ISH Journal of Hydraulic Engineering

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tish20

HYDROMETEROLOGICAL MODELING APPROACHES USING SUPPORT VECTOR REGRESSION (SVR) AND GENETIC PROGRAMMING (GP)

Rajib Maity M.ISH a, S. S. Kashid b & Ashish Bhatnagar M.ISH c

a Department of Civil Engineering, Indian Institute of Technology Bombay, Powai, Mumbai, 400076
b Department of Civil Engineering, Indian Institute of Technology Bombay, Powai, Mumbai, 400076
c Department of Civil Engineering, Indian Institute of Technology Bombay, Powai, Mumbai, 400076

Version of record first published: 07 Jun 2012

To cite this article: Rajib Maity M.ISH, S. S. Kashid & Ashish Bhatnagar M.ISH (2009): HYDROMETEROLOGICAL MODELING APPROACHES USING SUPPORT VECTOR REGRESSION (SVR) AND GENETIC PROGRAMMING (GP), ISH Journal of Hydraulic Engineering, 15:sup1, 244-257

To link to this article: http://dx.doi.org/10.1080/09715010.2009.10514978

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan,
sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
HYDROMETEROLOGICAL MODELING APPROACHES USING SUPPORT VECTOR REGRESSION (SVR) AND GENETIC PROGRAMMING (GP)

by

Rajib Maity\(^1\) MISH, S. S. Kashid\(^2\) and Ashish Bhatnagar\(^3\) MISH

ABSTRACT

Hydrometeorology deals with the problems involving hydrologic cycle, water budget, and rainfall statistics of storms and streamflows. These systems are quite complex and require complete understanding in order to analyze and manage water resources of a region. The rainfall and streamflow over the region are significantly influenced by large-scale coupled atmospheric-oceanic circulations patterns, such as El Niño-Southern Oscillation (ENSO), Equatorial Indian Ocean Oscillation (EQUINOO) etc. The local meteorological parameters like Outgoing Longwave Radiation (OLR), temperature, pressure, etc., also affect rainfall and basin scale streamflow. To model the complex relationship between these variables and basin scale stream-flow, artificial intelligence tools – Support Vector Regression (SVR) and Genetic Programming (GP) have been employed. The application of Support Vector Machine (SVM) to general regression problem to forecast future streamflow potentially improves the performance of monthly basin-scale streamflow prediction and gives better results than traditional methodologies such as ARIMA models. For multivariate inputs, GP-derived streamflow forecasting models are used. The GP model captured the performance of weekly basin-scale streamflow prediction and successfully improved upon existing methodologies. The observed and predicted stream flows, using SVR and GP, were found to correspond well with each other with a correlation coefficient of 0.78 and 0.81 respectively, which is reasonably good for such a complex system. This may however be noted that Support Vector Regression (SVR) used only single variable input and thus holds still greater promise to arrive at even better results.

1. Asst. Professor in Civil Engg., Department of Civil Engineering, Indian Institute of Technology Bombay, Powai, Mumbai-400076
2. Research Scholar, Department of Civil Engineering, Indian Institute of Technology Bombay, Powai, Mumbai-400076
3. P. G Student, Department of Civil Engineering, Indian Institute of Technology Bombay, Powai, Mumbai-400076
KEYWORDS: El Niño-Southern Oscillation (ENSO); Equatorial Indian Ocean Oscillation (EQUINOO); Mahanadi River; Support Vector Regression; Genetic Programming; Hydroclimatic Teleconnection; Streamflow; forecasting.

INTRODUCTION

Variation of hydrologic variables has significant influences on the socio-economic status of a country or at a region in particular. It is often desirable to have an understanding of the phenomenon governing these variations. A proper modeling of such variation of hydrologic variables using suitable techniques helps to understand the basic process that eventually provides the information to better utilize the water resources of a country. Prediction of streamflow is an important factor for water resources management of any region. It helps in flood control, devising agricultural strategy, reservoir operation, etc. Recently, it has been understood that the temporal structure of a hydrologic time series is significantly influenced by large-scale atmospheric circulations (Jain and Lall, 2001). However, it is scientifically and mathematically challenging to use such signals for the prediction of basin-scale hydrologic variables. In addition, the hydro-meteorological systems are quite complex and very difficult to model, as the mechanism behind the hydrologic cycle is spread over very large area and the physics behind all the atmospheric processes is not completely understood. Hence, the advanced techniques to understand the process are being investigated and also found to be more efficient as compared the traditional modeling approaches based upon statistical concepts. In this study two different approaches are discussed – Support Vector Machines (SVM) and Genetic Programming (GP). Both these methods are described in greater details along with their usefulness in streamflow prediction. The results are also compared with a traditional statistical modeling technique in the following sections.

SUPPORT VECTOR MACHINES

Support Vector Machines are a kind of supervised machine learning technique, which were developed by Vladimir Vapnik and associates (Vapnik, 1995). They belong to a family of generalized linear classifier and are considered a special case of Tikhonov regularization. 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 which equips SVMs with a greater potential to generalize.

The main advantage of SVMs is the ability to map the data to a higher dimensional 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.

The approach of SVMs is to seek for an optimal separating hyperplane, separating two sets of data points called vectors, possibly in a higher dimension where it is easier to separate them into two sections. The purpose of separating the vectors is to maximize the margin (distance) between two adjacent data points (vectors). An important and unique feature of this approach is that the solution is based only on those data points, which are at the margin. These points are called support vectors. The support vectors corresponding to two-dimensional system can also be extended to a nonlinear case when the problem is transformed into a feature space. In the feature space which can be very high dimensional, the data points can be separated linearly, thus reducing the problem to a simplified form. This flexibility to map the data into higher plane and simplify the problem is provided by the use of kernel functions and that is the unique feature of Support Vector Machines.

The solutions offered by traditional neural network models may tend to fall into a local optimal solution, whereas global optimum solution is guaranteed for SVM. The SVMs can be applied to both classification and regression problem. The traditional artificial neural networks (ANNs) have considerable subjectivity in model architecture, whereas for SVMs the learning algorithm automatically decides the model architecture (number of hidden units). Moreover, traditional ANN models do not give much emphasis on generalization of performance, while SVMs seek to address this issue in a rigorous theoretical setting.

The use of SVMs has seen progressively increasing in the field of hydrology. Asefa et al. (2005) used SVM for forecasting time series with time lags ranging from 2 weeks to months. The reported results were better than the conventional methods. She and Basketfield (2005) forecast streamflow in Pacific NW region using SVM, and showed superior results than other regression methods. Burger et al. (2007) modeled river flow using SVM and reported encouraging results as well. Most importantly Xinying and Liong (2007) concluded, based upon their research, that SVM is one of the most elegant method developed for time series analysis. Their research proved that higher prediction accuracy was obtained in time series analysis by using SVMs and the models were computationally faster and stable.

Support Vector Regression (SVR)

The SVM method can be applied in regression maintaining all the features that characterize the maximal margin algorithm (Vapnik, 1998). A nonlinear function is learned by a linear learning machine in a kernel induced feature space while the capacity of the system is controlled by a parameter that does not depend upon the dimensionality of the space. As in the classification case, the learning algorithm minimizes a convex function and its solution is sparse.
Let us consider the finite training sample pattern \((x_i, y_i)\), where \(x_i \in \mathbb{R}^n\) is a sample value of the input vector \(X\) consisting of \(N\) training patterns (i.e., \(X = [x_1, x_2, \ldots, x_n]\)) and \(y \in \mathbb{R}\) is the corresponding value of the desired model output. A nonlinear transformation function \(\phi(\bullet)\) is defined to map the input space to a higher-dimension feature space \(\mathbb{R}^N\). According to Cover's theorem, a linear function, \(f(\bullet)\), could be formulated in the high dimensional feature space to look for a nonlinear relation between inputs and outputs in the original input space, as shown below.

\[
\hat{y} = f(x) = w_i \phi(x) + b
\]  

In above equation \(\hat{y}\) denotes the actual model output, where the coefficients \(w\) and \(b\) are the adjustable model parameters.

Moreover, the model produced by support vector classification depends only on a subset of the training data, because the cost function for building the model does not care about training points that lie beyond the margin. Analogously, the model produced by SVR only depends on a subset of the training data, because the cost function for building the model ignores any training data that are close (within a threshold \(\varepsilon\)) to the model prediction.

Above mentioned approach can be thus applied to regression problems. The general field of hydrology is one such area which has greatly benefited from the application of support vector regression.

GENETIC PROGRAMMING (GP)

While the support vector regression (SVR), as described above, is a kernel base learning approach, Genetic Programming (GP) is based on evolutionary algorithm. In recent times, GP has been applied successfully to several water resources problems. Drecourt (1999) applied neural networks and genetic programming for rainfall-runoff modeling. Whigham and Crapper (2001) also used genetic programming for rainfall-runoff modeling. Muttill and Liong (2001) used genetic programming for improving runoff forecasting by input variable selection. Drunpob et al. (2005) applied genetic programming for streamflow rate prediction over a semi-arid coastal watershed in U.S.A. Makkeasorn et al. (2008) compared genetic programming and neural network models for short-term streamflow forecasting with global climate change implications.

According to Koza (1992) before applying GP to a problem, the user must perform five major preparatory steps. These five steps involve determining:
(i) The set of terminals,
(ii) The set of primitive functions
(iii) The fitness measure
(iv) The parameters for controlling the run, and
(v) The method for designating a result and the criterion for terminating a run.

The terminals can be viewed as the inputs to the as-yet-undiscovered computer program. The set of terminals (along with the set of functions) is the ingredients from which GP attempts to construct a computer program to solve, or approximately solve, the problem at hand. The choice of input variables is generally based on prior knowledge of casual variables and physical insight into the problem being studied and if the relationship to be modeled is not well understood, and then analytical techniques can be used. The aim of GP is to evolve a function that relates the input information to the output information, which is of the form:

\[ Y^m = f(X^n) \]  

(2)

where \( X^n \) an n-dimensional input is vector and \( Y^m \) is an m-dimensional output vector. The values of output can be hydrological responses such as runoff, streamflow, ordinates of a hydrograph, optimal pumping patterns, hydraulic conductivities, contaminant concentration etc. and the inputs may be selected based on the physical understanding of the respective process.

Background of GP

Any mathematical model can be represented in a simple tree structure. A typical model is shown in the Fig.1. It can be seen from the ‘Tree’ structure that there exists a clean hierarchical structure. The structure is made up of simple functions that can be easily encoded using a high-level language. Tree manipulation routines exist in several high-level languages.

\[ 2\pi + \left[ \frac{(X + 3) - \frac{Y}{s + 1}}{\pi} \right] \]

FIG. 1 SIMPLE ‘TREE’ STRUCTURE TO REPRESENT A MODEL.
Genetic programming breeds a population of computer programs to solve a problem. It iteratively transforms a population of computer programs into a new generation of programs by applying analogs of naturally occurring genetic operations. It genetically breeds a population of computer programs using the principles of Darwinian natural selection and biologically inspired operations. The operations include reproduction, crossover, mutation, and architecture altering operations patterned after gene duplication and gene deletion in nature as shown in the flow chart (Fig. 2). Crossover operates on two programs (a binary operator), and produces two child programs. Two random nodes are selected from within each program and then the resultant ‘sub-trees’ are swapped, generating two new programs. These new programs become part of the next generation of programs to be evaluated. Reproduction is performed by simply copying a selected member from the current generation to the next generation. Mutation becomes an important operator in Genetic Algorithms, which provides diversity to the population. However, mutation is relatively unimportant in the Genetic Programming, because the dynamic sizes and shapes of the individuals in the population already provide diversity, and as stated above, the population should not converge. Thus, mutation can be considered as a variation on the crossover operation.

Basic Genetic Programming Parameters

- **Population size**: The population size parameter sets the number of programs in the population that the GP is to evolve. A larger population can solve more difficult problems. The recommended size of the population can be in a range of 100 to 1000.

- **Mutation Rate**: The ‘Mutation Frequency’ parameter sets the overall probability of mutation of the programs that have been selected as winners in a tournament. The allowable range for the Mutation Frequency parameter is 0% to 100%. Most genetic programming systems use a very low mutation rate.

- **Crossover Rate**: Crossover operates by exchanging sequences of instructions between two tournament winners. The result of that exchange produces two offsprings that are then inserted into the population in place of the losers in the tournament. The ‘Crossover Frequency’ parameter sets the overall probability that crossover will occur between the two winners in a tournament. The allowable range for this parameter is 0% to 100%. Most genetic programming systems use a high crossover rate.

- **Reproduction Rate**: Reproduction just copies a program and places the copy into the population in addition to the original program. The reproduction rate in a run is what is left over after the application of the crossover and mutation operators. Where Reproduction rate = 100 – mutation rate – (crossover rate * [1-mutation rate]).
HYDROMETEOROLOGICAL MODELING APPROACHES USING SUPPORT VECTOR REGRESSION (SVR) AND GENETIC PROGRAMMING (GP)

Start

Create Initial Population

Generation K=0

A genetic loop

Evaluate fitness of each model of population probabilistically and execute one of the three branches below

Model N=0

Select genetic operation probabilistically and execute one of the three branches below

(1-Pc-Pm)

Select one model using Roulette wheel selection

Perform reproduction

Copy into new population

Models N=N+1

Select two models using Roulette wheel selection

Perform crossover

Inherit two offspring into new population

Models N=N+2

Select one model using Roulette wheel selection

Perform mutation

Inherit mutant into new population

Models N=N+1

Models=N?

No

Yes

Create N new models for the next generation

Max. generation?

No

Generation K=K+1

Yes

Return best model and Stop

FIG. 2 FLOWCHART FOR GENETIC PROGRAMMING (HONG AND BHAMIDIMARRI 2002)
Formulation of Problem with Genetic Programming

The genetic operators (crossover, reproduction, etc.) are performed on constructed random model trees. In order to create these individuals, two distinct sets are defined: the terminal set T, and the function set F. The terminals must be compatible and should be able to pass information between each other. In general, a random tree is designed by picking randomly from terminals and functions. One can skew the probabilities to obtain trees of certain sizes, as well as place absolute limits on the depth of trees. Initialization procedure is to be performed as many times as needed for generating population of programs.

Selection of the individuals for crossover and reproduction can be done by adopting the methodology of measuring fitness. The measurement of fitness is highly problem-dependent, which can be done by a process known as ‘scaling’. Scaling standardizes the measurement of fitness of a particular individual with respect to the rest of the population. Based on the fitness value, the selection for survival is done in one of two ways:

- Choose the individuals with the highest fitness for reproduction. ‘Only the strong survive.’
- Assign a probability that a particular individual will be selected for either reproduction or crossover.

Steps in Genetic Programming Run

Steps in Genetic Programming for a single run can be described as following:

1. **Initialize the population**: A population of programs is created randomly. The number of programs is set by the Population Size parameter. The average length of the programs is set by the program size initial parameter.

2. **Run a tournament**: Four programs are randomly picked out of the population of programs. It compares them and picks two winners and two losers based on fitness.

3. **Apply the Search Operators**: Search operators like crossover and mutation are then applied to the winners and this produces two “Children” or “Offsprings” as follows:
   a) Copy and crossover the copies of the winners,
   b) With Mutation Frequency, mutate one of the programs resulting from performing step 3a, and
   c) Mutate other of the programs resulting from performing step 3a.

4. **Replace the losers**: After the search operators have been applied to the copies
of the winners (the offspring), these offering replace the two losers in the tournament.

5. Repeat until termination: Steps 2 through 4 are repeated until the run is terminated.

The flowchart of Genetic Programming methodology is shown in Fig. 2.

Advantages of GP over Other Modeling Approaches

Genetic programming approach does not assume any a priori functional form of the solution unlike in a typical regression method. It also does not need to perform a task of initially defining the network structure as necessary neural networks. The building blocks (the input and target variables and the function set) are defined initially, and the learning method subsequently finds both the optimal structure of the model and its coefficients. GP evolves an equation or formula relating the input and output variables. A major advantage of the GP approach is its automatic ability to select input variables that contribute beneficially to the model and disregard those that do not. GP can thus substantially reduce the dimensionality of the input variables.

A BRIEF CASE STUDY FOR APPLICATION OF SVR AND GP FOR STREAMFLOW PREDICTION

Both the approaches, as described in the earlier sections, are applied to the problem of basin-scale streamflow prediction, both at monthly scale and at weekly scale. A case study describing the applications of these approaches for streamflow prediction of Mahanadi River of India is presented in this section. The input data pertains to Mahanadi River basin and the objective is to predict the future streamflow variables.

Study Area and Datasets

Mahanadi River, which encloses a drainage area of 51,000 square miles (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 560 miles (900 km). It’s one of the longest rivers in the country and drains a substantial part of peninsular India. The River covers the states of Chhattisgarh (75,136 km²), Orissa (65,580 km²), Bihar (635 km²) and Maharashtra (238 km²) for a total of 141589 km². Rainfall comes predominantly from the summer monsoon (June through September). The average annual rainfall in the basin is 1,463 mm.

Daily water stage data pertaining to Rainfall and consequent discharge (Streamflow) from Basantpur water level station (Station Code EM000R2) operated by the Water Resources Agency, was collected for this work.

Among these annual records, 13 (276 data) years’ data were used for calibration, and nine (102 data) years’ data were used for validation. The division of calibration
and validation sets is based on the chronology.

**Monthly Streamflow prediction using SVR**

**Input normalization**

The observed Stage discharge data consists of values varying over a significant range therefore, all input variables are normalized through Mean and Standard Deviation. This scheme prevents the model from being dominated by the variables with large values, and is commonly used in data-driven models, such as ANNs. Previous studies too, have shown that the SVM with normalized input data in the range from zero to one outperforms than that with un-scaled input data, Bray and Han (2004). Therefore, the SVR model is fed normalized data, and finally the model output stages are returned to their original scale. The results are then compared with the observed values.

The following equation is applied to normalize the data:

$$y_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j}$$  \hspace{1cm} (3)

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. $\mu_j$ and $\sigma_j$ are the mean and standard deviation for $j^{th}$ month respectively.

**Parameter calibration**

In this case study the SVR method is employed, so a kernel function had to be selected from the qualified functions. Among various kernel functions available, the radial basis function, which has a parameter $\sigma$, is adopted in this work.

The SVR model used herein has two parameters ($\gamma$, $\sigma$) to be determined. These parameters are interdependent, and their (near) optimal values are obtained by a trial and error method. The analyses and calculations of SVR herein are performed using LS-SVM software and based on above derived parameters, 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 tradeoff between $\gamma$ and $\sigma$. The resulting error parameters (Correlation Coefficient, Root Mean Square and Mean Absolute Error) between predicted and observed values were compared and the best combination was chosen for further testing and validation. Through the above process the values of $\gamma$ and $\sigma$ were found to be 2 and 1.5 respectively.
Results and Discussions

Once decided on the values of parameters, four different combinations of input data were tested to arrive at the best possible forecasting model. In case one, 276 training datasets containing coordinates with one input vector were fed to the model and performance of the model was recorded. Similar procedure was followed and datasets containing two, three and four input vectors as a single coordinate, were used to separately train the model. The model was then validated on remaining 102 validation datasets. The results were compared and tabulated. The performance statistics for de-normalized values for different sets of inputs for Support Vector Regression is given in Table 1. The model based upon two input vectors returned values which were slightly better than model developed on only one input vector and the model performance decreased substantially afterwards with each increase in numbers of input vectors. The results were also compared with another popular forecasting methodology, namely ARIMA and ARMA modeling. While applying this approach Autocorrelogram and Partial correlogram for the streamflow time series were obtained and the following models, ARIMA (5,0,1) and ARIMA (3,0,2) were chosen to forecast the streamflow. The correlation coefficient returned from ARIMA (5,0,1) and ARIMA (3,0,2) were in the range of 0.65 to 0.69 respectively as compared to 0.78 in case of Support Vector Regression.

The above results indicate that the performance of streamflow forecasting model developed from Support Vector Regression is better than that of other traditional methods like ARMA, ARIMA modeling approaches and further work using above discussed kernel based learning method “Support Vector Regression” can be taken up in the existing areas of hydrologic time series and forecasting.

<table>
<thead>
<tr>
<th>TABLE-1</th>
</tr>
</thead>
</table>

PERFORMANCE STATISTICS FOR DE-NORMALIZED VALUES FOR DIFFERENT SETS OF INPUTS FOR SUPPORT VECTOR REGRESSION

<table>
<thead>
<tr>
<th>No. of input vectors in a data set</th>
<th>Performance statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
</tr>
<tr>
<td>1</td>
<td>17983.33</td>
</tr>
<tr>
<td>2</td>
<td>18777.01</td>
</tr>
<tr>
<td>3</td>
<td>22014.20</td>
</tr>
<tr>
<td>4</td>
<td>209666.92</td>
</tr>
</tbody>
</table>
WEEKLY STREAMFLOW PREDICTION USING GENETIC PROGRAMMING (GP)

A GP model is developed to predict weekly streamflow at Basantpur stream gauging site across Mahanadi River in the state of Orissa, India. Total eight analyses were carried out to find out the most influential input variables and the best model for the prediction of weekly streamflow of Mahanadi River at Basantpur. The analyses were based on information of large-scale atmospheric circulations in form of weekly values of ENSO indices, EQUINOX indices, OLR anomalies for certain number of previous time steps.

Weekly data from January 1, 1990 to December 31, 2003 was used for this study. The period of analysis had to be limited to 1990 to 2003 due to non-availability of OLR data before 1990. Out of this period, weekly data 1990 through 1998 were used for training purpose and data from 1999 through 2003 were used for testing purpose.

The weekly streamflow is modeled as a function of (i) Historical average weekly streamflow for the particular week, (ii) Streamflow of certain number of previous weekly time steps, (iii) ENSO index of certain number of previous weekly time steps, (iv) EQUINOX index of certain number of previous weekly time steps, (v) Outgoing Longwave Radiation anomaly over the basin for certain number of previous weekly time steps. Thus,

\[ SF_t = f\{HSF_t, (SF_{t-1}, SF_{t-2}, \ldots ), (EN_{t-1}, EN_{t-2}, \ldots ), (EQ_{t-1}, EQ_{t-2}, \ldots ), (OLR_{t-1}, OLR_{t-2}, \ldots )\} \]

where \( SF \) stands for Streamflow, \( HSF \) stands for historical weekly average streamflow, \( EN \) stands for ENSO index, \( EQ \) stands for EQUINOX Index and \( OLR \) stands for outgoing longwave radiation anomaly. The optimum number of lags to be considered for each input variables was decided based on the ‘input impacts’ of that input variable during model calibration.

RESULTS

The model developed through Genetic Programming (GP) was used to capture the relationship between various inputs like local meteorological input (OLR) and large-scale circulation patterns (ENSO and EQUINOX) to predict and forecast weekly basin-scale streamflow of Mahanadi River basin. It was found that consideration of both large-scale and local influence (OLR) produced the best prediction performance among all other possibilities. The developed models were evaluated based on their performance through Correlation Coefficient (CC). The observed and predicted streamflows were found to have a correlation coefficient of 0.81, which is a highly encouraging result for such a complex system.
SUMMARY AND CONCLUSION

Hydrometeorological systems are significantly influenced by large scale atmospheric circulations patterns such as El Nino Southern Oscillation (ENSO), Equatorial Indian Ocean Oscillation (EQUINOO) etc, and hence as such are quite complex. To model the complexity of relationships between various input parameters, two Artificial Intelligence (AI) based approaches were presented in this paper. In the first approach, a kernel based modeling technique, Support Vector Regression (SVR) was applied to forecast the streamflow of Mahanadi river basin on a monthly basin-scale. The model contained only endogenous inputs and was fed four sets of data. The best result was obtained by a single input model and hence was selected for forecasting of monthly basin-scale streamflow. The resultant model was then compared with existing statistical (ARIMA, ARMA) Models. The SVM model returned a correlation coefficient of 0.78 as compared to correlation coefficient returned in the range of 0.65 and 0.69 by ARMA, and ARIMA models respectively.

In another approach described in this work, Genetic Programming was used to model the relationship between different input variables and subsequently forecast weekly basin scale stream-flow. The model successfully improved the performance of short term (Weekly) basin-scale streamflow prediction and gave encouraging results. It returned a correlation coefficient of 0.81, which is a highly commendable result for a complex system consisting of various input variables. The applications of both methodologies presented in this work demonstrate the tremendous potential held by algorithm based techniques in the general area of predictive hydrology and thus present a compelling case to further the research in predictive hydrology using above mentioned methods.

ACKNOWLEDGEMENTS

This work is partially supported by the Department of Science and Technology, Govt. of India, through project # SR/FTP/ETA-26/07 and Indian Institute of Technology Bombay, through project # 07IR020.

REFERENCES


