteria presented in Table 8 reconfirm 4, 8 and 16 days modes as dominant daily frequencies of the process for both study areas as obtained in the multi-step-ahead FFNN modeling (see section 4.1). Comparison of the results tabulated in Tables 7 and 8 reveals that although the multi-step-ahead EWANN could increase the prediction horizon with regard to the FFNN model (see Table 3), the values of *DC* were deteriorated with regard to the single-step-ahead EWANN modeling.

4.3. Comparison of the Models

In order to evaluate the efficiency of the proposed daily and monthly models, a comparison among the best results of EWANN, classic FFNN, ad hoc WANN (without screening step) and classic ARIMAX(p,d_d,q)$P(t)$ models was also conducted for both sub-basins and the results presented in Table 9.

In this study, precipitation ($P(t)$) and antecedents of runoff data were used as inputs to predict future runoff as the output, moreover various values were examined as the ARIMAX parameters p, q and d. Like other models, the ARIMAX model was first calibrated using the calibration and test data sets, and the calibrated model was then verified using the validation data. As an example, the ARIMAX(3, 0, 1)$P(t)$ refers to a model that contains p = 3 autoregressive parameter and q = 1 moving average parameter, which were computed for the time series after it was differenced d = 0 times.

Tabulated results in Table 9 indicate that the ARIMAX model, due to its linear inherence, was unable to reliably handle complex non-linear rainfall-runoff process. Although ad hoc FFNN model was more efficient than the ARIMAX model, it only considers short term autoregressive features of the process and could not capture long-term frequency. Therefore it led to lower performance with compared when the FFNN was linked to wavelet and feature extraction concepts as EWANN model.

According to the results presented in Table 9, because of several saasonalities involved in rainfall-runoff process, the structure of EWANN model is slightly complex than ad hoc FFNN model. Such complexity is acceptable since according to Table 9 drastically increased the values of DC and RMSE with regard to FFNN model. Furthermore, the basic superiority of the proposed EWANN approach over the WANN model is the optimum structure (the number of epochs and the neurons of input and hidden layers) of the model and thereinafter is the reduced numbers of modeling parameters (weights and bias). The other superiorities of the EWANN in comparison with classic WANN model are no necessity to the time consuming trial-error procedure to extract dominant sub-series and significant reduction in the noise of model due to the ability of entropy to detect the redundant information.

As a statistical clarification on the structural optimization occurred in the EWANN approach over the classic WANN model, an uncertainty analysis was performed on the best architectures of models presented in Figure 4 (utilizing the outcomes of db4 mother wavelet for instance). According to Figure 4 although WANN models led to extensive limitation of confidence band, EWANN models could reach more performance with less input and hidden neurons (optimized structure). Figure 4 clearly presents that confidence band of EWANN and WANN approaches overlap when the total numbers of input and hidden neurons were considered between 17 and 26. In such overlapping limitation, lower confidence band of proposed EWANN was slightly better than mean of WANN outcomes, the fact directly denote to the superiority of EWANN over WANN approach. Furthermore, WANN because of high number of input layer (16 sub-series in the current study) could not reach simple structures.

Although the study areas are climatically similar, the topography of the watersheds are markedly distinct (see section 2.4). Therefore as tabulated in Table 7, the proposed EWANN method could handle the properties of the flat sub-basin (Delaney Creek) slightly accurately than the wild watershed (Payne Creek). EWANN models contain dominant *MI*-based sub-series as input neurons of model to have insight into physics of rainfall-runoff process. Therefore Table 7 illustrates that in contrast to FFNN model (see Table 2) where the approach led to almost unreliable efficiency criteria for a sloppy and wild watershed (Payne Creek), the proposed EWANN approach could effectively decrease undesirable effects of topographic variety of the study area. Consequently the extracted features of process (dominant sub-series) could implicitly represent the physical conditions of the watersheds via a black box framework and cope with the topographic distinctions of the watersheds.

On the other hand, time of concentration as a hydrologic quantity depends on physical parameters of catchment. It seems, time of concentration in both catchments (i.e., Delaney Creek and Payne Creek) have direct relation with the lag times of input time series in FFNN models, and applied subseries, which are indicators of frequencies in structure of rainfall and runoff time series, in inputs of EWANN models. It is expected that according to fairly similar time of concentration at both sub-basins (i.e., *230 minutes* for Delaney Creek and *250 minutes*

for Payne Creek), FFNN and EWANN models' inputs for efficient DC criterion are the same, which the results in Tables 2 and 7 admit this expectation. It is noted that since in black box rainfall-runoff modeling elimination of base flow from runoff hydrograph or calculation of excess rainfall is not needed, so that the proportion of times of concentrations for the catchments in this study is determinant. If the modeling procedure is performed for an event of rainfall that excess rainfall hyetograph as well as direct runoff hydrograph have been computed, the lag times of models should be the same as concentration time not the proportion of concentration times.

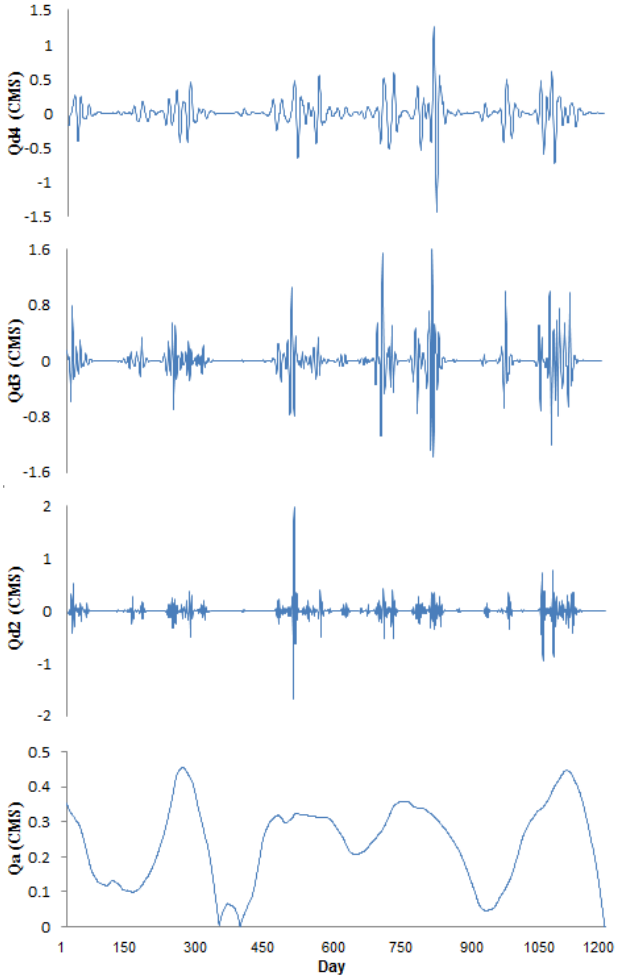

Figure 5. Validation daily data of dominant sub-series (derived by db4 mother wavelet and decomposition level 7) utilized in the EWANN model for the Delaney Creek Sub-basin.

The obtained results also indicated the sufficiency of proposed wavelet-based feature extraction method using *MI* with regard to other used criteria (i.e., *H*, *E* and *CC*) to select dominant input variables for FFNN modeling of rainfall-runoff process. The nature of *MI* is similar to FFNN models; supervised and non-linear. *CC* is also a supervised criterion but can only detect linear correlation among data and is more appropriate for the linear models. Thus, in wavelet-based dominant feature extraction for FFNN-based rainfall-runoff modeling, *CC* could not perform as well as *MI* but slightly reliable than unsupervised *E* and *H* criteria.

Besides, it could be deduced that although the data preprocessing procedure by the wavelet transform can improve the performance of EWANN models in both time scales with regard to classic FFNN model, the percentage of this improvement is more considerable in monthly scale modeling for both study areas (as presented in Table 9, percentage of the improvement in the EWANN modeling with regard to the FFNN approach is up to 23% for daily data and up to 35% in monthly time scale). Such outcome is reasonable since the autoregressive feature is more significant in the daily modeling whereas the frequency feature is dominant factor in the monthly modeling.

The most effective result of EWANN model in daily models (for instance) was obtained via db4 mother wavelet, decomposition level 7 and for the Delaney Creek Sub-basin with calibration, testing and validation *DC* values equal to 0.95, 0.92 and 0.94, respectively. Figure 5 presents extracted dominant sub-series of validation data for the mentioned model (i.e., *Qa, Qd2, Qd3* and *Qd4*) also the computed and observed runoff time series of the model for calibration, test and validation data sets have been presented in Figure 6. Furthermore, Figure 7 shows the scatter plot of observed and computed runoff values. Both figures denote the good agreement between the observed and computed runoff values using the proposed EWANN model.

To evaluate the performance of FFNN and EWANN models in multi-step-ahead modeling of the process, tabulated results in Tables 3 and 8 were compared. The link of wavelet and entropy concepts via the structure of multi-step-EWANN model could represent the properties of watersheds and rainfall-runoff process and yielded to improved values of efficiency criteria with regard to multi-step-FFNN approach. Therefore, the multi-step-EWANN could efficiently cope with the frequencies involved in rainfall-runoff process.

5. Concluding Remarks

In this study, the wavelet transform was used to decompose the Delaney Creek and Payne Creek Sub-basins rainfall and runoff daily and monthly time series at decomposition levels 3 and 7, respectively. Thereafter, the sub-series were ranked via two statistical entropy-based criteria (*MI* and *H*) and two classic criteria (*CC* and *E*) to determine dominant frequencies to be considered as the inputs of the FFNN to forecast runoff values. Four different measures (i.e., *H, MI, CC* and *E*) were led to different dominant sub-series in feature extraction step. The results show merit of *MI* criterion (non-linear and supervised) to select dominant wavelet-based inputs in both daily and monthly types for the non-linear FFNN rainfall–runoff modeling.

Ad hoc neural network models may simply be unable to cope with non-linearity and non-stationary if wavelet-based pre-processing of the input is not performed. Thus, the applied wavelet transform not only removes probable noise as a filter, but also distinguishes several features and frequencies invol-

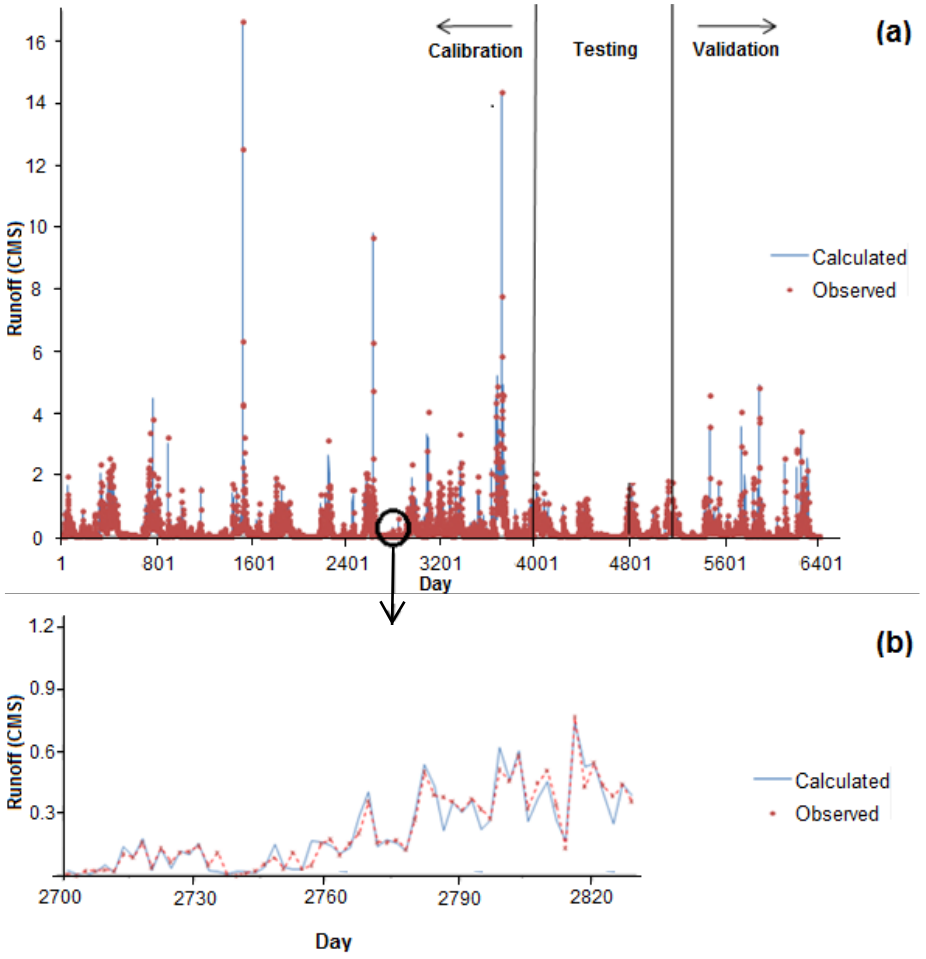

Figure 6. Observed and computed daily runoff values by EWANN model for the Delaney Creek Sub-basin, (a) general view, (b) a detailed view.

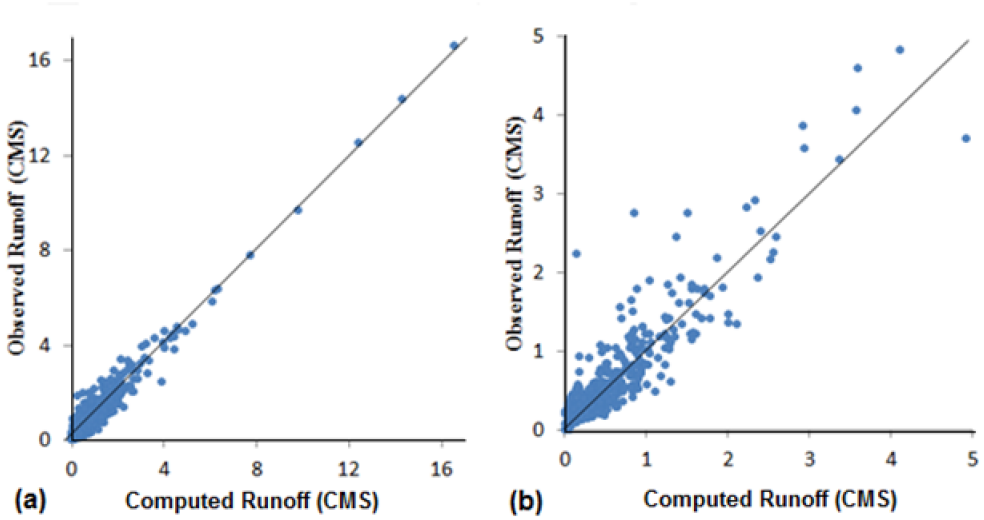

Figure 7. Scatter plots of the EWANN model for daily data of the Delaney Creek Sub-basin in a) Calibration, and b) Validation, steps.

ved in the process by decomposition of time series at different levels (scales).

Although the monthly EWANN showed lower performance with compared to daily EWANN model (may be due to more uncertainty involved in monthly data which could be detected by the concept of entropy), implication of wavelet analysis was more beneficial in monthly modeling than daily simulation. The reason may be due to that the frequency is the most governed pattern in monthly rainfall-runoff time series whereas the autoregressive pattern is the main trait in the daily time series

According to Table 9 all forecasting models which are developed for two sub-basins of the Delaney Creek and Payne Creek revealed more accurate results for Delaney Creek which involves flat topographic condition in comparison to Payne Creek. Since Payne Creek sub-basin includes the sloppy topography and wild situation of area with regard to the Delaney Creek sub-basin, the underlying physical parameters that are involved in transformation of rainfall to runoff are more complex. In as much as the approach of the black box based models are to extract underlying pattern in data, the more complex the physical parameters, the difficult the pattern extraction. Accordingly, the black box based models (i.e., ARIMAX, FFNN, WANN, and EWANN) could deal efficiently with the data of flat sub-basin than sloppy wild sub-basin. Amongst the several black box models of current research EWANN model revealed the highest efficiency, but it is considerable that although results of evaluation criteria in EWANN model for Delaney Creek sub-basin with flat topography and subsequently less physical complexity, is slightly good in comparison to Payne Creek, the difference between these two outcomes (i.e., Delany Creek 0.95 and Payne Creek 0.92 in daily DC calibration criterion) is not dominant. This denotes the capability of EWANN model in selecting appropriate inputs to cope with the complex physical inherent of the Payne Creek sub-basin.

The distinctive improvement of efficiency criteria and simplified structure of the proposed EWANN model (according to Table 9) were illustrated as preponderances of EWANN (single-step and multi-step-ahead) in comparison to the conventional FFNN (single-step and multi-step-ahead) and WANN models, respectively. Furthermore, EWANN methodology could decrease the undesirable effects of topographic variety on the performance of model which could be cited as the other authority of the proposed approach with regard to both of classic models.

As a research plan and due to social and economic importance of the study areas, it is suggested to couple the proposed ANN-based model to a geomorphology-based rainfall-runoff model (e.g., TOPMODEL, Nourani and Zanardo, 2014) to have a comprehensive semi-distributed rainfall-runoff modeling for the site.

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., and Wilby, R.L. (2012). Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. *Prog. Phys. Geogr.,* 36(4), 480-513. http://dx.doi.org/10.1177/0309133312444943

Adamowski, J. (2008). Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. *J. Hydrol.,* 353(3-4), 247-266. http://dx.doi.org/10.1016/j.jhydrol.2008.02.013

Adamowski, J., Chan, H.F., Prasher, S.O., and 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. Hydroinf.,* 14(3), 731-744. http://dx.doi.org/10.2166/hydro.2011.044

Amorocho, J., and Espildora, B. (1973). Entropy in the assessment of uncertainty in hydrologic systems and models. *Water Resour. Res.,* 9(6), 1551-1522. http://dx.doi.org/10.1029/WR009i006p01511

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000). Artificial neural networks in hydrology. II: Hydrologic applications. *J. Hydrol. Eng.,* 5(2), 124-137. http://dx.doi.org/10.1061/(ASCE)1084-0699(2000)5:2(124)

Aussem, A., Campbell, J., and Murtagh, F. (1998). Wavelet-based feature extraction and decomposition strategies for financial forecasting. *J. Comput. Intell. Finance,* 6(2), 5-12.

Box, G.E.P., and Jenkins, G. (1976). *Time Series Analysis: Forecasting and Control*, Second ed., San Francisco, Holden-Day.

Cannas, B., Fanni, A., See, L., and Sias, G. (2006). Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. *Phys. Chem. Earth (A,B,C),* 31(18), 1164-1171. http://dx.doi.org/10.1016/j.pce.2006.03.020

Caselton, W.F., and Husain, T. (1980). Hydrological networks: Information transmission. *J. Water Resour. Plann. Manage.,* 106(WR2), 503-520.

Chang, F.J., Chiang, Y.M., and Chang, L.C. (2007). Multi-step-ahead neural networks for flood forecasting. *Hydrol. Sci. J.,* 52(1), 114-130. http://dx.doi.org/10.1623/hysj.52.1.114

Cover, T.M., and Thomas, J.A. (1991). *Elements of Information Theory*, John Wiley & Sons, New York. http://dx.doi.org/10.1002/0471200611

Ebrahimi, N., Maasoumi, E., and Soofi, E.S. (1999). Ordering univariate distributions by entropy and variance. *J. Econ.,* 90(2), 317-336. http://dx.doi.org/10.1016/S0304-4076(98)00046-3

Gao, Z., Gu, B., and Lin, J. (2008). Monomodal image registration using mutual information based methods. *Image Vision Comput.,* 26(2), 164-173. http://dx.doi.org/10.1016/j.imavis.2006.08.002

Hagan, M.T., and Menhaj, M.B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Networks,* 5(6), 989-993. http://dx.doi.org/10.1109/72.329697

Harmancioglu, N.B., and Singh, V.P. (1998). *Entropy in Environmental and Water Resources*, Kluwer Academic Publishers, Dordrecht.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feed forward networks are universal approximators. *Neural Networks,* 2(5), 359-366. http://dx.doi.org/10.1016/0893-6080(89)90020-8

Krstanovic, P.F., and Singh, V.P. (1992). Evaluation of rainfall networks using entropy: I. Theoretical development. *Water Resour. Manage.,* 6(4), 279-293. http://dx.doi.org/10.1007/BF00872281

Labat, D. (2005). Recent advances in wavelet analyses: Part 1. A review of concepts. *J. Hydrol.,* 314(1-4), 275-288. http://dx.doi.org/10.1016/j.jhydrol.2005.04.003

Legates, D.R., and McCabe, G.J. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.,* 35(1), 233-241. http://dx.doi.org/10.1029/1998WR900018

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.,* 2(2), 164-168.

Maheswaran, R., and Khosa, R. (2012a). Multiscale nonlinear model for monthly stream flow forecasting: A wavelet-based approach. *J.*

Hydroinf., 14(2), 424-442. http://dx.doi.org/10.2166/hydro.2011.130

Maheswaran, R., and Khosa, R. (2012b). Comparative study of different wavelets for hydrologic forecasting. *Comput. Geosci.,* 46, 284-295. http://dx.doi.org/10.1016/j.cageo.2011.12.015

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

MathWorks, Inc. (2010). *MATLAB: User's Guide, Version 7*, The Math works, Inc., Natick.

May, R.J., Maier, H.R., Dandy, G.C., and Gayani Fernando, T.M.K. (2008). Non-linear variable selection for artificial neural networks using partial information. *Environ. Model. Software,* 23(10-11), 1312-1326. http://dx.doi.org/10.1016/j.envsoft.2008.03.007

Mishra, A.K., Özger, M., and Singh, V.P. (2009). An entropy-based investigation into the variability of precipitation. *J. Hydrol.,* 370(1-4), 139-154. http://dx.doi.org/10.1016/j.jhydrol.2009.03.006

Morchen, F. (2003). *Time Series Feature Extraction for Data Mining Using DWT and DFT*, Technical Report, Department of Mathematics and Computer Science, University of Marburg, Germany.

Nourani, V. (2010). Reply to comment on 'Nourani V, Mogaddam AA, Nadiri AO. 2008. An ANN-based model for spatiotemporal groundwater level forecasting. Hydrological Processes 22: 5054-5066'. *Hydrol. Process.,* 24(3), 370-371. http://dx.doi.org/10.1002/hyp.7469

Nourani, V., and Zanardo, S. (2014). Wavelet-based regularization of the extracted topographic index from high-resolution topography for hydro-geomorphic applications. *Hydrol. Process.,* 28(3), 1345-1357. http://dx.doi.org/10.1002/hyp.9665

Nourani, V., Komasi, M., and Mano, A. (2009). A multivariate ANN-wavelet approach for rainfall-runoff modeling. *Water Resour. Manage.,* 23(14), 2877-2894. http://dx.doi.org/10.1007/s11269-009-9414-5

Nourani, V., Kisi, O., and Komasi, M. (2011). Two hybrid artificial intelligence approaches for modeling rainfall-runoff process. *J. Hydrol.,* 402(1-2), 41-59. http://dx.doi.org/10.1016/j.jhydrol.2011.03.002

Nourani, V., Baghanam, A.H., and Gebremichael, M. (2012). Investigating the ability of Artificial Neural Network (ANN) models to estimate missing rain-gauge data. *J. Environ. Inf.,* 19(1), 38-50. http://dx.doi.org/10.3808/jei.201200207

Nourani, V., Baghanam, A.H., Adamowski, J., and Gebremichael, M. (2013). Using self-organizing maps and wavelet transforms for space-time pre-processing of satellite precipitation and runoff data in neural network based rainfall-runoff modeling. *J. Hydrol.,* 476, 228-243. http://dx.doi.org/10.1016/j.jhydrol.2012.10.054

Rajaee, T., Nourani, V., Zounemat-Kermani, M., and Kisi, O. (2011). River suspended sediment load prediction: Application of ANN and wavelet conjunction model. *J. Hydrol. Eng.,* 16(8), 613-627. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000347

Safehian, J., Akbarzadeh, A., and Imani, B.M. (2012). Application of wavelet thresholding filter to improve multi-step ahead prediction model for hydraulic system. *Adv. Mater. Res.,* 488-489, 1783-1787. http://dx.doi.org/10.4028/www.scientific.net/AMR.488-489.1783

Sang, Y.F. (2013). A review on the applications of wavelet transform in hydrology time series analysis. *Atmos. Res.,* 122, 8-15. http://dx.doi.org/10.1016/j.atmosres.2012.11.003

Shannon, C.E. (1948). A mathematical theory of communications I and II. *Bell Syst. Tech. J.,* 27(3), 379-443. http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shoaib, M., Shamseldin, A.Y., and Melville, B.W. (2014). Comparative study of different wavelet based neural network models for rainfall-runoff modeling. *J. Hydrol.,* 515, 47-58. http://dx.doi.org/10.1016/j.jhydrol.2014.04.055

Singh, V.P. (1997). The use of entropy in hydrology and water resources. *Hydrol. Process.,* 11(6), 587-626. http://dx.doi.org/10.1002/(SICI)1099-1085(199705)11:6<587::AID-HYP479>3.0.CO;2-P

Singh, V.P. (2011). Hydrologic synthesis using entropy theory: Review. *J. Hydrol. Eng.,* 16(5), 421-433. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000332

Singh, R.M. (2012). Wavelet-ANN model for flood events. *Adv. Intell. Soft Comput.,* 131, 165-175. http://dx.doi.org/10.1007/978-81-322-0491-6_16

Yang, H.H., Vuuren, S.V., Sharma, S., and Hermansky, H. (2000). Relevance of time-frequency features for phonetic and speaker-channel classification. *Speech Commun.,* 31(1), 35-50. http://dx.doi.org/10.1016/S0167-6393(00)00007-8