

Modelling runoff in an arid watershed through integrated support vector machine

Sandeep Samantaray and Dillip K. Ghose*

Department of Civil Engineering, NIT Silchar, Assam, India

*Corresponding author. E-mail: dillipghose_2002@yahoo.co.in

Abstract

Modelling of runoff is a significant practice in water resources engineering. Therefore, discovering consistent and advanced methods for prediction of runoff is crucial for hydrologic processes. Here, a narrative integrated intelligence model attached with PSR (phase space reconstruction) is anticipated to estimate runoff for five watersheds of Balangir, Odisha, India. Monthly monsoon precipitation, temperature, humidity data of five watersheds over 28 years (1990–2017) are employed and validated. Here, the proposed model is an integration of support vector machine (SVM) with firefly algorithm (FFA) and PSR. Various indices such as NSE (Nash–Sutcliffe), RMSE (root mean square error) and WI (Willmott's index) are used to find the performance of the model. The developed PSR-SVM-FFA model demonstrates pre-eminent WI value ranging from 0.97 to 0.98 while the SVM and SVM-FFA models encompass 0.92 to 0.93 and 0.94 to 0.95, respectively. Also, an assessment of data from the suggested model is schemed and validated. The proposed PSR-SVM-FFA model gives better accuracy results and error limiting up to 2–3%.

Key words: hydrology, PSR-SVM-FFA, runoff, SVM, SVM-FFA, watershed

Highlights

- The most relevant techniques are used for performing the output of rainfall-runoff process at a single outlet. The advances gained from a modelling point of view are considered and discussed.
- The models developed using three techniques are support vector machine (SVM), support vector machine with firefly algorithm (SVM-FFA) and support vector machine and firefly algorithm with Phase space reconstruction (PSR-SVM-FFA) in the study.
- The study is different from previous studies by including and comparing two new techniques, SVM-FFA and PSR-SVM-FFA, while past research has used hybridization of SVM for predicting runoff in a watershed.
- Interaction of all techniques along with five different models is presented.
- The model proposed reasonably predicts the runoff at a single outlet and improves the method of measuring runoff without the detailed study for all sub-outlets of the watershed.

INTRODUCTION

Improvement in water resources is essential for balancing water demand around the globe. Due to the complex nature and uncertainty in water parameters, traditional methods like statistical method,

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

Dickens formula (1865), Ryves formula (1984), and Khosla formula (1960) are insufficient to understand the cycle of rainfall-runoff (Das 2009). Innovative computational algorithms and their application to water resources engineering are necessary to understand the behaviour of the water cycle in the context of climate change. Computational algorithms in view of modern techniques like FFA, SVM and PSR are important for engineering applications. Attempts have also been made to apply these techniques to water resources engineering (Tayfur 2014).

Tabari *et al.* (2012) employed the SVM, ANFIS (adaptive neuro-fuzzy inference system), MLR (multiple linear regression) method utilizing six climate data variables as inputs to estimate evapotranspiration at Nozheh station, Iran. Outcomes suggested that SVM and ANFIS models perform better with greater precision than those of regression and climate-based models. Baydaroglu & Kocak (2014) applied support vector regression (SVR) to envisage water loss using solar radiation, relative humidity, wind speed, evaporation and temperature as input. SVR helps in predicting evaporation effectively since it has a good generalization ability with greater determination coefficients of 83 and 97% for univariate and multivariate time series entrenching. Raghavendra & Deka (2014) in their review found that SVM is more suitable for prediction than other techniques pertinent to hydrology. Cha *et al.* (2014) projected the SVM-FFA model for more precise forecast of malaria incidences in the Jodhpur and Bikaner area. Performance of the projected hybrid model was compared with ANN, autoregressive moving average as well as SVM and results indicated that the proposed integrated model provides forecasts with better accuracy as compared to traditional methods. The FNN (feed forward neural network) technique was used for monthly stream flow forecasting for a period of 53 years in the United States by Vignesh *et al.* (2015). Obtained outcomes are valuable in identifying suitable model intricacy at discrete stations, arrangements across regions and sub-regions, interpolating and extrapolating data, and classifying catchments. SVM-FFA and SVM-wavelet were used to evaluate precipitation trends in 29 meteorological stations in Serbia from 1946 to 2012 by Gocic *et al.* (2016). Estimation and prediction outcomes of hybrid models were contrasted and results showed that the SVM-wavelet approach gives improved prediction accuracy and ability to generalize. Moghaddam *et al.* (2016) implemented SVM-FFA for predicting the exhaustion lifetime of polyethylene terephthalate (PET) reformed asphalt mixture taking PET percentages, stress levels and environmental temperatures as inputs. Prediction results by SVM-FFA were compared against ANN and genetic programming (GP) and it was found that the hybrid model performs better. Al-Shammari *et al.* (2016) projected a hybrid method integrating SVM with FFA for predicting daily dew point temperature (T_{dew}) for an Iranian city and performance of the hybrid model is compared to SVM, ANN and GP. They observed that SVM-FFA is certainly very efficient in predicting T_{dew} with better accuracy and consistency. Integrated PSR-ANN and ANN techniques were utilized to predict daily river flow on river gauging stations in the USA by Delafrouz *et al.* (2017). The outcomes of PSR-ANN, ANN and gene expression programming models were compared and it was found that the projected hybrid model gives the best prediction accuracy for daily river flow. Mehr *et al.* (2018) employed integration of FFA and SVM for rainfall forecasting on a monthly basis at Tabriz and Urmia stations, Iran. The effectiveness of the hybrid model was cross-checked with SVR and GP-based forecasting models and it was found that the SVM-FFA model performs well with promising accuracy in terms of rainfall forecasting. Ghose & Samantaray (2019) employed LRNN (layer recurrent NN), RBFN (radial basis function NN) and FFBPN (feed forward back propagation NN) to evaluate runoff as a loss function of evapotranspiration, temperature and precipitation. Outcomes revealed that performance of the LRNN model is best as compared to FFBPN and RBFN for predicting runoff in the watershed which helps in planning, designing and managing hydraulic structures in the neighbourhood of the watershed. Zaini *et al.* (2018) used an SVM model and also a hybridized SVM-PSO (particle swarm optimization) model to scheme daily stream flow at Upper Bertam watershed, Malaysia. Hybridized models SVM-PSO1 and SVM-PSO2 perform best in comparison to SVM1 and SVM2 at forecasting river flow 1–7 days ahead. Tao *et al.* (2018) used SVM and SVM-FFA to anticipate rainfall

at Chhattisgarh. The proposed hybrid model significantly improved forecasting accuracy and can also be used for monthly rainfall forecasting in provincial areas of India. Gandomi *et al.* (2011) utilized FFA to solve mixed variable structural optimization problems and also the implications of FFA are thoroughly analysed in comparison with PSO, GA, simulated annealing for future research. Ji & Sun (2013) presented a method for multitasking multiple classes of SVM on the basis of minimizing regularization functions and found it to be best among other multitask learning methods. Shamshirband *et al.* (2015) proposed a hybrid SVM-FFA model for estimating monthly mean horizontal global solar radiation (HGSR). Performance of the hybrid model is evaluated by comparing it with ANN, GP and ARMA (autoregressive moving average) models and results revealed that the projected hybrid model performs best.

Several researchers have found different individual algorithms for predicting runoff. Some have also developed optimization methods and least integrated methods to predict runoff. Thus, an attempt has been made to think about the hybrid model via integral algorithms. The objective of this research is to compare results of the SVM-FFA, PSR-SVM-FFA model with the SVM and empirical models to tune up the magnitude of runoff for controlling flood in the region. Basically, it is applicable for efficient planning and management of water resources such as flood control and management of a watershed.

STUDY AREA

Balangir district is located in the west part of Odisha, India, with an area of 5,165 km², as shown in Figure 1. It lies between the latitudes of 20°59'00" and longitudes of 83°32'22". It has an average elevation of 115 metres (377 feet). The watershed is located towards the mid-north edge of Balangir district. Five gauging stations, Loisingha, Balangir, Tushura, Saintala and Patnagarh, are considered in the present research.

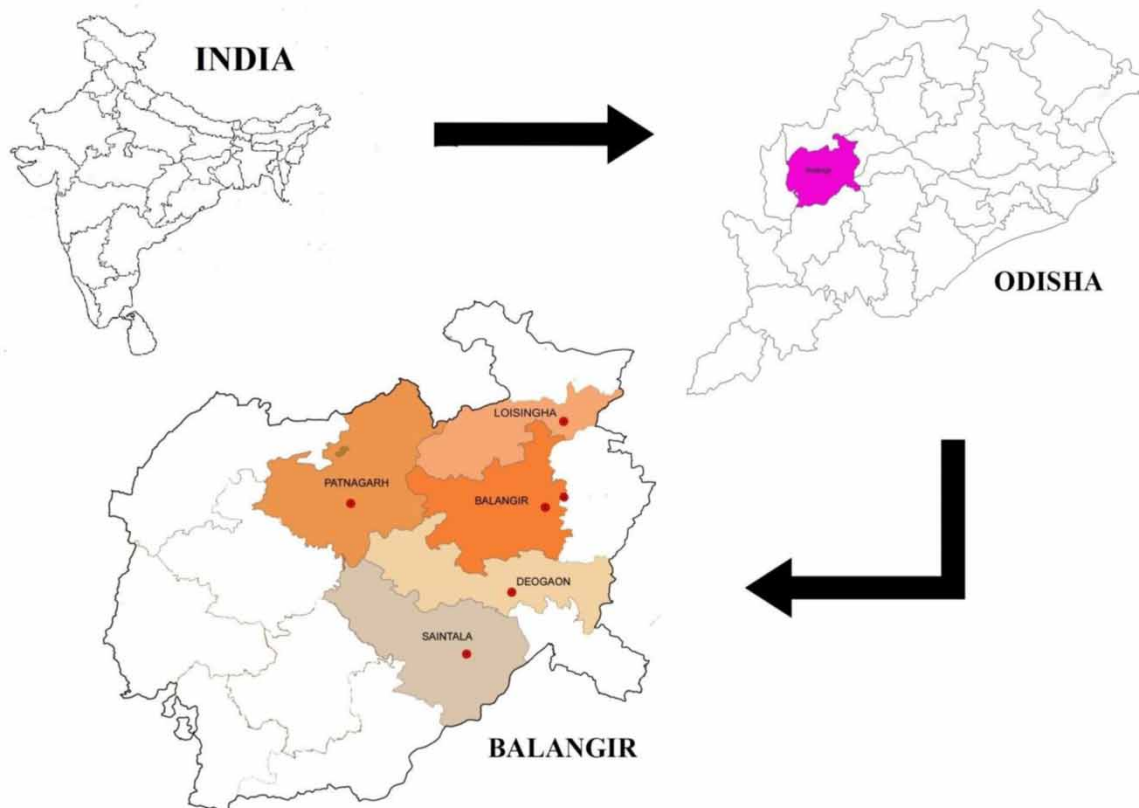


Figure 1 | Schematic diagram of the proposed watershed.

Physical and statistical characteristics of the gauging stations are described in Table 1. Here, prediction of runoff is assessed by taking the dataset from 1990 to 2017. Mean monthly precipitation and temperature data for the monsoon months (May to October) from the period 1990–2017, spanning over 28 years, are obtained from IMD (India Meteorological Department) Bhubaneswar. Runoff data are collected from the soil conservation office, Bolangir. Here, the runoff data are computed using Khosla's empirical equation to understand the coherence between developed empirical data and observed data collected from that department.

Table 1 | Physical and statistical characteristics of five stations in Balangir, Odisha, India

Station name	Drainage area (km ²)	Latitude	Longitude	Elevation (m)	Mean rainfall (mm)	Min.–Max. rainfall (mm)
Loisinga	317.6	20.86N	83.50E	162	91.22	7.83–158.8
Balangir	386.98	20.7N	83.5E	115	87.82	7.93–157.29
Deogaon	432.01	20.71N	83.49E	148	93.69	7.28–155.26
Saintala	454.43	20.26N	83.20E	182	87.795	7.45–159.56
Patnagarh	591.45	20.72N	83.15E	243	87.23	7.89–151.34

METHODOLOGY

Khosla formula

Khosla considered precipitation, discharge and temperature data for different catchments in India and the USA to arrive at a pragmatic connection as follows in Equation (1). It is circuitously anchored in the water-balance perception and mean monthly temperature to replicate losses due to evapotranspiration:

$$R_m = P_m - L_m \quad (1)$$

$$L_m = 0.48 T_m, \text{ for } T_m > 4.5^\circ\text{C}$$

where R_m is monthly runoff, P_m is monthly rainfall, L_m is monthly losses and T_m is mean monthly temperature of the catchment.

Support vector machine model

A technique with statistical learning theory and promising empirical results for recognition of pattern with high dimension data to avoid error with decision boundary using training in terms of support vector is known as a support vector machine (Cortes & Vapnik 1995).

A graphical description of the SVM model is shown in Figure 2. The principles of SVM include statistical learning known as structural risk minimization, which affords a higher boundary for generalization error of a classifier (R) concerning training error (R_e), model complexity (h) and number of training examples (N) to evaluate probability of error and is defined as (Chen *et al.* 2018; Thanh & Kappas 2018; Wang *et al.* 2018a):

$$R \leq R_e + \varphi\left(\frac{h}{N}, \frac{\log \eta}{N}\right) \quad (2)$$

where φ is monotonic increasing function of the model complexity.

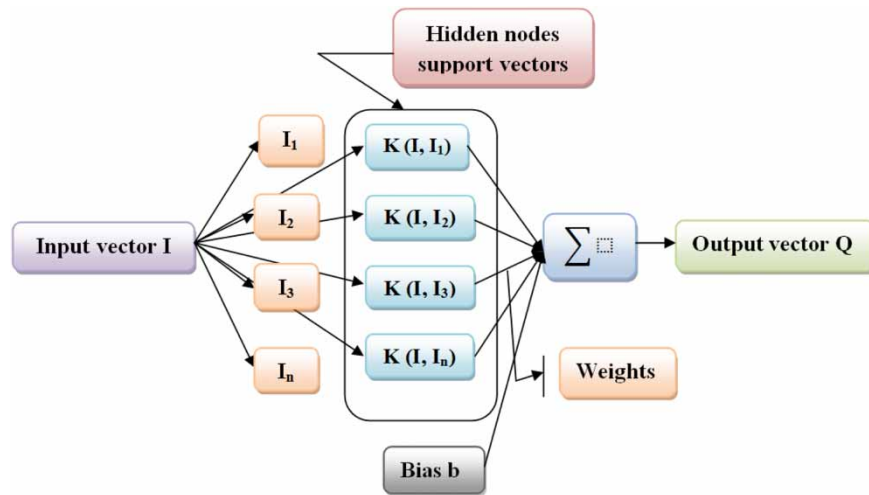


Figure 2 | Graphical description of SVM model.

This is enviable when designing a linear classifier which maximizes margins of their decision boundaries to minimize error. In this regard, SVM is an effective linear classifier which searches a hyper plane with the largest margin called the maximal marginal classifier. The decision boundary of the linear classifier is:

$$w \times x + b = 0 \quad (3)$$

The training phase of SVM involves estimation of constraints weight (w), bias (b) of decision boundary from the training data set. Parameters are preferred with the below conditions:

$$\begin{aligned} w \times x_i + b &\geq 1, \text{ if } y_i = 1 \\ w \times x_i + b &\leq -1, \text{ if } y_i = -1 \end{aligned} \quad (4)$$

The objective function is the Lagrangian multiplier and denoted as:

$$L_p = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \lambda_i \{y_i (w \times x_i + b) - 1\} \quad (5)$$

The linear programming is to be utilized through Karush–Kuhn–Tucker (KKT) conditions and the KKT condition is:

$$\begin{aligned} \lambda_i &\geq 0 \\ \lambda_i \{y_i (w \times x_i + b) - 1\} &= 0 \end{aligned} \quad (6)$$

Integration of SVM with FFA model fitness

The FFA is a natural algorithm based on the blinking manners of fireflies (Yang 2009; Abd-Elazim & Ali 2018; Lieu *et al.* 2018; Wang *et al.* 2018b, 2018c). Since the complex nature of fireflies induces global communication among swarming particles, effective multi-objective optimization is necessary to recognize the multifaceted algorithm of FFA. In this algorithm, fireflies use flashing light to search for mates, to attract potential prey for protection from their predators. In order to achieve efficient optimal solutions, flashing light intensity induces fireflies to a more attractive position associated with objective function of the problem.

There are some components in FFA, such as distance, attractiveness and movement.

Euclidian distance

Euclidian distance is the distance between two fireflies i and j at locations u_i and u_j , defined as Euclidean distance (r_{ij}) applying Equation (7), where $u_{i,k}$ is the k^{th} factor of spatial coordinate u_i of i^{th} firefly and d is dimension number:

$$r_{ij} = \|u_i - u_j\| = \sqrt{\sum_{k=1}^d (u_{i,k} - u_{j,k})^2} \quad (7)$$

Attractiveness

Attractiveness task of a firefly is revealed in Equation (8):

$$\beta_r = \beta_0 \times \exp(-\gamma r^m), \quad m \geq 1 \quad (8)$$

Movement

Progress of firefly i is towards an extra striking firefly j and is specified by Equation (9):

$$\text{updated } u_i = u_i + \beta_0 \times \exp(\gamma r_{ij}^2) \times (u_j - u_i) + \alpha \left(r - \frac{1}{2} \right) \quad (9)$$

where u_i is the current position of firefly, $\beta_0 \times \exp(\gamma r_{ij}^2) \times (u_j - u_i)$ is active for allowing a firefly's attractiveness to light intensity and $\alpha \left(r - \frac{1}{2} \right)$ is applied for random movement of a firefly.

The settings of firefly algorithm (FA) parameters: γ (light absorption coefficient) = 1.0, α (randomization parameter) = 0.3, β_0 (attractiveness value) = 1.0 and random number generator consistently disseminated in space $[0, 1] = 0.2$. A flowchart of the SVM-FFA model is shown in Figure 3.

Objective function

```

F(Y), Y = (Y1, Y2, Y3, ..... , Yk)Z
Fireflies' initialization, Yj (j = 1, 2, 3, 4, ..... , m)
Concentration of light Lj at Yj, f(Yj)
Light preoccupation coefficient β
While z < maximum generation do
    For j = 1:m all m fireflies do
        For i = 1:i all m fireflies do
            If Li > Lj then
                Transfer firefly j towards i in k dimension
            End if
            Attractiveness diverges with distance r through exp[-βr]
            Calculate high result and update intensity of light
        End for
    End for

```

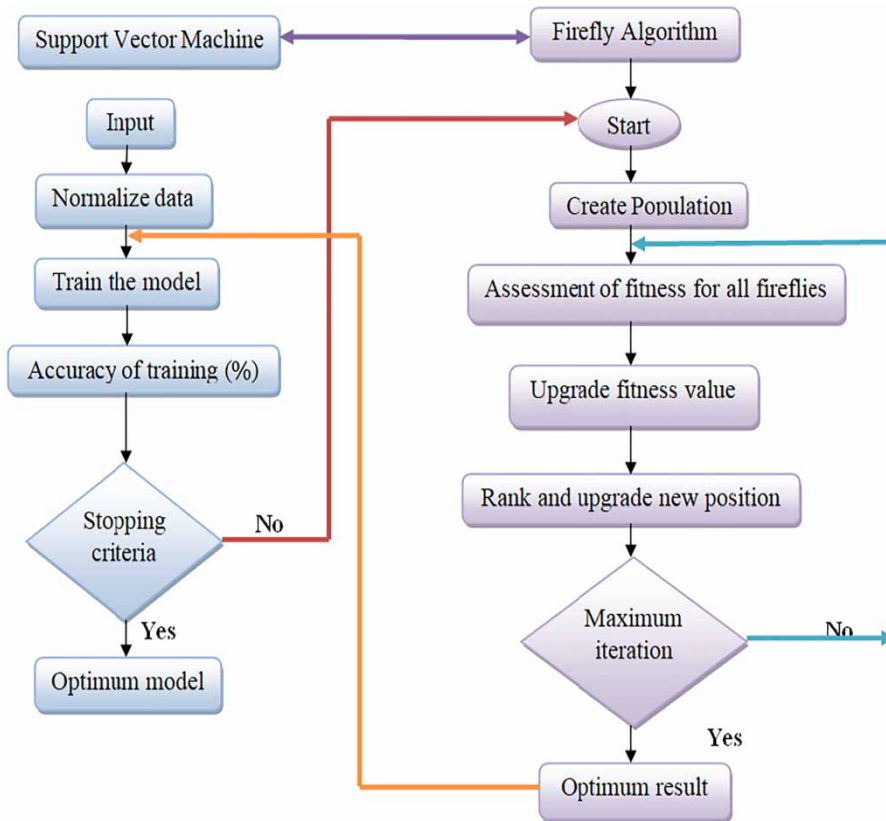



Figure 3 | Descriptive flowchart of SVM-FFA model.

Evaluate best value
End while
Final result and conception

Integration of PSR coupled with SVM-FFA model

Illustration of hybrid extravagance model is demonstrated in Figure 4. The models involve the subsequent modelling process (Dutta *et al.* 2018; Fan *et al.* 2018; Hajiloo *et al.* 2018; Sun & Wang 2018).

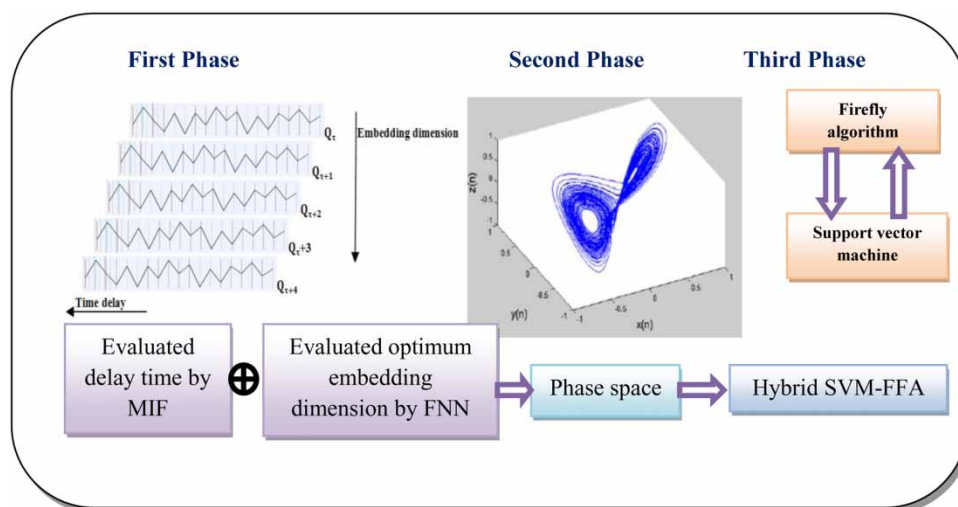


Figure 4 | Graphical representation of hybrid PSR-SVM-FFA.

- (i) First phase, optimal τ and m is established to accomplish MIF and FFN approaches for embattled index phase space reconstruction. In this manner, phase space will converge from input aspects and abstract critical dynamics of chaotic time series.
- (ii) Second phase, input matrix is assembled rooted in several dimensions $X_t, X_{t-\tau}, X_{t-2\tau}, \dots, X_{t-(m-1)\tau}$ by considering optimal τ and m , for targeted value X_{t+T} . According to the PSR technique, an input matrix is proposed (Sivakumar *et al.* 2001).
- (iii) Third phase, the hybrid SVM-FFA model is fed by the produced matrix using phase space signal.

Objective function

For univariate time signal $X_1 = (X_{1,1}, X_{1,2}, \dots, X_{1,N})^T$, PSR is

$$X'_{1i} = (x_{1,i}, x_{1,i+\tau_1}, \dots, x_{1,i+(m_1-1)\tau_1}), i = 1, 2, \dots, N - (m_1 - 1)\tau_1$$

where, N is time series, τ_1 is lag and m_1 is embedding dimension for X_1 .

For multivariate time series,

$$X = [X_1, X_2, X_3, \dots, X_s] = \begin{bmatrix} x_{1,1} & \dots & x_{s,i} \\ \vdots & & \vdots \\ x_{1,N} & \dots & x_{s,N} \end{bmatrix}$$

PSR computes the reconstruction for each time series:

$$X'_1 = (X'_{1,i}, X'_{2,i}, X'_{3,i}, \dots, X'_{s,i}) i = 1, 2, 3, \dots, N - (\max\{m_i\} - 1)(\max\{\tau_i\})$$

s: measurement number.

Delay estimation

For delay, PSR allows average mutual information (AMI).

For reconstruction, time delay is first local lowest AMI.

$$AMI \text{ is computed as: } AMI(T) = \sum_{i=1}^N p(x_i, x_{i+T}) \log_2 \left[\frac{p(x_i, x_{i+T})}{p(x_i)p(x_{i+T})} \right]$$

T = 1 [for max. lag].

Embedding dimension (ED) estimation

ED for PSR is appraised utilizing the false nearest neighbour (FNN) algorithm. For I point at d dimension, points X_i^r and adjacent point X_i^{r*} in reconstructed phase space $\{X_i^r\}$, $i = 1:N$, are false neighbours

if $\sqrt{\frac{R_i^2(d+1) - R_1^2(d)}{R_1^2(d)}} > \text{distance threshold}$ where, $R_1^2(d) = \|X_i^r - X_i^{r*}\|^2$ is distance metric, d is smallest value; $P_{fnn} < \text{per cent false neighbours}$ and P_{fnn} is ratio of FNN points to total PSR quantity.

Processing and preparation of data

Humidity, temperature and monthly average rainfall are composed of data from the meteorological department of India for a season of monsoon months (May to October) for 1990–2017. Data from

Table 2 | Input constraints for SVM, SVM-FFA model

Input	Model name (SVM)	Model name (SVM-FFA)
Q_{t-1}	SVM1	SVMI-FFA
Q_{t-1}, Q_{t-2}	SVM2	SVM2-FFA
$Q_{t-1}, Q_{t-2}, Q_{t-3}$	SVM3	SVM3-FFA
$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}$	SVM4	SVM4-FFA
$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}$	SVM5	SVM5-FFA

where, Q_{t-1} is one-month lag runoff, Q_{t-2} is two-month lag runoff, Q_{t-3} is three-month lag runoff, Q_{t-4} is four-month lag runoff and Q_{t-5} is five-month lag runoff.

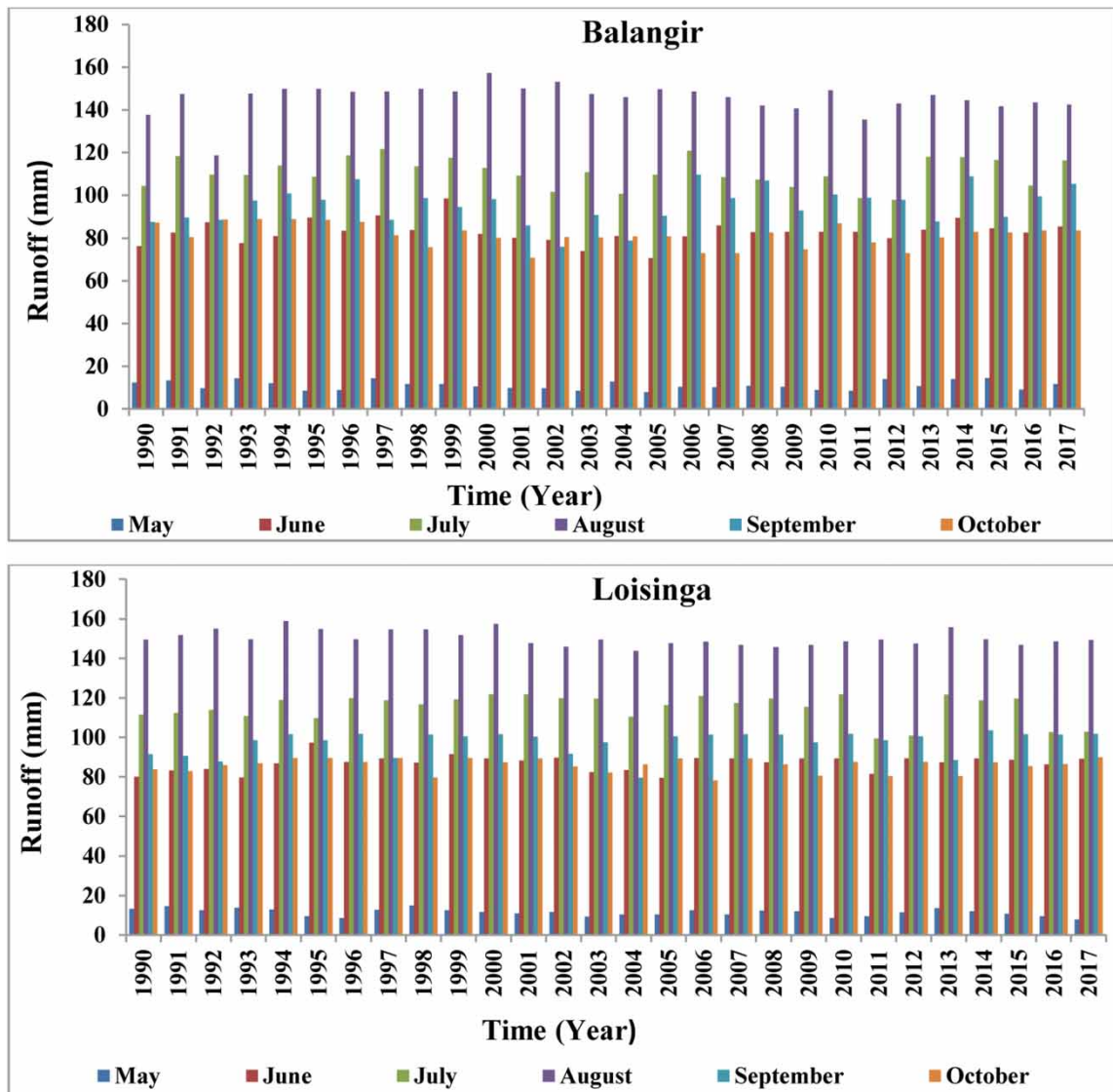


Figure 5 | Estimated mean monthly runoff of monsoon season for proposed watershed. (Continued.)

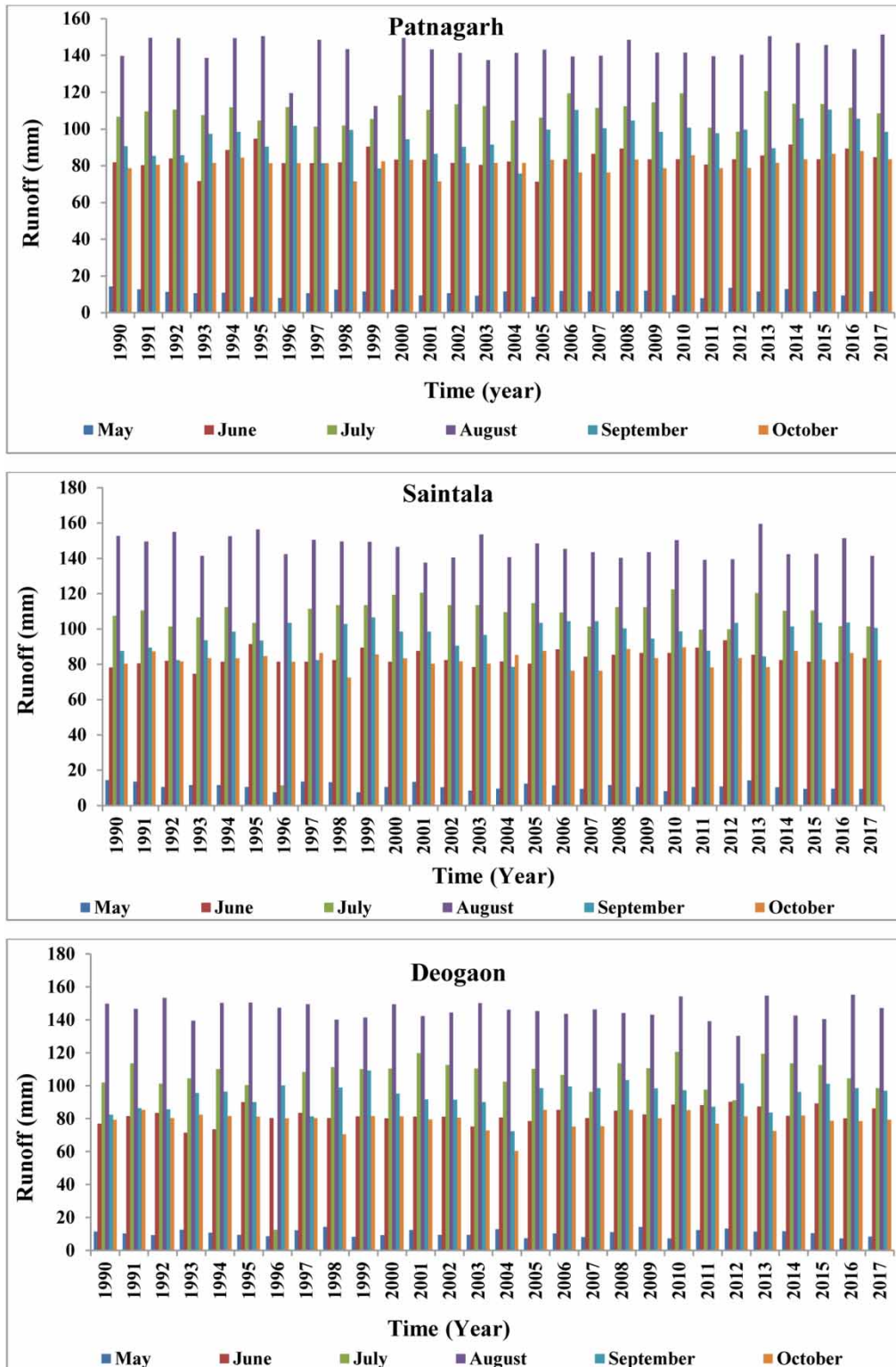


Figure 5 | Continued.

1990 to 2009 are employed for training and from 2010 to 2017 are considered for testing the model. Daily data are transformed into monthly data, which finally helps in training and testing the model. Subsequent runoff arrangements are employed as input (Table 2).

Model performance evaluation

The indicators NSE, RMSE and WI are used to assess the performance of model efficiency. The formulae can be articulated as:

$$NSE = 1 - \left[\frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O}_i)^2} \right] - \infty < NS \leq 1 \quad (10)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2} \quad (11)$$

$$WI = 1 - \left[\frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (|P_i - \bar{O}_i| + |O_i - \bar{O}_i|)^2} \right] \quad (12)$$

where, O_i and P_i are observed and predicted i^{th} runoff and \bar{O}_i is mean of observed runoff.

RESULTS AND DISCUSSION

Runoff is estimated by using the Khosla formula for five watersheds in the monsoon period (May–October) for 1990–2017. A graphical representation of mean monthly runoff is presented in Figure 5.

The performances of the SVM model with different inputs for the five proposed watersheds are presented in Table 3. Three evaluating parameters, NSE, RMSE and WI, are estimated for both testing and training phases as explained below.

Table 3 | Results of SVM model for runoff prediction

Positions	Model	Training period			Testing period		
		NSE	RMSE	WI	NSE	RMSE	WI
Balangir	SVM1	0.513	91.93	0.841	0.486	88.93	0.849
	SVM2	0.548	87.17	0.857	0.509	85.48	0.864
	SVM3	0.597	84.86	0.874	0.552	81.39	0.881
	SVM4	0.642	80.05	0.869	0.601	77.91	0.886
	SVM5	0.681	76.98	0.935	0.653	73.92	0.942
Loisinga	SVM1	0.497	93.78	0.826	0.451	91.95	0.845
	SVM2	0.518	89.42	0.831	0.486	88.12	0.857
	SVM3	0.561	85.18	0.862	0.537	83.85	0.884
	SVM4	0.598	77.04	0.886	0.552	75.93	0.905
	SVM5	0.636	72.73	0.919	0.601	70.21	0.925
Patnagarh	SVM1	0.505	92.75	0.818	0.478	90.68	0.832
	SVM2	0.531	88.51	0.835	0.507	85.08	0.856
	SVM3	0.574	84.94	0.862	0.541	80.59	0.881
	SVM4	0.605	81.62	0.881	0.576	78.38	0.907
	SVM5	0.648	76.83	0.927	0.613	73.41	0.931
Saintala	SVM1	0.527	92.85	0.795	0.485	90.03	0.821
	SVM2	0.551	88.22	0.829	0.517	86.62	0.857
	SVM3	0.605	83.94	0.857	0.554	81.79	0.882
	SVM4	0.663	79.05	0.874	0.617	75.91	0.914
	SVM5	0.692	73.82	0.938	0.658	69.73	0.947
Deogaon	SVM1	0.518	92.65	0.805	0.496	90.52	0.831
	SVM2	0.547	88.91	0.826	0.501	86.01	0.852
	SVM3	0.582	85.57	0.853	0.549	82.59	0.886
	SVM4	0.621	81.48	0.882	0.584	78.71	0.918
	SVM5	0.683	77.73	0.945	0.648	74.05	0.953

Bold font indicates best value of model performance.

Among the five models, the SVM5 model shows the best result with WI 0.935 and 0.942 for training and testing period, respectively, while Q_{t-1} , Q_{t-2} , Q_{t-3} , Q_{t-4} , Q_{t-5} are considered as input parameters for Balangir watershed. For the remaining watersheds, model SVM5 gives best WI value for both phases out of five simulations. In the case of the Loisinga watershed, the pre-eminent WI values for training and testing periods are 0.919 and 0.925 for model SVM5. For the training phase at Patnagarh, Saintala and Deogaon watersheds, principal WI values are 0.927, 0.938 and 0.945, respectively. The result for the SVM-FFA model based on NSE, RMSE and WI for testing and training periods is given in Table 4.

Table 4 | Outcomes of SVM-FFA model for runoff prediction

Positions	Model	Training period			Testing period		
		NSE	RMSE	WI	NSE	RMSE	WI
Balangir	SVM1-FFA	0.694	89.71	0.846	0.673	86.48	0.859
	SVM2-FFA	0.717	86.04	0.862	0.698	82.71	0.864
	SVM3-FFA	0.759	80.54	0.899	0.736	79.03	0.903
	SVM4-FFA	0.782	76.82	0.926	0.758	75.66	0.941
	SVM5-FFA	0.815	74.34	0.965	0.791	70.72	0.971
Loisinga	SVM1-FFA	0.681	90.55	0.835	0.662	90.55	0.841
	SVM2-FFA	0.708	86.81	0.847	0.689	86.27	0.865
	SVM3-FFA	0.746	82.59	0.892	0.716	80.91	0.901
	SVM4-FFA	0.772	76.68	0.926	0.751	73.78	0.937
	SVM5-FFA	0.807	70.39	0.958	0.783	68.49	0.961
Patnagarh	SVM1-FFA	0.679	91.84	0.818	0.649	88.81	0.825
	SVM2-FFA	0.692	87.96	0.858	0.662	83.96	0.864
	SVM3-FFA	0.737	83.68	0.885	0.709	78.78	0.899
	SVM4-FFA	0.769	78.17	0.929	0.726	75.93	0.937
	SVM5-FFA	0.791	74.09	0.953	0.754	70.07	0.962
Saintala	SVM1-FFA	0.688	90.38	0.825	0.648	88.81	0.836
	SVM2-FFA	0.711	86.17	0.859	0.672	84.35	0.868
	SVM3-FFA	0.745	81.04	0.907	0.705	79.16	0.913
	SVM4-FFA	0.781	77.91	0.924	0.751	73.49	0.935
	SVM5-FFA	0.813	70.55	0.955	0.793	67.88	0.966
Deogaon	SVM1-FFA	0.693	90.01	0.848	0.654	89.28	0.855
	SVM2-FFA	0.713	85.63	0.869	0.689	84.96	0.885
	SVM3-FFA	0.759	83.86	0.914	0.723	80.81	0.929
	SVM4-FFA	0.792	79.47	0.939	0.751	76.67	0.946
	SVM5-FFA	0.828	74.28	0.961	0.779	70.11	0.974

Here five different models are used to estimate NSE, RMSE and WI value for five projected watersheds. When Q_{t-1} , Q_{t-2} , Q_{t-3} , Q_{t-4} , Q_{t-5} is used as input scenario, WI value provides the best result. Considering Balangir, the best values for WI are 0.965 and 0.971 for the training and testing periods, respectively. For the training phase, the WI values for Loisinga, Patnagarh, Saintala, Deogaon watershed are 0.958, 0.953, 0.955, 0.961, respectively. Similarly, the WI values of the testing phase are 0.968, 0.962, 0.966 and 0.974 for Loisinga, Patnagarh, Saintala and Deogaon watershed, respectively. Results of the PSR-SVM-FFA model for both testing and training phases are presented in Table 5.

Table 5 | Optimal parameters of τ , m for proposed watersheds and results of PSR-SVM-FFA model for runoff prediction

Positions	Optimal parameters of τ , m for PSR	Training period			Testing period		
		NSE	RMSE	WI	NSE	RMSE	WI
Balangir	$(m = 14, \tau = 5)$	0.905	43.74	0.984	0.874	42.27	0.985
Loisinga	$(m = 14, \tau = 5)$	0.938	39.18	0.970	0.908	37.96	0.967
Patnagarh	$(m = 14, \tau = 5)$	0.912	40.38	0.971	0.893	39.08	0.974
Saintala	$(m = 14, \tau = 5)$	0.897	37.66	0.972	0.869	36.11	0.976
Deogaon	$(m = 14, \tau = 5)$	0.923	41.82	0.982	0.897	40.35	0.988

Assessment of results for recommended model

A comparison between the PSR-SVM-FFA, SVM-FFA and SVM models for testing of all projected watersheds is presented in Figures 6–8. The pre-eminent WI values for SVM, SVM-FFA and PSR-SVM-FFA models are 0.935, 0.965 and 0.984, correspondingly, for Balangir watershed. Similarly, for Loisinga station, the paramount WI value is 0.919, 0.958 and 0.9702 for SVM, SVM-FFA and

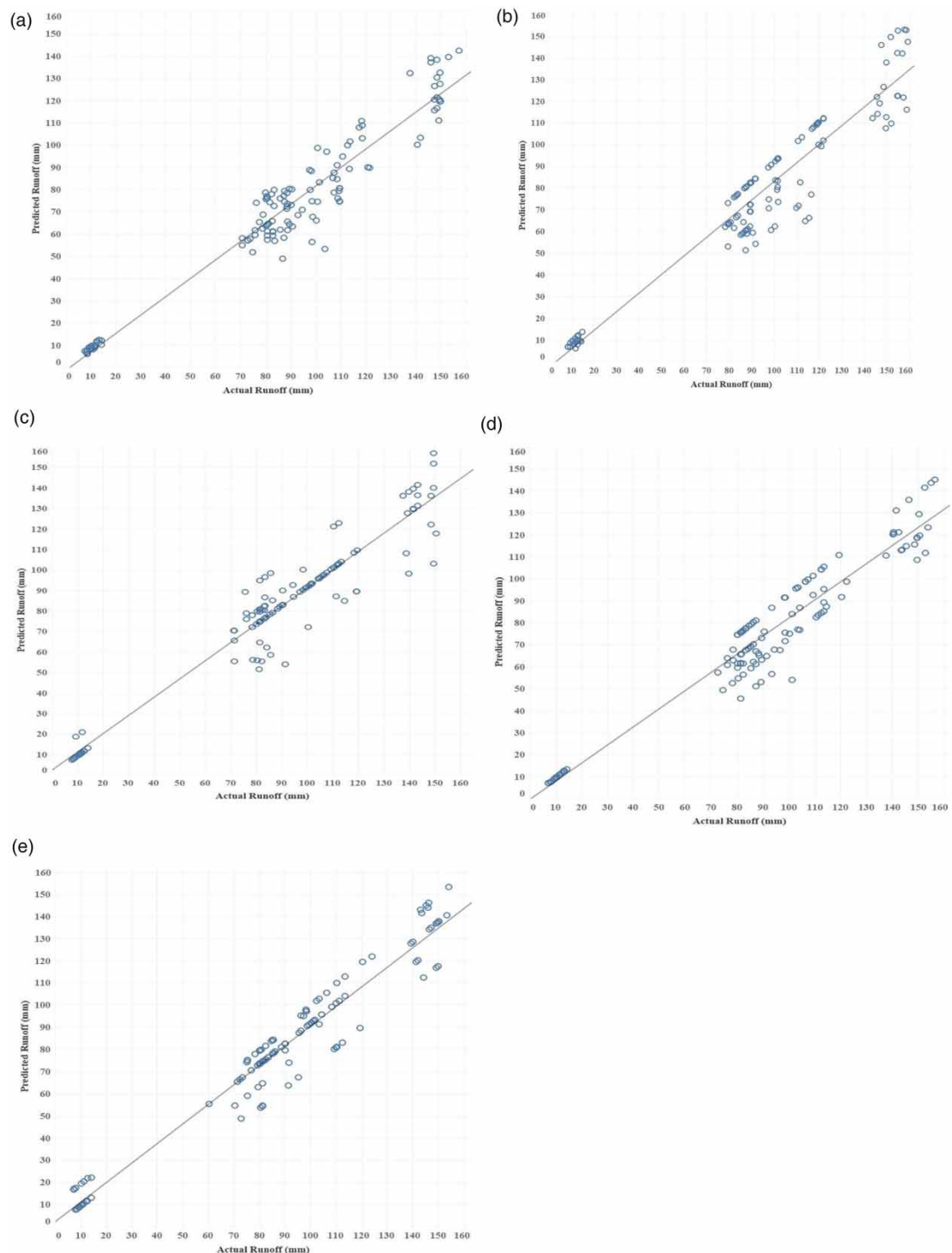


Figure 6 | Best fit lines by SVM model at (a) Balangir, (b) Loisinga, (c) Patnagarh, (d) Saintala and (e) Deogaon watersheds during training phase.

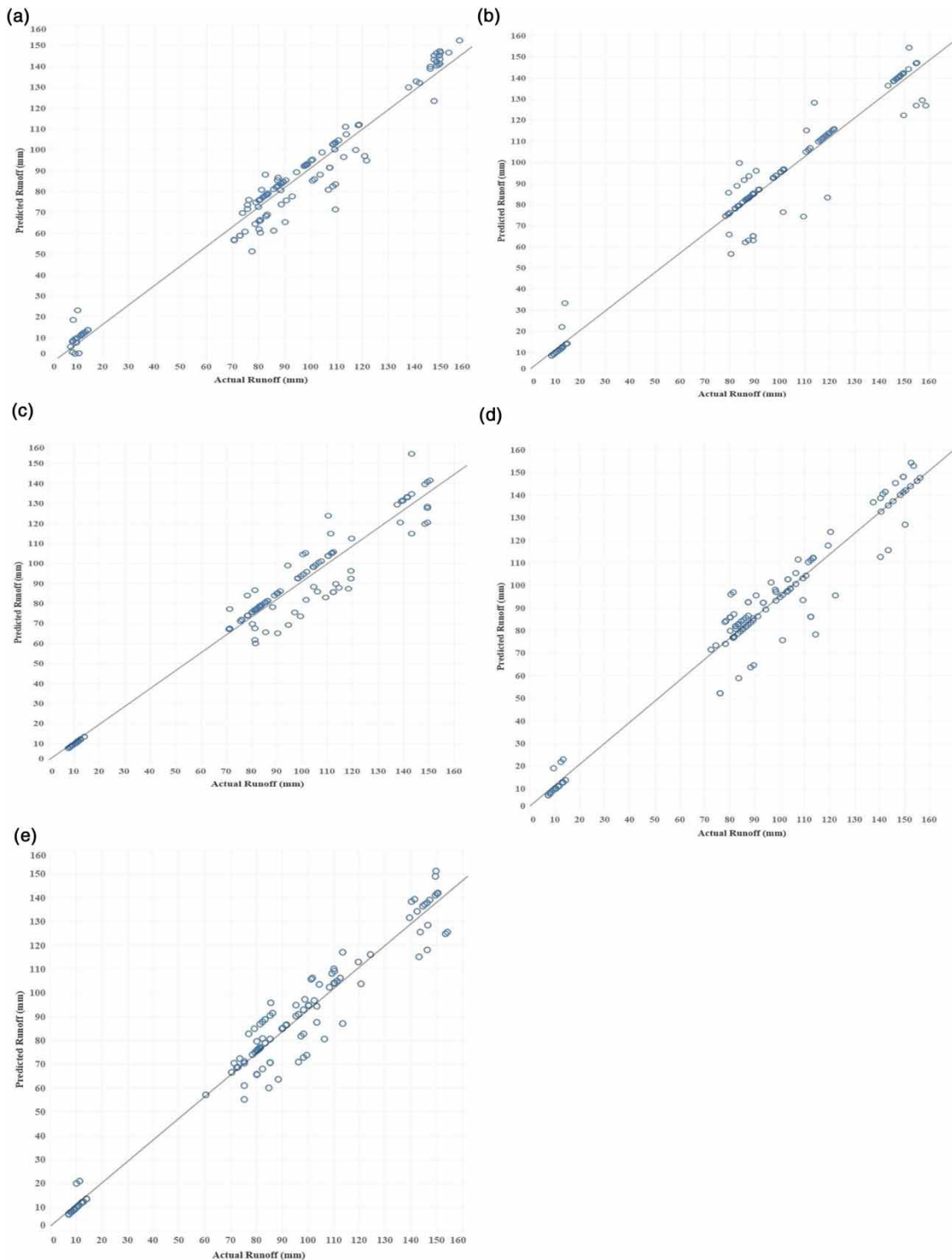


Figure 7 | Best fit lines by SVM-FFA model at (a) Balangir, (b) Loisinga, (c) Patnagarh, (d) Saintala and (e) Deogaon watersheds during training phase.

PSR-SVM-FFA models, respectively. Patnagarh, Saintala and Deogaon illustrate values for the SVM model of 0.9276, 0.9386 and 0.9459, respectively, during the testing phase. For the SVM-FFA model, the excellent values of WI are 0.953, 0.955 and 0.961 with respect to Patnagarh, Saintala and Deogaon. Similarly, the best values of the PSR-SVM-FFA model are 0.971, 0.972 and 0.982 for Patnagarh, Saintala and Deogaon, respectively.

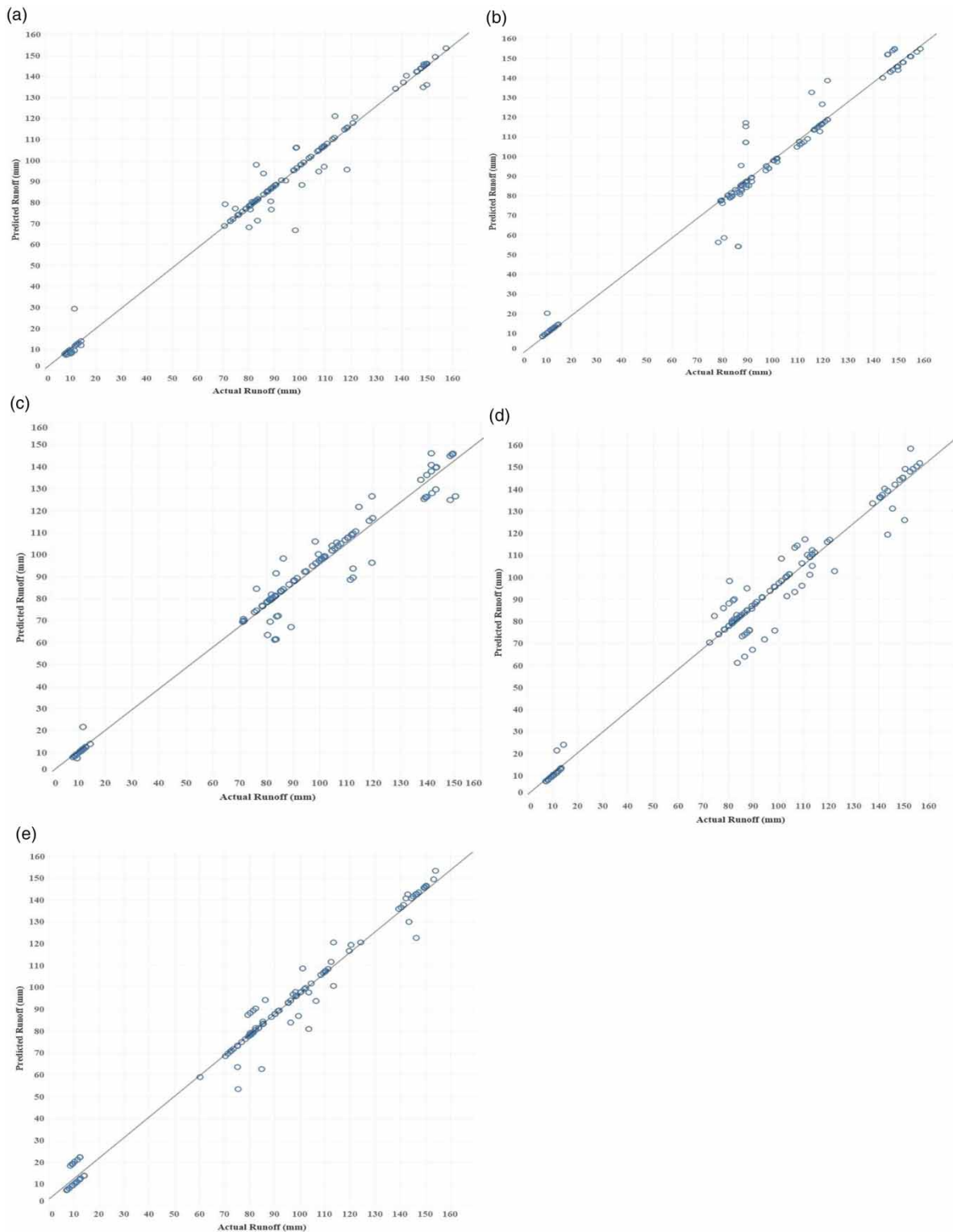


Figure 8 | Best fit lines by PSR-SVM-FFA model at (a) Balangir, (b) Loisinga, (c) Patnagarh, (d) Saintala and (e) Deogaon watersheds during training phase.

The linear scale plot of actual versus predicted monthly runoff for proposed model of projected area is shown in Figure 9. Results illustrate that estimated peak runoff is 141.208 mm, 142.313 mm and 149.272 mm for SVM, SVM-FFA and PSR-SVM-FFA against actual peak 157.22 mm for the Balangir watershed. The approximated peak runoffs are 126.4107 mm, 147.091 mm and 154.655 mm for SVM,

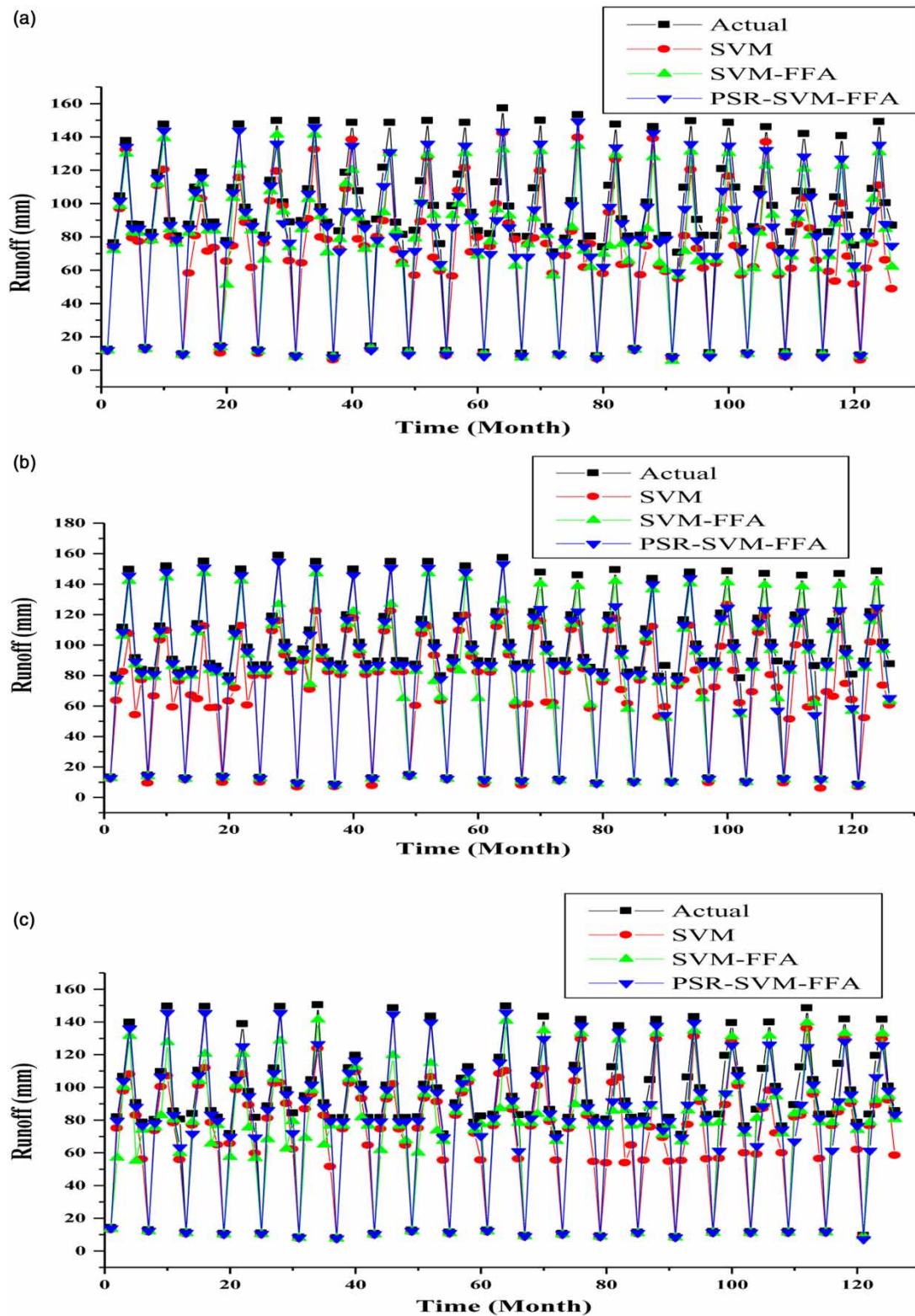


Figure 9 | Variation of measured runoff and predicted runoff at (a) Balangir, (b) Loisinga, (c) Patnagarh, (d) Saintala and (e) Deogaon watersheds during training phase. (Continued.)

SVM-FFA and PSR-SVM-FFA adjacent to the actual peak 158.8 mm for Loisinga division. For Patnagarh gauging station, tangible runoff is 150.43 mm aligned with predicted runoff of 136.032 mm, 141.329 mm and 145.744 mm for SVM, SVM-FFA and PSR-SVM-FFA, respectively. Correspondingly,

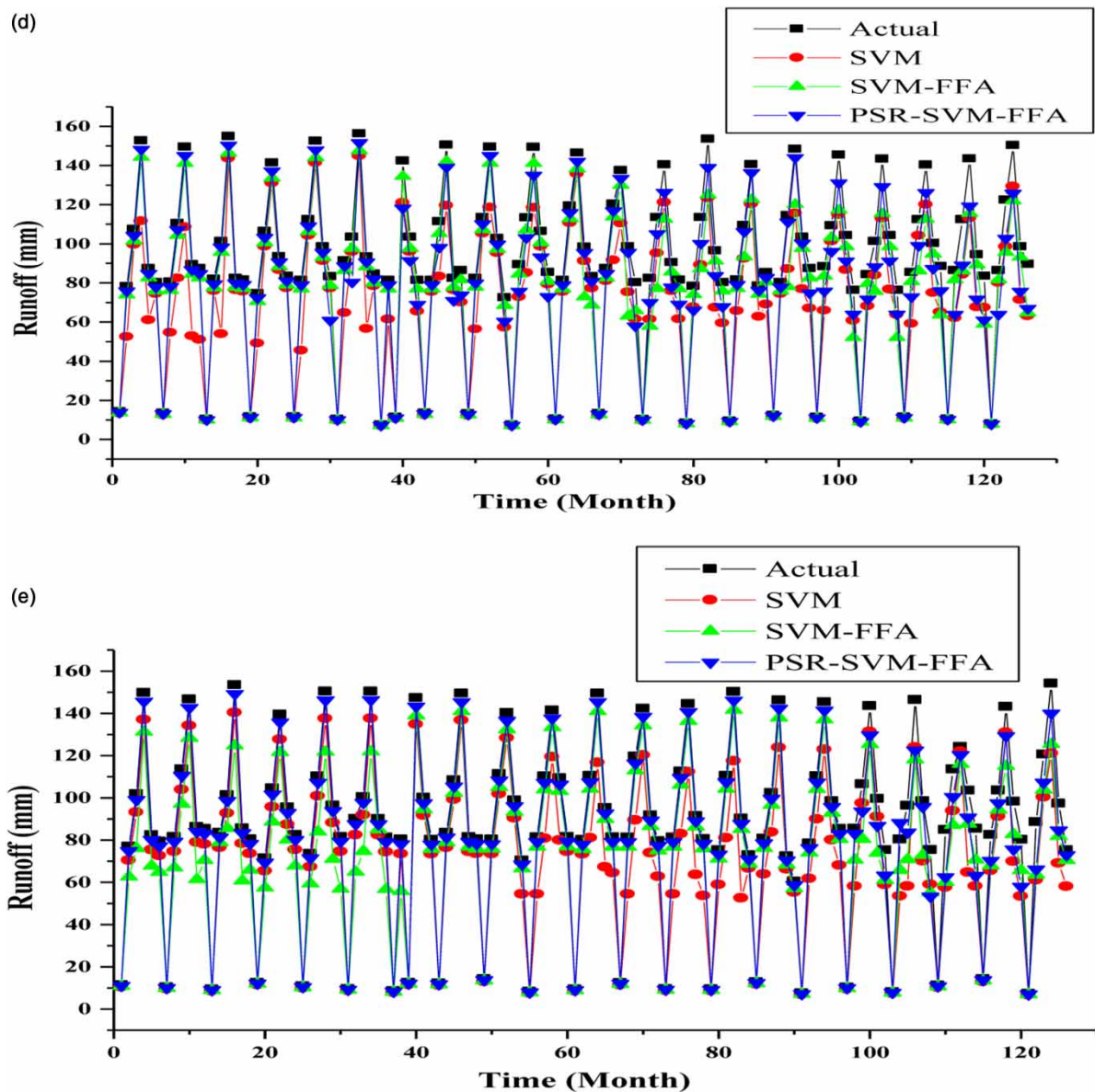


Figure 9 | Continued.

for Saintala watershed, the observed maximum runoff is 156.34 mm associated with estimated runoff of 151.728 mm, 147.522 mm and 144.911 mm for PSR-SVM-FFA, SVM-FFA and SVM, respectively. Similarly, evaluated peak runoffs are 149.322 mm, 141.591 mm and 140.376 mm, for PSR-SVM-FFA, SVM-FFA and SVM models with respect to genuine runoff of 154.23 mm for Deogaon watershed.

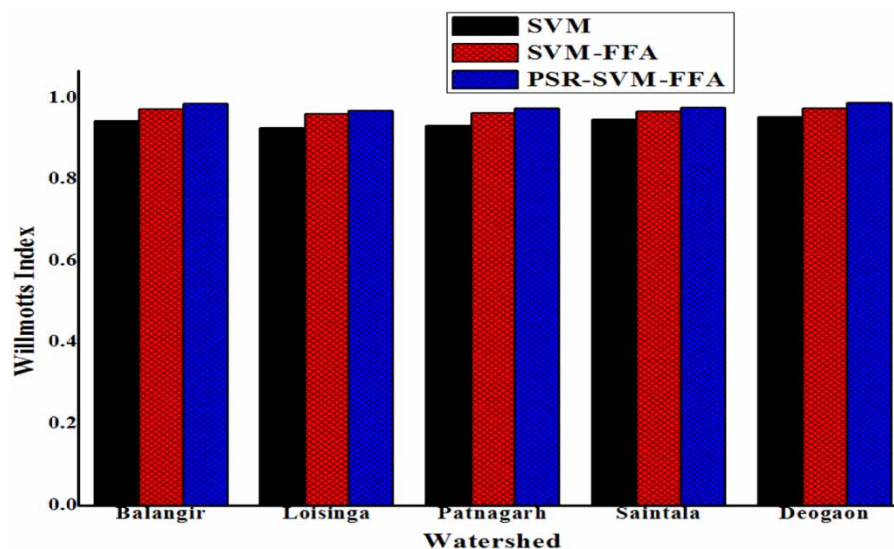
Comparison of best results

The SVM, SVM-FFA and PSR-SVM-FFA models are used to evaluate the performance of NSE, RMSE and WI indicators for five gauging watersheds. Assessments of performance indicators are specified in Table 6, which illustrates the efficiency of each model. Evaluating runoff is important and so the methods used here are significant for demonstrating runoff information. Therefore, calculation of RMSE, WI and NSE values are essential for predicting runoff. It is apparent that the PSR-SVM-FFA model is executed well, comparably to SVM-FFA and SVM. Accuracy of models is evaluated and assessed.

Assessment of each model is demonstrated in Figure 10 in terms of linear bar scheme.

Table 6 | Performance indices of different models for runoff forecasting

Stations	SVM			SVM-FFA			PSR-SVM-FFA		
	NSE	RMSE	WI	NSE	RMSE	WI	NSE	RMSE	WI
Balangir	0.653	73.92	0.942	0.791	70.72	0.971	0.874	42.27	0.985
Loisinga	0.601	70.21	0.925	0.783	68.49	0.961	0.908	37.96	0.967
Patnagarh	0.613	73.41	0.931	0.754	70.07	0.962	0.893	39.08	0.974
Saintala	0.658	69.73	0.947	0.793	67.88	0.966	0.869	36.11	0.976
Deogaon	0.648	74.05	0.953	0.779	70.11	0.974	0.897	40.35	0.988

**Figure 10** | Comparison of performance index for proposed watershed.

CONCLUSIONS

The efficiency of SVM, SVM-FFA and PSR-SVM-FFA approaches is explored for runoff prediction. Five gauging stations in Bolangir district, India are used for execution over a period of 28 years. The performance accuracy of PSR-SVM-FFA models is demonstrated in contrast to SVM-FFA and SVM models. It is appraised with various statistical indicators, WI, NSE and RMSE parameters, to reveal the projected PSR-SVM-FFA model is superior to SVM-FFA and SVM. It is observed that there is a significant improvement in the performance of the selected hybrid model for predicting the accuracy of each watershed. It is observed that the development of PSR coupled with the hybrid SVM-FFA model is superior to SVM models because of comprehensive withdrawal of input arrangements to produce targeted values. The PSR-SVM-FFA model is established to be a suitable model for runoff forecasting in arid watersheds. The research carried out over five gauge watersheds will help in the widespread good performance of non-gauging arid watersheds to predict runoff in similar conditions. Also, the present research findings can be used for irrigation and water resources engineering to design hydraulic structures regarding agriculture.

REFERENCES

- Abd-Elazim, S. M. & Ali, E. S. 2018 Load frequency controller design of a two-area system composing of PV grid and thermal generator via firefly algorithm. *Neural Computing and Applications* **30** (2), 607–616.

- Al-Shammari, E. T., Mohammadi, K., Keivani, A., Ab Hamid, S. H., Akib, S., Shamshirband, S. & Petkovi, D. 2016 Prediction of daily dew point temperature using a model combining the support vector machine with firefly algorithm. *Journal of Irrigation and Drainage Engineering* **142** (5), 1–9.
- Baydaroglu, O. & Kocak, K. 2014 SVR-based prediction of evaporation combined with chaotic approach. *Journal of Hydrology* **508**, 356–363.
- Cha, S., Sohanian, S. K., Kumara, D., Malikk, A., Chaharc, B. R., Namac, A. K., Panigrahi, B. K. & Dhima, R. 2014 A support vector machine-firefly algorithm based forecasting model to determine malaria transmission. *Neuro Computing* **129**, 279–288.
- Chen, W., Pourghasemi, H. R. & Naghibi, S. A. 2018 A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. *Bulletin of Engineering Geology and the Environment* **77** (2), 647–664.
- Cortes, C. & Vapnik, V. 1995 Support-vector networks. *Machine Learning* **20** (3), 273–297.
- Das, G. (ed.) 2009 Runoff. In: *Hydrology and Soil Conservation Engineering Including Watershed*, 2nd edn (G. Das, ed.). PHI Learning, New Delhi, India, pp. 79–98.
- Delafrouz, H., Ghaheri, A. & Ghorbani, M. A. 2017 A novel hybrid neural network based on phase space reconstruction technique for daily river-flow prediction. *Soft Computing* **22** (7), 2205–2215.
- Dutta, S., Singh, M. & Kumar, A. 2018 Classification of non-motor cognitive task in EEG based brain-computer interface using phase space features in multivariate empirical mode decomposition domain. *Biomedical Signal Processing and Control* **39**, 378–389.
- Fan, G. F., Peng, L. L. & Hong, W. C. 2018 Short term load forecasting based on phase space reconstruction algorithm and bi-square kernel regression model. *Applied Energy* **224**, 13–33.
- Gandomia, A. H., Yang, X. & Alavic, A. H. 2011 Mixed variable structural optimization using Firefly Algorithm. *Computers and Structures* **89**, 2325–2336.
- Ghose, D. K. & Samantaray, S. 2019 Estimating runoff using feed-forward neural networks in scarce rainfall region. In: *Smart Intelligent Computing and Application* (Satapathy, S. C., Bhatija, V. & Das, S., eds). Smart Innovation, system and technologies, Vol. 104. Springer, Singapore, pp. 53–64.
- Gocic, M., Shamshirband, S., Razak, Z., Petkovik, C. S. & Trajkovic, S. 2016 Long-term precipitation analysis and estimation of precipitation concentration index using three support vector machine method. *Advances in Meteorology*. **2016**, 7912357, 1–11.
- Hajiloo, R., Salarieh, H. & Alasty, A. 2018 Chaos control in delayed phase space constructed by the Takens embedding theory. *Communications in Nonlinear Science and Numerical Simulation* **54**, 453–465.
- Ji, Y. & Sun, S. 2013 Multitask multiclass support vector machines: model and experiments. *Pattern Organisation* **46**, 914–924.
- Lieu, Q. X., Do, D. T. & Lee, J. 2018 An adaptive hybrid evolutionary firefly algorithm for shape and size optimization of truss structures with frequency constraints. *Computers & Structures* **195**, 99–112.
- Mehr, A. D., Nourani, V., Khosrowshahi, V. K. & Ghorbani, M. A. 2018 A hybrid support vector regression–firefly model for monthly rainfall forecasting. *International Journal of Environmental Science and Technology* **16**, 335–346.
- Moghaddam, T. B., Soltani, M., Shahraki, H. S., Shamshirband, S., Noor, N. B. M. & Karim, M. R. 2016 The use of SVM-FFA in estimating fatigue life of polyethylene terephthalate modified asphalt mixtures. *Measurements* **90**, 526–533.
- Raghavendra, S. N. & Deka, P. C. 2014 Support vector machine applications in the field of hydrology: a review. *Applied Soft Computing* **19**, 372–386.
- Shamshirband, S., Mohammadi, K., Tong, C. W., Zamani, M., Motamedi, S. & Ch, S. 2015 A hybrid SVM-FFA method for prediction of monthly mean of global solar radiation. *Theoretical and Applied Climatology* **125** (1–2), 53–65.
- Sivakumar, B., Berndtsson, R. & Persson, M. 2001 Monthly runoff prediction using phase space reconstruction. *Hydrological Sciences Journal* **46** (3), 377–387.
- Sun, W. & Wang, Y. 2018 Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network. *Energy Conversion and Management* **157**, 1–12.
- Tabari, H., Kisi, O., Ezani, A. & Talaei, P. H. 2012 SVM, ANFIS, regression, and climate based model for reference evapotranspiration modelling using limited climatic data in semiarid highland environment. *Journal of Hydrology* **444–445**, 78–89.
- Tao, H., Sulaiman, S. O., Yaseen, Z. M., Asadi, H., Meshram, S. G. & Ghorbani, M. A. 2018 What is the potential of integrating phase space reconstruction with SVM-FFA data-intelligence model? Application of rainfall forecasting over regional scale. *Water Resources Management* **32**, 3935–3959.
- Tayfur, G. 2014 *Soft Computing in Water Resources Engineering: Artificial Neural Networks, Fuzzy Logic and Genetic Algorithms*. WIT Press, Southampton, UK.
- Thanh Noi, P. & Kappas, M. 2018 Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors* **18** (1), 18.
- Vignesh, R., Jothiprakash, V. & Sivakumar, B. 2015 Stream flow variability and classification using false neighbour method. *Journal of Hydrology* **531**, 706–715.
- Wang, H., Wang, W., Cui, Z., Zhou, X., Zhao, J. & Li, Y. 2018a A new dynamic firefly algorithm for demand estimation of water resources. *Information Sciences* **438**, 95–106.

- Wang, H., Zheng, B., Yoon, S. W. & Ko, H. S. 2018b A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research* **267** (2), 687–699.
- Wang, H., Wang, W., Cui, L., Sun, H., Zhao, J., Wang, Y. & Xue, Y. 2018c A hybrid multi-objective firefly algorithm for big data optimization. *Applied Soft Computing* **69**, 806–815.
- Yang, X. S. 2009 Firefly algorithms for multimodal optimization. In: *International Symposium on Stochastic Algorithms*. Springer, Berlin, Heidelberg, Germany, pp. 169–178.
- Zaini, N., Malek, M. A., Yusoff, M., Mardi, N. H. & Norhisham, S. 2018 Daily river flow forecasting with hybrid support vector machine-particle swarm optimization. *IOP Conference Series: Earth and Environmental Science* **140**, 012035.

First received 9 January 2020; accepted in revised form 12 March 2020. Available online 6 August 2020