Measurement 103 (2017) 294–301

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

# A hybrid wavelet-support vector machine model for prediction of Lake water level fluctuations using hydro-meteorological data

Basant Yadav *, Kh. Eliza

*Department of Civil Engineering, Indian Institute of Technology, Delhi, India*

ARTICLE INFO

*Article history:*
Received 16 January 2017
Received in revised form 28 February 2017
Accepted 1 March 2017
Available online 2 March 2017

*Keywords:*
WAVE-SVM
Lake water level
Prediction

ABSTRACT

Prediction of Lake water level fluctuation is immensely important to maintain the ecological conditions of the lake as well as for the development and management of water resources. Loktak Lake, a Ramsar convention recognised wetland have a unique ecosystem and is known for its floating vegetated islands. Fluctuation of water level in Loktak Lake primarily due to hydropower and agriculture have impacted the lake ecosystem significantly. The study employed a hybrid wavelet-support vector machine (WA-SVM) model to predict the daily lake water fluctuation. Daily lake water level along with other hydro-meteorological data have been used as input to predict the lake water level up to 20 days ahead. Various combinations of input data have been experimented to obtain the best model structure which is later extended to predict the lake water level. The predicted lake water are found to be close to observed values.



## 1. Introduction

Groundwater is an important water resource and its timely prediction with reasonable accuracy could provide help to policy makers and engineers. Prediction of daily lake water fluctuation is primarily important for hydropower plants, commercial navigation, domestic, agricultural, and industrial activities in many countries [14]. Mohanty et al. [21] suggest that an overall groundwater management strategy depends on various factors, including availability of accurate information, financial support, policy framing and implementation. However, another key component of groundwater management is being able to forecast groundwater levels with a high degree of accuracy. The precise prediction of daily lake water fluctuation can help policymakers resolve the best approach to groundwater management problems [10,12]. In past, various studies have been conducted to simulate and predict the groundwater level fluctuation of lakes, however, the interplay of many hydro-meteorological factors such as evaporation, rainfall, evapotranspiration makes it complex and highly nonlinear [37,12].

In the past decade, various studies have been conducted to investigate the advantages and disadvantages of conceptual-based models. Subsequently, their performance has been compared with data-driven models such as artificial neural networks (ANN) [19,9,20,22]. A comparative study suggests that conceptual-based models require many parameters for calibration and have large computation times. In practice, however, poor model performance and the associated uncertainties of such models can be attributed to data collection cost and time as well as inaccessibility of sites [15]. Data-driven models are able to develop interrelationships between input-output variables and are able to generalize complex phenomena with high accuracy and minimal computation time [23,38]. Adamowski and Chan [2] brought attention to the relative importance of numerical models (for capturing the complex underlying processes) and data-driven models (for accurate prediction with limited data set) when modeling a particular site. Further, the availability of sufficient and accurate data for numerical model development is also a common problem [39].

Recently, the use of machine learning approaches such as artificial neural networks (ANN), support vector machines (SVM), adaptive neuro-fuzzy inference system (ANFIS) and genetic programming (GP), have been proved efficient for use in water resources problems [32,10,29]. The application of GP in many water management problems suggest that it reduces the computational time significantly and yields results with acceptable accuracy [11]. Further, ANN has been successfully used in hydrologic field to predict various nonlinear and complex phenomenon such as precipitation, saltwater intrusion, groundwater level, lake water level, and river discharge [26,3,22,10,12,29].

SVM model is also as efficient as ANN in modeling nonlinear systems. The basis of support vector machines is developed by Vapnik [35]. The SVM approach overcomes common problems

* Corresponding author.
*E-mail address:* basant1488@gmail.com (B. Yadav).

http://dx.doi.org/10.1016/j.measurement.2017.03.003
0263-2241/

associated with ANN (local minimum and the network over fitting) as it is based on structural risk minimization (SRM) instead of the empirical risk minimization (ERM) concept of ANN. SVM model has been used by many researchers to solve hydrology, hydrogeology problems. The application of SVM in water resources problems is recent and has been found to be more efficient than the traditional soft computing techniques [4,39,41,25,13,33,12,38]. Further, a comparative study among ANN and SVM to predict the lake water level suggest that the results obtained by SVM is found to be better than those when ANN is used [8].

Though, ANN, SVM and other data based model have been successfully used to predict various hydrologic problems, these methods frequently have limitations with non-stationary data [5,34,2]. To handle such problems a method called wavelet analysis has been used in various hydrological studies. The non-stationarities of time series can be analyzed by using mathematical functions called wavelets that decompose a given time series into various components which later can be used as input for data based model like ANN [2]. Kisi [16,17] used wavelet decomposition in a hybrid mode with ANN to forecast daily and monthly flow. Shiri and Kisi [27,28] used wavelet decomposition to study the streamflow forecasting and suspended sediment load estimation. Adamowski and Sun [1] developed a wavelet neural network conjunction model to forecast daily river flow at various time leads. Adamowski and Chan [2] predicted groundwater level fluctuation using WA-ANN model and suggested further investigation of this technique. Kisi and Cimen [18] used wavelet support vector machine model for monthly streamflow prediction of Canakdere and Goksudere River, in Eastern Black Sea region of Turkey. Suryanarayana et al. [33] used WA-SVR model to predict monthly groundwater level and compared the results with SVR, ANN and ARIMA models. In general, the conducted studies concluded that the application of wavelet transform with data based models like ANN and SVM improves simulation and prediction accuracy.

Prediction of Lake water level fluctuation is of immense importance and to date, no work has been published which explores the applicability of wavelet-support vector machine for Lake water level prediction with various hydro-meteorological input variables. The focus of this research is to examine the ability of the coupled WA–SVM method for prediction of daily lake water level fluctuation and to compare it with normal support vector machine model. In this work, WA–SVM models have been developed and compared with SVM model for lead times of 5, 10, 15 and 20 days at Loktak Lake in Manipur, India. The data used include daily rainfall (*R*), evaporation (*EV*), evapotranspiration (*ET*) and past lake water level (*Lv*) data. The simulation and predictions results obtained using WA-SVM and SVM models are compared with the observed Lake water level data and evaluated based on quantitative standard statistical analysis.

## 2. Methodology

### 2.1. Support vector machine

Vapnik et al. [36] proposed a kernel-based algorithm as support vector machine (SVM) a data driven model used in surrogate modeling for the mapping of physical processes and to obtain the second level of fidelity mainly to reduce the simulation time and cost of physical simulation. Good generalization ability, less prone to over-fitting and simultaneous minimization of estimation errors allows SVM to be used in diverse scientific fields. Qu and Zuo [24] discussed the utility of kernel functions in SVM which make the original inputs linearly separable in mapped high dimensional feature space. Let $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ be assumed to be the given training data sets, where $x_i \subset R^n$ represents the input sample space and $y_i \subset R$ for i = 1, ..., $l$ denotes respective target output, elements in the training data set represented by $l$. Error tolerance level is fixed by a value of $\epsilon$ (errors < $\varepsilon$). The linear regression in SVM is estimated by solving Eq. (1).

$$\text{Minimize} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=0}^{n}(\xi + \xi^*) \tag{1}$$

$$\text{Subject to} \quad \begin{cases} y_i - \langle w, x_i \rangle - b \leqslant_i + \xi \\ \langle w, x_i \rangle + b - y_i \leqslant_i + \xi^* \\ \xi_i \xi_i^* \geqslant 0 \quad i = 1, \ldots, l \end{cases} \tag{2}$$

$w$ denotes the normal vector, $b$ is a bias, $C$ represents a regularization constant, $\epsilon$ is the error tolerance level of the function, and the $\xi$, $\xi^*$ are slack variables.

By converting the primal formulation into dual formulation using the mathematical manipulations the dual problem kept as an objective function for the quadratic programming used for solving SVM. The output of the SVM is critically dependent on the parameters as regularization constant ($C$), insensitive loss function ($\varepsilon$), and parameter of radial basis function ($\gamma$). Trial and error procedure was used in the present study to optimize these parameters based on the RMSE value. The trial continues by using different combinations of all three parameters till the value of RMSE was minimized.

### 2.2. Wavelet analysis

The data driven model depends on the hidden information in the time series used for the analysis of any physical phenomenon but sometimes due to variability in the frequency of the data, extraction of the relevant information becomes difficult. Wavelet decomposition is a technique used in case of non-periodic and transient signals to extract the relevant time-frequency information by disintegrating the data into low frequency and high frequency components. Considering the non-stationary behaviour of the Lake water level data series we have selected the Daubechies wavelets which is preferred for handling the non-stationary series. In this study we have employed the discrete wavelet transform (DWT)

$$Lv_{mn}^{w} = 2^{-(m/2)}\sum_{t=0}^{N-1} Lv_t W\left(\frac{t - n \cdot 2^m}{2^m}\right) \tag{3}$$

$$= 2^{-(m/2)}\sum_{t=0}^{N-1} Lv_t W_{mn}(t), \tag{4}$$

$W(\cdot))$ is the selected wavelet function, $Lv_t$ denotes the observed daily (t) Lake water level, $N$ is the length of the time series, $Lv_{mn}^{w}$ is the decomposition coefficient, m and n is the resolution level and position. At each resolution level the number of coefficient is obtained by $N/2^m$ considering N is divisible by $2^m$.

A multi resolution techniques has been used in this study where a father wavelet function used for the extraction of low frequency components while the high frequency component is extracted by using a complementary of the father wavelet, a mother wavelet function. The mathematical decomposition of the data series is represented by the approximation series $A_m$ and the detail series $D_m$.

$$A_m = \sum_n Lv_{mn}^{\Phi}\varphi_{mn}(t); \quad m = 1, \ldots, M \tag{5}$$

and

$$D_m = \sum_n Lv_{mn}^{\Psi}(t); \quad m = 1, \ldots, M \tag{6}$$

$\varphi_{mn}(t)$ denotes the father wavelet functions while $\Psi_{mn}(t)$ denotes the mother wavelet functions, and $Lv_{mn}^{\Phi} and Lv_{mn}^{\psi}$ are coefficients obtained through Eq. (2). The original Lake water level time series can be reconstructed as given in Eq. (7)

$$Lv_t = D_1 + \cdots + D_M + A_M \tag{7}$$

### 2.3. Case study

Loktak Lake is the largest freshwater wetland located in the state of Manipur, India (Fig. 1). This internationally recognised wetland (Ramsar Convention) covers an area of 287 $km^2$ and has a total catchment area of 4947 $km^2$. The Manipur River, which drains from the lake, is a tributary of the much larger Irrawaddy basin (total area 413,710 $km^2$). The lake is located within a central valley covering 28% of the total Loktak catchment. The elevation of the catchment varies from 800 m above mean sea level (a.m.s.l.) in the valley to over 2500 m a.m.s.l. in the surrounding mountains.

The catchment climate varies from tropical to semi-tropical climate at lower altitudes, while at higher altitudes it varies from semi-temperate to temperate [30]. Southwestern monsoon drives the climatic condition in this region. The time period starting from the month of June to September marks the rainy season in this region and this duration accounts almost 63% of the annual average catchment precipitation of 1409 mm (Loktak Development Authority, LDA). The mean annual temperature for meteorological stations within the valley is 20.5 °C. Mean summer (May–July) temperature is 24.0 °C, while the mean temperature in winter (November–January) is 14 °C. Mean annual potential evapotranspiration (PET) for the catchment is 1063 mm. Forest covers approximately 64% (3150 $km^2$) of the catchment, with the principal forest types comprising tropical semi-evergreen, subtropical pine and montane wet temperate forests [30].

Loktak Lake water level fluctuates between 0.5 and 4.5 m (LDA). Seasonal variations in water level indicates the seasonal variation in rainfall and, in turn, river discharge. The most prominent characteristic feature of this lake is the presence of floating heterogeneous masses of soil, vegetation and organic matter at various stages of decomposition, known locally as Phumdis. This unique feature of the lake has been included in the analysis by considering evaporation and evapotranspiration separately to conduct this study. The effect of wave exposure on the lake water level depends on both the size of the wave and the characteristics of the lake. The presence of Phumdis in the lake makes very difficult to measure the effects of wind waves on Lake water level hence the variable has not been considered in this study. The Keibul Lamjao National Park (KLNP), located in the southern part of the lake within an extensive areas of phumdis, is the only floating wildlife sanctuary in the world [30]. It is also the home of a brow-antlered deer or Sangai which is listed as the most endangered species in the International Union for Conservation of Nature (IUCN). Loktak Lake is of great importance to the socio-economy of the region. Its ecosystem services are associated with the utilization of the lake's water, fish and aquatic vegetation. In this way, it supports a human population of 279,935 living on it and around its margins [30]. The time series of rainfall, evaporation, evapotranspiration and lake water level from 1st June 2000 to 31st May 2003 were used to conduct this study. The daily Lake water level data (Fig. 2) of the Loktak Lake is provided by the Loktak Development Authority (LDA).

### 2.4. Model development

LIBSVM toolbox [6] has been used to develop the SVM models for simulation and prediction of Lake water level. Appropriate selection of kernel and its parameters generally affect the accuracy of SVM model. Researches in past have employed SVM to study various hydrologic problems considering various kernels and concluded that the radial basis function (RBF) is most accurate and efficient kernel function for SVM [31,7,40]. Therefore, the RBF is adopted in this work. The kernel parameters are obtained using a trial and error method, where the parameters such as $C$ and $\gamma$ have been used in various combinations until the error is minimized to a satisfactory level. Table 1 presents the optimal parameters of SVM while Table 2 presents the optimal parameters of WA-SVM for approximation as well as decomposed time series. The developed WA-SVM models comprises several combinations of the normalized data of daily rainfall ($R$), evaporation ($EV$), evapotranspiration ($ET$) and past Lake water level ($Lv$) at different lags. The daily data from 1st June 2000 to 31st May 2002 was used for training and 1st June 2002 to 31st May 2003 was used for testing the selected models. Time lag impact for each model input combination is tested from t (current day) to t − 4 (4-days earlier). To keep the model simple and fast the antecedent data of all the variables up to t − 2 (2-days earlier) have been used to predict the lake water level. The model structures has four variable, further addition of past data (up to 4 days) makes the structure complicated. Furthermore, wavelet decomposes these variables into one approximation and four detailed series. This complex structure makes the simulation process slow without adding to the prediction accuracy. Therefore, the model structure for both SVM and WA-SVM has been developed using the past data for all the variables up to two days time lag (t − 2) only. Fig. 3 depicts the architecture of hybrid wavelet support vector machine (WA-SVM) model. In order to assess the capabilities of applied models and to compare their performance for prediction of Lake water level, different statistical performance evaluation criteria such as Coefficient of Determination ($R^2$), Nash-Sutcliffe Coefficient (NS), Normalized Mean Square Error (NMSE) and Root-Mean-Square Error (RMSE) have been considered. The model with highest values of $R^2$ and NS along with minimum values of RMSE and NMSE is considered to be the best model which is later used for the prediction of Lake water level for Loktak Lake.

## 3. Application and results

### 3.1. Selection of best prediction model

The Lake water level of Loktal Lake is simulated and predicted using both the SVM and WA-SVM model considering daily rainfall ($R$), evaporation ($EV$), evapotranspiration ($ET$) and past Lake water level ($Lv$) data at different time lags. The wavelet-support vector machine (WA-SVM) models have been obtained combining two methods, wavelet and support vector machine. The hybrid WA-SVM model is an SVM model which uses decomposed series components (approximation and detail) obtained using DWT on original data. The architecture of the developed WA-SVM model is shown in Fig. 3. For the WA-SVM model inputs, the original time series are decomposed into a certain number of approximation and detail series components. Each decomposed series plays different role while providing the hidden information in the time series to the model. The WA-SVM is designed in such a way where the decomposed time series of original input times series are input of SVM and the original output time series are output of SVM. The performance statistics, such as the RMSE, NMSE, NS and $R^2$ of both the SVM and WA-SVM models, have been illustrated in Table 3. As it is clearly evident from Table 3 that, for SVM, model 1 performs better than the other combinations with minimum errors (NMSE, RMSE) along with maximum efficiency (NS) and $R^2$. However, for WA-SVM model, the structure with past Lake water level along with evaporation data (model 29) simulates with

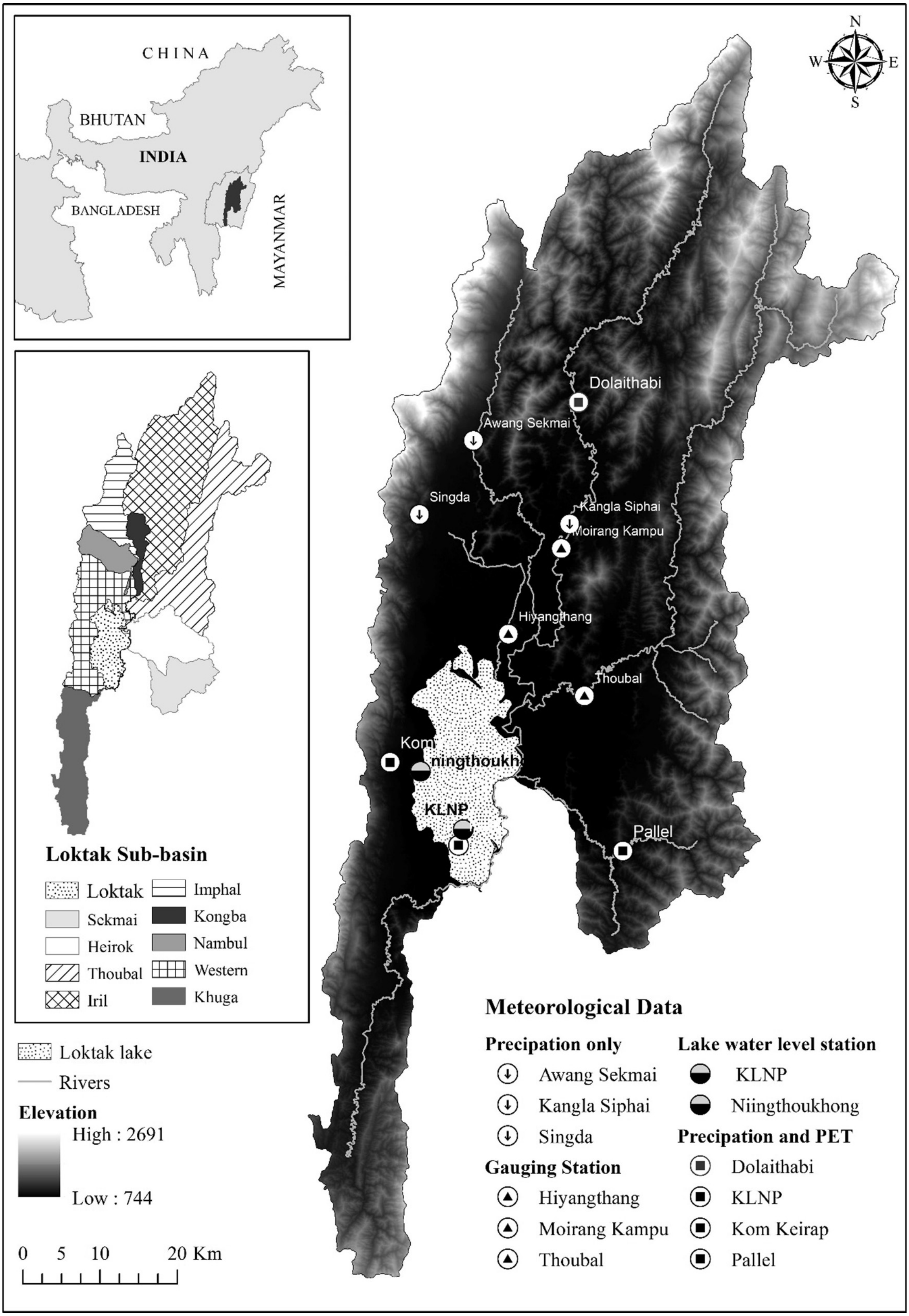


**Fig. 1.** Loktak Lake, its sub-catchments and location of hydro-meteorological stations.

highest accuracy and minimum error. The statics is almost same for some cases where the past Lake level data in combination with other hydro-meteorological variable is used. On the other hand, the accuracy is very low when the past information of Lake water level has not been used. Which suggest that the past Lake level data is utmost important for the simulation and prediction studies as it helps in understanding the underlying physics of the Lake system and to develop a relationship between input and output. In general, the statics in Table 3 and Fig. 4 suggest that the wavelet and support vector machine conjunction model (WA-SVM) that is

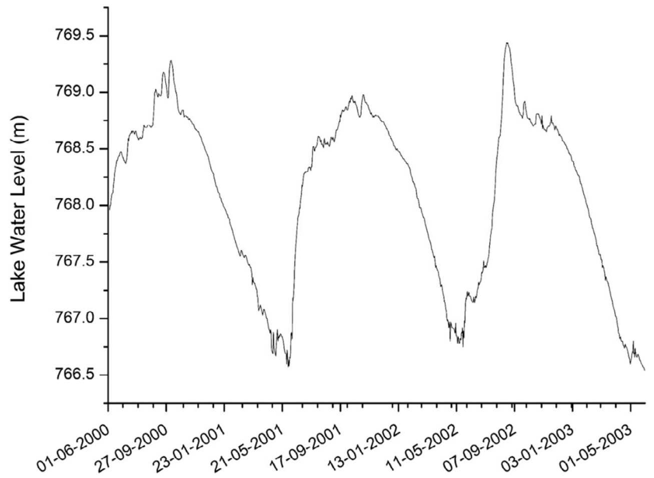


**Fig. 2.** Daily Lake water level for Loktak Lake.

**Table 1**
Optimal parameters of the SVM model.

| | Best $C$ Loktak Lake | Best $\gamma$ Loktak Lake |
|---|---|---|
| SVM | 3 | 0.0812 |

**Table 2**
Optimal parameters of the WA-SVM model for approximation and decomposition time series.

| Decomposition series | Best $C$ Loktak Lake | Best $\gamma$ Loktak Lake |
|---|---|---|
| Approximation series | 250 | 0.0812 |
| D1 series | 50 | 0.0081 |
| D2 series | 6 | 0.0081 |
| D3 series | 150 | 0.0081 |
| D4 series | 250 | 0.0812 |

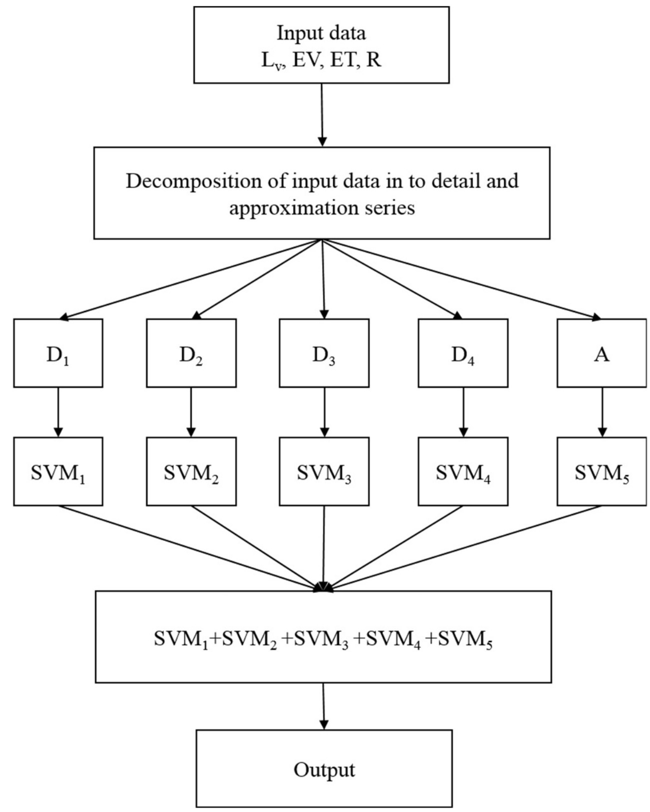


**Fig. 3.** Architecture of hybrid wavelet support vector machine (WA-SVM) model.

improved combining two methods, discrete wavelet transform (DWT) and SVM seems to be more adequate than the single SVM model for the simulation of Lake water level. The analysis of results also support the argument that the inclusion of various hydro-meteorological data may not necessarily adds to the accuracy of the selected model and hence models with simple structure can also be employed while studying such complex phenomenon.

Later, the best models of both SVM and WA-SVM (model 1 and model 29) are used to predict the daily Lake water level for a time lead of 5, 10, 15 and 20 days. Tables 4 and 5 show the statistical analysis of SVM and WA-SVM models while predicting the Lake water level. A general observation suggest that the prediction performance of hybrid WA-SVM model is significantly better than the single SVM model. Further, it also indicate that the prediction accuracy for both the models degrades as the time lead increases. The range of NMSE statistics for the SVM, and WA-SVM model for the time lead of 5–20 days is between 0.0348–0.2868, and 0.0170–0.1700, respectively. In terms of RMSE relative performance, among these two models, SVM predictions produced a much higher error ranging between 0.1547–0.4440, while in case of WA-SVM, it ranges between 0.1081–0.3418. The range of $R^2$ values for the SVM, and WA-SVM model for the time horizon of 5–20 days are between 0.9657–0.7227 and 0.9830–0.8331, respectively. In terms of NS statistics, the WA-SVM model showed a higher efficiency compared to the SVM model for all time leads. The overall statics indicate that for 5 days lead and onward prediction the WA-SVM significantly outperforms the SVM model.

The WA-SVM and SVM prediction performance in test period are shown in Figs. 5–8. It can be seen especially from the prediction plots that the WA-SVM maps the Lake water level time series better than the SVM. For 5 and 10 days lead time, WA-SVM overestimates the peak water level values, however it captures the trend better than the SVM model. Similarly for 15 and 20 days lead time the WA-SVM predict the output close to observed values. Such behaviour of overestimation could be attributed to the fact the training of data based models depends on the length of time series and if the data length is insufficient, model would overestimate/underestimate the target output. Apart from peak values, WA-SVM predicts the output better than SVM for all time leads. Moreover, it captures the total behaviour of Lake water level fluctuations significantly better than when SVM is used.

In general, the hybrid wavelet and support vector machine (WA-SVM) appears to be more adequate than the single SVM model for the simulation and prediction of Lake water level. The original data series is decomposed using a discrete wavelet transform function, which further provides the input to the SVM model at various resolution interval. Prediction accuracy of WA-SVM model is higher due to the fact that decomposed data series provides additional information than that obtained directly by original data series. This is why the WA-SVM model performs better than the SVM model. In practice, studying daily Lake water level fluctuation is difficult because this is affected by complex hydro-meteorological as well as anthropogenic factors. The variable such as use of Lake water for irrigation, hydropower generation, domestic as well industrial use could not be considered which could also be a reason for low prediction accuracy. However, decomposed time series provides additional advantage therefore, the WA-SVM model that uses multiple resolution components could be applied to study Lake water level fluctuation.

**Table 3**
Various models developed for daily Lake water level prediction using SVM and WA-SVM considering past Lake water level, rainfall, evaporation and evapotranspiration data as input from current day to previous two days during the testing period.

| Model no. | Model | NMSE | | $R^2$ | | NS | | RMSE | |
|---|---|---|---|---|---|---|---|---|---|
| | | SVM | WASVM | SVM | WASVM | SVM | WASVM | SVM | WASVM |
| **1.** | $\boldsymbol{Lv_{t+1} = Lv_t}$ | **0.006** | **0.005** | **0.995** | **0.998** | **0.993** | **0.998** | **0.0645** | **0.0317** |
| 2. | $Lv_{t+1} = R_t$ | 0.918 | 0.560 | 0.165 | 0.484 | 0.079 | 0.437 | 0.793 | 0.620 |
| 3. | $Lv_{t+1} = ET_t$ | 0.637 | 0.566 | 0.434 | 0.556 | 0.361 | 0.431 | 0.661 | 0.623 |
| 4. | $Lv_{t+1} = EV_t$ | 0.636 | 0.566 | 0.429 | 0.561 | 0.362 | 0.431 | 0.660 | 0.623 |
| 5. | $Lv_{t+1} = Lv_t + R_t$ | 0.007 | 0.002 | 0.994 | 0.997 | 0.992 | 0.997 | 0.069 | 0.039 |
| 6. | $Lv_{t+1} = Lv_t + ET_t$ | 0.012 | 0.001 | 0.990 | 0.998 | 0.987 | 0.998 | 0.091 | 0.028 |
| 7. | $Lv_{t+1} = Lv_t + EV_t$ | 0.011 | 0.001 | 0.990 | 0.998 | 0.988 | 0.998 | 0.090 | 0.028 |
| 8. | $Lv_{t+1} = Lv_t + EV_t + ET_t$ | 0.015 | 0.001 | 0.988 | 0.998 | 0.984 | 0.998 | 0.102 | 0.028 |
| 9. | $Lv_{t+1} = Lv_t + R_t + ET_t$ | 0.013 | 0.002 | 0.989 | 0.998 | 0.986 | 0.997 | 0.096 | 0.037 |
| 10. | $Lv_{t+1} = Lv_t + R_t + EV_t$ | 0.013 | 0.002 | 0.986 | 0.998 | 0.983 | 0.998 | 0.106 | 0.037 |
| 11. | $Lv_{t+1} = Lv_t + R_t + EV_t + ET_t$ | 0.016 | 0.002 | 0.987 | 0.998 | 0.983 | 0.998 | 0.107 | 0.037 |
| 12. | $Lv_{t+1} = Lv_t + Lv_{t-1}$ | 0.008 | 0.001 | 0.992 | 0.998 | 0.991 | 0.998 | 0.077 | 0.0283 |
| 13. | $Lv_{t+1} = R_t + R_{t-1}$ | 0.865 | 0.557 | 0.229 | 0.483 | 0.132 | 0.440 | 0.770 | 0.618 |
| 14. | $Lv_{t+1} = ET_t + ET_{t-1}$ | 0.606 | 0.531 | 0.475 | 0.585 | 0.392 | 0.466 | 0.644 | 0.604 |
| 15. | $Lv_{t+1} = EV_t + EV_{t-1}$ | 0.616 | 0.535 | 0.466 | 0.588 | 0.382 | 0.462 | 0.650 | 0.606 |
| 16. | $Lv_{t+1} = Lv_t + R_t + Lv_{t-1} + R_{t-1}$ | 0.012 | 0.003 | 0.990 | 0.997 | 0.987 | 0.996 | 0.092 | 0.046 |
| 17. | $Lv_{t+1} = Lv_t + ET_t + Lv_{t-1} + ET_{t-1}$ | 0.032 | 0.001 | 0.973 | 0.998 | 0.967 | 0.998 | 0.150 | 0.028 |
| 18. | $Lv_{t+1} = Lv_t + EV_t + Lv_{t-1} + EV_{t-1}$ | 0.031 | 0.001 | 0.975 | 0.998 | 0.968 | 0.998 | 0.147 | 0.028 |
| 19. | $Lv_{t+1} = Lv_t + EV_t + ET_t + Lv_{t-1} + EV_{t-1} + ET_{t-1}$ | 0.045 | 0.001 | 0.963 | 0.998 | 0.954 | 0.998 | 0.176 | 0.029 |
| 20. | $Lv_{t+1} = Lv_t + R_t + ET_t + Lv_{t-1} + R_{t-1} + ET_{t-1}$ | 0.043 | 0.002 | 0.965 | 0.997 | 0.956 | 0.997 | 0.172 | 0.041 |
| 21. | $Lv_{t+1} = Lv_t + R_t + EV_t + Lv_{t-1} + R_{t-1} + EV_{t-1}$ | 0.042 | 0.002 | 0.967 | 0.997 | 0.957 | 0.997 | 0.169 | 0.041 |
| 22. | $Lv_{t+1} = Lv_t + R_t + EV_t + ET_t + Lv_{t-1} + R_{t-1} + EV_{t-1} + ET_{t-1}$ | 0.059 | 0.003 | 0.954 | 0.998 | 0.940 | 0.997 | 0.201 | 0.042 |
| 23. | $Lv_{t+1} = Lv_t + Lv_{t-1} + Lv_{t-2}$ | 0.011 | 0.001 | 0.990 | 0.999 | 0.988 | 0.999 | 0.087 | 0.019 |
| 24. | $Lv_{t+1} = R_t + R_{t-1} + R_{t-2}$ | 0.825 | 0.551 | 0.276 | 0.490 | 0.172 | 0.447 | 0.752 | 0.614 |
| 25. | $Lv_{t+1} = ET_t + ET_{t-1} + ET_{t-2}$ | 0.610 | 0.508 | 0.494 | 0.610 | 0.387 | 0.490 | 0.647 | 0.590 |
| 26. | $Lv_{t+1} = EV_t + EV_{t-1} + EV_{t-2}$ | 0.615 | 0.517 | 0.488 | 0.613 | 0.383 | 0.481 | 0.649 | 0.595 |
| 27. | $Lv_{t+1} = Lv_t + R_t + Lv_{t-1} + R_{t-1} + Lv_{t-2} + R_{t-2}$ | 0.016 | 0.002 | 0.986 | 0.997 | 0.983 | 0.998 | 0.107 | 0.039 |
| 28. | $Lv_{t+1} = Lv_t + ET_t + Lv_{t-1} + ET_{t-1} + Lv_{t-2} + ET_{t-2}$ | 0.058 | 0.001 | 0.952 | 0.999 | 0.941 | 0.999 | 0.201 | 0.025 |
| **29.** | $\boldsymbol{Lv_{t+1} = Lv_t + EV_t + Lv_{t-1} + EV_{t-1} + Lv_{t-2} + EV_{t-2}}$ | **0.059** | **0.001** | **0.951** | **0.999** | **0.940** | **0.999** | **0.202** | **0.024** |
| 30. | $Lv_{t+1} = Lv_t + EV_t + ET_t + Lv_{t-1} + EV_{t-1} + ET_{t-1} + Lv_{t-2} + EV_{t-2} + ET_{t-2}$ | 0.095 | 0.001 | 0.922 | 0.999 | 0.904 | 0.999 | 0.255 | 0.025 |
| 31. | $Lv_{t+1} = Lv_t + R_t + ET_t + Lv_{t-1} + R_{t-1} + ET_{t-1} + Lv_{t-2} + R_{t-2} + ET_{t-2}$ | 0.088 | 0.003 | 0.929 | 0.997 | 0.912 | 0.997 | 0.246 | 0.042 |
| 32. | $Lv_{t+1} = Lv_t + R_t + EV_t + Lv_{t-1} + R_{t-1} + EV_{t-1} + Lv_{t-2} + R_{t-2} + EV_{t-2}$ | 0.085 | 0.002 | 0.932 | 0.997 | 0.915 | 0.997 | 0.242 | 0.043 |
| 33. | $Lv_{t+1} = Lv_t + R_t + EV_t + ET_t + Lv_{t-1} + R_{t-1} + EV_{t-1} + ET_{t-1} + Lv_{t-2} + R_{t-2} + EV_{t-2} + ET_{t-2}$ | 0.133 | 0.003 | 0.892 | 0.997 | 0.865 | 0.997 | 0.302 | 0.044 |

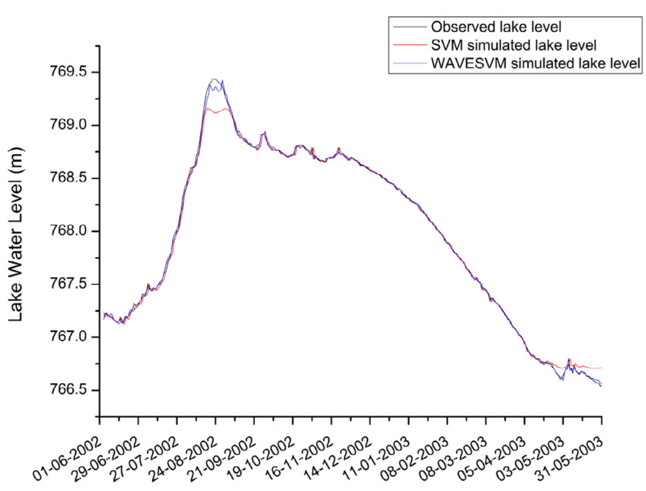


**Fig. 4.** Simulation of Lake water level using SVM and WA-SVM model.

**Table 4**
Prediction performance of best SVM for 5-day, 10-days, 15-days and 20-days lead.

| Prediction level | NMSE | RMSE | $R^2$ | NS |
|---|---|---|---|---|
| 5 day | 0.0348 | 0.1547 | 0.9657 | 0.9651 |
| 10 days | 0.0942 | 0.2546 | 0.9068 | 0.9055 |
| 15 days | 0.1793 | 0.3512 | 0.8255 | 0.8202 |
| 20 days | 0.2868 | 0.4440 | 0.7227 | 0.7124 |

**Table 5**
Prediction performance of best WAVE-SVM for 5-day, 10-days, 15-days and 20-days lead.

| Prediction level | NMSE | RMSE | $R^2$ | NS |
|---|---|---|---|---|
| 5 day | 0.0170 | 0.1081 | 0.9830 | 0.9830 |
| 10 days | 0.0522 | 0.1895 | 0.9477 | 0.9477 |
| 15 days | 0.1081 | 0.2727 | 0.8927 | 0.8915 |
| 20 days | 0.1700 | 0.3418 | 0.8331 | 0.8295 |

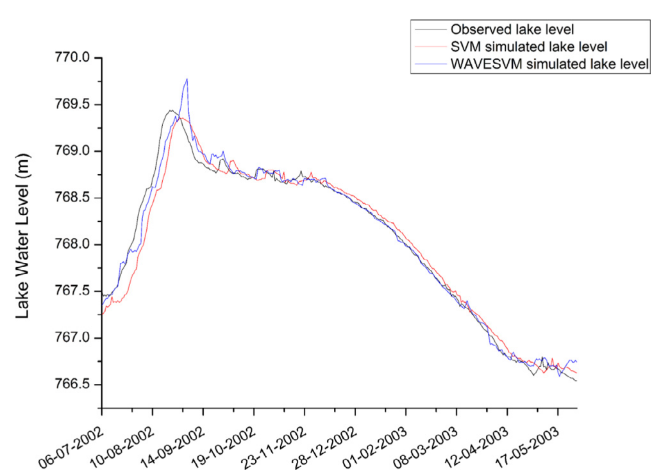


**Fig. 5.** Daily Lake water level prediction of SVM and WA-SVM models for a time lead of 5 days.

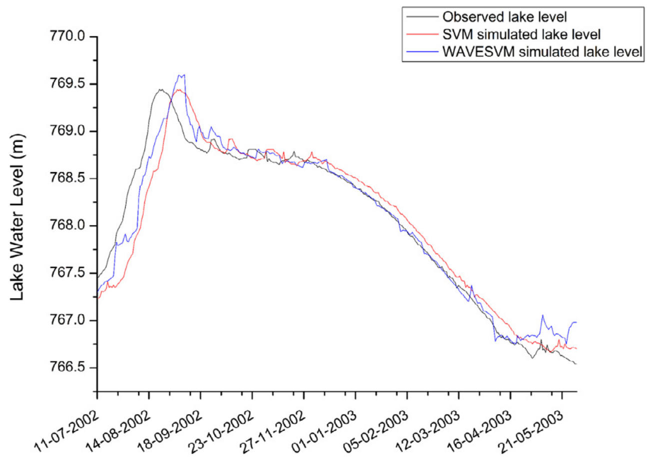


**Fig. 6.** Daily Lake water level prediction of SVM and WA-SVM models for a time lead of 10 days.

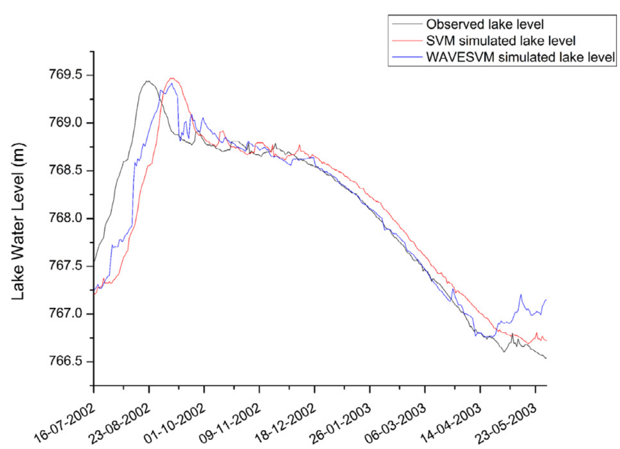


**Fig. 7.** Daily Lake water level prediction of SVM and WA-SVM models for a time lead of 15 days.

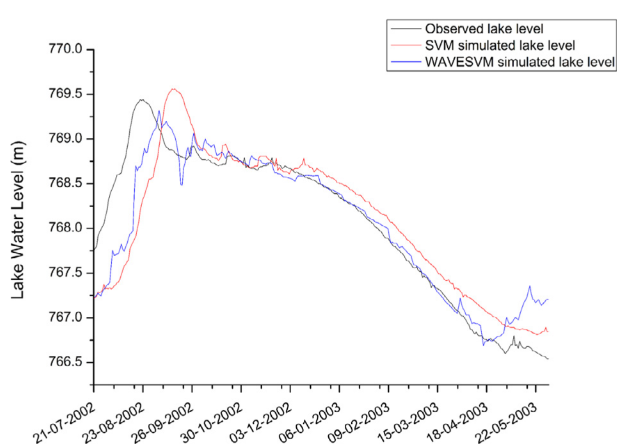


**Fig. 8.** Daily Lake water level prediction of SVM and WA-SVM models for a time lead of 20 days.

## 4. Conclusion

The accuracy of wavelet-support vector machine (WA-SVM) model has been evaluated for prediction of Lake water level in the present study. Two methods namely discrete wavelet transform and support vector machine were used to develop the hybrid WA-SVM models. Further, the developed WA-SVM models were compared with single SVM model while predicting the daily Lake water level of Loktak Lake up to 20 days lead time. Various hydro-meteorological variable were used as input for the SVM model, however for WA-SVM models the time series of those variables were further decomposed at sub-series level to provide more information about the physical dynamics. The comparative analysis suggest that the DWT could significantly enhance the simulation as well as prediction accuracy of the SVM model in predicting Lake water level. The WA-SVM models shows higher efficiency for all leads (5–20 days). The error values (NMSE, RMSE) is also significantly less for WA-SVM models than the SVM models. The results presented in the forms of statistical indicators and figures suggest that the WA-SVM approach provide a superior alternative to the SVM model for developing input–output simulations and predicting Lake water level in situations that do not require modeling of the internal structure of the watershed.

## Acknowledgements

Authors are thankful to the Loktak Development Authority (LDA) for providing the required data to conduct this research work.

## References

[1] J. Adamowski, K. Sun, Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds, J. Hydrol. 390 (1) (2010) 85–91.
[2] J. Adamowski, H.F. Chan, A wavelet neural network conjunction model for groundwater level forecasting, J. Hydrol. 407 (1) (2011) 28–40.
[3] P. Banerjee, V.S. Singh, K. Chatttopadhyay, P.C. Chandra, B. Singh, Artificial neural network model as a potential alternative for groundwater salinity forecasting, J. Hydrol. 398 (3) (2011) 212–220.
[4] M. Behzad, K. Asghari, E.A. Coppola Jr., Comparative study of SVMs and ANNs in aquifer water level prediction, J. Comput. Civ. Eng. 24 (5) (2009) 408–413.
[5] B. Cannas, A. Fanni, L. See, G. Sias, Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning, Phys. Chem. Earth 31 (18) (2006) 1164–1171.
[6] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. (TIST) 2 (3) (2011) 27.
[7] K.Y. Choy, C.W. Chan, Modelling of river discharges and rainfall using radial basis function networks based on support vector regression, Int. J. Syst. Sci. 34 (14–15) (2003) 763–773.
[8] M. Çimen, O. Kisi, Comparison of two different data-driven techniques in modeling lake level fluctuations in Turkey, J. Hydrol. 378 (2009) 253–262.
[9] I. Daliakopoulos, P. Coulibalya, I.K. Tsani, Groundwater level forecasting using artificial neural network, J. Hydrol. 309 (1–4) (2005) 229–240.
[10] S. Emamgholizadeh, K. Moslemi, G. Karami, Prediction the groundwater level of bastam plain (Iran) by artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS), Water Resour. Manage. 28 (15) (2014) 5433–5446.
[11] E. Fallah-Mehdipour, O.B. Haddad, M.A. Mariño, Prediction and simulation of monthly groundwater levels by genetic programming, J. Hydro-environ. Res. 7 (4) (2013) 253–260.
[12] Y. Gong, Y. Zhang, S. Lan, H. Wang, A comparative study of artificial neural networks, support vector machines and adaptive neuro fuzzy inference system for forecasting groundwater levels near Lake Okeechobee, Florida, Water Resour. Manage. 30 (1) (2016) 375–391.
[13] Z. He, X. Wen, H. Liu, J. Du, A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region, J. Hydrol. 509 (2014) 379–386.
[14] M.S. Khan, P. Coulibaly, Application of support vector machine in Lake water level prediction, J. Hydrol. Eng. 11 (3) (2006) 199–205.
[15] S.J. Kim, Y. Hyun, K.K. Lee, Time series modeling for evaluation of groundwater discharge rates into an urban subway system, Geosci. J. 9 (1) (2005) 15–22.
[16] O. Kisi, River flow forecasting and estimation using different artificial neural network techniques, Hydrol. Res. 39 (1) (2008) 27–40.
[17] O. Kisi, Neural networks and wavelet conjunction model for intermittent streamflow forecasting, J. Hydrol. Eng. 14 (8) (2009) 773–782.
[18] O. Kisi, M. Cimen, A wavelet-support vector machine conjunction model for monthly streamflow forecasting, J. Hydrol. 399 (1) (2011) 132–140.
[19] S. Maskey, Y.B. Dibike, A. Jonoski, D. Solomatine, Groundwater model approximation with artificial neural network for selecting optimal pumping strategy for plume removal, in: Workshop Proceedings in Artificial Intelligence Methods in Civil Engineering Applications, 2000, pp. 67–80.

[20] K. Mohammadi, Groundwater table estimation using MODFLOW and artificial neural networks, Water Sci. Technol. Libr. 68 (2) (2008) 127–138.

[21] S. Mohanty, M.K. Jha, A. Kumar, K.P. Sudheer, Artificial neural network modeling for groundwater level forecasting in a river island of eastern India, Water Resour. Manage. 24 (9) (2010) 1845–1865.

[22] S. Mohanty, M.K. Jha, A. Kumar, D.K. Panda, Comparative evaluation of numerical model and artificial neural network for simulating groundwater flow in Kathajodi-Surua Inter-basin of Odisha, India, J. Hydrol. 495 (2013) 38–51.

[23] J. Platt, Fast training of support vector machines using sequential minimal optimization, Adv. Kernel Methods—Support Vector Learn. 3 (1999).

[24] Jian. Qu, Ming J. Zuo, Support vector machine based data processing algorithm for wear degree classification of slurry pump systems, Measurement 43 (6) (2010) 781–791.

[25] H.R. Safavi, M. Esmikhani, Conjunctive use of surface water and groundwater: application of support vector machines (SVMs) and genetic algorithms, Water Resour. Manage. 27 (7) (2013) 2623–2644.

[26] R.R. Sethi, A. Kumar, S.P. Sharma, H.C. Verma, Prediction of water table depth in a hard rock basin by using artificial neural network, Int. J. Water Resour. Environ. Eng. 2 (4) (2010) 95–102.

[27] J. Shiri, O. Kisi, Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model, J. Hydrol. 394 (2010) 486–493.

[28] J. Shiri, O. Kisi, Estimation of daily suspended sediment load by using wavelet conjunction models, ASCE J. Hydrol. Eng. 17 (9) (2012) 986–1000.

[29] J. Shiri, S. Shamshirband, O. Kisi, S. Karimi, S.M. Bateni, S.H.H. Nezhad, A. Hashemi, Prediction of water-level in the Urmia Lake using the extreme learning machine approach, Water Resour. Manage. 30 (14) (2016) 5217–5229.

[30] C.R. Singh, J.R. Thompson, D.G. Kingston, J.R. French, Modelling water-level options for ecosystem services and assessment of climate change: Loktak Lake, northeast India, Hydrol. Sci. J. 56 (8) (2011) 1518–1542.

[31] C. Sivapragasam, S.Y. Liong, M.F.K. Pasha, Rainfall and runoff forecasting with SSA–SVM approach, J. Hydroinformatics 3 (3) (2001) 141–152.

[32] P.D. Sreekanth, P.D. Sreedevi, S. Ahmed, N. Geethanjali, Comparison of FFNN and ANFIS models for estimating groundwater level, Environ. Earth Sci. 62 (2010) 1301–1310.

[33] C. Suryanarayana, C. Sudheer, V. Mahammood, B.K. Panigrahi, An integrated wavelet-support vector machine for groundwater level prediction in Visakhapatnam, India, Neurocomputing 145 (2014) 324–335.

[34] M.K. Tiwari, C. Chatterjee, Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach, J. Hydrol. 1 (394) (2010) 458–470.

[35] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, USA, 1995, p. 314.

[36] V. Vapnik, S.E. Golowich, A. Smola, Support vector method for function approximation, regression estimation, and signal processing, Adv. Neural Inform. Process. Syst. (1996) 281–287.

[37] A.K. Verma, T.N. Singh, Prediction of water quality from simple field parameters, Environ. Earth Sci. 69 (2013) 821–829.

[38] B. Yadav, S. Ch, S. Mathur, J. Adamowski, Estimation of in-situ bioremediation system cost using a hybrid extreme learning machine (ELM)-particle swarm optimization approach, J. Hydrol. (2016).

[39] H. Yoon, S.C. Jun, Y. Hyun, G.O. Bae, K.K. Lee, A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer, J. Hydrol. 396 (1) (2011) 128–138.

[40] X. Yu, S.Y. Liong, V. Babovic, EC-SVM approach for real-time hydrologic forecasting, J. Hydroinformatics 6 (3) (2004) 209–223.

[41] H. Yoon, S.C. Jun, Y. Hyun, G.O. Bae, K.K. Lee, A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer, J. Hydrol. 396 (2011) 128–138.