Scientific Research and Essays Vol. 7(17), pp. 1690-1703, 9 May, 2012
Available online at http://www.academicjournals.org/SRE
DOI: 10.5897/SRE12.110
ISSN 1992-2248 © 2012 Academic Journals

Full Length Research Paper

Using wavelet transform to improve generalization capability of feed forward neural networks in monthly runoff prediction

Umut Okkan

Department of Civil Engineering, Balikesir University, Balikesir, Turkey. E-mail: umutokkan@balikesir.edu.tr. Tel: 0090 0266 612 11 94-95.

Accepted 23 April, 2012

In the study presented, a hybrid model is proposed for monthly runoff prediction by using wavelet transform and feed forward neural networks. Discrete wavelet transform (DWT) and Levenberg-Marquardt optimization algorithm based feed forward neural networks (FFNN) are considered for the modeling study. The study region covers the basins of Medar River which is located at the Aegean region of Turkey. Meteorological data, which represent the study region, were decomposed into wavelet sub-time series by DWT. Ineffective sub-time series were eliminated by using Mallow C_p coefficient based all possible regression method to prevent collinearity. Then, effective sub-time series components constituted the new inputs of FFNN. Some favorite evaluation measures, that is, determination coefficient (R^2), adjusted determination coefficient (Adj.R^2), Nash-Sutcliffe efficiency coefficient (NS), root mean squared error (RMSE), weighted mean absolute percentage error (WMAPE), were employed to assess modeling performances. The results determined in study indicate that the DWT based FFNN models (DWT-FFNN) are successful tools to model the monthly runoff series and can give good prediction performances than conventional methods.

Key words: Wavelet transform, feed forward neural networks, Levenberg-Marquardt algorithm, monthly runoff prediction.

INTRODUCTION

Over the past years, artificial intelligence methods have been widely used in the modeling and prediction of hydrological variables. Especially, artificial neural networks (ANN), which is a nonlinear black box model inspired by the biological learning process of the brain, have been accepted an effective tool for modeling a complex hydrologic system (Hsu et al., 1995; Dawson and Wilby, 1998; Campolo et al., 1999; Coulibaly et al., 2000; Cigizoglu, 2005; Kisi, 2004, 2008a; Cigizoglu and Alp, 2006; Diamantopoulou et al., 2007; Razavi and Araghinejad, 2009; Okkan, 2011; Fistikoglu and Okkan, 2011; Okkan and Dalkilic, 2011; Okkan and Serbes, 2011). A comprehensive review of the usage of ANN in hydrology is given by ASCE Task Committee on Application of the Artificial Neural Networks in Hydrology (ASCE, 2000a, b).

Although ANN methods had been used extensively as useful tools for prediction of hydrological variables, it has also many drawbacks to deal with non-stationary data (Cannas et al., 2006; Partal, 2009). Therefore, some hybrid modeling approaches which include different data-preprocessing and combine techniques have been developed to increase generalization capability of ANN. Approaches for dealing with non-stationary characteristics of data are not as highly generated, nor as well proved, as those for hydrological prediction problems. In the last years, there has been an interest in hybrid modeling techniques. Chaotic neural networks (Karunasinghe and Liong, 2006), set pair analysis (SPA) and principle component analysis (PCA) based neural networks (Wang et al., 2006a; Wang et al., 2006b; Wu et al., 2008), the threshold neural networks (Wang et al.,

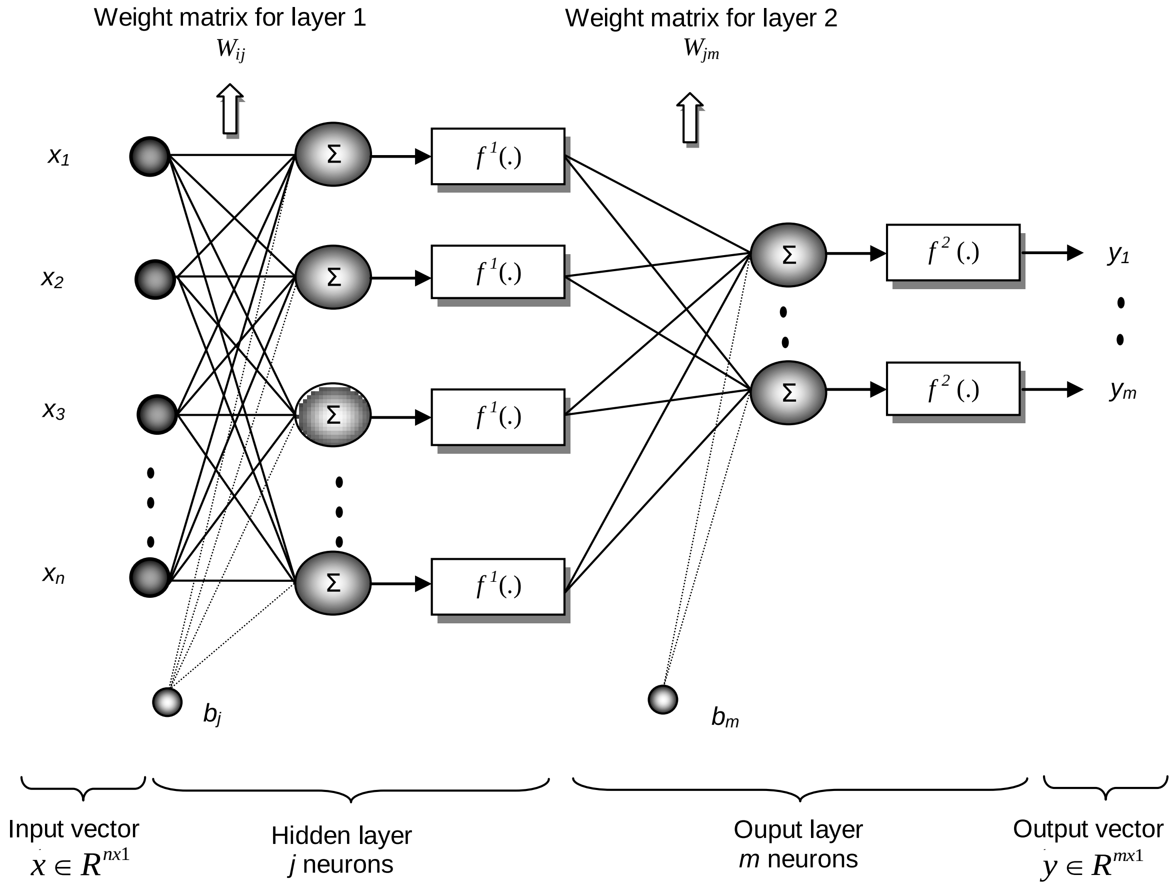

Figure 1. FFNN structure.

2006a), the cluster-based hybrid neural networks (Cigizoglu and Kisi, 2005; Wang et al., 2006a; Wu et al., 2008) were successfully applied to hydrological variables.

Recently, wavelet transform, which is another data-pre-processing technique, showed successful performance in hydrological applications. There are some appreciable studies of wavelet transform based neural network models (Li et al., 1999; Wang and Ding, 2003; Anctil and Tape, 2004; Cannas et al., 2006; Kisi, 2008b; Wang et al., 2009; Rajaee, 2010; Rajaee et al., 2011). The wavelet transform is also integrated with multiple linear regression (Kucuk and Agiralioglu, 2006; Kisi, 2009, 2010) and support vector machine approach (Kisi and Cimen, 2011). Each of these studies showed that different black box models trained or calibrated with decomposed data resulted in higher accuracy than the single models that were calibrated with an undecom-posed and noisy time series.

The main purpose of the study presented is to examine the applicability and generalization capability of the wavelet-feed forward neural networks combined model for the prediction of the monthly runoff values of a study region which is an important water resource for the Gediz Basin/Turkey, and to compare its performance with single feed forward neural networks, multiple linear regression (MLR), the wavelet-MLR combined model and another data pre-processing technique called principle component analysis based neural networks. Some favorite performance evaluation measures are employed to assess developed models.

MATERIALS AND METHODS

Feed forward neural networks

There are many papers and books which provide a detailed description of the FFNN (Haykin, 1994; Hagan and Menhaj, 1994; Ham and Kostanic, 2001; Partal et al., 2008; Fistikoglu and Okkan, 2011; Okkan and Serbes, 2011), and hence only a brief description of FFNN is given here. The running procedure of FFNN involves typically two phases; forward computing and backward computing.

In forward computing, each layer uses a weight matrix associated with all the connections made from the previous layer to the next layer (Figure 1).

The hidden layer has the weight matrix W_{ij} and activation function $f^{(1)}$; the output layer has the weight matrix W_{jm} and activation function $f^{(2)}$. Given the network input vector $x \in R^{nx1}$, the output of the output layer, which is the response (output) of the network $y \in R^{mx1}$, can be written as:

$$y_m = f^{(2)}\left\{\sum_{j=1}^{m}\left[f^{(1)}\left(\sum_{i=1}^{n}x_i W_{ij} + b_j\right)\right]W_{jm} + b_m\right\} \quad (1)$$

After the phase of forward computing, backward computing, which depends on the algorithms to adjust weights, is carried out. The process of adjusting these weights to minimize the differences between the actual and the desired output values is called training or learning of network. If these differences (errors) are higher than the desired values, the errors are passed backwards through the weights of the network. In ANN terminology, this phase is also called the back propagation algorithm.

Depending on the techniques to train FFNN models, different back propagation algorithms have been developed. In this study, the Levenberg-Marquardt back propagation algorithm was used in training of the FFNN. The Levenberg-Marquardt back propagation algorithm is a second-order nonlinear optimization technique that is usually faster and more reliable than any other back propagation techniques (Coulibaly et al., 2000; Kisi, 2004; Cigizoglu and Kisi, 2005; Fistikoglu and Okkan, 2011; Okkan, 2011).

The Levenberg-Marquardt optimization algorithm represents a simplified version of Newton method applied to the training of FFNN (Hagan and Menhaj, 1994). The training process can be viewed as finding a set of weights that minimize the error (e_p) for all samples in the training set (T). The performance function is a sum of squares of the errors as follows:

$$E(W) = \frac{1}{2}\sum_{p=1}^{P}(d_p - y_p)^2 = \frac{1}{2}\sum_{p=1}^{P}(e_p)^2,\ P = mT \quad (2)$$

where, T is the total number of training samples, m is the number of output layer neurons, W represents the vector containing all the weights in the network, y_p is the actual network output, and d_p is the desired output.

When training with the Levenberg-Marquardt algorithm, the changing of weights ΔW can be computed as follows:

$$\Delta W_k = -[J_k^T J_k + \mu_k I]^{-1} J_k^T e_k \quad (3)$$

Then, the update of the weights can be adjusted as follows:

$$W_{k+1} = W_k + \Delta W_k \quad (4)$$

where, J is the Jacobian matrix, I is the identify matrix, e is the network error, μ is the Marquardt parameter which is to be updated using the decay rate β depending on the outcome. In particular, μ is multiplied by the decay rate β ($0<\beta<1$) whenever $E(W)$ decreases, while μ is divided by β whenever $E(W)$ increases in a new step (Coulibaly et al., 2000; Fistikoglu and Okkan, 2011; Okkan, 2011; Okkan and Dalkilic, 2011).

Wavelet transform

The wavelet transform, developed during the last decades, is a decomposition method. This method provides an analyzing way of a signal in both time and frequency and appears to be a more successful than the conventional Fourier transforms that do not provide time-frequency analysis for the variables involve non-stationary signals (Daubechies, 1990; Kucuk and Agiralioglu, 2006; Partal and Cigizoglu, 2009; Partal, 2009; Kisi and Partal, 2011). Assuming a continuous time series x(t), wavelet function $\psi_{s,\tau}$(t), called the mother wavelet, can be defined as $\int_{-\infty}^{\infty}\psi(t)\,dt = 0$. $\psi_{s,\tau}$ can be obtained through compressing and expanding $\psi(t)$ (Equation 5).

$$\psi_{s,\tau}(t) = |s|^{-1/2}\psi\left(\frac{t-\tau}{s}\right) \quad \tau \in R,\ s \in R, s \neq 0 \quad (5)$$

where s is scale or frequency factor, τ is time factor, R is the domain of real number.

$\psi_{s,\tau}$ (t) must have zero mean and be localized in both time and Fourier space (Wang and Ding, 2003; Kucuk and Agiralioğlu, 2006; Partal, 2009; Kisi, 2009). Wavelet transform of x(t) is also written as:

$$\psi_{s,\tau}(t) = |s|^{-1/2}\int_R x(t)\,\overline{\psi}\left(\frac{t-\tau}{s}\right)dt \quad \tau \in R,\ s \in R, s \neq 0 \quad (6)$$

where $\overline{\psi}$ is complex conjugate functions of ψ. Equation (6) describes that wavelet transform is the decomposition of x(t) under different resolution scale (Wang and Ding, 2003; Kucuk and Agiralioglu, 2006; Kisi, 2011; Rajaee et al., 2011; Kisi and Partal, 2011).

For practical applications in hydrology, researchers have access to a discrete time signal, rather than to a continuous time signal (Rajaee et al., 2011). A discretization of Eq.(6) based on trapezoidal rule is perhaps the simplest discretization of the continuous wavelet transform. This transform produces N^2 coefficients from a data set of length N; hence unnecessary information is locked up within the coefficients; which may or may not be an appealing attributes (Rajaee et al., 2011).To overcome this uncertainty, discrete wavelet transforms (DWT) which present power of two logarithmic scaling of the translations can be used in practical applications (Mallat, 1989; Kucuk and Agiralioğlu, 2006; Partal, 2009).

According to the Mallat algorithm, the discrete wavelet transform of discrete time series x_i is written as:

$$W_{j,k} = 2^{-j/2}\sum_{i=0}^{N-1}x_i\psi(2^{-j}i - k) \quad (7)$$

where i is integer time steps, j and k are integers that control, respectively, the scale and time; $W_{j,k}$ is wavelet coefficient for the scale factor $s = 2^j$ and the time factor $\tau = 2^j k$.

In DWT method, the time series (x_i) passes through two filters and are decomposed into wavelet sub-time series components without losing the information about the instant of the element occurrence (Mallat, 1989; Kucuk and Agiralioglu, 2006; Kisi, 2009). The DWT converts a signal into father and mother wavelets. Father wavelets represent the high-scale, low frequency components (approximation (A) components). Mother wavelets are representations of the low-scale, high frequency components (detail (D) components). Thus, DWT allows one to study different investing behaviors in different time scales independently (Ma, 2006; Partal, 2009; Rajaee et al., 2011). The sub-time series can be computed by using Equation (7). The DWT decomposition of a time series is presented in Figure 2.

Study region and data

In hydrological modeling studies, the determination of suitable input variables would play an important role in their applications. In the

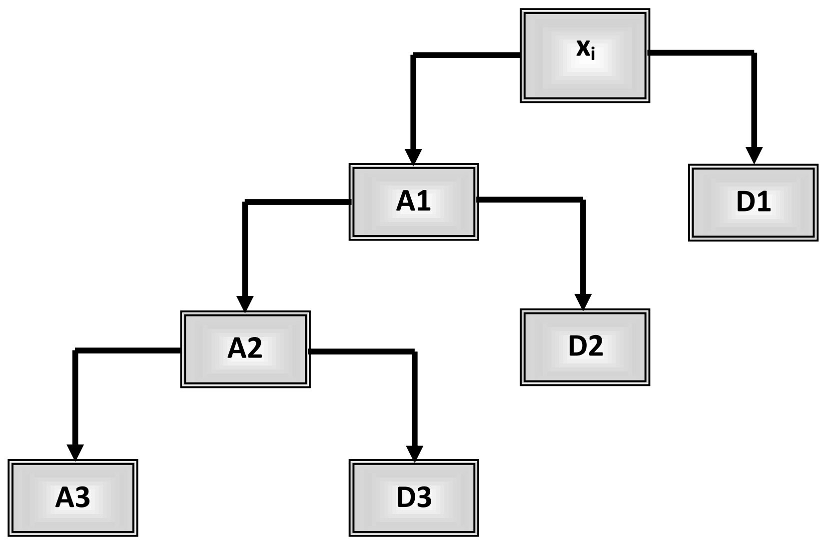

Figure 2. Decomposition of DWT.

study, modeling strategy that predicts outputs from inputs based on monthly rainfall and temperature data. In addition to concurrent values of these data, delaying process that rainfall transforms into runoff, in other words the representational monthly groundwater storage was considered and one-month-ahead rainfall values were also used in modeling studies. Thus, three input data (P_t, T_t, P_{t-1}) were prepared for the same periods of the runoff records (P_t: monthly rainfall; P_{t-1}: one-month-ahead rainfall T_t: monthly temperature).

The application area covers the drainage basin of Medar River, which is located in the Gediz Basin of Turkey. The monthly runoff data of Medar River were observed by Kayalioglu flow gauging station (EIE 509) of II. Regional Directorate of State Hydraulic Works for the period between 01.10.1961 and 01.09.2005 (1962-2005 water years). The monthly data of rainfall and temperature at Akhisar and Sarilar meteorological stations, which are represent the drainage basin of Medar River, were obtained from the State Meteorological Organization of Turkey. Next, Thiessen weighted precipitation values and arithmetical mean temperature values were prepared for monthly timescale, using records available at stations.

Modeling strategy developed in the study

In the study, discrete wavelet transform (DWT) was linked to feed forward neural networks (FFNN) for monthly runoff prediction. First, the input data (P_t, P_{t-1}, T_t) of training and testing periods were decomposed into a certain number of sub-time series components by DWT. The selection of the optimal decomposition level is one of the keys to determine the performance of model in wavelet domain. Decomposition level is generally based on signal characteristics and experiences to selection. For example, Chou and Wang (2004) showed that using only one decomposition level to model the streamflow time series does not easily represent the process. Kisi and Cimen (2011) preferred three decomposition levels in their monthly streamflow forecasting study. For the monthly scaled hydrological modeling studies, three decomposition levels can nearly represent annual period of the related time series. Similarly, three decomposition levels were considered in this study. Thus, meteorological input data were decomposed using appropriate wavelet functions (mother wavelets) and twelve sub-time series components (time series of 2-months mode (D1), 4-months mode (D2), 8-months mode (D3) and approximation mode (A3)) were obtained for the training and testing period.

This study also aimed at examining the effects of the employed mother wavelet type on the efficiency of proposed models. For this purpose, performances of four different kinds of wavelet, i.e., the Haar wavelet (simple wavelet-db1), Daubechies-2 (db2) wavelet (the most popular wavelet), Daubechies-6 (db6) wavelet and Daubechies-10 (db10) wavelet are investigated in the present study (Figure 3). For example, the three levels decomposition of the P_t and T_t signals that yield four sub-signals by the Daubechies-2 (db2) wavelet are shown in Figures 4 and 5.

After the decomposition processes, effective sub-time series components should be determined. Some correlated sub-time series components may reduce the generalization capabilities of the feed forward neural networks; thus, the effective sub-time series components that represented the inputs of the feed forward neural networks were treated so as to reduce collinearity as much as possible. In this study, this was carried out using Mallow's C_p based all possible regression method as this is an effective way to determine the subset of variables in cases where there are a large number of potential predictor variables (Efroymson, 1960; Mallows, 1973).

Mallow's C_p is a measure of the error in the best subset model, relative to the error incorporating all variables. Adequate models are those for which C_p is roughly equal to the number of variables in the model (Mallows, 1973). The C_p values can be computed using Equation (8).

$$C_p = (N-k)\frac{MSE_i^{\ 2}}{MSE_F^{\ 2}} - (N - 2i - 1) \quad (8)$$

where N is the number of data, MSE_i^2 is the mean of residual squares in the model with i variable, $MSE_F^{\ 2}$ is the mean of residual squares in the full model with k variable.

In some studies, new series which are obtained by summing the effective sub-time series were used as input to the models (Partal et al., 2008; Partal and Cigizoglu, 2009; Kisi, 2010; Kisi and Cimen, 2011). Unlike these studies, components of sub-time series determined with the all possible regression method are used as individual separate model inputs in this study.

The modeling strategy developed in this study is summarized in

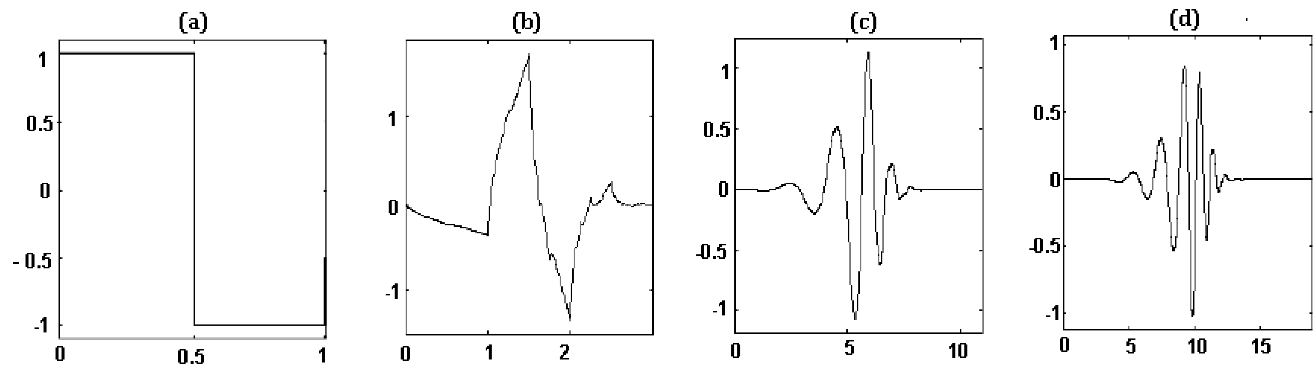

Figure 3. (a) The Haar wavelet; (b) Daubechies-2 (db2) wavelet; (c) Daubechies-6 (db6) wavelet; (d) Daubechies-10 (db10) wavelet.

Figure 6.

Data Normalization and Assessment of Model Performances

The input and output data are normalized to prevent the model from being dominated by the variables with large values, as is commonly used in artificial intelligence models. In this study, the normalization processes of all data were carried out using Equation (9).

$$z_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{9}$$

where z_i is the scaled value, x_i is the unscaled data, x_{min} and x_{max} are, respectively, the minimum and maximum values of the unscaled data.

Some favorite approaches are suggested for hydrological time series prediction accuracy assessment according to literature related to training and testing of models. In this study, five performance measures were considered.

All models with optimum structures of them provided the best training result in terms of the minimum root mean squared error (*RMSE*), weighted mean absolute percentage error (WMAPE) and the maximum determination coefficient (R^2), adjusted determination coefficient (*Adj.*R^2), and Nash-Sutcliffe coefficient (*NS*) were also employed for the testing period (Equations 10 to 14).

RMSE evaluates the residual between desired and output data, and WMAPE measures the mean absolute percentage error of the prediction. R^2 and *Adj.*R^2 evaluate the linear relation between desired and output data, while *NS* evaluates the capability of the model in simulating output data away from the mean statistics.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(d_t - y_t)^2} \tag{10}$$

$$\mathrm{WMAPE} = \frac{\sum_{t=1}^{n}\left|\frac{d_t - y_t}{d_t}\right| \mathrm{x}\, d_t}{\sum_{t=1}^{n} d_t} \tag{11}$$

$$\mathrm{R}^2 = \frac{\sum_{t=1}^{n}(d_t - d_{mean})^2 - \sum_{t=1}^{n}(d_t - y_t)^2}{\sum_{t=1}^{n}(d_t - d_{mean})^2} \tag{12}$$

$$AdjR^2 = 1 - \frac{(n-1)}{(n-p-1)}(1-R^2) \tag{13}$$

$$NS = 1 - \frac{\sum_{t=1}^{n}(d_t - y_t)^2}{\sum_{t=1}^{n}(d_t - d_{mean})^2} \tag{14}$$

where n is the number of training or testing samples, y_t is the model output, d_t is the observed (desired) data in the t^{th} time period, d_{mean} is the mean over the observed periods, and p is the number of the inputs of models.

RESULTS

In the application, a MATLAB code which involves Mallat's DWT algorithm and four different kinds of wavelet (db1, db2, db6 and db10) was prepared. Levenberg-Marquardt algorithm based FFNN model was also developed by a MATLAB code. To evaluate the generalization capability of all models, data set consisted of two periods. 44 years input-output data were used and divided into training and testing periods by proportions of 1/2 (October 1961 to September 1983) and 1/2 (October, 1984 to September, 2005), respectively. According to modeling strategy, input data of training and testing periods were decomposed into a certain number of sub-time series components by DWT firstly. Afterward, the effective variables were selected by the Mallows C_p coefficients for each mother wavelet type. To evaluate the strength and direction of the relations between variables, different linear regression analysis combinations in training period were obtained using the ''all possible regression method'' tool in Minitab software with twelve sub-time series components which derived from the input data. Performances of the models with optimum variables (seen as bold characters in Tables 1 to 4) are nearly the same as that of the full linear models (for k = 12) ; that is, the explained variance of the monthly runoff values, which have the minimum C_p values, are nearly equal to

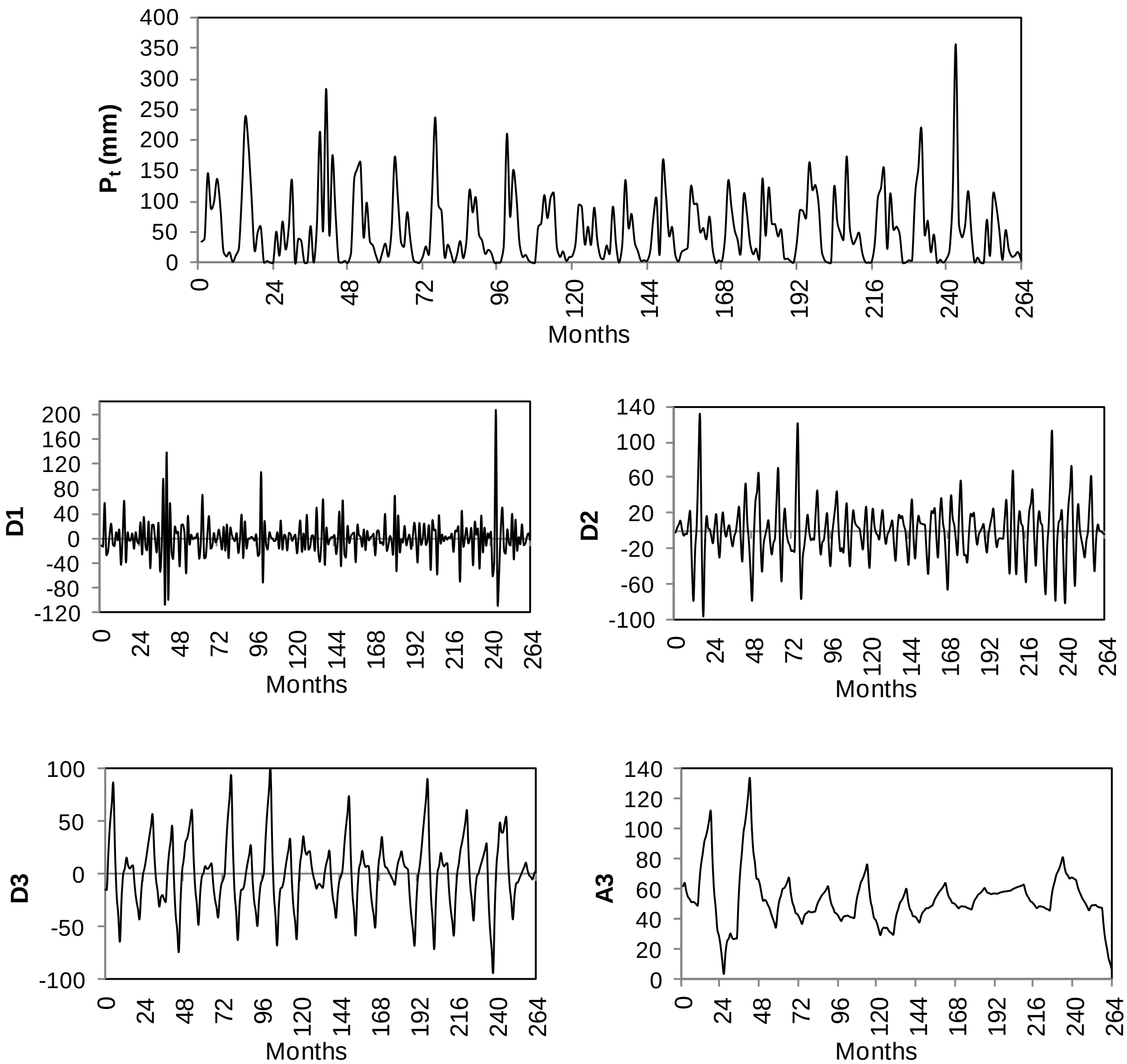

Figure 4. Original time series, 2-months mode of time series (D1), 4-months mode of time series (D2), 8-months mode of time series (D3) and approximation mode of time series (A3) of monthly rainfall for training period (mother wavelet type : db2).

that explained by the full linear models. Before presenting these determined effective sub-time series as input data to FFNN, the all data were normalized using Equation 9 to prevent the models from being dominated by the variables with the extreme values.

In the DWT based FFNN (DWT-FFNN) and single FFNN applications, the optimal number of neuron in the hidden layer was determined using a trial and error approach by varying the number of neurons from 2 to 20. Various types of the activation function are possible for FFNN modeling but sigmoid function was preferred for each layer in study; and the training epochs were set to 20. In the Levenberg-Marquardt algorithm based FFNN, the values of μ_0 and β can be taken to be 0.01 and 0.1 respectively. These values were fixed during the whole training process in this study. These values are also recommended by Hagan and Menhaj (1994). The DWT-FFNN results were compared with multiple linear regressions (MLR), DWT based MLR (DWT-MLR) and principle component analysis based FFNN (PCA-FFNN). According to the Eigen-analysis of the correlation matrix, Eigen values and variance proportions for the input vector (P_t, P_{t-1}, T_t) are computed as 2.011, 0.6553, 0.3329 and 0.671, 0.218, 0.111, respectively. Thus, cumulative proportions were obtained as 0.671 and 0.889 for the first and second components, which were considered in PCA-FFNN application. The description and application on the hydrology of PCA can be found in Wu et al. (2008).

Table 5 presents the results of the study region, in terms of different performance measures. It can be seen from Table 5 that DWT based FFNN models have good

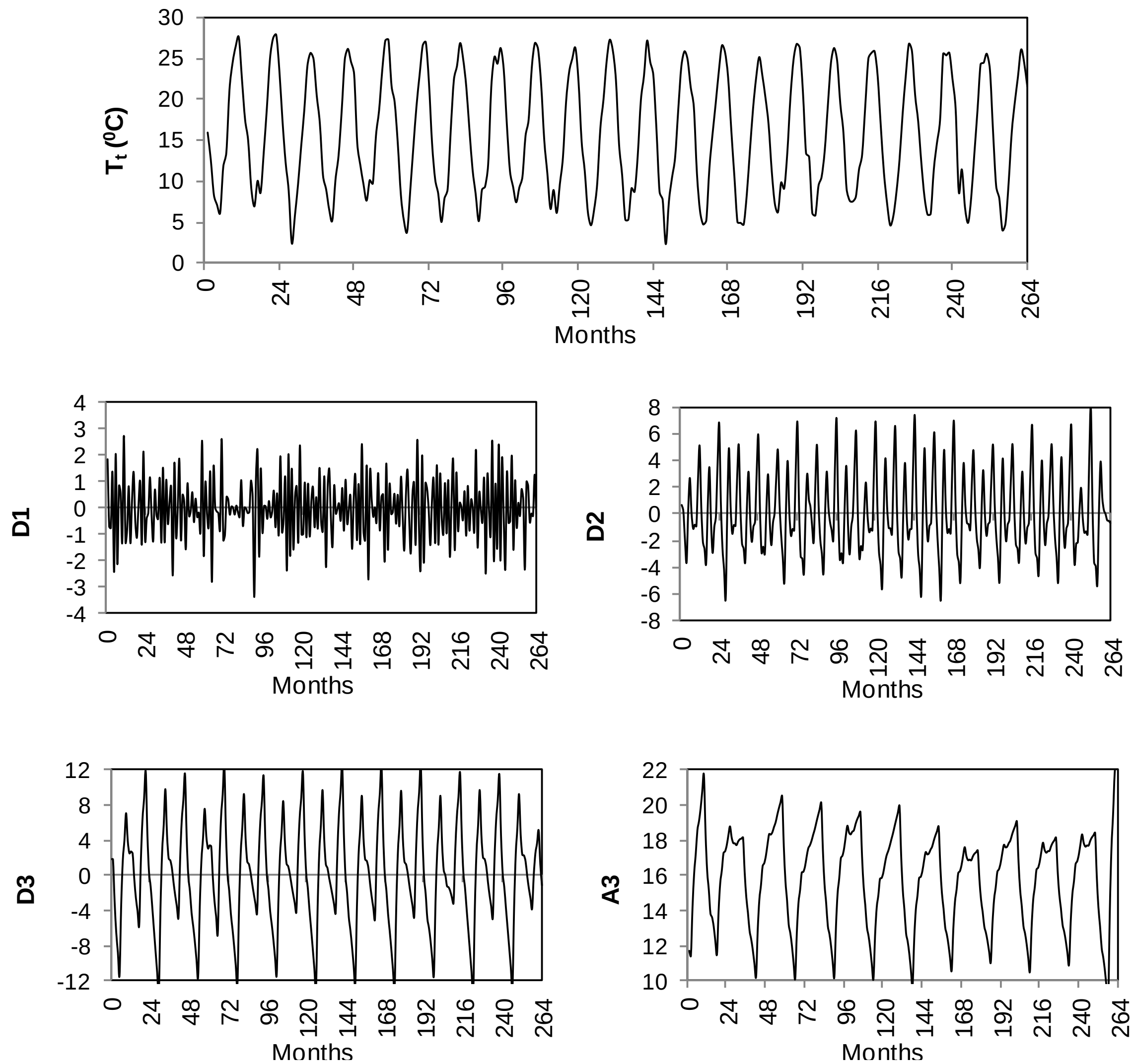

Figure 5. Original time series, 2-months mode of time series (D1), 4-months mode of time series (D2), 8-months mode of time series (D3) and approximation mode of time series (A3) of monthly temperature for training period (mother wavelet type : db2).

performances during both training and testing, and they outperform MLR, DWT-MLR and PCA-FFNN models in terms of the performance measures. In the training period, the DWT-FFNN (8, 6, 1) model with db-2 wavelet obtained the best RMSE, WMAPE, R^2, Adj.R^2, and NS statistics of 3.5146 mm, 0.2087, 0.9551, 0.9537, and 0.9548, respectively; while the FFNN (3, 6, 1) model obtained the best maximum value statistics of 117.26 mm. Investigating the results during testing period, it can be seen that the same model with the same mother wavelet type outperforms all other models.

According to these results, DWT-FFNN model with db-2 wavelet is able to obtain the better prediction accuracy in terms of different performance measures during the training and testing periods. The scatter plot and hydrograph of the best DWT-FFNN model and single FFNN model developed in this study during the testing period are shown in Figures 7 and 8. When the testing period scatter graphs of the models are examined, it is observed that the standard deviations around the y=x line are far less in the DWT-FFNN models. In other words, when $y=ax+b$ fitted lines in graph are examined, it is observed that, in DWT-FFNN model, "*a*" gets closer to the value 1, and "*b*" gets closer to the value 0, compared to the FFNN model.

In addition to long-term statistics of these models, seasonal box-plot presentations were also examined. Seasonal mean, minimum, maximum and, median statistics of the models and the observed data are shown in Figure 9. When the box-plot graph is examined, it can be seen that that DWT-FFNN model proves itself better than FFNN model for all seasons. When summer and

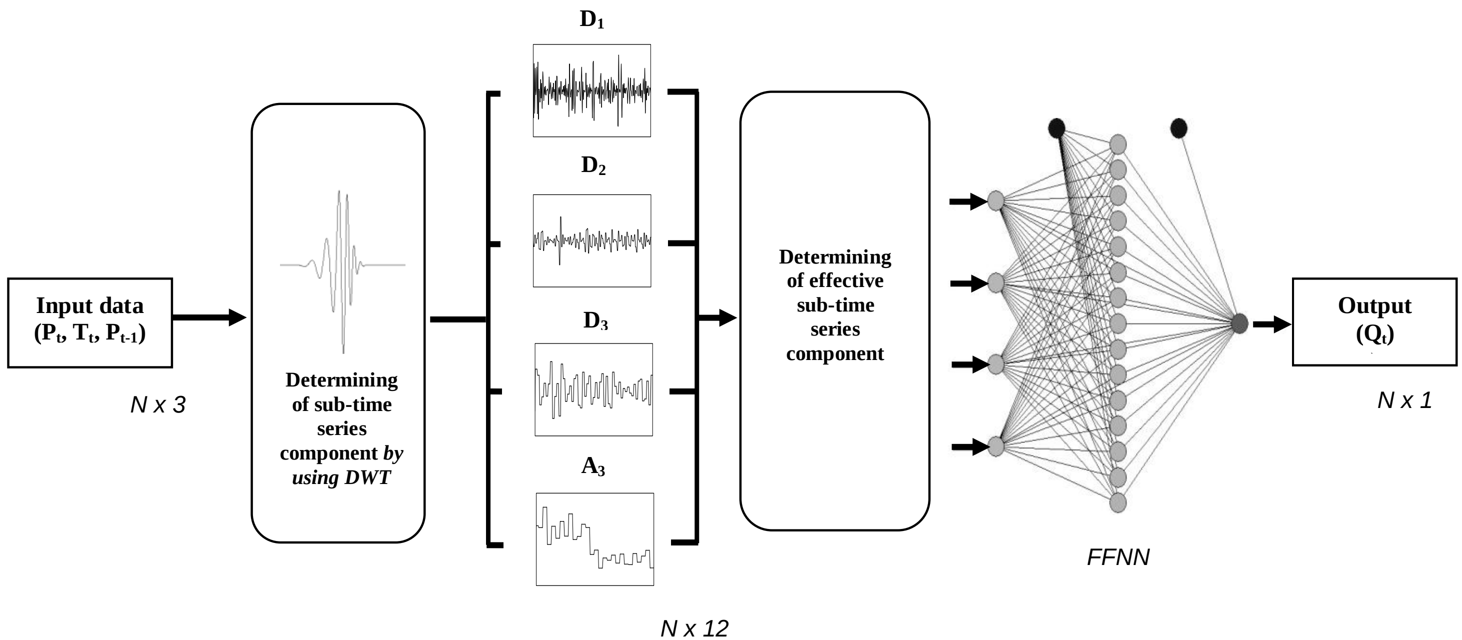

Figure 6. The modeling strategy developed in the study.

Table 1. Determination of Effective Sub-time Series Components for the Haar wavelet (db1).

Number of Components	R^2 (%)	Adj R^2 (%)	C_p	$D1\_P_t$	$D2\_P_t$	$D3\_P_t$	$A3\_P_t$	$D1\_P_{t-1}$	$D2\_P_{t-1}$	$D3\_P_{t-1}$	$A3\_P_{t-1}$	$D1\_T_t$	$D2\_T_t$	$D3\_T_t$	$A3\_T_t$
1	20.7	20.4	475							X					
2	40.7	40.3	291.2		X					X					
3	57.4	56.9	138.9		X					X	X				
4	64.3	63.7	76.8	X	X					X	X				
5	67.4	66.8	49.7	X	X				X	X	X				
6	69	68.2	37.7	X	X				X	X	X			X	
7	70.3	69.5	26.9	X	X				X	X	X			X	X
8	71.5	70.6	18	X	X	X			X	X	X			X	X
9	72.3	71.4	12.2	X	X	X		X	X	X	X			X	X
10	72.7	71.6	11.1	X	X	X		X	X	X	X		X	X	X
11	**72.9**	**71.7**	**11.1**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**	**X**
12	72.9	71.6	13	X	X	X	X	X	X	X	X	X	X	X	X

Table 2. Determination of Effective Sub-time Series Components for the Daubechies-2 (db2) wavelet.

Number of Components	R^2 (%)	Adj R^2 (%)	C_p	$D1\_P_t$	$D2\_P_t$	$D3\_P_t$	$A3\_P_t$	$D1\_P_{t-1}$	$D2\_P_{t-1}$	$D3\_P_{t-1}$	$A3\_P_{t-1}$	$D1\_T_t$	$D2\_T_t$	$D3\_T_t$	$A3\_T_t$
1	37.4	37.2	441.9							X					
2	55.4	55	242.5							X	X				
3	64	63.6	148	X						X	X				
4	72.5	72.1	54.5	X	X					X	X				
5	73.9	73.4	40.6	X	X					X	X				X
6	75	74.5	30	X	X				X	X	X		X		
7	76.5	75.8	16.2	X	X				X	X	X		X		X
8	**77.5**	**76.8**	**6.6**	**X**	**X**		**X**		**X**	**X**	**X**		**X**		**X**
9	77.5	76.7	8	X	X		X	X	X	X	X		X		X
10	77.6	76.7	9.8	X	X	X	X		X	X	X		X	X	X
11	77.6	76.6	11.2	X	X	X	X	X	X	X	X		X	X	X
12	77.6	76.6	13	X	X	X	X	X	X	X	X	X	X	X	X

Table 3. Determination of Effective Sub-time Series Components for the Daubechies-6 (db6) wavelet.

Number of Components	R^2 (%)	Adj R^2 (%)	C_p	$D1\_P_t$	$D2\_P_t$	$D3\_P_t$	$A3\_P_t$	$D1\_P_{t-1}$	$D2\_P_{t-1}$	$D3\_P_{t-1}$	$A3\_P_{t-1}$	$D1\_T_t$	$D2\_T_t$	$D3\_T_t$	$A3\_T_t$
1	42.8	42.6	369.6							X					
2	58.6	58.3	197.1							X	X				
3	67.9	67.5	97.6	X						X	X				
4	71.2	70.7	63.3	X					X	X	X				
5	72.6	72.1	49.3	X			X		X	X					X
6	74.2	73.6	33.7	X	X				X	X	X		X		
7	75.6	74.9	20.7	X	X		X		X	X			X		X
8	76.3	75.6	14.5	X	X		X	X	X	X			X		X
9	**76.9**	**76.1**	**10**	**X**	**X**		**X**	**X**	**X**	**X**	**X**		**X**		**X**
10	77	76	11.5	X	X		X	X	X	X	X		X	X	X
11	77.2	76.2	11.1	X	X	X	X	X	X	X	X		X	X	X
12	77.2	76.1	13	X	X	X	X	X	X	X	X	X	X	X	X

autumn statistics are taken as a basis specially, DWT-FFNN represents the distribution of runoff data well.

DISCUSSION

In this study, periodicity characters of the meteorological data were considered since time series decomposed by DWT produce detailed information about the endogenous structure of the

Table 4. Determination of Effective Sub-time Series Components for the Daubechies-10 (db10) wavelet.

Number of Components	R^2 (%)	Adj R^2 (%)	C_p	$D1\_P_t$	$D2\_P_t$	$D3\_P_t$	$A3\_P_t$	$D1\_P_{t-1}$	$D2\_P_{t-1}$	$D3\_P_{t-1}$	$A3\_P_{t-1}$	$D1\_T_t$	$D2\_T_t$	$D3\_T_t$	$A3\_T_t$
1	41.8	41.6	327.4							X					
2	57.2	56.8	174.5							X	X				
3	66.8	66.4	79.4	X						X	X				
4	71.8	71.3	31.1	X					X	X	X				
5	73.3	72.8	17.3	X			X		X	X					X
6	74.2	73.6	10.8	X	X		X		X	X					X
7	74.5	73.8	9.9	X	X				X	X	X		X		X
8	**74.7**	**73.9**	**9.7**	**X**	**X**		**X**		**X**	**X**	**X**		**X**		**X**
9	74.9	74	9.8	X	X		X		X	X	X		X	X	X
10	75	74	10.3	X	X	X	X		X	X	X		X	X	X
11	75.1	74	11.1	X	X	X	X	X	X	X	X		X	X	X
12	75.1	73.9	13	X	X	X	X	X	X	X	X	X	X	X	X

data. The effective ones of the decomposed series were determined by Mallow C_p based all possible regression analysis (Tables 1 to 4) and then imposed as an input vector to MLR and FFNN to model the rainfall-runoff relation of a basin in Turkey.

The Levenberg-Marquardt back propagation (BP) algorithm was used in training of the FFNN models. Other fast BP algorithms such as conjugate gradient and scaled conjugate gradient may be preferred.

The effective decomposed series can be also selected by other statistical approaches. However, Mallows C_p based all possible regression analysis is a quite effective way to determine the subset of decomposed series in cases where there are a large number of potential predictors.

The contribution of an appropriate mother wavelet type was also investigated. Four different popular wavelet types (db1, db2, db6 and db10) were used with three decomposition levels. In addition to these, other Daubechies's wavelets and some irregular wavelets such as Bior1.1, Rboi1.1, Coif1, Sym3, and Meyer wavelets may be used in similar applications.

Table 5 presents the predictions performed by the different kind of models. Because the phenomenon of the hydro-meteorological time series have the inherent complexities and nonlinearities, the performance of single MLR model was not suitable. On the contrary, DWT increased the efficiency of all MLR and FFNN models. The best model structures were also provided in Table 5. Here, DWT-FFNN model with db-2 wavelet denotes a three layered model comprising eight inputs, six hidden and one output node. This model is able to obtain the better prediction accuracy in terms of different performance measures and seasonal statistics during the training and testing periods.

Although FFNN has an ability to model complex and nonlinear relations, the structure of it is hard to determine and it can be determined using a trial-and-error approach. Therefore, DWT based MLR (DWT-MLR) models that may be much easier to interpret are used as an alternative way to FFNN for monthly rainfall-runoff modeling.

In addition to the input data used in the study, to increase the performance of the model, previous runoff series (Q_{t-1}, Q_{t-2},…Q_{t-p}) can be included into the models, considering the autocorrelation effect. The data used in the study are regarded to be sufficient considering the model parsimony for the models to enable them to function properly.

In the study, the Mallat algorithm was used for the DWT of monthly meteorological time-series. The other DWT algorithms (e.g., Trous algorithm) and the other types of wavelet transform (e.g., fast wavelet transform) may be used for the construction of the different neuro-computing models (e.g., support vector machines, other neural network algorithms, fuzzy logic) for the future studies. In addition to these, other data-preprocessing techniques (e.g., factor analysis, set pair analysis) may be compared with the wavelet transform.

Table 5. Prediction performance measures of different models for the training (a) and testing (b) periods.

(a) Training

Model	Mother wavelet type	RMSE (mm)	WMAPE	R^2	Adj. R^2	NS	Min (mm)	Max (mm)
Observed	-	-	-	-	-	-	0.37	119.00
MLR	-	9.9378	0.6014	0.6388	0.6347	0.6388	-7.28	75.13
DWT-MLR	Haar (db1)	8.6077	0.5340	0.7290	0.7172	0.7290	-10.17	81.72
DWT-MLR	db2	7.8463	0.4963	0.7749	0.7678	0.7749	-11.33	72.38
DWT-MLR	db6	7.9450	0.5207	0.7692	0.7610	0.7692	-12.89	69.84
DWT-MLR	db10	8.3221	0.5421	0.7467	0.7388	0.7467	-10.66	70.42
FFNN (3, 6, 1)	-	7.4692	0.4242	0.7961	0.7938	0.7960	1.30	**117.26**
PCA-FFNN (2, 9, 1)	-	7.4621	0.4079	0.8004	0.7988	0.7964	0.73	115.48
DWT-FFNN (11, 7, 1)	Haar (db1)	5.1559	0.2881	0.9036	0.8994	0.9028	1.06	114.08
DWT-FFNN (8, 6, 1)	db2	**3.5146**	**0.2087**	**0.9551**	**0.9537**	**0.9548**	**0.57**	117.09
DWT-FFNN (9, 9, 1)	db6	4.6746	0.2668	0.9225	0.9198	0.9201	0.72	103.80
DWT-FFNN (8, 7, 1)	db10	5.5132	0.3032	0.8900	0.8866	0.8888	0.72	98.80

(b) Testing

Model	Mother wavelet type	RMSE (mm)	WMAPE	R^2	Adj. R^2	NS	Min (mm)	Max (mm)
Observed	-	-	-	-	-	-	0.01	75.30
MLR	-	7.4743	0.8302	0.5731	0.5682	0.4210	-8.69	55.97
DWT-MLR	Haar (db1)	7.2792	0.8675	0.6364	0.6205	0.4509	-17.78	60.55
DWT-MLR	db2	6.5913	0.8270	0.6939	0.6843	0.5498	-11.58	49.94
DWT-MLR	db6	7.6831	0.9366	0.6218	0.6083	0.3882	-16.78	50.47
DWT-MLR	db10	7.8111	0.9999	0.5999	0.5873	0.3677	-13.79	48.18
FFNN (3, 6, 1)	-	6.4689	0.7215	0.6304	0.6262	0.5663	1.30	**74.75**
PCA-FFNN (2, 9, 1)	-	5.5330	0.5819	0.7035	0.7012	0.6827	0.69	64.11
DWT-FFNN (11, 7, 1)	Haar (db1)	6.4405	0.5791	0.6732	0.6589	0.5701	0.28	66.12
DWT-FFNN (8, 6, 1)	db2	**4.7174**	**0.4263**	**0.8047**	**0.7986**	**0.7694**	0.11	67.87
DWT-FFNN (9, 9, 1)	db6	6.1506	0.5365	0.7178	0.7078	0.6079	**0.03**	64.13
DWT-FFNN (8, 7, 1)	db10	5.3493	0.5031	0.7342	0.7259	0.7034	0.18	60.27

Conclusion

The current study presented the application of DWT based models (DWT-MLR and DWT-FFNN) compared with FFNN and MLR models used undecomposed data, for modeling of monthly runoff of Medar River, based on the meteorological data. The results determined in study indicate that the DWT based methods are successful tools to model the monthly runoff series of the study region and can give good prediction performances than conventional models. Although DWT-FFNN and FFNN methods are powerful artificial intelligence techniques, DWT-MLR makes the running time considerably faster with an appropriate accuracy. In terms of the best accuracy, the DWT-FFNN model with db-2 mother wavelet resulted in RMSE and WMAPE

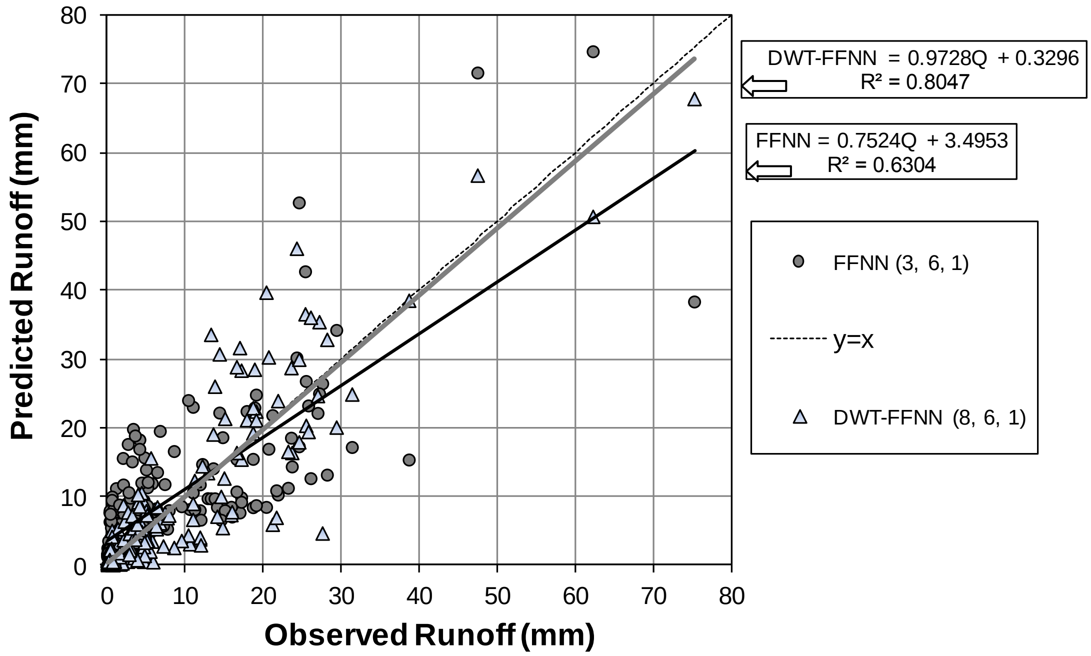

Figure 7. The scatter plots of FFNN (3, 6, 1) and DWT-FFNN (8, 6, 1) models in the testing period.

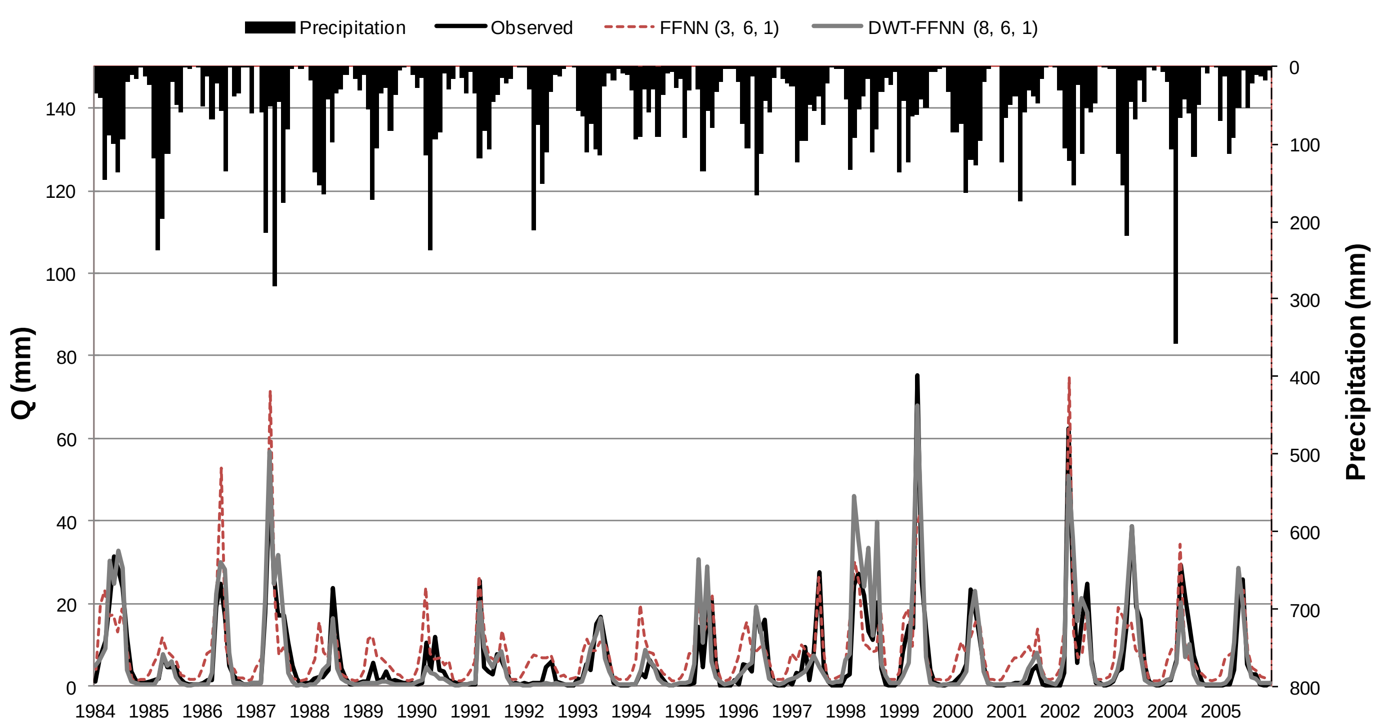

Figure 8. The hydrographs of FFNN (3, 6, 1) and DWT-FFNN (8, 6, 1) models in the testing period.

reductions and, R^2, Adj.R^2 and NS increases relative to that of the other models. This model has generalization capability and thus can more easily capture monthly runoff data. These results were improbable, however the explanation may be that it is able to simulate nonlinear of rainfall-runoff relationships in this arid region which is

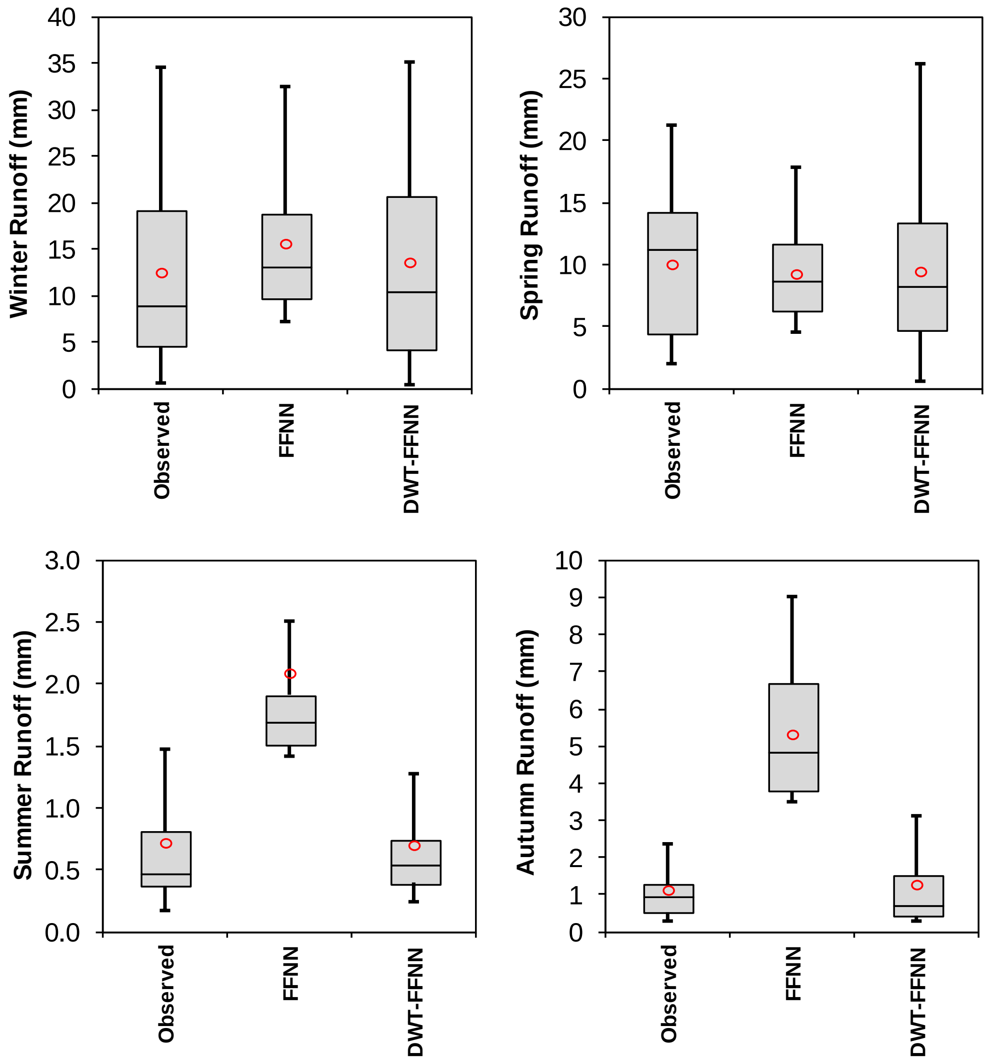

Figure 9. Box-plots of the observed and predicted seasonal mean runoff (o : mean value).

located at the Aegean coast of Turkey and have typical Mediterranean climate characteristics. For the future studies, the presented model structures could be used to model monthly/daily rainfall-runoff relations in some semi-arid regions of Turkey. The author also suggests that this technique can be also applied to other hydrological variables and other water resources problems to reconfirm the effectiveness of the approach.

REFERENCES

Anctil F, Tape DG (2004). An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition. J. Environ. Eng. Sci., 3: 121-128.
ASCE Task Committee (2000a). Artificial neural networks in hydrology – I: preliminary concepts, J. Hyd. Eng, ASCE, 5(2): 115-123.
ASCE Task Committee (2000b). Artificial neural networks in hydrology – II: hydrological applications, J. Hyd. Eng. ASCE, 5(2): 124-137.
Campolo M, Reussi P, Soldati A (1999). River flood forecasting with a neural network model. Water Res. Res., 35: 1191-1197.
Cannas B, Fanni A, See L, Sias G (2006). Data processing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. Phys. Chem. Earth. 31: 1164-1171.
Chou CM, Wang RY (2002). On-line estimation of unit hydrographs using the wavelet-based LMS algorithm. Hyd.Sci. J., 47(4): 721-738.
Cigizoglu HK, Alp M (2006). Generalized regression neural network in modelling river sediment yield. Adv. Eng. Software, 37: 63-68.
Cigizoglu HK, Kisi O (2005). Flow prediction by three back propagation techniques using k-fold partitioning of neural network training data. Nordic Hyd., 36(1): 49-64.
Cigizoglu HK (2005). Generalized regression neural networks in monthly flow forecasting. Civil Eng. Environ. Syst., 22(2): 71-84.
Coulibaly P, Anctil F, Bobee B (2000). Daily reservoir inflow forecasting using artificial neural Networks with stopped training approach. J. Hydrol. 230: 244-257.
Daubechies I (1990). The wavelet transform, time-frequency localization and signal analysis. IEEE Trans. Inf. Theory, 36(5): 961-1005.
Dawson CW, Wilby R (1998). An artificial neural network approach to rainfall-runoff modeling. Hydrol. Sci. J., 43: 47-66.
Diamantopoulou MJ, Georgiou PE, Papamichail DM (2007). Performance of neural network models with Kalman learning rule for flow routing in a river system. Fres. Environ. Bull., 16(11b): 1474-1484.

Efroymson MA (1960). Multiple regression analysis. Mathematical methods for digital computers, Ralston A and Wilf HS eds., Wiley, New York.

Fistikoglu O, Okkan U (2011). Statistical downscaling of monthly precipitation using NCEP/NCAR reanalysis data for Tahtali River Basin in Turkey. J. Hyd. Eng., 16(2): 157-164.

Hagan MT, Menhaj MB (1994). Training feed forward techniques with the Marquardt algorithm. IEEE Trans. Neural Networks, 5(6): 989-993.

Haykin S (1994). Neural Networks: A Comprehensive Foundation. MacMillan. New York.

Ham F, Kostanic I (2001). Principles of Neurocomputing for Science and Engineering. Macgraw-Hill. USA.

Hsu K, Gupta HV, Sorooshian S (1995). Artificial neural network modeling of the rainfall runoff process. Water Res., 31: 2517-2530.

Karunasinghe DSK, Liong SY (2006). Chaotic time series prediction with a global model: artificial neural network. J. Hydrol., 323: 92-105.

Kisi O, Cimen M (2011). A wavelet-support vector machine conjunction model for monthly streamflow forecasting. J. Hydrol., 399: 132-140.

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

Kisi O (2008a). River flow forecasting and estimation using different artificial neural network techniques. Hydrol. Res., 39(1): 27-40.

Kisi O (2008b). Streamflow forecasting using neuro-wavelet technique. Hyd. Process., 22(20): 4142-4152.

Kisi O (2009). Wavelet regression model as an alternative to neural networks for monthly streamflow forecasting. Hydrol. Process., 23: 3583-3597.

Kisi O (2010). Wavelet regression model for short-term streamflow forecasting. J. Hydrol., 389: 344-353.

Kisi O (2011). Evapotranspiration modeling using a wavelet regression model. Irr. Sci., 29: 241-252.

Kisi O, Partal T (2011). Wavelet and neuro-fuzzy conjunction model for streamflow forecasting. Hydrol. Res., 42(6): 447-456.

Kucuk M, Agiralioğlu N (2006). Wavelet regression technique for streamflow prediction. J. Appl. Stat., 33(9): 943-960.

Li X, Ding J, Li H (1999). Combining neural network models based on wavelet transformation. J. Hydrol. Eng., 2: 1-4.

Ma PY (2006). A fresh engineering approach for the forecast of financial index volatility and hedging strategies. PhD thesis, Quebec University, Montreal, Canada.

Mallat SG. (1989). A theory for multi resolution signal decomposition: the wavelet representation. IEEE Trans. Pattern. Anal. Mach. Intell., 11(7): 674-693.

Mallows CL (1973). Some comments on Cp. Technometrics, 15(4): 661-675.

Okkan U, Dalkilic HY (2011). Reservoir inflows modeling with artificial neural networks: the case of Kemer Dam in Turkey. Fres. Environ. Bull., 20(12): 3110-3119.

Okkan U (2011). Application of Levenberg-Marquardt optimization algorithm based multilayer neural networks for hydrological time series modeling. Int. J. Opt. Cont., 1(1): 53-63.

Okkan U, Serbes ZA (2011). Modeling of daily inflows of Keban Dam using artificial neural networks and stochastic methods. J. NWSA. 6(4): 1526-1539.

Partal T, Cigizoglu HK (2009). Estimation and forecasting of daily suspended sediment data using wavelet-neural networks. J. Hydrol., 358: 317-331.

Partal T (2009). Modelling evapotranspiration using discrete wavelet transform and neural networks. Hydrol. Process., 23: 3545-3555.

Partal T, Kahya E, Cigizoglu HK (2008). Estimation of precipitation data using artificial neural networks and wavelet transform. *itüdergisi/d,* 7(3): 73-85.

Rajaee T (2010). Wavelet and neuro-fuzzy conjunction approach for suspended sediment prediction. Clean: Soil, Air, Water, 38(3): 275-286.

Rajaee T, Nourani V, Mohammad ZK, Kisi O. (2011). River Suspended Sediment Load Prediction: Application of ANN and Wavelet Conjunction Model. J. Hydrol. Eng., 16(8): 613-627.

Razavi S, Araghinejad S (2009). Reservoir inflow modeling using temporal neural networks with forgetting factor approach. Water Res. Manag., *23*: 39-55.

Wang HF, Huang WJ, Wang WS (2006b). Cuntan station of the Yangtze River annual runoff forecasting with set pair analysis method. J. Heilongjiang Hydraul. Eng. Coll., 33(4): 3-5.

Wang W, Ding J (2003). Wavelet network model and its application to the prediction of hydrology. Nat. Sci. 1: 67-71.

Wang W, Jin J, Li Y (2009). Prediction of Inflow at Three Gorges Dam in Yangtze River with Wavelet Network Model. Water Res. Manag., 23: 2791-2803.

Wang W, Van Gelder P, Vrijling JK, Ma J (2006a). Forecasting daily streamflow using hybrid ANN models. J. Hydrol., 324: 383-399.

Wu CL, Chau KW, Li YS (2008). Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. Water Res., 45(8): 1-23.