Streamflow forecasting by combining neural networks and fuzzy models using advanced methods of input variable selection

A. B. Dariane and Sh. Azimi

ABSTRACT

In this paper the performance of extreme learning machine (ELM) training method of radial basis function artificial neural network (RBF-ANN) is evaluated using monthly hydrological data from Ajichai Basin. ELM is a newly introduced fast method and here we show a novel application of this method in monthly streamflow forecasting. ELM may not work well for a large number of input variables. Therefore, an input selection is applied to overcome this problem. The Nash–Sutcliffe efficiency (NSE) of ANN trained by backpropagation (BP) and ELM algorithm using initial input selection was found to be 0.66 and 0.72, respectively, for the test period. However, when wavelet transform, and then genetic algorithm (GA)-based input selection are applied, the test NSE increase to 0.76 and 0.86, respectively, for ANN-BP and ANN-ELM. Similarly, using singular spectral analysis (SSA) instead, the coefficients are found to be 0.88 and 0.90, respectively, for the test period. These results show the importance of input selection and superiority of ELM and SSA over BP and wavelet transform. Finally, a proposed multistep method shows an outstanding NSE value of 0.97, which is near perfect and well above the performance of the previous methods.

Key words | adaptive neuro-fuzzy inference system (ANFIS), extreme learning machine, input selection, neural network, singular spectral analysis, wavelet transform

A. B. Dariane
Sh. Azimi (corresponding author)
Department of Civil Engineering,
K.N. Toosi University of Technology,
Tehran,
Iran
E-mail: shazimi@mail.kntu.ac.ir

INTRODUCTION

Streamflow forecasting is a critical problem in hydrology and water management because of a wide range of variabilities in space and time. Many water resources system rule curves require monthly streamflow forecasts for their operation. So far, various methods have been proposed which are capable of forecasting streamflows with different accuracy levels under different conditions. Forecasting methods have considerably evolved from simple linear equations to very complicated methods during the last half century. Thus, a wide range of models, including stochastic (e.g., AR-autoregressive and ARMA-autoregressive moving average), conceptual (e.g., HEC-HMS and HBV) and physical based (e.g., SWAT) have been considered by researchers for streamflow forecasting. For instance, Adamowski et al. (2012)

indicated that, due to the complex relationship between rainfall–runoff variables and the lack of sufficient hydrological data in mountainous watersheds, data-driven models, such as artificial neural network (ANN), are more suitable than the process-based models in forecasting streamflows. In fact, there is no single unique method which could be used for all types of systems and basins. During this period, hydrologists have exploited the new techniques available with the progress of science, to build more reliable forecasting methods. Artificial intelligence methods have attracted great attention, due to their ability to address the nonlinear and complex relationship between input and output variables. Since then, ANNs have been widely applied by various researchers for streamflow forecasting (e.g., Zealand

et al. 1999; Wang *et al.* 2006; Islam 2010; Wei *et al.* 2012; Sudheer *et al.* 2013; Zounemat-Kermani 2014).

As any model, ANNs must be trained by the local data. Backpropagation (BP) is a standard method usually used for training neural networks (Rumelhart *et al.* 1986). In BP, least mean square error between the target and the network output is propagated backward to train the network. This is done by adjusting the network weights and biases which together form the model parameters with an objective to minimize the total error. Although BP works well on simple training problems, as the problem complexity grows the performance of the method falls (Dariane & Karami 2014). Meanwhile, the application of evolutionary methods has been suggested as an alternative to overcome some of these deficiencies (Montana & Davis 1989). However, evolutionary methods are generally time-consuming and may not be suitable for occasions where time is important (i.e., large problems). The method of extreme learning machine (ELM) was introduced by Huang *et al.* (2006) as a fast and simple way for training ANN models.

In the last couple of years, some researchers have reported the successful application of ELM in water resources problems. For example, Li & Cheng (2014) applied a combination of wavelet transform and neural network trained by extreme learning machine (WANN-ELM) to forecast one month ahead discharge. According to their results, ANN trained by ELM showed slightly better performance than the support vector machine (SVM), while WANN-ELM demonstrated the most accurate performance among the three models. Deo & Shahin (2015) compared the application of ANN trained by BP and ELM algorithms in predicting a monthly effective drought index. According to their results, the learning speed of ELM is 32 times faster than the BP algorithm. Wang *et al.* (2015) developed a hybrid model using ELM and seasonal auto-regression integrated moving average (SARIMA) to forecast wind speed. They applied ANN trained by ELM algorithm to simulate the wind speed. According to their results, the hybrid model proposed overtakes the single ones such as the BP neural network and the classic models (e.g., SARIMA) in accuracy of the predictions. Also, an online sequential ELM model was applied by Yadav *et al.* (2016) in a flood forecasting problem in Germany. According to their findings, for the considered case, the ELM-based model was more accurate than SVM, genetic programming (GP), and ANN trained by traditional methods.

According to the simple computational nature of ELM and the need for the definition of more hidden nodes than the gradient-based algorithms (e.g., BP), it would be inappropriate to use large number of inputs. Our review of the literature in streamflow forecasting fields showed that there is no application of ANN-ELM that uses an evolutionary algorithm to select effective features (i.e., input variables). However, input selection methods have been used in combination with other methods. For example, Asadi *et al.* (2013) applied the combination of data preprocessing and ANNs to predict basin runoff and indicated the effectiveness of the input selection approach in improving the results. Bowden *et al.* (2005) used two input determination approaches including partial mutual information, and the hybrid genetic algorithm (GA) and general regression neural network. They concluded that both input selection approaches provide more accurate results for river salinity forecasting. Moreover, Dariane & Azimi (2016) developed a combination of wavelet neuro-fuzzy models based on genetic input selection algorithm to forecast streamflow. They demonstrated that input selection approaches are more useful for large basins where effective input variables' determination is manually a time-consuming task. Their results indicated that by using genetic input selection algorithm and wavelet transform, considerable improvements would be achieved.

Suitable data preprocessing can help a predictor model to forecast the main features of a time series (main oscillations) more accurately (Tiwari & Chatterjee 2010; Wei *et al.* 2012). Singular spectral analysis (SSA) and wavelet transform are two types of preprocessing techniques. SSA is based on the decomposition of the original series into a sum of series including the trend, oscillation, and noise components. This is provided by the reconstruction of the original series (Golyandina *et al.* 2001). SSA has been employed in some hydrological problems (e.g., Sivapragasam *et al.* 2001; Wu & Chau 2011). Sivapragasam *et al.* (2001) applied SVM to model rainfall–runoff process by using SSA preprocessing technique. They demonstrated that their combined model has better performance than the single SVM without any data preprocessing. Wu & Chau (2011) used a conjunction of ANNs and SSA and indicated that using SSA provided a significant improvement in model accuracy. Wavelet transform is used more widely than the SSA to develop hydrologic forecasting models

(e.g., Kim & Valdes 2008; Rajaee *et al.* 2011; Wei *et al.* 2013; Santos & Silva 2014). Similar to SSA, wavelets help models to perform more accurately. For instance, Santos & Silva (2014) compared the performance of single ANN and wavelet-ANN for 1, 3, 5, and 7 day ahead streamflow forecasting. The wavelet-ANN models performed better than ANN. This superior performance became more evident for longer lead times (i.e., 5 and 7 day ahead forecasts). In addition, they employed a trial and error process for selecting appropriate input variables. Their results showed that using the approximations of the first five decomposition levels as inputs produced the best results and incorporation of detail signals did not directly play an important role in improving the results. Also, Tiwari & Chatterjee (2010) applied a wavelet-ANN hybrid model to forecast daily streamflow. They used a correlation analysis to select and then recombine significant wavelet components. These components were applied as the new input variables to the ANN model. Their proposed hybrid model was able to catch the peak flows more accurately than the simple ANN model.

In fact, streamflow forecasting helps decision-makers to adjust their actions according to the state of the forecasted streamflow. With one month lead time, reservoir operators are able to better decide about their releases and storages. Otherwise, one must use the long-term mean streamflow in the absence of any forecast that would yield considerable errors. On the other hand, using more lead time (e.g., two or more months) provides extra flexibility and more time for system operation adjustment but also would have higher risk. Overall, a monthly horizon with zero lead time (forecasting the flow of next month) is common in most reservoir operation systems. However, both lead time and forecast horizon may vary from one system to another. Herein, to forecast monthly streamflow, a proposed method is investigated where after the initial variable selection using GA, the variables are decomposed using both the wavelet transform and the SSA. Then, GA input selection is used once again to pick out the best sub-signals from the pool of all decomposed signals. In the next step, these sub-signals are set as the input variables of RBF-ANN trained by ELM (ANN-ELM) and ANN trained by BP (ANN-BP). Finally, a hybrid model is defined where the outputs of the aforementioned ANN-ELM and ANN-BP models are used as the inputs of an adaptive neuro-fuzzy inference system (ANFIS) model.

METHODOLOGY

Following the authors' previous study (Dariane & Azimi 2016), in this paper, the capabilities of ANNs have been examined in Ajichai streamflow forecasting. Due to some challenges and problems in application of ANNs (mentioned later), a proposed hybrid model which benefits from two signal decomposition techniques, two network training methods, and using twice an input selection algorithm was applied. Here, a brief presentation of applied methods, including ELM algorithm, SSA method, wavelet transform, and genetic input selection is given. Also, a short description of the proposed method is given. Readers can obtain detailed information about neuro-fuzzy and gradient-based training algorithms from Jang (1993) and Hagan *et al.* (1996).

ELM training algorithm

The ELM developed by Huang *et al.* (2006) is a novel and fast converging algorithm for training a single hidden-layer, feed forward neural network. For n input variables, N hidden nodes, and M training cases, the model is presented as follows:

$$\sum_{i=1}^{N} \beta_i g(w_i x_j + b_i) = o_j, j = 1, \ldots, M, N \le M \tag{1}$$

where x_j, o_j, w_i, b_i, and β_i are training input datasets, training output datasets, random input weights, random input biases, and weights of output layer, respectively. $g(x)$ is an activation function. Here, the Gauss RBF kernel function (Equation (2)) is applied as the activation function of the ANN-ELM model:

$$G(x, xi) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \tag{2}$$

where x_i and σ are constant parameters with values found to be 0.05 and 1.5 using a sensitivity analysis.

If $N = M$, the model can estimate the training target datasets with zero error:

$$\sum_{j=1}^{M} \|o_j - y_j\| = 0, \sum_{i=1}^{N} \beta_i g(w_i x_j + b_i) = y_j \tag{3}$$

where y is training target dataset. Also, these functions can be written as (Huang *et al.* 2006):

$$H\beta = Y, H = \begin{bmatrix} g(w_1*x_1 + b_1) & . & . & g(w_N*x_1 + b_N) \\ . & . & & . \\ & & . & . \\ g(w_1*x_M + b_1) & . & . & g(w_N*x_M + b_N) \end{bmatrix}_{M \times N}$$

(4)

where H is the output matrix of the hidden layer.

However, in most cases, the number of hidden nodes is much less than the number of distinct training samples, so the training error is not equal to zero:

$$\sum_{j=1}^{M} \|o_j - y_j\| < \varepsilon$$

(5)

Thus, the weights vector, β, can be obtained by solving the following equation:

$$Z = \min\|H\beta - Y\|$$

(6)

The solution of Equation (6) is $\hat{\beta} = H^*.Y$, where H^* is the Moore–Penrose generalized inverse of the hidden layer matrix H and Y is training target dataset.

The ELM algorithm can be summarized as follows:

1. Determine input weights and biases randomly.
2. Calculate the hidden layer output matrix H (from Equation (4)).
3. Calculate the output weight $\beta = H^*Y$.

Singular spectral analysis

The main purpose of SSA is to decompose the time series into some sub-series including appropriate details. This is done by reconstruction of the time series and application of them instead of the original time series, as input variables of predictor models. Here, the employed SSA follows the methodology of Golyandina *et al.* (2001). Data preprocessing consists of decomposition and construction stages. In the first decomposition step, the embedding procedure transfers the time series $x = (x_0, \ldots, x_{N-1})$ of length N into a sequence of L-dimensional vectors $x_i = (x_{i-1}, \ldots, x_{i+L-2})^T$, $(I = 1, \ldots, K = N-L+1)$:

$$X = \begin{bmatrix} x_0 & . & x_{K-1} \\ . & . & . \\ x_{L-1} & . & x_{N-1} \end{bmatrix}$$

(7)

In the next step of the decomposition stage, XX^T is calculated and its eigen triple (s_i, u_i, v_i) is determined by the singular value decomposition (SVD). Where s_i is i^{th} singular value of X, u_i and v_i are i^{th} left and right eigen vectors of XX^T, respectively. The trajectory matrix X can be as follows:

$$X = X_1 + X_2 + \ldots\ldots + X_L$$

(8)

$$X_i = s_i u_i v_i^T$$

The next two steps (grouping and diagonal averaging) provide reconstruction stages. In the grouping procedure, the indices $j = 1, \ldots, L$ are partitioned into M disjoint groups $I_1, I_2, \ldots, I_M$, so that the elementary matrices, X_i, in Equation (8) are split into M groups. Each group consists of indices as $I = \{i_1, \ldots, i_p\}$. The resultant matrix X_I is defined as: $X_{In} = X_{i1} + \ldots\ldots + X_{ip}$ and so the matrix X is presented as the sum of M resultant matrices:

$$X = X_{I1} + X_{I2} + \ldots\ldots + X_{IM}$$

(9)

In the last step of SSA, for diagonal averaging, each resultant matrix transforms into a new time series of length N. Let X be a $(L \times k)$ matrix with elements x_{ij}. Make $L^* = \min (L, k)$, $K^* = \max (L, k)$. Let us define $x_{ij}^* = x_{ij}$, if $L < K$; otherwise $x_{ij}^* = x_{ji}$. Diagonal averaging transfers matrix X to a series $g_0, \ldots, g_{N-1}$ using the following formula:

$$g_k = \begin{cases} \dfrac{1}{k+1}\displaystyle\sum_{m=1}^{k+1} x_{m,k-m+2}^* & 0 \le k < L^* - 1 \\ \dfrac{1}{L^*}\displaystyle\sum_{m=1}^{L^*} x_{m,k-m+2}^* & L^* - 1 \le k < K^* - 1 \\ \dfrac{1}{N-K}\displaystyle\sum_{m=k-k^*+2}^{N-k+1} x_{m,k-m+2}^* & K^* - 1 \le k < N \end{cases}$$

(10)

Equation (10) corresponds to averaging the elements along diagonals $i + j = k + 2$. This diagonal averaging is applied to a resultant matrix X_{In}. Thus, the original time series is decomposed into the sum of M series, and can be derived as follows:

$$\hat{x} = x_1 + \ldots\ldots + x_M$$

(11)

Wavelet transform

Wavelet transform, which is an important technique in the field of signal processing, has attracted a great deal of

attention since its introduction in the early 1980s. There are two types of wavelet transform: continuous and discrete. The first one deals with continuous functions and the second one is applied for discrete functions or time series. Most of the hydrological time series are measured in discrete time steps. Therefore, DWT would be a suitable method for decomposition and reconstruction of these series. This way, the signal is divided into two parts including 'approximation' and 'details' and the original signal is broken down into lower resolution components. These components explain a better behavior and reveal more information about the process than the original time series. Therefore, they can help the forecasting models to predict with a higher accuracy (Remesan *et al.* 2009; Adamowski & Chan 2011).

The time-scale wavelet transform of continuous time series, $x(t)$, is defined as (Mallat 1998):

$$C_\varphi^x = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\varphi^* \left(\frac{t - \tau}{a} \right) dt \qquad (12)$$

where * denotes conjugate complex function. $\varphi(t)$ is wavelet function or mother wavelet, a is scale or frequency factor also called dilation factor, and τ is the time factor. The term 'scale' refers to extending or compressing the wavelet. Using small scale causes the wavelet to be compressed and in the case of large scale the wavelet is extended. Large scale values are not able to show the details whereas small scales are applied to reveal more details.

Discrete wavelets have the following general form (Grossman & Morlet 1984):

$$\psi_{m,n}(t) = a_0^{-m/2} \psi \left(\frac{t - n\tau_0 a_0^m}{a_0^m} \right) \qquad (13)$$

where m and n are integers that control, respectively, the wavelet dilation and translation; a_0 is a specified fixed dilation step with a value greater than 1; and τ_0 is the location parameter which must be greater than zero. Scales and positions are usually based on powers of two (dyadic scales and positions), making it more efficient for practical cases (Mallat 1998). The most common (and simplest) choice for the parameters a_0 and τ_0 are 2 and 1, respectively. By this way, for a discrete time series x_i, which occurs at different time t, the DWT is defined as (Mallat 1998):

$$W_{m,n} = 2^{-m/2} \sum_{i=0}^{N-1} x_i \psi(2^{-m}i - n) \qquad (14)$$

where $W_{m,n}$ is the wavelet coefficient for the discrete wavelet of scale $a = 2^m$ and location $\tau = 2^m n$.

GA for input selection

Input variable selection approaches are generally partitioned into three main groups of wrapper, embedded, and filter methods. The GA input selection algorithm which is applied here can be considered as a wrapper input selection method. In order to obtain more comprehensive information about these classes of input selection methods, readers are referred to Kohavi & John (1997), Blum & Langley (1997) and Guyon & Elisseeff (2003).

Herein, the whole data were divided into train (70%), validation (15%), and test (15%). For input selection, the objective function of GA was defined as $0.75*\text{MSE}_{\text{TRAIN}} + 0.25*\text{MSE}_{\text{VALIDATION}}$ which uses both training and validation periods to evaluate the function. Using a weighted combination of fitness function helps in selecting better input features. Thus, for training and validation MSE, the weights were proposed equal to 0.75 and 0.25, respectively, which were determined through a trial and error process. It should be mentioned that after selecting the input variables of the ANN model, the train and validation data were used as usual to find the weights and biases of the ANN model after which they are evaluated by the test period data.

The main GA operators are applied and the new generation is produced. Then stopping conditions are checked and the decision is made to continue the loop or to end the process. The details of input selection algorithm are available in our last study (Dariane & Azimi 2016).

The proposed model

In this study, a hybrid forecasting model has been developed based on data preprocessing, input selection, and data-driven methods. In this model, a GA input selection

method is applied in the initial stage to choose the most proper variables from the available ones. Precipitation at Sarab, Esbaghran and Gushchi stations, along with temperature at Mirkuh plus Vanyar streamflow (all lagged by one time period) are selected as the final inputs. Then, the selected input variables are decomposed using wavelet transform and SSA independently. Next, the decomposed series are combined together to form a selection pool from which a second GA input selection algorithm is applied to choose the best decomposed sets. These sets are then used as the optimum input variables for ANN-BP and ANN-ELM, independently. Finally, a hybrid model is used, where the outputs of ANN-BP and ANN-ELM are set as the input for an ANFIS model, to generate the forecasted flow (Figure 1).

CASE STUDY

The data used for developing the models belong to the Ajichai basin. Ajichai is a sub-basin in the larger Urmia Lake Basin. Urmia Lake Basin which is mainly located within two Azerbaijan provinces in northwest Iran has an area of 51,876 km^2 (Figure 2). A small portion of the basin is located in Kurdistan province. Ajichai, with an area of 7,675 km^2 above Vanyar station, is considered a relatively large basin in the northeastern part of the lake. A monthly time step with zero lead time was used in developing monthly streamflow forecasting models in this research. Nine precipitation stations considered for this basin include Bostanabad, Ghurigul, Ghushchi, Nahand, Saeedabad, Sarab, Esbaghran, Sohzab, and Vanyar (Figure 2). In addition, temperature
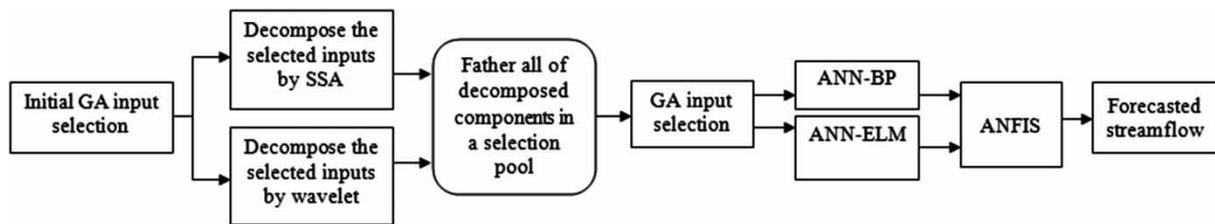

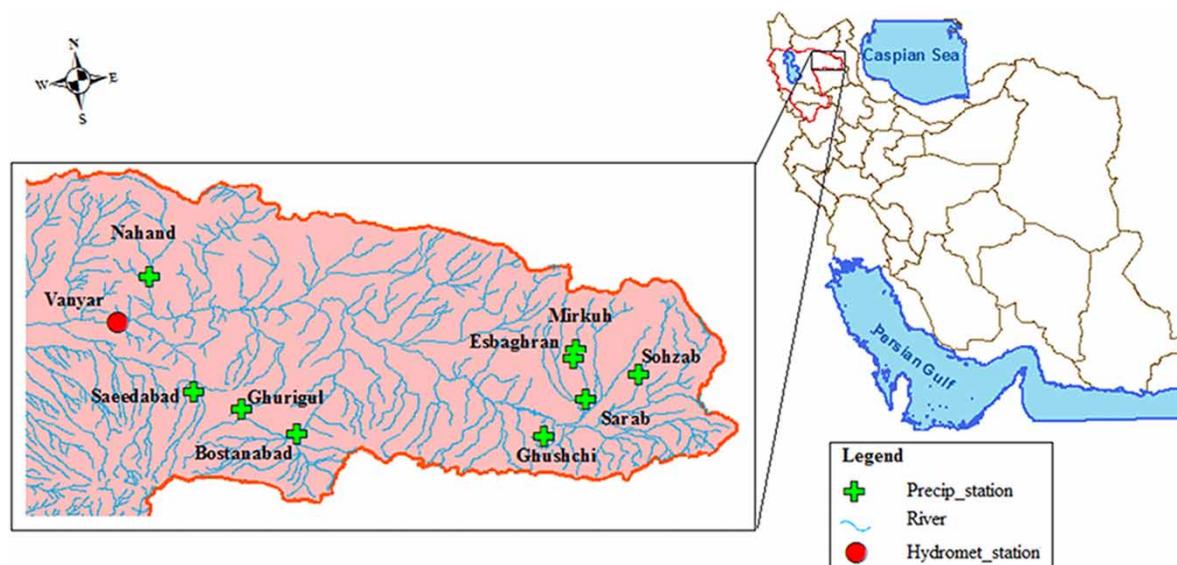**Figure 1** | Configuration of proposed model.

Figure 2 | Ajichai sub-basins in Urmia Lake Basin, Iran.

data from Sohzab, Mirkuh, and Ghurigul stations are used for developing the monthly streamflow forecasting models. The statistical properties of all available time series are shown in Table 1. Figure 3 presents the hydrograph of Ajichai River at Vanyar station. As can be seen from Figure 3, there are high fluctuations in streamflow data. In some periods, mainly in summer months, the river runs dry.

EXPERIMENTAL SETUP

In this section, the experimental setup is presented. As was mention in the section 'Case study', there are 13 potential input variables of different kinds in this basin. In order to

Table 1 │ Statistical properties of available data

Data	Station	Max.	Min.	Mean	St. dev.
Rainfall (mm)	Bostanabad	124.9	0	21.4	23.7
	Ghurigul	146.5	0	25.2	25.6
	Ghushchi	133.5	0	20.8	23.9
	Nahand	126.5	0	22.3	22.2
	Saeedabad	294	0	33.1	39.6
	Sarab	139.4	0	20.5	20.2
	Esbaghran	153	0	23.9	24.7
	Sohzab	166	0	25.4	25.7
	Vanyar	136.6	0	17.5	21.2
	Mirkuh	–	–	–	–
Temperature (°C)	Ghurigul	22.8	−16.8	6.9	9.6
	Sohzab	22.5	−17	8.1	8.9
	Mirkuh	24.9	−12.8	8.9	8.7
Runoff (m^3/s)	Vanyar	178.3	0	12.6	20.7

decrease the number of input variables, an initial GA-based input selection is applied. Thus, four out of thirteen input variables, including precipitation at Sarab, Esbaghran, and Gushchistations, along with temperature at Mirkuh (all lagged by one time period) are selected. Also, the lagged streamflow of Vanyar was directly added to the selected input variables. For all available data, 41 years (1966–2006, inclusive) of monthly data are available, of which 70% (344 datasets) are used for training the applied data-driven models and the remaining 30% (147 datasets) are applied for validation and testing them. It is worthwhile to mention that the test and validation periods were selected from the middle parts of the data duration based on some initial assessments. Root mean square error (RMSE), coefficient of determination (R^2), and Nash–Sutcliffe efficiency (NSE) index, as three performance measures, were selected to evaluate the models' performance. These parameters are commonly applicable for estimating the performance of data-driven models (Yadav et al. 2016).

RESULTS AND DISCUSSION

As was mentioned earlier, ANN-ELM suffers from weak performance when there are a large number of input variables. Using wavelet transform and SSA to decompose the input data into further input series would generate more input variables which could aggravate the aforementioned problem. In other words, by application of these transforms the final performance would become worse than they were
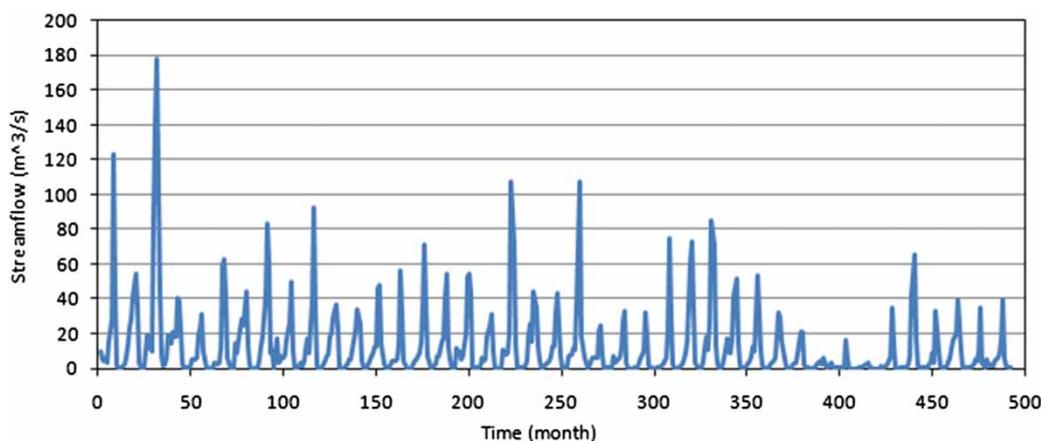

Figure 3 │ Hydrograph of Ajichai River at Vanyar station.

before the decomposition, thus undermining the decomposition process. This problem could be overcome by using an input selection method where only suitable input variables are chosen from among many data series. In the following, the performance of decomposition methods as well as the ELM training method along with the input selection algorithm is evaluated. Finally, the results of the proposed method are presented and discussed.

ELM versus BP

In this part of the article, the performance of ANN trained by ELM is evaluated through comparison with the one trained by BP algorithm. For this purpose, a neural network model is defined for forecasting the monthly river discharges at Vanyar station, Ajichai Basin. There are potentially 13 input variables of different kinds in this basin. These include monthly precipitation, temperature, and streamflow, all lagged by one time interval.

In order to reduce the number of variables, a GA-based input selection algorithm is applied. Table 2 shows

Table 2 │ Parameters of GA-based selection algorithm

Number of generation	500
Size of population	20
Rate of crossover	0.75
Rate of mutation	0.08

configuration parameters of the GA selection method. All these parameters have been obtained through a trial and error process. It should be noted that the model training, verification, and testing is carried out using 70, 15, and 15% of available data, respectively. During the input selection process, 70% of monthly data were used for training the model and selecting suitable input variables. Then, these selected variables are verified using another 15% of available data which do not participate in the ANN training process.

Clearly, the input selection methods would probably eliminate some of the appropriate inputs in favor of the lagged streamflow variable due to the high cross correlation between the streamflow and most of those variables. Therefore, to avoid this discrepancy, the lagged streamflow was excluded from the input selection test and was directly added to the selected variables' collection. The application of GA input selection algorithm resulted in choosing four out of thirteen input variables including precipitation at Sarab, Esbaghran, and Gushchistations, along with temperature at Mirkuh, all lagged by one time period.

In the next step, these variables plus the lagged streamflow at Vanyar were applied as the inputs to the ANN-ELM and ANN-BP models. Figure 4 shows a comparison of the results during the test period. As can be seen from Table 3, the test NSE indices as well as other criteria indicate the superiority of ANN-ELM over ANN-BP. Regardless of the fact that the ANN-ELM uses many more neurons in the hidden layer (20 versus 5 used by ANN-BP), it trains the network ten times faster than the ANN-BP model.
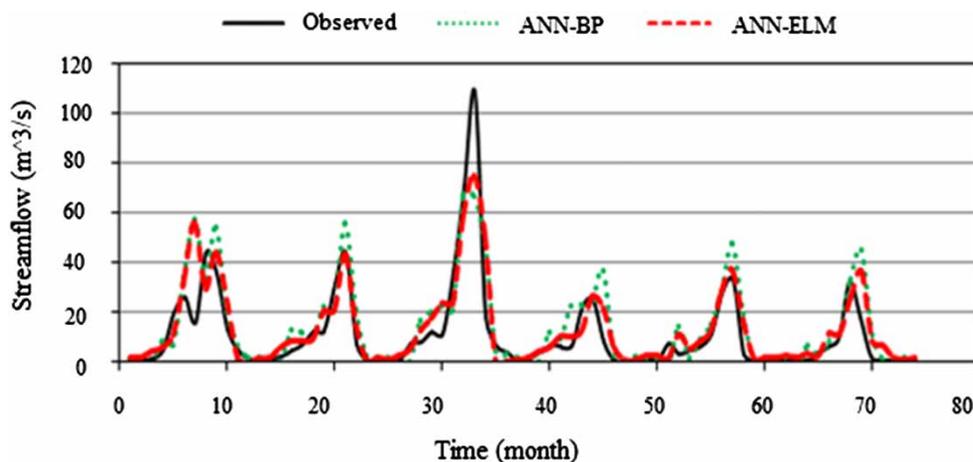

Figure 4 │ ANN-ELM and ANN-BP model output during the test period for Ajichai.

Table 3 | Comparison of models performance during train and test period

Model	Train			Test		
	R^2	NSE	RMSE	R^2	NSE	RMSE
ANN-BP	0.67	0.67	0.067	0.66	0.60	0.067
ANN-ELM	0.73	0.73	0.051	0.72	0.71	0.053

Thus, the speed is a great advantage of the ELM over BP, bearing in mind that the performance of ANN-ELM is also much better than the ANN-BP.

As can be seen from Figure 4, although the ELM method has been able to improve the performance of the ANN model, there are still instances (mainly in peak discharges) where more accuracy is needed. For instance, none of the methods were able to catch the main peak flow in the third year while errors in some other peaks are also significant. In the next step, data preprocessing methods, including wavelet transform and SSA methods are investigated for further enhancement.

Data preprocessing approaches

It is clear that redundant and irrelevant variables lead to a poor generalization performance, add error and noise to the model, and prevent correct learning process (May et al. 2011). Thus, one of the main problems which might arise during the application of a data-driven model is to detect the appropriate input variables. In general, according to the law of parsimony, the number of model inputs should be as few as possible. This is more emphasized when ELM is applied which is highly sensitive to the number of input variables. In comparison to ANN-BP, the ANN-ELM method suffers from weak performance when there is a large number of input variables. ELM-based ANN requires more hidden neurons than the BP-based ANN to train the network. This leads to poor performance of the method in large networks as compared to the BP-based ANN. Therefore, after decomposing the initially selected input variables by the wavelet and SSA methods, a GA input selection algorithm is applied to extract more important subseries and limit their numbers and thus the corresponding hidden neurons in order to avoid large networks. More details are presented in the following sections.

Using wavelet transform

Wavelet transforms are used to achieve more reliable and accurate outputs. In general, a suitable level of decomposition is selected with respect to the nature of time series. Usually, in order to select the best mother wavelet, the apparent similarity between mother wavelet and the time series should be considered; but some researchers use their own experience (e.g., Wei et al. 2013) and some others apply sensitivity analysis to choose a suitable mother wavelet (Nourani et al. 2011). In our application, we also used sensitivity analysis where db4 mother wavelet with two decomposition levels was determined suitable for the Ajichai time series. Therefore, one approximate sub-signal (a2) of the original signal and two detailed sub-signals (d1, d2) are generated for further application.

In the next step, the GA input selection is applied and nine appropriate inputs are selected from 15 generated subsignals (Table 4). These inputs are then used by ANN-BP (using six neurons) and ANN-ELM (using 60 neurons) to forecast monthly streamflow at Vanyar station. As can be seen from Table 5, by applying wavelet transform all evaluation parameters are improved, especially during the test period. Meanwhile, the network trained by the ELM outperforms the one by BP, which was also observed earlier when no data preprocessing method was applied (see Table 3). According to Tables 3 and 5, the result of the network trained by ELM has not considerably improved by only using the wavelet transform. This unexpected result is caused by the large number of input variables of WANN-ELM before applying input selection. However, by using GA input selection, the NSE index of WANN-ELM increases substantially from 0.71 to 0.85. This shows the importance of using an input selection method in ELM-based networks.

Table 4 | Selected decomposed sub series by GA

Variables	Selected components
$P_{t-1}^{Gushchi}$	d2
P_{t-1}^{sarab}	d2
$P_{t-1}^{Esbaghran}$	d1
T_{t-1}^{Mirkuh}	a2, d2, d1
Q_{t-1}^{vanyar}	a2, d2, d1

Table 5 | Results of WANN-ELM and WANN-BP with/without GA input selection

Model	Number of inputs	Train			Test		
		R²	NSE	RMSE	R²	NSE	RMSE
WANN-BP (no Inp. Sel.)	15	0.82	0.82	0.049	0.73	0.71	0.052
WANN-BP (with Inp. Sel.)	9	0.85	0.85	0.045	0.76	0.75	0.047
WANN-ELM (no Inp. Sel.)	15	0.93	0.93	0.031	0.73	0.72	0.051
WANN-ELM (with Inp. Sel.)	9	0.93	0.93	0.030	0.86	0.85	0.038

As can be seen from Figure 5, although there is some improvement when compared to the previous results in Figure 4, there is still a need for further improvement, especially for peak flow forecasts. Also, it shows that WANN-ELM using GA input selection has more accurate forecasts than the WANN-BP-GA. Clearly, WANN-ELM-GA is more successful in peak flows' prediction as compared to WANN-BP-GA. Consequently, GA input selection provides more improvement for the ELM-based network than the one trained by the BP.

Using SSA

A similar approach is used to develop models by SSA. In this regard, signals are decomposed into three levels and suitable input variables are selected by the GA input selection model. The results are presented in Table 6. As before, according to Table 6, the superiority of the ELM over BP is revealed. In addition, by applying the GA input selection model, the results of ELM- and BP-based networks show an increase of 17 and 9% in NSE during the test period, respectively. A similar trend can

be observed in R^2 and RMSE indices. Therefore, the GA input selection has more positive impact on the ELM-based network than the one trained by BP. Also, by trial and error process, the optimum number of neurons for SANN-ELM and SANN-BP, using the input selection approach, is obtained as 65 and 6, respectively. Moreover, a comparison of results in Tables 5 and 6 reveals that the application of SSA decomposition method has been able to improve the performance of both ELM- and BP-based models as compared to the wavelet transform. In other words, SSA extracts more appropriate details for the Ajichai time series compared to the wavelet transform.

So far, the findings are in support of Figure 6. As can be seen from this figure, data preprocessing by using the SSA leads to more accurate predictions compared to the wavelet transform. Figure 6 shows that the SSA method has been able to forecast the main peak with much higher accuracy. Nevertheless, learning from these experiences, we propose a multistep model that is able to forecast the streamflow with higher accuracy, as described in the following section.
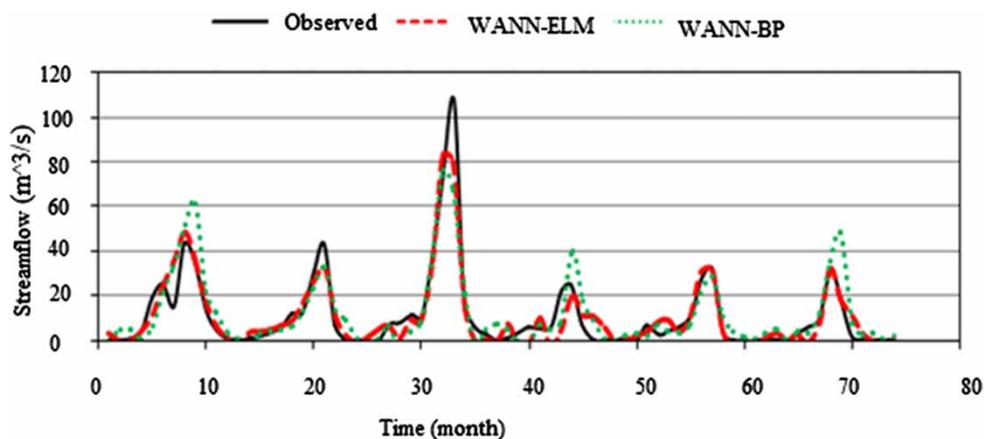

Figure 5 | ANN-ELM and ANN-BP model output using GA input selection and wavelet transform during the test period for Ajichai.

Table 6 │ Results of SANN-ELM and SANN-BP with/without GA input selection

Model	Number of input var.	Train			Test		
		R²	NSE	RMSE	R²	NSE	RMSE
SANN-BP (no Inp. Sel.)	15	0.88	0.88	0.039	0.77	0.78	0.044
SANN-BP (with Inp. Sel.)	10	0.89	0.89	0.032	0.88	0.87	0.031
SANN-ELM (no Inp. Sel.)	15	0.94	0.94	0.027	0.74	0.73	0.051
SANN-ELM (with Inp. Sel.)	10	0.95	0.95	0.022	0.90	0.90	0.031
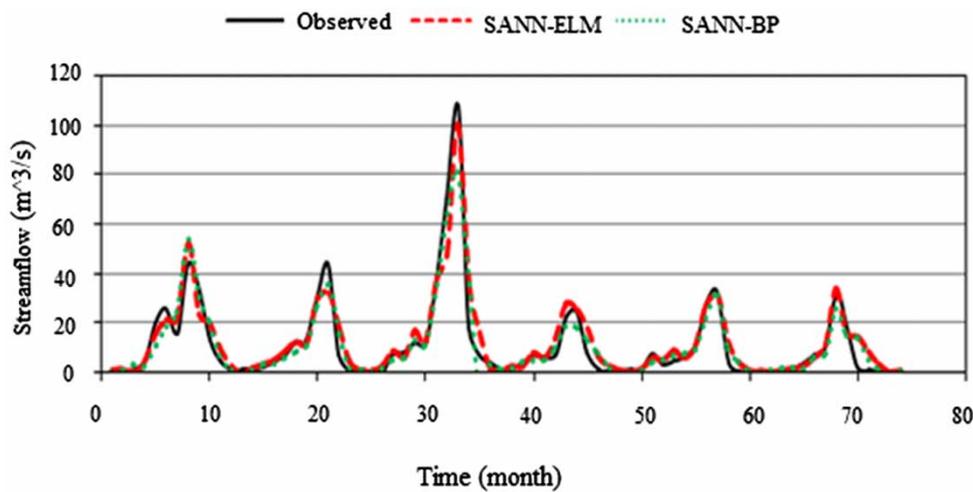
Figure 6 │ Comparison of models' output using GA input selection and SSA during the test period.

We can conclude from the aforementioned experiments that: (a) preprocessing methods, i.e., wavelet and SSA both have positive impact on model results, (b) ELM is a better and faster method for training neural networks than the commonly used BP method, and (c) GA-based input selection helps to improve model performance, especially those trained by ELM. There is also a great amount of research supporting the idea of implementing a hybrid approach by combining data-driven models. In addition, the literature shows that in many cases ANFIS performs better than simple ANN models. Therefore, putting these together we propose a multistep model in order to improve further the accuracy of streamflow forecasts. In other words, the proposed model benefits from preprocessing methods, input selection procedure, and using both ANN and ANFIS models in a hybrid configuration, as illustrated by Figure 1.

As was mentioned in the section 'The proposed model', the two-step GA input selection process (in the proposed model) was applied to select the final appropriate input variables (Table 7).

According to Table 7, the trend sub-signal in decomposition level 2 (i.e., L2) using SSA is analogous to the approximate sub-signal in level 2 (a2) using wavelet transform. Similarly, the noise sub-series in level 2 using SSA is analogous to details of the sub-signal in level 2 using wavelet.

Table 8 shows the results of the proposed model. As can be seen from this table and also Figure 7, the results

Table 7 │ Selected decomposed sub-series by two-step GA input selection

Variables	Selected wavelet components	Selected SSA components
$P_{t-1}^{Gushchi}$	a2	–
P_{t-1}^{sarab}	–	–
$P_{t-1}^{Esbaghran}$	–	–
T_{t-1}^{Mirkuh}	d2, a2	trend (L2), noise (L2)
Q_{t-1}^{vanyar}	d2, a2	trend (L2), noise (L2)

Table 8 | Results of the proposed model

Model	Number of input var.	Train			Test		
		R^2	NSE	RMSE	R^2	NSE	RMSE
Proposed model	9	0.97	0.97	0.020	0.97	0.97	0.021
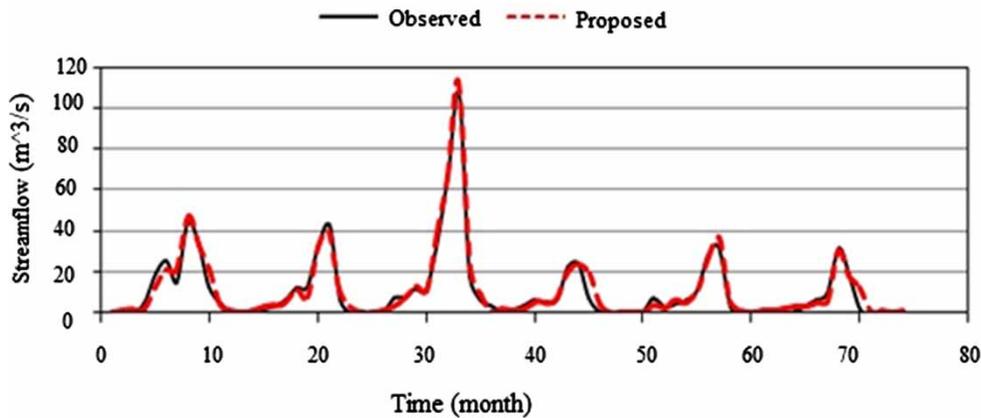
Figure 7 | The proposed model output during the test period for Ajichai.

obtained through the proposed method are near perfect during the test period. The model seems to be very robust and reliable. Achieving a coefficient of determination and NSE equal to 0.97 and RMSE equal to 0.021 as the average of ten independent runs shows how well the model is performing during the independent test period. In addition, a look at Figure 7 reveals that the proposed method has been able to catch almost all variations of the streamflow data in the forecast model with one month lead time. The value of these results is more appreciated by noting that the measured data in this part of the world are, to a large extent, inaccurate as is the case in many other countries as well. Therefore, the proposed model could be used as a framework for other parts of the country as well as other similar regions.

CONCLUSIONS

Monthly streamflow forecasting helps reservoir operators make better decisions about releases and storages. Using data from Ajichai Basin above Vanyar discharge station, the impact of data preprocessing methods, input selection algorithm, and hybridization in data-driven models were evaluated. Of the two preprocessing methods, it was shown that SSA, in general, outperforms the more commonly used wavelet transform method. Also, in our application, a comparison of commonly used BP network training method with recently introduced ELM method indicated the superiority of ELM both in accuracy and in speed. It was next shown that the deficiency of ELM with regards to a large number of variables could be easily overcome by applying a GA-based input selection model. Finally, a multistep data-driven model was proposed that uses both wavelet and SSA, as well as BP and ELM training algorithm along with the GA input selection and the capabilities of ANFIS model in a hybrid framework to yield near perfect outputs with substantial improvements over the previous method. Results indicated that streamflow forecasting could improve the system performance with NSE values well above zero and near one, where water resources system operators would have sufficient time to make proper decisions based on reliable forecasts.

REFERENCES

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

Adamowski, J., Chan, H. F., Prasher, S. O. & Sharda, V. N. 2012 Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data. *J. Hydroinform.* **14** (3), 731–744.

Asadi, S., Shahrabi, J. & Abbaszadeh, P. 2013 A new hybrid artificial neural networks for rainfall–runoff process modeling. *Neurocomputing* **121**, 470–480.

Blum, A. & Langley, P. 1997 Selection of relevant features and examples in machine learning. *Artif. Intell.* **97**, 245–271.

Bowden, G. J., Maier, H. R. & Dandy, G. C. 2005 Input determination for neural network models in water resources applications, part 2. Case study: forecasting salinity in a river. *J. Hydrol.* **301**, 93–107.

Dariane, A. B. & Azimi, Sh. 2016 Forecasting streamflow by combination of genetic input selection algorithm and wavelet transform using ANFIS model. *Hydrolog. Sci. J.* **61**, 585–600.

Dariane, A. B. & Karami, F. 2014 Deriving hedging rules of multi-reservoir system by online evolving neural networks. *Water Resour. Manage.* **28**, 3651–3665.

Deo, R. C. & Shahin, M. 2015 Application of the extreme learning machine algorithm for the prediction of monthly effective drought index in eastern Australia. *Atmos. Res.* **153**, 512–525.

Golyandina, N., Nekrutkin, V. & Zhigljavsky, A. 2001 *Analysis of Time Series Structure: SSA and Related Techniques.* Chapman and Hall/CRC, New York, USA.

Grossman, A. & Morlet, J. 1984 Decompositions of hardy functions into square integrable wavelets of constant shape. *SIAM J. Mathematical Analysis* **15**, 723–736.

Guyon, I. & Elisseeff, A. 2003 An introduction to variable and feature selection. *J. Machine Learning Res.* **3**, 1157–1182.

Hagan, M. T., Demuth, H. B. & Beale, M. H. 1996 *Neural Network Design.* PWS Publishing, Boston, MA, USA.

Huang, G. B., Zhu, Q. Y. & Siew, C. K. 2006 Extreme learning machine: theory and applications. *Neurocomputing* **70**, 489–501.

Islam, A. S. 2010 Improving flood forecasting in Bangladesh using an artificial neural network. *J. Hydroinform.* **12**, 351–364.

Jang, J. S. R. 1993 ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man. Cybern.* **23**, 665–685.

Kim, T. W. & Valdes, J. B. 2008 Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. *J. Hydrol. Eng.* **8**, 319–328.

Kohavi, R. & John, G. 1997 Wrappers for feature selection. *Artif. Intell.* **97**, 273–324.

Li, B. & Cheng, C. 2014 Monthly discharge forecasting using wavelet neural networks with extreme learning machine. *Science China Technological Sciences* **57**, 2441–2452.

Mallat, S. G. 1998 *A Wavelet Tour of Signal Processing*, 2nd edn. Academic Press, San Diego, CA, USA.

May, R., Dandy, G. & Maier, H. 2011 Review of input variable selection methods for artificial neural networks. In: *Artificial Neural Networks – Methodological Advances and Biomedical Applications* InTech, Rijeka, Croatia, pp. 19–44.

Montana, D. J. & Davis, L. 1989 Training feedforward neural networks using genetic algorithms. In: *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, Morgan Kaufmann, San Mateo, CA, pp. 762–767.

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

Rajaee, T., Nourani, V. & Zounemat-Kermani, M. 2011 River suspended sediment load prediction: application of ANN and wavelet conjunction model. *J. Hydrol. Eng.* **16**, 613–627.

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

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. 1986 Learning representations by back-propagating errors. *Nature* **323**, 533–536.

Santos, C. A. G. & Silva, G. B. L. 2014 Daily streamflow forecasting using a wavelet transform and artificial neural network hybrid models. *Hydrol. Sci. J.* **59**, 312–324.

Sivapragasam, C., Liong, S. Y. & Pasha, M. F. K. 2001 Rainfall and runoff forecasting with SSA-SVM approach. *J. Hydroinform.* **3**, 141–152.

Sudheer, C. H., Shirvastava, N. A., Panigrahi, B. K. & Mathur, S. H. 2013 Streamflow forecasting by SVM with quantum behaved particle swarm optimization. *Neurocomputing* **101**, 18–23.

Tiwari, M. K. & Chatterjee, C. 2010 A new wavelet-bootstrap ANN hybrid model for daily discharge forecasting. *J. Hydroinform.* **13** (3), 500–519.

Wang, W., van Gelder, P. H. A. J. M., Vrijling, J. K. & Ma, J. 2006 Forecasting daily streamflow using hybrid ANN models. *J. Hydrol.* **324**, 383–399.

Wang, J., Hi, J., Ma, K. & Zhang, Y. 2015 A self-adaptive hybrid approach for wind speed forecasting. *Renewable Energ.* **78**, 374–385.

Wei, S. K., Zuo, D. P. & Song, J. X. 2012 Improving prediction accuracy of river discharge time series using a wavelet-NAR artificial neural network. *J. Hydroinform.* **14** (4), 974–991.

Wei, S., Yang, H., Song, J., Abbaspour, K. & Xu, Z. 2013 A wavelet-neural network hybrid modelling approach for estimating and predicting river monthly flows. *Hydrol. Sci. J.* **58**, 374–389.

Wu, C. L. & Chau, K. W. 2011 Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. *J. Hydrol.* **399**, 394–409.

Yadav, B., Ch, S., Mathur, S. & Adamowski, J. 2016 Discharge forecasting using an online sequential extreme learning machine (OS-ELM) model: a case study in Neckar River, Germany. *Measurement* **92**, 433–445.

Zealand, C. M., Burn, D. H. & Simonovic, S. P. 1999 Short term streamflow forecasting using artificial neural networks. *J. Hydrol.* **214**, 32–48.

Zounemat-Kermani, M. 2014 Principal component analysis (PCA) for estimating chlorophyll concentration using forward and generalized regression neural networks. *J. Applied Artificial Intelligence* **28**, 16–29.