Neural Comput & Applic (2017) 28 (Suppl 1):S15–S28
DOI 10.1007/s00521-016-2293-9

ORIGINAL ARTICLE

Predicting river daily flow using wavelet-artificial neural networks based on regression analyses in comparison with artificial neural networks and support vector machine models

Maryam Shafaei[1] · **Ozgur Kisi**[2]

Received: 4 February 2015 / Accepted: 30 March 2016 / Published online: 18 April 2016

Abstract This study investigates the ability of wavelet-artificial neural networks (WANN) for the prediction of short-term daily river flow. The WANN model is improved by conjunction of two methods, discrete wavelet transform and artificial neural networks (ANN) based on regression analyses, respectively. The proposed WANN models are applied to the daily flow data of Vanyar station, on the Ajichai River in the northwest region of Iran, and compared with the ANN and support vector machine (SVM) techniques. Mean square error (MSE), mean absolute error (MAE) and correlation coefficient (R) statistics are used for evaluating precision of the WANN, ANN and SVM models. Comparison results demonstrate that the WANN model performs better than the ANN and SVM models in short-term (1-, 2- and 3-day ahead) daily river flow prediction.

Keywords Artificial neural networks · Discrete wavelet transform · Support vector machine · River flow · Forecasting

List of symbols

$x(t)$	Signal
t	Integer time stages
j	Integers that control the scale
k	Integers that control the time
$W_{j,k}$	Wavelet coefficient
s	Scale parameter
τ	Time parameter
C_{p}	Mallows' coefficient
i	ith iteration
η	Learning rate
θ	Momentum value
E	The sum of squared errors between observed and predicted data
ξ_k	Slack variables
ε	Insensitivity loss function
C	Positive trade-off factor
K	Kernel function
n	Number of support vectors
σ	SVM model parameter
X_{mean}	Overall mean
S_x	Standard deviation
C_{sx}	Skewness
X_{min}	Minimum value
X_{max}	Maximum value
Q_t	Observed flow
Q_{min}	Minimum flow
Q_{max}	Maximum flow
Z_i	Normalized flow

✉ Maryam Shafaei
maryamshafaei65@yahoo.com

Ozgur Kisi
okisi@basari.edu.tr

[1] Water Engineering Department, Tabriz University, Tabriz, Iran

[2] Civil Engineering Department, Faculty of Architecture and Engineering, Canik Basari University, Samsun, Turkey

1 Introduction

Precision in flow forecasting is necessary for both water quantity and quality management in a catchment area. Two possible methods are commonly used for flow forecasting. The first method is the process modeling that includes the study of rainfall–runoff processes and modeling physical principles [16]. The rainfall–runoff process is affected by

parameters such as weather conditions, land use/land cover, water absorption by soil and evapotranspiration. Thus, it requires many reduction presumptions or extensive data about the basin. The second method of flow forecasting is the structure identification method which attempts to identify flow structure based on their past records. In this method, complete conception of the physical principles is not required, and the data requisites are not as large as for the model process [23]. Artificial intelligence (AI) techniques such as artificial neural network (ANN) have been extensively used to model wide range of hydrologic processes [6, 8, 9]. The predictive performance of ANN was found better than empirical methods for different hydrologic problems, but performance of ANN models is affected by several user-defined parameters. The theory of a support vector machine (SVM) has been recently introduced by Vapnik [36]. Researchers employed SVM for different points. For example, using SVM, Cimen [5] anticipated suspended sediment load in two streams, Kisi and Cimen [13] displayed reference evapotranspiration and Radhika and Shashi [28] modeled daily extreme air temperature.

In the last decade, wavelet analysis, which is a superior alternative data preprocessing method to Fourier methods, showed successful implementation in various water resource engineering applications due to its ability to decompose a signal in both time and frequency domain. Wavelet and ANN (WANN) combination has drawn increasing interest and has showed benefits over single ANN, ANFIS, SVM and ARMA models in terms of exact fitting and forecasting [3, 11, 14, 22, 33, 38].

In hydrology and water resource management, Wang and Ding [38] used WANN to model shallow groundwater level in Beijing and daily flow of the Yangtze River in China and suggested that proposed model could expand prediction precision and increase the time horizon of forecasting. Kim and Valdes [11] used dyadic wavelet transforms and ANN to predict droughts in the Conches River basin in Mexico and showed that the integrated model successfully enhanced the capability of ANN for predicting the regional drought. Zhou et al. [42] applied a wavelet transform predictor–corrector method to break up a signal into high- and low-frequency sub-series. Afterward, apiece of sub-series was forecasted utilizing an ARMA model, and a correction method was performed for the total of the forecasting time series. At last, prediction flow using ARMA, seasonal ARIMA and neural network methods, they resulted in the wavelet predictor–corrector method to have the best prediction precision. What's more, the decomposition level demonstrated no clear impact on the forecasting for the monthly flow data. So as to predicting monthly flow, Kisi [12] applied a wavelet-ANN (WANN) hybrid model. Comparison of the results of wavelet-ANN model forecasting with those of a multilayer perceptron (MLP), a MLR and AR models indicated that the WANN model performed better than the other methods. Wang et al. [40] used the multi-scale features of wavelet analysis and the nonlinear ability of ANN to forecast inflow of Dams in Yangtze River, China. Utilizing both a WANN model and a kind of threshold AR model to forecast runoffs, they resulted the WANN model was more precise than the ANN model. Adamowski and Sun [1] employed wavelet-neural networks (WA-ANN) and single ANN methods for flow prediction at lead times of 1 and 3 days for two rivers in Cyprus. They concluded that coupled model forecasts flow more accurately than the single ANN model. More recently, many other researches and studies have been undertaken in the field of hydrology and river basin management. For example, Wei et al. [41] employed wavelet-ANN (WANN) models with distinctive wavelet decomposition levels to forecast river flows in 48 months ahead of time. Examination of results showed that the execution of the ANN model was enhanced by WNN models. Kalteh [10] studied the relative accuracy of ANN and support vector regression (SVR) models combined with wavelet transform in monthly river flow prediction and compared them with ANN and SVR methods. The comparison of the outcomes indicated that both single ANN and SVR models combined with wavelet transform were capable to give more precise results than the customary ANN and SVR models [25]. Okkan and Serbes [26] combined wavelet analysis and black box methods [multiple linear regression (MLR), feed-forward neural networks (FFNN) and least square support vector machines (LSSVM)] for modeling the reservoir inflow. They watched that combined model gave better accuracy than black box models. Seo et al. [31] developed and applied two WANFIS and WANN coupled models for daily water-level prediction. They found that combination of wavelet transform and AI models can increase accurate prediction daily water level and can give better performance than the conventional prediction methods. Seo et al. [32] used wavelet-ANN (WANN) and the wavelet-ANFIS (WANFIS) models for flood prediction. First, wavelet transform was applied to break up the signal into approximation and detailed sub-series. Then, the obtained sub-series were applied as inputs of the ANN and ANFIS models in the WANN and WANFIS methods, respectively. Their results showed that hybrid models performed superior to ANFIS and ANN for various lead times. They also found that WANN and WANFIS methods gave comparable results for all lead times. Shafaei and Kisi [33] employed three WANFIS (wavelet-ANFIS), WSVR (wavelet-SVR) and WARMA (wavelet-ARMA) hybrid procedures for

estimating monthly lake-level changes. They found that three hybrid models forecasted more accurately than the single models.

All these researches indicated that WANN is a useful tool for accurately locating irregularly distributed multi-scale characteristics of hydrologic time series in space and times, and also, ANN and SVM models are effective methods for forecasting non-stationary time series such as hydrologic time series. Thus, keeping in view the improved performances by different modeling techniques, this study is planned to compare the performance of WANN with the SVM and ANN for short-term daily flow forecasting of Vanyar station, on Ajichai River, located in northwest region of Iran.

2 Methodology

2.1 Wavelet transforms

Wavelet transform is utilized to analyze data in view of its ability to draw out the suitable time frequency information from non-stationary and temporary signals. Wavelet decomposes the frequency constituents of signals. Wavelet functions disintegrate the data into various frequency constituents and then study each component with a resolution corresponding to its scale. The lower scale points in the compressed wavelet are capable to extract the high-frequency constituent or the quickly altering details of the signal. The higher scales indicate the stretched version of a wavelet, and the equivalent coefficients reveal the slowly altering characteristic of a low-frequency constituent [15]. There are two main wavelet transformation approaches: continuous wavelet transform (CWT) and discrete wavelet transform (DWT). Since hydrologists do not have at their disposal a continuous time series but rather a discrete time series, thus for practical applications in hydrology, researchers have used discrete time signal rather than a continuous time signal. Moreover, continuous wavelet transform provides redundant information which may not be helpful for analyzing the hydrologic time series, while discrete wavelet transform produces N^2 coefficients from a time series with length N [29]. In this study, the Mallat's algorithm was applied for the DWT of daily time series. According to the Mallat's algorithm, the discrete wavelet transform of signal $x(t)$ is presented as [20]:

$$W_{j,k} = 2^{-j/2} \sum_{t=0}^{N-1} \psi\left(2^{-j}t - k\right) x_t \tag{1}$$

where t is integer time stages, and j and k are integers that control the scale and time, respectively. $W_{j,k}$ is the wavelet coefficient for the scale parameter, $s = 2^j$, and the time parameter, $\tau = 2^j k$. The time series is broken up into two components: one included the low frequency (approximation), and one included the high frequencies and the fast events (the detail). In the present study, the detail and approximation sub-time series are generated by using Eq. (1) [14]. The more information about wavelet transform can be found in Wei et al. [41].

After the discrete wavelet transform, effective details and approximation should be defined since correlated sub-series constituents may decrease the generalization abilities of the models. For this purpose, all possible regression methods are executed using Mallows' C_p based since this is an efficient approach to define the subset of main sub-series in cases where there are various potential predictor variables [21]. Mallows' C_p is a parameter of the error in the best subset model, in respect of the error including all variables. C_p parameter is nearly equal to the number of variables in the appropriate models. The C_p parameter can be defined [21]:

$$C_p = (N-k)\frac{\mathrm{MSE}_i}{\mathrm{MSE}_f} - (N - 2i - 1) \tag{2}$$

N the number of data, MSE_i the mean of residual squares in the model with (i) variable and MSE_f the mean of residual squares in the full model with k variable. Effective details and approximation time series determined with all the possible regression methods were used as inputs in this study.

2.2 Artificial neural networks (ANNs)

The ANN is an estimating framework designed after the treatment of biological neural networks. In the present research, ANN models were built by a multilayer perceptron network (MLPN) with one hidden layer trained by the back-propagation algorithm (BPA) [30]. This model has been most commonly applied in the ANN structure [18, 19, 20]. The mathematical formulation of the MLPN feed-forwarding process is as defined as follows:

$$Y_N = F\left(\sum_{M=1}^{L} W_{NM} X_M + B_N\right) \tag{3}$$

where M nodes previous layer, N nodes present layer, X nodal values in the previous layer, Y nodal values in the present layer, W weight values between X and Y, B bias in the present layer, L the number of node in the previous layer and F activation function.

Log-sigmoid function was employed in this study as an activation function. The weight-updating rule of the ANN training operation by BPA is explained as follows:

$$W^{i+1} - W^i = \theta\left(W^i - W^{i-1}\right) + (1-\theta)\eta\left(\frac{-\partial E^i}{\partial W^i}\right) \tag{4}$$

where i ith iteration, η learning rate, θ momentum value and E the sum of squared errors between measured and

predicted data. With a specific end goal to obtain the ANN parameters (the number of hidden nodes, the learning rate and the momentum value), the data set of the training period is shared into two stages of training and test sets. The model parameters were chosen by the trial-and-error method, and one hundred sets of random primitive weights were assumed to prevent the ANN model being captivated in a local minimum [34]. Model performing process was carried out by MATLAB Software. Optimal selection of parameters largely effects on the efficiency of ANN's model. It is important to determine the number of neurons in the hidden layer. ANN models were applied in this research include an input layer with the neurons changing from 1 to 4, an output layer and a hidden layer. The optimal number of neurons in the hidden layer is obtained by the trial-and-error operation by changing the number of neurons from 3 to 15. In the modeling ANN, the input and output data are normalized to certify that all variables obtain equal consideration, and the performance of the training network is increased [7]. Overfitting is a main issue that happens during neural network training. Regularization is one of the procedures of improving generalization, which includes improving the efficiency function, usually by modifying the mean sum of squares of the network error (MSE) by adding a phrase including the summation of squares of the network weights and biases [2]:

$$\mathrm{MSE}_{\mathrm{rgn}} = \beta\,\mathrm{MSE} + (1-\beta)\mu \tag{5}$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n}\omega_i^2 \tag{6}$$

where $\mathrm{MSE}_{\mathrm{rgn}}$, β, ω and μ are the regularization MSE, the performance ratio, the weights of the network MSE and their mean, respectively.

2.3 Support vector machine (SVM)

The SVM has become a popular and promising estimator in data-driven study fields, of which key thought and theory have been presented by Vapnik [36, 37]. The primary methodology of SVM model development includes choice support vectors which support the model structure and define their weights. The process of an SVM estimator (f) on regression can be defined as follows:

$$f(x) = w\varphi(x) + b \tag{7}$$

where w and b indicate a weight vector and bias, separately. Likewise, j is a nonlinear transfer function mapping the input space into a high-dimensional specific space. Vapnik [36] displayed an object function of convex optimization with ε-insensitivity loss function, whose mathematical explanation is disclosed:

$$\begin{aligned} &\min_{w,b,\xi,\xi^*} \quad \frac{1}{2}w^2 + C\sum_{k=1}^{N}\left(\xi_k + \xi_k^*\right) \\ &\text{Subject to} \quad \begin{cases} y_k - W^T\varphi(X_k) - b \le \varepsilon + \xi_k \\ W^T\varphi(X_k) + b - y_k \le \varepsilon + \xi_k^* \\ \xi_k, \xi_k^* \ge 0 \end{cases} \\ &k = 1, 2, \ldots, N \end{aligned} \tag{8}$$

where ξ_k and ξ_k* denote slack variables penalizing estimation error by the ε-insensitivity loss function, and C is a positive trade-off operator for the evaluation of the empirical error. In this research, for computing this optimization issue, the successive minimize optimization (SMO) algorithm, displayed by Platt [27], was utilized. The object function of Eq. (9) can be reformulated utilizing Lagrangian multipliers of (α, α*), as characterized below:

$$\begin{aligned} &\max_{\alpha,\alpha^*} \quad \begin{cases} -1/2\sum_{k,l=1}^{N}\left(\alpha_k - \alpha_k^*\right)\left(\alpha_l - \alpha_l^*\right)K(X_k, X_l) \\ -\varepsilon\sum_{k=1}^{N}\left(\alpha_k + \alpha_k^*\right) + \sum_{k=1}^{N} y_k\left(\alpha_k - \alpha_k^*\right) \end{cases} \\ &\text{Subject to} \quad \begin{cases} \sum_{k=1}^{N}\left(\alpha_k - \alpha_k^*\right) = 0 \\ 0 \le \alpha_k, \alpha_k^* \le C \end{cases} \end{aligned}$$

to acquire

$$\sum_{k=1}^{N}\left(\alpha_k - \alpha_k^*\right)K(X, X_k) + b \tag{9}$$

where K and n mean a kernel function and the number of support vectors, individually. In this study, a radial basis kernel function with coefficient σ was used, which can be clarified as

$$K(X_k, X_l) = \exp\left(-\frac{X_k - X_l^2}{2\sigma^2}\right) \tag{10}$$

The SVM model parameters trained by SMO are C, ε and σ. The model parameter choice rule and the stopping criteria were like the ANN. The SVR codes, downloadable from the LIBSVM Web site (http://www.csie.ntu.edu.tw/cjlin/libsvm), were utilized for SVM modeling in this study. Detailed information and overfitting issue can be obtained from Platt [27].

3 Data and statistical analysis

The daily flow data of Vanyar station, on Ajichai River, located in northwest region of Iran, is used in this study (Fig. 1). Table 1 summarizes the geographical location of the Vanyar station. The observed time series is 34 years (12,494 days) long with statistical period between January 1975 and March 2009. No formula has been expressed for data division between training and testing period. In this research, the first 29 years of daily flow data (10,620 days,

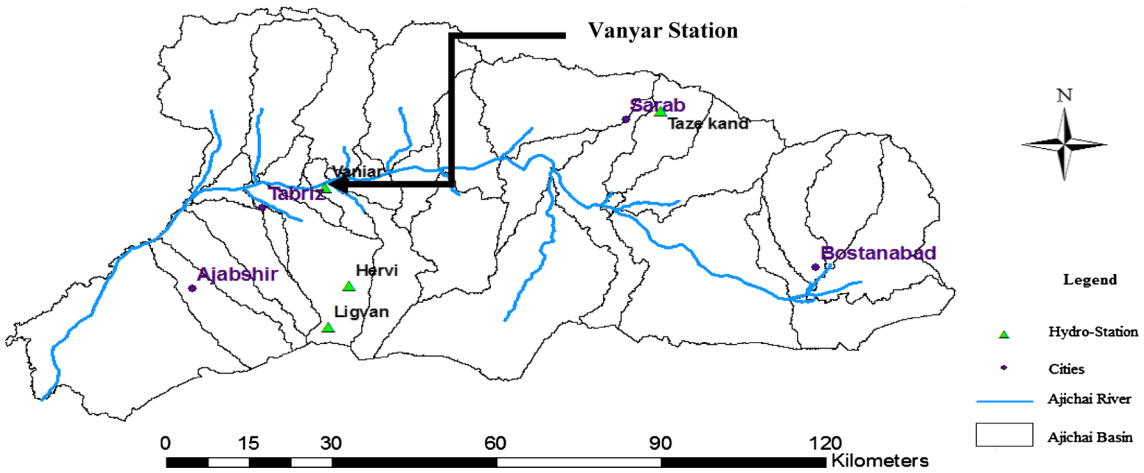

Fig. 1 The location of the Ajichai Basin, Ajichai River and Vanyar station

Table 1 Summary of Ajichai Basin

Location	Ajichai Basin
Latitude (°N)	38°7′
Longitude (°E)	46°24′
Altitude (meters above mean sea level)	3882
Basin area (km^2)	5.528
Slope around the site (%)	13.30
River length to Vanyar station (km)	227.2

85 % of the entire data set) are utilized for training, and the remaining 5 years (1874 days, 15 % of the entire data set) are utilized for testing. The daily statistics of the data sets are introduced in Table 2. In this table, X_{mean}, S_x, C_{sx}, X_{min} and X_{max} indicate the general mean, standard deviation, skewness, minimum and maximum, separately. The observed daily flow data indicate rather high-positive skewness values ($C_{sx} = 3.95$) showing scattered distribution.

The purpose of this study was to forecast short-term river flow of Vanyar station on Ajichai River (Q_{t+1}, Q_{t+2}, Q_{t+3}, Q_{t+4}) at time ($t + 1$, $t + 2$, $t + 3$, $t + 4$), respectively. The flow value is considered as a function of finite sets of antecedent flow observations. In modeling by WANN, ANN and SVM methods, four different model structures (different combinations) can be expressed as

(1) $Q_{t+1} = f(Q_t)$
(2) $Q_{t+1} = f(Q_t, Q_{t-1})$
(3) $Q_{t+1} = f(Q_t, Q_{t-1}, Q_{t-2})$
(4) $Q_{t+1} = f(Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3})$

Q_t, Q_{t-1}, Q_{t-2} and Q_{t-3} denote different lag time series of river flow at time t, $t - 1$, $t - 2$ and $t - 3$, respectively. In modeling the input and output, data are normalized so as to provide equal consideration to all variables [39].

$$Z_i = \frac{(Q_t - Q_{min})}{(Q_{max} - Q_{min})} \tag{11}$$

where Q_t, Q_{min}, Q_{max} and Z_i denote observed flow, minimum flow, maximum flow and normalized flow, respectively. In this research, the following error statistics were employed to assess the models, where n and t are the number of data and time, respectively. Q_{ot} is the observed flow, $\widehat{Q_{pt}}$ is the predicted flow, $\overline{Q_{ot}}$ and $\overline{\widehat{Q_{pt}}}$ are the mean of the observed flow and the mean of the predicted flow, respectively:

Correlation coefficient (R):

$$R = \frac{\sum_{t=1}^{n}\left(Q_{ot} - \overline{Q_{ot}}\right)\left(\widehat{Q_{pt}} - \overline{\widehat{Q_{pt}}}\right)}{\sqrt{\sum_{t=1}^{n}\left(Q_{ot} - \overline{Q_{ot}}\right)^2 \sum_{t=1}^{n}\left(\widehat{Q_{pt}} - \overline{\widehat{Q_{pt}}}\right)^2}} \tag{12}$$

Table 2 Statistical parameters of the used data set

Data period	X_{min} (m^3/s)	X_{max} (m^3/s)	X_{mean} (m^3/s)	S_x (m^3/s)	C_{sx}
Total study period	0	342	10.38	19.79	3.92
Training period	0	342	11.85	21.38	3.72
Testing period	0	134	5.954	12.98	3.95

Mean squared error (MSE):

$$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n}\left(Q_{ot} - \widehat{Q_{pt}}\right)^2 \quad (13)$$

Mean absolute error (MAE):

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|Q_{ot} - \widehat{Q_{pt}}\right| \quad (14)$$

Nash–Sutcliffe (NS):

$$\mathrm{NS} = \left(1 - \frac{\sum_{t=1}^{n}\left(Q_{ot} - \widehat{Q_{pt}}\right)^2}{\sum_{t=1}^{n}\left(Q_{ot} - \overline{Q_{ot}}\right)^2}\right) \times 100 \quad (15)$$

4 Application and results

4.1 One-day-ahead flow forecasting using SVM and ANN models

Multilayer perceptron network (MLPN) with one hidden layer trained by the back-propagation algorithm (BPA) was used in the present study for short-term river flow forecasting of Vanyar station, on Ajichai. This MLPN model has been successfully used in several studies [8, 35]. Logistic sigmoid transfer function was applied in the hidden layer. The network was trained in 1000 epochs utilizing the Levenberg–Marquardt learning algorithm with a learning rate of 0.001 and a momentum value of 0.7. In ANN modeling, the number of neurons in the input and output layers is determined depending on the number of input and output variables of the system under study, respectively. The SVM model optimal parameters trained by SVM are chosen using rule and the stopping criteria.

The RMSE, MAE, NS and R statistics of ANN and SVM models for training and testing data sets are provided in Table 3. Results from Table 3 suggest that the SVM model has, trained using input comprising 1-day flow (input combination 4), performed the best with respect to the R, NS and MAE values. The results from Table 3 also suggest that ANN (4, 9, 1) provides the best performance among other ANN models with respect to the R, NS, MAE and MSE criteria. However, ANN (3, 6, 1) model approximately has similar R, NS and MSE values with ANN (4, 9, 1) model in 1-day-ahead flow forecasting. Comparison of the optimal ANN (input combination 4) and SVM (input combination 4) models indicates that the both ANN and SVM models have similar results in the test period.

4.2 One-day-ahead flow forecasting using wavelet-ANN (WANN) model

Wavelet-ANN (WANN) model was also used to forecast daily flow data of only one station data (Vanyar station). The WANN approach uses both discrete wavelet transformed (DWT) data and ANN. DWT decomposes the time series data into sub-series including an approximation sub-series (A_i) with low frequency and several detailed sub-series (D_1, D_2, …, D_i) with high frequency, where i is the number of decomposition levels. One of the major issues in

Table 3 Performance results for 1-day-ahead flow forecasting by ANN and SVM models

Model inputs	Best structure (m, n, p)[a]	Training			Testing			
		MSE $(m^3/s)^2$	R	MAE (m^3/s)	MSE $(m^3/s)^2$	R	MAE (m^3/s)	NS (%)
ANN								
Q_t	(1, 5, 1)	13.0	0.938	1.02	14.80	0.921	1.43	85.1
Q_t, Q_{t-1}	(2, 10, 1)	14.0	0.930	1.41	14.04	0.929	1.41	86.0
Q_t, Q_{t-1}, Q_{t-2}	(3, 6, 1)	13.0	0.946	0.90	13.87	0.930	1.19	86.2
Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}	**(4, 9, 1)**	**14.0**	**0.941**	**1.51**	**13.80**	**0.931**	**1.14**	**86.3**
Model inputs	Best structure (C, ε, σ)[b]	Training			Testing			
		MSE $(m^3/s)^2$	R	MAE (m^3/s)	MSE $(m^3/s)^2$	R	MAE (m^3/s)	NS (%)
SVM								
Q_t	(10, 0.001,0.26)	19.3	0.933	1.36	25.88	0.924	1.43	84.6
Q_t, Q_{t-1}	(15, 0.01, 0.54)	14.7	0.956	1.11	18.99	0.945	1.21	88.6
Q_t, Q_{t-1}, Q_{t-2}	(90, 0.002, 0.34)	14.7	0.954	1.11	19.38	0.943	1.26	88.5
Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}	**(100, 0.001, 0.19)**	**11.4**	**0.964**	**1.00**	**14.87**	**0.955**	**1.10**	**88.7**

[a] m, n, p denote the number of input neurons, hidden neurons and output neurons, respectively

[b] C, ε, σ denote the optimal parameters of SVM model

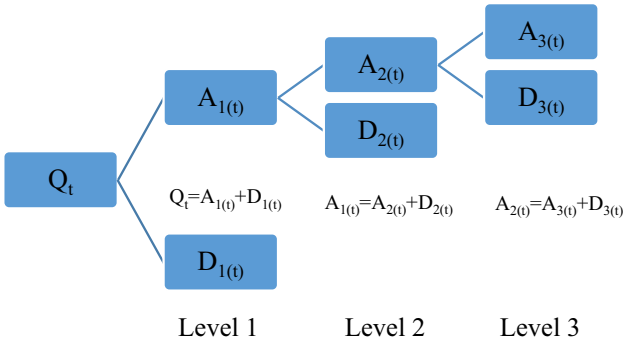

Fig. 2 Mallat's algorithm demonstration for three-level decomposition of a signal

the use of DWT is to select the type of mother wavelet. Daubechies wavelets (dbN) are one of the wavelets extensively applied in this type of studies [41]. Prosperous usage of the dbN for hydrologic data decomposition has also been reported by Kisi [12] and Nourani et al. [22, 24]. In the present study, Daubechies wavelet 5 (db_5) was used as the mother wavelet to decompose the time series data. Time series at time t (Q_t) was decomposed to three decomposition levels, and $A_{3(t)}$, $D_{3(t)}$, $D_{2(t)}$ and $D_{1(t)}$ sub-series were obtained by MATLAB codes which include Mallat's DWT algorithm. Mallat's DWT algorithm demonstration for three-level decomposition of a signal (time series) is presented in Fig. 2. Then, various linear regression analyses were performed using the "best subsets regression" tool of Minitab software to define the effective details and approximation sub-series since some correlated sub-series may decrease the generalization abilities of the model. So, the effective details and approximation were selected by the Mallows coefficients (C_p) as this is an efficient approach to determine the subset of effective sub-series in cases where there are a number of potential predictor variables [21]. Although all regressions with various sub-series combinations are assessed, only the most accurate sub-series combinations are provided in Table 4. Subsequently, Q_{t-1}, Q_{t-2} and Q_{t-3} (responses) were decomposed into four sub-series, and different linear regression analyses were performed for each response. Results of linear regression analyses are also provided in Table 4. The best combinations of the effective sub-series specified by bold and highlighted characters have the minimum C_p values and almost explain full linear model. For example, Q_t was first decomposed into $A_{3(t)}$, $D_{3(t)}$, $D_{2(t)}$ and $D_{1(t)}$ series. Then, the effective variables were chosen by assessing Mallows C_p values. Performance of the model with optimum combination of the input variables is almost the same as that of the full linear model with three variables. That is, the explained variance of Q_t values, which have the minimum C_p, ($C_p = 4.8$), is nearly equal to that explained by the full linear model. According to these results, $D_{1(t)}$ sub-series was eliminated. Summed series at time t (TW_t) was computed by summation of the $A_{3(t)}$, $D_{3(t)}$ and $D_{2(t)}$ effective sub-series. Subsequently, the effective sub-series of Q_{t-1}, Q_{t-2}, Q_{t-3} and values was selected from the results of Table 4, according to minimum Mallows C_p computed from regression analysis. The effective sub-series of Q_t, Q_{t-1}, Q_{t-2} and Q_{t-3} responses was defined as follows:

$$Q_t = f(A_{3(t)}, D_{3(t)}, D_{2(t)})$$
$$Q_{t-1} = f(D_{2(t-1)}, D_{3(t-1)}, A_{3(t-1)})$$
$$Q_{t-2} = f(A_{3(t-2)}, D_{3(t-2)}, D_{2(t-2)})$$
$$Q_{t-3} = f(A_{3(t-3)}, D_{3(t-3)}, D_{2(t-3)})$$

Therefore, TW_{t-1}, TW_{t-2} and TW_{t-3} that were summed series at time $t-1$, $t-2$ and $t-3$ were obtained by adding the effective sub-series of the Q_{t-1}, Q_{t-2}, Q_{t-3} and Q_{t-4} values.

$$TW_t = A_{3(t)} + D_{3(t)} + D_{2(t)}$$
$$TW_{t-1} = D_{2(t-1)} + D_{1(t-1)} + A_{3(t-1)}$$
$$TW_{t-2} = A_{3(t-2)} + D_{3(t-2)} + D_{2(t-2)}$$
$$TW_{t-3} = A_{3(t-3)} + D_{3(t-3)} + D_{2(t-3)}$$

For modeling by WANN model, 85 and 15 % of the input data set [(1) TW_t, (2) TW_t and TW_{t-1}, (3) TW_t, TW_{t-1} and TW_{t-2}, (4) TW_t, TW_{t-1}, TW_{t-2} and TW_{t-3}] were used for training and testing, respectively. Optimal values of hidden layer neuron numbers were obtained by the trial-and-error method. Logistic sigmoid transfer function was applied in the hidden layer. Before providing these summed series (TW_t, TW_{t-1}, TDW_{t-2} and TW_{t-3}) as input values to ANN, all inputs were normalized using Eq. (11). Figure 3 indicates the schematic structure of the WANN model with an input of Q_t.

The performance of the WANN models in training and testing is provided in Table 5. Table 5 shows that the WANN (2, 9, 1) provides better performance in comparison with other input combinations with respect to the R, NS and MSE values. Table 5 suggests that WANN models with various input combinations approximately have similar performance in terms of various criteria. It can be found from Tables 3 and 5 that the predictive efficiency of the WANN models is much better than the single ANN and SVM models with respect to the R, MSE and MAE values in Vanyar station.

The ANN and SVM models perform relatively poor for flow modeling in comparison with WANN model with the used data set in our case study. A possible reason may be using only flow data as input values for modeling, whereas each flow time series includes various frequency components. Utilizing only one resolution component to

Table 4 Summary of linear regression analysis the responses Q_t, Q_{t-1}, Q_{t-2} and Q_{t-3}

Variables	R	$R_{(adj)}$	C_p	S (m^3/s)	$A_{3(t)}$	$D_{3(t)}$	$D_{2(t)}$	$D_{1(t)}$
Response (Q_t)								
1	88.3	88.3	9000.6	6.75	*			
1	3.9	3.8	164,620.3	19.4		*		
2	92.2	92.2	1901.8	5.53	*	*		
2	89	89	7803.9	6.56	*		*	
3	**93.8**	**93.8**	**4.8**	**5.13**	*	*	*	
3	92.6	92.6	1201.7	5.39	*	*		*
4	93.2	93.2	5	5.15	*	*	*	*
Variables	R	$R_{(adj)}$	C_p	S (m^3/s)	$A_{3(t-1)}$	$D_{3(t-1)}$	$D_{2(t-1)}$	$D_{1(t-1)}$
Response (Q_{t-1})								
1	86.2	86.1	2273.4	7.36	*			
1	1.5	1.5	92,479.3	19.63		*		
2	87.7	87.7	652.8	6.94	*	*		
2	86.7	86.7	1664.8	7.21	*		*	
3	**88.2**	**88.2**	**5.2**	**6.76**	*	*	*	
3	87.7	87.7	613.7	6.93	*	*		*
4	88.1	88.1	5.6	6.77	*	*	*	*
Variables	R	$R_{(adj)}$	C_p	S (m^3/s)	$A_{3(t-2)}$	$D_{3(t-2)}$	$D_{2(t-2)}$	$D_{1(t-2)}$
Response (Q_{t-2})								
1	82.7	82.7	747.9	8.21	*			
1	0.9	0.9	63,511.7	19.64			*	
2	83.7	83.7	27.3	7.991	*		*	
2	82.8	82.8	723.7	8.21	*	*		
3	**83.7**	**83.7**	**4.8**	**7.98**	*	*	*	
3	83.7	83.7	29.2	7.99	*		*	*
4	83.7	83.7	5	7.98	*	*	*	*
Variables	R	$R_{(adj)}$	C_p	S (m^3/s)	$A_{3(t-3)}$	$D_{3(t-3)}$	$D_{2(t-3)}$	$D_{1(t-3)}$
Response (Q_{t-3})								
1	79	78.9	317.7	9.07	*			
1	0.5	0.5	48,057.6	19.79		*		
2	79.4	79.4	24.3	8.97	*	*		
2	79	79	296.6	9.07	*		*	
3	**79.5**	**79.5**	**3.2**	**8.96**	*	*	*	
3	79.4	79.4	26.1	8.97	*	*		*
4	79.5	79.5	5	8.96	*	*	*	*

Asterisk symbols indicate considered sub-series in each combination

Bold symbols show the best combinations of the effective sub-series for prediction of flow

forecast, the flow data do not readily describe the internal characteristics of the time series [4]. Therefore, WANN model with different frequency sub-series of time series is prosperous in forecasting of the applied time series. For the flow data, the relative MSE, R, NS and MAE differences between the optimum result of WANN (2, 9, 1) and that of ANN(4, 9, 1) are −122.6, 3.92, 7.99 and −12.87 in the test period, respectively. Also, the relative MSE, R, NS and MAE differences between the best result of WANN (2, 9, 1) and that of SVM (input combination (4)) are −139.95, 1.44, 5.43 and −8.91 % in the test period, respectively.

4.3 Forecasting 2-, 3- and 4-day-ahead flows using WANN, SVM and ANN models

In the next section of this study, ANN and SVM models were applied for 2-, 3- and 4-day-ahead flow forecasting of Vanyar station. For the ANN and SVM modeling, four input combinations (1), (2), (3) and (4) are imported as

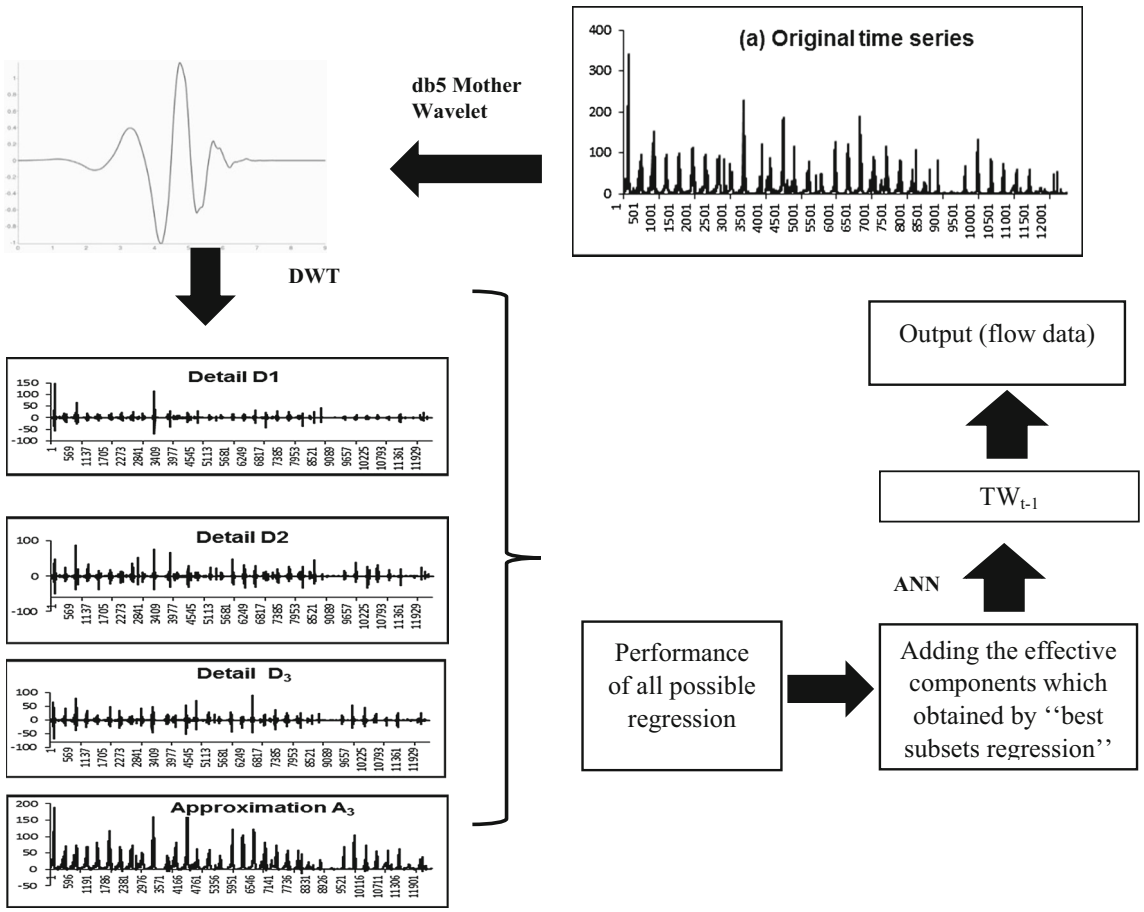

Fig. 3 The WANN model structure

Table 5 Performance results for 1-day-ahead flow forecasting by WANN model

Model inputs	Best structure (m, n, p)[a]	Training				Testing			
		MSE $(m^3/s)^2$	R	MAE (m^3/s)	NS (%)	MSE $(m^3/s)^2$	R	MAE (m^3/s)	NS (%)
TW_t	(1, 5, 1)	6.29	0.972	0.95	92.7	7.58	0.963	0.87	92.4
TW_t, TW_{t-1}	**(2, 9, 1)**	**6.14**	**0.972**	**1.01**	**92.9**	**6.20**	**0.969**	**1.01**	**93.8**
TW_t, TW_{t-1}, TW_{t-2}	(3, 3, 1)	5.99	0.971	0.91	93.6	6.78	0.966	1.29	93.2
TW_t, TW_{t-1}, TW_{t-2}, TW_{t-3}	(4, 12, 1)	6.23	0.981	1.20	94.4	6.54	0.968	0.91	93.5

Bold symbol indicates the best combination for flow forecasting

[a] m, n, p denote the number of input neurons, hidden neurons and output neurons, respectively

input to the ANN and SVM models, and the flows in the output neurons correspond to time $t + 2$, $t + 3$ and $t + 4$ for the 2-, 3- and 4-day-ahead forecasting, respectively. Results of forecasts using these models are given in Tables 6 and 7. Bold symbol indicates the best combination for flow forecasting.

In general, the optimum performances of the ANN and SVM models in Tables 6 and 7 confirm that the ANN models have better testing results than the SVM model in 2-, 3- and 4-day-ahead forecasting: higher R, NS and smaller MSE, MAE. Afterward, the WANN model was applied for forecasting flow data. For this purpose, the same inputs and parameters used in 1-day-ahead flow forecasting in WANN model were applied for short-term flow forecasting. The output in this case consists of flow values at time $t + 2$, $t + 3$ and $t + 4$. The results of 2-, 3- and 4-day-ahead flow forecasting for testing period are provided in Table 8. It is clear from Table 8 that the WANN (3, 5, 1) performs better than the other applied combinations according to different performance criteria in 2-day-ahead flow forecasting. Table 8 indicates that the WANN (3, 9, 1) model, whose inputs are the flow of

Table 6 Performance results for 2-, 3- and 4-day-ahead flow forecasting by ANN model

Model inputs	Best structure (m, n, p)[a]	Training				Testing			
		MSE $(m^3/s)^2$	R	MAE (m^3/s)	NS (%)	MSE $(m^3/s)^2$	R	MAE (m^3/s)	NS (%)
2 Days ahead forecast									
Q_t	(1, 7, 1)	26.6	0.881	1.21	73.6	26.9	0.857	2.12	73.4
Q_t, Q_{t-1}	**(2, 4, 1)**	**26.5**	**0.899**	**0.91**	**73.7**	**26.8**	**0.865**	**1.83**	**73.5**
Q_t, Q_{t-1}, Q_{t-2}	(3, 5, 1)	26.4	0.855	2.11	73.5	38.6	0.798	2.76	61.7
Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}	(4, 3, 1)	26.0	0.888	1.11	73.6	26.8	0.856	2.12	72.3
3 Days ahead forecast									
Q_t	(1, 10, 1)	27.0	0.863	1.81	73.2	39.5	0.794	2.49	60.9
Q_t, Q_{t-1}	(2, 7, 1)	30.0	0.836	2.41	68.3	44.8	0.758	2.53	55.6
Q_t, Q_{t-1}, Q_{t-2}	**(3, 6, 1)**	**26.9**	**0.869**	**1.81**	**73.5**	**35.5**	**0.805**	**2.12**	**64.8**
Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}	(4, 5, 1)	26.4	0.900	0.70	73.4	39.5	0.799	2.49	60.9
4 Days ahead forecast									
Q_t	(1, 12, 1)	33.1	0.770	2.82	55.7	44.8	0.751	2.82	55.4
Q_t, Q_{t-1}	(2, 15, 1)	30.0	0.835	2.41	61.0	44.7	0.760	3.11	55.5
Q_t, Q_{t-1}, Q_{t-2}	(3, 4, 10)	34.0	0.827	2.71	56.0	44.4	0.753	2.82	56.0
Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}	**(4, 9, 1)**	**29.0**	**0.832**	**2.42**	**63.0**	**37.9**	**0.800**	**2.21**	**62.5**

Bold symbol indicates the best combination for flow forecasting

[a] m, n, p denote the number of input neurons, hidden neurons and output neurons, respectively

current and two previous days (input combination 3), has the highest R (0.919) and NS (83.2 %) and the smallest MSE [15.89 $(m^3/s)^2$] and MAE [1.65 (m^3/s)] in comparison with other models in 3-day-ahead flow forecasting. However, WANN (2, 7, 1) and WANN (3, 9, 1) models approximately have similar R and NS values, but WANN (2, 7, 1) has the smallest MAE [1.49 (m^3/s)] when compared to other models in 3-day-ahead flow forecasting. Also, the WANN (2, 4, 1) model with R (0.911), MSE [17.52 $(m^3/s)^2$], NS (82.6 %) and MAE [1.52 (m^3/s)] performs better than the other models with input combinations 1, 3 and 4 in 4-day-ahead flow forecasting.Bold symbol indicates the best combination for flow forecasting.

Comparing Tables 6, 7 and 8 indicates that the WANN models have better forecasting abilities than the ANN and SVM models according to different performance criteria in our case study. The comparison of the optimum results of 2-, 3- and 4-day-ahead flow forecasting by ANN and SVM models shows that the ANN model performs better than the SVM model from the various criteria viewpoints. For the WANN and ANN models, overfitting was a much more substantial matter due to the much smaller errors during training, but largest errors of all during testing. In this study, regularization method was applied for preventing from overfitting, as regularization method does not require to validate data set separated from the training data set and thus can use all available data [2]. The obstacle with regularization is that it is difficult to compute the best value for the performance ratio parameter, while Bayesian regularization (BR) [17] can prevail from this obstacle, which can compute optimal regularization parameters in an automated fashion. Wei et al. [41] also used ANN and WANN models for estimating river monthly flows and then applied regularization method for preventing from overfitting applied. Also, comparison of optimum results of 2-, 3- and 4-day-ahead flow forecasting of Vanyar station by WANN, ANN and SVM shows that increasing the prediction horizon decreases the accuracy forecasting. For example, WANN model preforms 1-day-ahead flow forecasting with MSE = 6.20 $(m^3/s)^2$, $R = 0.969$, NS = 93.8 % and MAE = 1.01 m^3/s and 4-day-ahead flow forecasting with MSE = 15.89 $(m^3/s)^2$, $R = 0.919$, NS = 82.6 % and MAE = 1.66 m^3/s. In ANN and SVM models, the relative MSE, R, NS and MAE differences between 2- and 4-day-ahead flow forecasting are 29.28, −8.12, −17.19, 17.6 % and 33.98, −15.70, −30.77, 16.70 %, respectively. WANN estimates the maximum peak as 62.42 m^3/s instead of observed 75.4 m^3/s, with an underestimation of 17.21 %, while the ANN results in 57.4 m^3/s, with an underestimation of 23.8 %, and SVM estimates 61.83, with an underestimation of 17.9 %. Also, the WANN forecasting of the second max peak, 60.2 m^3/s, is 53.26 m^3/s, with an underestimation of 11.5 %, while the ANN and SVM are obtained as 49.16 and 35.89 m^3/s, respectively. It can be resulted that WANN model estimates peak flows better than the ANN and SVM methods in our case study. Hydrographs and scatter plots of the optimal WANN, ANN and SVM models in 2-day-ahead flow forecasting in test period are displayed in Fig. 4. It can be found that the WANN model forecasts the flow data and

Table 7 Performance results for 2-, 3- and 4-day-ahead flow forecasting by SVM model

Model inputs	Best structure (C, ε, σ)	Training				Testing			
		MSE $(m^3/s)^2$	R	MAE (m^3/s)	NS (%)	MSE $(m^3/s)^2$	R	MAE (m^3/s)	NS (%)
2 Days ahead forecast									
Q_t	(22, 0.005, 0.7)	35.585	0.885	1.753	78.9	39.495	0.854	1.969	75.3
Q_t, Q_{t-1}	**(100, 0.06, 0.34)**	**30.945**	**0.891**	**1.569**	**81.0**	**34.018**	**0.862**	**1.894**	**72.5**
Q_t, Q_{t-1}, Q_{t-2}	(120, 0.022, 0.10)	20.364	0.914	1.362	83.1	42.458	0.842	2.047	74.8
Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}	(96, 0.09, 0.49)	29.256	0.892	1.502	81.1	45.124	0.835	2.144	73.2
3 Days ahead forecast									
Q_t	(43, 0.004, 0.12)	45.156	0.831	2.236	73.3	59.074	0.757	2.549	63.3
Q_t, Q_{t-1}	(59, 0.048, 0.54)	39.145	0.854	1.904	75.2	56.461	0.776	2.507	66.5
Q_t, Q_{t-1}, Q_{t-2}	**(290, 0.51, 0.15)**	**45.253**	**0.835**	**2.146**	**75.7**	**42.718**	**0.789**	**2.327**	**64.6**
Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}	(110, 0.037, 0.66)	41.33	0.852	2.175	77.4	60.2079	0.746	2.651	64.3
4 Days ahead forecast									
Q_t	(76, 0.0021, 0.33)	30.537	0.893	1.512	79.1	59.7744	0.711	3.00	59.8
Q_t, Q_{t-1}	(54, 0.092, 0.14)	46.236	0.832	2.193	73.5	54.627	0.737	2.88	61.6
Q_t, Q_{t-1}, Q_{t-2}	(480, 0. 32, 0.23)	38.547	0.790	2.129	69.1	56.465	0.720	2.915	60.6
Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}	**(200, 0.043, 0.97)**	**48.026**	**0.821**	**2.250**	**71.0**	**51.530**	**0.745**	**2.736**	**62.1**

Bold symbol indicates the best combination for flow forecasting

Table 8 Performance results for 2-, 3- and 4-day-ahead flow forecasting by WANN model

Model inputs	Best structure (m, n, p)[a]	Training				Testing			
		MSE $(m^3/s)^2$	R	MAE (m^3/s)	NS (%)	MSE $(m^3/s)^2$	R	MAE (m^3/s)	NS (%)
2 Days ahead forecast									
TW_t	(1, 6, 1)	14.7	0.923	1.525	84.8	24.5	0.875	2.088	75.7
TW_t, TW_{t-1}	(2, 10, 1)	12.3	0.937	1.365	85.7	15.3	0.923	1.489	84.8
TW_t, TW_{t-1}, TW_{t-2}	**(3, 5, 1)**	**5.71**	**0.975**	**1.059**	**94.0**	**6.91**	**0.965**	**1.152**	**93.1**
TW_t, TW_{t-1}, TW_{t-2}, TW_{t-3}	(4, 11, 1)	11.5	0.949	1.265	86.2	15.5	0.922	1.435	84.6
3 Days ahead forecast									
TW_t	(1, 5, 1)	17.4	0.913	1.598	83.0	23.1	0.885	1.823	77.9
TW_t, TW_{t-1}	(2, 7, 1)	11.5	0.940	1.236	86.5	17.8	0.911	1.495	82.3
TW_t, TW_{t-1}, TW_{t-2}	**(3, 9, 1)**	**10.4**	**0.948**	**1.114**	**86.8**	**15.9**	**0.919**	**1.659**	**83.2**
TW_t, TW_{t-1}, TW_{t-2}, TW_{t-3}	(4, 6, 1)	18.3	0.904	1.742	82.1	29.4	0.849	2.062	70.8
4 Days ahead forecast									
TW_t	(1, 6, 1)	23.1	0.880	0.012	80.1	23.1	0.887	1.806	78.0
TW_t, TW_{t-1}	**(2, 4, 1)**	**8.88**	**0.941**	**0.007**	**91.4**	**17.5**	**0.911**	**1.524**	**82.6**
TW_t, TW_{t-1}, TW_{t-2}	(3, 9, 1)	14.4	0.921	0.008	84.7	22.4	0.891	1.596	77.8
TW_t, TW_{t-1}, TW_{t-2}, TW_{t-3}	(4, 5, 1)	22.6	0.885	0.012	76.8	38.1	0.805	2.468	62.2

Bold symbol indicates the best combination for flow forecasting

[a] m, n, p denote the number of input neurons, hidden neurons and output neurons, respectively

peak flows more accurately than the ANN and SVM models in Vanyar station. It is obvious from Fig. 4 that the estimates of the WANN model are less scattered and closer to the trend line than those of the ANN and SVM models.

The scatter plots of optimum 3- and 4-day-ahead flow forecasting by WANN, SVM and ANN methods in testing period are indicated in Figs. 5 and 6, respectively. According to the R criterion and fit line equations, the WANN model forecasts better than the ANN and SVM models in 3- and 4-day-ahead flow forecasting in our case study. Increasing the time horizon significantly influences the forecast accuracy. Considering optimum results of 3- and 4-day-ahead flow forecasting, increasing prediction time length from the current day to 4-day causes to decline

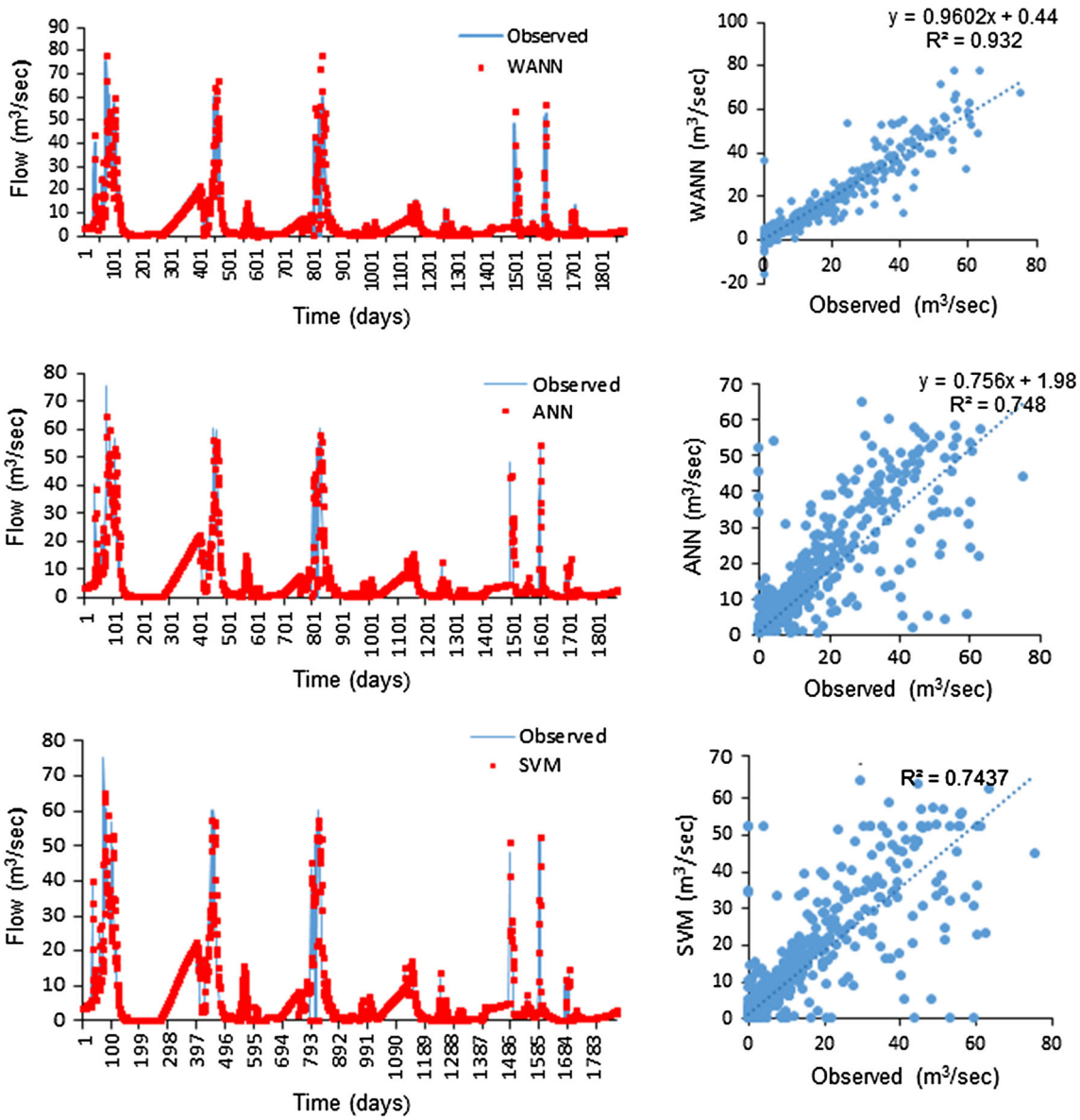

Fig. 4 The scatter plots and hydrographs of 2-day-ahead flow forecasts by WANN, SVM and ANN models in testing

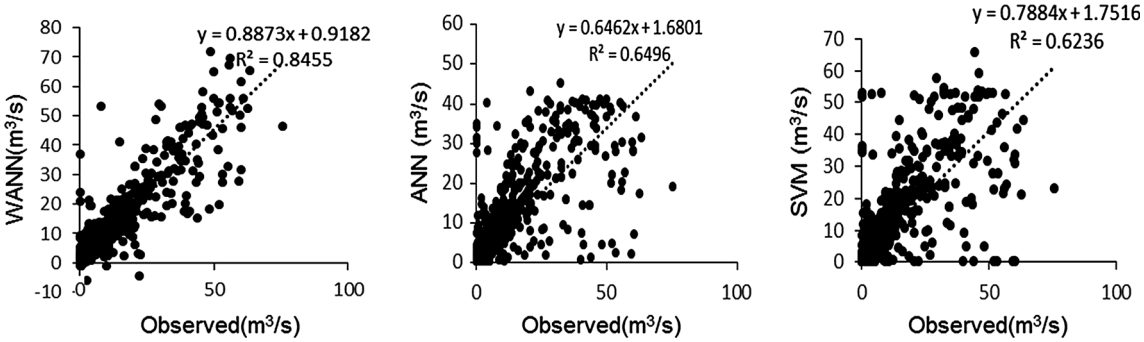

Fig. 5 The scatter plots of 3-day-ahead flow forecasts of the WANN, SVM and ANN models test period

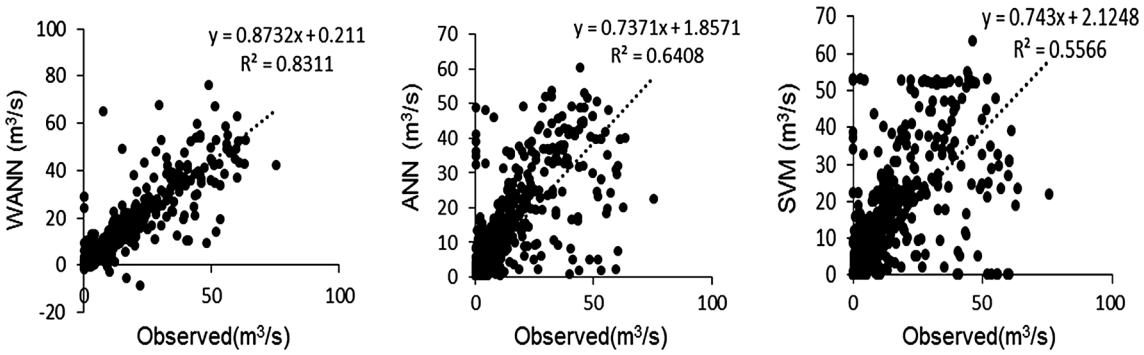

Fig. 6 The scatter plots of 4-day-ahead flow forecasts of the WANN, SVM and ANN models test period

the model accuracy. Figures 4, 5 and 6 show that the WANN models are better than the ANN and SVM models in Vanyar station, not only for 1-day-ahead flow prediction but also for 2-, 3- and 4-day-ahead prediction.

5 Conclusions

This study applied a WANN hybrid modeling technique for short-term river flow forecasting of Vanyar station on Ajichai River in Iran. This hybrid method was successfully used to estimate flow time series of Vanyar Station on Ajichai River. The time series was decomposed into four various frequency components at three decomposition levels using DWT and mother wavelet of Daubechies 5 (db_5). The results demonstrated that correlated frequency components (A_i and D_i) in regression analyses decrease the generalization ability of a model. So, the correlated components were removed using by all possible regression approaches. Importing effective frequency components as inputs, neurons of the ANN models makes the collinearity reduction. Mallows' C_p based on all possible regression analyses is an applied method to define the subset of decomposed series in cases where there are number of potential predictor variables. Then, summation of effective frequency components with different combinations was used as inputs of ANN model. The test results of WANN were compared with those of the SVM and ANN models. The results indicated that the WANN models having capability to model complex, non-stationary and nonlinear relations performed better than the SVM and ANN models in Vanyar station not only for 1-day-ahead flow predicting but also for 2-, 3- and 4-day-ahead and peak flow prediction. Comparing best results of the WANN, ANN and SVM models demonstrated that the 1-day-ahead flow forecasting of WANN model with MSE = 6.20 $(m^3/s)^2$, $R = 0.969$, NS = 93.8 % and MAE = 1.01 m^3/s has the best prediction ability than the others. In all cases, increasing forecasting horizon from 1-day to 4-day ahead caused to decrease the models' accuracy. Comparison of the test results of the optimal SVM and ANN models showed that ANN model has better accuracy than the SVM model in our case study not only in 1-day-ahead forecasting but also in 2-, 3- and 4-day-ahead flow forecasting.

References

1. 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
2. Beale MH, Hagan MT, Demuth MH (2010) Neural network toolbox 7 user's guide. Math Works Inc, Natick
3. Belayneh A, Adamowski J (2012) Standard precipitation index drought forecasting using neural networks, wavelet neural networks, and support vector regression. Appl Comput Intell Soft Comput. doi:10.1155/2012/794061
4. Chou CM, Wang RY (2002) On-line estimation of unit hydrographs using the wavelet-based LMS algorithm. Hydrol Sci J 47(5):721–738
5. Cimen M (2008) Estimation of daily suspended sediments using support vector machines. Hydrol Sci J 53(3):656–666
6. Cimen M, Kisi O (2009) Comparison of two different data-driven techniques in modeling lake level fluctuations in Turkey. J Hydrol 378:253–262
7. Dawson CW, Wilby RL (2001) Hydrological modelling using artificial neural networks. Prog Phys Geogr 25(1):80–108
8. Deswal S, Pal M (2008) Artificial neural network based modeling of evaporation losses in reservoirs. World Acad Sci Eng Technol 39:279–283
9. Dolling OR, Varas EA (2002) Artificial neural networks for streamflow prediction. J Hydraul Res 40(5):547–554
10. Kalteh AM (2013) Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. Comput Geosci 54:1–8
11. Kim TW, Valdes JB (2003) Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. J Hydrol Eng 8(6):319–328
12. Kisi O (2008) Stream flow forecasting using neuro-wavelet technique. Hydrol Process 22(20):4142–4152
13. Kisi O, Cimen M (2009) Evapotranspiration modelling using support vector machines. Hydrol Sci J 54(5):918–928
14. Kisi O, Cimen MA (2011) Wavelet-support vector machine conjunction model for monthly streamflow forecasting. J Hydrol 399:132–140

15. Kucuk M, Agiralioglu N (2006) Wavelet regression techniques for streamflow predictions. J Appl Stat 33(9):943–960
16. Kuchment LS, Demidov VN, Naden PS, Cooper DM, Broadhurst P (1996) Rainfall–runoff modelling of the Ouse basin, North Yorkshire: an application of a physically based distributed model. J Hydrol 181(1–4):323–342
17. MacKay DJC (1992) A practical Bayesian framework for back propagation networks. Neural Comput 4:448–472
18. Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: are view of modeling issues and applications. Environ Model Softw 15:101–124
19. Maier HR, Jain A, Dandy DC, Sudheer KP (2010) Methods used for the development of neural networks for the prediction of water resource variables in rivers system: current status and future directions. Environ Model Softw 25:891–909
20. Mallat SG (1989) A theory for multi resolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal 11(7):674–693
21. Mallows CL (1973) Some comments on C_p. Technometrics 15(4):661–675
22. 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
23. Nourani V, Kisi O, Komasi M (2011) Two hybrid artificial intelligence approaches for modeling rainfall–runoff process. J Hydrol 402:41–59
24. 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
25. Nourani N, Hosseini Baghanam A, Adamowski A, Kisi O (2014) Applications of hybrid wavelet–artificial intelligence models in hydrology: a review. J Hydrol 514:358–377
26. Okkan U, Serbes ZA (2013) The combined use of wavelet transform and black box models in reservoir inflow modeling. J Hydrol Hydromech 61(2):112–119
27. Platt JC (1999) Fast training of support vector machines using sequential minimal optimization. In: Schölkopf B, Burges CJC, Smolar AJ (eds) Advances in kernel methods—support vector learning. MIT Press, Cambridge
28. Radhika Y, Shashi M (2009) Atmospheric temperature prediction using support vector machines. Int J Comput Theory Eng 1(1):55–58
29. 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
30. Rumelhart DE, McClelland JL, The PDP Research Group (1986) Parallel distributed processing: explorations in the micro structure of cognition. MIT Press, Cambridge
31. Seo Y, Kim S, Kisi O, Singh VP (2014) Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. J Hydrol 520:224–243
32. Seo Y, Kim S, Singh VP (2015) Multistep-ahead flood forecasting using wavelet and data-driven methods. KSCE J Civil Eng 19(2):401–417
33. Shafaei M, Kisi O (2016) Lake level forecasting using wavelet-SVR, wavelet-ANFIS and wavelet-ARMA conjunction models. Water Resour Manag 30(1):79–97. doi:10.1007/s11269-015-1147-z
34. Shiri J, Kisi O, Yoon H, Lee K, Nazemi AH (2013) Predicting ground water level fluctuations with meteorological effect implications—a comparative study among soft computing techniques. Comput Geosci 56:32–44
35. Singh KK, Pal M, Singh VP (2010) Estimation of mean annual flood in Indian catchments using back propagation neural network and M5 model tree. Water Resour Manag 24:2007–2019
36. VapnikV N (1995) The nature of statistical learning theory. Springer, New York
37. Vapnik VN (1998) Statistical learning theory. Wiley, New York
38. Wang W, Ding J (2003) Wavelet network model and its application to the prediction of hydrology. Nat Sci 1(1):67–71
39. Wang W, Van Gelder PHAJM, Vrijling JK, Ma J (2006) Forecasting daily stream flow using hybrid ANN models. J Hydrol 32:383–399
40. Wang W, Jin J, Li Y (2009) Prediction of inflow at three Gorges Dam in Yangtze River with wavelet network model. Water Resour Manag 23:2791–2803
41. Wei S, Song J, Khan NI (2012) Simulating and predicting river discharge time series using a wavelet-neural network hybrid modelling approach. Hydrol Process 26(2):281–296
42. Zhou HC, Peng Y, Liang GH (2008) The research of monthly discharge predictor corrector model based on wavelet decomposition. Water Resour Manage 22:217–227