Hybrid wavelet-support vector machine approach for modelling rainfall–runoff process
Mehdi Komasi and Soroush Sharghi

ABSTRACT

Because of the importance of water resources management, the need for accurate modeling of the rainfall–runoff process has rapidly grown in the past decades. Recently, the support vector machine (SVM) approach has been used by hydrologists for rainfall–runoff modeling and the other fields of hydrology. Similar to the other artificial intelligence models, such as artificial neural network (ANN) and adaptive neural fuzzy inference system, the SVM model is based on the autoregressive properties. In this paper, the wavelet analysis was linked to the SVM model concept for modeling the rainfall–runoff process of Aghchai and Eel River watersheds. In this way, the main time series of two variables, rainfall and runoff, were decomposed to multiple frequent time series by wavelet theory; then, these time series were imposed as input data on the SVM model in order to predict the runoff discharge one day ahead. The obtained results show that the wavelet SVM model can predict both short- and long-term runoff discharges by considering the seasonality effects. Also, the proposed hybrid model is relatively more appropriate than classical autoregressive ones such as ANN and SVM because it uses the multi-scale time series of rainfall and runoff data in the modeling process.

Key words | Aghchai watershed, Eel River watershed, rainfall and runoff modeling, support vector machine, wavelet transform

INTRODUCTION

The accurate and exact modeling of hydrological processes such as rainfall–runoff can give impressive information for urban planning, land use, flood and water resources management for a given watershed. Therefore, many hydrological models have been developed in order to simulate this complex process and a comprehensive classification of these models has been presented by Nourani et al. (2007). Also, in a comparative study, Mehrotra & Singh (1998) examined relative performances of different rainfall–runoff models. Classic time series models involving autoregressive integrated moving averages are widely used for forecasting the hydrological time series (Salas et al. 1980). However, they are basically linear models assuming that data are stationary and have a limited ability to capture non-stationary and non-linearity data regarding the hydrologic data. So, the next generation of a black box model, including artificial neural network (ANN), was used by researchers. Considering some preliminary studies, Hsu et al. (1995), Minns & Hall (1996) and Shamseldin (1997) applied ANN to rainfall–runoff modeling, and recently, such researchers as Rajurkar et al. (2004) and Garbrecht (2006) have investigated the rainfall–runoff relationship using the ANN approach. Also, Giustolisi & Simeone (2006) used ANN by a multiobjective strategy to predict groundwater level. The ANN offers an effective approach to handle large amounts of dynamic, non-linear and noisy data, especially when the underlying physical relationships are not fully understood. This makes them well suited to time series modeling problems of a data-driven nature. As a new artificial intelligence (AI) model, the adaptive neural-fuzzy inference system (ANFIS) was used by Shafie et al. (2007) to forecast the inflow of Nile River at Aswan dam. The hybrid ANN and fuzzy systems may be a research focus which can use the advantages of both ANN and fuzzy systems, namely ANFIS (Nourani et al. 2011). This hybrid model is capable of combining the benefits of both fields in a single framework. In some studies, the application of the ANFIS model in rainfall–runoff modeling has been investigated (e.g. Gautam & Holz 2001; Nayak et al. 2004; Tayfur & Singh 2006; Jothiprakash et al. 2009). Nowadays, the support...
vector machine (SVM) model is proposed as a new tool in time series modeling. Researches have shown that the SVM approach is trained faster than the ANN and ANFIS models. In addition, the results obtained from the SVM approach are more accurate than the ANN approach (Choy & Chan 2005; Lin et al. 2009). Compared to the most common neural network, the SVM based on the structural risk minimization principle and linearly constrained quadratic programming theory can obtain better generation performance. In addition, the solution of the SVM model is unique and globally optimal (Tang et al. 2009).

The SVM model is based on the statistical learning theory developed by Vapnik & Cortes (1995). This model is a supervised learning technique that provides the input-output mapping functions with respect to a set of training data (Wang 2005). The advantage of SVM modeling in comparison with the other methods is that this modeling works with less training data and variables (Choy & Chan 2005). Compared to a traditional neural network, the SVM method replaces traditional empirical risk with structural risk minimization and solves a quadratic optimization problem which can get the optimal global solution, in theory (Liu et al. 2011). The SVM model uses a device called kernel mapping in order to map data in an input space to a high dimensional featured space in which the problem becomes linearly separable (Burges 1999). In the SVM training, the decision boundaries are directly determined by the training data. This learning strategy is based on statistical learning theory and minimizes the classification errors of the training data and the unknown data (Abe 2010).

In spite of suitable flexibility of the SVM in modeling the hydrologic time series such as runoff (Zhang & Dong 2001), there is sometimes a shortage when signal fluctuations are highly non-stationary and operate under a large range of scales varying from one day to several decades. Also, a lot of time series are always the superposition of periodic and trend terms with regard to the changes of both short-term and long-term characteristics. In such a situation, the SVM model may not be able to cope with non-stationary data if pre-processing the input and/or output data is not performed (Cannas et al. 2006; Xiaohong et al. 2011). Therefore, data pre-processing can be performed by the decomposition of a time series into its subcomponents using wavelet transform analysis. If this method is used, the impact of short-term characteristics will be weakened or even disappear. Wavelet transform provides useful decompositions of main time series so that wavelet-transformed data can improve the ability of a forecasting model by capturing useful information on various resolution levels. The wavelet-based decomposition of non-stationary time series into different scales provides an interpretation of series structure and extracts considerable knowledge about its history in time and frequency domains (Rajae et al. 2010; Nourani et al. 2012). Hence, a hybrid wavelet-support vector machine (WSVM) which uses multi-scale signals as input data may present more probable forecasting than a single pattern input.

In this paper, the WSVM model was proposed for the rainfall–runoff modeling of two watersheds with different climatologic characteristics. In this model, the daily rainfall and runoff time series of watersheds are decomposed into sub-signals with various resolution levels and periodicity scales by wavelet; then, these sub-signals are inserted into the SVM model to reconstruct the forecasted runoff time series. The other sections of this paper are organized as follows. In section 2, the concepts of wavelet transform, SVM and ANN models are briefly reviewed. Then, the structure of the proposed hybrid model was presented. In section 3, the study areas are introduced. In section 4, the performance of classical SVM and WSVM models are evaluated and discussed with respect to different structures. Concluding remarks are in the final section of paper.

METHODS

Wavelet transform

The wavelet transform has increased in usage and popularity in recent years since its inception in the early 1980s, yet still does not enjoy the widespread usage of the Fourier transform. Fourier analysis has a serious drawback. In transforming to the frequency domain, time information is lost. When looking at a Fourier transform of a signal, it is impossible to tell when a particular event took place but wavelet analysis allows the use of long time intervals where we want more precise low-frequency information, and shorter regions where we want high-frequency information. Figure 1 compares Fourier transform and wavelet transform.

In the field of earth sciences, Grossmann & Morlet (1984), who worked especially on geophysical seismic signals, introduced the wavelet transform application. A comprehensive literature survey of wavelet in geosciences can be found in Foufoula-Georgiou & Kumar (1995) and Labat (2005). As there are many good books and articles introducing the wavelet transform, this paper will not
delve into the theory behind wavelets and only the main concepts of the transform are briefly presented; recommended literature for the wavelet novice includes Mallat (1998) or Labat et al. (2000). The time-scale wavelet transform of a continuous time signal, \( x(t) \), is defined as (Mallat 1998):

\[
T(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} g\left( \frac{t-b}{a} \right) x(t) \cdot dt
\]  

(1)

where \( g \) corresponds to the complex conjugate and \( g(t) \) is called wavelet function or mother wavelet. The parameter \( a \) acts as a dilation factor, while \( b \) corresponds to a temporal translation of the function \( g(t) \), which allows the study of the signal around \( b \). The main property of wavelet transform is to provide a time-scale localization of processes, which derives from the compact support of its basic function. This is opposed to the classical trigonometric function of Fourier analysis. The wavelet transform searches for correlations between the signal and wavelet function. This calculation is done at different scales of \( a \) and locally around the time of \( b \). The result is a wavelet coefficient \( (T(a,b)) \) contour map known as a scalogram. So the original signal may be reconstructed using the inverse wavelet transform as:

\[
x(t) = \frac{1}{c_t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{a^m}} \left( \frac{t-b}{a^m} \right) T(a, b) \frac{da \cdot db}{a^2}
\]  

(2)

For practical applications, the hydrologist does not have at his or her disposal a continuous-time signal process but rather a discrete-time signal. A discretization of Equation (2) based on the trapezoidal rule may be the simplest discretization of the continuous wavelet transform. This transform produces \( N^2 \) coefficients from a data set of length \( N \); hence redundant information is locked up within the coefficients, which may or may not be a desirable property (Addison et al. 2001). To overcome the mentioned redundancy, logarithmic uniform spacing can be used for the \( a \) scale discretization with correspondingly coarser resolution of the \( b \) locations, which allows for \( N \) transform coefficients to completely describe a signal of length \( N \). Such a discrete wavelet has the form (Mallat 1998):

\[
g_{m,n}(t) = \frac{1}{\sqrt{a_0^{m}}} \left( \frac{t-nb_0a_0^{m}}{a_0^{m}} \right) g(t)
\]  

(3)

where \( m \) and \( n \) are integers that control the wavelet dilation and translation, respectively; \( a_0 \) is a specified fined dilation step greater than 1; and \( b_0 \) is the location parameter and must be greater than zero. The most common and simplest choice for parameters are \( a_0 = 2 \) and \( b_0 = 1 \).

This power-of-two logarithmic scaling of the translation and dilation is known as the dyadic grid arrangement.

**Support vector machine**

SVMs are kernel machines useful for classification and regression problems (Giustolisi 2006). Much research and investigation has so far been carried on the theory of the SVM (Kecman 2001; Theodoridis et al. 2010). Therefore, only a summary explanation of the SVM model is given in this article. Suppose we are given training data \( \{(x_1,y_1), \ldots \} \).
(x,y)) ⊂ X×R, where x and y define the space of input patterns and target values. The purpose of the SVM is to find a function \( f(x) \) that has the greatest \( \epsilon \) deviation from actual targets \( y_i \) for all training data, and it should be as flat as possible. In other words, errors are not considered as long as they are less than \( \epsilon \), but any deviation larger than \( \epsilon \) will not be accepted. Linear functions of \( f(x) \) are described as follows (Sadeghpour et al. 2014):

\[
f(x) = \langle w, x \rangle + b, \quad w \in X, \ b \in R
\]  

(4)

where \( \langle w, x \rangle \) defines the dot product in X. Flatness in the case of Equation (4) means that a small \( w \) is sought. One procedure to ensure \( w \) is to minimize the norm i.e. \( ||w||^2 = \langle w, w \rangle \). This problem can be written as a convex optimization problem:

Minimize \[
\frac{1}{2} ||w||^2 + c \sum_{i=1}^{l} (\xi_i + \xi_i^*)
\]

Subject to \[
\{ \begin{align*}
y_i - \langle w, x_i \rangle - b &\leq \epsilon + \xi_i \\
\langle w, x_i \rangle + b - y_i &\leq \epsilon + \xi_i^* \\
\xi_i, \xi_i^* &\geq 0
\end{align*} \}
\]

(5)

The assumption in Equation (5) is that such a function \( f(x) \) that approximates all pairs \( (x_i, y_i) \) with \( \epsilon \) precision and accuracy actually exists, and that the convex optimization problem can be solved. On occasion, this may not be the case, or it may also be permitted for some errors. It can be introduced slack variables \( \xi_i, \xi_i^* \) to cope with otherwise unsolvable limitations of the optimization problem Equation (6). Thus, the formulation can be described as follows:

Minimize \[
\frac{1}{2} ||w||^2 + c \sum_{i=1}^{l} (\xi_i + \xi_i^*)
\]

Subject to \[
\{ \begin{align*}
y_i - \langle w, x_i \rangle - b &\leq \epsilon + \xi_i \\
\langle w, x_i \rangle + b - y_i &\leq \epsilon + \xi_i^* \\
\xi_i, \xi_i^* &\geq 0
\end{align*} \}
\]

(6)

The constant \( C > 0 \) specifies the trade-off between the flatness of \( f(x) \) and the amount up to which deviations larger than \( \epsilon \) are tolerated. This corresponds to the situation with the so called \( \epsilon - \) insensitive loss function \( |\xi|_{\epsilon} \) explained in Equation (7) as follows (Sadeghpour et al. 2014):

\[
|\xi|_{\epsilon} = \begin{cases} 0 & \text{if } |\xi| < \epsilon \\ |\xi| - \epsilon & \text{otherwise} \end{cases}
\]

(7)

The optimization problem in Equation (7) can be solved in its dual formulation. The main plan is to make a Lagrange function from the primal target function and the corresponding limitations, by introducing a dual set of variables. It has a saddle point with respect to the primal and dual variables at the solution. This function is shown by Equation (8):

\[
L = \frac{1}{2} ||w||^2 + c \sum_{i=1}^{l} (\xi_i + \xi_i^*)
\]

\[
- c \sum_{i=1}^{l} (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^{l} a_i (\epsilon + \xi_i - y_i - \langle w, x_i \rangle + b)
\]

\[
- \sum_{i=1}^{l} a_i^* (\epsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b)
\]

(8)

The term \( L \) is the Lagrangian and \( \eta_i, \eta_i^*, a_i, a_i^* \) are factors of Lagrange multipliers. Therefore, dual variables in Equation (9) have to allow for constraints.

\[
\eta_i, \eta_i^*, a_i, a_i^* \geq 0
\]

(9)

It follows the saddle point condition that the partial derivatives of \( L \) with respect to the primal variables \( (w, b, \xi_i, \xi_i^*) \) have to disappear for optimality.

\[
\frac{\partial L}{\partial b} = \sum_{i=1}^{l} (a_i^* - a_i) = 0
\]

\[
\frac{\partial L}{\partial W} = W - \sum_{i=1}^{l} (a_i^* - a_i) x_i = 0
\]

\[
\frac{\partial L}{\partial \xi_i} = C - (a_i^* - \eta_i^*) = 0
\]

(10)

Substituting Equation (10) into Equation (8) yields the dual optimization problem and eliminates dual variables. Equation (11) is rewritten as follows:

\[
W = \sum_{i=1}^{l} (a_i - a_i^*) x_i \quad \text{so that} \quad f(x) = \sum_{i=1}^{l} (a_i - a_i^*) \langle x_i, x \rangle + b
\]

(11)

This is the so-called support vector expansion. The \( w \) could be explained as a linear algorithm that only depends on dot products between training patterns \( w \). Linear model is not appropriate for numerous hydrological events. Consequently, it becomes suitable by converting Kernel for putting data in a space with more dimensions and then by applying the standard support vector regression algorithm. These interpretations will be convenient and suitable for the formulation of a nonlinear extension. This could be done by pre-processing the training
patterns \( x_i \) by a map: \( x \rightarrow \mathcal{C} \) into some feature space \( \mathcal{C} \). Thus, it is adequate to recognize Kernel function \( k(x, x_i) = \langle \varphi(x), \varphi(x_i) \rangle \) rather than \( \varphi \) explicitly. The Kernel function provides the opportunity for using a nonlinear function in input space for varying to linear function in characteristics space. Kernel function benefits to untreated high dimensional feature space explicitly. This technique is named Kernel trick and is shown as follows (Vapnik et al. 1997; Smola & Scholkopf 2004; Wang 2005; Kakaeei et al. 2012):

\[
f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha^*) k(x, x_i) + b \tag{12}
\]

The standard conversion of Kernel function, which is most often used in regression and modeling, is given in Table 1 (Wang 2005).

### ANN

It has proved that back-propagation (BP) algorithm network model with three-layer is satisfied for the forecasting and simulating in the science of water (Hornik 1989; Rajaee et al. 2009).

As shown in Figure 2, three-layered feed forward neural networks (FFNNs), which have been usually used in forecasting hydrologic time series, provide a general framework for representing nonlinear functional mapping between a set of input and output variables. Three-layered FFNNs are based on a linear combination of the input variables, which are transformed by a nonlinear activation function.

In the Figure 2, \( i, j \) and \( k \) denote input layer, hidden layer and output layer neurons, respectively, and \( w \) is the applied weight of the neuron. The term ‘feed-forward’ means that a neuron connection only exists from a neuron in the input layer to other neurons in the hidden layer, or from a neuron in the hidden layer to neurons in the output layer and the neurons within a layer are not interconnected to each other. The explicit expression for an output value of FFNNs is given by (Kim & Valdes 2003):

\[
\hat{y}_k = f_o \left[ \sum_{i=1}^{M} w_{ki} \cdot f_h \left( \sum_{j=1}^{N} w_{ji} x_i + w_{jo} \right) + w_{ko} \right] \tag{13}
\]

where \( w_{ji} \) is a weight in the hidden layer connecting the \( i \)th neuron in the input layer and the \( j \)th neuron in the hidden layer, \( w_{jo} \) is the bias for the \( j \)th hidden neuron, \( f_h \) is the activation function of the hidden neuron, \( w_{ki} \) is a weight in the output layer connecting the \( j \)th neuron in the hidden layer and the \( k \)th neuron in the output layer, \( w_{ko} \) is the bias for the \( k \)th output neuron, \( f_o \) is the activation function for the output neuron, \( x_i \) is \( i \)th input variable for input layer and \( \hat{y}, y \) are computed and observed output variables. The
weights are different in the hidden and output layers, and their values can be changed during the process of network training.

**WSVM model**

As shown in Figure 3, the WSVM and SVM have a similar operation on data modeling but the WSVM modeling made the raw data decomposed into sub-signal by wavelet transform. In fact, the wavelet coefficients provide the detail signals, which can capture small features of interpretational value in the rainfall and runoff data.

The Matlab programming was used to achieve the results of the modeling. In the proposed method, the rainfall \((I(t))\) and runoff \((Q(t))\) signals were firstly decomposed into sub-signals with different scales, i.e. a large-scale sub-signal and several small-scale sub-signals in order to catch temporal characteristics of the input time series. If \(i\) and \(j\) were supposed as the decomposition levels of the rainfall and runoff time series, respectively, the number of data in the input layer would be determined with \(i + j + 2\) because the model uses two variables and each time series would be decomposed into \(i + 1\) or \(j + 1\) sub-signals. This study deals with some irregular mother wavelets such as Haar, db2, sym3 and coif1 which are shown in Figure 4.

The structure architecture that yielded the best results in terms of determination coefficient and root mean square error on the training and verifying steps may be specified through the trial and error process. For this purpose, the data set was divided into two parts: the first 75% of total data were used as a training set and the second 25% were used for verifying the hybrid model (WSVM). Also, the following measures of evaluation have been used to compare the efficiency of the different models:

\[
DC = 1 - \frac{\sum_{i=1}^{N} (Q_{obsi} - Q_{comi})^2}{\sum_{i=1}^{N} (Q_{obsi} - Q_{obs})^2}
\]  

(14)

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Q_{obsi} - Q_{comi})^2}{N}}
\]  

(15)

where \(DC\), \(RMSE\), \(N\), \(Q_{comi}\), \(Q_{obsi}\) and \(Q_{obs}\) are determination coefficient, root mean squared error, number of observations, observed data, predicted values and mean of observed data, respectively. Also, the Equation (16) can be used to compare the ability of different models in capturing the peak values in runoff time series as similar as Equation (14) for the total data.

\[
DC_{peak} = 1 - \frac{\sum_{i=1}^{N} (Q_{pci} - Q_{poi})^2}{\sum_{i=1}^{N} (Q_{pci} - Q_{poi})^2}
\]  

(16)

where \(DC_{peak}\) is the determination coefficient for peak values, \(N\) is the number of peak values, \(Q_{poi}\), \(Q_{pci}\) and \(Q_{po}\) are observed data, computed values and mean of observed data for peak values, respectively. The RMSE was used to measure forecast accuracy, which produces a positive value by squaring the errors. The RMSE increases from zero for perfect forecasts through to large positive values as the discrepancies between forecasts and observations become increasingly large. Obviously, a high value for \(DC\) (up to one) and a small value for RMSE indicate high efficiency of the model.

![Figure 3](https://iwaponline.com/wst/article-pdf/73/8/1937/462465/wst073081937.pdf)  
*Figure 3* | The structure of the hybrid WSVM model.
Study area and data

Data and information on Aghchai watershed in Azerbaijan province, Iran and Eel River watershed in California, USA were selected for this study.

Aghchai watershed

Aghchai watershed is located between 38°40’ and 39°30’ northern latitude and 44°10’ and 44°57’ eastern longitude. The watershed area is 1,440 km² (Figure 5). Watershed elevation varies between 1,168 and 3,280 m above sea level and its longest waterway has the length of 64.88 km. The topography is steep with an average slope of 25%. The time series data for 12 years (1995–2007) concerning this watershed were used in the modeling process (the first 8 years for training and then 4 years for verification).

Eel River watershed

Eel River watershed, with a drainage area of 8,051 km², is located between 39°14’ 25” and 40°29’ 43” northern latitude and 122°39’ 58” and 124°08’ 51” western longitude. It drains a rugged area between the Sacramento Valley and the ocean, crossing Humbolt, Lake, Mendocino and Trinity counties. Figure 6 shows the map of study area. The historical daily discharge data are available for nine stations inside the watershed (USGS website). Each station was considered as an outlet for the sub-basin. Also, daily

![Figure 5](https://iwaponline.com/wst/article-pdf/73/8/1937/462465/wst073081937.pdf)
rainfall data are available for three rain gauges (HydroLab website) in which the average distance-weighted rainfall values recorded in these gauges were considered as the daily rainfall over each sub-basin. In the current research, the models were calibrated using the Eel River data from December 1966 to September 1969 and were verified by the data set gathered during October 1969 to September 1970. The data of outlet station ID: 11477000 were used for the modeling process.

The statistic characteristics of rainfall and runoff for both watersheds in daily time scale are categorized in Table 2. The daily rainfall and runoff time series which were used in this study are presented in Figures 7–10.

<table>
<thead>
<tr>
<th>Step</th>
<th>Case study</th>
<th>Rainfall time series (mm)</th>
<th>Runoff time series (m³/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration</td>
<td>Aghchaid</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>Eel River</td>
<td>317</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Verification</td>
<td>Aghchaid</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>Eel River</td>
<td>307.67</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

RESULTS AND DISCUSSION

The SVM model consists of some structure parameters including Kernel function with various types. So, in the first step, the purpose is to select the best Kernel function to be applied in the hybrid model. In the next step, the wavelet transform is joined with the SVM model while selecting the best wavelet function and decomposition level.

Classical SVM

As it has been mentioned earlier, the SVM algorithm consists of Kernel function. Also, there are different kinds of Kernel
functions. In this paper, three kinds of Kernel functions involving RBF-kernel, Poly-kernel and MLP-kernel have been used. In the first step, the purpose is to select the best Kernel function as a high performance function regarding data modeling in comparison with the other Kernel functions. For this purpose, as it is shown in Table 3, the SVM modeling is performed by three kinds of Kernel functions with specific data on rainfall and runoff at time $t$ ($Q(t)$, $I(t)$) for Aghchaei and Eel River watersheds. Hence, DC and RMSE were used as the model efficiency criteria in the calibration and verification of the model. As described previously, a high value for DC (up to one) and a small value for RMSE indicate high efficiency of the desired model. On the other hand, one of the preferences of the SVM model as compared to the other models is that there is no need for the time series normalization of SVM modeling. Therefore, RMSE shows the real value of error according to discharge.

As shown in Table 3, the DC values for RBF-kernel function are 0.64 and 0.65 in calibration and verification steps, respectively, for Aghchaei watershed. Also, at Eel River watershed these values for RBF-kernel function are 0.65 and 0.52. As a result, the DC values for RBF-kernel function are more than the other kernels for both watersheds. So, the RBF-kernel function is more efficient than the Poly and MLP
kernel functions for both watersheds. Also, RBF-kernel function is taken into consideration in this paper because of the following:

1. In contrast to linear kernel function, it can hold the case when the communication between class labels and attributes is nonlinear.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Time series results of Aghchel and Eel River watersheds regarding information of Q(t, lII) for selecting the best Kernel function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aghchel</td>
<td>DC</td>
</tr>
<tr>
<td>Kernel function</td>
<td>Calibration</td>
</tr>
<tr>
<td>Poly-kernel</td>
<td>0.52</td>
</tr>
<tr>
<td>RBF-kernel</td>
<td>0.64</td>
</tr>
<tr>
<td>MLP-kernel</td>
<td>0.55</td>
</tr>
</tbody>
</table>

**Figure 9** | Runoff time series of Aghchel watershed.

**Figure 10** | Runoff time series of Eel River watershed.
2. This kernel function has a smaller amount of hyper parameters which control the difficulty of model selection.


The equations of kernel functions like RBF, Poly and MLP are presented in Table 1. Now, the details of the best kernel function, namely RBF-kernel, have been explained.

In machine learning, the (Gaussian) radial basis function kernel or RBF-kernel is a popular kernel function used in the classification of SVMs (Chang et al. 2010). \( k(x, x_i) = e^{-\gamma \|x - x_i\|^2} \) ranges between zero and one. In a more accurate explanation, \( k(x, x_i) = e^{-1/2\sigma^2 \|x - x_i\|^2} \) so that the value of \( \gamma \) in the first equation is \( \gamma = 1/2\sigma^2 \). \( x \) and \( x_i \) denote two sample data. On the other hand, \( x \) and \( x_i \) represent feature vectors in some input space. \( \sigma \) is a free parameter and single value (and, subsequently, \( \gamma \)) which needs to be tuned considering a validation or tuning data set (Vert et al. 2004).

The daily rainfall and runoff modeling is a Markovian process so that the process state at previous times \((Q_{t-1}), (I_{t-1}), (Q_{t-2}), (I_{t-2}), \ldots\) was used to forecast the \( Q_{t+1} \). The mathematic relation can be defined as Equation (17) while helping the model as an autoregressive model to achieve the best result.

\[
Q_{t+1} = f_{svm}(Q_t, I_t, Q_{t-1}, I_{t-1}, Q_{t-2}, I_{t-2}, \ldots)
\]  

Table 4 | Time series results of Aghchail and Eel River watersheds with different combinations of input data by RBF-kernel function

<table>
<thead>
<tr>
<th>Combinations</th>
<th>Aghchail</th>
<th>Eel River</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comb. (1)</td>
<td>DC 0.53</td>
<td>RMSE 0.22</td>
</tr>
<tr>
<td></td>
<td>0.59</td>
<td>0.86</td>
</tr>
<tr>
<td>Comb. (2)</td>
<td>DC 0.64</td>
<td>RMSE 0.15</td>
</tr>
<tr>
<td></td>
<td>0.69</td>
<td>0.79</td>
</tr>
<tr>
<td>Comb. (3)</td>
<td>DC 0.62</td>
<td>RMSE 0.27</td>
</tr>
<tr>
<td></td>
<td>0.42</td>
<td>1.22</td>
</tr>
<tr>
<td>Comb. (4)</td>
<td>DC 0.62</td>
<td>RMSE 0.29</td>
</tr>
<tr>
<td></td>
<td>0.33</td>
<td>1.49</td>
</tr>
<tr>
<td>Comb. (5)</td>
<td>DC 0.63</td>
<td>RMSE 0.31</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>1.74</td>
</tr>
</tbody>
</table>

Figure 11 | Computed and observed runoff time series of Aghchail watershed obtained by RBF-kernel function with combination (2) of input data.
Comb. (4): \( Q(t), Q(t-1), Q(t-2), Q(t-3), I(t), I(t-1), I(t-2), I(t-3), I(t-4) \)

Comb. (5): \( Q(t), Q(t-1), Q(t-2), Q(t-3), Q(t-4), I(t), I(t-1), I(t-2), I(t-3), I(t-4) \)

In this step, the purpose was to select the best combination that gives a high performance concerning the data modeling, in comparison with the other combinations. For this purpose, the rainfall and runoff modeling for each watershed was done by the Kernel function obtained from the obvious stage (RBF-kernel) and different following input combinations; then, the values of DC and RMSE were used to specify the accuracy of calibration and verification similar to the previous stage. The results obtained by the SVM modeling are presented in Table 4.

As shown in Table 4, the DC values for combination (2) are 0.64 and 0.69 in calibration and verification steps, respectively, for Aghchay watershed. Also these values for combination (2) are 0.64 and 0.58 for Eel River watershed. As a result, the DC values for combination (2) are more than the other combinations for both watersheds. So, the results indicate the improved performance of the combination (2) for both watersheds. Computed and observed runoff time series of Aghchay and Eel River watersheds.
obtained by RBF-kernel function with combination (2) of input data are shown in Figures 11 and 12.

**WSVM method**

In this step, the pre-processed data were inserted in the SVM model in order to examine the effects of wavelet analysis on the modeling process. For this purpose, the discrete wavelet transform was performed. Wavelet algorithm process data at different temporal scales (levels) allow gross and small features of a signal to be separated. In this study, attempts were made to distinguish the effects of the used mother wavelet type and decomposition level on the model efficiency. Hence, time series were decomposed to one, two, three and

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Result of the WSVM model by RBF-kernel function with different mother wavelets and decomposition levels</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mother wavelet type</strong></td>
<td><strong>Decomposition level (i – j)</strong></td>
</tr>
<tr>
<td><strong>Aghchai watershed</strong></td>
<td></td>
</tr>
<tr>
<td>Haar</td>
<td>1</td>
</tr>
<tr>
<td>Haar</td>
<td>2</td>
</tr>
<tr>
<td>Haar</td>
<td>3</td>
</tr>
<tr>
<td>Haar</td>
<td>4</td>
</tr>
<tr>
<td>db2</td>
<td>1</td>
</tr>
<tr>
<td>db2</td>
<td>2</td>
</tr>
<tr>
<td>db2</td>
<td>3</td>
</tr>
<tr>
<td>db2</td>
<td>4</td>
</tr>
<tr>
<td>sym1</td>
<td>1</td>
</tr>
<tr>
<td>sym1</td>
<td>2</td>
</tr>
<tr>
<td>sym1</td>
<td>3</td>
</tr>
<tr>
<td>sym1</td>
<td>4</td>
</tr>
<tr>
<td>coif1</td>
<td>1</td>
</tr>
<tr>
<td>coif1</td>
<td>2</td>
</tr>
<tr>
<td>coif1</td>
<td>3</td>
</tr>
<tr>
<td>coif1</td>
<td>4</td>
</tr>
<tr>
<td><strong>Eel River watershed</strong></td>
<td></td>
</tr>
<tr>
<td>Haar</td>
<td>1</td>
</tr>
<tr>
<td>Haar</td>
<td>2</td>
</tr>
<tr>
<td>Haar</td>
<td>3</td>
</tr>
<tr>
<td>Haar</td>
<td>4</td>
</tr>
<tr>
<td>db2</td>
<td>1</td>
</tr>
<tr>
<td>db2</td>
<td>2</td>
</tr>
<tr>
<td>db2</td>
<td>3</td>
</tr>
<tr>
<td>db2</td>
<td>4</td>
</tr>
<tr>
<td>sym1</td>
<td>1</td>
</tr>
<tr>
<td>sym1</td>
<td>2</td>
</tr>
<tr>
<td>sym1</td>
<td>3</td>
</tr>
<tr>
<td>sym1</td>
<td>4</td>
</tr>
<tr>
<td>coif1</td>
<td>1</td>
</tr>
<tr>
<td>coif1</td>
<td>2</td>
</tr>
<tr>
<td>coif1</td>
<td>3</td>
</tr>
<tr>
<td>coif1</td>
<td>4</td>
</tr>
</tbody>
</table>
four levels by four different kinds of wavelet transforms involving 1-Haar wavelet, a simple wavelet; 2-Daubechies wavelet (db2), a most popular wavelet; 3-sym3 wavelet with three sharp peaks; and 4-coif1 wavelet (Mallat 1998). These wavelets are shown in Figure 4. For instance, the decomposition level 3 of main rainfall signal for Aghchay watershed data with regard to total raw signal which yields four sub-signals (approximation at level 3 and details at levels 1, 2, and 3) by Haar wavelet is presented in Figure 13.

Next, the rainfall and runoff values of a distinct day from each sub-signal selected from the calibration data set were considered as an input in order to predict the runoff one day ahead (as output). Then, the trained model was validated by the verification data set. The obtained results for each watershed by the best Kernel function (RBF-kernel) are added up in Table 5 for all cases. When multilevel sub-signals were inserted into the model as an input, the applied weights to them by the SVM will be different at different decomposition levels (Figure 3). Thus, those high weights will be applied to the worthy signal level. For example, using the order of two for decomposition level which yields three sub-signals for both rainfall and runoff time series, \( Q(t+1) \) is more related to \( I_d2(t) \) rather than \( I_d1(t) \) because \( I_d2(t) \) (detailed sub-signal of rainfall) is a short period in rainfall time series and plays a significant role in predicting runoff at \( t+1 \) (\( Q(t+1) \)). Therefore, the structure magnifies its weight as comparatively as the other sub-signals.

Based on Table 5, the db2 wavelet function and decomposition 3 produced better results for two watersheds. According to the structure of db2 wavelet (Figure 4(b)), which is similar to the rainfall (I) and runoff (Q) signals, it could capture the signal features, especially peak values and yield comparatively high efficiency. The calibration and verification time series of the WSVM modeling in order to achieve the best results including high efficiency on data modeling for each watershed are shown in Figures 14 and 15. To better understand the computed and observed data, a specified range shown in Figure 14 is represented in Figure 16.

By comparing the obtained results (Table 5), it can be clearly seen in the calibration phase that 1, 2, 3, and 4 decomposition levels give relatively equal performance but in the verification phase, there was no clear trend in different wavelet functions. Furthermore, in Aghchay watershed, the difference between the decomposition levels 1 and 2 was perceptible for the verification step but this is not confirmed for levels 3 and 4.

The most important factor in the runoff management and engineering design of hydraulic equipment is the determination of runoff peak values. Hence, in this step, the abilities of SVM and WSVM models in determining peak values have been specified. For this purpose, peak values were sampled by considering the top 5% threshold of data from the original runoff time series contractually. Using Equation (16), the performances of SVM and WSVM models were evaluated and given as follows. For the SVM modeling, the DCpeak values of Aghchay and Eel River watersheds are 0.68 and 0.47, respectively. For WSVM modeling, these values are 0.94 and 0.68, respectively. According to the DCpeak values, it can be concluded that the seasonal model (i.e. WSVM) is more efficient than the autoregressive model; it refers to the SVM model in monitoring peak values. It is evident that extreme or peak values in the rainfall and runoff time series, which occur in a periodic pattern, can

![Figure 14](https://iwaponline.com/wst/article-pdf/73/8/1937/462465/wst073081937.pdf)
be accurately two watersheds located in various places of word. Regarding this comparison, it is proved that the results of data modeling by black box models including the SVM and WSVM ones are not related to the specific basin data and are attributed to physical characteristics.

In the final part, the classical SVM and hybrid WSVM models are compared with the other classical and hybrid models, namely ANN and wavelet artificial neural network (WANN). The results are shown in Table 6.

According to DC values, the WSVM model is more accurate than WANN one. Also, in each model case, the hybrid model showed a better performance than the single one due to appropriate performance of wavelet transform on input data; in other words, by comparing the single and hybrid modeling results (Table 6), it is concluded that using multi-resolution rainfall and runoff data as the input in SVM algorithm caused better results when compared to the use of raw (un-preprocessed) rainfall and runoff data.

**CONCLUDING REMARKS**

In this paper, an improved hybrid model involving WSVM was formed in order to model data. For this purpose, the classical SVM was first used for data modeling. The Kernel function was employed for data classification in the SVM. There were
various kinds of Kernel functions. In this paper, the SVM modeling is done by three kinds of Kernel functions with respect to rainfall and runoff data concerning Aghchai and Eel River watersheds. The obtained results indicated that the RBF-kernel function has the highest efficiency as compared to Poly- and MLP-kernel functions for both watersheds. So, the RBF-kernel function was employed in the next steps.

Afterwards, some input combinations (combs. 1–5) were used as the inputs of SVM model; then, the performance of each combination was measured. The results showed that the combination 2 has a better compatibility as the SVM input for both watersheds.

The most significant part of the study occurred when the wavelet transform was combined with the SVM. So, the single SVM modeling gets promoted by the wavelet transform. In this step, the time series of each watershed were calibrated and verified by the WSVM model with the different kinds of mother wavelet types including Haar, db2, sym1, and coif1 and different decomposition levels (1–4). Finally, the db2 mother wavelet and the decomposition level 3 were selected as high efficiency ones for both watersheds and their calibration and verification results were illustrated in the given figures. Finally, the results indicate high ability of WSVM to recognize peak data in hydrological time series and to observe long-term runoff discharges while considering the seasonality effects. In addition, the hybrid model is more appropriate because it uses the multi-scale time series of rainfall and runoff data in the SVM model. In the final part of this article, the WSVM model was compared to WANN and high relative accuracy of WSVM model was proved.

As it has been suggested for further studies, it is recommended to use the presented methodology to forecast the runoff 2, 3, … days ahead and also to model the rainfall–runoff process of watershed through adding the other hydrologic time series and variables. The WSVM model can be investigated for the other hydrological processes such as sedimentation and drought.

### Table 6

<table>
<thead>
<tr>
<th>Model</th>
<th>Model type</th>
<th>Aghchai (DC calibration)</th>
<th>Aghchai (DC verification)</th>
<th>Eel River (DC calibration)</th>
<th>Eel River (DC verification)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>Single</td>
<td>0.61</td>
<td>0.52</td>
<td>0.51</td>
<td>0.44</td>
</tr>
<tr>
<td>WAAN</td>
<td>Hybrid</td>
<td>0.92</td>
<td>0.82</td>
<td>0.83</td>
<td>0.77</td>
</tr>
<tr>
<td>SVM</td>
<td>Single</td>
<td>0.64</td>
<td>0.69</td>
<td>0.64</td>
<td>0.58</td>
</tr>
<tr>
<td>WSVM</td>
<td>Hybrid</td>
<td>0.96</td>
<td>0.92</td>
<td>0.93</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Note: In this table, the best result for each model is presented.

### REFERENCES

Abe, S. 2010 *Support Vector Machines for Pattern Classification*. Springer-Verlag, London, UK.


Liu, Z., Wang, X., Cui, L., Lian, X. & Xu, J. 2009 Research on water bloom prediction based on least squares support vector machine, in computer science and information engineering. WRI World Congress on IEEE.


First received 3 September 2015; accepted in revised form 30 November 2015. Available online 25 January 2016