

A Hybrid Model for Stream Flow Forecasting Using Wavelet and Least Squares Support Vector Machines

Ani Shabri*

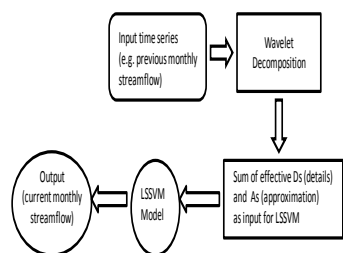
Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor Malaysia

*Corresponding author: ani@utm.my

Article history

Received: 14 August 2014
Received in revised form:
6 November 2014
Accepted: 1 February 2015

Graphical abstract



Abstract

This paper proposed a hybrid wavelet-least square support vector machines (WLSSVM) model that combine both wavelet method and LSSVM model for monthly stream flow forecasting. The original stream flow series was decomposed into a number of sub-series of time series using wavelet theory and these time series were imposed as input data to the LSSVM for stream flow forecasting. The monthly stream flow data from Klang and Langat stations in Peninsular Malaysia are used for this case study. Time series prediction capability performance of the WLSSVM model is compared with single LSSVM and Autoregressive Integrated Moving Average (ARIMA) models using various statistical measures. Empirical results showed that the WLSSVM model yield a more accurate outcome compared to individual LSSVM, ANN and ARIMA models for monthly stream flow forecasting.

Keywords: Wavelet, least square support vector machines, artificial neural network, ARIMA, SVM

© 2015 Penerbit UTM Press. All rights reserved.

1.0 INTRODUCTION

Stream flow forecasting is an important issue in hydrologic engineering where it provides basic information on a wide range of problems for the optimal management of water resources related to the design and operation of a river system. In Malaysia, short-term forecasting like hourly and daily forecasting is crucial for flood warning and defense while medium- to long-term forecasting, at weekly, monthly, seasonal, or even annual time series stream flow data, is particularly useful in reservoir operation and irrigation management decisions such as drought mitigation and managing river treaties. However, accurate time series forecasting is one of the greatest challenges in operational hydrology because hydrological systems are dynamic with time varying system for inputs and outputs, large temporal variability, and commonly demonstrate non-linear responses to system inputs. It is relevant to note that it can established better models by using meteorological and hydrological variables such as precipitation, evapotranspiration, temperature and geomorphologic characteristics data as input parameters for the applied models. Although incorporating other variables may improve the prediction accuracy, in practice, especially in developing countries like Malaysia, such information is often either not available or difficult to obtain. Owing to the complexity of this process, many researchers are beginning to focus on stream flow forecasting which only considers past stream flow data. Therefore, the current work applies the stream flow values as inputs for the models used.

In general, the stochastic models such as autoregressive integrated moving average (ARIMA) is widely used for hydrologic time series forecasting.¹⁻⁶ The popularity of ARIMA models is due to its statistical properties, such as the well-known Box-Jenkins methodology, forecasting capabilities and richness of information on time-related changes. However, ARIMA models are basically linear models using data that are assumed to be stationary, and have a limited capability to capture non-stationarities and non-linearities in hydrologic data.

In data-based forecasting, artificial neural network (ANN) model is probably the most common method for modeling and forecasting non-linear hydrologic time series. Recently, ANN has been introduced for modeling complex hydrologic system as it has been successfully employed in the modeling of various aspects of hydrologic processes.⁶⁻¹³ Although ANN has the advantages of generating accurate forecasting, their performance in some specific situation is inconsistent. The network structure of ANN is hard to determine and usually done by using a trial-and-error approach. In addition, the ANNs do have some drawbacks, such as over-fitting, convergence to local minimum and learning slowly, which make it difficult to gain satisfactory performance when dealing with complex hydrological processes.¹⁴

A more advanced Artificial Intelligence (AI) approach is the least square support vector machines (LSSVM) method which is a new universal learning machines proposed by Suykens and co-workers.¹⁵ The LSSVM algorithm provides a computational advantage over standard support vector machines (SVM) by converting quadratic optimization problem into a system of linear

equations. This method uses equality constraints instead of inequality constraints and adopts the least squares linear system as its loss function, which is computationally attractive. Furthermore, LSSVM also has good convergence and high precision. It is one of the soft computational techniques with a powerful methodology and has been successfully applied to solve various case studies.¹⁶⁻¹⁸ In the water resource field, the LSSVM method has received very little attention and only a few applications of LSSVM in modeling environmental and ecological systems such as water quality prediction¹⁹ rainfall-runoff modeling²⁰ and stream flow forecasting^{18,21,22} have been carried out.

Wavelet transform has been studied for many years by mathematicians and has become a useful technique for analyzing variations, periodicities and trends in different areas of hydrology and water resources.²³⁻²⁷ Recently, new hybrid models on wavelet transform process have been improved for forecasting.²⁵⁻³⁶ They observed the use of wavelet techniques to pre-process time series data into decomposed wavelet coefficients of different decompositions produced significant results than the original time series when used as input. Such hybrid models show significant advantages over the traditional AI models. This indicates that the wavelet can be a promising tool in the decomposition of time series.

In this paper, a hybridization of wavelet and LSSVM model (WLLSVM) has been proposed to forecast monthly stream flow. This paper aims to demonstrate the ability of WLLSVM model to forecast stream flow based on previously measured flow values. To show the application of this model, the stream flow data from Klang and Langat stations in Peninsular Malaysia is chosen as the case study. Finally, this study also aims to evaluate the proposed model's ability based on its performance by comparing with other models' performance, such as the performance of single LSSVM and ARIMA models.

2.0 THE LEAST SQUARE VECTOR MACHINES

LSSVM is a new version of SVM modified by Suykens *et al.*¹⁵ Furthermore, LSSVM involves the solution of a quadratic optimization problem (QOP) with a least squares loss function and equality constraints instead of inequality constraints. Instead of solving a QOP problem as in SVM, LSSVM can obtain the solutions of a set of linear equation. In this section, we briefly introduce the basic theory of LSSVM in time series forecasting. A training sample set (x_i, y_i) with input $x_i \in R^n$ and output $y_i \in R$ are used. The following regression model can be constructed by using non-linear mapping function

$$y(x) = w^T \phi(x) + b \quad (1)$$

where w is the weight vector, b is the bias terms and the nonlinear mapping $\phi(x)$ maps the input data into a higher dimensional feature space. LSSVM introduces a least square version to SVM regression by formulating the regression problem as

$$\min R(w, e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^n e_i^2 \quad (2)$$

subject to the equality constraints

$$y(x) = w^T \phi(x_i) + b + e_i, \quad i = 1, 2, \dots, n \quad (3)$$

where e_i represents error variables and γ is a regularization parameter. To solve this optimization problem, Lagrange function is constructed as

$$L(w, b, e, \alpha) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^n e_i^2 - \sum_{i=1}^n \alpha_i \{w^T \phi(x_i) + b + e_i - y_i\} \quad (4)$$

where α_i is Lagrange multipliers. The solution of (2) can be obtained by partially differentiating with respect to w, b, e_i and α_i

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^n \alpha_i \phi(x_i) \quad (5)$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^n \alpha_i = 0 \quad (6)$$

$$\frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i \quad (7)$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \phi(x_i) + b + e_i - y_i = 0, \quad i = 1, 2, \dots, n \quad (8)$$

Then the weight w can be written as a combination of the Lagrange multipliers with the corresponding training data x_i .

$$w = \sum_{i=1}^n \alpha_i \phi(x_i) = \sum_{i=1}^n \gamma e_i \phi(x_i) \quad (9)$$

Putting the result of Eq. (9) into Eq. (1), then the following result is obtained:

$$y(\mathbf{x}) = \sum_{i=1}^n \alpha_i \phi(x_i)^T \phi(x_i) + b = \sum_{i=1}^n \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b \quad (10)$$

where a positive definite kernel is defined as follows:

$$K(\mathbf{x}_i, \mathbf{x}) = \phi(x_i)^T \phi(x_i)$$

The α vector and b can be found by solving a set of linear equations:

$$\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \phi(x_i)^T \phi(x_j) + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (11)$$

where $y = [y_1; \dots; y_n]$, $\mathbf{1} = [1; \dots; 1]$, $\alpha = [\alpha_1; \dots; \alpha_n]$. This finally leads to the following LSSVM model for function estimation:

$$y(\mathbf{x}) = \sum_{i=1}^n \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b \quad (12)$$

where α_i, b are the solution to the linear system. Kernel function, $K(\mathbf{x}_i, \mathbf{x})$ represents the high dimensional feature space that is nonlinearly mapped from input space x . The selection of an appropriate kernel function plays an important role during LSSVM modeling. Types of kernel have been developed such as linear kernel, polynomial kernel, Radial basis function (RBF) kernel and multi-layer perception kernel. The RBF kernel has received more attention from the machine learning community.³⁸⁻³⁹ The RBF kernel is defined as

$$K(\mathbf{x}_i, \mathbf{x}) = \exp(-\|\mathbf{x}_i - \mathbf{x}\|^2 / \sigma^2), \quad \sigma > 0 \quad (13)$$

where σ^2 is a positive real constant.

3.0 WAVELET ANALYSIS

Wavelet analysis is a multi-decomposition analysis, which provide information for both time and frequency domains of the signal and it is the important derivative of the Fourier transform. Wavelet transforms provide useful decompositions of the original time series, so that the wavelet-transformed data will improve the ability of a forecasting model by capturing useful information on various decomposition levels.⁴⁰

Wavelet will become an important tool in time series forecasting. The basic objective of wavelet transformation is analyzing the time series data, both in time and frequency domain by decomposing the original time series in different frequency bands using wavelet functions. Compared to the Fourier transform, time series are analyzed using sine and cosine functions. Wavelet transformations provide useful decomposition of the original time series by capturing useful information on various decomposition levels.

Assuming a continuous time series $f(t)$, $t \in [-\infty, \infty]$, a wavelet function can be written as

$$\psi(t, s) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \quad (14)$$

where t stands for time, τ for the time step in which the window function is iterated, and $s \in [0, \infty]$ for the wavelet scale. $\psi(t)$ called the mother wavelet can be defined as $\int_{-\infty}^{\infty} \psi(t) dt = 0$. The continuous wavelet transform (CWT) is given by

$$W(\tau, s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} f(t) \overline{\psi}\left(\frac{t-\tau}{s}\right) dt \quad (15)$$

where $\overline{\psi}(t)$ represents the complex conjugation of $\psi(t)$. $W(\tau, s)$ represents the overall sum of the time series multiplied by scale and the shifted version of wavelet function $\psi(t)$. The use of continuous wavelet transform in forecasting is not practically possible because calculating wavelet coefficient at every possible scale is time consuming and it generates abundance of data.

Therefore, discrete wavelet transformation (DWT) is preferred in most forecasting problems because of its simplicity and ability to compute with less time. The DWT involves choosing scales and position on powers of 2 so called dyadic scales and translations, then the analysis will be much more efficient as well as more accurate. The DWT can be defined as

$$\psi_{m,n}\left(\frac{t-\tau}{s}\right) = \frac{1}{\sqrt{s_0^{m/2}}} \psi\left(\frac{t-n\tau_0 s_0^m}{s_0^m}\right) \quad (16)$$

where m and n are integers that control the scale and time, respectively; s_0 is a specified, fixed dilation step greater than 1; and τ_0 is the location parameter, which must be greater than zero. The most common choices for the parameters $s_0 = 2$ and $\tau_0 = 1$. For a discrete time series $f(t)$ where $f(t)$ occurs at discrete time t , the DWT becomes

$$W_{m,n} = 2^{-m/2} \sum_{t=0}^{N-1} \psi(2^{-m} t - n) f(t) \quad (17)$$

where $W_{m,n}$ is the wavelet coefficient for the discrete wavelet at scale $s = 2^m$ and $\tau = 2^m n$. In Eq. (17), $f(t)$ is the time series ($t = 1, 2, \dots, N-1$), and N is an integer power of 2 ($N = 2^M$); n is the time translation parameter, which changes in the ranges $0 < n < 2M - m$, where $1 < m < M$.

According to Mallat's theory, the original discrete time series $f(t)$ can be decomposed into a series of linearly independent approximation and detail signals by using the inverse DWT. The inverse DWT is given by⁴¹

$$f(t) = T + \sum_{m=1}^M \sum_{t=0}^{2^{M-m-1}} W_{m,n} 2^{-m/2} \psi(2^{-m} t - n) \quad (18)$$

or in a simple format as

$$f(t) = A_M(t) + \sum_{m=1}^M D_m(t) \quad (19)$$

which $A_M(t)$ is called approximation sub-series or residual term at levels M and $D_m(t)$ ($m = 1, 2, \dots, M$) are detail sub-series which can capture small features of interpretational value in the data.

4.0 STUDY AREA

The stream flow data used in this study were obtained from the Department of Irrigation and Drainage Malaysia. Monthly stream flow data of Klang station (Station no: 3116430) and Langat station (Station no: 2816441) of Peninsular Malaysia. The locations of these stations are shown in Fig. 1.

Klang River flows through Kuala Lumpur and Selangor in Malaysia and eventually flows into the Straits of Malacca. It is approximately 120 km in length and drains a basin of about 1288 square kilometers. Klang River has 11 major tributaries that the river flows through the capital city of Kuala Lumpur which is a heavily populated area of more than four million people. It is considerably polluted.

The Langat River has a total catchment area of approximately 1815 km². The main river length at 141 km where it is mostly situated about 40 km east of Kuala Lumpur. The Langat River catchment straddles the main urban conurbation in the Klang Valley forming parts of the growing urban complex in Selangor. The Langat River is situated south and adjacent to the Klang Valley, Malaysia's highly developed urban conurbation where the nation's capital Kuala Lumpur is located.

In this study, the data sample of Klang station consisted of 34 years (1975–2008) of stream flow record. The first 30 years of flow data (360 months, 80% of the whole data set) were used for training the network to obtain parameters model and another dataset consisting of 84 monthly records (20% of the whole data) was used for testing.



Figure 1 Location map of the study area

While for Langat station, the observed contained 47 years of data (564 months) with an observation period between 1961 and 2007. In the application, the first 38 years of flow data (456 months), 80% of the whole data set are used for training the network to obtain the model parameters. Another set of data consists of 108 monthly records (20% of the whole data) are used for testing.

The performances of the presented models were evaluated based on their root mean-square error (RMSE), mean absolute error (MAE) and correlation coefficient (R) for one step ahead. The RMSE, MAE and R are defined as

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t^o - y_t^f)^2$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t^o - y_t^f|$$

(20)

$$R = \frac{\frac{1}{n} \sum_{t=1}^n (y_t^o - \bar{y}_t^o)(y_t^f - \bar{y}_t^f)}{\sqrt{\frac{1}{n} \sum_{t=1}^n (y_t^o - \bar{y}_t^o)^2} \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t^f - \bar{y}_t^f)^2}}$$

where y_t^o and y_t^f are the observed and forecasted values at time t , respectively and n is the number of data points. The RMSE and MAE provide different types of information about the predictive capabilities of the model. The RMSE and MAE evaluate on how close the predictions match the observations. The criteria in deciding the best model are based on how small the MAE and RMSE found in both training and testing of the data. The correlation coefficient (R) measures how well the predicted flows correlate with the observed flows and shows the degree to which the two variables are linearly related. R value close to unity indicates a satisfactory result, while a low value or one that is close to zero implies an inadequate result.

5.0 RESULTS AND DISCUSSIONS

5.1 Fitting LSSVM To The Data

The selection of appropriate input data sets is an important part in LSSVM modelling. In this study, five model structures were developed to investigate the model performance of the input variables. The inputs represent the previous monthly stream flows (at time $t-1, t-2, t-3, t-4$ and $t-5$) and the output corresponds to the monthly stream flow at time t . Thus, the five different input models evaluated for stream flow forecasting are as follows:

M1-input was monthly flow data at lags 1: $y_t = f(y_{t-1})$

M2- input was monthly flow data at lags 1 and 2:

$$y_t = f(y_{t-1}, y_{t-2})$$

M3- input was monthly flow data at lags 1, 2 and 3:

$$y_t = f(y_{t-1}, y_{t-2}, y_{t-3})$$

M4-input was monthly flow data at lags 1, 2, 3 and 4:

$$y_t = f(y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4})$$

M5- input was monthly flow data at lags 1, 2, 3, 4 and 5:

$$y_t = f(y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5})$$

Secondly, the part that needs to be considered is what values of the parameters (γ, σ^2) are to be used. The parameter γ controls the penalty degree, and the parameter σ^2 represents the kernel function parameter. There is no designated rule in choosing the optimal parameters of a LSSVM model. In order to obtain the optimal model parameters of LSSVM, a grid search algorithm and cross-validation method were employed. The grid search method is a common method that was applied to calibrate these parameters more effectively and systematically to overcome the potential shortcomings of the trails and error method. It is a straightforward and exhaustive method to search parameters. In this study, a grid search of (γ, σ^2) within the γ range from 10 to 1000 and σ^2 in the range of 0.01 to 1.0 was conducted to find the optimal parameters. In order to avoid the risk of over fitting, the cross-validation scheme is used to evaluate the model performance. The entire dataset was randomly partitioned into 10 equal-size subsets. During each run, one of the partition was chosen for testing, while the rest were used for training. This procedure was repeated 10 times. For each hyper parameter pair (γ, σ^2) in the search space, 10-fold cross validation on the training set was performed to predict the prediction error. This process achieved through a program written in MATLAB. The best fit model structure for each model is determined according to the criteria of the performance evaluation.

Table 1 below presents the performance results obtained in the training and testing period of the regular LSSVM approach (i.e. those using original data) for Klang and Langat station, respectively. For Klang station, the LSSVM model where the inputs are one previous month (M1) has the best accuracy in training period and M4 has the best accuracy in testing period. While for Langat station, the model input M5 gave the best performance for LSSVM model in training period and M2 in testing period.

Table 1 Forecasting Performance Indices of LSSVM for Klang and Langat Stations

Station	Model Inputs	Training			Testing		
		RMSE	MAE	R	RMSE	MAE	R
Klang	M1	5.706	4.016	0.789	3.980	3.263	0.759
	M2	6.730	4.396	0.686	3.665	2.966	0.801
	M3	6.791	4.483	0.678	3.538	2.863	0.821
	M4	6.449	4.265	0.717	3.468	2.825	0.841
	M5	7.274	4.713	0.615	3.660	2.941	0.824
Langat	M1	17.067	12.261	0.517	19.974	13.440	0.519
	M2	16.813	11.821	0.538	19.653	13.204	0.542
	M3	16.567	11.629	0.556	19.886	13.457	0.523
	M4	16.652	11.705	0.551	19.901	13.628	0.523
	M5	16.481	11.478	0.564	19.767	13.415	0.532

5.2 Fitting Hybrid Wavelet LSSVM To The Data

A hybrid wavelet LSSVM (WLSSVM) model is obtained by combining two methods, DWT and LSSVM. Before LSSVM application is applied, the original time series data were decomposed into periodic components (DWs) by Mallat’s DWT algorithm.⁴¹ The observation series was decomposed into a number of wavelet components, depending on the selected decomposition levels. Deciding the optimal decomposition level of the time series data in wavelet analysis plays an important role in preserving the information and reducing the distortion of the datasets. However, there is no existing theory to tell how many decomposition levels are needed for any time series. To select the number of decomposition levels, the following formula is used to determine the decomposition level²⁶

$$M = \log(N)$$

where N is length of the time series and M is decomposition level. In this study, $N = 408$ and 564 monthly data are used for Klang and Langat rivers, respectively, which approximately gives $M = 3$ decomposition levels. Three decomposition levels are employed in this study; similar to the studies employed by Kisi²⁷. The observation time series of discharge flow data was decomposed at 3 decomposition levels (2-4-8 months).

The effectiveness of discrete wavelet components (DWC) is determined based on the correlation between the observed stream flow data and the wavelet coefficients of different decomposition levels.³² Table 2 below shows the correlations between each wavelet component time series and the original monthly stream flow data.

In Table 2 above, the D1 component shows low correlations for both stations. The correlation between the wavelet component D2 and D3 of the monthly stream flow and the observed monthly stream flow data show significantly higher correlations compared to the D1 components. The results of the correlation analysis showed that D2 and D3 are the most effective components to be considered for forecasting. According to the correlation analyses, the effective components D2 and D3 were selected as the dominant wavelet components. Afterward, the significant wavelet components D2, D3 and approximation (A3) component were added to each other to constitute the new series.

Table 2 The correlation coefficients between each of sub-time series and original monthly stream flow data (Q_t)

Station	DWC	Correlations between Q_t and					Mean
		D_{t-1}	D_{t-2}	D_{t-3}	D_{t-4}	D_{t-5}	
Klang	D1	0.079	-0.103	-0.032	0.155	-0.044	0.083
	D2	0.248	0.092	-0.134	-0.380	-0.335	0.238
	D3	-0.285	-0.286	-0.193	0.033	0.214	0.202
	A3	0.358	0.402	0.451	0.497	0.540	0.450
Langat	D1	-0.086	0.026	0.062	0.032	-0.122	0.066
	D2	0.405	0.182	-0.219	-0.524	-0.409	0.348
	D3	-0.366	-0.361	-0.230	0.069	0.247	0.255
	A3	0.135	0.192	0.240	0.294	0.349	0.242

For the WLSSVM model, the new series (components D2, D3 and A3) were used as inputs to the LSSVM model. Thus, the five different input models for WLSSVM models for stream flow forecasting are as follows:

Mw1- input was monthly flow data at lags 1: $y_t = f(Dw_{t-1})$

Mw2- input was monthly flow data at lags 1 and 2:

$$y_t = f(Dw_{t-1}, Dw_{t-2})$$

Mw3- input was monthly flow data at lags 1, 2 and 3:

$$y_t = f(Dw_{t-1}, Dw_{t-2}, Dw_{t-3})$$

Mw4- input was monthly flow data at lags 1, 2, 3 and 4:

$$y_t = f(Dw_{t-1}, Dw_{t-2}, Dw_{t-3}, Dw_{t-4})$$

Mw5- input was monthly flow data at lags 1, 2, 3, 4 and 5:

$$y_t = f(Dw_{t-1}, Dw_{t-2}, Dw_{t-3}, Dw_{t-4}, Dw_{t-5})$$

where $Dw_t = A3_t + D2_t + D3_t$

Figure 2 shows the structure of the WLSSVM model.

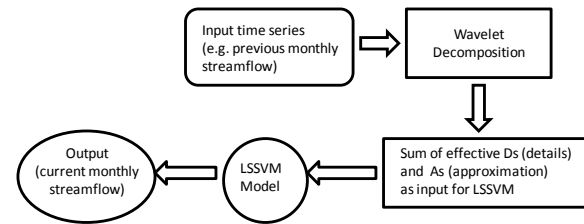


Figure 2 The structure of the WLSSVM model

A program code that include with wavelet toolbox was written in MATLAB language for the development of LSSVM model. The forecasting performances of the WLSSVM model are presented in Table 3 for Klang and Langat stations.

Table 3 below shows that WLSSVM model has a significant positive effect on stream flow forecast. From Table 3, we can see that in Klang station, the Mw4 model has the best accuracy in both training and testing period. In Langat station, the Mw5 has the best performance criteria in training. However, for the testing phase, the best RMSE (11.796) and R (0.865) was obtained from the model Mw4 and MAE (8.734) from model Mw5.

Table 3 Forecasting Performance Indices of WLSSVM model.

Station	Model Input	Training			Testing		
		RMSE	MAE	R	RMSE	MAE	R
Klang	Mw1	6.795	4.453	0.677	2.969	2.275	0.872
	Mw2	5.164	3.481	0.829	2.991	2.334	0.871
	Mw3	6.449	4.265	0.717	3.468	2.825	0.841
	Mw4	4.564	3.039	0.872	2.709	2.071	0.891
	Mw5	4.708	3.151	0.867	2.978	2.305	0.867
Langat	Mw1	14.848	11.002	0.667	17.829	12.089	0.646
	Mw2	10.477	7.850	0.851	13.098	10.018	0.829
	Mw3	9.617	7.103	0.876	12.557	9.439	0.844
	Mw4	8.847	6.524	0.896	11.796	8.947	0.865
	Mw5	8.480	6.146	0.905	11.805	8.734	0.865

For further analysis, the best performance of the LSSVM and WLSSVM models in terms of the RMSE, MAE and R of testing phase are compared. For Klang station, the best correlation coefficient (R) obtained by LSSVM model is 0.841, while the best R value of WLLSVR model increased to 0.891. The RMSE obtained by LSSVM model is 3.468, with WLSSVM model this value is decreased to 2.709. Similarly, while the MAE of LSSVM obtained is 2.825, the MAE value of WLSSVM model is decreased to 2.071. The proposed WLSSVM model improved the LSSVM forecast to about 28.02% and 36.41% reduction in RMSE and MAE values, respectively; and improvements of the R value was approximately 5.61%.

For the Langat station, the WLSSVM model reduced the RMSE and MAE by 66.48% and 51.18%, respectively, and

increased the R by 37.27% with respect to the single LSSVM model. This results show that the new series (DWT) have significance of extremely positive effect on LSSVM model results.

Figure 3 and 4 below show the hydrograph and scatter plot for the LSSVM and WLSSVM models for testing period in Klang and Langat stations, respectively. It can be seen that the WLSSVM model forecasts quite close to the observed data for both stations. The performance of WLSSVM in predicting the streamflow is superior to individual LSSVM model. As seen from the fit line equations (assume that the equation is $y = a + bx$) in the scatterplots that a and b coefficients for the LSSVM and WLSSVM models, respectively, the WLSSVM has less scattered estimates and the R value of WLSSVM model close to 1 ($R = 0.891$ for Klang and $R = 0.864$ for Langat) compared to the LSSVM model. Overall, it can be concluded that the WLLSVM model at both studies provided more accurate forecasting results than the LSSVM model for streamflow forecasting.

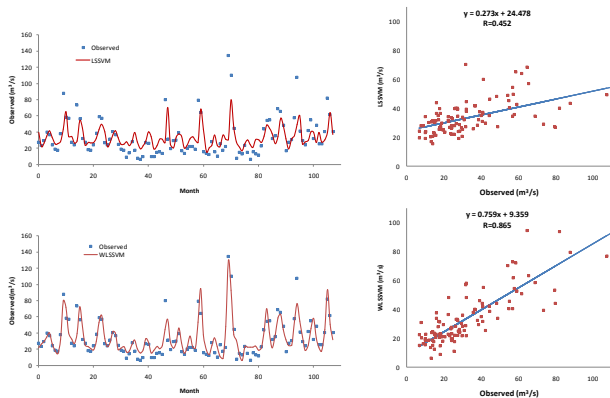


Figure 3 Optimal LSSVM and WLSSVM models during the test period for monthly streamflows of Klang River

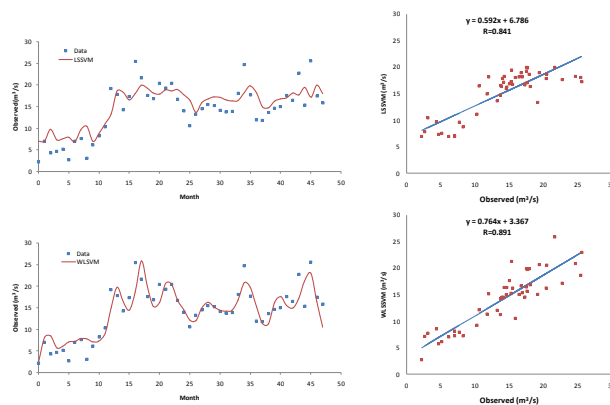


Figure 4 Optimal LSSVM and WLSSVM models during the test period for monthly streamflows of Langat River

In addition, the streamflow forecasting has been carried out by ARIMA models for the purpose of comparison. For illustration, an example from Klang station was described briefly. The sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) for Klang series is plotted in Fig. 5.

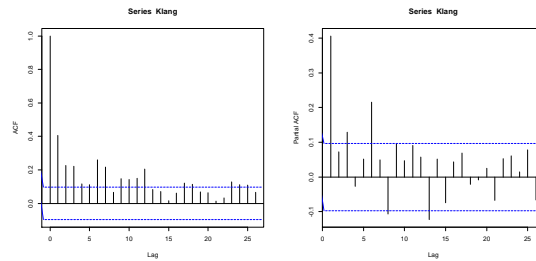


Figure 5 ACF and PACF plots used for the selection of ARIMA model for Klang station.

The ACF is damping out in exponential waves with significant spikes at lag 1, 2, 3, 6, 7 and 9 to 12. The PACF has significance values at lag 1, 3, 6, 8 and 13. This may imply the presence of seasonal and non-seasonal AR and MA operators for the monthly series. The best model from different candidate models was identified using the minimum Akaike Information Criterion (AIC). Different ARIMA models along with their Ljung-Box, Q(r) test and AIC values are shown in Table 4 below. It was observed that the selected ARIMA models passed all diagnostic checks and the ARIMA (3,0,3)×(2,0,0)₆ is the best model. The residual ACF (RACF) and the PACF (RPACF) of the best model are demonstrated in Fig. 6. The RACF and RPACF lie within the confidence limits, which clearly supports the fact that the residuals from the best model are white noise.

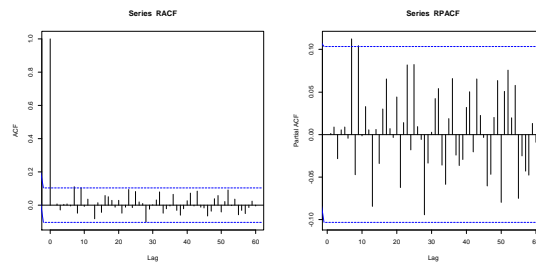


Figure 6 ACF and PACF of residuals for ARIMA (3,0,3)×(2,0,0)₆ model

For Langat station, three models were initially selected based on the AIC and the Ljung-Box statistic. The identification of the best model for stream flow series based on minimum AIC and Ljung-Box statistics is shown in Table 4. The best model is based on the AIC shows that ARIMA(3,0,3)×(2,0,0)₆ and ARIMA (0,0,6)×(0,1,4)₁₂ is the best model for Klang station and Langat station, respectively. Inspection of the Ljung-Box test confirmed that the best model is adequate.

Table 4 Comparison of AIC and Ljung-Box Statistics for selected ARIMA model

Station	ARIMA model	AIC	Q(r)	p-values
Klang	(3,0,0)×(2,0,0) ₆	2548.79	53.74	0.126
	(3,0,3)×(2,0,0)₆	2543.92	44.54	0.287
	(3,0,3)×(2,0,2) ₆	2547.53	42.08	0.299
Langat	(1,0,6)×(4,1,0) ₁₂	3717.7	49.91	0.076
	(0,0,6)×(4,1,0)₁₂	3716.1	50.00	0.092
	(0,0,3)×(0,1,4) ₁₂	3686.0	44.99	0.203

For further analysis, the best performance of the ARIMA, LSSVM and WLSSVM models in terms of the MSE, MAE and R are

compared. Table 5 represents the results of Klang and Langat stations study site in terms of various performance statistics.

Table 5 The Performance Results of ARIMA, LSSVM and WLSSVM Approach During Training and Testing Period.

Data	Model	Training			Testing		
		RMSE	MAE	R	RMSE	MAE	R
Klang	ARIMA	8.05	5.24	0.48	4.63	3.56	0.66
	LSSVM	6.45	4.27	0.72	3.47	2.83	0.84
	WLSSVM	4.56	3.04	0.87	2.71	2.07	0.89
Langat	ARIMA	14.44	9.48	0.70	16.54	10.29	0.71
	LSSVM	16.81	11.82	0.54	19.65	13.20	0.54
	WLSSVM	8.480	6.146	0.91	11.80	8.73	0.87

Table 5 above clearly shows that WLSSVM model performs much better than the single LSSVM and ARIMA models especially for the long term streamflow forecasting.

6.0 CONCLUSION

The potential of wavelet least square support vector machines (WLSSVM) model for 1-month ahead stream flow forecast has been presented. The proposed model based on the WLSSVM model were developed by combining two methods which are discrete wavelet transforms (DWT) and least square support vector machines (LSSVM) model. The monthly stream flow time series was decomposed at different decomposition level by DWT. Each of the decomposition carried most of the information and plays a distinct role in original time series. The correlation coefficients between each of sub-series and original series were used for the selection of the LSSVM model inputs and for the determination of the effective wavelet components on stream flow. For the input of LSSVM model, sum of effective details and the approximation component are used. The WLSSVM model were trained and tested by applying different input combinations of monthly stream flow data of Klang and Langat stations in Peninsular Malaysia. The performance of the proposed WLSSVM model was compared to individual LSSVM model, and was also compared with the conventional ARIMA models for monthly stream flow forecasting.

Comparison results indicate that the WLSSVM model was found to be significantly superior compared to the one that obtained by conventional LSSVM and ARIMA models. The study concludes that the forecasting abilities of the LSSVM model are found to be improved when the wavelet transformation technique is adopted for the data pre-processing. Moreover, the decomposed periodic components that obtained from the DWT technique are found to be most effective in yielding accurate forecast when used as inputs in the LSSVM model. Thus, the accurate forecasting results for both stations indicate that WLSSVM model provides a superior alternative to LSSVM and ARIMA models, and potentially a useful tool for stream flow forecasting.

Acknowledgement

The authors thankfully acknowledged the financial support that afforded by Universiti Teknologi Malaysia under FRGS Grant (vot. 4F088). Besides that, the authors would like to thank the Department of Irrigation and Drainage, Ministry of natural Resources and Environment, Malaysia in helping us to provide the data.

References

- [1] Huang, W., Bing, Xu, B., and Hilton, A. 2004. Forecasting flow in apalachicola river using neural networks. *Hydrological Processes*. 18: 2545–2564.
- [2] Yurekli, K., Kurunc, A., and Simsek, H. 2004. Prediction of daily streamflow based on stochastic approaches. *Journal of Spatial Hydrology* 4(2):1–12.
- [3] Muhamad, J.R., and Hassan, J.N. 2005. Khabur River flow using artificial neural networks. *Al-Rafidain Engineering*. 13(2): 33 – 42.
- [4] Modarres, R. 2007. Streamflow drought time series forecasting. *Stoch. Environ. Res. Risk Assess.* 21: 223–233.
- [5] Fernandez, C., and Vega, J. A. 2009. Streamflow drought time series forecasting: a case study in a small watershed in north west Spain. *Stoch. Environ. Res. Risk Assess.* 23: 1063–1070.
- [6] Wang, W.C., Chau, K.W., Cheng, C.T., and Qiu, L. 2009. A comparison of performance of several Artificial Intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*. 374: 294–306.
- [7] Maier, H.R., and Dandy, G.C. 2000. Neural networks for the production and forecasting of water resource variables: a review and modelling issues and application. *Environmental Modelling and Software*. 15: 101–124.
- [8] Shamseldin, A., Ahmed, E., Nasr, and O'Connor, K.M. 2002. Comparison of different forms of the multi-layer feed-forward neural network method used for river flow forecasting. *Hydrol. and Earth Syst. Sci.* 6(4): 671–684.
- [9] Cigizoglu, H. K. 2003. Incorporation of ARMA models into flow forecasting by artificial neural networks. *Environmetrics*. 14(4): 417–427.
- [10] Gopakumar, R., Takara, K., and James, E.J. 2007. Hydrologic data exploration and river flow forecasting of a Humid Tropical River Basin using artificial neural networks. *Water Resoure Management*. 21: 1915–1940.
- [11] Firat, M. 2008. Comparison of Artificial Intelligence Techniques for river flow forecasting. *Hydrol. Earth Syst. Sci.* 12: 123–139.
- [12] Keskin, M.E., and Taylan, D. 2009. Artifical models for interbasin flow prediction in southern Turkey. *Journal Hydrologic Engineering*. 14(8): 752–758.
- [13] Wu, C.L., and Chau, K.W. 2010. Data-driven models for monthly streamflow time series prediction. *Engineering Applications of Artificial Intelligence*. 23: 1350–1367.
- [14] Guo, J., Zhou, J., Qin, H., Zou, Q., and Li, Q. 2011. Monthly streamflow forecasting based on improved support vector machine model. *Expert Systems with Applications*. 38: 13073–13081.
- [15] Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B., and Vandewalle, J. 2002. *Least Squares Support Vector Machines*. World Scientific, Singapore.
- [16] Hanbay, D. 2009. An expert system based on least square support vector machines for diagnosis of the valvular heart disease. *Expert Systems with Applications*. 36(3): 4232–4238.
- [17] Kang, Y.W., Li, J., Cao, G.Y., Tu, H.Y., Li, J., and Yang, J. 2008. Dynamic temperature modeling of an SOFC using least square support vector machines. *J. Power Sources*. 179: 683–692.
- [18] Ismail, S., Shabri, A., and Samsudin, R. 2011. A hybrid model of self-organizing maps (SOM) and least square support vector machine (LSSVM) for time-series forecasting. *Expert Systems with Applications*. 38(8): 10574–10578.
- [19] Yunrong, X., and Liangzhong, J. 2009. Water quality prediction using LS-SVM with particle swarm optimization. *Second International Workshop on Knowledge Discovery and Data Mining*. 900–904.
- [20] Okkan, U., and Serbes, A.Z. 2012. Rainfall-runoff modeling using least squares support vector machines. *Environmetrics*. 23(6):549–564.
- [21] Ismail, S., Samsudin, R., and Shabri, A. 2010. River flow forecasting: a hybrid model of self organizing maps and least square support vector machine. *Hydrol. Earth Syst. Sci. Discuss.* 7: 8179 – 8212.
- [22] Shabri, A., and Suhartono. 2012. Stremflow forecasting using least-squares support vector machines. *Hydrological Science Journal*. 57(7): 1275–1293.
- [23] Wang, W., and Ding, J. 2003. Wavelet network model and its application to the prediction of the hydrology. *Nat. Sci.* 1(1): 67–71.
- [24] Labat, D., Ababou, R., and Mangin, A. 2000. Rainfall–runoff relations for karstic springs. Part II. Continuous wavelet and discrete orthogonal multidecomposition analyses. *Journal of Hydrology*. 238(3–4):149–178.
- [25] Kisi, O. 2008. River flow forecasting and estimation using different artificial neural network techniques. *Hydrology Research*. 39(1): 27– 40.
- [26] Kisi, O. 2009. Neural network and wavelet conjunction model for modeling monthly level fluctuations in Turkey. *Hydrological Processes*. 23: 2081–2092.
- [27] Kisi, O. 2010. Wavelet regression model for short-term streamflow forecasting. *Journal of Hydrology*. 389: 344–353.

- [28] Partal, T., and Cigizoglu, K. 2008. Estimation and forecasting of daily suspended sediment data using wavelet-neural network. *Journal of Hydrology*. 358: 317–331.
- [29] Pramanik, N., Panda, R. K., and Singh, A. 2010. Daily river flow using wavelet ANN hybrid models. *Journal of Hydroinformatics*. 1–15.
- [30] Cannas, B., Fanni, A., See, L., and Sias, G. 2006. Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. *Physics and Chemistry of the Earth*. 31: 1164–1171.
- [31] Adamowski, J., and Sun, K. 2010. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*. 390(1–2): 85–91.
- [32] Partal, T., and Kisi, O. 2007. Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. *Journal of Hydrology*. 342:199–212.
- [33] Rajae, T., Mirhagher, S.A., Nourani, V., and Alikhani, A. 2010. Prediction of daily suspended sediment load using wavelet and neuro-fuzzy combined model. *Int. J. Environ. Sci. Tech.* 7(1): 93–110.
- [34] Nourani, V., Alami, M.T., and Aminfar, M.H. 2009. A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. *Engineering Applications of Artificial Intelligence*. 22: 466–472.
- [35] Nourani, V., Komasi, M., and Mano, A. 2009. A multivariate ANN-Wavelet approach for rainfall-runoff modeling. *Water Resour Manage.* 23: 2877–2894.
- [36] Kisi, O., and Cimen, M. 2012. Precipitation forecasting by using wavelet-support vector machine conjunction model. *Engineering Applications of Artificial Intelligence*. 25(4): 783–792.
- [37] Forecasting. *Water Resource Management*. 25(2): 579–600.
- [38] Liu, L., and Wang, W. 2008. Exchange rates forecasting with least squares support vector machines. *International Conference on Computer Science and Software Engineering*. 1017–1019.
- [39] Gencoglu, M.T., and Uyar, M. 2009. Prediction of flashover voltage of insulators using least square support vector machines. *Expert Systems with Applications*. 36: 10789–10798.
- [40] Kim, T.W., and Valdes, J.B. 2003. Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. *Journal Hydrologic Engineering*. 8(6): 319–328.
- [41] Mallat, S.G. 1989. A theory for multi decomposition signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 11(7): 674–693.