

Successive-station monthly streamflow prediction using neuro-wavelet technique

Ali Danandeh Mehr · Ercan Kahya ·
Farzaneh Bagheri · Ekin Deliktas

Received: 15 August 2013 / Accepted: 16 December 2013 / Published online: 28 December 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract This study investigates the effect of discrete wavelet transform data pre-processing method on neural network-based successive-station monthly streamflow prediction models. For this aim, using data from two successive gauging stations on Çoruh River, Turkey, we initially developed eight different single-step-ahead neural monthly streamflow prediction models. Typical three-layer feed-forward (FFNN) topology, trained with Levenberg-Marquardt (LM) algorithm, has been employed to develop the best structure of each model. Then, the input time series of each model were decomposed into subseries at different resolution modes using Daubechies (db4) wavelet function. At the next step, eight hybrid neuro-wavelet (NW) models were generated using the subseries of each model. Ultimately, root mean square error and Nash-Sutcliffe efficiency measures have been used to compare the performance of both FFNN and NW models. The results indicated that the successive-station prediction strategy using a pair of upstream-downstream records tends to decrease the lagged prediction effect of single-station runoff-runoff models. Higher performances of NW models compared to those of FFNN in all combinations demonstrated that the db4 wavelet transform function is a powerful tool to capture the non-stationary feature of the successive-station streamflow process. The comparative performance analysis among different combinations showed that the highest improvement for FFNN occurs when simultaneous lag-time is considered for both stations.

Keywords Feed-forward neural networks · streamflow prediction · successive-station forecasting · wavelet transform

1. Introduction

Conventional black-box regression models such as auto-regressive, auto-regressive moving average and the likes are linear models, which were widely used for hydrological forecasting (Abrahart and See 2000; Guang-Te and Singh 1994; Mondal and Wasimi 2006). Based on a basic necessity of being stationary, these models have limited ability to predict non-linear hydrologic time series (Nourani et al. 2011). It is the reason behind a great deal of research into the application of artificial intelligence (AI) techniques in hydrological forecasting. Artificial neural network (ANN) is one of the most popular AI techniques, which has been employed in various fields of hydrological forecasting and successful results have been reported (Altunkaynak 2007; Aytok et al. 2008; Besaw et al. 2010; Can et al. 2012; Dahamshe and Aksoy 2009; Kakahaji et al. 2013; Taormina et al. 2012). A comprehensive review concerning the application of neural networks in river forecasting was presented by Abrahart et al. (2012).

Both short-term and long-term streamflow predictions are required to plan, operate and optimize the activities associated with water resource system (Kisi 2010). Each of them has their own benefits and applications in operational hydrology. Monthly period prediction, as a long-term or transition from short- to long-term period, is useful for many water resource applications such as environmental protection, drought management, and optimal reservoir operation (Wang et al. 2009). However; Short-term forecasts, with lead times of hours or days, are necessary for flood warning systems and real-time reservoir operation (Chang et al. 2004). Based upon the aim of forecasting issue, different kinds of daily, monthly, and annual streamflow prediction models were developed individually by researchers

A. Danandeh Mehr (✉) · E. Kahya
Istanbul Technical University, Civil Engineering Department,
Istanbul, Turkey
e-mail: danandeh@itu.edu.tr

F. Bagheri
Department of Electrical Engineering, Azarbaijan Shahid Madani
University, Tabriz, Iran

E. Deliktas
Faculty of Science and Letters, Istanbul Technical University,
Istanbul, Turkey

without considering the efficiency of the proposed model in other time horizons and successful results were reported. For instance, genetic programming model (Makkeasorn et al. 2008), wavelet regression model (Küçük and Agiralioğlu 2006, Kisi 2010), stand-alone wavelet and cross-wavelet model (Adamowski 2008), and wavelet-ANN model (Kisi 2009, Krishna 2013) were suggested for daily streamflow forecasting. Moreover, an adaptive-network-based fuzzy inference system (Lin et al. 2005), a single station neuro-wavelet model (Cannas et al. 2006), and a wavelet predictor–corrector model (Zhou et al. 2008) were also suggested for monthly streamflow prediction. Coulibaly and Burn (2004) used wavelet analysis method with climatic patterns for both describing inter-annual features and quantifying the temporal variability of Canadian annual streamflow. Their results indicated that streamflow regionalization can be refined based on their wavelet spectra.

As known, streamflow series are highly non-stationary and quasi-periodic signals contaminated by various noises at different flow levels (Wu et al. 2009a). Unsatisfactory prediction results of ANN models have been widely reported for this kind of signals (Cannas et al. 2006; Nourani et al. 2011). ANN models were not also performed good enough in the case of runoff-runoff streamflow prediction due to the lagged prediction effect, which has been mentioned by some researchers (Chang et al. 2007; De Vos and Rientjes 2005; Muttil and Chau 2006; Wu et al. 2009a). In such cases, application of data pre-processing methods, such as moving average, singular spectrum analysis, and continuous/discrete wavelet transform (WT) integrated with ANN models have been suggested to improve the accuracy of the model (Adamowski and Sun 2010; Cannas et al. 2006; Kisi 2008; Krishna 2013; Niu and Sivakumar 2013; Nourani et al. 2011, Wu et al. 2009b). Application of the abovementioned data pre-processing methods along with some other feasible alternatives was discussed in detail by Chau and Wu (2011).

Our review in the relevant literature showed that hybrid single-station daily/monthly streamflow prediction models have received tremendous attention of researchers (e.g. Cannas et al. 2006; Adamowski 2008; Kisi 2009; Shiri and Kisi 2010; Krishna 2013). However, at the best of our knowledge, generalization of NW techniques in successive-station monthly streamflow prediction was not discussed. Thus, the main challenge of this study is to investigate the effect of wavelet-based data pre-processing method on the performance of ANN-based successive-station monthly streamflow predictions models in a perennial river. The successive-station prediction strategy using a pair of upstream-downstream records tends to decrease the lagged prediction effect of commonly proposed single-station runoff-runoff modes, while the wavelet component of the model provides a powerful tool to capture the non-stationary feature of the process. Following this, in the first phase of the current research, we put forward eight black-box ANN-based prediction models using streamflow records from two successive gauging stations on Çoruh River, Turkey. Typical three-layer feed-

forward neural network (FFNN) algorithm has been applied to develop the best structure for each model. In the second phase, we applied discrete wavelet decomposition procedure to generate our NW models. Finally, a comparative efficiency analysis between ad hoc FFNN and hybrid NW models has been done in terms of accuracy and simplicity.

Overview of FFNN

FFNN is one of the commonly used ANN techniques that typically involve three parts (American Society of Civil Engineers, ASCE Task committee 2000): including input layers with a number of input nodes, one or several hidden layers, and a number of output nodes. The number of hidden layers and nodes in each layer are two key design parameters. The input nodes do not perform any transformation upon the input data sets. They only send their initial weighted values to the nodes in the hidden layer. The hidden layer nodes typically receive the weighted inputs from the input layer or a previous hidden layer, perform their transformations on it and pass the output to the next adjacent layer which is generally another hidden layer or the output layer. The output layer consists of nodes that receive the hidden-layer outputs and send it to the modeller. Initial weight values are progressively corrected during a training process (at each iteration) that compares predicted outputs with corresponding observations and back-propagates any errors to determine the fitting weights which is required to minimize the errors.

A critical issue in ANN modeling is to avoid likely undertraining and overtraining problems. In the former situation, the network may not be possible to fully detect all attributes of a complex time series, while overtraining may reduce the model capacity for generalization properties. Application of cross-validation or selection of appropriate number of neurons in hidden layer using the trial-and-error procedure with a confined training iteration number (epoch), which were commonly suggested to prevent these problems (Cannas et al. 2006; Elshorbagy et al. 2010a; Wu et al. 2009b; Kisi 2008; Krishna 2013, Nourani et al. 2008 and 2009; Principe et al. 2000). The design issues, training mechanisms and application of FFNN in hydrological studies have been the subject of different studies (ASCE Task committee 2000, Abrahart et al. 2012) and the explicit expression for an output value of a three-layered FFNN networks is given by (Nourani et al. 2013). Therefore, to avoid duplication, we only introduced the main concepts of the FFNN here. The FFNN modelling attributes used in this study will be given in Section 5.1.

Wavelet transform

Wavelet transforms (WT) have recently begun to be used as a beneficial data pre-processing tool for hydrological time series.

The term wavelet means small wave. The wave refers to the condition that it is oscillatory and the smallness refers to the condition that the function is of finite length (Cannas et al. 2006). By a WT, a time series decomposes into multiple levels of details, sub-time series, which provide an interpretation of the original time series structure and history in both the time and frequency domains using a few coefficients (Nourani et al. 2013). Wavelets are waving-like mathematical functions with amplitude begin at zero, increases, and then decreases back to zero. Figure 1 represents three types of the most popular wavelet functions. As it clear from the figure, a wavelet tends to be irregular and asymmetric unlike the sine waves (Ozger, 2010).

WT appears to be more effective tool than the Fourier transform (FT) in studying non-stationary time series (Partal and Küçük 2006; Partal and Kisi 2007). The main advantage of WT is their ability to simultaneously obtain information on the location and frequency of a signal, while FT separates a time series in to sine waves of various frequencies, WT separates it into shifted and scaled waves using a predefined (or mother) wavelet (Ozger, 2010). WT Advantages compared to FT was discussed in detail by Sifuzzaman et al. (2009).

Continuous wavelet transforms (CWT)

In mathematics, an integral transform (Tf) is particular kind of mathematical linear operator, which has the following form:

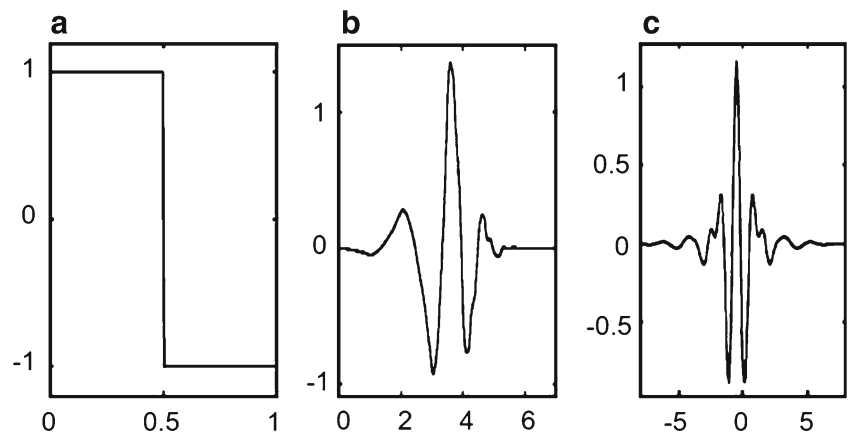
$$Tf(u) = \int_{t_1}^{t_2} K(t, u) f(t) dt \tag{1}$$

where $f(t)$ is an square-integrable function such as a continuous time series and K is a two variable, t and u , function called kernel.

According to Eq. 1, any integral transform is specified by a choice of the kernel function. If function K is chosen as wavelet function, then CWT is (Mallat, 1998):

$$T(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} \psi^* \left(\frac{t-b}{a} \right) f(t) dt \tag{2}$$

Fig. 1 Schematics of **a** Harr wavelet, **b** db4 wavelet, and **c** Meyer Hat wavelet



where $T(a, b)$ is the wavelet coefficients, $\psi(t)$ is a mother wavelet function, in time and frequency domain, and $*$ denotes operation of complex conjugate.

The parameter a , is scaling factor which acts as a dilation ($a > 1$) or a contraction ($a < 1$) coefficient of the mother wavelet. When the scaling factor is less than 1, the mother wavelet is more contracted which results more detailed sub-time series. In contrast, the scaling factor greater than 1 means the stretched mother wavelet which results less detail sub-time series. The parameter b is the translation (or position) value and also called time factor of the mother wavelet, which allows the study of the signal $f(t)$ locally around the time b (Küçük and agiralioglu, 2006). The main property of wavelets is localized in both time (b) and frequency (a) whereas the Fourier transform is only localized in frequency.

The CWT calculation produce N^2 coefficients from a data set of length N ; hence redundant information is locked up within the coefficients which may or may not be a desirable property (Addison et al., 2001; Krishna 2013). As an alternative, the discrete wavelet transform is usually preferred for practical applications in hydrological time series decomposition (Rajaei et al., 2011; Rathinasamy and Khosa, 2012).

Discrete wavelet transforms (DWT)

Hydrologic data usually are recorded in discrete time intervals. Hence, DWT is preferred to CWT for practical applications (Nourani et al. 2013). The wavelet function in its discrete form can be represented as:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{a} \right) \tag{3}$$

where a and b are scaling and position parameters, respectively.

In DWT, wavelet coefficients are commonly calculated at every dyadic step, i.e. the operation of WT is carried out at dyadic dilation ($a=2^m$) and integer translations ($b=2^m n$);

where m and n are integers that control the wavelet dilation and translation respectively. Therefore, the dyadic wavelet function can be obtained by Eq. (4) and DWT coefficients, $T_i(m, n)$, for a time series such as $f(t)$ can be defined as Eq. (5).

$$\psi_{m,n}(t) = 2^{-m/2} \psi(2^{-m}t - n) \quad (4)$$

$$T_i(m, n) = \sum_{i=0}^{N-1} \psi_i \cdot (t) f_i(t) = 2^{-m/2} \sum_{i=0}^{N-1} \psi(2^{-m}t - n) f_i(t) \quad (5)$$

Study area and data

Our study area is Çoruh River, the perennial river in eastern Black Sea region, Turkey (Fig. 2). The river springs from Mescit Mountains in Bayburt and reaches the Black Sea in Batumi City of Georgia after a course of 431 kms. Mean annual flow of the river before leaving Turkey's border is about $200 \text{ m}^3/\text{s}$. General characteristics of the river catchment were already described by Can et al. (2012).

The reliability of any hydrological prediction model will be crucially dependent on the quality of the underlying data. It is therefore important that sufficient attention is placed on gathering good quality data before they are used in prediction analysis. There are 11 successive gauging stations, with different recording time period, on the main course of the Çoruh River. Respect to the spatial and temporal consistency of the stations, we selected our study gauges (Fig. 2 stations 2322 and 2315) and observation period (1972–2000) from available data in such a way that provide both the longest and the most reliable records concurrently. Monthly streamflow time series of stations at 29-year period has been depicted in Fig. 3. Table 1 presents the statistical characteristics of the relevant data.

Çoruh is a shared river between Turkey and Georgia. Monthly streamflow prediction at the lower reach of the river, trans-boundary reach, will help both countries' water resources managers make suitable decisions in dry or wet spells or to resolve probable conflicts about sharing of river water.

Prior to the training of our ANN and NW models, the normalization was applied for the data. The main goal of data normalization is to scale the data within a certain range in

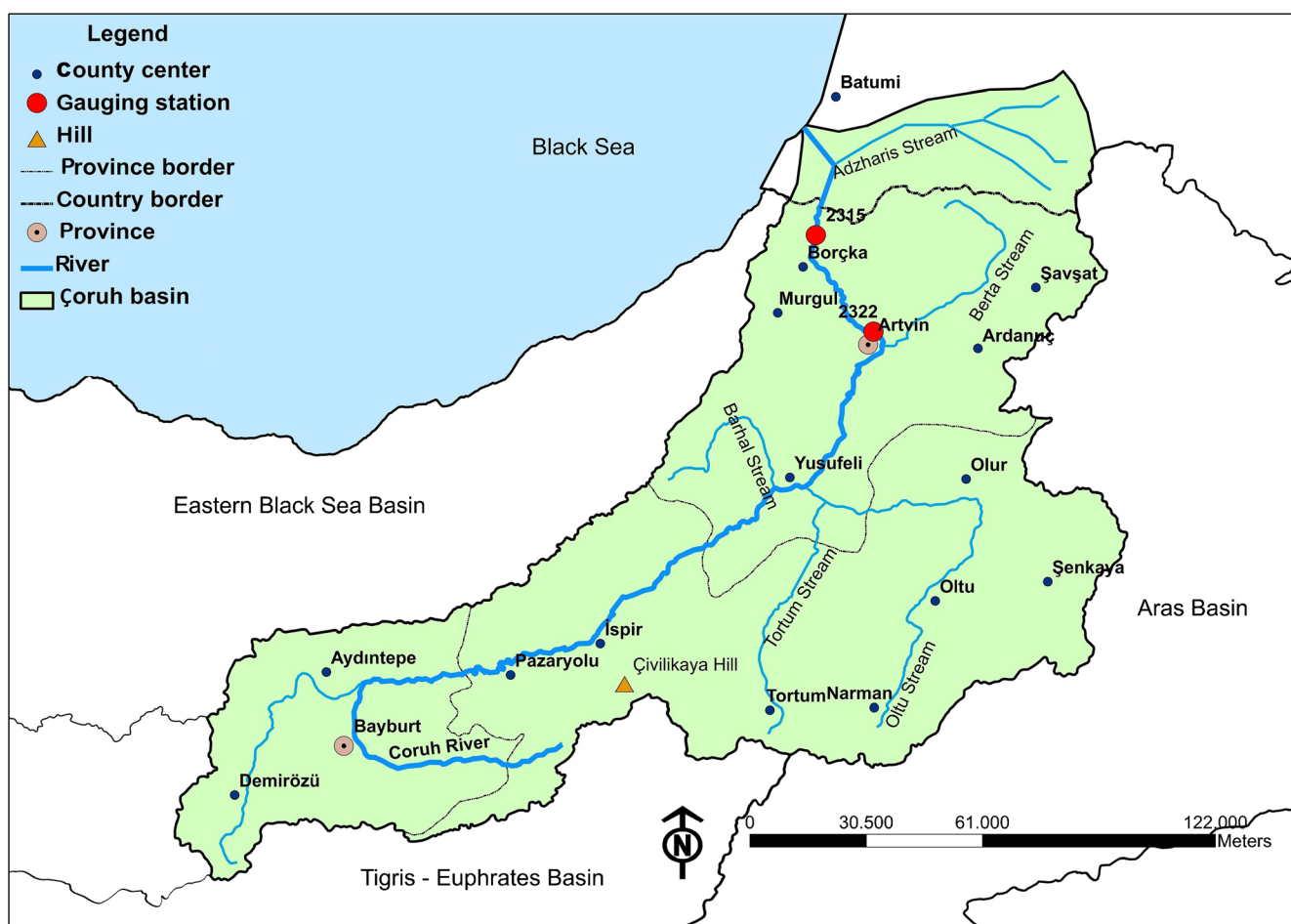
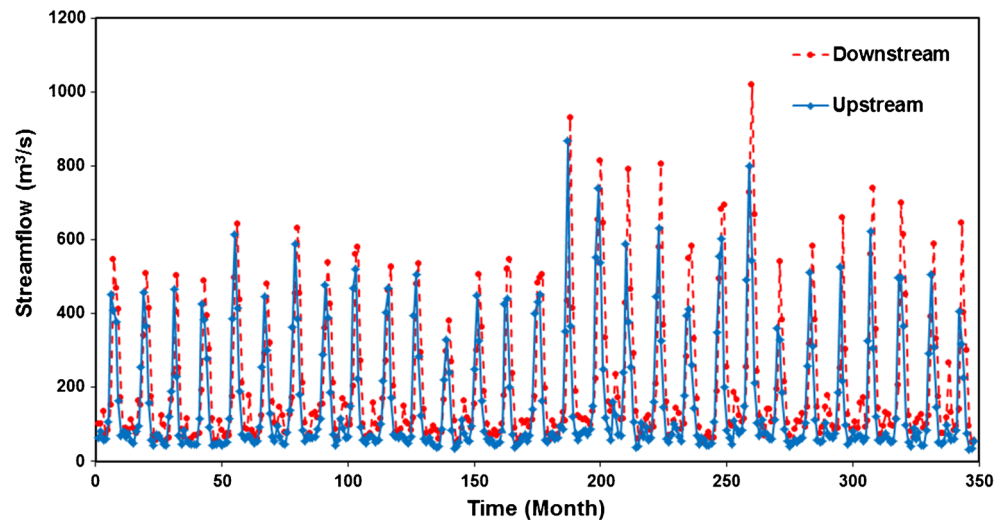


Fig. 2 Location of study area (Çoruh River Catchment)

Fig. 3 Monthly streamflow time series for the 29-year period



order to minimize bias within the neural networks. In this study, we applied the following formula for the min-max normalization method (Eq. (6)). This formula rescales target values to lie within a range of 0.1 to 1.0 and produces the arithmetic mean of target values larger than that of common min-max normalization method and at the same time being closer to the mean of activation factor (=0.5 in our networks). In commonly used min-max normalization method with a range of 0.0 to 1.0 (Priddy and Keller 2005), inefficient learning process of ANN is reported (Sajikumar and Thandaveswara 1999; Tahershamsi et al. 2012).

$$X_n = 0.1 + 0.8 \times \left(\frac{X}{X_{\max} - X_{\min}} \right) \tag{6}$$

where X_n =normalized value of X , X_{\max} =maximum value and X_{\min} =minimum value of each variable of the original data.

After training, the model that yields the best results in terms of Nash-Sutcliffe efficiency (NSE) and root mean squared error (RMSE) can be selected as the most efficient model.

NSE (Eq. (7)) is a normalized indicator of the model’s ability to predict about the 1:1 line between observed and predicted data. RMSE (Eq. (8)) measures the root average of the squares of errors. The error is the amount by which the value implied by the estimator differs from the target or quantity to be estimated.

$$NSE = 1 - \frac{\sum_{i=1}^n (X_i^{obs} - X_i^{pre})^2}{\sum_{i=1}^n (X_i^{obs} - X_{mean}^{obs})^2} \tag{7}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_i^{obs} - X_i^{pre})^2}{n}} \tag{8}$$

where X_i^{obs} = observed value of X , X_i^{pre} = predicted value X_{mean}^{obs} = mean value of observed data and n =number of observed data.

Table 1 The monthly statistical parameters of observed streamflow data

Statistical parameter	Station			
	Upstream (2322)		Downstream (2315)	
	Raw	Normalized	Raw	Normalized
Number of data (X)	348	348	348	348
X_{\max} (m ³ /s)	867.4	0.9321	1018.2	0.9404
X_{\min} (m ³ /s)	33.5	0.1321	48.9	0.1404
X_{mean} (m ³ /s)	158.6	0.2522	207.1	0.2709
Standard deviation (m ³ /s)	159.1	0.1526	184.4	0.1522
Coefficient of Skewness	1.7271	1.7271	1.7001	1.7001

Table 2 Statistical parameters of subsets

Parameters	Entire data	Training set	Validation set
Number of data (X)	348	244	104
X_{\max} (m ³ /s)	1,018	930	1,018
X_{\min} (m ³ /s)	33.5	36	48
X_{mean} (m ³ /s)	158.6	158	223
Standard deviation (m ³ /s)	159.1	156	204
Coefficient of Skewness	1.7271	1.71	1.71

Proposed models and results

At many of the hydrological modeling studies, modelers have developed different empirical relationships between input and output variables based upon single- or multi-step-ahead forecasting scenarios (Chang et al. 2007; Kisi 2004 and 2008; Krishna 2013; Nourani et al. 2009 and 2012). Sudheer et al. (2002) suggested a statistical procedure, based on cross-correlation, autocorrelation and partial autocorrelation properties of the time series for identifying the appropriate input vector for the model. However, Modarres (2009) showed that global statistics are insufficient indicators for the best ANN. In this study, a number of successive-station combinations (Models (1) to (8)) were trained and compared with each other to identify the appropriate input vector. Such an identification methodology has previously applied by Kisi (2004 and 2007) for single-statin daily and monthly streamflow prediction. The structure of the proposed models expressed as follows.

- (1) Model (1) $D_t = f(D_{t-1}, \epsilon_t)$
- (2) Model (2) $D_t = f(U_t, \epsilon_t)$
- (3) Model (3) $D_t = f(D_{t-1}, U_t, \epsilon_t)$
- (4) Model (4) $D_t = f(U_{t-1}, D_{t-1}, \epsilon_t)$
- (5) Model (5) $D_t = f(U_{t-1}, U_t, D_{t-1}, \epsilon_t)$
- (6) Model (6) $D_t = f(U_{t-1}, U_t, D_{t-2}, D_{t-1}, \epsilon_t)$
- (7) Model (7) $D_t = f(U_{t-2}, U_{t-1}, D_{t-2}, D_{t-1}, \epsilon_t)$
- (8) Model (8) $D_t = f(U_{t-2}, U_{t-1}, U_t, D_{t-2}, D_{t-1}, \epsilon_t)$

where D_t and U_t represent downstream and upstream monthly streamflow respectively and ϵ_t is an uncertainty term (to be minimized).

FFNN modelling results

Our data record is composed of 348 monthly streamflow observations at each station. Due to the temporal consequence of hydrological process, such as streamflow between gauging stations, it is recommended to use the first part of observed time series for training and the rest for the verification (Kisi 2007; Krishna 2013; Nourani et al. 2012). In this study, the first 70 and the rest 30 percentage of 348 observations were selected for training and validation of the networks, respectively. Therefore, the entire data set were divided into two subsets. The statistical parameters of each subset were presented in Table 2.

At the first stage of streamflow prediction, several three-layer FFNNs with the sigmoid transfer function in hidden layer and linear transfer function in output layer has been developed as a nonlinear modelling structure. It has been proved that three-layer FFNNs are satisfied for the hydrological forecasting (ASCE Task committee 2000; Chau et al. 2005; Nourani et al. 2008; Rezaeianzadeh 2013a). The three-layer FFNN network, which is well known as universal approximator (Hornik et al. 1989), is probably the most popular ANN in the case of nonlinear mapping (Tahershamsi et al. 2012, Krishna 2013). The satisfactory application of the sigmoid and linear transfer functions in hidden and output layers, respectively for streamflow prediction was extensively reported (e.g. Chang et al. 2007, Krishna 2013; Rezaeianzadeh et al. 2013b). Following the structure definition, we trained our models using the Levenberg-Marquardt (LM) algorithm with a useful toolbox available in the MATLAB® software. The LM is one of the Newtonian optimization techniques and more powerful than conventional gradient descent techniques (Kisi, 2007), which widely used in FFNNs (Haykin, 1999). The successful implementation of LM algorithm to train

Table 3 Performance comparison of FFNN and NW techniques at validation period

Model	Prediction scenario	FFNN			NW		
		NHL ^a	RMSE ^b	NSE	NHL ^a	RMSE ^b	NSE
1	$D_t = f(D_{t-1})$	3	0.131	0.355	2	0.130	0.364
2	$D_t = f(U_t)$	3	0.130	0.363	4	0.117	0.482
3	$D_t = f(D_{t-1}, U_t)$	4	0.079	0.766	4	0.078	0.772
4	$D_t = f(U_{t-1}, D_{t-1})$	5	0.039	0.943	6	0.032	0.962
5	$D_t = f(U_{t-1}, U_t, D_{t-1})$	4	0.037	0.948	5	0.034	0.955
6	$D_t = f(U_{t-1}, U_t, D_{t-2}, D_{t-1})$	6	0.046	0.921	6	0.044	0.925
7	$D_t = f(U_{t-2}, U_{t-1}, D_{t-2}, D_{t-1})$	7	0.043	0.918	9	0.040	0.922
8	$D_t = f(U_{t-2}, U_{t-1}, U_t, D_{t-2}, D_{t-1})$	9	0.033	0.922	9	0.031	0.926

^a Nodes in hidden layer (NHL), the result has been presented for the best structure

^b The RMSE is dimensionless

ANN-based hourly, daily, and monthly streamflow prediction models is frequently reported (e.g. De Vos and Rientjes 2005; Chau et al. 2005; Cannas et al. 2006; Kisi 2010)

As broadly suggested in the literature (e.g. Cheng et al. 2005; Chang et al. 2007; Taormina et al. 2012), in this

research, a fixed number of input nodes, equals to the number of input time series at each model, a single hidden layer with dynamic nodes varying from 1 to 10 with a step size of 1 in each trial, and one output node (the desired output) was adopted to train the proposed models. Determination of

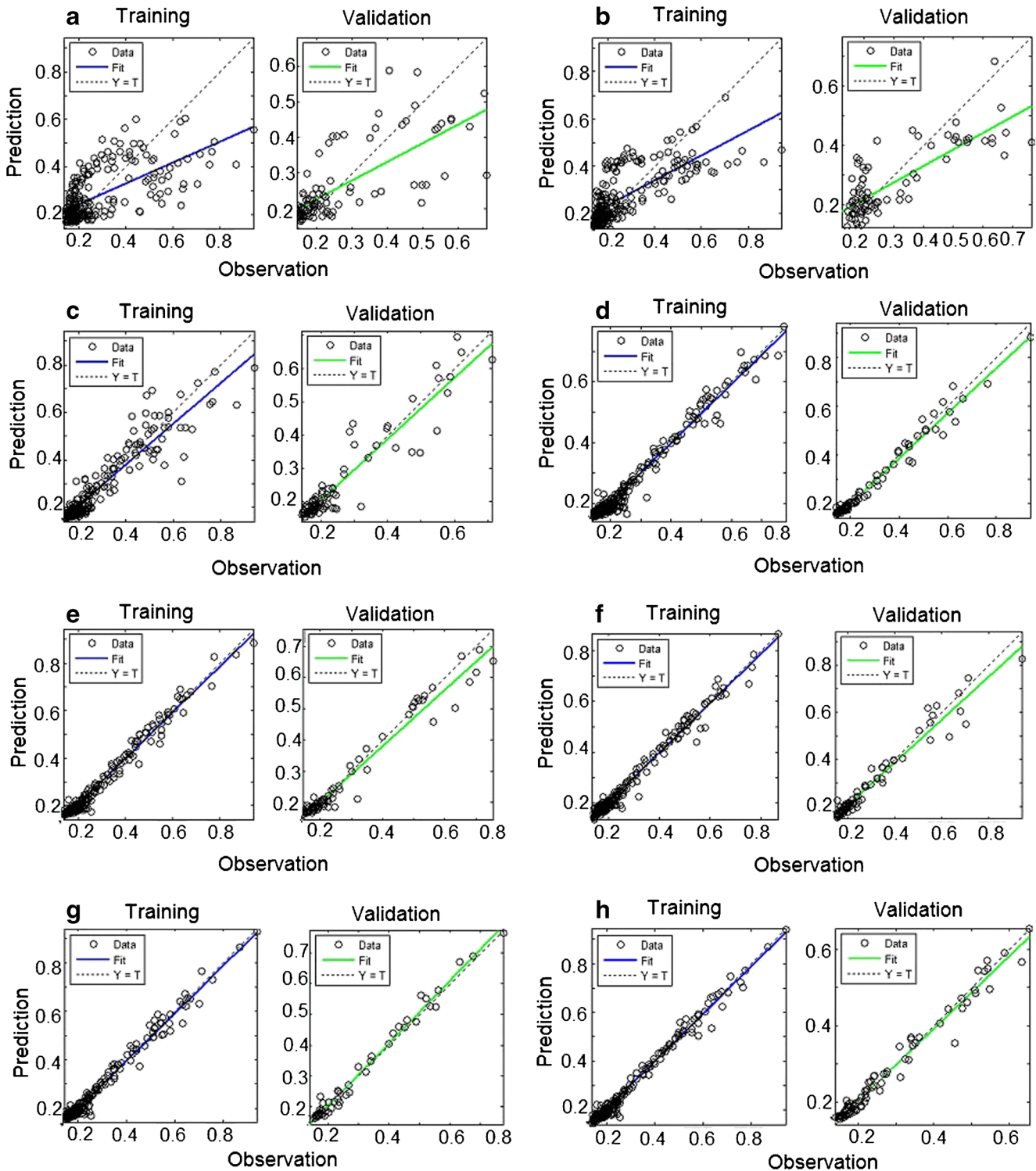


Fig. 4 Comparison of dimensionless predicted streamflow for the Calibrations and validations by different models; **a** Model 1, **b** Model 2, **c** Model 3, **d** Model 4, **e** Model 5, **f** Model 6, **g** Model 7, **h** Model 8

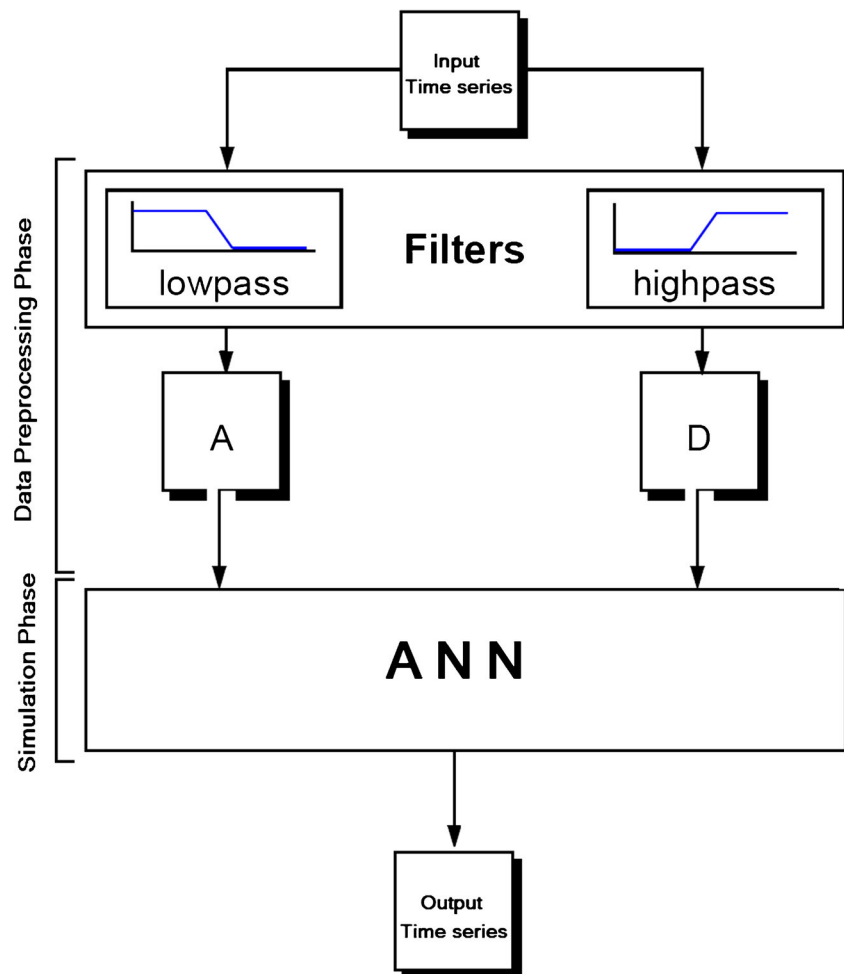
optimum number of neurons in hidden layer is an important aspect of an efficient network. As it mentioned previously, in order to check any under- or over-fitting problem during the training process, a commonly suggested trial-and-error approach (Kisi 2008) is utilized to select the optimum number of neurons in the hidden layer. If the number of nodes in the hidden layer is too small, the network may not have sufficient degrees of freedom to learn the process correctly. On the contrary, if it is too high, the training will take a long time and the network may sometimes overfit (Kisi 2004). Cannas et al. (2006) found that the optimal FFNN models are constructed with a few number of hidden neurons. As it mentioned previously, in present study, the upper hidden neurons threshold was adopted to be 10. The mean square error value at validation step in each trial is used as a criterion for selecting the optimal number of hidden neurons. No significant improvement in model performance was found when the number of hidden neurons was increased from the limit, which is similar to the experiences of other researchers (Cheng et al. 2005; Cannas et al. 2006; Wu et al. 2009b). Results of the best FFNN structure for each model and their performance levels have been tabulated in Table 3. Scatterplots of the best

predictions compared to their corresponding observed values were illustrated in Fig. 4.

NW modelling results

The proposed NW models are hybrid-ANN models, which mean the pre-processed data via wavelet transform, are entered to the ANN model in order to improve the accuracy of the formerly structured FFNN models. The schematic structure of the proposed NW model is illustrated in Fig. 5. The structure includes two phases. In the pre-processing phase, the original monthly streamflow time series including D_t , U_t and their previous observations are decomposed into sub-series of approximations (A) and details (D) through the high-pass and low-pass filter coefficients of a chosen mother wavelet. In this manner each sub-series may represent a special level of the temporal characteristics of the original input time series. Appropriate mother wavelet may be selected in this phase according to shape pattern similarity between the mother wavelet and the investigated time series (Onderka et al. 2013). The Brute-force search method can also be

Fig. 5 Schematic view of proposed NW model



applied as an alternative to choose an appropriate mother wavelet (Nourani et al. 2013).

Our time series is characterized by a very irregular signal shape, fast decaying and oscillating, thus an irregular mother wavelet, the Daubechies wavelet with four vanishing moments (db4, Fig. 1b) was selected for use (Daubechies, 1990). Nourani et al. (2011) compared the effect of db4 with three different irregular mother wavelets, comprising Haar, Sym3 and Coif1 wavelets, on daily and monthly streamflow prediction and showed that the db4 provides the best performance. The effective application of the db4 for decomposition of non-stationary monthly streamflow time series has also been reported by Cannas et al. (2006). Therefore, in the present study, we applied dyadic DWT with the db4 mother wavelet at different resolution modes (i.e. 2, 4, 8, 16, and 32-month mode) to decompose the original observed streamflow time series. Each resolution demonstrates specific time feature of the investigated time series. In decomposition level 5 (i.e. dyadic dilation $a=32$), there is five details (or resolution modes) and one approximation sub-signal. We wrote a special code to use the MATLAB® software for NW simulation in this research. For instance, the approximation (a_5) and detail sub-series (d_1 through d_5) for the streamflow time series of downstream station has been illustrated in Fig. 6.

In the second phase, simulation phase, at first, the three-layer FFNN models are built afresh such that the sub-series, a_5 and d_1 through d_5 in our experiment, of

the original time series are the input neurons to the FFNN. Then, training and validation process are performed using the input sub-series to determine output time series. The number of input layer neurons was determined according to the used decomposition level and the output layer neuron was a fixed one. Results for the best NW structure and their performance level compared to those of FFNN are presented in Table 3.

Discussion

The efficiency results of the best developed FFNN and NW models for all proposed combinations are compared in Table 3 using NSE and RMSE indices at the validation period. It can be implied from the table that Model (1) resulted in the lowest achieved performance level. Suggested decomposition process did not also provide any noteworthy improvement in the results of corresponding FFNN. It may be because of introducing insufficient lag time (input neurons) to the models. It indicated that this single-step-ahead monthly streamflow prediction scenario could not provide a suitable prediction model for the river.

Based upon aforementioned successive-station prediction strategy, in the second scenario upstream data records was considered as input neurons instead of increasing in the lag-time of downstream station records. Model (2), provides a

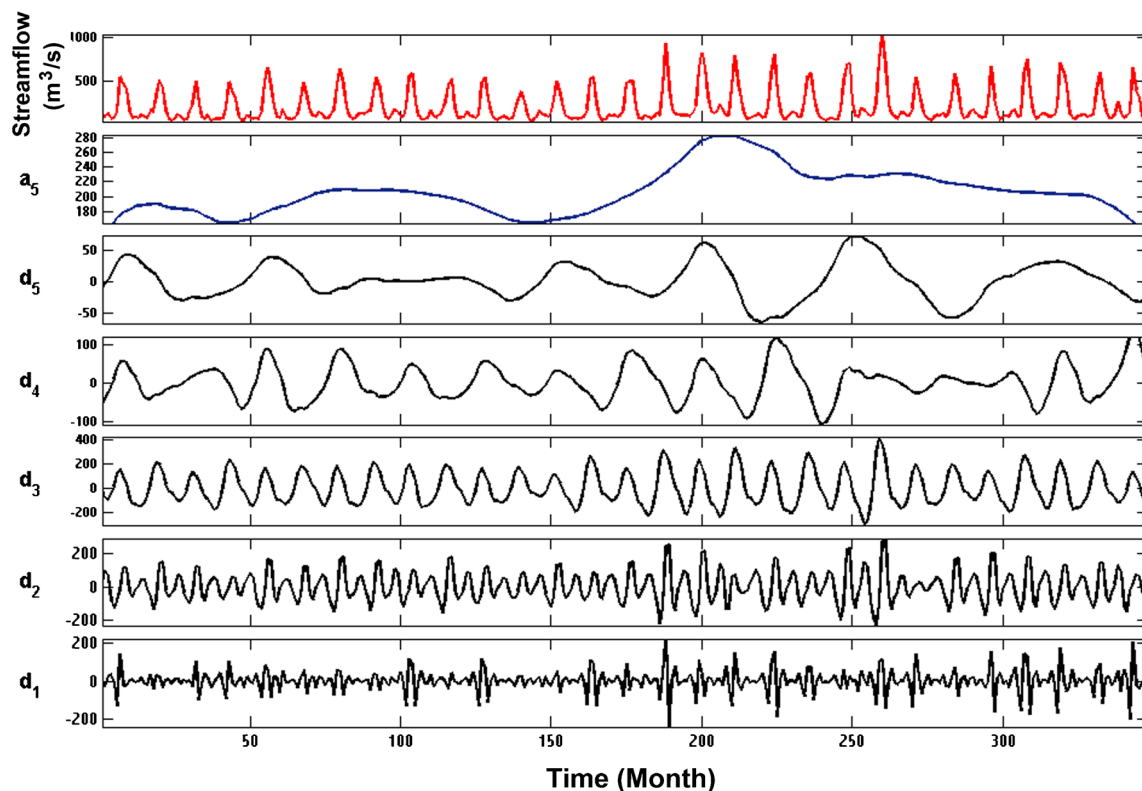


Fig. 6 Approximation and detail sub-series of streamflow at station 2315

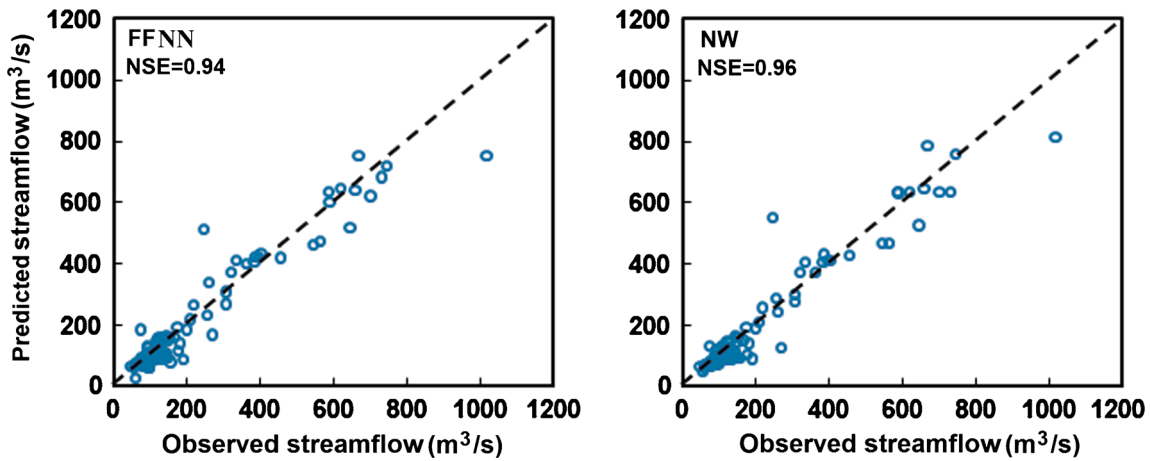


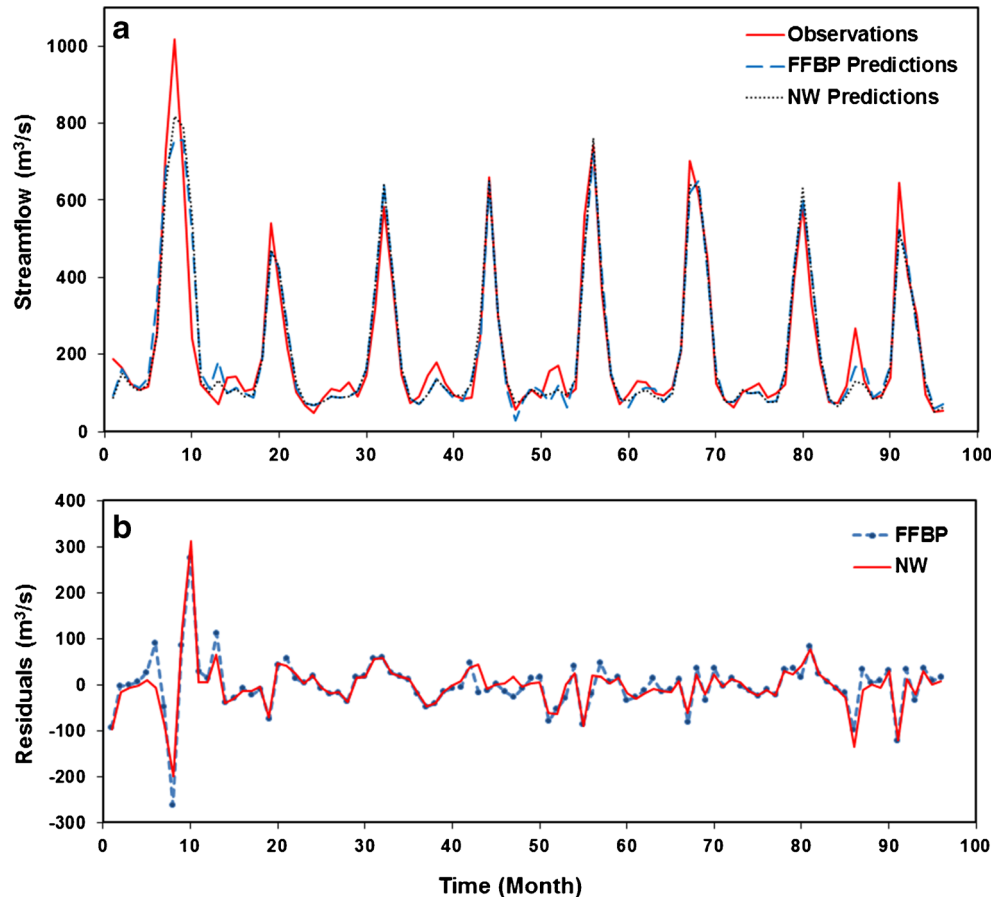
Fig. 7 Scatter plots of predicted data at validation period (station 2315)

downstream flow prediction scenario using only simultaneous upstream station records. Although it makes a considerable improvement in the performance levels of corresponding NW-based model compared with those of Model (1), it does not provide a suitable streamflow prediction model yet (NSE=0.482). The reason is obviously related to the fact that the mean monthly flow in the study reach has a spatially increasing regime which is also distinguishable in Fig. 3. Comparing with all other scenarios, Model (2) shows that the highest

effect of wavelet decomposition on the accuracy of FFNN models occurs in this case. It is due to the fact that temporal characteristics of the downstream time series are thought to be highly simultaneously correlated with time features of upstream sub-series. The results of these first two models also indicated the necessity of additional forecasting lead time and/or efficient input variables.

The first combination of upstream and downstream flow, Model (3), resulted in significant improvement in the

Fig. 8 Predicted and observed monthly streamflow (a) and prediction residuals (b) at station 2315 for validation period



performance levels. It implies that that existing sub-basins between the stations have considerable physical effects (i.e., increasing drainage area) on the occurrence of flow at downstream station. This is the reason why we intentionally keep downstream flow at time ($t-1$), D_{t-1} , in constructing of the other combinations (Models (4) to (8)). Similar to Model (1), Although NW slightly outperformed FFNN; wavelet decomposition technique did not provide considerable improvement for the ad hoc FFNN in Model (3). It may be due to the presence of inefficient decomposed sub-series within the input nodes of NW. Therefore, it is suggested that suitable filtering strategies would be implemented on the choice of effective decomposed sub-series in future studies. For instance, correlation between decomposed sub-series with original input time series can be used to distinguish the effective discrete wavelet components (Krishna 2013).

Models (4) to (8) represent five effective prediction successive-station combinations with high performance levels, which all perform similar to each other. NW provided comparatively better outcomes than FFNN in all cases. Due to remarkable coherence between the observed and predicted values in these models, there is a very strict competition among them to be chosen as an optimum streamflow prediction model of the river. From Table 3, it is observed that Models (4) and (5) are the most appropriate combinations for streamflow prediction by FFNN and NW. Both of them use downstream and upstream flows at time ($t-1$) (D_{t-1} and U_{t-1}); however, the former also contains an extra input variable namely upstream flow at time (t) (U_t). Adding one more variable into the input layer of Model (4) intentionally, resulted in imperceptible improvement in FFNN and reduction in NW efficiency. Such a diminishing accuracy is also observed in NW results of Models (6) to (8). This drawback of NW might be due to increasing number of input neurons. Adding one more input variable in our FFNN models generates six more input sub-signals within the corresponding NW structure that significantly increases the complexity of the model. It is important to note that inevitable errors of each input series (sub-series) magnify the total error of the NW, and thus the model efficiency diminishes. Therefore, we suggest the NW modellers to be very careful when selecting the input variables whether or not it is inefficient or efficient multivariable NW model.

The efficiency values of NW-based Model (4) (RMSE=0.032 and NSE=0.962) demonstrate that successive-station strategy using only 1-month lag provides the most accurate monthly streamflow prediction model on Çoruh River. It provides more than 60 percentage improvement in NSE value of single-station-1-month-ahead model (i.e. Model (1)). It is why in single-station runoff-runoff models application of at least 3 to 4 lag-times in order to achieve the best modelling structure were commonly recommended (Wu et al. 2009a, b; Kisi 2010; Shiri and Kisi 2010).

Considering all the aforementioned results and the concept of simplicity and applicability as a main issue in modelling, Model (4) has been selected as the best monthly streamflow prediction at station 2315 on Çoruh River. In order to present detailed comparison between the NW and FFNN models, the scatterplots and the plots of the observed and predicted time series of 1-month-ahead forecast for Çoruh River (Model (4)) at validation period were illustrated in Figs. 7 and 8, respectively. The residuals of the models have also depicted in the Fig. 8b.

These figures show the high ability of both proposed FFNN and NW models to predict the low and medium monthly streamflow ($150 < D_t < 400 \text{ m}^3/\text{s}$). NW is more capable of capturing local extremes (i.e. minima and maxima), global maximums (or annual peaks) than FFNN and generated $5 \text{ m}^3/\text{s}$ mean absolute error less than those of FFNN which warrant its superiority to FFNN in overall sense. Furthermore, Fig. 8 shows that some of the FFNN model's timing of the peaks is lagged. This lagged prediction effect is the result of using antecedent observed values as FFNN inputs without considering any data pre-processing approach, which is consistent with the findings of other researchers (De Vos and Rientjes 2005; Wu et al. 2009a, b). It reveals the fact that the chosen mother wavelet, db4, has performed well as expected.

Conclusion

In this study, we attempted to investigate how and how much a wavelet-based data pre-processing method can improve the performance of the common tree-layer FFNN models in successive-station monthly streamflow prediction. For this aim, we firstly developed eight different FFNN prediction models using historical observations from two successive gauging stations on Çoruh River, Turkey. Then, we tried to optimize our models via a discrete wavelet transform method. Wavelet-transformed data in conjunction with ANN structures generated our proposed hybrid NW models that were capable of capturing useful information about history of the observations on various time resolution levels with the use of only a few coefficients.

The obtained results indicated a promising role of NW in successive-station monthly streamflow prediction, which led all FFNN models to more accurate predictions. Our proposed NW model provides flexible choice of mother wavelet selection to decompose the input time series. The results showed that Daubechies mother wavelet with four vanishing moments provides acceptable outcomes for monthly streamflow decomposition. Therefore, its application to decompose similar monthly streamflow signals is suggested. The comparative analysis between different input combinations pointed out that only 1 month-lagged records of both downstream and

upstream stations are enough to establish the best prediction model for the study catchment.

Owing to our study limited to the implementation of a distinct catchment data, applying this approach to pairs of stations from different catchments is suggested to strengthen these conclusions. We evaluated a single ANN algorithm along with a single db4 mother wavelet in order to propose our NW model. Other ANN techniques, such as generalized regression and/or radial basis function algorithms using different mother wavelets (e.g., Harr, Mayer, and others) can also be investigated in order to improve the efficiency of NW-based successive-station streamflow prediction models.

Acknowledgments We wish to express our sincere gratitude to the anonymous reviewers whose suggestions and comments have greatly helped us to improve the quality of the manuscript. We also thank Assoc. Prof. A. Atunkaynak, Istanbul Technical University, for his valuable guidance in improving the work.

References

- Abrahart RJ, See L (2000) Comparing neural network (NN) and autoregressive moving average (ARMA) techniques for the provision of continuous river flow forecasts in two contrasting catchments. *Hydrol Process* 14:2157–2172
- Abrahart RJ, Ancil F, Coulibaly P, Dawson CW, Mount NJ, See LM, Shamseldin AY, Solomatine DP, Toth E, Wilby RL (2012) Two decades of anarchy? Emerging themes and outstanding challenges for neural network modelling of surface hydrology. *Prog in Phys Geogr* 36(4):480–513
- 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* 353:247–266
- Adamowski J, Sun K (2010) Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *J Hydrol* 390:85–91
- Addison PS, Murrery KB, Watson JN (2001) Wavelet transform analysis of open channel wake flows. *J Eng Mech* 127(1):58–70
- Altunkaynak A (2007) Forecasting surface water level fluctuations of Lake Van by artificial neural networks. *Water Resour Manage* 21:399–408
- ASCE Task committee (2000) Artificial neural networks in hydrology: hydrologic applications. *J Hydrol Eng* 5(2):124–137
- Aytek A, Guven A, Yuce MI, Aksoy H (2008) An explicit neural network formulation for evapotranspiration. *Hydrol Sci J* 53(4):893–904
- Besaw LE, Rizzo DM, Bierman PR, Hackett WR (2010) Advances in ungauged streamflow prediction using artificial neural networks. *J Hydrol* 386:27–37
- Can İ, Tosunogulu F, Kahya E (2012) Daily streamflow modelling using autoregressive moving average and artificial neural networks models: case study of Çoruh basin, Turkey. *Water and Environ J* 26:567–576
- Cannas B, Fanni A, See L, Sias G (2006) Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. *Phys Chem Earth, Parts A/B/C* 31(18):1164–1171
- Chang LC, Chang FJ, Chiang YM (2004) A two-step ahead recurrent neural network for streamflow forecasting. *Hydrol Processes* 18:81–92
- Chang FJ, Chiang YM, Chang LC (2007) Multi-step-ahead neural networks for flood forecasting. *Hydrol Sci J* 52(1):114–130
- Chau KW, Wu CL (2011) Hydrological predictions using data-driven models coupled with data pre-processing Lap Lambert academic publishing. Saarbrücken, Germany
- Chau KW, Wu CL, Li YS (2005) Comparison of several flood forecasting models in Yangtze River. *J Hydrol Eng* 10(6):485–491
- Cheng CT, Chau KW, Sun YG, Lin JY (2005) Long-term prediction of discharges in Manwan Reservoir using artificial neural network models. *Lect Notes Comput Sc* 3498:1040–1045
- Coulibaly P, Burn HD (2004) Wavelet analysis of variability in annual Canadian streamflows. *Water Resour Res* 40, W03105
- Dahamshe A, Aksoy H (2009) Artificial neural network models for forecasting intermittent monthly precipitation in arid regions. *Meteorol Appl* 16:325–337
- Daubechies I (1990) The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans Inf Theory* 36(5):961–1005
- De Vos NJ, Rientjes THM (2005) Constraints of artificial neural networks for rainfall–runoff modeling: trade-offs in hydrological state representation and model evaluation. *Hydrol Earth Syst Sci* 9:111–126
- Elshorbagy A, Corzo G, Srinivasulu S, Solomatine DP (2010) Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: Concepts and methodology. *Hydrol Earth Syst Sci* 14:1931–1941
- Guang-Te W, Singh VP (1994) An autocorrelation function method for estimation of parameters of autoregressive models. *Water Resour Manage* 8:33–56
- Haykin S (1999) Neural networks, a comprehensive foundation. Prentice Hall, Upper Saddle River, NJ
- Hornik K, Stinchcombe M, White M (1989) Multi-layer feed forward networks are universal approximators. *Neural Networks* 2(5):359–366
- Kakahaji H, Dehghan Banadaki H, Kakahaji A, Kakahaji A (2013) A Prediction of Urmia Lake water-level fluctuations by using analytical, linear statistic and intelligent methods. *Water Resour Manage* 27:4469–4492
- Kisi O (2004) River flow modeling using artificial neural network. *J Hydrol Eng* 9(1):60–63
- Kisi O (2007) Streamflow forecasting using different artificial neural network algorithms. *J Hydrol Eng* 12(5):532–539
- Kisi O (2008) Stream flow forecasting using neuro-wavelet technique. *Hydrol Process* 22:4142–4152
- Kisi O (2009) Neural networks and wavelet conjunction model for intermittent streamflow forecasting. *J Hydrol Eng* 14(8):773–782
- Kisi O (2010) Wavelet regression model for short-term streamflow forecasting. *J Hydrol* 389:344–353
- Krishna B (2013) Comparison of wavelet based ANN and Regression models for Reservoir Inflow Forecasting. *J Hydrol Eng*. doi:10.1061/(ASCE)HE.1943-5584.0000892
- Küçük M, Agiralioglu N (2006) Regression technique for stream flow prediction. *J Appl Stat* 33(9):943–960
- Lin JY, Cheng CT, Sun YG, Chau KW (2005) Long-term prediction of discharges in Manwan Hydropower using adaptive-network-based fuzzy inference systems models. *Lect Notes Comput Sc* 3612:1152–1161
- Makkeasorn A, Chang NB, Zhou X (2008) Short-term streamflow forecasting with global climate change implications – A comparative study between genetic programming and neural network models. *J Hydrol* 352:336–354
- Mallat S (1998) A Wavelet tour of signal processing, 2nd edn. Academic Press, San Diego, CA
- Modares R (2009) Multi-criteria validation of artificial neural network rainfall-runoff modeling. *Hydrol Earth Syst Sci* 13:411–421
- Mondal MS, Wasimi SA (2006) Generating and forecasting monthly flows of the Ganjes River with PAR Model. *J Hydrol* 232:41–56
- Muttill N, Chau KW (2006) Neural network and genetic programming for modelling coastal algal blooms. *Int J Environ Pollut* 28(3–4):223–238

- Niu J, Sivakumar B (2013) Scale-dependent synthetic streamflow generation using a continuous wavelet transform. *J Hydro* 496:71–78
- Nourani V, Mogaddam AA, Nadiro AO (2008) An ANN based model for spatiotemporal groundwater level forecasting. *Hydrol Process* 22: 5054–5066
- Nourani V, Alami MT, Aminfar MH (2009) A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. *Eng Appl Artif Intell* 22(3):466–472
- Nourani V, Kisi Ö, Komasi M (2011) Two hybrid Artificial Intelligence approaches for modelling rainfall–runoff process. *J Hydrol* 402:41–59
- Nourani V, Komasi M, Alami MT (2012) Hybrid Wavelet–Genetic Programming Approach to Optimize ANN Modelling of Rainfall–Runoff Process. *J Hydrol Eng* 17(6):724–741
- Nourani V, Hosseini Baghanam A, Adamowski J, 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
- Onderka M, Banzhaf S, Scheytt T, Krein A (2013) Seepage velocities derived from thermal records using wavelet analysis. *J Hydrol* 479:64–74
- Ozger M (2010) Significant wave height forecasting using wavelet fuzzy logic approach. *Ocean Eng* 37:1443–1451
- Partal T, Kisi O (2007) Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. *J Hydrol* 342:199–212
- Partal T, Küçük M (2006) Long-term trend analysis using discrete wavelet components of annual precipitations measurements in Marmara region (Turkey). *Phys Chem Earth* 31:1189–1200
- Priddy KL, Keller PB (2005) *Artificial neural networks, an introduction*. SPIE Press, Washington
- Principe JC, Euliano NR, Curt Lefebvre W (2000) *Neural and Adaptive Systems*. Wiley & Sons.
- Rajae 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
- Rathinasamy M, Khosa R (2012) Multi-scale nonlinear model for monthly streamflow forecasting: a wavelet-based approach. *J Hydroinform* 14(2):424–442
- Rezaeianzadeh M, Stein A, Tabari H, Abghari H, Jalalkamali N, Hosseinipour EZ, Singh VP (2013a) Assessment of a conceptual hydrological model and artificial neural networks for daily outflows forecasting. *Int J Environ Sci Technol* 10(6):1181–1192
- Rezaeianzadeh M, Tabari H, Abghari H (2013b) Prediction of monthly discharge volume by different artificial neural network algorithms in semi-arid regions. *Arab J Geosci* 6:2529–2537
- Sajikumar N, Thandaveswara BS (1999) A non-linear rainfall-runoff model using an artificial network. *J Hydrol* 216(4):32–55
- Shiri J, Kisi O (2010) Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model. *J Hydrol* 394: 486–493
- Sifuzzaman M, Islam MR, Ali MZ (2009) Application of wavelet transform and its advantages compared to Fourier transform. *J Physical Sciences* 13:121–134, www.vidyasagar.ac.in/journal
- Sudheer KP, Gosain AK, Ramasastri KS (2002) A data driven algorithm for constructing artificial neural network rainfallrunoff models. *Hydrol Process* 16(6):1325–1330
- Tahershamsi A, Majdzade Tabatabai MR, Shirkhani R (2012) An evaluation model of artificial neural network to predict stable width in gravel bed rivers. *Int J Environ Sci Technol* 9:333–342
- Taormina R, Chau KW, Sethi R (2012) Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. *Eng Appl Artif Intel* 25(8):1670–1676
- Wang WC, Chau KW, Cheng CT, Qiu L (2009) A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *J Hydrol* 374:294–306
- Wu CL, Chau KW, Li YS (2009a) Methods to improve neural network performance in daily flows prediction. *J Hydrol* 372: 80–93
- Wu CL, Chau KW, Li YS (2009b) Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. *Water Resour Res* 45, W08432
- Zhou HC, Peng Y, Liang G (2008) The research of monthly discharge predictor–corrector model based on wavelet decomposition. *Water Resour Manage* 22(2):217–227