

Estimation of Discharge Using LS-SVM and Model Trees

V. G. Sayagavi^{1, *}, Shrikant Charhate², Rajendra Magar³

¹Civil Engineering Research Center, Datta Meghe College of Engineering, , Navi Mumbai, India

²Department of Civil Engineering, Pillai HOC College of Engineering & Technology, Rasayani, India

³Department of Civil Engineering, AIKTC School of Engineering & Technology, New Panvel, India

Email address:

sayagavi_vg@yahoo.co.in (V. G. Sayagavi), scharhate@mes.ac.in (S. Charhate), rajendramagar69@gmail.com (R. Magar)

*Corresponding author

To cite this article:

V. G. Sayagavi, Shrikant Charhate, Rajendra Magar. Estimation of Discharge Using LS-SVM and Model Trees. *Journal of Water Resources and Ocean Science*. Vol. 5, No. 6, 2016, pp. 78-86. doi: 10.11648/j.wros.20160506.11

Received: August 28, 2016; **Accepted:** October 13, 2016; **Published:** November 7, 2016

Abstract: In planning and management of any water resource systems prediction or estimation of runoff over the catchment is considered as a crucial factor. Many researchers over the past two decades addressed these problems by traditional methods as well as with some new techniques. This paper is describable and is focused on the capability of some data driven techniques such as Least Square Support Vector Machines (LS-SVM) and Model Trees with M5 algorithm. These methods were used to estimate runoff of various stations in the catchment area in Upper Krishna basin, Maharashtra State, India, and discussed here two stations namely Shigaon and Gudhe. The specialty of these catchment areas is Shigaon has large area and Gudhe has small area. This was done to see the model performance in both conditions. Additionally metrological data was used in the process to see the performance of models. The quantitative analysis was carried out to check the performance of the accuracy by considering standard statistical performance evaluation metrics and the scatter plots are drawn for evaluating qualitative performances of the developed models. The effect of the various metrological parameters as an input parameter for the rainfall was also investigated. The performance of both the tools was judged with various performance measures and it is found that the results are quite encouraging. LS-SVM models performed better since it has captured all the higher peak discharges for both catchments, indicating LS-SVM is best suited for large sized catchments and MT tool is best suited for the smaller sized catchments. However LS-SVM performance is better as compared to MT as modeling approaches are examined, using the long-term observations of yearly river flow discharges.

Keywords: Hydrological Processes, Runoff, Least Square Support Vector Machines, Model Trees

1. Introduction

In hydrology, prediction or estimation of runoff is most complex hydrological phenomena due to temporal and spatial variability of watershed as well as number of variables are involved in the process of rainfall-runoff. In last two decade data driven techniques are being used as an alternative approach for developing the models such as “ANN, Fuzzy logic, and GP, SVM and MT [3]”. Researchers mentions that amongst these there is no doubt that ANN and GP approaches have gained significant importance and popularity for developing rainfall-runoff models but use of tools like LS-SVM and MT can also yield good results and can further be explored as an alternative tool for developing rainfall-runoff models “[5], [15]”.

The main objective of the study was to explore the potentiality of developing the streamflow estimation models based upon Least Square Support Vector Machines (LS-SVM) and Model Trees (MT) M5P modeling techniques at daily scale using hydrological and metrological data. Many researchers have employed ANN and GP tools and have considered only stream flow data to predict runoff which describes the phenomena of the rainfall-runoff process Very few researchers has considered LS-SVM and MT for making a comparative evaluation of the data driven modeling techniques along with few metrological parameters for prediction of rainfall-runoff “[13], [1]”.

This study considers hydrological and metrological data as variables to describe the physical phenomena of the rainfall-runoff process, in order to estimate runoff (Q). While

developing the models, metrological parameters were added to the input to examine the effect to improve the accuracy of the model. Hence to achieve this objective, attempt has been made by developing the LS-SVM and MT models that have various input structures and applied for runoff estimation for the selected study area.

2. Methodology

In this paper we have used two methods of artificial intelligence, namely LS-SVM and Model Trees for two catchments and evaluate their performances with error criteria and model building with qualitative and quantitative analysis. The two methods adopted in the study are discussed below.

2.1. LS-SVM Approach

Support Vector Machines has gained significant importance and has been employed by many researchers “[2], [11], [19]”. When one has to use large data-set for prediction of the rainfall-runoff, use of SVM tool is not advised because of large number of parameters and of high level of computational efforts. Hence to overcome this difficulty “[16]” proposed modifications in SVM which has lead to LS-SVM. In this revision one finds the solution by solving a set of linear equations instead of a convex Quadratic Programming (QP) problem for classical SVM’s. LS-SVMs are a class of kernel-based learning methods.

Basically LS-SVM has been introduced in the framework of statistical learning theory for the purpose of functional estimation and pattern recognition. Least squares support vector machines (LS-SVM) involves equality instead of inequality constraints and works with a least squares cost function.

LS-SVM method received very little attention in the field of water resources “[18]”. Later there was an increase in the use of LS-SVMs in the modeling and forecasting of hydrological processes. “[2]”, for enhancing the prediction accuracy, employed grid search and cross-validation techniques to investigate the ability of LS-SVM, “[13]”, tried to explore the potentiality of LS-SVM model by adding the hydro-climatic variables, as well as streamflow values of the previous days as an input parameter and RBF kernel parameters are estimated based on the performance of the model for validation their results shown good performance of LS-SVM; “[19]”, considered LS-SVM for forecasting future stream flow discharge and their results mentions that LS-SVM based predictive models and training algorithm ensures accurate prediction by association with any natural measurable system. “[14]”, forecasted stream flow using PCA and LS-SVM by using monthly stream flow data and their results states that LS-SVM with PCA model performs better than only LS-SVM models. Later on “[11]”, developed nonlinear auto regression exogenous model (NARX0 and LS-SVM models for forecasting the future stream flow based on the previous day stream flow and the performances of the developed models were compared with Recurrent Neural Network models trained with Levenberg-Marquardt back propagation algorithm and proved that the performances of the models developed by using LS-SVM

tool was most suitable for their study area whereas “[19]”, considered three techniques namely LS-SVR, M5 Model Tree and Multivariate adaptive Regression Splines (MARS) for streamflow forecasting and prediction of monthly streamflows. In the first stage all models were compared with MLR in forecasting one month ahead of each station individually and in the second stage the models were evaluated and compared in predicting stream flow of one station using data of nearby station and their results indicates that LS-SVR model performed well in comparison with MARS and M5 MT.

2.2. Model Trees Approach

For learning, use of decision tree models is commonly used which is based on divide and conquer. Decision tree models are a non-parametric supervised learning method which is used for classification and regression. Each leaf in the tree contains a linear regression model, which is used to predict the target variable at that leaf, and the resulting model is called as Model Tree and was first introduced by “[9]” and was applied to hydrological modelling by many researchers, “[5], [4]”.

M5 model tree is a machine learning technique which uses an idea of splitting the parameter space into areas or subspaces and then it builds in each of subspaces a linear regression model. The splitting of the Model trees follows an idea which is used in building a decision tree, but instead of the class labels, it will be having linear regression functions at its leaves, which will be capable of predicting continuous numeric attributes. Hence Model Trees generalizes the concept of regression which has constant values at their leaves. So they are analogous to piecewise linear function and hence they are non-linear “[8]”. The M5 tree being a piecewise linear model, are having certain advantages such as MT’s are more transparent and very fast in training and they always converge; model trees are very much smaller than the regression trees, and the strength of the decision is clear and the regression functions do not normally involve much variables. “[3]”. M5 model tree algorithm was originally developed by “[9]”. M5P is a reconstruction of Quinlan’s M5 algorithm for inducing trees of regression models. Hence M5P combines a conventional decision tree with the possibility of linear regression functions at the nodes. For detailed procedure one can refer “[9]”. The application of M5 model trees in the field of hydrology for rainfall- runoff modelling is very limited; however some researchers who have considered MT tool for rainfall-runoff modelling are enumerated in the subsequent paragraph.

Reference “[15]” developed MT, ANN and GP models for streamflow forecasting considering daily streamflow as an input parameter for two catchments in the Narmada basin of India. Their work mentions that the GP models performed better as compared to ANN and MT models marginally however, the performance of ANN models were considerably better than the MT models. However “[17]”, analyzed the ability of data driven techniques such as ANN and MT to predict the next time step rainfall using lagged time series of observed daily rainfall It was observed that TLRN has captured the pattern in a better way and in Model of MT various

trials on pruning and smoothening were carried and found that un-pruned and un-smoothened MT performed better and the performance of both the models was found better. Thereafter, “[6]”, considered MLP-ANN and M5P for stream flow predictions and their results proved that M5P model tree were found to predict the flows significantly well and mentions that M5P MT seems to be sensitive to data splitting. Later on “[7]”, considered M5 MT tool for validation of simplified discharge prediction model by using precipitation and stream flow data for a catchment in Ireland and the results of M5 MT were significantly good. Recently, “[10]”, compared different data-driven techniques namely ANN and M5 Model Tree for the Chaskaman reservoir of Maharashtra, India and found that the M5 Model Tree technique performed reasonably well and gave more accurate results than other techniques.

3. Study Area and Data Analysis

This section discusses about the data collected from various locations and the analysis of the database was carried out to know certain statistical parameters.

3.1. Study Area

The study area selected was Upper Krishna basin which is located on the western regions of the Maharashtra State, India, lying between latitudes $13^{\circ}07'$ and $19^{\circ}20'N$ and longitudes $73^{\circ}22'$ E and $81^{\circ}10'E$. The average rainfall in the Krishna basin is 784 mm. The South West monsoon sets in the middle of June and withdraws during mid of October. About 90% of the rainfall occurs during the monsoon period of which more than 70% of the annual rainfall occurs during June, July, August, September and October.

The location of the study area comprises of two catchments namely, Shigaon (Fig.1) and Gudhe (Fig.2) located on Krishna River in the districts of Satara and Sangli of Maharashtra State of India. The data collected for the study area was obtained from Hydrological Data Center, Nashik, and official department of Government of Maharashtra, India.

Totally 9 years of data from 1/6/2002 to 31/10/2010 was collected for, hydrological and metrological parameters namely observed daily rainfall, runoff, daily pan evaporation, daily maximum and minimum temperatures, daily wind speed and maximum humidity.



Figure 1. Details of Shigaon catchment.

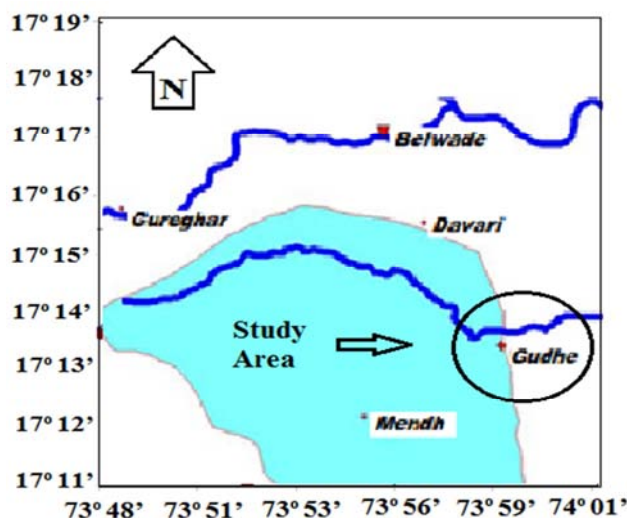


Figure 2. Details of Gudhe catchment.

3.2. Statistical Analysis

For the study area, rainfall, runoff and metrological data and information over a period of 9 was collected. From the data we intended to observe weather there was increase or decrease in the observed series to understand the probability distributions. Hence for monsoon periods of the data we have calculated certain statistical parameters viz. mean, standard deviation, skewness using Mat lab tools. The results of the same are graphically shown in Fig. 3:

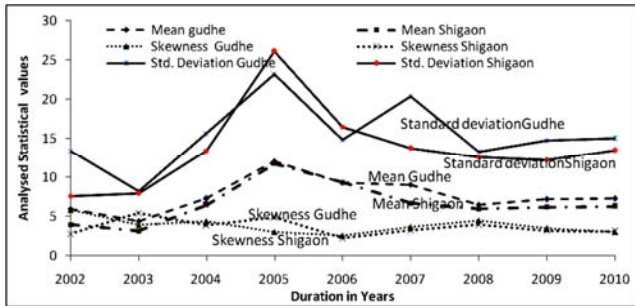


Figure 3. Graphical representation of the statistical Parameters for the study area.

Statistical parameters for the measured database in Gudhe and Shigaon catchments indicate that the minimum mean rainfalls of 5.949 mm and 3.1588 mm are in the year 2002 and 2003 respectively. Even though the rainfall was less, the rainfall was almost evenly distributed and the maximum mean rainfall of 12.0183 mm and 11.7438 mm was seen in the year 2005 in both of the catchments which indicates that the rainfall was not evenly distributed and for the remaining years we can observe that there is variation. In Gudhe catchment the standard deviation is lowest i.e. 8.2423mm is minimum for the year 2003 which indicates that the data points tend to be very close to the mean; and it can be seen that the standard deviation of 23.2271mm is maximum in the observed data set is more in the year 2005 which also shows that the data points does not tend to be very close to its mean but it has spread out over a large range of values. In Shigaon catchment also we can see that the minimum standard deviation is for year 2002 and maximum standard deviation is for the year 2005 which is 7.5604mm (minimum) and 26.0848 mm (maximum) respectively which indicates that data point tends to be very close to its mean and it also indicates that the data points are spread out over a large range of values. The skewness for Gudhe catchment is positive and is varying from 3.0 mm to 6.0 mm which

show that the data is positively skewed and the tail on the right side is longer or fatter than the left side. In Shigaon catchment the minimum skewness is observed in the year 2006 and maximum in the year 2003 which is 2.2599 mm and 5.4441 mm respectively which indicates that the skewness is positive and tail on the right side is longer or flatter than the right side of the mean. Hence by observing the qualitative analysis from Fig. 3 and after performing the calculation for different parameters of rainfall it is observed that the intensity of rainfall in some years is more and hence the standard deviation is more and also the skewness is leaning towards right side. In time series it is observed that the data consists of a systematic pattern (usually a set of identifiable components) and random noise (error), which usually makes the pattern difficult to identify.

4. Model Development

By observations of the data it is noticed that the average values of the discharges differs considerably. Because of variation in the stream inflow data between the two seasons that is monsoon and non monsoon seasons, it was decided to develop the models purely for the monsoon season months i.e. from June to October. Various models were developed initially using the values of rainfalls as an input parameter and metrological parameters in increasing number of order were added along with the rainfall as an input parameters to observe the accuracy of the developed models. By considering various options the models were developed. Table 1 gives the information on the data availability and its utility along with the methodology for Gudhe and for Shigaon catchment area respectively. Alternative LS-SVM and MT models were built as per the following functional relations for Gudhe catchment (P₁ to P₄) and for Shigaon catchment (P₁ to P₉)

$$\text{Model 1: } Q = f(P_{avg}) \tag{1}$$

$$\text{Model 2: } Q = f(P_1 + P_2 + P_3 + P_4 + \dots + P_9) \tag{2}$$

$$\text{Model3: } Q=f(P_1+..+P_9+MEP+MTN+MTX+MHS+MWS) \tag{3}$$

Where, Rainfall as ‘P’ is an input parameter and Runoff as ‘Q’ as output parameter MEP as Pan Evaporation as, MTX as Maximum Temperature and MTN as minimum Temperatures, MHS as Maximum Humidity and MWS as Maximum Wind speed as input parameters.

Table 1. Data usage and methodology adopted for Study area.

Available Data set	Duration and Data set considered for developing Models		Input Parameters for Gudhe Catchment Models			Input Parameters for Shigaon Catchment Models			Remark
	Training	Testing	Model 1	Model 2	Model 3	Model 1	Model 2	Model3	Model Output
9Years 1377	7 Years 963 (70%)	2 Years 414 (30%)	P _{avg.}	P ₁ + P ₂ + P ₃ + P ₄ .	P ₁ + P ₂ +P ₃ +P ₄ +MEP+MT N+ MTX+MHS+MWS	P _{avg.}	P ₁ + P ₂ +P ₃ +P ₄ +...+P ₉	P ₁ +P ₂ +P ₃ +P ₄ + ...+P ₉ +MEP+ MTN+ MTX+MHS+ MWS	Q

4.1. LS-SVM Model Calibration and Parameter Estimation

The LS-SVM models has two regularization parameters namely c , σ which are to be determined. The regularization parameter c determines the trade- off between the fitting error minimization and smoothness of the estimated function and σ is the RBF kernel parameter. To achieve maximum performance of the LS-SVM models these two parameters have to be calibrated during the time of model development because it is not known before hand which c and σ^2 are best suited for a particular problem to achieve maximum performance with LS-SVM models. As these parameters are independent their optimum (near) values are often obtained by trial and error method. For finding of all these parameters grid search method is employed in parameter space.

While developing the models care has been taken to main-

4.2. Performance Indicators (tabulated below in Table 2)

tain the same amount of the division of data for training and for testing various combination were tried to achieve maximum accuracy, however by trial and error it was found that, 70:30 i.e. 70% data for model building and 30 % for testing the model was performing accurately. This was in line with the study carried out by Londhe and Charhate (2010) in both the techniques for effective comparison of the accuracy of the models. The different models are developed initially using the values of rainfalls as an input parameter and addition of meteorological parameters which affect the process as per the functional relationship shown in eqn. 1, 2 and 3. For achieving these objectives, the LS-SVM toolbox based on MATLAB was used for developing the LS-SVM models and for M5 Model Trees the software WEKA developed by The University of Waikato, New Zealand was used to develop the models.

Table 2. Performance criteria used for qualitative analysis.

Sr. No.	Performance evaluation criteria	Formula
1	Pearson's Correlation Coefficient (R)	$\frac{\sum_{t=1}^N [Q_{obs}(t) - \overline{Q_{obs}}][Q_{est}(t) - \overline{Q_{est}}]}{\sqrt{\sum_{t=1}^N [Q_{obs}(t) - \overline{Q_{obs}}]^2} \sqrt{\sum_{t=1}^N [Q_{est}(t) - \overline{Q_{est}}]^2}}$
2	Root Mean Square Error (RMSE)	$\sqrt{\frac{\sum_{t=1}^N \{Q_{est}(t) - Q_{obs}(t)\}^2}{N}}$
3	Mean Absolute Error (MAE)	$\frac{\sum_{t=1}^N Q_{est}(t) - Q_{obs}(t) }{N}$
4	Normalized Mean Square Error (NMSE) (1-Isu et al, 1995)	$\frac{\sqrt{[(1/n) \sum_{t=1}^N \{Q_{est}(t) - Q_{obs}(t)\}^2]}}{(1/n) \sum_{t=1}^N [Q_{obs}(t)]}}$
5	Mean Error in Estimating Peak Parameter (%MF)	$\frac{Q_{estmax} - Q_{obsmax}}{E_{obsmax}} \times 100$

Where $Q_{obs}(t)$ is the observed value of discharge at time "s", $Q_{est}(t)$ the estimated value of discharge at time "s", 'N' the total number of data points, $\overline{Q_{obs}}$ the mean value of observed discharge, $\overline{Q_{est}}$ the mean value of estimated discharge, Q_{estmax} the maximum value of estimated discharge, and Q_{obsmax} the maximum value of observed discharge

LS-SVM tool, two parameters namely regularization parameter c and the RBF Kernel parameter σ^2 was calibrated by trial and error method by considering parameters grid search method. Range of c is considered to be 1 to 1000 with a resolution of 1 and σ^2 in the range of 0.01 to 1 with a resolution of 0.01 and the developed model's performance is to be then assed using the data set for testing period.

5. Results and Discussion

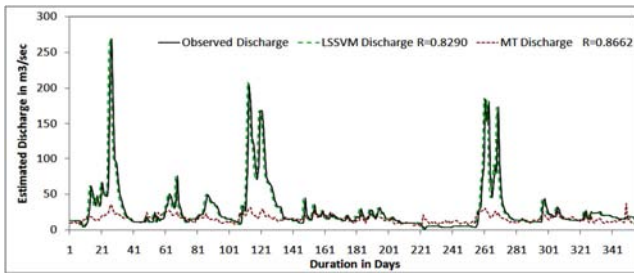
Input data for both the catchments are given as mentioned in Table 1. Before feeding the data for testing considering

Number of trials was considered for developing the models and their performances are discussed below and the one which had showed better performances is considered and the Results are tabulated in Table 3 and Table4.

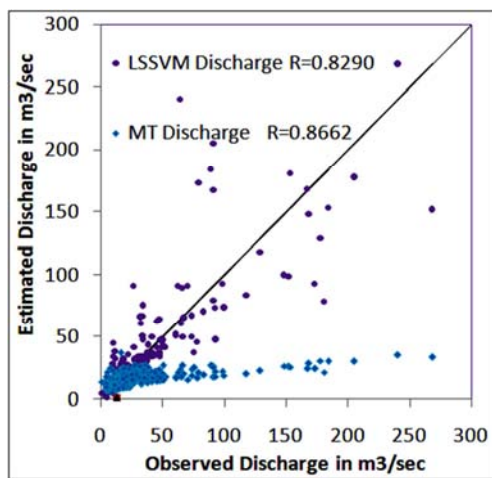
Table 3. Results of model testing using LS-SVM and tools for Gudhe catchment.

Model No.	Input Parameters	Model	R	RMSE	MAE	NMSE	%MF
1	Model1 P_{avg}	LS-SVM	0.856	19.429	-0.1413	19.767	-0.0002
		MT M5 P	0.864	23.048	16.391	63.379	90.40
2	Model2 $P_1 + P_2 + P_3 + P_4$	LS-SVM	0.833	19.429	-0.1413	19.767	-0.0002
		MT M5 P	0.827	30.520	16.238	44.748	-84.293
3	Model 3 $P_1 + P_2 + P_3 + P_4 + MEP + MTN + MTX + MHU + MWS$	LS-SVM	0.829	19.429	-0.141	19.767	-0.0002
		MT M5 P	0.866	22.893	13.689	61.692	-86.191

The time series plot and scattered plot for best models in Gudhe and Shigaon catchments are shown in the Fig. 4 a, b, and Fig. 6 a, b, respectively.



(a)



(b)

Figure 4. (a) Time series plot of Gudhe Catchment area, (b) Scatter plot of Gudhe Model 3.

Observations:

The time series plot (Fig. 4 a) for the Gudhe model 3 exhibits a good performance of M5P MT in testing with an R value of 0.8662 between the observed and predicted values than that of LS-SVM model 3.model having correlation coefficient 0.829. The scatter plot (Fig. 4 b) indicates a balanced scatter except at that of high measured values of discharges for LS-SVM and it also indicates that M5P MT models have far under estimated its peak discharge. LS-SVM model has predicted maximum peak discharge of 268.51 m³/sec which is equal to that of observed maximum discharge indicating that LS-SVM has exactly predicted the peak discharge where as MT model 3 is showing that the maximum predicted discharge is 37.07 m³/sec indicating MT has very much under predicted the discharge indicating the effect of temperature and evaporation is playing major roll in its under prediction. For Gudhe catchment area it may be noted that LS-SVM and M5 P model trees model have shown equal performances with respect to correlation for all the three Models. However, for this catchment M5P Model trees Model 3 is working well in comparison with the Model 1 and Model 2 with reference to correlation coefficient and other parametric evaluation measurers has also considerably reduced. The same can be revealed from MAE

and NMSE also. This might be because of the varying characteristics of the catchment with respect to its size, shape and other factors affecting runoff. By referring the Table 3 one can observe that the performances of LS-SVM seems to be good in comparison with the MT models but the predicted discharges are indicating certain amount of lag in Model 1 and in Model 2, it has over predicted and its performance with respect correlation coefficient went on decreasing by a lesser value. Hence the performances of LS-SVM might not be considered as good in this catchment Hence taking into considerations the parametric evaluation criteria it can be mentioned that the performance of M5P Model trees increased after adding metrological indicating characteristics of the catchment plays an important model in increasing efficiency of the model and it also indicates that the performance of LS-SVM is reduced because of the effect of the metrological parameters on the observed values of rainfall and runoff might have lead to inaccurate prediction of the discharge. Another reason for ineffective prediction of runoff value by LS-SVM tool is that rainfall might not occur on all the days of the monsoon and as such there will be zero values of measured rainfall in some of the raingauge stations resulting in less accuracy. Finally after comparing both the models it can be stated that for small sized catchment M5P model trees is best suited. The M5P MT algorithm 1 and the equations of LM for Model 3 of Gudhe catchment are as shown below along with the Pruned model tree (Fig. 5) obtained by using M5Pmodelling approach.

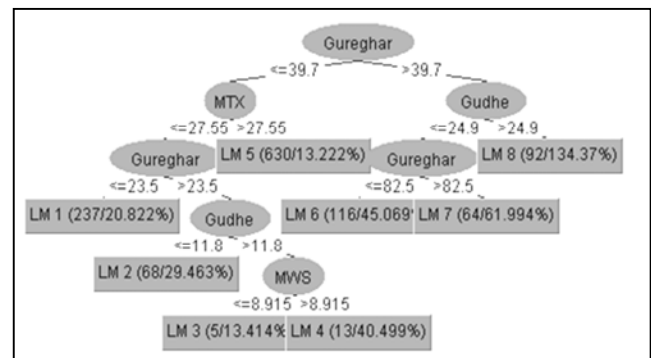


Figure 5. Pruned MT obtained by using M5P modelling approach for Model 3 of Gudhe catchment.

Algorithm 1: Tree-structured regression model generated by M5 algorithm using training data for Gudhe Catchment is *M5 pruned model tree: (using smoothed linear models) Number of Rules: 8

Gureghar \leq 39.7:

||MTX \leq 27.55:

|||Gureghar \leq 23.5: LM1 (237/20.822%)

|||Gureghar $>$ 23.5:

||||Gudhe \leq 11.8: LM2 (68/29.463%)

||||Gudhe $>$ 11.8: ||||MWS \leq 8.915: LM3 (5/13.414%)

||||MWS $>$ 8.915: LM4 (13/40.499%)

|||MTX $>$ 27.55: LM5 (630/13.222%)

Gureghar $>$ 39.7:

||Gudhe \leq 24.9:

||Gureghar<= 82.5: LM6 (116/45.069%)

||Gureghar> 82.5: LM7 (64/61.994%)

|Gudhe> 24.9: LM8 (92/134.37%)

LM num: 1 Runoff or discharge (HQC) = -0.2578 * MHS + 0.8908 * MWS - 1.7075 * MTX + 1.942 * MTN - 14.5651 * MEP + 0.0096 * Gudhe + 0.1752 * Mendh + 0.2178 * Dawari + 0.0339 * Gureghar + 65.9426

LM num: 2 Runoff or discharge (HQC) = - 0.0513 * MHS + 0.4598 * MWS - 1.1211 * MTX + 0.9806 * MTN - 2.7566 * MEP + 0.0096 * Gudhe + 0.106 * Mendh + 0.1044 * Dawari+ 0.0595 * Gureghar + 40.6549

LM num: 3 Runoff or discharge (HQC) = 0.9195 * MHS + 0.8477 * MWS - 2.1586 * MTX + 1.9009 * MTN - 2.7566 * MEP - 0.2938 * Gudhe + 0.0263 * Mendh + 0.5502 * Dawari + 1.5652 * Gureghar - 81.8037

LM num: 4 Runoff or discharge (HQC) = 1.6576 * MHS + 0.8477 * MWS- 2.1586 * MTX + 1.9009 * MTN - 2.7566 * MEP - 0.2071 * Gudhe + 0.0263 * Mendh + 0.9143 * Dawari

+ 3.2043 * Gureghar - 209.2988

LM num: 5 Runoff or discharge (HQC) = 0.0867 * MHS + 0.3679 * MWS - 0.8181 * MTX + 0.0102 * MTN - 10.5611 * MEP + 0.0952 * Gudhe + 0.0038 * Mendh + 0.0086 * Dawari + 0.0123 * Gureghar + 53.0049

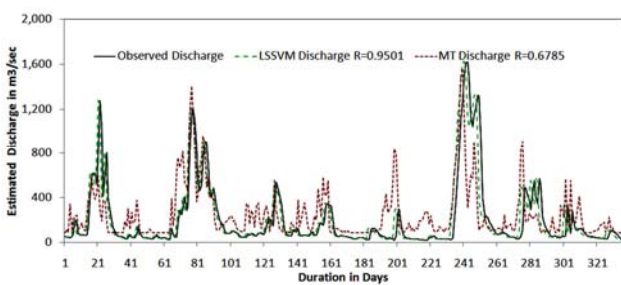
LM num: 6 Runoff or discharge (HQC) = - 0.0318 * MHS + 1.4593 * MWS - 3.342 * MTX - 1.1128 * MTN - 13.9689 * MEP + 0.1678 * Gudhe + 0.1282 * Mendh + 0.439 * Dawari + 0.0713 * Gureghar + 155.7341

LM num: 7 Runoff or discharge (HQC) = - 0.0318 * MHS + 2.6361 * MWS - 0.9989 * MTX- 14.6365 * MTN - 20.2014 * MEP + 1.8963 * Gudhe + 0.0395 * Mendh + 0.3512 * Dawari + 0.0824 * Gureghar + 390.0938

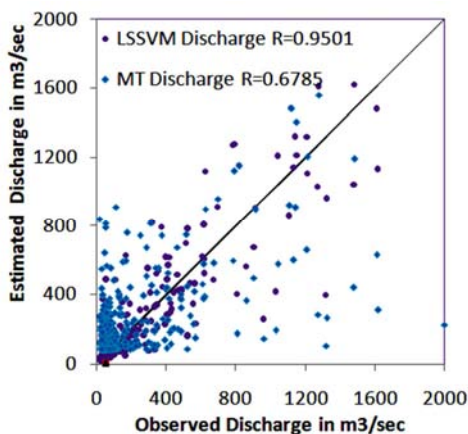
LM num: 8 Runoff or discharge (HQC) = - 2.6986 * MHS + 0.3361 * MWS - 22.2465 * MTX - 1.0533 * MTN - 67.2169 * MEP + 0.1421 * Gudhe + 0.0042 * Mendh + 0.0918 * Dawari + 0.6227 * Gureghar + 1021.656

Table 4. Results of model testing using LS-SVM and MT techniques for Shigaon catchment.

Model No.	Input Parameters	Tool	R	RMSE	MAE	NMSE	%MF
1	Model1	LS-SVM	0.95	106.912	-0.069	55.03	0.00011
	P _{avg}	MT M5 P	0.6785	260.189	33.613	382.6	-9.7382
2	Model2	LS-SVM	0.86145	106.913	-0.070	55.03	0.00011
	P ₁ + P ₂ +P ₃ +P ₄ +P ₅ + P ₆ +P ₇ + P ₈ + P ₉	MT M5 P	0.7328	242.247	153.26	322.2	-10.714
3	Model3	LS-SVM	0.86021	106.913	-0.070	55.03	0.00011
	P ₁ +P ₂ +P ₃ +P ₄ +P ₅ +P ₆ +P ₇ +P ₈ +P ₉ +MEP+MTN+MTX+MHU+MWS	MT M5 P	0.7328	245.946	154.25	322.2	-10.714

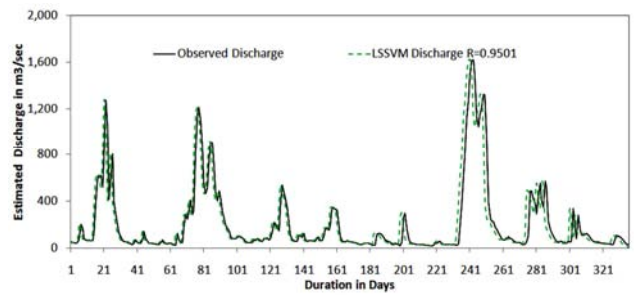


(a)

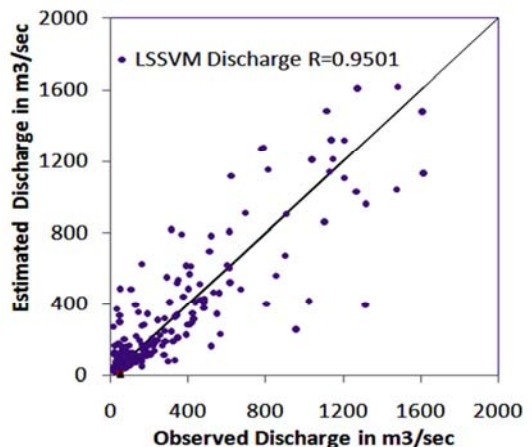


(b)

Figure 6. (a) Time series plot of Shigaon Catchment area, (b) Scatter plot of Shigaon Model 1.



(a)



(b)

Figure 7. (a) Time series plot of Shigaon catchment area, (b) Scatter plot of Shigaon model 1 using LS-SVM tool.

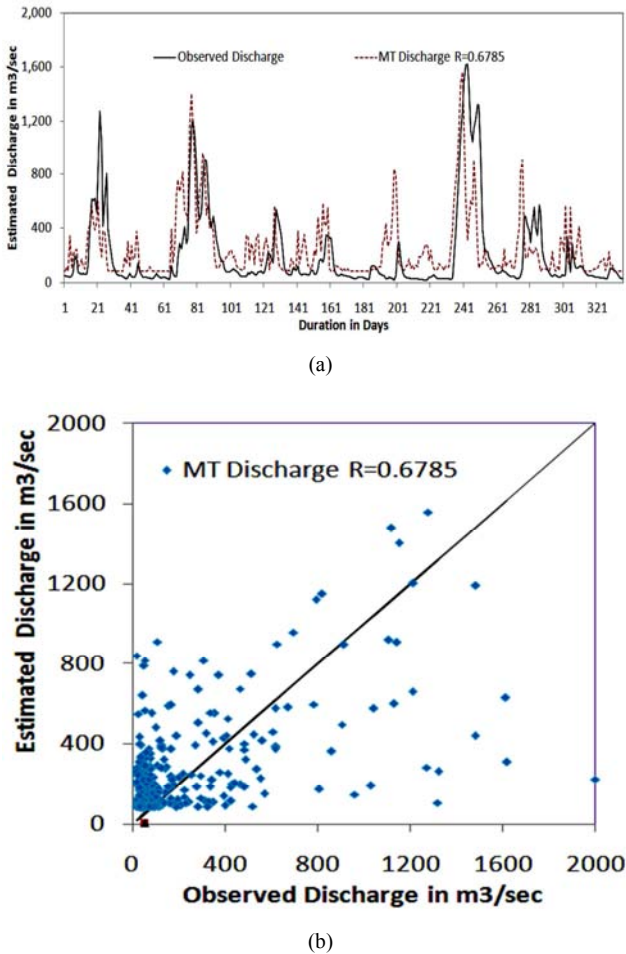


Figure 8. (a) Time series plot of Shigaon catchment area, (b) Scatter plot of Shigaon model 1 using MT tool.

Observations: For Shigaon model 1, by observing the time series plot (Fig.6 a), the Shigaon LS-SVM model 1 shows good performance having correlation coefficient (R) of 0.95 and MT model 1 performance is very less in its comparison having correlation coefficient 0.6785 between the observed and estimated values. The scatter plot (Fig. 6 b) confirms this with a balanced scatter for LS-SVM model. It has also kept the MAE and NMSE minimum. From the Shigaon time series model 1 (Fig.6 a) LS-SVM model has predicted its maximum discharge 1617.8 m³/sec which is same as that of observed maximum discharge 1617.8 m³/sec indicating that LS-SVM has exactly predicted the discharge (Fig. 7 a, b) and MT model is showing that the maximum predicted discharge as 1560.936m³/sec in comparison of observed discharge of 1617.79 m³/sec (Fig.8 a, b). This indicates that MT has a bit under predicted the higher peak discharges and over predicted its lower peaks discharges (Fig. 6 b). The performance evaluation metrics also reveals the same as shown in Table 4. The performances of LS-SVM models are consistent for Model 2 and Model 3 might be because of the characteristics of the size and shape of the catchments and at the same one can note that it has reduced its performance as compared to Model 1, indicating the effect of climatic conditions on the predicted discharges. Hence one can state that the LS-SVM modelling tools

works good for larger sized catchments as compared to small sized catchments.

6. Conclusions

The main objective of the study was to develop the rainfall-runoff models with various combinations of input parameters and compare the results of LS-SVM and MT tools for the study area. Both the models performed reasonably good during testing with little bit exceptions. The performance of the LS-SVM models was better for large sized catchments as compared to M5P model trees showing better correlation coefficient, minimum other errors especially Root Mean Square Error. Even though in some situations LS-SVM has over predicted the runoff still it can be considered as good because it has captured both the higher and lower peaks reasonably well. MT has performed well for Gudhe catchment area indicating the influence of metrological parameters on increasing the accuracy of prediction. It can be concluded that for both catchments the effect of characteristics of the catchment and the influence of the metrological parameters were observed in predicting the runoff as discussed above. However, the techniques like LS-SVM and Model trees need to be further explored in the fields of Water resources for sustainable development of the water resources projects

References

- [1] A. Elshorbagy *et al.* (2010), "Experimental investigation of the predictive capabilities "part 1 Hydrol. Earth Syst. Sci., 14, 1931–1941, 2010.
- [2] A. Shabri, and Suhartona, (2012), "Streamflow forecasting using least-square support vector machines", Hydrological Sciences Journal, 57 (7), pp 1275-1293.
- [3] B. Bhattacharya, and D.P. Solomatine, (2005) "Neural networks and M5 model trees in modelling water level–discharge relationship", Neurocomputing. 63, 381–396, 2005.
- [4] D. P. Solomatine, & Xue, Y. (2004b). "M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China". Journal of Hydrologic Engineering, 9, (6)491-501.2004.
- [5] D. P. Solomatine, and K. Dulal, (2003), "Model tree as an alternative to neural network in rainfall-runoff modeling" Hydrological Sciences, 48(3), 399–41, 2003.
- [6] E. K. Onyari and F. M. Ilunga (2013), "Application of MLP Neural Network and M5P Model Tree in Predicting Streamflow: A Model Study of Luvuvhu Catchment, South Africa." International Journal of Innovation, Management and Technology, Vol. 4, No. 1, February 2013.
- [7] H. Zia, Nick Harris, Geoff Merritt, Mark Rivers. (2015) "Validation of a Low Complexity Machine Learning Discharge Predictive Model.
- [8] I. H. Witten and E. Frank, (2005) "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations," Morgan Kaufmann, San Francisco, California, USA, 2005.

- [9] J. R. Quinlan, (1992) "Learning with continuous classes", Proc., 5th Australian Joint Conf. on Artificial Intelligence, Adams & Sterling, eds., World Scientific, Singapore, 343-348, 1992.
- [10] M Pawar, Samveda Mohite, Rushikesh Deshmukh, Nivedita Bhirud (2016) "Comparison of Various Data-Driven Modelling Techniques for Inflow Analysis". IOSR Journal of Computer Engineering (IOSR-JCE) Volume 18, Issue 2, Ver. V (Mar-Apr. 2016), PP 01-02.
- [11] N Zhang, Tilaye Alemayehu, Pradeep Behera (2015), "Non-linear Autoregressive (NAR) Forecasting Model for Potomac River Stage using Least Squares Support Vector Machines (LS-SVM)." International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-4 Issue-9, February 2015.
- [12] Nian Zhang, Charles Williams and Pradeep Behera (2014), "Water Quantity Prediction Using Least Squares Support Vector Machines (LSSVM) Method" Systemics Cybernetics and Informatics Volume 12-Number 4-Year 2014.
- [13] P. Bhagwat and Maity R., (2013), "Hydro-climatic stream flow prediction using Least Square Support Vector Regression," ISH Journal of Hydraulic Engineering, vol. 19, No. 3, pp 320-328.
- [14] S Ismail and AniShabri, (2014) "Stream Flow Forecasting using Principal Component Analysis and Least Square Support Vector Machine." J. Appl. Sci. & Agric., 9(11): 170-180, 2014.
- [15] S. N. Londhe, and S. B. Charhate, (2010) "Comparison of data driven modeling techniques for river flow forecasting", Hydrological Sciences. 55(7), pp1163-1174, 2010.
- [16] Suykens, K. A. J, Brabanter, D. J., Lukas, L. and Vandewalle, J. (2001), "Weighted least support vector machines: robustness and sparse approximation", ELSEVIER Science Neurocomputing: 48, 0925-2312/02, pp 85-105. Doi: 10.1023/A:101862860974.
- [17] T Mandal and V. Jothiprakash. (2012) "Short-term rainfall prediction using ANN and MT techniques" ISH Journal of Hydraulic Engineering Volume 18, Issue 1, 2012 pages 20-26.
- [18] X. Yunrong and J. Liangzhong (2009), "Water Quality Prediction Using LS - SVM and Particle Swarm Optimization".
- [19] Zaher Mundher Yaseen, Ozgur Kisi, Vahdettin Demir (2016), "Enhancing Long-Term Streamflow Forecasting and Predicting using Periodicity Data Component: Application of Artificial Intelligence" Journal of Water Resources Management pp 1-27 on line July 2016 DOI 10.1007/s11269-016-1408-5.