

# Wavelet and cuckoo search-support vector machine conjugation for drought forecasting using Standardized Precipitation Index (case study: Urmia Lake, Iran)

Mehdi Komasi, Soroush Sharghi and Hamid R. Safavi

## ABSTRACT

In this study, wavelet-support vector machine (WSVM) is proposed for drought forecasting using the Standardized Precipitation Index (SPI). In this way, the SPI time series of Urmia Lake watershed is decomposed to multiple frequency time series by wavelet transform. Then, these time sub-series are applied as input data to the support vector machine (SVM) model to forecast drought. Also, a cuckoo search (CS)-based approach is proposed for parameter optimization of SVM, finding the best initial constant parameters of the SVM algorithm. The obtained results indicate that the radial basis function (RBF)-kernel function of the SVM algorithm has high efficiency in the SPI modeling, resulting in a determination coefficient (DC) of 0.865 in verification step. In the WSVM model, the Coif1, which is considered as a mother wavelet function with decomposition level of five, shows a better performance with DC of 0.954 in verification step, revealing that the proposed hybrid WSVM model outperforms the single SVM model in forecasting SPI time series. Also, DC of cuckoo search-support vector machine (CS-SVM) is calculated to be 0.912 in verification step, indicating the fact that the proposed CS-SVM model shows better efficiency than single SVM model.

**Key words** | cuckoo search, drought forecasting, SPI, SVM, Urmia Lake watershed, wavelet transform

**Mehdi Komasi** (corresponding author)  
Faculty of Engineering,  
University of Ayatollah Ozma Borujerdi,  
Borujerd,  
Iran  
E-mail: komasi@abru.ac.ir

**Soroush Sharghi**  
Water Resources Management Engineering,  
University of Tehran,  
Tehran,  
Iran

**Hamid R. Safavi**  
Department of Civil Engineering,  
Isfahan University of Technology,  
Isfahan,  
Iran

## INTRODUCTION

Unlike other natural menaces, droughts have a slow evolution time such that the outcomes of droughts take a considerable amount of time to come into effect. As a result, the ability to forecast and model the characteristics of drought, especially their initiation, frequency, and severity is important in order to manage water resources for agricultural and industrial uses. The conventional method to monitor drought conditions is a drought index. Some drought indices, such as the Palmer Drought Severity Index (PDSI) and the Standardized Precipitation Index (SPI), are more commonly used than others. A major advantage of the SPI is that it makes the description of droughts on multiple time scales possible (Cacciamani *et al.* 2007). One of the differences between the PDSI and

SPI is that the PDSI index has a complex structure with a very long memory whereas the SPI is an easily interpreted and simple-moving average process (Tsakiris & Vangelis 2004). Furthermore, unlike the PDSI, the characteristics of SPI are constant site to site and the calculations of which only take precipitation data (Belayneh & Adamowski 2013; Blagojevic *et al.* 2016). As a result, SPI is used as an appropriate drought index in this study.

The SPI needs machine learning tools to be forecasted. Today, artificial intelligence (AI) models such as the artificial neural network (ANN) and adaptive network-based fuzzy inference system (ANFIS) have been used in several studies to forecast the hydrological, geological, and

meteorological characteristics (Jalalkamali *et al.* 2011; Zanganeh *et al.* 2016; Dariane & Azimi 2017).

Another machine learning tools suggested for time series modeling is the support vector machine (SVM). The SVM learning or data-driven model has increasingly become popular in the hydrologic forecasting, mainly due to effectiveness in dealing with non-linear characteristics of hydrologic data. Research has shown that the SVM approach is trained faster than ANN and ANFIS models (Lin *et al.* 2009; Noori *et al.* 2015). Moreover, the results obtained from the SVM approach are more accurate than those of ANN approach (Lin *et al.* 2009). In recent years, the SVM model has been tested for hydrological and climatological applications due to its superior potential versus ANN model (Wang *et al.* 2008; Behzad *et al.* 2009; Huang *et al.* 2017). Nikbakht Shahbazi *et al.* (2011) used SVM model to forecast SPI in four reservoir basins supplying the water demand of Tehran, Iran. They concluded that the SVM model has enough accuracy to be used in long-term water resources planning and management compared to ANN. Zahraei & Nasser (2014) also used SVM to develop some models for forecasting seasonal SPI. They concluded that the SPI values can be forecasted by the proposed model with two to five months of lead-time with high accuracy.

Hybrid approaches have been extensively used in research to improve modeling accuracy (Vojinovic *et al.* 2003; Sannasiraj *et al.* 2004; Yu *et al.* 2004; Sun *et al.* 2010). For example, in spite of the suitable flexibility of SVM models for modeling hydrologic time series, there is a shortage arising when signal fluctuations are highly non-stationary and a physical-hydrologic process operates under a large range of scales varying from one day to several decades. In such a situation, the SVM model may not be able to cope with non-stationary data if pre-processing of the input and/or output data is not performed (Cannas *et al.* 2006). Therefore, data pre-processing can be performed by the decomposition of time series into its subcomponents while conducting a wavelet transform (WT) analysis. The WT provides useful decomposition of time series. Thus, it may improve the ability of a forecasting model by capturing useful information on various resolution levels. Hence, a hybrid wavelet-support vector machine (WSVM) model which uses multi-scale signals as input data rather than a single pattern input may present more accurate forecasting.

Furthermore, cuckoo search (CS) has recently been introduced as an optimization algorithm for parameter estimation in some applications, such as forecasting the output energy values of wind parks in Texas and Montana (Barbosa & Vasconcelos 2016) and applying a cuckoo search-support vector machine (CS-SVM) model for predicting dynamic measurement errors of sensors (Jiang *et al.* 2016). However, no attention has been paid to forecasting hydrological time series using the CS algorithm. As a result, CS may help the SVM model to forecast hydrological time series (e.g., SPI time series) efficiently by optimizing and setting appropriate constant parameters of the SVM algorithm.

In this paper, the WSVM model was proposed for drought modeling based on SPI obtained by a corresponding precipitation value from the same month in the previous year for Urmia Lake watershed in Iran. In this model, SPI data 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 a multi-scale model for forecasting SPI time series. In addition, another aim of this study is to evaluate and validate the efficiency of CS-SVM model for modeling and forecasting SPI time series.

The other sections of this paper are organized as follows. Below, the SPI methods, WT, SVM, CS models, and the case study are briefly described. Then, the structures of the proposed hybrid model WSVM and CS-SVM are presented. In the section after, the classical SVM, WSVM, and CS-SVM model performances are evaluated and discussed with different structures. Concluding remarks are in the final section of the paper.

## METHODS

### Standardized Precipitation Index

The SPI was developed by McKee *et al.* (1993). It is exclusively based on precipitation that makes its calculation comparatively straightforward. Standardization of a drought index confirms the independence from the geographical position, affording the desired index to be calculated with respect to the average precipitation in the same place (Cacciamani *et al.* 2007).

SPI values are positive or negative for greater or less values than mean precipitation, respectively. Positive SPI values for wet conditions are greater than mean precipitation and negative SPI values for dry conditions are lower than mean precipitation. Variance from the mean is a probability indication of the severity of humidity or drought that can be employed for the risk assessment. Table 1 indicates SPI drought classes.

For more details, please see Abramowitz & Stegun (1965), Cacciamani et al. (2007), and Belayneh & Adamowski (2012).

**Wavelet transform**

The WT is an analytical appliance, providing a time-frequency representation of a signal in the time domain. The time scale WT of a continuous time series,  $x(t)$ , is defined as Equation (1) (Mallat 1998):

$$T(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} g^* \left( \frac{t-b}{a} \right) x(t).dt \tag{1}$$

where \* refers to the complex conjugate and  $g$  is called wavelet function or mother wavelet. The parameter  $a$  is introduced as the scale parameter indicating the range and duration of the desired time series. Meanwhile, the  $b$  parameter is transmission parameter determining wavelet position on the time axis. As a special case of this study, the WT concept is proposed in order to decompose initial raw signal of the SPI time series into several sub-series using mother wavelet functions extended by  $a$  and  $b$  parameters. This research deals with some mother wavelet

functions such as Haar, Db2, Sym3, Coif1, Mexican hat, and Morlet, which are illustrated in Figure 1.

**Support vector machine**

SVM is a kernel structure which is generally applied for solving classification and regression problems. The design of the SVM model was developed by Vapnik & Cortes (1995). It was formulated based on the structure risk minimization (SRM) principle. It has been proved that the application of SRM in SVM leads to better efficiency than the application of traditional empirical risk minimization (ERM) principle used in traditional models. SRM minimizes an upper bound on the expected risk whereas ERM minimizes the error for the training data (Haykin 2003). SVM can be used for regression and classification.

The SVM model detects the optimal solution of the following primary problem (Zahraei & Nasser 2014).

Minimize:

$$\phi(w, \xi, \xi^{(*)}) = \frac{1}{2} w^T w + C \cdot \frac{1}{L} \sum_{i=1}^L (\xi_i + \xi_i^*),$$

Subject to:

$$\begin{cases} (\langle w, x_i \rangle + b) - y_i \leq \varepsilon + \xi_i, \\ y_i - (\langle w, x_i \rangle + b) \leq \varepsilon + \xi_i^*, \\ \xi_i \geq 0 \end{cases} \tag{2}$$

$$w \in X. \xi^{(*)} \in R^m. b \in R. i = 1, \dots, L$$

where  $L$  is the number of data points in the training data set,  $C$  is the model parameter,  $x_i$  is the  $i_{th}$  feature space data point,  $w$  is one of the optimization problem’s decision variables, and  $\xi$  is the model residual defined by  $(\xi_i = y_i - f(x_i))$ .  $\xi_i$  and  $\xi_i^*$  are positive slack variables and  $C$  is a positive real value and predetermined constant. The constant  $C$  specifies the amount to which deviations from  $\varepsilon$  are tolerated. Deviations above  $\varepsilon$  are defined by  $\xi_i$  whereas deviations below  $\varepsilon$  are defined by  $\xi_i^*$ .  $C$ , which is always positive, is the penalty parameter on the training error. As a result, the value of  $C$  and  $\varepsilon$  must be properly selected by the user.

**Table 1** | Drought classification based on SPI (McKee et al. 1993)

SPI values	Class
>2	Extremely wet
1.5-1.99	Very wet
1.0-1.49	Moderately wet
-0.99-0.99	Near normal
-1-1.49	Moderately dry
-1.5-1.99	Very dry
<-2	Extremely dry

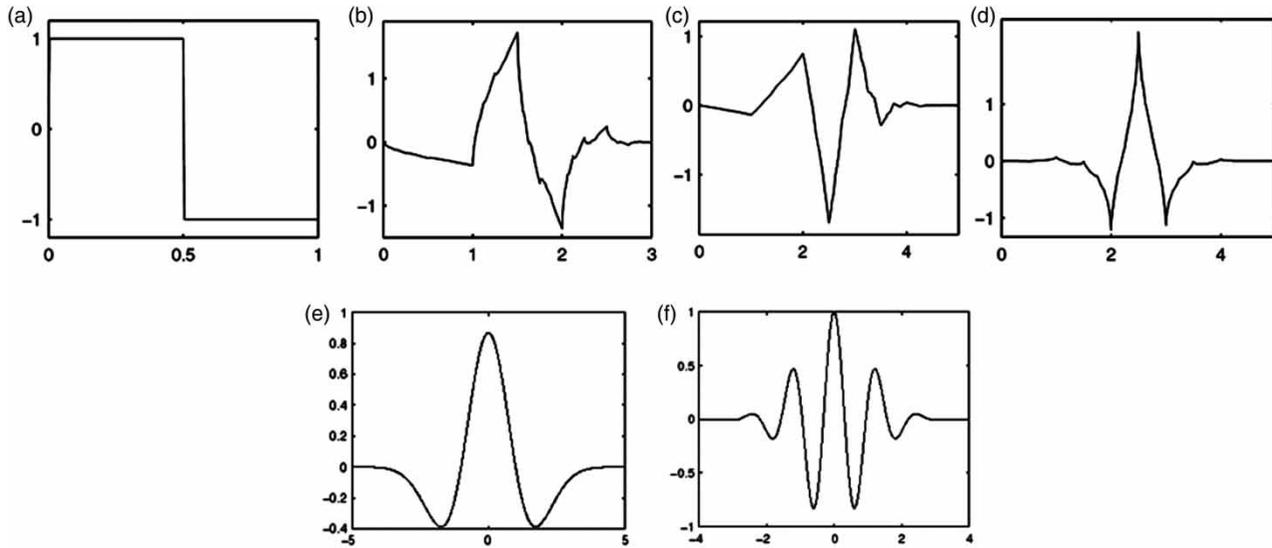


Figure 1 | (a) Haar, (b) Db2, (c) Sym3, (d) Coif1, (e) Mexican hat, (f) Morlet.

In this research, the SVM method is employed in order to carry out regression and forecast the monthly time series of SPI. The transformation of initial data from the input space to a new space (feature space) is another quality of SVM, leading to better recognition of the data structure and making linear-regression possible (Figure 2). For this purpose, a non-linear transformation function, called kernel function, is indicated to map the input space to a higher dimension feature space.

### Cuckoo search

CS algorithm is inspired by the unique behavior of the cuckoo species and Levy flight. Yang & Deb (2009) proposed the CS algorithm. Cuckoos lay their eggs in other birds' nests when the host birds leave the nest unguarded.

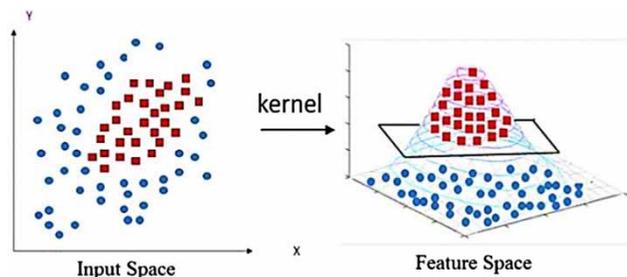


Figure 2 | Data transmission from the input space to feature space (Dibike et al. 2001).

In the process, some of these eggs, which are similar to the host bird's eggs, hatch and grow into adult cuckoos. If the host birds detect that the eggs are not their own, they will expel the alien eggs or leave their nest and find another place to build a new nest. Each egg in a nest suggests a solution, and a cuckoo egg represents a new solution. The purpose of the CS algorithm is to apply the new and potentially better solutions (cuckoos) to replace the not-so-good solutions in the nests (Jiang et al. 2016). The CS algorithm has the following three rules (Yang & Deb 2013):

1. Each cuckoo only lays one egg (one solution) at a time, and it puts the eggs in a haphazard determined nest.
2. In these nests, the best nest, with high quality eggs (solutions), will carry over to the next generation.
3. The total number of available host nests is fixed. A host bird can detect an alien egg with probability. In this case, the host bird may either expel the egg or leave and establish a new nest in a new location.

Based on the above three rules, the CS algorithm updates the bird nest locations. Its search path can be expressed as follows:

$$X_i^{t+1} = X_i^t + \alpha \oplus L \tag{3}$$

where  $X_i^t$  represents the location of the  $i$ th nest at iteration  $t$ . The product  $\oplus$  means entry-wise multiplication, and  $\alpha$  is the

step size, which is subject to a normal distribution.  $L$  is the Levy random search path.

**Model development**

**Wavelet-support vector machine model**

The hybrid WSVM idea is proposed when the SVM model uses the pre-processed data generated with a WT. As shown in Figure 3, the WSVM and SVM model have the same process for the data modeling, but the WSVM model uses the decomposed input data (that are generated with a WT) instead of raw ones.

MATLAB programming was used to obtain the modeling results. In the proposed approach, the SPI signals were first decomposed into sub-signals with different scales, such as one large scale sub-signal ( $SPI_a$ ) and several small scale sub-signals ( $SPI_{di}$ ) in order to catch the temporal characteristics of input time series (see Figure 3). Then, the SVM model was ready for data modeling in a space that was created by the kernel function ( $K(x, x_i)$ ) to forecast the one time ahead of the SPI ( $SPI_{(t+1)}$ ).

**Cuckoo search-support vector machine model**

The general procedure of CS-SVM is illustrated in the flowchart in Figure 4. The CS algorithm applied to optimize the SVM parameters  $C$  and  $\epsilon$  is organized as follows (Jiang et al. 2016):

1. Initialize the CS algorithm and set the number of nests,  $N$ , the probability parameters,  $p_a$ , the maximum iterations,  $t_{max}$ , and the ranges of  $C$  and  $\epsilon$ .
2. Randomly generate nest positions using  $q_i^0 = [x_1^0, x_2^0, \dots, x_n^0]^T$ . Each nest corresponds to a set of parameters ( $C, \epsilon$ ). The fitness evaluation function is defined as follows:

$$I = \frac{\sum_{i=1}^n (\bar{Y}_i - Y_i)^2}{n} \tag{4}$$

where  $Y_i$  is the actual value and  $\bar{Y}_i$  is the prediction value of the model, and  $n$  is the number of training samples.

3. Evaluate the fitness value of each nest, discover the current best solution, and record the minimum fitness value and its corresponding position.
4. Keep the best solutions from the previous generation, and update the position of the other nests using Equation (3). Then, evaluate the fitness value of the new position.
5. Replace the best solution of the previous generation if the fitness value of the new generation is better than that of the previous generation, and record the position of the best nest.
6. Set a random number (*random*) as the probability of egg detection. Compare it with  $p_a$ . If  $random > p_a$ , change the position of the nest randomly to obtain a new set of positions.
7. Find the best nest position in Step 6. Stop searching when the maximum iteration limit is reached, and output the best position to achieve the optimal parameter value; otherwise, return to Step 3.

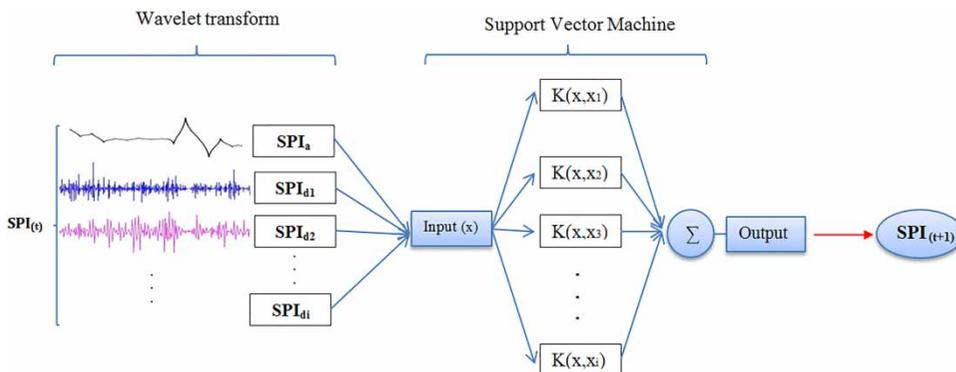


Figure 3 | The architecture of the hybrid WSVM model.

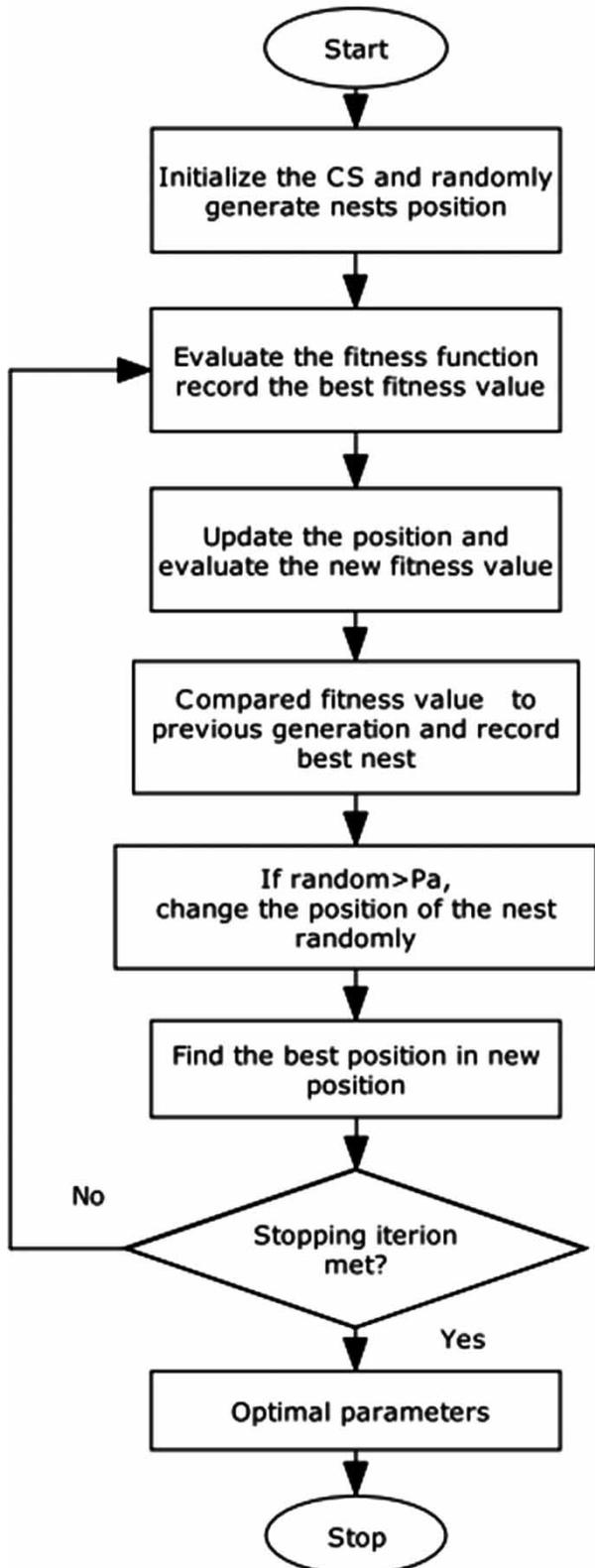


Figure 4 | Flow chart of CS algorithm for SVM parameter selection (Jiang et al. 2016).

### Case study

Urmia Lake is a lake in northwestern Iran and is reportedly the largest lake in the Middle East (between 45°03'00" and 37°40'00" east longitude and north latitude, respectively). It covers an area varying from 5,200 to 6,000 km<sup>2</sup>. The lake is about 140 km long and 40–55 km wide with a maximum depth of 16 m (Figure 5). The lake's water levels are below the critical level and the groundwater levels in some parts of the basin have decreased by 16 meters. The annual mean temperature of the basin is 11 °C and 2.5 °C around the lake and in the mountainous areas, respectively. In this study, the average monthly precipitation from eight precipitation stations (illustrated in Figure 5) were prepared during 1969–2009 for developing the Urmia Lake simulation model. Thus, the SPI time series obtained from precipitation data in the 12-month period have been computed (Figure 6).

### Efficiency criteria

The data set was divided into two parts: first, 75% of total data were used as the training set and second, 25% were used for verifying the WSVM and CS-SVM models. Also, the determination coefficient (*DC*) and root mean square error (*RMSE*) have been used to compare the efficiency of different models:

$$DC = 1 - \frac{\sum_{i=1}^N (SPI_{obs_i} - SPI_{com_i})^2}{\sum_{i=1}^N (SPI_{obs_i} - SPI_{obs})^2} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (SPI_{obs_i} - SPI_{com_i})^2}{N}} \quad (6)$$

where *DC* stands for DC, *RMSE* indicates root mean squared error, *N* denotes the number of observations, *SPI<sub>obs<sub>i</sub></sub>* is observed data, *SPI<sub>com<sub>i</sub></sub>* is forecasted values, and the mean of observed data are given by  $\overline{SPI}_{obs}$ . Moreover, Equation (7) can be used to compare the ability of different models in capturing the peak values in SPI time series similar to Equation (5) for

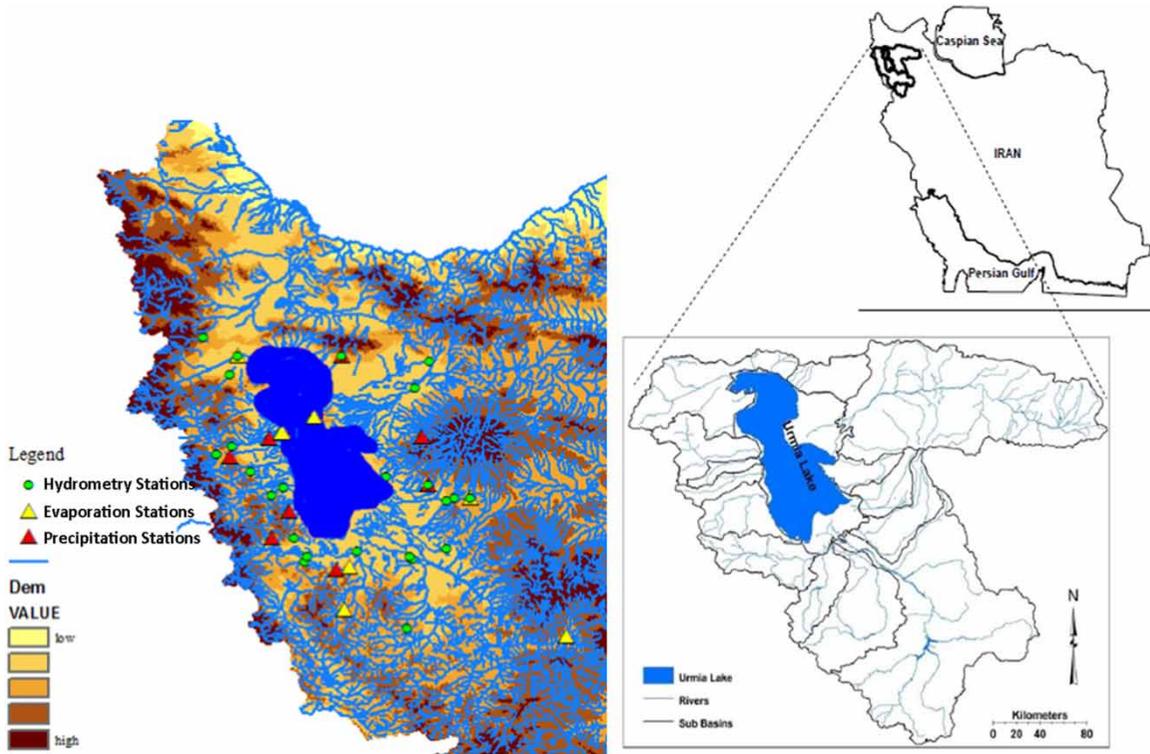


Figure 5 | Location of Urmia Lake and drainage sub-basins.

the total data.

$$DC_{peak} = 1 - \frac{\sum_{i=1}^N (SPI_{PCi} - SPI_{POi})^2}{\sum_{i=1}^N (SPI_{POi} - SPI_{PO})^2} \quad (7)$$

where  $DC_{peak}$  is the DC for the peak values.  $N$  is the number of peak values and  $SPI_{poi}$ ,  $SPI_{pci}$ , and  $\overline{SPI}_{obs}$ .

are the observed data, computed values, and mean of observed data for peak values, respectively. The  $RMSE$  was used to measure the forecast accuracy which produces a positive value by squaring the errors. The  $RMSE$  increases from zero for perfect forecasts through large positive values as the discrepancies between forecasts and observations become increasingly large.

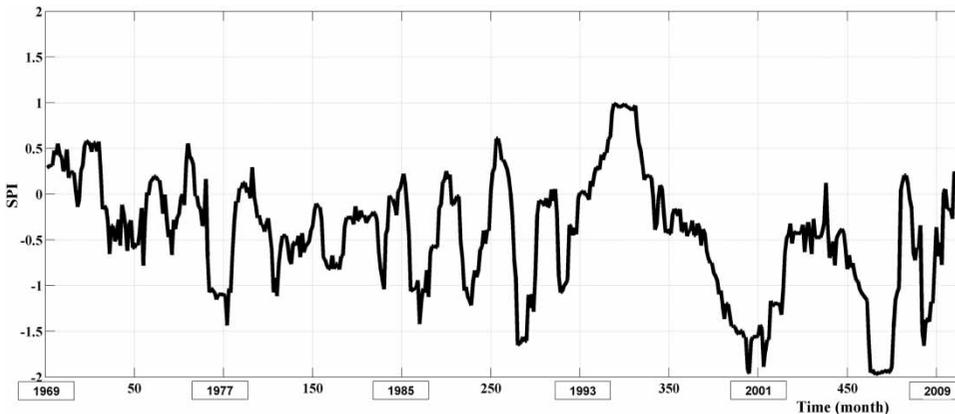


Figure 6 | SPI time series of Urmia Lake watershed.

Obviously, a high value for *DC* (up to one) and a small value for *RMSE* indicate high performance of the model.

## RESULTS AND DISCUSSION

The SVM model consists of some structure parameters like the kernel function with various types. Hence, in the first step, the purpose is to select the best kernel function. In the next step, the WT is joined with the SVM model for choosing the best wavelet function and decomposition level.

### Results of single SVM

The SVM model uses a device called kernel mapping to map the data in input space to a high-dimensional feature space in which the problem becomes linearly separable. 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). There are different kinds of kernel function. In the first step, the purpose is to achieve the best function as a high efficiency of data modeling compared with the other kernel functions. Hence, the performance of four kinds of kernel functions to forecast SPI was evaluated in terms of *DC* and *RMSE* criteria. The results are illustrated for different kernel functions in Table 2. Because SPI data are the normalized dimensionless ratios, *RMSE* values are dimensionless numbers.

As is clear in Table 2, comparing *DC* and *RMSE* values in the verification (simulation) phase reveals the radial basis function (RBF)-kernel function to have a better performance than the other kernel functions. Therefore, the RBF-kernel

function was used for the next modeling process. RBF-kernel function is utilized in this paper. This kernel function has suitable characteristics, as follows (Xu et al. 2012):

- a. In contrast with linear kernel, RBF kernel can hold the case when the communication between class labels and attributes is nonlinear.
- b. RBF-kernel function has a smaller amount of hyper parameters which control the difficulty of model selection. Also, it has less numerical complexities.

Now, the details of the best kernel function, namely, RBF kernel are explained. In machine learning, the (Gaussian) radial basis function kernel, or RBF kernel, is a popular kernel function used in SVM classification  $k(x, x_i) = e^{-\gamma \|x-x_i\|^2}$  which ranges between zero and one. In a more accurate explanation,  $k(x, x_i) = e^{-(1/2\sigma^2)\gamma \|x-x_i\|^2}$  so, the value of  $\gamma$  in the first equation is  $\gamma = -(1/2\sigma^2)$ .  $x$  and  $x_i$  denote two sample data (Wang 2005). 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 then needs to be tuned on a validation or tuning data set (Vert et al. 2004).

In the next step, the previous monthly SPI time data ( $SPI_{(t-1)}, SPI_{(t-2)}, SPI_{(t-3)} \dots$ ) of Urmia Lake were used to forecast the  $SPI_{(t+1)}$  (i.e., the SPI of the next month). The goal of time series prediction or forecasting can be formulated as follows (Babovic & Keijzer 2000):

$$SPI_{(t+1)} = f_{svm}(SPI_{(t)}, SPI_{(t-1)}, SPI_{(t-2)}, \dots) \tag{8}$$

It is supposed that the series is some sampling of a continuous system which may be either stochastic, chaotic, or deterministic. Thus, the following input combinations were defined as Combs. (1)–(6) as depicted below:

- Comb. (1):  $SPI_{(t)}$
- Comb. (2):  $SPI_{(t)}, SPI_{(t-1)}$
- Comb. (3):  $SPI_{(t)}, SPI_{(t-1)}, SPI_{(t-2)}$
- Comb. (4):  $SPI_{(t)}, SPI_{(t-1)}, SPI_{(t-2)}, SPI_{(t-3)}$
- Comb. (5):  $SPI_{(t)}, SPI_{(t-1)}, SPI_{(t-2)}, SPI_{(t-3)}, SPI_{(t-4)}$
- Comb. (6):  $SPI_{(t)}, SPI_{(t-1)}, SPI_{(t-2)}, SPI_{(t-3)}, SPI_{(t-4)}, SPI_{(t-5)}$

Like the previous step, to find the best result, each Comb. was modeled and its performance was assessed with *DC* and *RMSE* values (Table 3).

Table 2 | Result of SVM model for finding the best kernel function

Kernel function	Calibration		Verification	
	DC	RMSE	DC	RMSE
RBF-kernel	0.883	0.154	0.865	0.237
Poly-kernel	0.765	0.302	0.761	0.341
MLP-kernel	0.791	0.210	0.771	0.302
Lin-kernel	0.784	0.295	0.698	0.386

**Table 3** | Result of SVM model with different Combs of input data by RBF-kernel function

Combs	Calibration		Verification	
	DC	RMSE	DC	RMSE
Comb. (1)	0.781	0.323	0.763	0.341
Comb. (2)	0.883	0.190	0.865	0.237
Comb. (3)	0.820	0.201	0.856	0.241
Comb. (4)	0.786	0.305	0.761	0.345
Comb. (5)	0.887	0.156	0.670	0.447
Comb. (6)	0.888	0.144	0.641	0.455

As Table 3 suggests, Comb. (2) shows the best efficiency as compared to the other Combs. Also, it can be concluded from Table 3 that high degrees of data combination, i.e., Comb. (5) and Comb. (6), resulted in over-fitting (over-training) in the calibration step such that the accuracy of results was reduced in the verification step. This item is in accordance with the results of other research when the autoregressive rule is used (Nourani et al. 2011a, 2011b, 2012). Simulated and actual SPI time series of Urmia watershed which were obtained by RBF-kernel function with Comb. (2) of input data are shown in Figure 7.

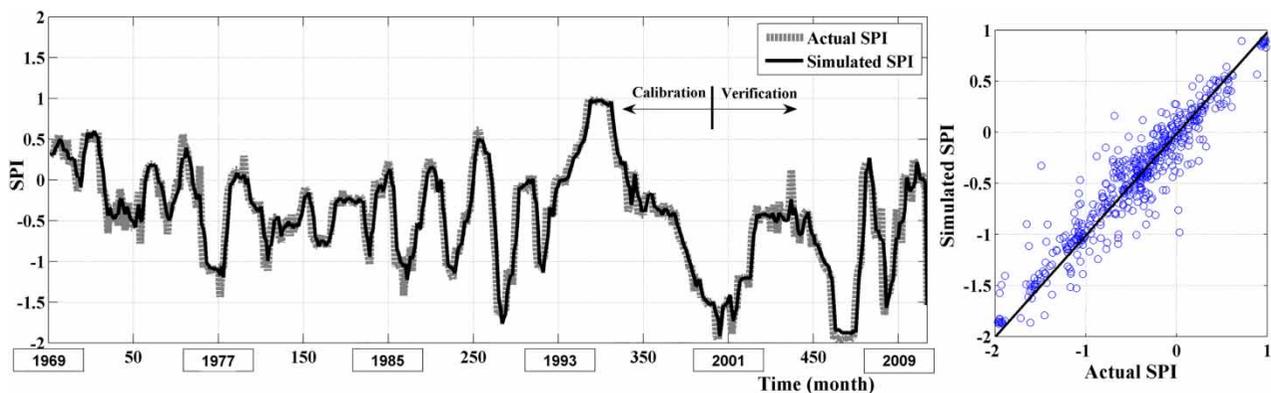
### Results of WSVM hybrid model

In this section, the effects of WT on data modeling are determined. For this purpose, the pre-processing of SPI data by a wavelet was inserted in the SVM model and the performance of the model was determined by DC and RMSE criteria. Hence, time series were decomposed to one to

seven levels by four different kinds of wavelet transforms, i.e., 1-Sym3 wavelet with three sharp peaks, 2-Daubechies wavelet (Db2), a most popular wavelet, 3-Haar wavelet, a simple wavelet, and 4-Coif1 wavelet. Hence, the SPI data were decomposed in seven sub-signals by the WT. Then, the decomposed SPI data play the roles of SVM inputs and the performance of each case was determined by the mentioned criteria (Table 4).

Based on Table 4, the Coif1 wavelet function and decomposition level 5 produced better results. According to the structure of Coif1 wavelet (Figure 1(d)) which is similar to the SPI signal, it could capture the signal features, especially peak values, and yield comparatively high efficiency. The reason for the superiority of decomposition level 5 may be hidden in the fact that decomposition level 5 includes a  $2^5$ -month mode which is approximately nearby annual mode. This periodicity is very important in a hydrologic time series, referring to the fact that the drought growth depends on the annual changes in Urmia Lake watershed. According to Table 4, high degrees of data decomposition (6 and 7) resulted in over-fitting in the calibration step so that the accuracy of results was reduced in the verification step as in the previous section. The calibration and verification time series of WSVM modeling with the best results, having high efficiency on data modeling, are shown by Figures 8 and 9.

By comparing the SVM and WSVM modeling results, it is concluded that using multi-resolution SPI as the input in the SVM algorithm leads to achieving better results as compared to the results obtained when using raw SPI data. In other words, when the WT converts SPI data into multi-

**Figure 7** | Simulated and actual SPI time series for Comb. (2) as input data and RBF-kernel function.

**Table 4** | Result of the WSVM model by RBF-kernel function with different mother wavelets and decomposition levels

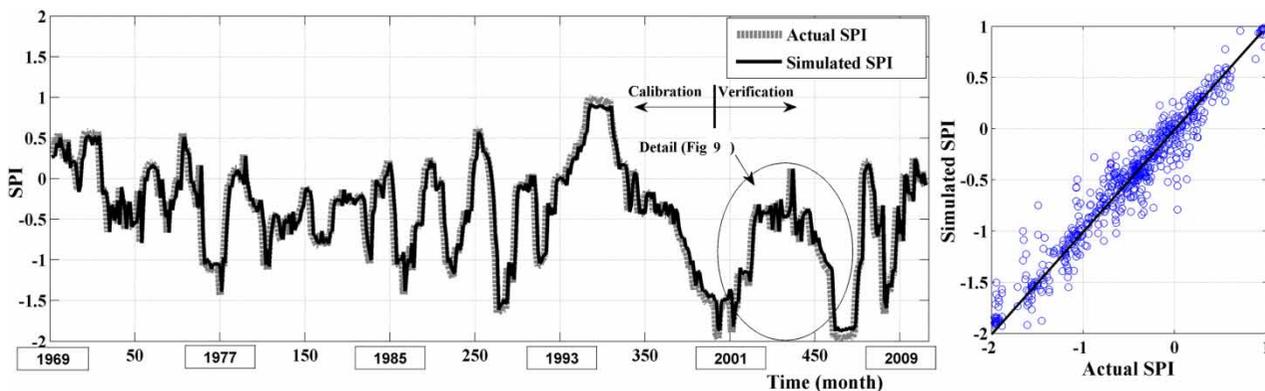
Mother wavelet type	Decomposition level	Calibration		Verification	
		DC	RMSE	DC	RMSE
Sym3	1	0.926	0.058	0.890	0.072
Sym3	2	0.926	0.058	0.890	0.072
Sym3	3	0.926	0.058	0.892	0.071
Sym3	4	0.926	0.058	0.892	0.071
Sym3	5	0.924	0.059	0.890	0.072
Sym3	6	0.932	0.057	0.377	0.172
Sym3	7	0.933	0.057	0.375	0.174
Db2	1	0.936	0.053	0.890	0.072
Db2	2	0.940	0.052	0.909	0.065
Db2	3	0.941	0.052	0.907	0.066
Db2	4	0.942	0.051	0.909	0.065
Db2	5	0.942	0.051	0.910	0.065
Db2	6	0.953	0.051	0.020	0.216
Db2	7	0.955	0.051	0.019	0.218
Haar	1	0.926	0.058	0.890	0.072
Haar	2	0.926	0.058	0.890	0.072
Haar	3	0.926	0.058	0.892	0.071
Haar	4	0.926	0.058	0.892	0.071
Haar	5	0.924	0.059	0.890	0.072
Haar	6	0.936	0.053	0.377	0.172
Haar	7	0.938	0.052	0.375	0.173
Coif1	1	0.928	0.057	0.929	0.058
Coif1	2	0.931	0.056	0.930	0.057
Coif1	3	0.932	0.055	0.930	0.057
Coif1	4	0.932	0.055	0.931	0.057
Coif1	5	0.943	0.055	0.954	0.056
Coif1	6	0.983	0.043	0.780	0.102
Coif1	7	0.984	0.042	0.771	0.101

resolution data, the SVM algorithm can perform data regression with minimum error. This process was done by the kernel function (see Figure 10). In this regard, Figure 10 schematically shows the regression on raw and decomposed SPI data by the SVM algorithm. As is clear, the regression accuracy of multi-resolution data may be better than raw ones.

The most important factor in drought management is the determination of extreme values in SPI time series which indicate the further potentials of dryness and drought. Since the feasible estimation of the peak values is usually the most important factor in any water resource management, a key point when comparing different models is the capability of the models in estimating peak values. For this purpose, peak values were sampled by considering the threshold of the top 5% of the data from the original SPI time series contractually. The results show that for the SVM modeling, the  $DC_{peak}$  value was 0.62 while for WSVM modeling, it was given as 0.93. According to the  $DC_{peak}$  values, it can be concluded that the seasonal model (i.e., WSVM) is more efficient than the autoregressive model which means SVM obtains the extreme values. It is obvious that the extreme values in SPI time series which occur in a periodic pattern can be correctly calculated by the seasonal model.

**Results of CS-SVM model**

SVM hyper-parameter  $C$  and  $\epsilon$  (Equation (2)) are two main free parameters of the SVM algorithm that should appropriately be set by a user. To accomplish this aim, SPI time



**Figure 8** | Simulated and actual SPI series by Coif1 wavelet function and decomposition level of five.

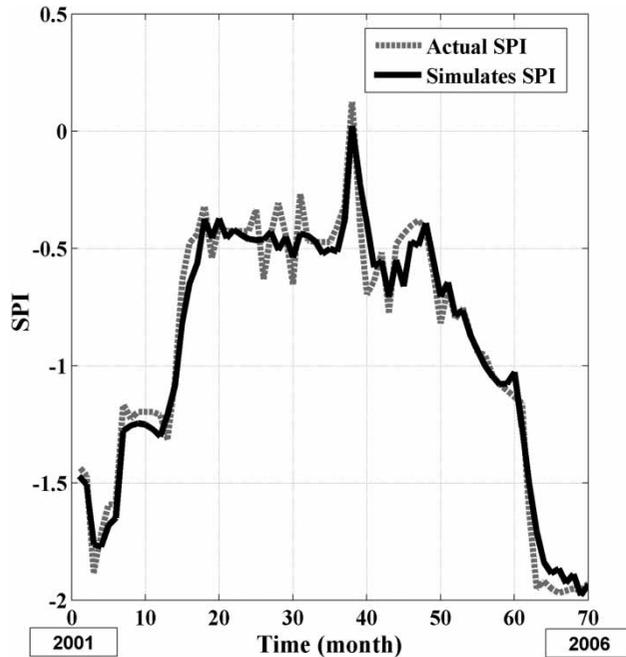


Figure 9 | Simulated and actual SPI series specified in Figure 8.

series is modeled by CS-SVM algorithm in MATLAB programming. Comb. (2) and RBF-kernel function is selected based on previous sections. In this way, the parameter optimizing and SPI modeling were carried out for 15 iterations. As illustrated in Figure 11, the value of the RMSE as fitness function is plotted against the number of iterations. The fitness function values converge to the desired value (minimum value) so that the SVM parameters are finally optimized to be selected as follows:  $C = 901$ ;  $\varepsilon = 0.95$ . It is

notable that over-fitting problem is likely to occur in higher iteration values than 15.

The results of verifying the CS-SVM model are obtained as follows. DC and RMSE criteria are calculated to be 0.899 and 0.093, respectively, in the calibration step and 0.912 and 0.065, respectively, in the verification step. While, for the single SVM model, DC and RMSE are about 0.865 and 0.237, respectively, in the verification step. The finding provides evidence that CS-SVM shows more efficiency than the SVM model due to using the CS parameter optimizer in the SVM algorithm. Figure 12 illustrates the simulated and actual SPI time series of Urmia Lake watershed modeled by CS-SVM.

In the final part, the single SVM, hybrid WSVM, and CS-SVM models are compared with each other, as shown in Table 5.

As a result, the hybrid model showed a better performance than the single one. In addition, the CS-SVM model is more efficient than SVM model according to appropriate setting of SVM parameters  $C$  and  $\varepsilon$ . The optimized parameters lead to fitting the model to general patterns in the time series. In other words, the trained model based on optimized parameters can recognize and reflect the fluctuations in time series data. Hence, the over-fitting problem is not likely to occur in calibration data. Since the length of data in the verification period is less than the calibration period, the accuracy improvement for the verification step is more than calibration step. However, this type of improvement cannot be a general rule for every time series modeling

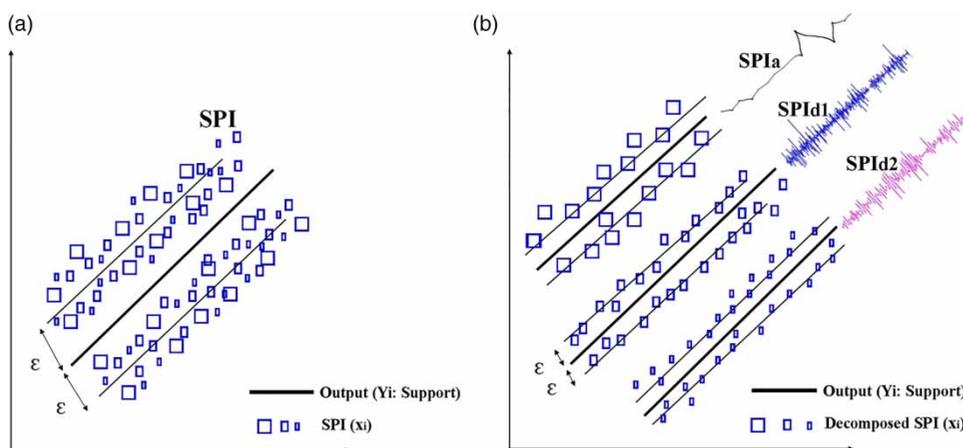


Figure 10 | SPI raw data (a) and multi-resolution data (b) regression.

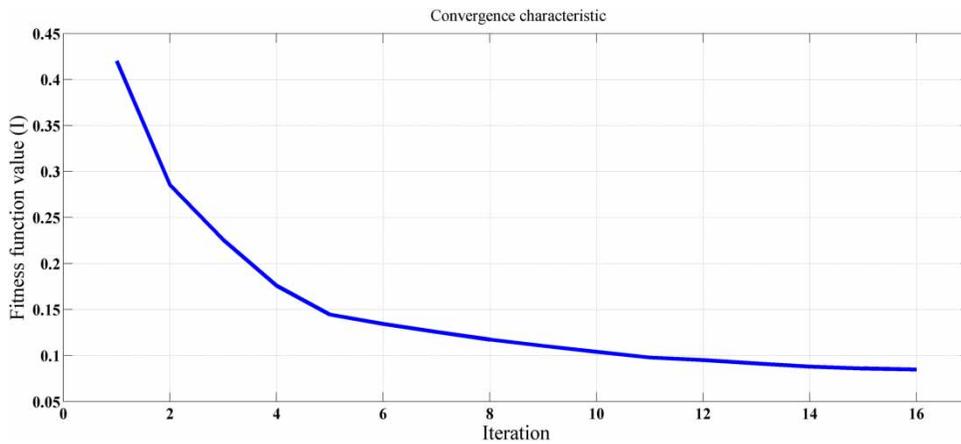


Figure 11 | Fitness function values convergence through 15 iterations.

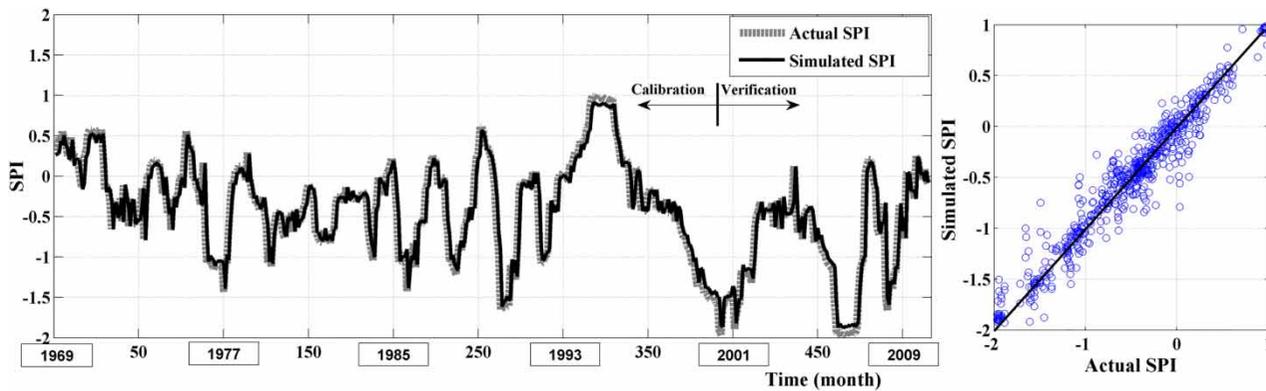


Figure 12 | Simulated and actual SPI series modeled by CS-SVM.

Table 5 | Comparison of SVM, WSVM, and CS-SVM modeling

Model	Model type	DC calibration	DC verification
SVM	Single	0.883	0.865
WSVM	Hybrid	0.943	0.954
CS-SVM	Parameters optimized	0.899	0.912

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

and it is dependent on the existing patterns in calibration and verification data sets.

## CONCLUSIONS

In this paper, models including single SVM, CS-SVM, and WSVM were employed for forecasting one of the drought indices called SPI at one lead time for Urmia Lake

watershed. In the first step, the purpose was to select the best kernel function. Thus, four kinds of kernel functions were used and the modeling performance of each function was determined by DC and RMSE criteria. Results show that the RBF-kernel function is more accurate than the others. Hence, the RBF-kernel function was used in the next steps of the study. Afterwards, the following input combinations (Combs. (1)–(6)) were used as inputs of the SVM model to forecast  $SPI_{(t+1)}$ . Then, the performance of each case of data combinations was examined. The results showed that Comb. (2) has better compatibility as the SVM input. In the next section of this study, a WT was designed. To achieve this goal, the SPI data were decomposed to the sub-signals with different degrees of decomposition and WT functions. Then, these sub-signals were used as the inputs of the SVM algorithm, the results of DC and RMSE criteria were calculated and the

best-performing wavelet transformation and decomposition approaches were delineated. Accordingly, Coif1 WT with five degrees of decomposition level presented good results. Also, the overall results revealed the WSVM's superiority in assessing the extreme data in SPI time series and in observing long-term SPI by considering the seasonality effects. Furthermore, the hybrid model was shown to be more appropriate because it used the multi-scale time series of SPI data in the SVM model. Meanwhile, the evolutionary CS algorithm was applied to optimize the  $C$  and  $\epsilon$  parameters of the SVM model, resulting in an improvement of accuracy of the SVM model.

As a suggestion for further studies, it is recommended to use the presented methodology to forecast the SPI 2, 3 and more lead times, and also to model the SPI data of the watershed by adding other time series of hydrologic variables, such as temperature. In addition, a further study could assess the effect of another evolutionary algorithm, such as ant colony optimization algorithm, to optimize constant parameters of the SVM model.

## REFERENCES

- Abe, S. 2010 *Support Vector Machines for Pattern Classification*, (Vol. 2). Springer, London.
- Abramowitz, M. & Stegun, A. 1965 *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables* (Vol. 55). Courier Corporation, New York.
- Babovic, V. & Keijzer, M. 2000 Forecasting of river discharges in the presence of chaos and noise. *Nato Science Series 2 Environmental Security* **71**, 405–420.
- Barbosa, C. E. M. & Vasconcelos, G. C. 2016 Cuckoo search optimization for short term wind energy forecasting. In: *IEEE Congress on Evolutionary Computation (CEC)*, Vancouver, BC, pp. 1765–1772.
- Behzad, M. K., Asghari, M., Eazi, M. & Palhang, M. 2009 Generalization performance of support vector machines and neural networks runoff modeling. *Expert Systems with Applications* **36** (4), 7624–7629.
- Belayneh, A. & Adamowski, J. 2012 Standard precipitation index drought forecasting using neural networks wavelet neural networks, and support vector regression. In: *Applied Computational Intelligence and Soft Computing*, p. 6.
- Belayneh, A. & Adamowski, J. 2013 Drought forecasting using new machine learning methods. *Journal of Water and Land Development* **18** (9), 3–12.
- Blagojevic, B., Srdjevic, Z., Bezdan, A. & Srdjevic, B. 2016 Group decision-making in land evaluation for irrigation: a case study from Serbia. *Journal of Hydroinformatics* **18** (3), 579–598.
- Cacciamani, C., Morgillo, A., Marchesi, S. & Pavan, V. 2007 Monitoring and forecasting drought on a regional scale: Emilia-Romagna region. In: *Methods and Tools for Drought Analysis and Management* (G. Rossi, T. Vega & B. Bonaccorso, eds). Water Science and Technology Library book series, Vol. 62, pp. 29–48.
- Cannas, B., Fanni, A., See, L. & Sias, G. 2006 Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. *Physics and Chemistry of the Earth* **31** (18), 1164–1171.
- Darlane, A. B. & Azimi, S. 2017 Streamflow forecasting by combining neural networks and fuzzy models using advanced methods of input variable selection. *Journal of Hydroinformatics* **20** (2), 520–532. doi:10.2166/hydro.2017.076.
- Dibike, Y. B., Velickov, S., Solomatine, D. & Abbott, M. B. 2001 Model induction with support vector machines: introduction and applications. *Journal of Computing in Civil Engineering* **15** (3), 208–216.
- Haykin, S. 2003 *Neural Networks: A Comprehensive Foundation*. Prentice Hall, New Delhi.
- Huang, F., Huang, J., Jiang, S. H. & Zhou, C. 2017 Prediction of groundwater levels using evidence of chaos and support vector machine. *Journal of Hydroinformatics* **19** (4), 586–606.
- Jalalkamali, A., Sedghi, H. & Manshour, M. 2011 Monthly groundwater level prediction using ANN and neuro-fuzzy models: a case study on Kerman plain, Iran. *Journal of Hydroinformatics* **13** (4), 867–876.
- Jiang, M., Luo, J., Jiang, D., Xiong, J., Song, H. & Shen, J. 2016 A cuckoo search-support vector machine model for predicting dynamic measurement errors of sensors. *IEEE Access* **4**, 5030–5037.
- Lin, G., Chen, G., Huang, P. & Chou, Y. 2009 Support vector machine-based models for hourly reservoir inflow forecasting during typhoon-warning periods. *Journal of Hydrology* **372** (1), 17–29.
- Mallat, S. G. 1998 *A Wavelet Tour of Signal Processing*, 2nd edn. Academic Press, San Diego, CA.
- McKee, T. B., Doesken, N. J. & Kleist, J. 1993 The relationship of drought frequency and duration to time scales. In: *Proceedings of the 8th Conference on Applied Climatology*, American Meteor Society, Anaheim, CA, pp. 179–184.
- Nikbakht Shahbazi, A., Zahraei, B., Sadghi, H., Manshour, M. & Nasser, M. 2011 Seasonality meteorological drought prediction using support vector machine. *World Applied Sciences Journal* **13** (6), 1387–1397.
- Noori, R., Deng, Z., Kiaghadi, A. & Kachooangi, F. T. 2015 How reliable are ANN, ANFIS, and SVM techniques for predicting longitudinal dispersion coefficient in natural rivers. *Journal of Hydraulic Engineering* **142** (1), 04015039.
- Nourani, V., Goli Ejlali, R. & Taghi Alami, M. 2011a Spatiotemporal groundwater level forecasting in coastal aquifers by hybrid artificial neural network-geostatistics model: a case study. *Environmental Engineering Science* **28** (3), 217–228.

- Nourani, V., Kisi, Ö. & Komasi, M. 2011b Two hybrid artificial intelligence approaches for modeling rainfall-runoff process. *Journal of Hydrology* **402** (1), 41–59.
- Nourani, V., Komasi, M. & Taghi Alami, M. 2012 Hybrid wavelet-genetic programming approach to optimize artificial neural network modeling of rainfall-runoff process. *Journal of Hydrology* **17** (6), 724–741.
- Sannasiraj, S. A., Zhang, H., Babovic, V. & Chan, E. S. 2004 Enhancing tidal prediction accuracy in a deterministic model using chaos theory. *Advances in Water Resources* **27** (7), 761–772.
- Sun, Y., Babovic, V. & Chan, E. S. 2010 Multi-step-ahead model error prediction using time-delay neural networks combined with chaos theory. *Journal of Hydrology* **395** (1), 109–116.
- Tsakiris, G. & Vangelis, H. 2004 Towards a drought watch system based on spatial SPI. *Water Resources Management* **18** (1), 1–12.
- Vapnik, V. N. & Cortes, C. 1995 Support vector networks. *Machine Learning* **20** (3), 273–297.
- Vert, J. P., Tsuda, K. & Schölkopf, B. 2004 *A Primer on Kernel Methods. Kernel. Methods in Computational Biology*. MIT Press, Cambridge, MA, pp. 35–70.
- Vojinovic, Z., Kecman, V. & Babovic, V. 2003 Hybrid approach for modeling wet weather response in wastewater systems. *Journal of Water Resources Planning and Management* **129** (6), 511–521.
- Wang, L. 2005 *Support Vector Machines: Theory and Applications*. Springer Science & Business Media, Berlin, p. 177.
- Wang, W., Men, C. & Lu, W. 2008 Online prediction model based on support vector machine. *Neuro Computing* **71** (4), 550–558.
- Xu, Y., Chen, X. & Li, Q. 2012 INS/WSN-Integrated navigation utilizing LS-SVM and  $H_{\infty}$  filtering. *Mathematical Problems in Engineering* **20** (12), 1–19.
- Yang, X. S. & Deb, S. 2009 Cuckoo search via Lévy flights. In: *World Congress on Nature & Biologically Inspired Computing*, Coimbatore, India.
- Yang, X. S. & Deb, S. 2013 Multi-objective cuckoo search for design optimization. *Computers & Operations Research* **40** (6), 1616–1624.
- Yu, X., Liong, S. Y. & Babovic, V. 2004 EC-SVM approach for real-time hydrologic forecasting. *Journal of Hydroinformatics* **6** (3), 209–223.
- Zahraei, B. & Nasser, M. 2014 Basin scale meteorological drought forecasting using support vector machine. In: *International Conference on Drought Management Strategies in Arid and Semi-Arid Regions*.
- Zanganeh, M., Yeganeh-Bakhtiary, A. & Yamashita, T. 2016 ANFIS and ANN models for the estimation of wind and wave-induced current velocities at Joetsu-Ogata coast. *Journal of Hydroinformatics* **18** (2), 371–391.

First received 17 September 2017; accepted in revised form 20 February 2018. Available online 17 April 2018