Journal of Modeling and Simulation of Materials
ISSN: 2582-2365
Volume 5, Issue 1, pp. 7-20, 2022
DOI: https://doi.org/10.21467/jmsm.5.1.7-20
RESEARCH ARTICLE

Comparative Analysis of Rainfall-Runoff Modeling Using Support Vector Machines for Two Dams in Uttarakhand

Basant Ballabh Dumka* and Pravendra Kumar

Department of Soil & Water Conservation Engineering, G.B. Pant University of Agriculture and Technology, College of Technology, Pantnagar – 263145, Uttarakhand, India.

*Corresponding Author e-mail: dumkabasant@gmail.com

Received: 11 August 2021 / Revised: 21 December 2022 / Accepted: 21 December 2022 / Published: 30 December 2022

ABSTRACT

The main objective of this study was to evaluate and compare the performances of rainfall-runoff models that were developed by using support vector machines (SVMs). Rainfall and runoff data of Haripura and Baur dams were adopted on daily basis from Irrigation Division Rudrapur in Uttarakhand. In this study, radial kernel function was used. As the values of Cost function (C), γ and ε varies, performances of the models can be altered. So, at optimum values of these variables, there exists a best correlation between rainfall and runoff. It can be inferred from the study that SVM models provide satisfactory results for both dams. These results can be used for runoff prediction for various purpose such as irrigation etc.

Keywords: Support Vector Machines; Radial Kernal Function; Pooled Average Relative Error.

1 Introduction

Water resources affect every part of the atmosphere, lithosphere, and hydrosphere, including living and non-living organisms on earth. Naturally, water is distributed in a different form at different sources like oceans, seas, bays, rivers, lakes, ponds, canals, springs etc. It is used for domestic, agriculture, industrial and commercial purposes [1]. As available water on earth varies with place and time, it is a topic for researchers to save the water because pure water present on the earth is in very little quantity. The COVID-19 pandemic has disrupted daily activities across multiple sectors globally [2], as the industries and human resources were affected by it. It has led to a dramatic loss of human life worldwide and presents an unprecedented challenge to public health, food systems and the world of work. The most common and recent water and COVID-19 relation is the threat this pandemic poses to the ability of water service providers (both formal and informal) to guarantee supply of water of a suitable quality, to enable sanitation and hygiene practices that limit the spread of the virus [3]. Water is a key resource for all types of activities on the earth. The quantity of water available on earth is nowadays under heavy stress due to high demand but limited availability. There is a need for sustainable water management that can ensure a narrow gap between the demand and supply of water resources. This inspires the researchers to the forecasting of water resources such as runoff, rainfall etc. In this study Gamma test was performed having various input combinations and various SVM models were used having rainfall and previous day's runoff and rainfall as inputs and runoff as output and compared their performances using some hydrological and statistical parameters.

Gamma test is generally used to find out the best input combination out of a number of input combinations. It can be defined as a non-parametric test which is based on some trial-and-error method. The Gamma test is a non-linear modeling and analysis tool to test the relationship between input and output variables on the numerical dataset [4]. The main principle in this test is that if there exists two input quantities m and m' that lie close to each other in an input space then the respective output quantities n and n' should also be close to each other in their given output space. One parameter gamma test in hydrology has limited use because it has relative inflexibility in fitting to frequency distributions of hydrologic variables but the two or three-parameter gamma distributions are generally adopted in place of the log-normal distribution function with two and three parameters. The selection of input variables builds a model structure, alters the weighted coefficient and influences the results of forecasting and prediction [5].

SVMs are advanced supervised learning machines which are basically used for classification, regression analysis and outlier detection. SVM algorithms can be employed for regression analysis tasks, but in practice they are mostly used for classification applications. In addition, there are some other types of algorithms in machine learning such as decision trees, random forest, and K-NN which are employed also for both classification and regression analysis tasks [6]. For a given set of training data in a two-dimensional learning task, an SVM learning algorithm develop a model that assigns new observation points to one of the two classes on both sides of a hyperplane that make it a non-probabilistic binary linear classifier. Support vector machines (SVMs) can be known as the most successful machine learning methods that are applied in the area of data mining [7]. The algorithms that are capable of learning inductively by using the examples have been used to various difficult, nonlinear, real-world problems of practical interest [8]. During the validation, the SVM model is better in performance as compared to other models [9]. Recently many researchers applied SVM successfully in rainfall- runoff modeling [10]–[12].

The present study has been carried out to develop rainfall-runoff models using the SVM technique for study areas after selecting the best inputs combination using the Gamma test (GT). The best model was selected by evaluating the performance and adequacy of the developed models.

2 Materials and Methods

2.1 General Description of Study Area

The Haripura and Baur dams are earthen dams, built on the Bhakhara and Baur (Junar) rivers respectively in Udham Singh Nagar district of Uttarakhand. The location of these dams lies between 29^0 8' N latitude and 79^0 20' E longitude and 29^0 8' N latitude and 79^0 18' E longitude respectively (as shown in Figure 1).

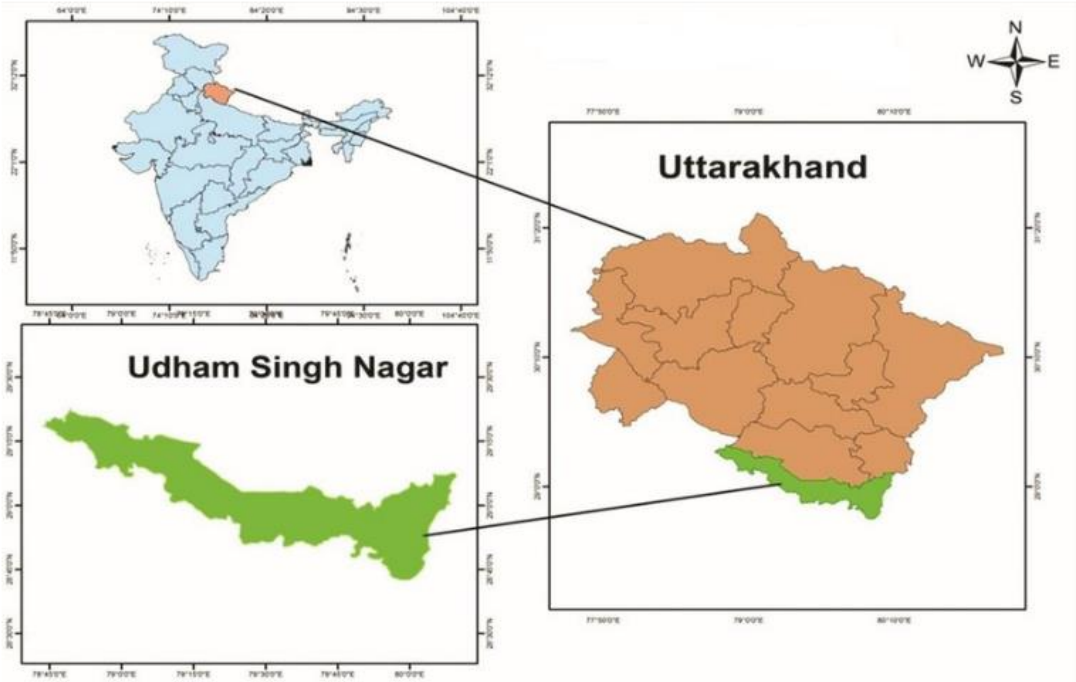


Figure 1: *Location map of the study area*

The catchment areas of these dams are 294.4 km^2 and 307.2 km^2 respectively that are hilly and partially plain. The reduced level of cut in between Baur and Haripura dams is 238.81 m. The climate of the study area is warm, and the average annual temperature is 24.3 °C. The normal monsoon rainfall is 1500 mm. The main crops in this region are rice, soyabean, urad, moong etc. (Kharif) and wheat, barley, gram, masoor, mustard etc. (Rabi). Soil pH is found slightly alkaline in Rudrapur, Gadarpur and Sitarganj region [13].

2.2 Data Collection

Data is obtained from the Irrigation division Rudrapur located at Rudrapur (U. S. Nagar) in Uttarakhand. Daily rainfall and corresponding runoff data for four months of monsoon season (June, July, August and September) were obtained for 20 years starting from 1996 to 2015 for Haripura dam and 8 years starting from 2006 to 2013 for Baur dam. Table 1 and Table 2 shows the statistical parameters of the data used for runoff prediction for Haripura and Baur dams respectively.

Table 1: *Statistical parameters of the dataset (Haripura dam) used for daily runoff prediction*

Statistical parameter	Whole data (1995-2015)		Training data		Testing data	
	P_t (mm)	Q_t (cusecs)	P_t (mm)	Q_t (cusecs)	P_t (mm)	Q_t (cusecs)
Mean	8.785	67.459	8.799	63.150	8.752	77.513
Standard Deviation	22.245	63.134	22.761	63.351	20.996	61.547
Kurtosis	45.831	0.526	54.426	0.986	17.250	-0.243
Skewness	5.186	0.773	5.677	0.937	3.695	0.428
Range	368.000	380.000	368.000	380.000	190.000	325.000
Minimum	0.000	0.000	0.000	0.000	0.000	0
Maximum	368.000	380.000	368.000	380.000	190.000	325.000
Count	2440	2440	1708	1708	732	732

Table 2: *Statistical parameters of the dataset (Baur dam) used for daily runoff prediction*

Statistical parameter	Whole data (2006-2013)		Training data		Testing data	
	P_t (mm)	Q_t (cusecs)	P_t (mm)	Q_t (cusecs)	P_t (mm)	Q_t (cusecs)
Mean	7.839	142.138	7.241	142.829	9.232	140.529
Standard Deviation	19.424	145.574	18.084	145.001	22.207	147.139
Kurtosis	19.930	0.4647	22.936	0.1458	20.071	1.180
Skewness	3.926	0.7467	3.834	0.8707	3.895	0.4780
Range	190	640	157	640	190	500
Minimum	0	0	0	0	0	0
Maximum	190	640	190	640	190	500
Count	976	976	683	683	293	293

2.3 Methodology

For obtaining an optimum and efficient training between input and output data, all input and output data were normalized by using a standard normal variable (z). It provides simple and fast training convergence within a small range during model development. It also eliminates dimensions thus give equal weightage to all variables. The standard normal variable is defined as,

$$z = \frac{x-\mu}{\sigma} \qquad \ldots (1)$$

Where μ = mean of the observed variable and σ = standard deviation of observed variable.

For Haripura dam 2440 sets of input and output data were distributed as 1708 sets (70%) for training and 732 sets (30%) for testing purposes. Similarly, for Baur dam 976 sets were distributed as 683 sets (70%) for training and 293 sets (30%) for testing purposes. These data were applied for development of models using support vector machines (SVMs).

2.4 Support Vector Machines (SVMs)

Support vector machines (SVMs) are advanced supervised learning techniques that are used to recognize patterns and analyze data. The support vector regression is an advanced form of the classification problem where a model provides a continuous valued output. In other words, a regression model finds out a continuous valued multivariate function. SVMs can solve binary classification problems by formulating them as convex optimization problems [14]. The optimization problem associated with finding out the maximum margin separating the hyperplane, whereas correctly classifying as many training points is possible. SVMs represent this optimal hyperplane with support vectors. The accurate solution and good generalization of the SVM leads to using it in regression problems. SVM generalization to SVR is accomplished by introducing an ε -insensitive region around the function, called the ε -tube. This tube helps in reformulating the optimization problem such that best approximation of the continuous-valued function can be obtained. At the same time balancing model complexity and minimizing prediction error are also functions of ε -tube.

If the training data coordinates, *T*, are denoted as

$$T = \{(x_1, y_1), (x_2, y_2), (x_m, y_m)\} \quad \dots \quad (2)$$

where $x \in X \subset R^n$ are the training inputs and $y \in Y \subset R^n$ are the training outputs.

Assume a non-linear function $f(x)$ is given by,

$$f(x) = w^T\Phi(x_i) + b \quad \dots \quad (3)$$

where w is the weight vector, b is the bias and $\Phi(x_i)$ is the high dimensional feature space. Now the aim is to fit the dataset T by finding a function $f(x)$ which has the largest deviation ε from the actual targets for all the training data T, and at the same time is as small as possible.

Here min. $(\frac{1}{2}w^Tw)$ is subject to

$$\{y_i - (w^T\Phi(x_i) + b) \leq \varepsilon \; y_i - (w^T\Phi(x_i) + b) \geq \varepsilon \quad \dots \quad (4)$$

And min. $(\frac{1}{2}w^Tw + C\sum_{i=1}^{m}(\xi_i^+ + \xi_i^-))$ is subjected to

$$\{y_i - w^T\Phi(x_i) - b \leq \varepsilon + \xi_i^+ \; w^T\Phi(x_i) + b - y_i \leq \varepsilon + \xi_i^- \; \xi_i^+, \xi_i^-\} \geq 0 \quad \dots\dots\dots(5)$$

where w^Tw represents the regularization term, $C\sum_{i=1}^{m}(\xi_i^+ + \xi_i^-)$ is called empirical term and measure ε-insensitive loss function, slack variables *i.e.,* ξ_i^+ *and* ξ_i^- represents upper and lower deviations, respectively. The final expression for SVM becomes,

$$f(x) = \sum_{i=1}^{m}(\alpha_i^+ - \alpha_i^-)K(x_i, x_j) + b \quad \dots \quad (6)$$

where α_i^+ *and* α_i^- are Lagrange multipliers. Therefore, the support vectors are points where exactly one of the Lagrange multipliers is greater than zero [15].

2.4.1 Radial basis kernel function (RBF) (Gaussian kernel)

Gaussian RBF (Radial basis function) is a well-known kernel function used in SVM models for modeling. RBF kernel is a function whose value depends on the distance from the origin or from some point. Gaussian Kernel is of the following format as,

$$K(X_1, X_2) = \exp(-\gamma \parallel X_1 - X_2 \parallel^2) \quad \dots \quad (7)$$

where $||X_1 - X_2||$ = Euclidean distance between X_1 & X_2.

2.5 Gamma test (GT)

For determining gamma (Γ) value the least square regression line is drawn such as,

$$y = A\delta + \Gamma \quad (8)$$

The intercept on vertical axis indicates Γ value.

Another standard term used for this purpose is v ratio which can be defined as,

$$V_{ratio} = \frac{\Gamma}{\sigma^2(y)} \quad \ldots \quad (9)$$

Value of V_{ratio} varies between 0 and 1. Its value closer to 1 denotes the higher degree of predictability of output.

2.6 Development of SVM models

Different SVM models were developed for hydrological modeling using daily rainfall and corresponding runoff data of monsoon season for the prediction of daily runoff in Haripura and Baur dams. In this study, it was considered that present day runoff is a function of present-day rainfall and past day's rainfall and runoff. The relationship is shown schematically in Figure 2 and also expressed functionally as,

$$Q(t) = f(P(t),\ P(t-1), P(t-2), \ldots\ldots, P(t-m), Q(t-1), Q(t-2), \ldots\ldots, Q(t-n)) \quad \ldots \quad (10)$$

where $P(t)$ is present day rainfall, $P(t-1)$ is previous one day rainfall, $P(t-m)$ is previous m days rainfall, $Q(t)$ is present day runoff, $Q(t-1)$ is previous one day runoff and $Q(t-n)$ is previous m days runoff.

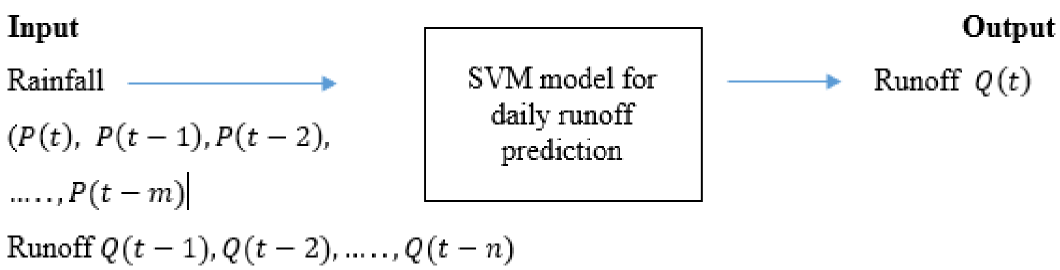


Figure 2: *Schematic representation of SVM models for rainfall-runoff modeling*

2.7 Performance Indicators of Models

2.7.1 Qualitative evaluation

The qualitative evaluation includes visual inspection of graphs prepared with observed and predicted runoff data. The comparison of graphs represents either under-prediction or over-prediction of the models or degree of under-prediction or over-prediction. In the present study time series plots and scatter plots were used as qualitative performance of the model.

2.7.2 Quantitative evaluation

For selecting a best performance model with some quantitative values various statistical such as correlation coefficient (r) and root mean square error (RMSE) and hydrological parameters such as Nash-Sutcliff coefficient of efficiency (NSCE) and pooled average relative error (PARE) were used.
Pooled average relative error is the indication of under-prediction and over-prediction performance of developed models. It can be expressed as,

$$PARE\ (\%) = \frac{1}{N}\left\{\frac{\sum_{i=1}^{N}(Q_P - Q_0)}{\sum_{i=1}^{N} Q_0}\right\} \times 100 \quad \ldots (11)$$

The positive value of PARE shows over-prediction and the negative value of PARE shows under-prediction performance of the developed model.

3 Results and Discussions

3.1 Gamma Test

Various possible combinations using different variables were developed for Haripura and Baur dams as presented in Table 3 and Table 4, respectively. Table 3 shows model M25 (P_t, P_{t-1}, Q_{t-1}, Q_{t-2}, Q_{t-3}) with mask 1100111 represented minimum Gamma value (0.0312) and minimum V-Ratio (0.1249) value as compared to other models hence it was selected as the best model for input selection in rainfall-runoff modelling for daily runoff prediction for Haripura dam. Table 4 shows model M25 (P_t, P_{t-1}, Q_{t-1}, Q_{t-2}, Q_{t-3}) with mask 1100111 represented minimum Gamma value (0.0572) and minimum V-Ratio (0.2289) value as compared to other models hence it was selected as the best model for input selection in rainfall-runoff modeling for daily runoff prediction for Baur dam. These selected input combinations for both dams increase the efficiency of the models. Figure 3 shows the flow chart of complete process of modeling.

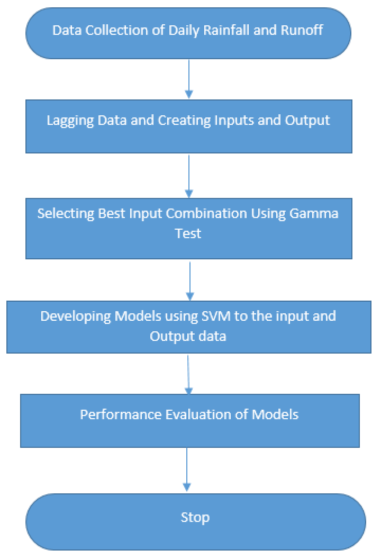


Figure 3: *Flow Chart of Complete Process of Modeling*

Table 3. *Selection of the best input combination for rainfall-runoff modeling for Haripura dam data using Gamma test* $\{Q_t = f(P_t, P_{t-1}, P_{t-2}, P_{t-3}, Q_{t-1}, Q_{t-2}, Q_{t-3})\}$

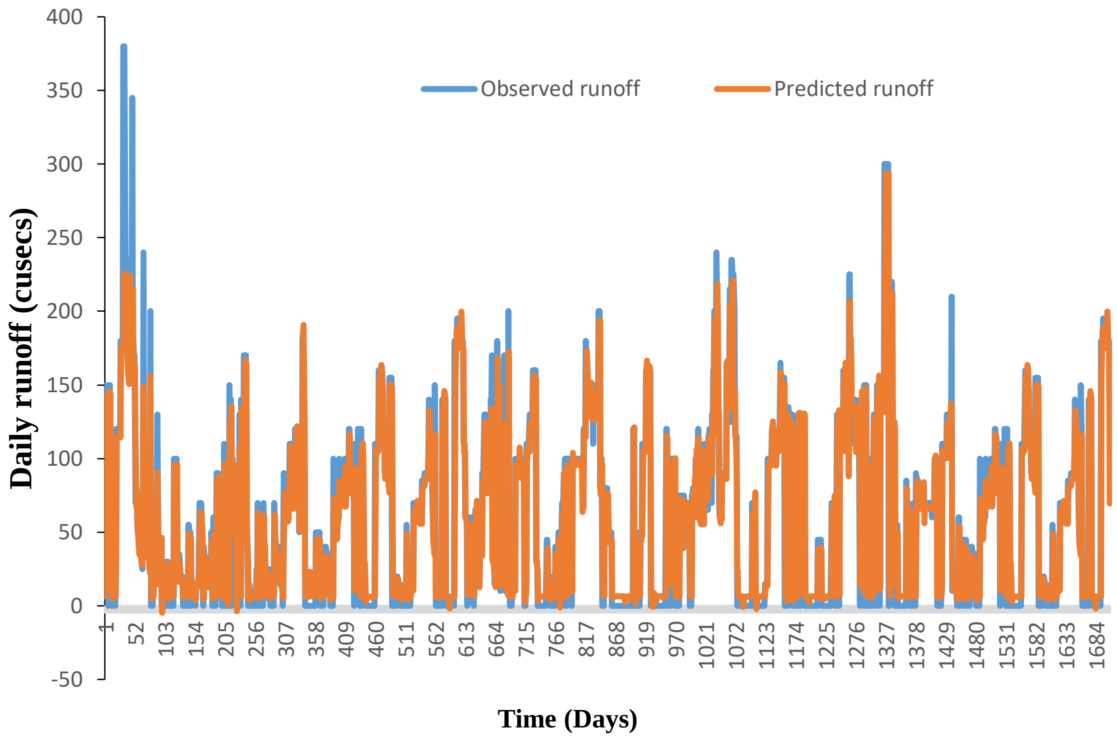
Model	Model input combination	Mask	Gamma	V-Ratio
M1	P_t	1000000	0.2429	0.9716
M2	P_t, P_{t-1}	1100000	0.2490	0.9960
M3	P_t, P_{t-1}, P_{t-2}	1110000	0.2499	0.9998
M4	$P_t, P_{t-1}, P_{t-2}, P_{t-3}$	1111000	0.2477	0.9910
M5	$P_t, P_{t-1}, P_{t-2}, P_{t-3}, Q_{t-1}$	1111100	0.0685	0.2744
M6	$P_t, P_{t-1}, P_{t-2}, P_{t-3}, Q_{t-1}, Q_{t-2}$	1111110	0.0443	0.1774
M7	$P_t, P_{t-1}, P_{t-2}, P_{t-3}, Q_{t-1}, Q_{t-2}, Q_{t-3}$	1111111	0.0512	0.2050
M8	Q_{t-1}	0000100	0.0752	0.3010
M9	Q_{t-1}, Q_{t-2}	0000110	0.0652	0.2610
M10	$Q_{t-1}, Q_{t-2}, Q_{t-3}$	0000111	0.0704	0.2817
M11	P_{t-1}, Q_{t-1}	0100100	0.0745	0.2982
M12	$P_{t-1}, Q_{t-1}, Q_{t-2}$	0100110	0.0674	0.2698
M13	$P_{t-1}, Q_{t-1}, Q_{t-2}, Q_{t-3}$	0100111	0.0713	0.2853
M14	P_{t-2}, Q_{t-1}	0010100	0.1155	0.4621
M15	$P_{t-2}, Q_{t-1}, Q_{t-2}$	0010110	0.0816	0.3267
M16	$P_{t-2}, Q_{t-1}, Q_{t-2}, Q_{t-3}$	0010111	0.0715	0.2863
M17	P_{t-3}, Q_{t-1}	0001100	0.0779	0.3116
M18	$P_{t-3}, Q_{t-1}, Q_{t-2},$	0001110	0.0676	0.2705
M19	$P_{t-3}, Q_{t-1}, Q_{t-2}, Q_{t-3}$	0001111	0.0703	0.2812
M20	P_t, P_{t-2}	1010000	0.2217	0.8869
M21	P_t, Q_{t-1}, Q_{t-2}	1000110	0.0602	0.2410
M22	$P_t, Q_{t-1}, Q_{t-2}, Q_{t-3}$	1000111	0.0649	0.2599
M23	P_t, P_{t-1}, Q_{t-1}	1100100	0.0715	0.2865
M24	$P_t, P_{t-1}, Q_{t-1}, Q_{t-2}$	1100110	0.0433	0.1735
M25	$\mathbf{P_t, P_{t-1}, Q_{t-1}, Q_{t-2}, Q_{t-3}}$	**1100111**	**0.0312**	**0.1249**
M26	$P_t, P_{t-1}, P_{t-2}, Q_{t-1}$	1110100	0.0796	0.3184
M27	$P_t, P_{t-1}, P_{t-2}, Q_{t-1}, Q_{t-2}$	1110110	0.0534	0.2137
M28	$P_t, P_{t-1}, P_{t-2}, Q_{t-1}, Q_{t-2}, Q_{t-3}$	1110111	0.0484	0.1937

Table 4.: *Selection of the best input combination for rainfall-runoff modeling for Baur dam data using Gamma test* $\{Q_t = f(P_t, P_{t-1}, P_{t-2}, P_{t-3}, Q_{t-1}, Q_{t-2}, Q_{t-3})\}$

Model	Model input combination	Mask	Gamma	V-Ratio
M1	P_t	1000000	0.6878	0.9751
M2	P_t, P_{t-1}	1100000	0.0820	0.3283
M3	P_t, P_{t-1}, P_{t-2}	1110000	0.0747	0.2989
M4	P_t, P_{t-1}, P_{t-2}, P_{t-3}	1111000	0.0724	0.2573
M5	P_t, P_{t-1}, P_{t-2}, P_{t-3}, Q_{t-1}	1111100	0.0624	0.2499
M6	P_t, P_{t-1}, P_{t-2}, P_{t-3}, Q_{t-1}, Q_{t-2}	1111110	0.0715	0.2863
M7	P_t, P_{t-1}, P_{t-2}, P_{t-3}, Q_{t-1}, Q_{t-2}, Q_{t-3}	1111111	0.0699	0.2799
M8	Q_{t-1}	0000100	0.2012	0.8049
M9	Q_{t-1}, Q_{t-2}	0000110	0.2095	0.8383
M10	Q_{t-1}, Q_{t-2}, Q_{t-3}	0000111	0.2130	0.8523
M11	P_{t-1}, Q_{t-1}	0100100	0.1867	0.7470
M12	P_{t-1}, Q_{t-1}, Q_{t-2}	0100110	0.1548	0.6195
M13	P_{t-1}, Q_{t-1}, Q_{t-2}, Q_{t-3}	0100111	0.1609	0.6437
M14	P_{t-2}, Q_{t-1}	0010100	0.2047	0.8188
M15	P_{t-2}, Q_{t-1}, Q_{t-2}	0010110	0.1638	0.6554
M16	P_{t-2}, Q_{t-1}, Q_{t-2}, Q_{t-3}	0010111	0.1364	0.5459
M17	P_{t-3}, Q_{t-1}	0001100	0.2180	0.8722
M18	P_{t-3}, Q_{t-1}, Q_{t-2},	0001110	0.2251	0.9004
M19	P_{t-3}, Q_{t-1}, Q_{t-2}, Q_{t-3}	0001111	0.1268	0.5057
M20	P_t, P_{t-2}	1010000	0.2409	0.9639
M21	P_t, Q_{t-1}, Q_{t-2}	1000110	0.0639	0.2558
M22	P_t, Q_{t-1}, Q_{t-2}, Q_{t-3}	1000111	0.0626	0.2504
M23	P_t, P_{t-1}, Q_{t-1}	1100100	0.0972	0.3891
M24	P_t, P_{t-1}, Q_{t-1}, Q_{t-2}	1100110	0.0809	0.3237
M25	**P_t, P_{t-1}, Q_{t-1}, Q_{t-2}, Q_{t-3}**	**1100111**	**0.0572**	**0.2289**
M26	P_t, P_{t-1}, P_{t-2}, Q_{t-1}	1110100	0.0822	0.3291
M27	P_t, P_{t-1}, P_{t-2}, Q_{t-1}, Q_{t-2}	1110110	0.0934	0.3736
M28	P_t, P_{t-1}, P_{t-2}, Q_{t-1}, Q_{t-2}, Q_{t-3}	1110111	0.0636	0.2547

3.2 Rainfall-Runoff Modelling using SVM

In the present study, R software was used for prediction of runoff for 70% data as training and 30% for testing purposes. Package e1071 was installed in R for running the program. The radial function SVM (SVM-RF) was used for runoff prediction. The three parameters of SVM i.e., C, γ *and* ε were varied based on trial-and-error methods for development of SVM-RF models. The observed and predicted data were then compared, and the efficiency of the developed model was then found out. After comparing the prediction performances of all the models developed for Haripura dam shown in Table 5, it can be found out that models SVM-RF-4 (C = 10, $\gamma = 0.333, \varepsilon = 0.01$) for training and SVM-RF-4 (C = 10, $\gamma = 0.333, \varepsilon = 0.01$) for testing performed the best. It can be observed from Table 5 that for SVM-RF-4, value of RMSE is 34.38 cusec, correlation coefficient is 0.842, NSCE is 0.705 and PARE is -0.98$\times 10^{-3}$ for training data and 31.65 cusec, 0.859, 0.734 and -1.45$\times 10^{-3}$ for testing, respectively. Figure 4 and Figure 5 show under-predicted models for training and testing, respectively as observed and predicted runoff values represented to be in close agreement.



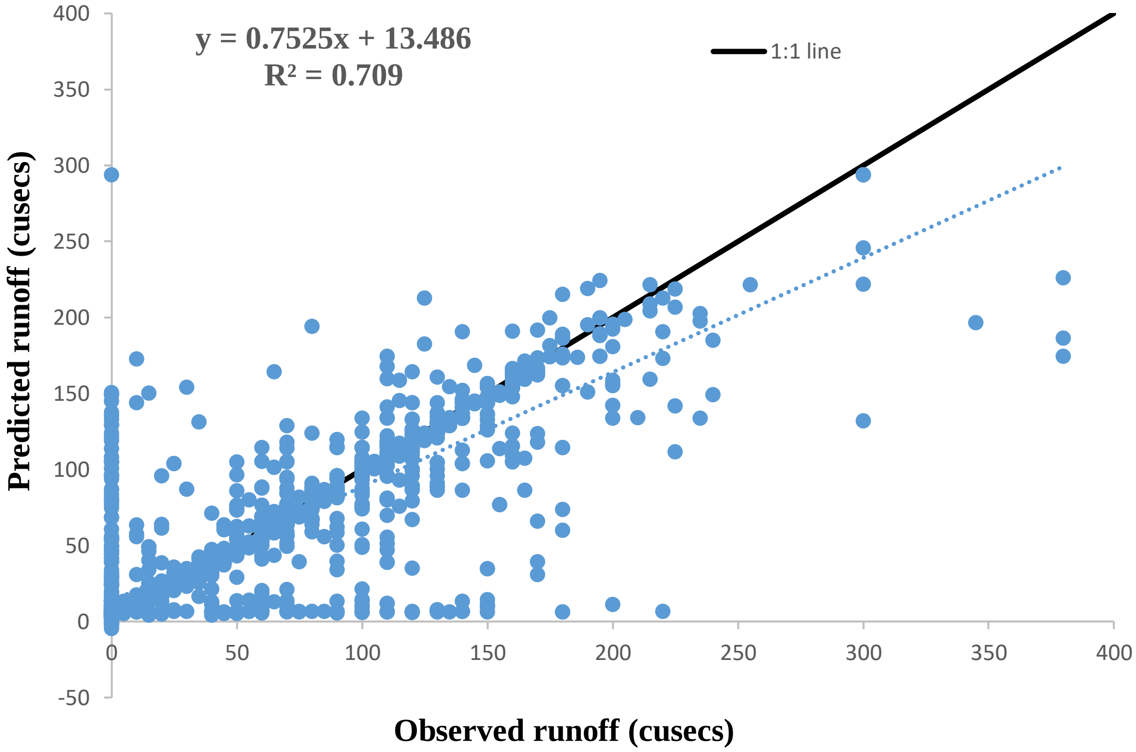
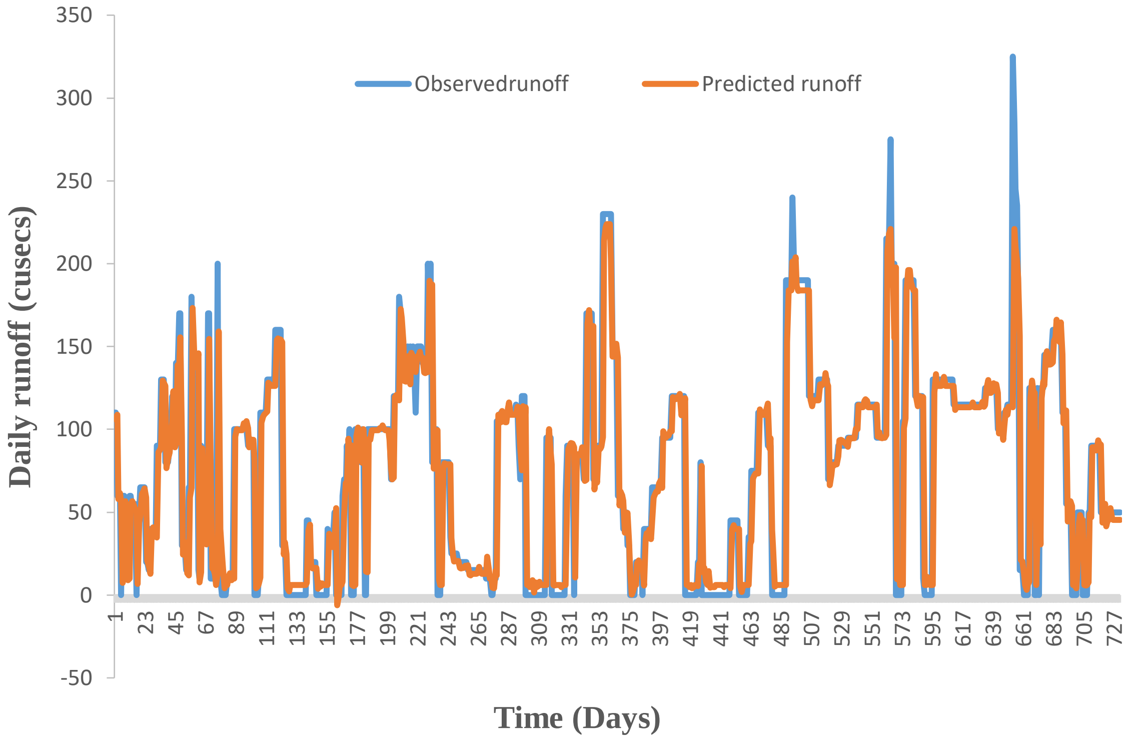


Figure 4: *Time series and scatter plots of predicted and observed runoff for Support vector machine model (SVM-RF-4) during training period for Haripura dam*



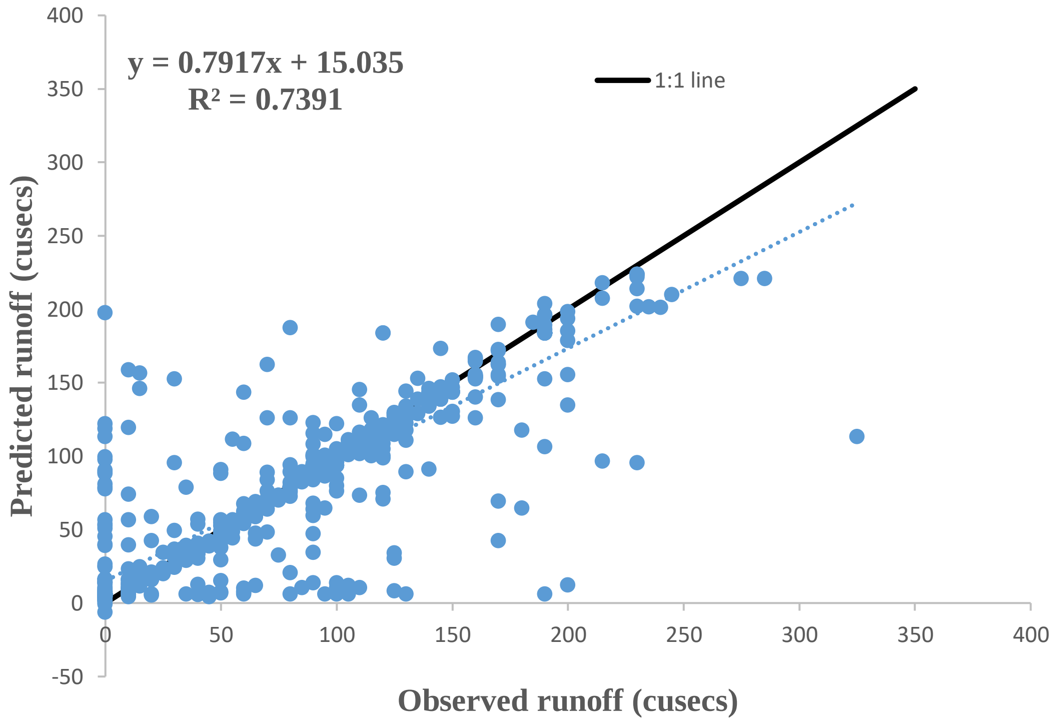


Figure 5: *Time series and scatter plots of predicted and observed runoff for Support vector machine model (SVM-RF-4) during testing period for Haripura dam*

Table 5: *Results of different performance indicators for support vector machine-based runoff (Prediction models of Haripura dam)*

Model	Architecture	Training				Testing			
		RMSE	r	NSCE	PARE (10^{-3})	RMSE	r	NSCE	PARE (10^{-3})
SVM-RF-1	C = 10, $\gamma = 0.01$, $\varepsilon = 0.05$	36.23	0.820	0.672	0.56	42.35	0.730	0.525	-6.91
SVM-RF-2	C = 10, $\gamma = 0.9$, $\varepsilon = 0.01$	36.10	0.821	0.674	-1.45	42.30	0.734	0.526	1.75
SVM-RF-3	C = 10, $\gamma = 0.333$, $\varepsilon = 0.1$	35.33	0.830	0.688	-2.89	43.48	0.719	0.499	-7.76
SVM-RF-4	C = 10, $\gamma = 0.333$, $\varepsilon = 0.01$	**34.38**	**0.842**	**0.705**	**-0.98**	**31.65**	**0.859**	**0.734**	**-1.45**
SVM-RF-5	C = 10, $\gamma = 0.333$, $\varepsilon = 0.001$	35.46	0.829	0.686	-2.67	42.65	0.726	0.518	-1.54
SVM-RF-6	C = 10, $\gamma = 0.5$, $\varepsilon = 0.1$	35.66	0.827	0.682	-1.66	42.42	0.731	0.524	-2.34

Table 6: *Results of different performance indicators for support vector runoff prediction models of Baur dam*

Model	Architecture	Training				Testing			
		RMSE	r	NSCE	PARE (10^{-3})	RMSE	r	NSCE	PARE (10^{-3})
SVM-RF-1	C = 10, $\gamma = 0.01$, $\varepsilon = 0.05$	61.98	0.780	0.546	-5.96	78.23	0.609	0.668	-9.98
SVM-RF-2	C = 10, $\gamma = 0.9$, $\varepsilon = 0.01$	64.87	0.770	0.447	-2.18	81.34	0.599	0.763	-2.98
SVM-RF-3	C = 10, $\gamma = 0.244$, $\varepsilon = 0.1$	**60.16**	**0.814**	**0.712**	**-1.02**	**75.19**	**0.770**	**0.669**	**-0.62**
SVM-RF-4	C = 10, $\gamma = 0.244$, $\varepsilon = 0.01$	67.98	0.739	0.699	-3.17	83.98	0.614	0.712	-3.57
SVM-RF-5	C = 10, $\gamma = 0.244$, $\varepsilon = 0.001$	65.58	0.757	0.589	-2.45	77.90	0.634	0.509	0.91
SVM-RF-6	C = 10, $\gamma = 0.5$, $\varepsilon = 0.1$	60.56	0.786	0.487	-4.56	90.45	0.566	0.665	-7.67

In case of Baur dam shown in Table 6, it can be found out that models SVM-RF-3 (C = 10, $\gamma = 0.244$, $\varepsilon = 0.1$) for training and SVM-RF-3 (C = 10, $\gamma = 0.244, \varepsilon = 0.1$) for testing performed the best. It can be observed from Table 6 that for SVM-RF-3, value of RMSE is 60.16 cusec, correlation coefficient is

0.814, NSCE is 0.712 and PARE is -1.02$\times 10^{-3}$ for training data and 75.19 cusec, 0.770, 0.669 and -0.62$\times 10^{-3}$ for testing, respectively. Figure 6 and Figure 7 show under-predicted models for training and testing, respectively as observed and predicted runoff values represented to be in sufficiently close agreement.

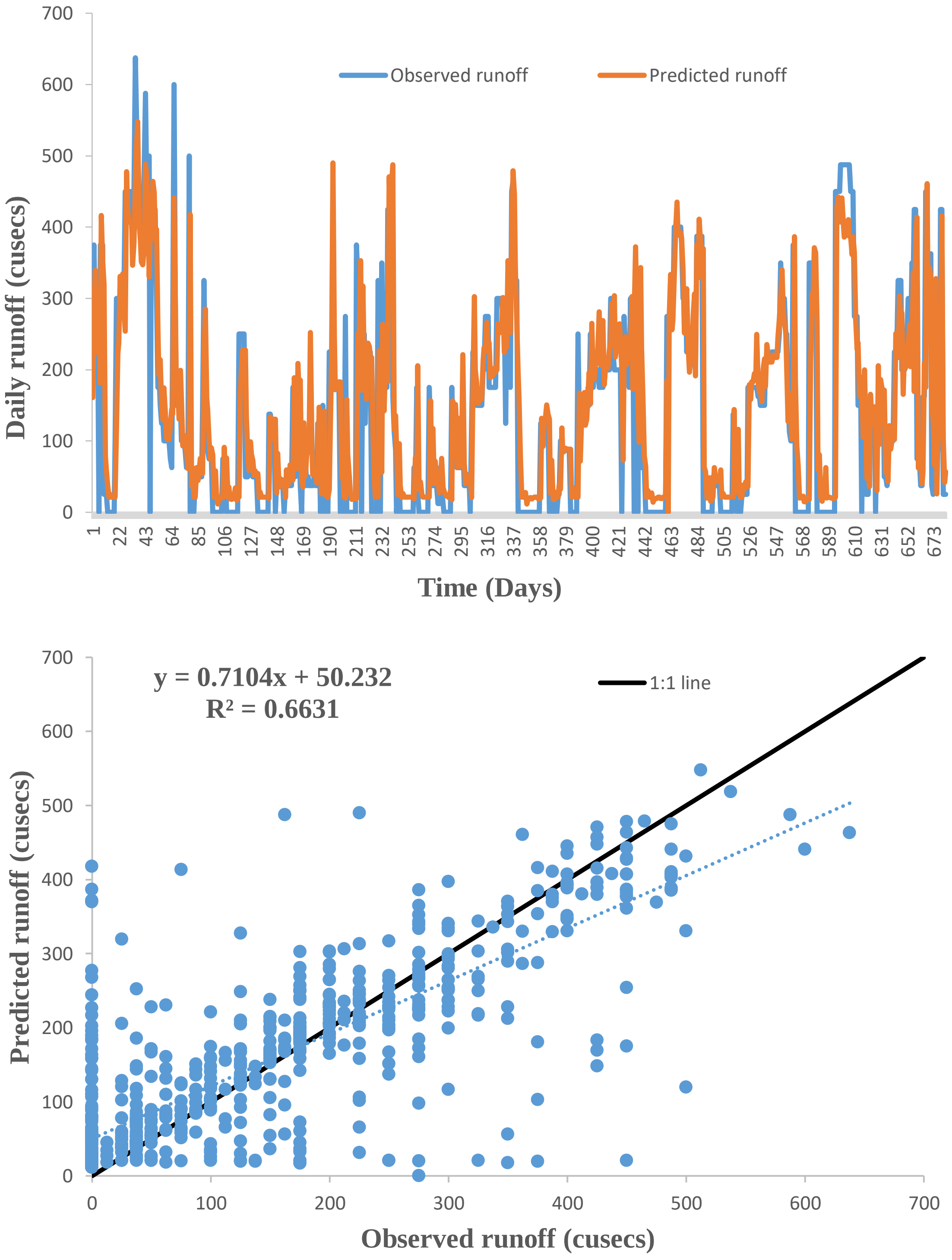
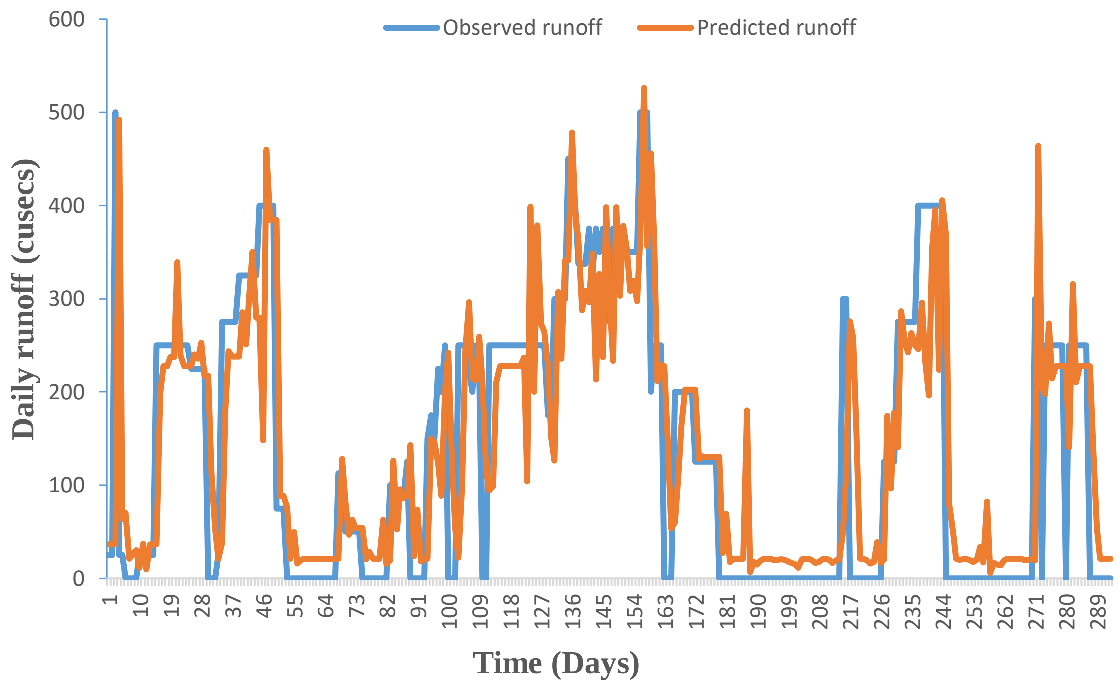


Figure 6.: *Time series and scatter plots of predicted and observed runoff for Support vector machine model (SVM-RF-3) during training period for Baur dam*



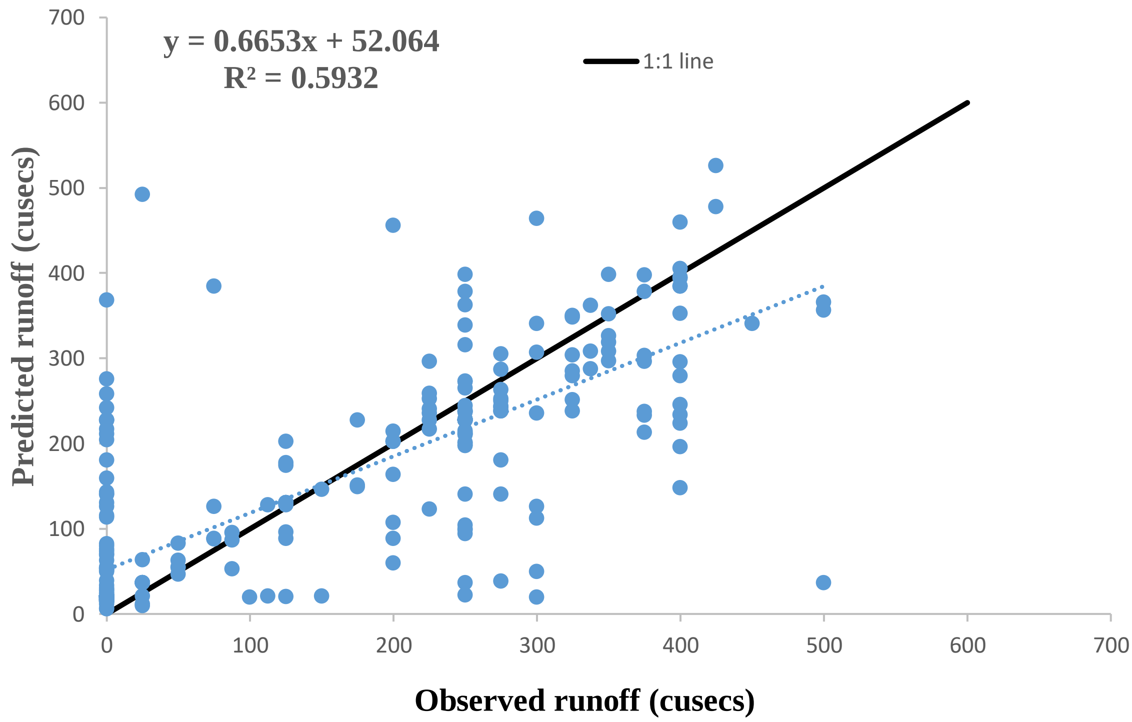


Figure 7: *Time series and scatter plots of predicted and observed runoff for Support vector machine model (SVM-RF-3) during testing period for Baur dam*

The lower value of RMSE shows the on an average there is lesser difference in estimated and actual value of output. Higher value of correlation coefficient represents a better correlation between inputs and output. Larger value of NSCE represents the input and output data tends to be in perfect fitting. Positive value of PARE represents over prediction and negative value represents under prediction of output.
In all models cost function (C) remains constant. Error tolerance (ε) and Gaussian Kernel parameter (γ) take different values to optimize the performance. After performing several iterations, it was found that for a particular combination of these values, the model gives best performance. These obtained results are in

line with the results of support vector machine in Kumar *et al.* [16] who analyzed the stage-discharge-sediment modelling using SVM and ANN techniques and found SVM is a good technique for stage-discharge modelling.

The statistical data represents that the best models for Haripura and Baur dams as mentioned above, are sufficiently good for runoff forecasting for the study area. As these dams are mainly aimed to irrigation purpose, hence one can use the models for future availability of water for crops and management of water as per sustainable need for public. These models can be efficiently used to the other areas having similar watershed and climatic characteristics. It can be more generalized as the models does not include the soil parameters of study area.

4 Conclusion

Since the past few decades SVM techniques have been used in rainfall-runoff modeling. These computing techniques are superior to conventional computing methods for runoff prediction. They have capability to capture nonlinear and non-stationary behavior of time series data. Support vector machine (SVM) models have satisfactory performance because they have better generalization capacity and rapid learning than conventional methods. As the study areas are close to each other, their climatic and meteorological conditions and even soil characteristics are the same. Hence both dams are comparable, and a common model can be suggested for both Haripura and Baur dams. In case, if data is unavailable for one of the dams it can be generated by using data from another dam by various methods.

5 Declarations

5.1 Acknowledgement

The authors would like to thank the Irrigation Division Rudrapur for providing the rainfall and runoff data for Haripura and Baur dams used in this study. They also are thankful to College of Technology Pantnagar for providing software facilities.

5.2 Competing Interests

The authors declared that no conflict of interest exist in this work.

5.3 Publisher’s Note

AIJR remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

How to Cite this Article:

B. B. Dumka and P. Kumar, “Comparative Analysis of Rainfall-Runoff Modeling Using Support Vector Machines for Two Dams in Uttarakhand”, *J. Mod. Sim. Mater.*, vol. 5, no. 1, pp. 7–20, Dec. 2022. https://doi.org/10.21467/jmsm.5.1.7-20

References

[1] P. G. Shinde, “WATER SCENARIO 2025,” in *National Level Conference on Water Management Scenario 2025 Problems, Issues and Challenges*. Accessed: Jan. 09, 2023. [Online]. Available: https://vpmthane.org/publication-wms/WMScenario%202025.pdf

[2] M. Balamurugan, K. S. Kasiviswanathan, I. Ilampooranan, and B. S. Soundharajan, “COVID-19 Lockdown Disruptions on Water Resources, Wastewater, and Agriculture in India,” *Frontiers in Water*, vol. 3, p. 24, Mar. 2021, doi: 10.3389/FRWA.2021.603531/BIBTEX.

[3] M. J. Neal, “COVID-19 and water resources management: reframing our priorities as a water sector,” *Water Int*, vol. 45, no. 5, pp. 435–440, Jul. 2020, doi: 10.1080/02508060.2020.1773648.

[4] Y. Li, G. Li, L. Guo, L. Hernández-Callejo, S. Nesmachnow, and S. G. Saavedra, “Feature Selection for Regression Based on Gamma Test Nested Monte Carlo Tree Search,” *Entropy 2021, Vol. 23, Page 1331*, vol. 23, no. 10, p. 1331, Oct. 2021, doi: 10.3390/E23101331.

[5] V. Güldal and H. Tongal, “Comparison of recurrent neural network, adaptive neuro-fuzzy inference system and stochastic models in eğirdir lake level forecasting,” *Water Resources Management*, vol. 24, no. 1, pp. 105–128, Jan. 2010, doi: 10.1007/S11269-009-9439-9.

[6] B. M. Gupta, S. M. Dhawan, and G. M. Mamdapur, “Support Vector Machine (SVM) Research in India: A Scientometric Evaluation of India’s Publications Output during 2002-19,” *The Journal of Indian Library Association*, vol. 57, no. 3, pp. 12–25, 2021.

[7] B. Pang, S. Guo, L. Xiong, and C. Li, "A nonlinear perturbation model based on artificial neural network," *J Hydrol (Amst)*, vol. 333, no. 2–4, pp. 504–516, Feb. 2007, doi: 10.1016/J.JHYDROL.2006.09.015.

[8] T. Roshni, M. K. Jha, and J. Drisya, "Neural network modeling for groundwater-level forecasting in coastal aquifers," *Neural Comput Appl*, vol. 32, no. 16, pp. 12737–12754, Aug. 2020, doi: 10.1007/S00521-020-04722-Z/METRICS.

[9] Z. Hassan *et al.*, "Comparison of Artificial Neural Network and Support Vector Machine for Long-Term Runoff Simulation," *Earth Environ. Sci.*, vol. 476, no. 1, p. 012119, Jun. 2020, doi: 10.1088/1755-1315/476/1/012119.

[10] M. Bray and D. Han, "Identification of support vector machines for runoff modelling," *Journal of Hydroinformatics*, vol. 6, no. 4, pp. 265–280, Oct. 2004, doi: 10.2166/HYDRO.2004.0020.

[11] Y. Han, S.-L. Lau, M. Kayhanian, and M. K. Stenstrom, "Characteristics of Highway Stormwater Runoff," *Water Environment Research*, vol. 78, no. 12, pp. 2377–2388, Nov. 2006, doi: 10.2175/106143006X95447.

[12] F. Sedighi, M. Vafakhah, and M. R. Javadi, "Rainfall–Runoff Modeling Using Support Vector Machine in Snow-Affected Watershed," *Arab J Sci Eng*, vol. 41, no. 10, pp. 4065–4076, Oct. 2016, doi: 10.1007/S13369-016-2095-5/METRICS.

[13] A. Tripathi, S. Maithani, and S. Kumar, "X-band persistent SAR interferometry for surface subsidence detection in Rudrapur City, India," in *Proc. SPIE 10793, Remote Sensing Technologies and Applications in Urban Environments III*, Oct. 2018, vol. 10793, pp. 105–115. doi: 10.1117/12.2326267.

[14] V. Vapnik, "The Support Vector Method of Function Estimation," in *Nonlinear Modeling*, J. A. K. Suykens and J. Vandewalle, Eds. Springer, Boston, MA, 1998, pp. 55–85. doi: 10.1007/978-1-4615-5703-6_3.

[15] T. Evgeniou and M. Pontil, "Support vector machines: Theory and applications," in *Machine Learning and Its Applications*, vol. 2049 LNAI, Springer Verlag, 2001, pp. 249–257. doi: 10.1007/3-540-44673-7_12/COVER.

[16] M. Kumar, P. Kumar, A. Kumar, A. Elbeltagi, and A. Kuriqi, "Modeling stage–discharge–sediment using support vector machine and artificial neural network coupled with wavelet transform," *Appl Water Sci*, vol. 12, no. 5, May 2022, doi: 10.1007/S13201-022-01621-7.