The Pharma Innovation Journal 2021; SP-10(1): 149-154

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.03
TPI 2021; SP-10(1): 149-154

www.thepharmajournal.com
Received: 10-11-2020
Accepted: 12-12-2020

Manish Kumar
Department of Soil and Water Conservation Engineering, College of Technology, G.B.P.U.A.&T., Pantnagar, Uttarakhand, India

Pravendra Kumar
Department of Soil and Water Conservation Engineering, College of Technology, G.B.P.U.A.&T., Pantnagar, Uttarakhand, India

Corresponding Author:
Manish Kumar
Department of Soil and Water Conservation Engineering, College of Technology, G.B.P.U.A.&T., Pantnagar, Uttarakhand, India

Stage-discharge-sediment modelling using support vector machine

Manish Kumar and Pravendra Kumar

Abstract
The sediment modelling are important aspects in planning and management of river dynamics. The stage and discharge play crucial role in sediment flow through a river channel. The present study carried out for stage-discharge-sediment modelling using support vector machine (SVM) and wavelet-based support vector machine (WSVM) for Adityapur site of Jharkhand, India. The best inputs variable was selected based on the gamma test. SVM and WSVM model were developed based on training (70% of total dataset) and testing (30% of total dataset). The developed models were analyzed through root mean square error (RMSE, g/l), Pearson correlation coefficient (PCC), Wilmott index (WI), line diagram and scatter plots. Based on line diagram and scatter plots, the results are under-predicted and over-predicted the SSC values. The results showed that the values of R^2, RMSE, PCC and WI were found to be 0.3560 and 0.4078, 0.090 g/l and 0.077 g/l, 0.597 and 0.639, 0.744 and 0.767 for SVM and WSVM models, respectively during testing period. Therefore, WSVM model found to be superior and can be applied to predict daily SSC for Adityapur site.

Keywords: stage, discharge, SSC, support vector machine, wavelet transform, Adityapur

Introduction

For the measurement of sediment transport in rivers, the construction of dams, reservoirs and channels, environmental effect evaluation, and the determination of the effectiveness of watershed management and other catchment treatment, the assessment of the amount of sediments being transported by a river is relevant. To measure the sediment load being carried by a flow, sediment rating curves based on regression analysis are commonly used. Regression and curve-fitting approaches are ineffective considering the severity of the problem (Kisi, 2005) [6]. The high degree of dispersion, which can be decreased but not removed, is an issue inherent in the rating curve system (Jain, 2001) [4]. Various studies performed to compare the actual and projected suspended concentration show that traditional rating curves will dramatically under-predict actual sediment concentrations (Asselman, 2000) [1]. Several attempts have been made to resolve statistical inaccuracies correlated with curve fitting, and numerous approaches have been proposed to optimise traditional sediment rating curves by adding different statistical correction variables, utilising nonlinear regression, or classifying discharge and sediment data into distinct classes (Singh and Durgunoglu, 1989; Phillips, *et al.*, 1999; Asselman, 2000) [18, 15, 1].

SVM is a system in which the strong points of conventional statistical approaches are used, which are technically based and analytically straightforward. In the fields of hydrology and time series forecasting, the SVM method has been widely applied. Liong and Sivapragasam (2002) [10] used the flood-predicting process. By integrating Chaos Theory and the SVM system, Yu *et al.*, (2004) [20] introduced a method for predicting the daily runoff. The method of support vector regression (SVR) has recently been developed based on SVM and demonstrates dominance in hydrological process prediction. Kalteh (2013) [5] revealed that both models coupled with wavelet transformation provided more reliable results than the regular models by applying an Artificial Neural Network (ANN) and SVR to monthly streamflow recorded at two separate stations. In contrast to ANN models, the findings have specified that SVR models had better performance. In order to predict river flow one day ahead in two studied places, Londhe and Gavraskar (2015) [12] used the SVR model. According to the low values of the measuring parameters, the model findings were favorable.

More recently, combined wavelet and artificial neural network (WNN) research has attracted growing attention and has demonstrated advantages in terms of strong fitting and prediction precision over a single ANN model and other modelling approaches. Wavelet Analysis (WA)

is becoming a popular technique of analysis due to its ability to expose both spectral and temporal information simultaneously within a single signal (Nourani *et al.*, 2009; Kumar and Kumar 2020; Kumari *et al.*, 2020) [14, 7, 8, 9]. In modelling studies on hydrometeorological variables, WA was used: precipitation, flow, wind and suspended sediment brought inside the flow (Unal *et al.*, 2004) [19]. Cannas *et al.*, (2006) [2] integrated continuous and discrete wavelet transforms to evaluate the effects of data pre-processing for river flow forecasting using neural networks. The findings found that pre-processed data-trained networks worked better than networks trained on unrecompensed signal data. Keeping these facts, the present study utilized following objectives: (i) to select best input variables based on gamma test, (ii) to apply data-driven techniques for modelling stage-discharge-sediment relationships, (iii) to evaluate performance evaluation of developed models.

Materials and Methods

Study area and data descriptions

The Adityapur study site is located in Saraikela Kharsawan district of Jharkhand, India. It is situated at latitude of 22° 47'29" N and longitude 86° 10'06" E. The study area is a small part of Subernrekha river basin. It is generally affected by south west monsoon and receive annual rainfall about 1400 mm. The temperature variation in winter were 10 °C to 15 °C while 35 °C to 40 °C during summer season. The present study used daily data of sage (m), discharge (Q, cumec) and suspended sediment concentration (SSC, g/l) for the duration of 10 year (2004-2013). The total datasets were divided into two sets namely, training and testing. The training dataset utilizes 70 % of the total datasets while rest 30 % were comes under testing phases.

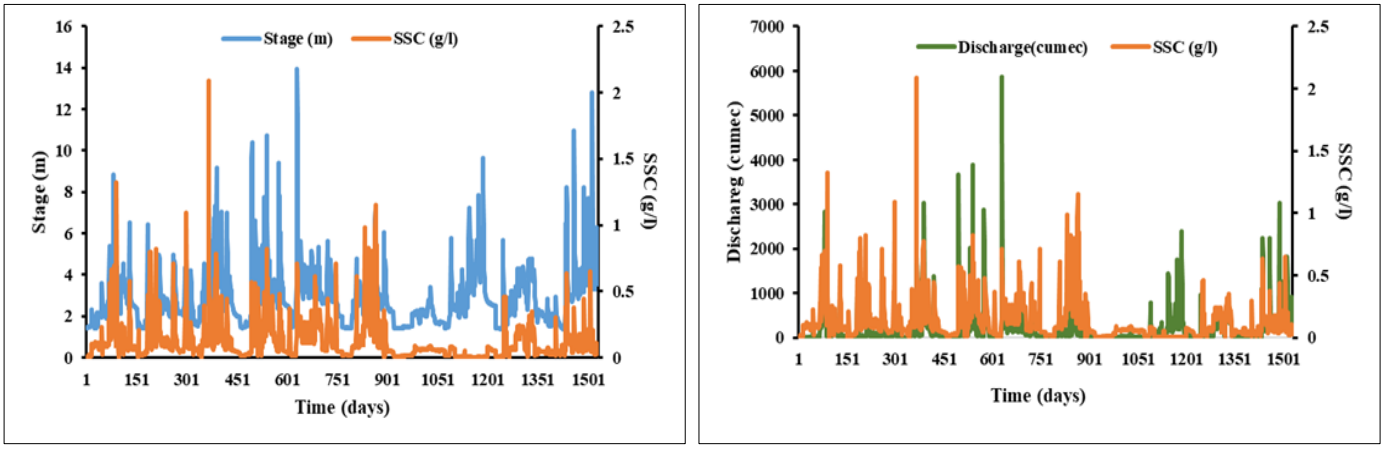

Fig 1: Time series data of stage, discharge and SSC for Adityapur site

Statistical analysis

The statistical analysis of daily stage (m), discharge (cumec), and suspended sediment concentration (SSC, g/l) for Adityapur site (Jharkhand, India) is presented in Table 1. The statistical analysis includes mean, median, minimum, maximum, standard deviation (Std. Dev.), coefficients of variations (C.V.), and skewness values. The skewness values in Table 1 shows that the distribution is highly skewed (Rajaee *et al.*, 2011) [16]. The skewness coefficients for discharge values are more significant, followed by SSC and stage values.

Table 1: Statistical analysis performed for stage (m), discharge (m^3/s), and SSC (g/l) for the Adityapur site

Statistical Parameters	Mean	Median	Minimum	Maximum	Std. Dev.	C.V.	Skewness
Stage	2.911	2.600	1.380	13.950	1.417	8.425	2.189
Discharge	200.2	46.99	1.061	5855.7	422.8	54.22	6.113
SSC	0.123	0.071	0.000	2.085	0.160	23.014	3.648

Support Vector Machine (SVM)

SVM is a recognized technique for classification and regression [40]. Generally, regression-based SVM is called SVR. To solve complex problems effectively, SVR is constructed based on minimizing the structural risk. The insensitive loss function (ε−) is identified as the model tolerating errors up to in the training data. Thus, the SVR ε pursues a linear function as follows:

$$f(x) = w^T\Phi(x_i) + b \quad (1)$$

where w and b represent the coefficients of the weight vector and $\mathbf{\Phi(x_i)}$ is the high dimensional feature space. This can be clarified as the following problem:

$$\text{Minimize: } \frac{1}{2}w^Tw + C\sum_{i=1}^{m}(\xi_i^+ + \xi_i^-)$$

$$\text{Subject to: } \begin{cases} y_i - w^T\Phi(x_i) - b \le \varepsilon + \xi_i^+ \\ w^T\Phi(x_i) + b - y_i \le \varepsilon + \xi_i^- \\ \xi_i^+, \xi_i^- \ge 0 \end{cases} \quad (2)$$

where C > 0 is a penalty parameter which has to be selected earlier. The constant C can grade the experimental error. Moreover, ξ_i^+, ξ_i^-, which are known as slack variables, indicate the distance between real values and the corresponding boundary values of ε-tube. Hence, in order to minimize Equation (2) subject to Equation (3), the function is given by;

$$f(x) = \sum_{i=1}^{m}(\alpha_i^+ - \alpha_i^-)K(x_i, x_j) + b \quad (3)$$

Where $K(x_i, x_j)$ is the kernel function, α_i^+, α_i^- i ≥ 0 are the Lagrange multipliers and b is a bias term. The present study used radial based kernel function and is given by

$$K(x_i, x_j) = \exp\left(-\gamma \left|\left|x_i - x_j\right|\right|^2\right) \quad (4)$$

Where γ is the bandwidth of the kernel function and C, γ, and ε are three predefined parameters.

Wavelet transforms (WT)

Wavelet transform (WT) is a popular time-frequency approach for signal analysis that has more benefits than conventional Fourier analysis (FA) or its "Short-Time Fourier Transform" (Misiti *et al*. 2008). The WT is an advanced variant of the short-term Fourier transformation used in data to track time characteristics (Daubechies 1990, Rioul and Vetterli 1991, Bayazit *et al*., 2001) [3, 17]. It is a variable-sized windowing technique, enabling lengthy time spans for low-frequency information to be used, and shorter regions for high-frequency information. There are two primary wavelet transformation methods: continuous wavelet transformation (CWT) and discrete wavelet transformation (DWT).

$$T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} g^* \left(\frac{t-b}{a}\right) x(t) \cdot dt \quad (5)$$

Where T(a,b) is wavelet coefficient, x(t) is present input signals, g(t) is a base wavelet function, referred to as the mother wavelet, ∗ corresponds to the complex conjugate of g(t), a is a scaling factor stretching or compressing the mother wavelet to a frequency of the signal; and b is a translating factor shifting the mother wavelet to a time domain of the signal. DWT calculates the wavelet coefficients using the simplest and most efficient way, where scales (a) and position (b) are selected based on the powers of 2, called dyadic scales and positions (Mallat 1989) [13]. The discrete wavelet is given in the form of;

$$g_{m,n}(t) = \frac{1}{\sqrt{a_0^m}} g\left(\frac{t - nb_0 a_0^m}{a_0^m}\right) \quad (6)$$

Wavelet coupled support vector machine (WSVM)

The wavelet decomposed inputs timeseries data of current day stage, discharge and previous days stage, discharge and SSC data used as input variables for SVM techniques. The coupling of wavelet decomposed with SVM hybridized to form WSVM. The structure used for WSVM formation for SSC prediction is shown in Fig. 2.

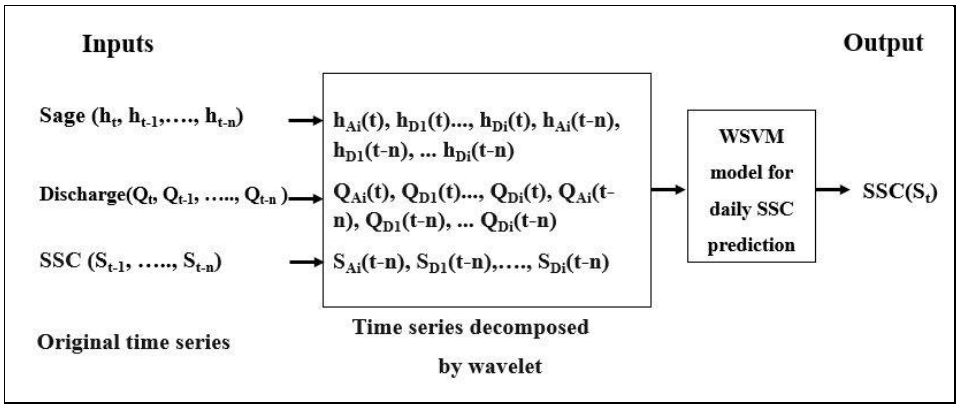

Fig 2: Structure for WSVM model

Gamma test

Gamma test was used to select best input combination. Gamma Test is a flexible and unbiased tool for evaluating the substantial potential of each input parameter. The Gamma test was used for estimating a minimum standard error for continuous nonlinear models for any input-output dataset. A linear regression line is built to measure gamma as:

$$Y = \Delta A + \Gamma \quad (7)$$

Where Y is the regression line's output vector, Δ t is gradient, and Γ t is the regression line's intercept. The value of Γ is corresponding to the output at Δ = 0. The smaller value of Γ (close to zero), is acceptable.

Model development and performance evaluation

SVM and WSVM, and MLR techniques were used for the development and validation of the model. The developed models were evaluated through different performance indicators like root mean square error (RMSE, g/l), Pearson correlation coefficient (PCC) and Wilmot index (WI) and also with visual inspection includes time-series plot and scatter plots. The input variables for the development of SVM and WSVM models were selected using the gamma test. The formula for different performance indicators is given below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (S_{obs,i} - S_{pre,i})^2}{N}} \quad (8)$$

$$PCC = \frac{\sum_{i=1}^{N} (S_{obs,i} - S_{obs,i})(S_{pre,i} - \bar{S}_{pre,i})}{\sqrt{\sum_{i=1}^{N} (S_{obs,i} - \bar{S}_{obs,i})^2 \sum_{i=1}^{N} (S_{pre,i} - \bar{S}_{pre,i})^2}} \quad (9)$$

$$WI = 1 - \frac{\sum_{i=1}^{N} (S_{p_{obs,i}} - S_{p_{pre,i}})^2}{\sum_{i=1}^{N} (|S_{p_{pre,i}} - \bar{S}_{obs,i}| + |\bar{S}_{pre,i} - \bar{S}_{obs,i}|)^2} \quad (10)$$

Where, $S_{obs,i}$ is observed SSC values, $S_{pre,i}$ is the predicted SSC values.

Results and Discussion

Selection of input using gamma test

The selection of input variables before modelling processes is primary processes. The prediction of current suspended sediment concentrations not only depends current day stage (h), discharge (Q) but also depends on previous days stage, discharge and SSC (Lohani *et al*., 2007) [11]. Therefore, present study considers current day (h_t) and previous one day (h_{t-1}) time step for stage, current day discharge (Q_t), previous one day discharge (Q_{t-1}) and previous two-day discharge (Q_{t-2}), previous one SSC (S_{t-1}) and previous two-day SSC (S_{t-2}) for predictions of current day SSC (S_t). The selection of lag days is based on trial-and-error method. The various combinations using these input variables was used to performed gamma test. The gamma test was performed in Wingamma software. The results for various combinations of input variables using gamma test is given in Table 2. The gamma (Γ) values of seven different combinations were obtained in the range of 0.0813 to 0.0915. The variance ratio (V_{ratio}) values obtained in the range of 0.3253 to 0.3660. The selection of best input combination is based on the minimum gamma and V_{ratio} values (Noori *et al*. 2011; Rashidi *et al*. 2016). Therefore, it is apparent from Table 2 that the minimum gamma and V_{ratio} values were 0.0813 and 0.3253, respectively for M-5 model with combination h_t, h_{t-1}, Q_t, Q_{t-1}, Q_{t-2} and S_{t-1}. Therefore, M-5 model combination used as input for prediction of current SSC using support vector machine (SVM) and wavelet-based support vector machine (WSVM).

Table 2: Gamma test results based on different input combinations for study area

Model	Combination	Γ	V_{ratio}
M-1	h_t h_{t-1} Q_t S_{t-1}	0.0833	0.3334
M-2	h_t h_{t-1} Q_t S_{t-1} S_{t-2}	0.0901	0.3604
M-3	h_t h_{t-1} Q_t Q_{t-1} S_{t-1}	0.0911	0.3646
M-4	h_t h_{t-1} Q_t Q_{t-1} S_{t-1} S_{t-2}	0.0915	0.3660
M-5	h_t h_{t-1} Q_t Q_{t-1} Q_{t-2} S_{t-1}	0.0813	0.3253
M-6	h_t h_{t-1} Q_t Q_{t-1} Q_{t-2} S_{t-1} S_{t-2}	0.0904	0.3615

Trial selection

The results on SVM and WSVM model were analyzed in two different phases namely, training phases and testing phase. The training phases includes 70% of the total data and testing phases used rest 30% of the total dataset. The model performance of developed model was evaluated using minimum values of root mean square error (RMSE) while higher values of Pearson correlation coefficient (PCC) and Wilmot index (WI). Several trials were performed for both SVM and WSVM techniques in order to select best model. The radial basis kernel function was used for SVM analysis. The selection of SVM trials is based on SVM-γ, SVM-C, and SVM-e parameters. The cost parameter 'C' value was taken 10 as it varied from 1 to 10 and found to be the same results. The selection of best model considers testing results because during training phases there may chances of over fitting of data.

Quantitative and qualitative evaluation of developed models

The performance of developed model was evaluated through quantitative and qualitative analysis. The quantitative evaluation used RMSE (g/l), PCC and WI for present study. The quantitative evaluation included line diagram and scatter plots. The results of quantitative evaluation of developed model are presented in Table 3. It is apparent from Table 3 that the values of RMSE (g/l) for SVM model found in the range of 0.101 to 0.114 g/l and 0.090 to 0.123 g/l during training and testing phases, respectively. The values for PCC were obtained in the range of 0.773 to 0.833 and 0.426 to 0.597 during training and testing periods, respectively. Further, the values of WI ranged from 0.852 to 0.883 and 0.607 to 0.744, respectively during training and testing phases. Therefore, it is depicted from Table 3 that the trial-3 (γ = 0.1667, ε = 0.1) found to be superior among SVM models with RMSE, PCC and WI values of 0.090 g/l, 0.597 and 0.744, respectively during testing phases. For wavelet based SVM models, the values of RMSE obtained in the range of 0.101 to 0.108 g/l and 0.077 to 0.131 g/l during training and testing phases, respectively. The values of PCC were ranged from 0.810 to 0.834 and 0.299 to 0.639, respectively during training and testing phases. Furthermore, the values of WI were found in the range of 0.855 to 0.880 and 0.527 to 0.767 during training and testing phases, respectively. Therefore, it is apparent from Table 3 that the trial-1 (γ = 0.2, ε = 0.1) found to be superior among all WSVM models with RMSE, PCC and WI values of 0.077 g/l, 0.639 and 0.767 during testing phase, respectively. The comparative evaluation of SVM and WSVM shows the superiority of WSVM model.

Table 3: Performance evaluation of SVM and WSVM models during training and testing phases

Model	Architecture	Training			Testing		
		RMSE	PCC	WI	RMSE	PCC	WI
SVM							
Trial-1	γ = 0.2, ε = 0.1	0.105	0.813	0.876	0.119	0.442	0.620
Trial-2	γ = 0.9, ε = 0.01	0.101	0.833	0.883	0.123	0.426	0.607
Trial-3	γ = 0.1667, ε = 0.1	0.114	0.773	0.852	0.090	0.597	0.744
Trial-4	γ =0.1667, ε =0.01	0.104	0.814	0.879	0.116	0.461	0.640
WSVM							
Trial-1	γ = 0.2, ε = 0.1	0.102	0.831	0.880	0.077	0.639	0.767
Trial-2	γ = 0.9, ε = 0.01	0.102	0.831	0.880	0.095	0.524	0.697
Trial-3	γ = 0.1667, ε = 0.1	0.108	0.810	0.855	0.131	0.299	0.527
Trial-4	γ =0.1667, ε =0.01	0.101	0.834	0.880	0.095	0.524	0.697

The predicted values were plotted with observed values with the help of line diagram and scatter plot as shown in Fig. 3 and Fig. 4 for SVM and WSVM models, respectively. It is clear from Fig. 3 and Fig. 4 that the predicted values are under predicted for larger values of SSC while over-predicted for smaller values of SSC during both SVM and WSVM models. For SVM models, the values for R^2 obtained as 0.5970 and 0.3560, respectively during training and testing phases. For

WSVM models, the values of R^2 during training and testing was observed to be 0.6902 and 0.4078, respectively. Therefore, it is revealed that the WSVM model found to be superior over SVM model.

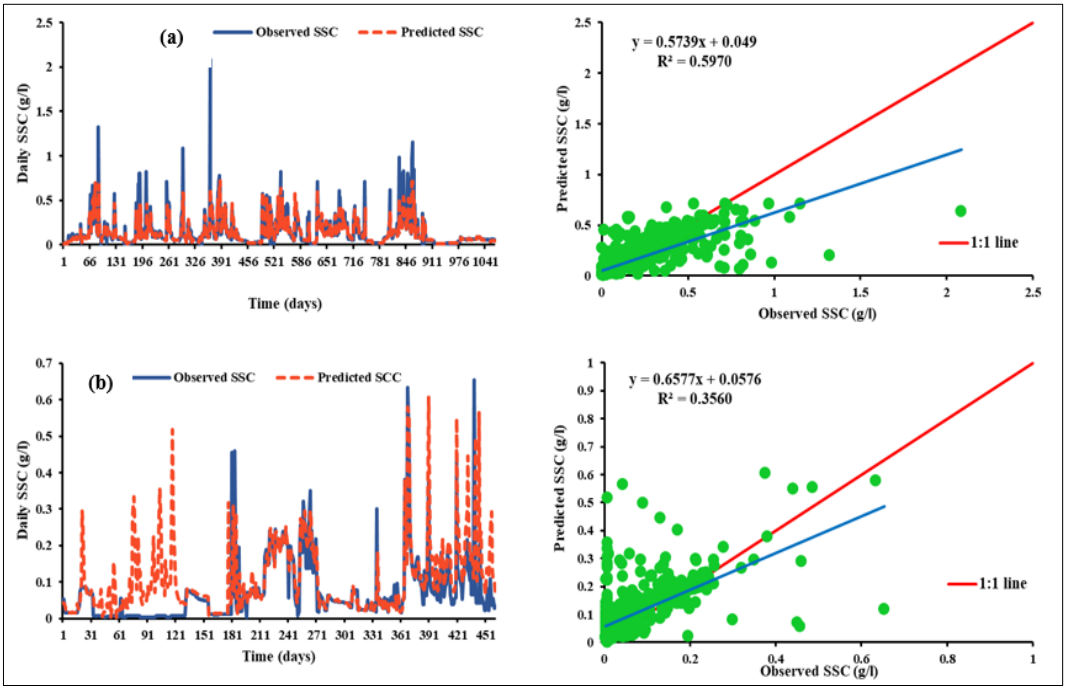

Fig 3: Observed versus predicted SSC using support vector machine (SVM) during (a) training (b) testing phases at study location

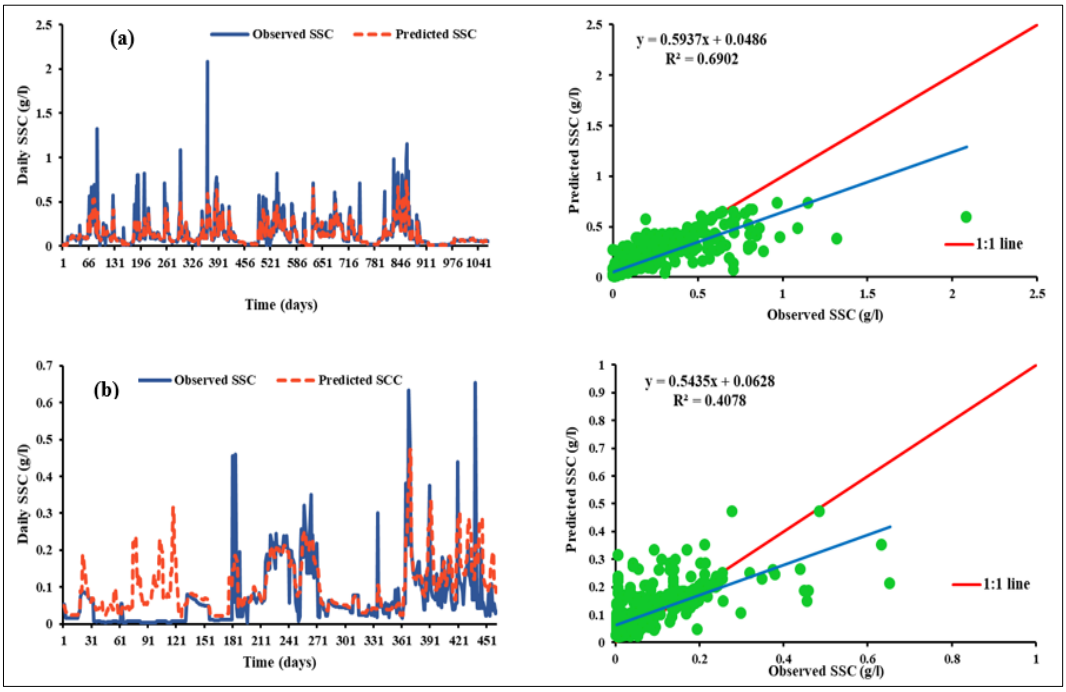

Fig 4: Observed versus predicted SSC using wavelet-based support vector machine (WSVM) during (a) training (b) testing phases at study location

In comparison of both models based on quantitative and qualitative analysis, the WSVM model superior than SVM model. This is because the wavelet transformation decomposes the input data into several frequencies which gathered sufficient information of given data. Therefore, WSVM model has advantages over SVM model to predict daily SSC at Adityapur sites.

Conclusions

The present study was carried out on stage-discharge-sediment modelling using data driven techniques for Adityapur site located in Jharkhand, India. The support vector machine (SVM) and wavelet-based support vector machine (WSVM) were used to predict daily suspended sediment concentration (SSC). The present-day SSC not only depends on present day stage (h), discharge (Q) but also on previous days stage, discharge and SSC. Keeping these facts, various combinations was formed to select best input combination using gamma test. Based on gamma test, the minimum gamma and V_{ratio} values were 0.0813 and 0.3253, respectively for M-5 model with combination h_t, h_{t-1}, Q_t, Q_{t-1}, Q_{t-2} and S_{t-1}. Therefore, M-5 model used as inputs for prediction of current

SSC using SVM and WSVM. Thereafter, SVM and WSVM model were developed based on training (70% of total dataset) and testing (30% of total dataset). The developed models were analyzed through root mean square error (RMSE, g/l), Pearson correlation coefficient (PCC), Wilmott index (WI), line diagram and scatter plots. Based on the quantitative and qualitative evaluation, the results of trial-3 (γ = 0.1667, ε = 0.1) found to be superior among SVM models with R^2, RMSE, PCC and WI values of 0.3560, 0.090 g/l, 0.597 and 0.744, respectively during testing phases. Among results of WSVM models, the trial-1 (γ = 0.2, ε = 0.1) found to be superior among all WSVM models with R^2, RMSE, PCC and WI values of 0.4078, 0.077 g/l, 0.639 and 0.767 during testing phase, respectively. The comparative model analysis showed that the WSVM model found superior than SVM model. This is because the wavelet transformation decomposes the input data into several frequencies which gathered sufficient information of given data. Therefore, WSVM model has advantages over SVM model to predict daily SSC at Adityapur sites.

References

1. Asselman NEM. Fitting and interpretation of sediment rating curves. Journal of Hydrology 2000;234(3-4):228-248.
2. Cannas B, Fanni A, See L, Sias G. Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. Physics and Chemistry of the Earth, Parts A/B/C 2006;31(18):1164-1171.
3. Daubechies I. The wavelet transform, time-frequency localization and signal analysis. IEEE transactions on information theory 1990;36(5):961-1005.
4. Jain SK. Development of integrated sediment rating curves using ANNs. Journal of hydraulic engineering 2001;127(1):30-37.
5. Kalteh AM. Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. Computers & Geosciences 2013;54:1-8.
6. Kisi O. Suspended sediment estimation using neuro-fuzzy and neural network approaches/Estimation des matières en suspension par des approches neurofloues et à base de réseau de neurones. Hydrological Sciences Journal 2005, 50(4).
7. Kumar M, Kumari A, Kushwaha DP, Kumar P, Malik A, Ali R *et al*., Estimation of Daily Stage–Discharge Relationship by Using Data-Driven Techniques of a Perennial River, India. Sustainability 2020;12(19):7877.
8. Kumar M, Kumar P. Daily Suspended-sediment Concentration simulation using ANN and Wavelet ANN models. International Archive of Applied Sciences and Technology 2020;11(3):60-69.
9. Kumari A, Kumar M, Kumar A, Amitabh A. Daily Gauge-Discharge Estimation using ANN and Wavelet-ANN Models for Muri Station, Jharkhand. Indian Journal of Ecology 2020;47(3):645-649.
10. Liong SY, Sivapragasam C. Flood stage forecasting with support vector machines 1. JAWRA Journal of the American Water Resources Association 2002;38(1):173-186.
11. Lohani AK, Goel NK, Bhatia KS. Deriving stage-discharge-sediment concentration relationships using fuzzy logic. Hydrological Sciences Journal 2007;52(4):793-807.
12. Londhe S, Gavraskar SS. Forecasting one day ahead stream flow using support vector regression. Aquatic Procedia 2015;4:900-907.
13. Mallat SG. A wavelet tour of signal processing, Academic Press, San Diego 1998, 557.
14. Nourani V, Alami MT, Aminfar MH. A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Engineering Applications of Artificial Intelligence 2009;22(3):466-472.
15. Phillips JM, Webb BW, Walling DE, Leeks GJL. Estimating the suspended sediment loads of rivers in the LOIS study area using infrequent samples. Hydrological processes 1999;13(7):1035-1050.
16. Rajaee T, Nourani V, Zounemat-Kermani M, Kisi O. River suspended sediment load prediction: application of ANN and wavelet conjunction model. Journal of Hydrologic Engineering 2011;16(8):613-627.
17. Rioul O, Vetterli M. Wavelets and signal processing. IEEE signal processing magazine 1991;8(4):14-38.
18. Singh KP, Durgunoglu A. Developing accurate and reliable stream sediment yields. IAHS Publication 1989;184:193-99.
19. Ünal NE, Aksoy H, Akar T. Annual and monthly rainfall data generation schemes. Stochastic Environmental Research and Risk Assessment 2004;18(4):245-257.
20. Yu X, Liong SY, Babovic V. EC-SVM approach for real-time hydrologic forecasting. Journal of Hydroinformatics 2004;6(3):209-223.