Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/090622

A Hybrid Model for Monthly Time Series Forecasting

*Siraj M. Pandhiani** *and Ani Bin Shabri**

Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia, Skudai, Malaysia.

Received: 21 Feb. 2015, Revised: 21 Apr. 2015, Accepted: 22 Apr. 2015
Published online: 1 Nov. 2015

Abstract: This study aims to propose a hydrological model for estimating the future value for monthly river flow. The proposed model was constructed by combining three components: i.e. Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA) and Least Square Support Vector Machine (LSSVM). The first two components, i.e. the wavelets and the PCA, are meant for preparing input data. Wavelets were employed to obtain a certain level of data decomposition, and in this case, a three level decomposition was employed. The output from the wavelets was given to PCA. This component simply picks up the important components from the given data, i.e. it addresses the issues relating to the dimensionality of the data. For approximating the desired value, LSSVM was employed for training, using the data derived from Wavelets and PCA models. For testing stability and reliability of the proposed model monthly data from two Pakistani rivers was collected. The reliability was measured by employing well known reliability measuring methods, i.e. Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Correlation (R). All performance measuring methods concluded that the proposed model is stable, reliable and produced an appreciative level of accuracy.

Keywords: Discrete Wavelet Transform, Principle Component Analysis, Least Square Support Vector Machine, Root Mean Square Error, Mean Absolute Error, Coefficient of Correlation, hybrid model

1 Introduction

Hydrological processes are well known for critically studying the temporal and spatial variability of water. Certain important factors like climatic changes and physiographic elements are complicating hydrological processes. Various conventional approaches have been used to model hydrological processes. These conventional approaches either aim to model the process by considering physical characteristics of the system or they depend upon experimental observations to model hydrological process [1]. For managing available water resources, all components forming a hydrological structure need careful examination. River flow forecasting is one such component. Since rivers are directly or indirectly connected to the streams, there is a strong need to have some mechanism that can provide a reliable estimation of the water flowing in a stream. This estimation is expected to contribute significantly in addressing a number of hydraulic issues. For example, these estimations would help in coming up with a good, easy to build and easy to understand water supply plan and its operation. Besides, they can contribute to addressing hydraulic engineering related problems. For example, designing dams for balancing water flow or for defining various hydraulic structures [2].

Presently, a number of approaches are used to solve the problem of river flow process. These approaches are described in the form of mathematiFcal models. These models are found to be compatible with the other relevant models [3]. This compatibility creates chances for introducing hybrid models. Hybrid models are acknowledged for ensuring a higher degree of accuracy and reliability. According to [4], a number of approaches were introduced in the past few decades to address problems belonging to diversified fields, including river flow forecasting. These approaches include regression analysis and ANN, stochastic models, fuzzy mathematics, etc.

The present study uses time series data. Until now, wavelets are the best known tool for analyzing this type of data. Its contribution to modeling hydrological resources can be seen in the last few years [5], [6]. These include meteorological pollution simulation [7], [8], open channel wake flows analysis [9], and groundwater level time series modeling [10]. Recently, wavelet theory has been introduced in the field of hydrology [11], [12]. Wavelet models, because of their natural ability to analyse a signal

* Corresponding author e-mail: pandhiani@hotmail.com ani@utm.my

in time and frequency domains, are becoming a general choice for researchers addressing issues related to hydrological models. [13], [14]. They are reliable for analyzing non-stationary signals [15].

Wavelets and Principle Component Analysis (PCA) carry a higher degree of mutual compatibility. Wavelets are able to decompose the given data, which ensures computational time reduction. PCA is another important component, which is known for dimensionality reduction [16], [17]. When both of these components are joined together, they form an optimized system. This system can successfully reduce the dimension of wavelets coefficients, with the reduced set of coefficients; it forms a time efficient system. This characteristic creates an opportunity for designing and developing real time systems.

The third component of the proposed model is Least Square Support Vector Machine (LSSVM). LSSVM originates from the Support vector machine (SVM). SVM works on the idea of supervised learning. Supervised learning needs some training and testing data. The system is trained with the help of training data and an estimated value is obtained on the basis of this trained data. This is similar to the concept that is applied in Artificial Neural Network (ANN). Like ANN, SVMs were also successfully applied to facilitate hydrological structures for solving their issues [18], [19], flood stage forecasting [20] and rainfall-runoff modeling [21], [22]. Though SVMs were good and produced encouraging results, what discourages their frequent use is that they are computationally expensive. They depend upon complex quadratic programming (QP) [23].

In order to make SVMs computationally inexpensive, without compromising its reliability and accuracy, LSSVM were introduced [24]. This approach depends on computing the least square error by considering the input vectors and the obtained vectors (results). The inclusion of this step ensures a higher level of accuracy in comparison to SVM. Besides, LSSVM is found suitable for solving linear equations, which is a much needed characteristic. Unlike SVM, LSSVM works under the influence of equality constraints. Equality constraints are instrumental in reducing computational speed. LSSVM provides an appreciative level of precision and carries good convergence [25], [26]. LSSVM is still considered to be in its developing stages and is rarely used in addressing the problems of hydrological modeling [23]. However, the model was successfully applied for solving regression, pattern recognition problems [27], [28] and for modeling ecological and environmental systems [29]. Technically both of these methods, i.e. SVM and LSSVM are equally effective. In terms of implementation LSSVM is comparatively easier than SVM. In terms of their generalized performance both of them are comparable [30] and are found reliable.

However, LSSVM is found to be good in terms of stability and accuracy and generally, it seems to be a good choice for training data. The only concern is that for getting generalized results, like any other AI based technique, LSSVM also requires a huge amount of data for training purposes. This need for a large amount of data makes LSSVM difficult to use as a training tool. Therefore, there is a strong need to utilize the capability of LSSVM by optimizing the input data. The present research addresses this issue and successfully optimizes the training and testing data.

This research aimed to propose a hybrid model for addressing the hydrological issue of estimating future values. It attempted to achieve this by combining three independent components. All of the chosen components are compatible with each other and produced appreciative results. In order to establish a clear difference various configurations were tried. For example, the required values were independently estimated by using LSSVM, by using wavelets combined with LSSVM and finally by combining wavelets, PCA and LSSVM. Results from these models were recorded. A mutual comparison was drawn and it was concluded that the WPLSSVM performed well and produced better results.

2 Methods and Materials

2.1 Data Preparation

The experimental data for carrying out this research was collected by recording the monthly stream flow of two rivers in Pakistan, namely; Jhelum and Chenab. The collected data ranges from 1971 to 2010 for Jhelum and 1956 to 2010 for Chenab. The input data which ensures the reliability of the proposed model is downloaded from Pakistan Meteorological Department (PMD). This is benchmarked data and is frequently cited in the research literature regarding water flow forecasting. The data is being used by a Government supervised research center. This research center specifically carries research on managing water resources and studies, investigates and proposes models for addressing water flow. This is a raw set of data and needs normalization. This study normalizes the collected data with the help of a mixture of Gaussians. This normalization technique is chosen for its simplicity, effectiveness, easy implementation and its widespread recognition. In short, this is a well-studied, well researched and well established method for doing statistical analysis.

2.2 Wavelet Analysis

Wavelets are becoming an increasingly important tool in time series forecasting. The basic objective of wavelet transformation is to analyse the time series data, both in the time and frequency domains, by decomposing the original time series in different frequency bands, using

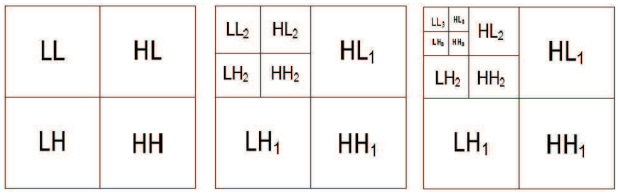

Fig. 1 Discuss with diagram DWT and Multiresolution Analysis.

wavelet functions. A wavelet function for a continuous time signal can be described with the help of Equation (1)

$$\psi(t,s) = \frac{1}{\sqrt{x}}\psi(\frac{t-\tau}{sn}) \tag{1}$$

Where $'t'$ represents time factor, $t-\tau$ represents the shifted version for studying the input signal at different time interval, $'S'$ represents the scale and $\psi(t)$ represents the mother wavelet. Wavelets are analyse the input signal by using $'shift'$ and $'scale'$ functions. The shifting and scaling of the input data is achieved with the help of Equation 2 and 3

$$\phi(t) = \sum_{n=-\infty}^{\infty} g_n 2^{\frac{1}{2}} \phi(2t-n) \tag{2}$$

$$\psi(t) = \sum_{n} d(n)\varphi(2t-n) \tag{3}$$

Since this study uses discrete data therefore, DWT is employed to get the required level of decomposition. A three level decomposition is shown in Fig 1. The decomposition of data at various levels creates an opportunity for critically analyzing the data at different levels of resolution. This is what is generally known as the concept of multiresolution. Advantages of multiresolution are easily visible. It ensures every single bit of information and the probability for the loss of some data value is almost negligible. As a result, the true strength of wavelets can be realized, i.e. decomposing the data at various levels, ensuring maximum level of information and maintaining reliability.

2.3 Principal Component Analysis

PCA is generally used for reducing the dimensionality of the data. The need for reducing the dimensionality of the data arises from the actual volume of the data. If the whole volume of the data is used for training LSSVM then it is likely to cause two important impacts. (i) The system would be overloaded in terms of main memory. (ii) It would require a larger amount of time to train the system. Practically, both of these impacts are unaffordable. Therefore in order to fix it the present research makes use of PCA for reducing the coefficients of wavelets. These outputs produced by PCA are later on supplied to LSSVM to produce a trained data file. To achieve the required PCAs, the study follows a certain procedure. The data used for preparing PCAs is shown according to Equation 4.

$$Y_k = a_{k1}X_1 + a_{k2}X_1 + ... + a_{kk}X_k \tag{4}$$

Since PCA focuses on capturing a large variance, it computes the variance of the data with the help of Equation (5)

$$Var[y] = E((y-\overline{y})^2) = \frac{1}{n}\sum_{k=1}^{n}(a_1^T x_k - n_1^T\overline{x})^2 = \frac{1}{n}\sum_{k=1}^{n}(a_1^T(x_k-\overline{x})(x_k-\overline{x})^T a_1 Var[y] = (a_1^T x_k - n_1^T S a_1) \tag{5}$$

The study computes the co-variance of input data with the help of Equation (5)

$$S = \frac{1}{n}\sum_{k=1}^{n}(a_1^T(x_k-\overline{x})(x_k-\overline{x})\ is the covariance \tag{6}$$

The covariance matrix of Y is a diagonal matrix, cov $(Yi, Yi+1) = 0$;

The study computes the Eigenvectors of input data with the help of Equation (5)
Let λ be a Lagrange multiplier

$$L = a_1^T S a_1 - \lambda(a_1^T a_1 - 1) \Longrightarrow \nabla_{a_1} L = Sa_1 - \lambda a_1 = 0 \Longrightarrow (S-\lambda I_p)a_1 = 0, \tag{7}$$

where I_p is a $p \times p$ identity matrix. Therefore a1 is an eigenvector of S.
It computes the Eigenvalues of input data with the help of Equation (7) the ordered p characteristic roots (eigenvalues) as:

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \tag{8}$$

The study collects the principal components in a matrix as shown below.
Let X be the covariance matrix of input samples $x = (x_1 x_2 x_3 ... x_k)^T$ there eigenvalues and eigenvector can be expressed as $(\lambda_1, e_1), (\lambda_2, e_2), ..., (\lambda_k, e_k)$; so the ith principle component is:

$$Y_i = \varepsilon_i^T X \tag{9}$$

Then $Var(Y_i) = \varepsilon_i^T \sum \varepsilon_i = \lambda_i, (i = 1,2,...,k)$. The eigenvalue illustrates the proportion of the information that the correlative principle component includes, namely; it denotes the variation extent of the samples in this direction. Eigenvalues measure the amount of the variation explained by each PC and will be largest for the first PC and smaller for the subsequent PCs. Thus, the autocorrelation of space will be strengthened after using it as a variable. Finally, the total contribution rate of the $i-th$ principle component is defined by the following relation.

$$\beta_i = \frac{\sum_{p=1}^{i}\lambda_k}{\sum_{p=1}^{k}\lambda_k} \tag{10}$$

Where, the value of variable p predicated from the first k factors, λ_k , refers to the kth eigen value.

The output obtained from PCA which is in the form of wavelets coefficients is used as an input to train LSSVM. This trained data file helps to estimate new data values. The following section explains steps for understanding mathematical structure of LSSVM.

2.4 The Least Square Support Machines Model

LSSVM optimizes SVM by replacing complex quadratic programming. It achieves this by using least squares loss function and equality constraints. For the purpose of understanding the construction of the model, we might consider a training sample set represented by (x_i, y_i) where x_i represents the input training vector. Suppose that this training vector belongs to $'n'$ dimensional space, i.e. R^n, so we can write $x_i \in R$. Similarly, suppose that yi represents the output and this output can be described as, $y_i \in R$.
SVM can be described with the help of Equation (11)

$$y(x) = w^T \phi(x) + b \tag{11}$$

Where $\phi(x)$ is a function that ensures the mapping of nonlinear values into higher dimensional space.
LSSVM formulates the regression problem according to Equation (12)

$$minR(w,\varepsilon) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum \lim_{i=1}^{n} \varepsilon_i^2 \tag{12}$$

The regression model shown in Equation (11) works under the influence of equality constraints

$$y(x) = w^T \phi(x_i) + b + \varepsilon_i, i = 1,2,...,n \tag{13}$$

It introduces Lagrange multiplier for

$$L(w,b,\varepsilon,\alpha) = \tfrac{1}{2} w^T w + \tfrac{\gamma}{2} \sum \lim_{i=1}^{2} \varepsilon_i^2 - \sum \lim_{i=1}^{n} \alpha_i w^T \phi(x_i) + b + \varepsilon_i - y_i \tag{14}$$

Where, α represents Lagrange multipliers. Since Equation (14) involves more than one variable, a partial differentiation of Equation (15-18) is required for studying the rate of change. Therefore, differentiating Equation (14) with respect to w, b, ε_i and α_i and equating them with zero, yields the following set of Equations.

$$\frac{\partial L}{\partial w} = 0 \longrightarrow w = \sum_{i=1}^{n} \alpha_i \phi(x_i) \tag{15}$$

$$\frac{\partial L}{\partial b} = 0 \longrightarrow w = \sum_{i=1}^{n} \alpha_i = 0 \tag{16}$$

$$\frac{\partial L}{\partial \varepsilon_i} = 0 \longrightarrow \alpha_i = \gamma \varepsilon_i \tag{17}$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \longrightarrow w^T \phi(x_i) + b + \varepsilon_i - y_i = 0, i = 1,2,...,n \tag{18}$$

Substituting Equation (15-18) for Equation (14) we get the value of w. This w is described according to Equation (19)

$$w = \sum \lim_{i=1}^{n} \alpha_i \phi(x_i) = \lim_{i=1}^{n} \gamma \varepsilon_i \phi(x_i) \tag{19}$$

Putting Equation (19) in Equation (13)

$$y(x) = \sum \lim_{i=1}^{n} \alpha_i \phi(x_i)^T \phi(x_i) + b = \lim_{i=1}^{n} \alpha_i K(x_i, x) + b \tag{20}$$

Where, $K(xi,x)$, represents a kernel such that:

$$K(x_i, x) = \phi(x_i)^T \phi(x_i) \tag{21}$$

The α vector which is a Lagrange Multiplier can be computed by solving a set of linear equations shown in Equation (22)

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

Where, $y = [y_1; ...; y_n], 1 = [1; ...; 1], \alpha = [\alpha_1; ...; \alpha_n]$ This eventually constitutes LSSVM model which is described according to Equation (23).

$$y(x) = \sum \lim_{i=1}^{n} \phi_i K(x_i, x) + b \tag{22}$$

The model shown in Equation (23) deals with a linear system and solution. This linear system is provided by α_i, b. The high dimensional feature space is defined by a function generally known as $'kernel\,function'$ and is represented by $K(x_i, x)$ There are various choices available for picking up this function.

3 Accuracy of the Proposed Model

The strength of the proposed WPLSSVM when compared with other methods was assessed in the light of three well-known performance measuring methods. The first method computes the Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and correlation (R). The mathematical representation of all these performance measuring methods is given as follows:-

$$MSE = \frac{1}{n} \sum_{t=1}^{n} |y_t^0 - y_t^f|^2 \tag{23}$$

$$MAE = \frac{1}{n} \sum_{t=1}^{n} (y_t^0 - y_t^f)^2 \tag{24}$$

$$R = \frac{\frac{1}{n} \sum_{t=1}^{n} (y_t^0 - \overline{y}_t^0)(y_t^f - \overline{y}_t^f)}{\sqrt{\frac{1}{n} \sum_{t=1}^{n} (y_t^0 - \overline{y}_t^0)^2} \frac{1}{n} \sum_{t=1}^{n} (y_t^f - \overline{y}_t^f)^2} \tag{25}$$

Where y_t^0 and y_t^f are the observed and forecasted values at time t, respectively and n is the number of data points. The criteria to judge for the best model are relatively small of MAE and MSE in the training and testing. The correlation coefficient measures how well the flows predicted correlate with the flows observed.
Clearly, the R-value close to unity indicates a satisfactory result, whiles a low value or close to zero implies an inadequate result. The MAE is related with the prediction bias whereas the RMSE is associated with the model error variance. Both MAE and RMSE evaluate how closely the predictions match the observations by judging the best model based on the relatively small MAE and RMSE values. Clearly, the R value close to unity indicates a satisfactory result, while a low value or close to zero implies an inadequate result. R ranges from -1 to $+1$ for a perfect model.

4 Fitting of the Data

This section describes the fitting of the data to the three different configurations. In the first configuration only LSSVM was used. In the second configuration DWT is combined with LSSVM and in the third configuration, which is the proposed configuration of the present research, DWT is combined with PCA which is further connected with LSSVM. This is how it composes a new model and at the same time it draws mutual comparison among the contemporary models. The proposed arrangement is found to be reliable, as it comfortably deals with non-linear data. On top of it, it has also brought forward promising results. With this arrangement a specimen of six different combinations of input data were prepared in table 1. This data was used for training LSSVM.

Table 1 The Input structure of the Models for Forecasting.

Input	Input Structure
M1	$y_t = f(x_{t-1}, x_{t-2})$
M2	$y_t = f(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4})$
M3	$y_t = f(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6})$
M4	$y_t = f(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}, x_{t-8})$
M5	$y_t = f(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}, x_{t-8}, x_{t-9}, x_{t-10})$
M6	$y_t = f(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}, x_{t-8}, x_{t-9}, x_{t-10}, , x_{t-11}, x_{t-12})$

4.1 Fitting LSSVM to the Data

The performance of LSSVM depends upon the training data. The training data, which is prepared carefully, contributes to ensuring the reliability of the system. The training and testing data is divided into different ranges. For example, it was divided from M1 – M6. Optimal model parameters ensure the success of the LSSVM. These parameters include (γ, σ^2). Where Gamma ranges from 10 to 1000 and σ^2 ranges from 0.01 to 1.0. These parameters were obtained by using grid search algorithm and cross validation methods. LSSVM needs a kernel function to produce an output. As mentioned above there are various choices of kernels. All of them are equally good [31], [32]. This study uses radial basis function (RBF) as a kernel function for stream flow forecasting. The purpose of employing cross validation is to calibrate the model parameters. Once these model parameters are calibrated successfully they save the model from undergoing over fitting. In order to predict the prediction error for each hyper parameter pair a 10 fold cross validation was performed. This produces a best fit model in the light of performance measuring criteria.
Table 2 and 3 shows the performance results obtained in the training and testing period of the LSSVM approach (i.e. those using original data). For the training and testing phase in Jhelum stations, the best values of the MSE (0.0048 and 0.0052), MAE (0.0481 and 0.0541) and R (0.9104 and 0.8353) respectively, were obtained using ML6 (training) and ML5 (testing). For the Chenab River training and testing phase, the observed value of MSE and MAE were obtained using ML5 (training) and ML6 (testing). In the training and testing model had the smallest MAE (0.0031 and 0.0021) and MSE (0.0345 and 0.0317) whereas it had the highest value of the R (0.9442 and 0.9457).

Table 2 Forecasting performance indicates of LSSVM for Jhelum River of Pakistan

	Training			Testing		
Model	MSE	MAE	R	MSE	MAE	R
ML1	0.0073	0.0589	0.8586	0.0072	0.0613	0.7827
ML2	0.0060	0.0543	0.8857	0.0058	0.0583	0.8199
ML3	0.0060	0.0536	0.8856	0.0054	0.0572	0.8360
ML4	0.0054	0.0498	0.8998	0.0059	0.0573	0.8168
ML5	0.0051	0.0497	0.9043	**0.0052**	**0.0541**	**0.8353**
ML6	**0.0048**	**0.0480**	**0.9104**	0.0053	0.0543	0.8341

Table 3 Forecasting performance indicates of LSSVM for Chenab River of Pakistan

	Training			Testing		
Model	MSE	MAE	R	MSE	MAE	R
ML1	0.0073	0.0589	0.8586	0.0072	0.0613	0.7827
ML2	0.0060	0.0543	0.8857	0.0058	0.0583	0.8199
ML3	0.0060	0.0536	0.8856	0.0054	0.0572	0.8360
ML4	0.0054	0.0498	0.8998	0.0059	0.0573	0.8168
ML5	**0.0051**	**0.0497**	**0.9043**	0.0052	0.0541	0.8353
ML6	0.0048	0.0480	0.9104	**0.0053**	**0.0543**	**0.8341**

The discussion presented in the preceding paragraphs shows the performance of LSSVM as in the original model, i.e. it is not combined with any other

component(s). The performance produced by this model is recorded for drawing comparisons with the other competing model. In this regard the next section combines LSSVM with DWT. As established earlier these combinations are attempted under a single objective, i.e. how to improve the performance of proposed models without losing the level of accuracy.

4.2 Fitting Wavelet and LSSVM to the Data

In an effort to improve the performance of the model used, this brings to light the performance of the model which is made by combining two components. The candidate components are DWT and LSSVM. An important and immediate benefit of DWT is its ability to decompose the given data. With this decomposition data integrity is ensured and the original data is recovered at certain stage of execution. The decomposed data is expected to put a minimal load on system resources, both in terms of software and hardware. Once this load is properly managed it ensures system optimization.
Before training LSSVM, DWT [33], [34] was employed to decompose the given time series data. The decomposition of data is a finite step and it is expected to provide acceptable results by using a specific level of decomposition. In order to know that specific level, generally the relation, $M = log(n)$ is used [35]. This relationship describes length of time series data in terms of n and levels of decompositions in terms of M. For testing the aforementioned model, this study keeps $n = 480$ and 550. This represents the monthly data for each river, i.e. for Jhelum and Chenab. With these values of $'n'$ the decomposition levels $'M'$ are found to be 3 [36], [37].

The decomposed components produced by DWT are of two types known as $'significant'$ components and $'approximated'$ components. The significant components represented by $'Ds'$ whereas, approximated components are represented by $'As'$. $'D1'$ represents the time series data covering two months' time, $'D2'$ represents the time series data describing four months' time and $'D3'$ represents the same data covering eight months' time. The training data for LSSVM is prepared by adding together $'D2'$, $'D3'$ with $'A3'$. Figure 2 and Figure 3 show the original streamflow data time and their Ds. From this combination $'D1'$ is dropped mainly because it produces low co-relation. This combination is found to be effective. The results obtained by employing this combination are discussed in Table 4 and 5. It shows that the WLSSVM model has a significant positive effect on the streamflow forecast. As seen from Table 4, for the Jhelum station, the MWL4 model has the smallest MSE (0.0017) and MAE (0.0306) and the highest R (0.9788) in the training phase. However, for the testing phase, the best MSE (0.0027), MAE (0.0383) and R (0.9517) were obtained for the model input combination MW4. From Table 9 for Chenab station, the MWL6 model has the smallest MSE (0.0007) and MAE (0.0172) and the highest R (0.9880) in the training phase. However, for the testing phase, the best MSE (0.0011) and MAE (0.0248) and R (0.9729) were obtained for the model input combination MW6.

Table 4 Forecasting performance indicates of Wavelet + LSSVM (WLSSVM) for Jhelum River of Pakistan

	Training			Testing		
Model	MSE	MAE	R	MSE	MAE	R
MWL1	0.0330	0.1519	0.4601	0.0209	0.1226	0.5575
MWL2	0.0026	0.0381	0.9692	0.0034	0.0441	0.9415
MWL3	0.0025	0.0366	0.9696	0.0029	0.0405	0.9481
MWL4	**0.0017**	**0.0306**	**0.9788**	**0.0027**	**0.0383**	**0.9517**
MWL5	0.0024	0.0361	0.9711	0.0029	0.0395	0.9487
MWL6	0.0024	0.0356	0.9716	0.0030	0.0387	0.9480

Table 5 Forecasting performance indicates of Wavelet + LSSVM (WLSSVM) for Chenab River of Pakistan

	Training			Testing		
Model	MSE	MAE	R	MSE	MAE	R
MWL1	0.0198	0.1071	0.5775	0.0151	0.0969	0.5355
MWL2	0.0013	0.0251	0.9788	0.0016	0.0304	0.9618
MWL3	0.0008	0.0185	0.9861	0.0011	0.0251	0.9730
MWL4	0.0011	0.0204	0.9825	0.0014	0.0279	0.9678
MWL5	0.0012	0.0227	0.9804	0.0014	0.0276	0.9653
MWL6	**0.0007**	**0.0172**	**0.9880**	**0.0011**	**0.0247**	**0.9735**

4.3 Fitting Wavelet, PCA and LSSVM to the Data (WPLSSVM)

The objective is to optimize the system in term of computational resources and accuracy. The same time series data as mentioned in the preceding models, i.e. the models which are discussed above. The results obtained by using WPLSSVM were found to be highly accurate; this is found to be reliable and stable. The mutual comparison of WPLSSVM with LSSVM and WLSSVM showed that the proposed model, over above its performance superiority, has the capability of generalization.
The significant difference between WPLSSVM, LSSVM and WLSSVM is that the proposed model holds an additional component, i.e. PCA. The objective of PCA as mentioned above is just to reduce the dimensionality of the data. In the present experimental setup PCA is supposed to reduce the coefficient of wavelets. Looking at the performance of these three models, it was concluded that WPLSSVM produces high accuracy 99% and error rate 1%. This level of accuracy and error rate is highly desirable. The idea of this arrangement, i.e. WPLSSVM

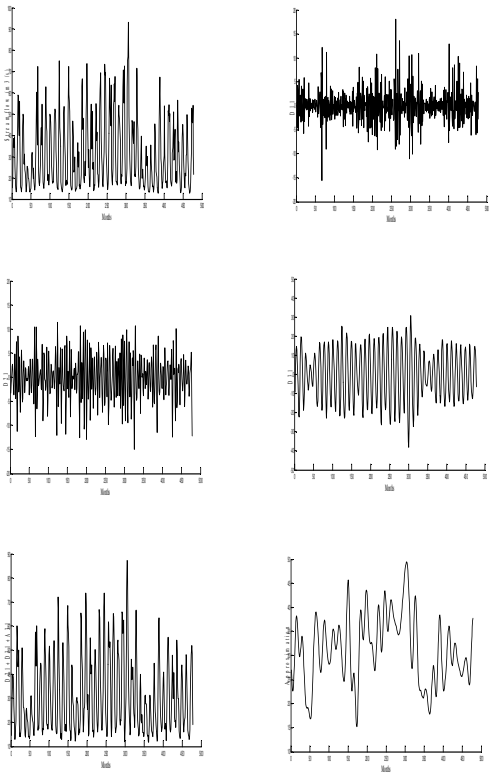

Fig. 2 Decomposition wavelet sub-series components (Ds) of streamflow data of Jhelum Station.

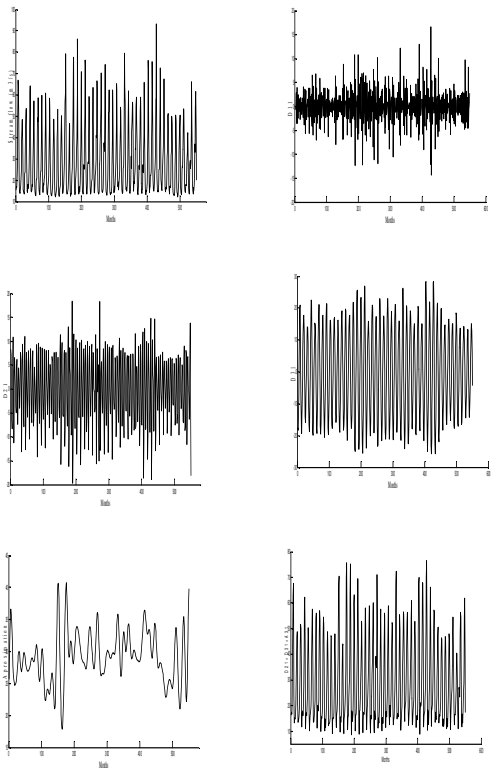

Fig. 3 Decomposition wavelet sub-series components (Ds) of streamflow data of Chenab Station.

is depicted in Figure 4. According to this figure time series data is collected. To optimize the system this data is decomposed by DWT. The decomposed data produced coefficients of DWT. These coefficients are given to PCA which picks up the principle components and prepares a new set of data. This set of data is finally used to train LSSVM which is meant for estimating future data value.

The idea shown above is simulated in MATLAB program. According to this program the forecasting performances of the PCA-wavelet-LSSVM (PWLSSVM) models are presented in Table 6, 7 in terms of MSE, MAE and R in training and testing periods. From Table 6 for the Jhelum station, the MWPL2 model has the smallest MSE (0.0002) and MAE (0.0071) and the highest R (0.9984) in the training phase. However, for the testing phase, the best MSE (0.0002), MAE (0.0120) and R (0.9937) were obtained for the model input combination MWPL2. From Table 7 for Chenab station, the MWPL2 model has the smallest MSE (0.0002) and MAE (0.0106) and the highest R (0.9963) in the training phase. However, for the testing phase, the best MSE (0.0003) and MAE (0.0137) and R (0.9901) were obtained for the model input combination MWPL2.

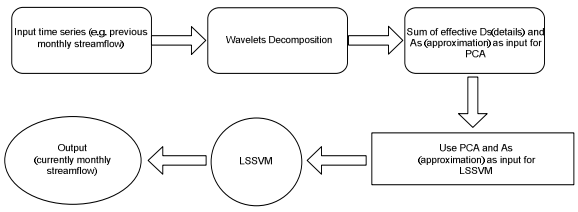

Fig. 4 The structure of the WPLSSVM.

Table 6 Forecasting performance indicates of Wavelet + LSSVM (WLSSVM) for Jhelum River of Pakistan

	Training			Testing		
Model	MSE	MAE	R	MSE	MAE	R
MWPL1	0.0020	0.0320	0.9634	0.0035	0.0415	0.8901
MWPL2	**0.0002**	**0.0071**	**0.9984**	**0.0002**	**0.0120**	**0.9937**
MWPL3	0.0003	0.0132	0.9945	0.0005	0.0175	0.9833
MWPL4	0.0004	0.0161	0.9916	0.0006	0.0196	0.9800
MWPL5	0.0008	0.0214	0.9847	0.0010	0.0248	0.9625
MWPL6	0.0009	0.0229	0.9832	0.0011	0.0259	0.9642

Table 7 Forecasting performance indicates of Wavelet + LSSVM (WLSSVM) for Chenab River of Pakistan

	Training			Testing		
Model	MSE	MAE	R	MSE	MAE	R
MWPL1	0.0008	0.0205	0.9854	0.0010	0.0230	0.9736
MWPL2	**0.0002**	**0.0106**	**0.9963**	**0.0003**	**0.0137**	**0.9901**
MWPL3	0.0004	0.0140	0.9926	0.0007	0.0176	0.9820
MWPL4	0.0005	0.0156	0.9914	0.0007	0.0189	0.9791
MWPL5	0.0009	0.0214	0.9841	0.0005	0.0174	0.9833
MWPL6	0.0008	0.0202	0.9857	0.0008	0.0195	0.9780

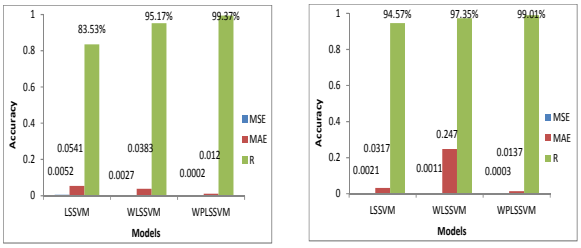

Fig. 5 Accuracy data of Jhelum Station and Chenab Station using MSE, MAE and R.

5 Discussions

The performance of the models investigated during this study is presented in the tables from $1-5$. This performance was measured by using well known performance measuring methods, i.e. MSE, MAE, R. Figure 5 shows graphical representation of the computing models. The graphs compare the accuracy of MSE, MAE and R.

In Table 8 and 9 and Figure 5, shows that WPLSSVM has good performance, and when compared with LSSVM and WLSSVM. The correlation coefficient (R) for Jhelum River and Chenab River data obtained by LSSVM models is 0.8353 and 0.9457 and by WLSSVM models is 0.9515 and 0.9735 respectively, with WPLSSVM model, the R value is increased to 0.9937 and 0.9901. The MSE obtained by LSSVM models is 0.0052 and 0.0021 for both data sets respectively with WPLSSVM model this value is decreased to 0.0002 and 0.0003. Similarly, while the MAE obtained by LSSVM is 0.0541 and 0.0317, the MAE value of WPLSSVM model is decreased to 0.0120 and 0.0137. The WPLSSVM model obtained the best value of MSE and MAE decrease 96% and 86%, respectively. The best of R increases by 19% compared with single LSSVM model for Jhelum data. For Chenab data the best of R increases by 5% and the best value obtained for MSE and MAE decreases 78% and 57%, which shows a substantial improvement in the estimation in comparison to the LSSVM and WLSSVM

Table 8 The performance results LSSVM, WLSSVM and WPLSSVM approach during testing period

Jhelum River			
Model	MSE	MAE	R
LSSVM	0.0052(96%)	0.0541(78%)	0.8353(19%)
WLSSVM	0.0027(93%)	0.0383(69%)	0.9517(4%)
WPLSSVM	**0.0002**	**0.0120**	**0.9937**

Table 9 The performance results LSSVM, WLSSVM and WPLSSVM approach during testing period

Chenab River			
Model	MSE	MAE	R
LSSVM	0.0021(86%)	0.0317(57%)	0.9457(5%)
WLSSVM	0.0011(73%)	0.0247(45%)	0.9735(2%)
WPLSSVM	**0.0003**	**0.0137**	**0.9901**

Figure 6 and Figure 7 shows the hydrograph and scatter plot for the LSSVM, WLSSVM and WPLSSVM models for the testing period. It can be seen that the WPLSSVM forecasts closer to the observed data for both station.

6 Conclusions

We propose WPLSSVM model based on DWT, PCA and LSSVM for forecasting streamflows. The monthly streamflow time series is decomposed at three levels by DWT. Each level carries most of the information and plays a distinct role in original time series. Sub-series are used as inputs to PCA to minimize the dimensionality of high dimensional input vectors 90% the dimensions, which were original, and the dimensions, which we finally decided to use. Finally LSSVM which is the last component of the proposed model produces the required estimated values. The WPLSSVM is trained and tested by different combinations of monthly streamflow data of Chanari station in Jhelum River and Marala station in Chenab in Punjab of Pakistan. Then, LSSVM and WLSSVM models are constructed with new series as

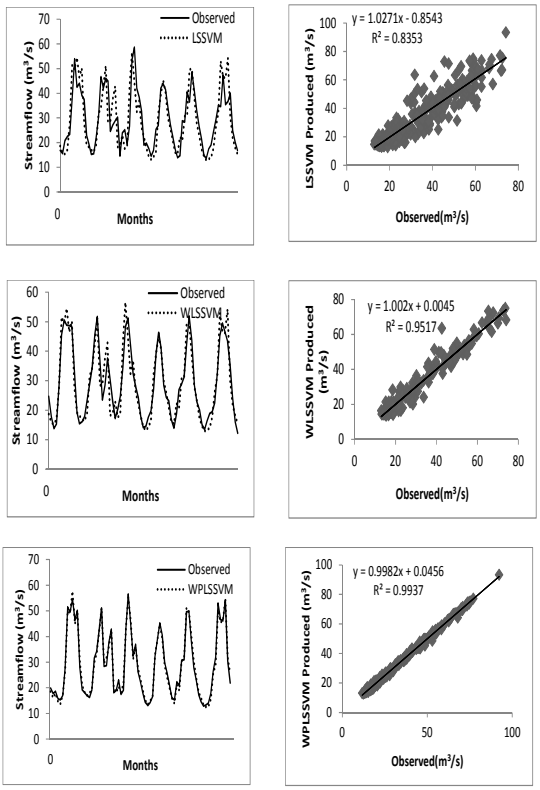

Fig. 6 Predicted and observed streamflow during testing period by LSSVM, WLSSVM and WPLSSVM for Jhelum Station.

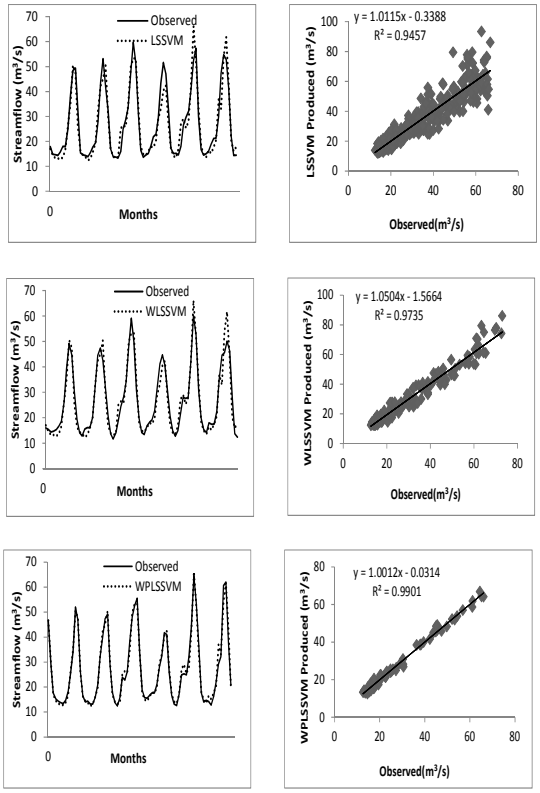

Fig. 7 Predicted and observed streamflow during testing period by LSSVM, WLSSVM and WPLSSVM for Chenab Station.

inputs and original streamflow time series as output. The performance of the proposed WPLSSVM model is then compared to the conventional LSSVM and WLSSVM models for using the same set of data. The proposed WPLSSVM was evaluated in terms of its performance by using commonly used methods, i.e. MSE, MAE and R. The results obtained from these performance measuring methods confirmed that the proposed model has encouraging results, 99.37% and 99.01 respectively. Besides, during the process, another significant observation was made: that the forecasting abilities of the LSSVM and WLSSVM model are found to be improved when the principle component analysis is adopted for the data pre-processing. The decomposed periodic components obtained from the PCA technique are found to be most effective in yielding accurate forecast when used as inputs in the WLSSVM models. The present results not only depict the fact that the present model, WPLSSVM is time efficient, but also that it is almost error free, more reliable and the most stable among the existing models (LSSVM and WLSSVM).

References

[1] S. Narain and A. Jain, "Artificial neuron models for hydrological modeling," in *Neural Networks, 2007. IJCNN 2007. International Joint Conference on*. IEEE, 2007, pp. 1338–1342.

[2] E. Khadangi, H. R. Madvar, and M. M. Ebadzadeh, "Comparison of anfis and rbf models in daily stream flow forecasting," in *Computer, Control and Communication, 2009. IC4 2009. 2nd International Conference on*. IEEE, 2009, pp. 1–6.

[3] B. Cannas, A. Fanni, M. Pintus, and G. Sechi, "Neural network models to forecast hydrological risk," in *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on*, vol. 1. IEEE, 2002, pp. 423–426.

[4] K. Wang and S. Liang, "Global atmospheric downward longwave radiation over land surface under all-sky conditions from 1973 to 2008," *Journal of Geophysical Research: Atmospheres (1984–2012)*, vol. 114, no. D19, 2009.

[5] O. Kisi, "Wavelet regression model for short-term streamflow forecasting," *Journal of hydrology*, vol. 389, no. 3, pp. 344–353, 2010.

[6] R. Maheswaran and R. Khosa, "Comparative study of different wavelets for hydrologic forecasting," *Computers & Geosciences*, vol. 46, pp. 284–295, 2012.
[7] S. Osowski and K. Garanty, "Forecasting of the daily meteorological pollution using wavelets and support vector machine," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 6, pp. 745–755, 2007.
[8] M. Onderka, S. Banzhaf, T. Scheytt, and A. Krein, "Seepage velocities derived from thermal records using wavelet analysis," *Journal of Hydrology*, vol. 479, pp. 64–74, 2013.
[9] P. S. Addison, K. B. Murray, and J. N. Watson, "Wavelet transform analysis of open channel wake flows," *Journal of engineering mechanics*, vol. 127, no. 1, pp. 58–70, 2001.
[10] W. Wang and J. Ding, "Wavelet network model and its application to the prediction of hydrology," *Nature and Science*, vol. 1, no. 1, pp. 67–71, 2003.
[11] B. Krishna and R. YR Satyaji, "Time series modeling of river flow using wavelet neural networks," *Journal of Water Resource and Protection*, vol. 2011, 2011.
[12] D. Labat, R. Ababou, and A. Mangin, "Rainfall–runoff relations for karstic springs. part ii: continuous wavelet and discrete orthogonal multiresolution analyses," *Journal of hydrology*, vol. 238, no. 3, pp. 149–178, 2000.
[13] L. C. Smith, D. L. Turcotte, and B. L. Isacks, "Stream flow characterization and feature detection using a discrete wavelet transform," *Hydrological processes*, vol. 12, no. 2, pp. 233–249, 1998.
[14] B. Krishna and R. YR Satyaji, "Time series modeling of river flow using wavelet neural networks," *Journal of Water Resource and Protection*, vol. 2011, 2011.
[15] E. Swee and S. Elangovan, "Applications of symlets for denoising and load forecasting," in *Higher-Order Statistics, 1999. Proceedings of the IEEE Signal Processing Workshop on*. IEEE, 1999, pp. 165–169.
[16] Y. Zhang, H. Li, A. Hou, and J. Havel, "Artificial neural networks based on principal component analysis input selection for quantification in overlapped capillary electrophoresis peaks," *Chemometrics and Intelligent Laboratory Systems*, vol. 82, no. 1, pp. 165–175, 2006.
[17] X. Wang, S. Chen, D. Lowe, and C. Harris, "Sparse support vector regression based on orthogonal forward selection for the generalised kernel model," *Neurocomputing*, vol. 70, no. 1, pp. 462–474, 2006.
[18] T. Asefa, M. Kemblowski, M. McKee, and A. Khalil, "Multi-time scale stream flow predictions: the support vector machines approach," *Journal of Hydrology*, vol. 318, no. 1, pp. 7–16, 2006.
[19] G.-F. Lin and M.-C. Wu, "A hybrid neural network model for typhoon-rainfall forecasting," *Journal of Hydrology*, vol. 375, no. 3, pp. 450–458, 2009.
[20] P.-S. Yu, S.-T. Chen, and I.-F. Chang, "Support vector regression for real-time flood stage forecasting," *Journal of Hydrology*, vol. 328, no. 3, pp. 704–716, 2006.
[21] Y. B. Dibike, S. Velickov, D. Solomatine, and M. B. Abbott, "Model induction with support vector machines: introduction and applications," *Journal of Computing in Civil Engineering*, vol. 15, no. 3, pp. 208–216, 2001.
[22] A. Elshorbagy, G. Corzo, S. Srinivasulu, and D. Solomatine, "Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology-part 2: Application," *Hydrology and Earth System Sciences*, vol. 14, no. 10, pp. 1943–1961, 2010.
[23] S. Ismail, R. Samsudin, and A. Shabri, "River flow forecasting: a hybrid model of self organizing maps and least square support vector machine," *Hydrology and Earth System Sciences Discussions*, vol. 7, no. 5, pp. 8179–8212, 2010.
[24] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.
[25] B. George, *Time Series Analysis: Forecasting & Control, 3/e*. Pearson Education India, 1994.
[26] N. C. Matalas, "mathematical assessment of symmetric hydrology,," *Water Resources Research*, vol. 3, no. 4, pp. 937–945, 1967.
[27] D. Hanbay, "An expert system based on least square support vector machines for diagnosis of the valvular heart disease," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4232–4238, 2009.
[28] Y.-W. Kang, J. Li, G.-Y. Cao, H.-Y. Tu, J. Li, and J. Yang, "Dynamic temperature modeling of an sofc using least squares support vector machines," *Journal of Power sources*, vol. 179, no. 2, pp. 683–692, 2008.
[29] X. Yunrong and J. Liangzhong, "Water quality prediction using ls-svm and particle swarm optimization," in *Knowledge Discovery and Data Mining, 2009. WKDD 2009. Second International Workshop on*. IEEE, 2009, pp. 900–904.
[30] H. Wang and D. Hu, "Comparison of svm and ls-svm for regression," in *Neural Networks and Brain, 2005. ICNN&B'05. International Conference on*, vol. 1. IEEE, 2005, pp. 279–283.
[31] P.-S. Yu, S.-T. Chen, and I.-F. Chang, "Support vector regression for real-time flood stage forecasting," *Journal of Hydrology*, vol. 328, no. 3, pp. 704–716, 2006.
[32] L. Liu and W. Wang, "Exchange rates forecasting with least squares support vector machine," in *Computer Science and Software Engineering, 2008 International Conference on*, vol. 5. IEEE, 2008, pp. 1017–1019.
[33] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 7, pp. 674–693, 1989.
[34] O. Kisi, "Wavelet regression model for short-term streamflow forecasting," *Journal of hydrology*, vol. 389, no. 3, pp. 344–353, 2010.
[35] Ö. Kişi, "Wavelet regression model as an alternative to neural networks for monthly streamflow forecasting," *Hydrological processes*, vol. 23, no. 25, pp. 3583–3597, 2009.
[36] P. Y. Ma, "A fresh engineering approach for the forecast of financial index volatility and hedging strategies," Ph.D. dissertation, École de technologie supérieure, 2006.
[37] S. M. Pandhiani and A. B. Shabri, "Time series forecasting using wavelet-least squares support vector machines and wavelet regression models for monthly stream flow data," *Open Journal of Statisitcs*, vol. 3, no. 3, pp. 183–194, 2013.

Siraj Muhammed Pandhiani is a final year PhD student at Universiti of Teknologi Malaysia (UTM). He received his MSc and M.Phil from University of Sindh in 2001. His research interests include Time Series Analysis, Least Square Support Machine, Wavelets, Principle Components Analysis and Kernel Principle Components Analysis. He is currently preparing to take his final PhD defense. He has published research articles in reputed international journals of mathematical and engineering sciences. He is referee and editor of mathematical journals.

Ani Bin Shabri received the PhD degree in Statistics at Univeristi Kebangsaan Malaysia of Malaysia. He has published research articles in reputed international journals of mathematical and engineering sciences. His main research interests are: time series forecasting using statistical method and artificial intelligence methods such as neural network, support vector machine, ANFIS etc. and flood frequency analysis.