HYDROLOGICAL PROCESSES
Hydrol. Process. **24**, 917–923 (2010)
Published online 11 December 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/hyp.7535

Potential of support vector regression for prediction of monthly streamflow using endogenous property

Rajib Maity,[1]* Parag P. Bhagwat[1] and Ashish Bhatnagar[2]
[1] *Department of Civil Engineering, Indian Institute of Technology, Kharagpur, Kharagpur, West Bengal, India*
[2] *Department of Civil Engineering, Indian Institute of Technology, Bombay, Powai, India*

Abstract:

In the recent past, a variety of statistical and other modelling approaches have been developed to capture the properties of hydrological time series for their reliable prediction. However, the extent of complexity hinders the applicability of such traditional models in many cases. Kernel-based machine learning approaches have been found to be more popular due to their inherent advantages over traditional modelling techniques including artificial neural networks(ANNs). In this paper, a kernel-based learning approach is investigated for its suitability to capture the monthly variation of streamflow time series. Its performance is compared with that of the traditional approaches. Support vector machines (SVMs) are one such kernel-based algorithm that has given promising results in hydrology and associated areas. In this paper, the application of SVMs to regression problems, known as support vector regression (SVR), is presented to predict the monthly streamflow of the Mahanadi River in the state of Orissa, India. The results obtained are compared against the results derived from the traditional Box–Jenkins approach. While the correlation coefficient between the observed and predicted streamflows was found to be 0·77 in case of SVR, the same for different auto-regressive integrated moving average (ARIMA) models ranges between 0·67 and 0·69. The superiority of SVR as compared to traditional Box-Jenkins approach is also explained through the feature space representation. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS Kernel-based learning approach; least square support vector machine (LS-SVM); support vector regression (SVR); streamflow prediction; Mahanadi River

Received 17 April 2009; Accepted 12 October 2009

INTRODUCTION

Natural complexity of hydrologic variables drives the application of statistical approaches to analyse their statistical behavior. The need to improve such methods to capture the inherent complexity of hydrologic variables, such as streamflow, rainfall, etc. has been recognized for centuries. It is acknowledged that a reliable prediction of streamflow is essential to effectively manage the available water resources. It helps in reservoir operation, flood control and warning, planning reservoir capacity, and economical usage of water such as in agriculture, industry, power generation, navigation, etc. However, depending on the inherent complexity, traditional statistical methods, such as transfer function model, Box–Jenkins approach, etc., were found to be inadequate. As a consequence, many new methodologies have been introduced to understand the variations of hydrological variables and to predict their future time steps.

In the last decade, machine learning techniques such as artificial neural networks (ANNs), fuzzy logic, genetic programming, etc. have been widely used in the modelling and prediction of hydrological variables. Among these techniques, support vector machines (SVMs) have gained popularity in many traditionally ANN-dominated fields. SVMs are learning algorithms that use a hypothesis space comprising linear functions in a higher dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory (Cristianini and Shawe-Taylor, 2000).

* Correspondence to: Rajib Maity, Department of Civil Engineering, Indian Institute of Technology, Kharagpur, Kharagpur—721302, West Bengal, India. E-mail: rajib@civil.iitkgp.ernet.in; rajibmaity@gmail.com

SUPPORT VECTOR MACHINES

SVMs in their present form were first developed for classification problems in the early 1990s by Vapnik and others at AT&T Bell Laboratory (Boser *et al*., 1992; Guyon *et al*., 1993). SVM was introduced for regression in 1995, and the first applications were reported in the late 1990s (Drucker *et al*., 1997, Vapnik, 1998, 2000). Since then, applications of SVMs have seen a rapid rise in fields as varied as text classification to pattern recognition to forecasting in hydrology and other areas.

Though the applications of SVMs in hydrology have been limited in terms of absolute numbers, the results have been encouraging. (Dibike *et al*. (2001) applied SVM in remotely sensed image classification and regression (rainfall/runoff modelling) problems. (Liong and Sivapragasam (2002) also reported a superior SVM performance compared to ANN in forecasting flood stage. (Bray and Han (2004) used SVMs to identify a suitable model structure and its parameters for rainfall runoff modelling. The model was compared with a transfer function model and the study outlined a promising area of research for further application of SVMs in unexplored areas. (She and Basketfield (2005) also reported superior results in forecasting spring and fall season streamflows in the Pacific Northwest region of the United States using SVM. In 2006, to forecast seasonal and hourly flows, new data-driven SVM-based models were presented. Empirical results obtained from these models showed promising performance in solving site-specific, real-time water resources management problems. In addition, seasonal volume predictions were improved using SVMs (Asefa *et al.*, 2005). (Khadam and Kaluarachchi (2004) discussed the impact of accuracy and reliability of hydrological data on model calibration. This, coupled with application of SVMs, was used to identify faulty model calibration which would have been undetected otherwise. (Qin *et al.* (2005) used least squares support vector machines (LS-SVMs), a non-linear kernel-based machine, to demonstrate the excellent generalization property of SVMs and its potential for further applications in area of general hydrology. Applicability of SVMs was also demonstrated in downscaling general circulation models (GCMs), which are among the most advanced tools for estimating future climate change scenarios. The results presented SVMs as a compelling alternative to traditional ANN to conduct climate impact studies (Tripathi *et al.*, 2006). Anandhi *et al.* (2008) downscaled monthly precipitation to basin scale using SVMs and reported the results to be encouraging in their accuracy while showing large promise for further applications.

Overall, the comparison between SVM and ANN has shown the superior performance of SVM in regression. In this paper, the application of support vector regression (SVR) to forecast the streamflow of the Mahanadi River is presented. The results are compared with the traditional Box–Jenkins approach and the potential of SVR in monthly streamflow prediction is demonstrated. Before describing the methodology, a brief description of feature space and kernel functions is presented in the following section.

METHODOLOGY

Supervised machine learning

SVMs are a kind of supervised machine learning technique which belongs to a family of generalized linear classifier. The formulation embodies the structural risk minimization (SRM) principle, as opposed to the empirical risk minimization (ERM) approach commonly employed within statistical learning methods. SRM minimizes an upper bound on the generalization error, as opposed to ERM which minimizes the error on the training data. It is this difference that equips SVMs with a greater potential to generalize. Moreover, the solutions offered by traditional neural network models may tend to fall into a local optimal solution, whereas a global optimum solution is guaranteed for SVM. SVMs can be applied to both classification and regression problems.

Radial basis function as kernel

A brief and basic introduction to karnel function and feature space is provided in Appendix. The flexibility of the SVM is provided by the use of kernel functions that implicitly map the data to a higher dimensional feature space. A linear solution in the higher dimensional feature space corresponds to a non-linear solution in the original, lower dimensional input space. This makes SVM a feasible choice for solving a variety of problems in hydrology, which are non-linear in nature. There are methods available that use non-linear kernels in their approach towards regression problems while applying SVMs. One such approach involves using the radial basis function (RBF) and is called LS-SVMs. The main advantage of LS-SVM is that it is computationally more efficient than the standard SVM method, since the training of LS-SVM requires only the solution of a set of linear equations instead of the long and computationally demanding quadratic programming problem involved in the standard SVM (Suykens and Vandewalle, 1999). In comparison with some other feasible kernel functions, the RBF is a more compact, supported kernel and able to shorten the computational training process and improve the generalization performance of LS-SVM, a feature of great importance in designing a model. (Dibike *et al.* (2001) applied different kernels in SVR to rainfall–runoff modelling and demonstrated that the RBF outperforms other kernel functions. (Han and Cluckie (2004) indicated that the centralized feature of the RBF enables it to effectively model the regression process. Also, many works on the use of SVR in hydrological modelling and forecasting have demonstrated the favourable performance of the RBF (Liong and Sivapragasam, 2002; Choy and Chan, 2003; Yu and Liong, 2007). Therefore, the RBF, which has a parameter σ, is adopted in this study and its mathematical form is presented later.

Support vector regression (SVR)

In SVR, $\{x_i, \ y_i\}_{i=1}^N$ is considered as a training set, in which $\mathbf{x}_i \in \Re^p$ represents a p-dimensional input vector and $y_i \in \Re$ is a scalar measured output, which represents the system output. The goal is to construct a function $y = f(\mathbf{x})$ which represents the dependence of the output y_i on the input $\mathbf{x}_i$. The form of this function is

$$y = w^{\mathrm{T}}\phi(x) + b \tag{1}$$

where $\mathbf{w}$ is known as the weight vector and b the bias.

This regression model can be constructed using a non-linear mapping function $\boldsymbol{\phi}(\bullet)$. By mapping the original input data onto a high-dimensional space, the non-linear separable problem becomes linearly separable in space. The function $\boldsymbol{\phi}$ $(\bullet)$: $\Re^p \rightarrow \Re^h$ is a mostly non-linear function which maps the data into a higher, possibly infinite, dimensional feature space. The main difference from the standard SVM is that LS-SVM involves equality constraints instead of inequality constraints, and works with a least squares cost function. The optimization problem and the equality constraints are defined by the following equations:

$$\min J(w, e) = \frac{1}{2}w^{\mathrm{T}}w + \gamma\frac{1}{2}\sum_{i=1}^{N} e_i^2 \tag{2}$$

subject to

$$y_i = w^{\mathrm{T}}\boldsymbol{\phi}(x_i) + b + e_i,\ \ i = 1, \ldots, N \tag{3}$$

where e_i is the random error and $\boldsymbol{\gamma} \in \Re^+$ is a regularization parameter in optimizing the trade-off between minimizing the training errors and minimizing the model's complexity. The objective is now to find the optimal parameters that minimize the prediction error of the regression model. The optimal model will be chosen by minimizing the cost function where the errors e_i are minimized. This formulation corresponds to the regression in the feature space and, since the dimension of the feature space is high, possibly infinite, this problem is difficult to solve. Therefore, to solve this optimization problem, the following Lagrange function is given:

$$L(w, b, e; \alpha) = J(w, e) - \sum_{i=1}^{N} \alpha_i\{w^{\mathrm{T}}\boldsymbol{\phi}(x_i) + b + e_i - y_i\} \tag{4}$$

The solution of Equation (4) can be obtained by partially differentiating with respect to $\mathbf{w}$, b, e_i and α_i, i.e.

$$\frac{\partial L}{\partial w} = 0 \longrightarrow w = \sum_{i=1}^{N} \boldsymbol{\alpha}_i\boldsymbol{\phi}(x_i) \tag{5}$$

$$\frac{\partial L}{\partial b} = 0 \longrightarrow b = \sum_{i=1}^{N} \boldsymbol{\alpha}_i = 0 \tag{6}$$

$$\frac{\partial L}{\partial e_i} = 0 \longrightarrow \boldsymbol{\alpha}_i = \boldsymbol{\gamma} \quad e_i,\ \ i = 1, \ldots, N \tag{7}$$

$$\frac{\partial L}{\partial x_i} = 0 \longrightarrow w^{\mathrm{T}}\boldsymbol{\phi}(x_i) + b + e_i - y_i = 0, \quad i = 1, \ldots, N \tag{8}$$

Finally, the estimated values of b and $\boldsymbol{\alpha}_i$, i.e. $\hat{b}$ and $\hat{\boldsymbol{\alpha}}_i$, can be obtained by solving the linear system and the resulting LS-SVM model can be expressed as

$$y = f(x) = \sum_{i=1}^{N} \hat{\boldsymbol{\alpha}}_i K(x, x_i) + \hat{b} \tag{9}$$

where $K(\mathbf{x},\mathbf{x}_i)$ is a kernel function. Here, the non-linear RBF kernel is defined as:

$$K(x, x_i) = \exp\left(-\frac{1}{\boldsymbol{\sigma}^2}||x - x_i||^2\right) \tag{10}$$

where $\boldsymbol{\sigma}$ is the kernel function parameter of the RBF kernel.

The regularization parameter $\boldsymbol{\gamma}$ is also necessary in LS-SVM model and determines the trade-off between the fitting error minimization and smoothness of the estimated function. It is not known beforehand which $\boldsymbol{\gamma}$ and σ are the best for a particular application problem to achieve the maximum performance with LS-SVM models. Thus, the regularization parameter $\boldsymbol{\gamma}$ and the value of σ from the kernel function have to be tuned during model calibration. In this work, a grid-search technique is used for tuning these two parameters, using cross-validation on the training set to find out the optimal parameter values. The LS-SVM model thus obtained is used to estimate the desired output which is finally used to predict the monthly streamflow.

PERFORMANCE OF SVR FOR STREAMFLOW PREDICTION

Study area and datasets

The Mahanadi River, which encloses a drainage area of 132 100 square km, is located in the eastern part of India. The Mahanadi rises in the highlands of Chhattisgarh and flows through Orissa to reach the Bay of Bengal slowly for 900 km. It is one of the longest rivers in the country and drains a substantial part of peninsular India. Rainfall comes predominantly from the summer monsoon (June through September). The average annual rainfall in the basin is 1463 mm. Figure 1 shows the locations of the Mahanadi River Basin and the Basantpur station, from where the streamflow data was collected for this study.

Streamflow data from the Basantpur station (Station Code EM000R2), operated by the Water Resources Agency, were obtained from Central Water Commission, Govt. of India. Among these records, data for 23 years (June 1972 to May 1995) were used for calibration, and data for 9 years (June 1995 to May 2004) were used to test the model performance.

Data normalization and parameter calibration

The observed streamflow data is normalized to prevent the model from being dominated by the variables with large values, as is commonly used in data-driven models. The performance of LS-SVM with normalized input data in the range from 0 to 1 has shown to outperform the same with unscaled input data (Bray and Han, 2004). Therefore, the data is normalized and finally the model outputs are back-transformed to their original scale. The normalization (also back-transformed) is carried out using

$$y_{i,j} = \frac{x_{i,j} - \boldsymbol{\mu}_j}{\boldsymbol{\sigma}_j} \tag{11}$$

where $y_{i,j}$ is the normalized value for i-th year and j-th month; $x_{i,j}$ is the observed value for i-th year and j-th month; $\boldsymbol{\mu}_j$ and $\boldsymbol{\sigma}_j$ are the mean and standard deviation, respectively, for j-th month.

The SVR model used herein has two parameters ($\boldsymbol{\gamma}$, σ) to be determined. These parameters are interdependent, and their (near) optimal values are often obtained by a trial-and-error method. The analyses and calculations of SVR herein are performed using LS-SVM and based

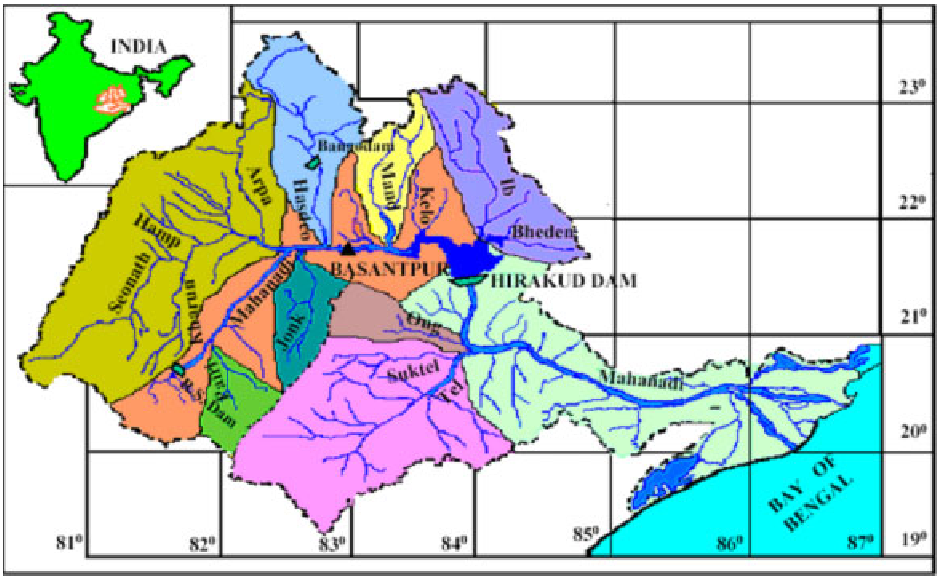

Figure 1. Location map of catchment and sub-basins of the Mahanadi River (Source: Central Water Commission, Bhubaneswar)

on the above-derived parameters, and the model is used to perform stage discharge forecasting. The model used four different combinations of datasets to calibrate and develop the model. One of these four datasets (containing one input vector, with a lag of 1) was used as a sample set to estimate the trade-off between γ and σ. The resulting error parameters (correlation coefficient (CC), root mean square and Nash–Sutcliffe efficiency (NSE) coefficient) between the predicted and observed values were compared and the best combination was chosen for further testing and validation. The results are shown in Table I. From top to bottom in each cell of Table I, the statistics are (1) CC between the observed and predicted value of streamflow; (2) the root mean square error (RMSE) between the observed and predicted value of streamflow; and (3) the NSE coefficient. A higher value of the first statistic indicates a better model performance. Performance statistics in the cell corresponding to the best combinations are shown in boldface, which are found to be 2 and 1·5 for γ and σ, respectively.

Table I. Performance statistics for different combinations of kernel parameters γ and σ[a]

σ	γ					
	0·1	2	4	6	8	10
0.5	0·7264 0·9928 0·301	0·7091 0·8847 0·4449	0·7022 0·884 0·4458	0·6993 0·8841 0·4456	0·6979 0·8841 0·4456	0·6971 0·884 0·4458
1	0·7361 0·9776 0·3222	0·7207 0·8835 0·4464	0·7141 0·8807 0·4499	0·7101 0·8804 0·4503	0·7074 0·8806 0·45	0·7054 0·881 0·4495
1.5	0·7391 0·9704 0·3322	**0·7231** **0·889** **0·4395**	0·7206 0·8847 0·4449	0·7191 0·8825 0·4477	0·7178 0·8812 0·4493	0·7166 0·8804 0·4503
2	0·741 0·966 0·3382	0·7217 0·8924 0·4352	0·72 0·8895 0·4389	0·7194 0·8878 0·441	0·7191 0·8866 0·4426	0·7189 0·8855 0·4439
2.5	0·7426 0·963 0·3424	0·7204 0·894 0·4331	0·7188 0·8919 0·4358	0·7183 0·8908 0·4372	0·7182 0·8899 0·4383	0·7181 0·8892 0·4392
3	0·7438 0·9608 0·3453	0·7193 0·895 0·4319	0·7175 0·8934 0·434	0·7172 0·8925 0·4351	0·7171 0·8918 0·4359	0·7172 0·8913 0·4366

[a] Each cell shows CC, RMSE and NSE coefficient between the normalized values of observed and predicted streamflow during model development period for one input.

Results and Discussion

Having decided on the values of parameters, four different cases of input variables were tested to decide the optimum number of inputs for the best possible results. Two hundred and seventy-six training datasets containing one, two, three and four input vectors (previous time steps) as a single coordinate were used to separately train each model. Validation of each model was carried out with the remaining 102 validation datasets. The results were compared and some interesting observations were made. Model performances are shown in Table II in terms of different error parameters, i.e., RMSE, CC and NSE coefficient, for different sets of data, containing multiple inputs. The statistics are obtained between the normalized observed and predicted streamflow values. From Table II, it is observed that the best performance is obtained for the 'two inputs' case. Beyond this, the performance decreases as the number of inputs increases. Thus, the optimum number of inputs is two for the given streamflow series. A plot between the observed and predicted streamflow using two inputs is shown in Figure 2 for the model testing period. The CC between the observed and predicted values is found to be 0·77. However, it is observed from the Figure 2 that for some years the peak values during monsoon period are not predicted properly (e.g. 2000, 2003). There might be some other reasons apart from the endogenous property, which is extracted from the time series itself. To investigate the relative skill of

Table II. RMSE, correlation coefficient and Nash coefficient for Normalized model outputs

No. of inputs	Performance statistics		
	RMSE	Correlation coefficient	NSE coefficient
1	0·8890	0·7230	0·4395
2	0·8716	0·7475	0·4612
3	1·0193	0·5793	0·2632
4	1·0745	0·5232	0·1812

SVR compared to other models that use the endogenous property of the time series, the model performance is compared with that of the Box–Jenkins approach (Box *et al*., 1994).

Different types of Box–Jenkins models are known, such as auto-regressive (AR) and auto-regressive moving average (ARMA) or ARIMA. The model order and parameter values depend on the statistical properties of time series. For example, an AR model is a linear regression of the current value of the series against one or more prior values of the series. Thus, the AR model for univariate time series can be expressed as

$$x_t = \delta + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + A_t \quad (12)$$

where $\boldsymbol{\delta} = \left(1 - \sum_{i=1}^{p} \boldsymbol{\phi}_i\right) \boldsymbol{\mu}$, with $\boldsymbol{\mu}$ denoting the process mean, x_t is the observed value and A_t is white noise at time t. The value of p is called the order of the AR model. Similarly, the order of the moving average is denoted as q. The term 'integrated' deals with the stationary property of the time series and its order is denoted as d. If the time series is stationary, $d = 0$. Details of the ARIMA approach can be found elsewhere (Makridakis *et al*., 1998).

Autocorrelogram and partial autocorrelogram for the streamflow time series were obtained, and based on these correlograms, ARIMA(5,0,1) and ARIMA(3,0,2) were tentatively chosen to be competent. The CC between the normalized observed and predicted streamflow were found to be 0·67 to 0·69 for ARIMA (5,0,1) and ARIMA (3,0,2) respectively, during the testing period. It is observed that ARIMA (3,0,2) shows better results than the ARIMA (5,0,1). However, the performance of SVR is even better than both the ARIMA models. The prediction performance of ARIMA (3,0,2) is shown in Figure 2. A plot of the predicted streamflow is shown along with the SVR results for the model testing period. It is observed that the peaks are not captured well in the case of ARIMA (3,0,2) as compared to SVR even though the CCs are close to each other (0·77 for SVR and 0·75 for ARIMA). However, it can also be observed that there is a phase shift between observed and predicted peak streamflow values. This is due to the fact that only the previous step streamflow values are used as the input (endogenous inputs). Consideration of other inputs (exogenous input) which have more immediate effect on streamflow, such as rainfall, may improve the performance further. The reason behind the superior results from SVR can be explained by the basic concept of mapping from the input space to the feature space, which is developed during model calibration. For 'two inputs' case (which is also the best case for the time series analysed), the input and output space can be shown in a three-dimensional figure (Figure 3). A 'smoothened' and 'fitted through observations' surface is also shown in this figure. The 'smoothness' and 'fitting through observations' are 'conflicting-with-each-other' properties and controlled by the parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\sigma}$ as explained before. These parameters are obtained for the specific time series being analysed as was done before. The important aspect is its basic difference from ARIMA approach, in which the fitted surface is linear in nature

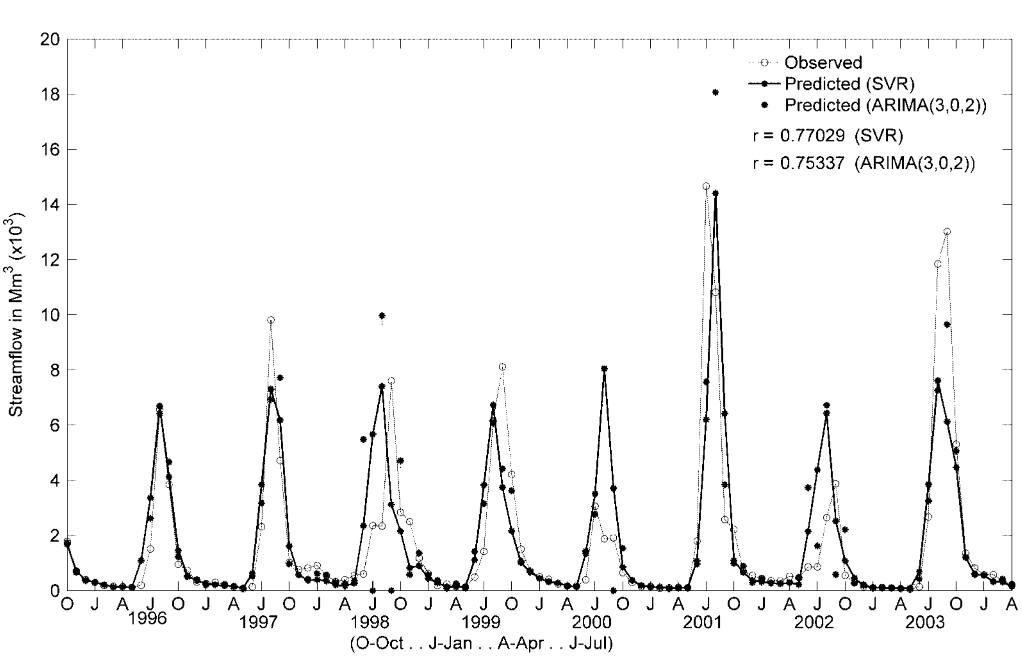

Figure 2. Comparison between observed and predicted streamflow for SVR and ARIMA (during testing period)

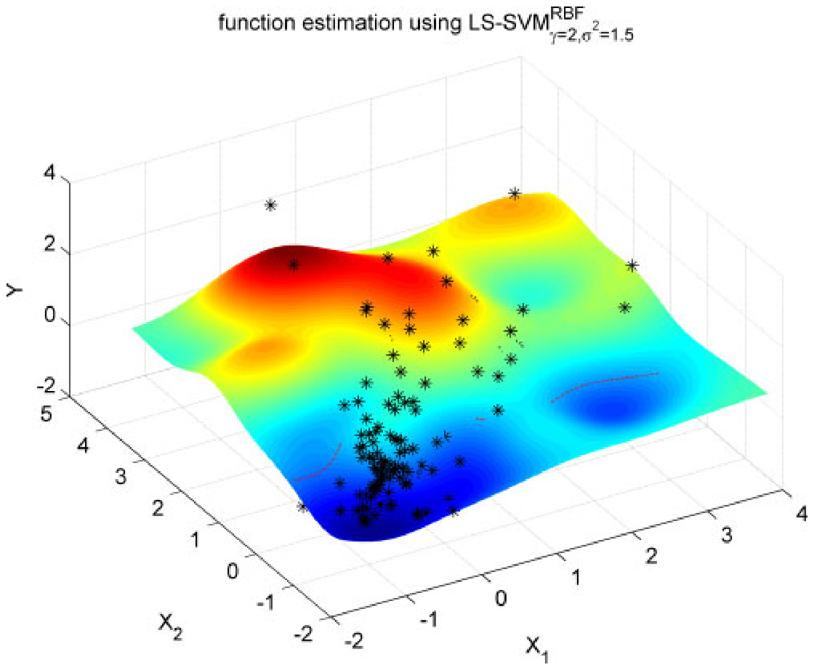

Figure 3. Input–output space for the 'two inputs' case developed by LS-SVM

(a linear surface for 'two inputs' case). It might be possible to get a non-linear surface in case of non-linear ARIMA. However, even this non-linear surface has a regular pattern presumed by the modeller. Observation to the fitted surface in case of SVR indicates a more complex nature of non-linearity, through different peaks and troughs. This yields the better suitability for the time series which is possibly having non-linearity. Thus, the prediction performances for a linear time series by both SVR and ARIMA approaches might be equal; however, it will be better in case of SVR if non-linearity exists in the time series, which is captured through the feature space. With the increase in the number of inputs (if found suitable for other cases), the dimension of the feature space increases, which is not possible to visualize. However, keeping the basic concept of feature space the same, the characteristics of the time series is captured and used for its prediction.

CONCLUSIONS

Prediction of monthly streamflow using SVR was of interest in this work. The study was carried out using data collected from the Basantpur station on the Mahanadi River in India. LS-SVM was used for parameter calibration and model development. The problem-specific subjective parameters γ and σ were obtained based upon the model performance during the calibration period. It was found that $\gamma = 2$ and $\sigma = 1{\cdot}5$ provided the best performance for the model. While considering the model parsimony, it was found that the 'two input' case yielded the best performance. The model was tested for the period October 1995 to April 2004 and different error statistics (RMSE, CC and NSE coefficient) were obtained. These statistics confirm the good correspondence between the observed and predicted streamflow values even though for some years the peak values are not well captured. However, a comparison between the Box–Jenkins approach and the SVR approach indicates the superiority of the latter. The reason behind the strong potential of SVR lies in the non-linear nature of the feature space captured and utilized by SVR. Thus, the approach can be applied to other hydrological time series with a non-linear nature to achieve a better prediction performance than would be obtained from linear modelling approaches, such as the Box–Jenkins approach. However, there are a few other issues involved in SVR approach, such as the use of grid-search technique to fine-tune the kernel variables and the use of exogenous inputs, which may be of further research interest.

ACKNOWLEDGEMENT

This work is supported by the SERC division, Department of Science and Technology, Government of India, through project SR/FTP/ETA-26/07.

APPENDIX

Feature space and kernel functions

The basic working principle of the SVMs is to map the data in some other dot product space (called the feature space) via a non-linear mapping and perform the linear algorithm in the feature space. As the evaluation of a dot product is involved, the feature space is high dimensional and thus requires high computational resources and time. In some cases, however, a simple kernel can be formulated and its efficiency evaluated.

Real-world complex problems require a more expressive hypothesis space than linear functions, as the available linear learning machines are limited by their computational powers. In other words, the target data cannot be expressed as a simple linear combination of the given attributes. One important property of linear learning machines is that they can be expressed in a dual representation. This means that the hypothesis can be expressed as a linear combination of the training points, so that the decision rule can be evaluated using just the inner products between the test point and the training points. If a way of computing the inner product in feature space directly as a function to the original input points is available, it becomes possible to build a non-linear learning machine and it is known as direct computation method of kernel function, denote it by K. In other words, a kernel function can be defined as a function K, such that for all $x, z \in X$,

$$K(x, z) = \langle \boldsymbol{\phi}(x) \bullet \boldsymbol{\phi}(z) \rangle \tag{A1}$$

There are two basic characteristic of a kernel function, (1) the function must be symmetric, i.e.

$$K(x, z) = \langle \boldsymbol{\phi}(x) \bullet \boldsymbol{\phi}(z) \rangle = \langle \boldsymbol{\phi}(z) \bullet \boldsymbol{\phi}(x) \rangle = K(z, x)) \tag{A2}$$

and (2) it must satisfy Cauchy–Schwartz inequality (Cristianini and Shawe-Taylor, 2000)

$$K(x, z)^2 = \langle \boldsymbol{\phi}(x) \bullet \langle \boldsymbol{\phi}(z) \rangle^2 \leq ||\boldsymbol{\phi}(x)||^2 ||\boldsymbol{\phi}(z)||^2 \tag{A3}$$

Above equations, though necessary, however, are not sufficient to guarantee a feature space as described by the kernel function. However, once characterized, kernel representations offer an alternative solution by projecting the data into a high-dimensional feature space to increase the computational power of the linear learning machines. Out of the various kernel functions available to develop a model, non-linear kernel functions are more efficient in analysing complex relations between real-world problems and are thus used in present work. This study makes use of LS-SVM, a kind of SVM learning approach which consists of RBF kernel, to develop a model and forecast the streamflow.

REFERENCES

Asefa T, Kemblowski M, Lall U, Urroz G. 2005. Support vector machines for nonlinear state space reconstruction: Application to the Great Salt Lake time series. *Water Resources Research* **41**: W12422. DOI:10.1029/2004WR003785.

Anandhi A, Srinivas VV, Nanjundiah RS, Nagesh Kumar D. 2008. Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine. *International Journal of Climatology* **28**(3): 401–420. DOI: 10.1002/joc.1529.

Boser BE, Guyon I, Vapnik V. 1992. A training algorithm for optimal margin classifiers. *Proceedings Fifth annual Workshop on Computational Learning Theory*, Pittsburgh, 144–152. DOI: 10.1145/130385.130401.

Box GEP, Jenkins GM, Reinsel GC. 1994. *Time Series Analysis, Forecasting and Control*. Pearson Education: India (Indian reprint).

Bray M, Han D. 2004. Identification of support vector machines for runoff modelling. *Journal of Hydroinformatics* **6**: 265–280.

Cristianini N, Shawe-Taylor J. 2000. *An Introduction to Support Vector Machines: and Other Kernel-based Learning Methods*. Cambridge University Press: Cambridge.

Choy KY, Chan CW. 2003. Modelling of river discharges and rainfall using radial basis function networks based on support vector regression. *International Journal of Systems Science* **34**(14–15): 763–773.

Dibike YB, Velickov S, Slomatine D, Abbott MB. 2001. Model induction with support vector machines: introduction and applications. *Journal of Computing in Civil Engineering* **15**(3): 208–216. DOI: 10.1061/(ASCE)0887–3801(2001)15 : 3(208).

Drucker HD, Burges CJC, Kaufman L, Smola A, Vapnik V. 1997. Support vector regression machines. In *Advances in Neural Information Processing Systems*, vol. 9, Mozer MC, Jordan MI, Petsche T (eds). Morgan Kaufmann: San Mateo; 155–161.

Guyon I, Boser B, Vapnik V. 1993. Automatic capacity tuning of very large VC-dimension classifiers. In *Advances in Neural Information Processing Systems*, vol. 5, José Hanson S, Cowan JD, Lee Giles C (eds). Morgan Kaufmann: San Mateo; 147–155.

Han D, Cluckie I. 2004. Support vector machines identification for runoff modeling. In *Proceedings of the Sixth International Conference on Hydroinformatics*, Liong SY, Phoon KK, Babovic V (eds). World Scientific Publishing Co.: Singapore.

Khadam IM, Kaluarachchi JJ. 2004. Use of soft information to describe the relative uncertainty of calibration data in hydrologic models. *Water Resources Research* **40**: W11505. DOI: 10.1029/2003WR002939.

Liong S-Y, Sivapragasam C. 2002. Flood stage forecasting with support vector machines. *Journal of the American Water Resources Association* **38**(1): 173–196.

Makridakis S, Wheelwright SC, Hyndman RJ. 1998. *Forecasting Methods and Applications*, 3rd edn, John Wiley & Sons: New York.

Qin Z, Yu Q, Li J, Wu Zhi-yi, Hu Bing-min. 2005. Application of least squares vector machines in modelling water vapor and carbon dioxide fluxes over a cropland. *Journal of Zhejiang University Science B* **6**(6): 491–495. DOI: 10.1631/jzus.2005.B0491.

She N, Basketfield D. 2005. Long range forecast of streamflow using support vector machine. In *Proceedings of the World Water and Environment Resources Congress* ASCE, Raymond Walton (ed). May 15–19, 2005, Anchorage: Alaska, USA. DOI:10.1061/40792(173)481.

Suykens JAK, Vandewalle J. 1999. Least squares support vector machine classifiers. *Neural Processing Letters* **9**: 293–300.

Tripathi S, Srinivas VV, Nanjundian RS. 2006. Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of Hydrology* **330**(3–4): 621–640. DOI: 10.1016/j.jhydrol.2006.04.030.

Vapnik VN. 1998. *Statistical Learning Theory*. John Wiley & Sons: New York.

Vapnik VN. 2000. *The Nature of Statistical Learning Theory*, 2nd edn, Springer: New York.

Yu X, Liong S-Y. 2007. Forecasting of hydrologic time series with ridge regression in feature space. *Journal of Hydrology* **332**(3–4): 290–302. DOI: 10.1016/j.jhydrol.2006.07.003.

