Modeling the damming effect on hydrological alteration and prediction of discharge in Padma River by proposing PSO based novel hybrid machine learning algorithm

Abstract

This paper quantified the hydrological alteration of the Padma River basin caused by the construction of Ferakka Barrage (FB) using innovative trend analysis (ITA), range of variability approach (RVA), and continuous wavelet analysis (CWA). We also predict flow regime by proposing particle swarm optimization (PSO) based novel hybrid machine learning algorithms. Results of the ITA showed the negative trend of the average discharge in the dry season (January-May), while the RVA analysis indicated that average discharge was lower than environmental flows. The CWA demonstrated a substantial effect of the FB on the periodicity of the streamflow regime. Results showed that PSO-Reduced Error Pruning Tree (REPTree), PSO-random forest (RF), and PSO-M5P were the optimal fit for average, maximum, and minimum discharge prediction (RMSE = 0.14, 0.3, 0.18) respectively.

Keywords: Farakka barrage, Bangladesh, Innovative Trend, RVA, CWA, Hybrid machine learning model

1. Introduction

One of the greatest modifications of the fluvial landscape on earth is the construction of artificial structures, such as dams and/or barrages over the river. Hydrological alteration in the downstream of the river, after the construction of the barrage, is one of the major challenges (Arévalo-Mejía et al., 2020). In Bangladesh, change of discharge downstream of the river due to the barrage is unequivocal.

Hydrological alteration signifies the indication of changes either by magnitude or by timing or both of the factors of a river system. Anthropogenic activities, such as dams, embankment, settlements and resource abstraction have the potential to alter the natural stream flow pattern (Gain and Giupponi 2014; Poff et al., 1997). Land use land cover changes through the construction of artificial structures are the key anthropogenic factor for the hydrological alteration of a river. The Padma river basin has been considering as a huge dynamic river system and also affected by the excessive human interventions, such as hydropower generation, as well as the negative impacts from the changing climatic regime in this region (Mirza, 1998; Nawfee et al., 2018). Until 1975, the river was its natural flow state. However, on 21 April 1975, the FB barrage was installed over Ganga River by the Indian government. The barrage location is situated in Farakka, West Bengal at the upstream (16.5 km) from the Indo-Bangladesh border. Following its operation, dry season flow reduced significantly (Mirza, 1998), while the monsoonal discharge increased in the downstream part of Bangladesh. Due to the barrage, there is a significant change in the river flow pattern which affected the ecological and social systems, evidenced by existing works (Gain and Giupponi, 2014; Talukdar and Pal, 2020). Besides, the withdrawal of water in the dry period induced hydro-geomorphological alterations . The barrage also affected the downstream floodplain regime, including its connectivity, nutrients dynamics, and sediment influxes (McCartney, 2009).

The hydrological alteration of the river due to the construction of barrage or dam on the river is well documented (Graf, 2006; Islam, 2016; Talukdar and Pal, 2017). To investigate the impact of Farakka barrage on the downstream part of the river in Bangladesh, a good numbers of research have already been carried out (Rahman and Rahman, 2017; Gain and Giupponi, 2014), but predicting the future flow by using machine learning algorithms is still scarce. Thus, it is essential to predict the stream-flow by using machine learning algorithms in the Padma River basin for sustainable water resource management

Water resource management planners depend on the stream-flow regime and the potential flow of basin, establish an ecological flow condition, and warrant sustainable catchment management (Akhter et al., 2019). However, researchers, especially in developing countries, encounter the problems of data scarcity like unavailability of long-term time series data, because very few gauge stations have been installed over river, while many rivers do not have gauge station. Similarly, Padma River has only single gauge station. Therefore, water resource management planners must rely on the applications of sophisticated statistical and machine learning techniques for quantifying the trend, stream-flow pattern, periodicity and predicting streamflow for incoming days to conserve/ restore the fluvial system and sustainable development (Razavi and Coulibaly, 2012). Enhancing the forecasting of streamflow in an ungagged stream has been a major goal for hydrologists in the hydrology field (Blöschl, 2016). Recently, the machine learning algorithms have been gaining attention to the researchers and scientists for forecasting future flow scenarios under different hydrological conditions. For instance, Worland et al. (2018) applied a set of eight machine learning algorithms to compare the prediction of low-flow indices for an ungauged river in the USA.

Of late, some researchers have examined the possibilities of addressing water engineering problems by using advanced soft computing models. Machine learning algorithms, such as Random forest, and random trees (RT), artificial neural network (ANN), support vector machine (SVM), bagging etc. have been progressively employed in solving hydraulic engineering issues. Many recent works successfully applied the machine learning algorithms to the field of groundwater hydrology

(Sepahvand et al., 2019; Sihag et al., 2019), water resources (Adnan et al., 2017) and engineering (Tien Bui et al., 2020). Ensemble machine learning algorithms have been widely used in different fields, but few studies related to stream-flow prediction have been conducted very recently (Granata et al., 2018).

Keeping the research gaps in mind, this paper deals with the development and application of three ensemble machine learning algorithms, such as PSO-RF, PSO-REPtree, and PSO-M5P to forecast the stream-flow. So far, there is no study on the prediction of streamflow in the Padma River basin by using machine learning algorithms. To the best of authors' knowledge, the development and application of PSO-M5P, PSO-RF, and PSO-REPtree along with ITA, RVA, and CWT methods, have not yet been applied for quantifying the hydrological alteration and predicting the stream-flow in relation to barrage or dam construction. Therefore, this study is an attempt to quantify the hydrological alteration due to the FB as well as predicting the stream-flow by using ensemble machine learning algorithms for the first time in Bangladesh.

2. Materials and methodology

2.1 Description of the study area

The Padma river is the major downstream stretch of the Ganges River, which flows more than 2562 km^2, originating from the Gangotri glacier of the Himalaya. The basin area of the Ganges river is considered as one of the densely populated inhabitants on the Earth. About 407 million population of two countries, namely India, and Bangladesh either directly or indirectly obtain benefits from the Ganges and its downstream Padma river. This river plays a pivotal role in the socio-ecological settings of these countries. The stretch of the Padma river runs for 108 km^2 before the confluence with river Meghna at Chandpur, Bangladesh. The total discharge of the

Ganges river and Brahmaputra flows through the course of the Padma River is 30,000 m^3/s^{-1} and sometimes, it reaches to 75,000 m^3/s^{-1} during bank full stage (Dewan et al., 2017). The elevation of the river course decreases by 5 cm/km (Sarker and Thorne, 2006). This study has been conducted on the Padma river basin, which covers almost 8 districts, including Pabna, Shirajganj, Natore, Bogura, Jaypurhat, Naogaon, Rajshahi, Chapainawabganj (Fig. 1). Annually, 900 metric tons of sediment load passes through the river, out of which, 60% of sediment is either silt or clay, while the rest is bed load (Hossain, 2010). Dewan et al. (2017) described the floodplain of the river as a 'wandering' pattern. Apart from this, this basin reports extreme variability of flow regime (water and sediment), triggering from monsoonal precipitation and the melting of the Himalayan ice, which causes frequent large floods of high magnitude in Bangladesh. Four seasons, such as summer (March to May), monsoon (June to September), post-monsoon (October to November), and winter (December to February) have predominated in this region with significant temperature and precipitation differences.

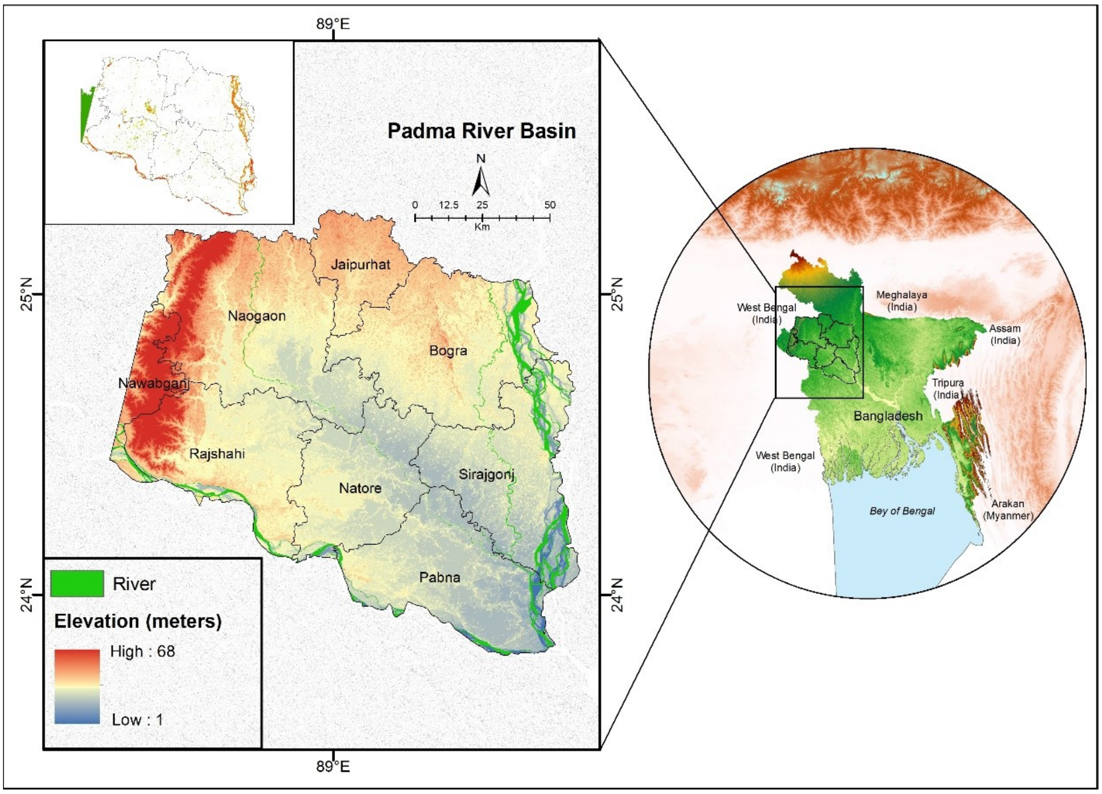

Figure 1: Map showing the location of the Padma River basin, Bangladesh (SRTM 30m DEM, source: USGS).

2.2 Materials

Daily river water flow data (1970 to 2018) of the Padma river basin were collected from the representative gauge station (Hardinge Gauge station over Padma River, Pabna).

2.3 Methods for innovative trend analysis

In this paper, we used ITA, which was first proposed by Sen (2012). Sen (2012) reported some limitations in most commonly employed Mann-Kendall and Spearman's rho tests, for instance, they need normally distributed datasets, long-term time series datasets, independent variables etc. (Kişi et al., 2018), which works in accordance with a Cartesian coordinate system. For the ITA approach, the time series data have been needed to be classified into two classes. The original trend indicator has been expressed by the following equation 1:

$$\Phi = \frac{1}{n}\sum_{1}^{n}\frac{10X_j - X_i}{\mu} \tag{1}$$

where, Φ means the indicator of trend, number of total observations are denoted by n; Xi denotes the data of first sub-series; Xj and μ are represented as the data of the second sub-series and mean value of the data of the first sub-series, respectively. The increasing and decreasing trend can be identified by the positive and negative value of Φ.

2.4 Methods for hydrologic alteration using the range of variability approach

Large numbers of flora and fauna form the ecosystem of a river and its floodplain. The particular amount of stream-flow is required for their survival and good health (Richter et al., 1996). If the required water availability is not available in the river, the species do not get adequate water and then it becomes difficult for their survival. In this study, we attempted to identify the ecological range of flow, which can be considered as suitable for sustaining ecosystems and also presented the effect of stream-flow change on the ecological state. It is very difficult to calculate the threshold water volume level needed for the useful existence of the riparian ecosystem.

Therefore, Richter et al. (1996) recommended several dispersion measures (e.g., ±1 or ±2 standard deviation, and eightieth percentile) for assessing primary threshold levels of stream-flow condition. Here, the standard deviation has driven RVA for the monthly average, the maximum and minimum water level are calculated. Gain and Giupponi (2014) reported ±1 SD-based RVA was computed for different months to determine exceedance flow condition over time using the following Eq.2:

$$(\text{Mean} - \text{SD}) \leq \text{Parameter} \leq (\text{mean} + \text{SD}) \tag{2}$$

Richter et al. (2003) described that setting a flow range (RVA) is not the plan to formulate all values of the downstream of river lie within the range, however the principle goal of RVA is to count the downstream flow of post change years, which lie within this flow range with the same frequencies, which were observed during pre-change period. In this way, the frequency of pre and post change years were counted for every IHA parameters. The frequency of years during the pre-change period that attains the threshold limits of RVA is regarded as the "expected" values. During the post Farakka barrage period, the frequency of years occupies the threshold levels of RVA, considered as the “observed” values. In light of this computation, hydrologic alteration is evaluated (Pal and Talukdar, 2020). In this manner, the degree of hydrological alteration (DHA) for each variable is computed by the below Eq. 3:

$$Degree\ of\ hydrological\ alteration(\%) = \left\{\frac{Observed \text{ frequnecy} - Expected \text{ frequency}}{Expected \text{ frequency}}\right\} \times 100$$

(3)

The stream-flow alteration could be zero, positive, and negative values. When it equals zero, indicating the present condition is within the anticipated range and no alteration happens with good ecosystem health. When it equals a positive value, expressing the observed values of the variables attained the threshold limit more times than the expected, which indicates a good sign for the ecosystem, although it will not be good for some species as they prefer to survive in optimal range. However, the negative value of DHA indicates that the observed value does not attain the threshold lime as expected value, which represents the bad ecosystem health. Consequently, the species living there may have faced critical situations for survival. The HDA of maximum, average, and minimum stream-flow was represented in the graphical form of the heat map.

2.5 Periodicity analysis by using continuous wavelet transform

The periodicity analysis on the time series data using wavelet transformation is not very old and is a modified form of Fourier transformation. In the case of the Fourier transform, it reveals the information in one dimension. If it gives information about the time domain, the scale domain will be lost and vice versa. While the wavelet transformation is able to give both time and scale domains information (Liu et al., 2016;). This is why wavelet transformation has become a popular technique to solve the time series problems . The non-stationary time series data (mean, variance, covariance, and autocorrelation are changed over time and not able to get back to its original state) are most appropriate to use the wavelet analysis to explore the volatility of the data. Therefore, hydro-meteorological time series data is one kind of non-stationary data, consequently, the wavelet transformation is a useful method. Goupillaud et al. (1984) first used wavelets as a family of functions derived from the translations and dilations of a single function, which is known as ‘‘mother wavelet”.

2.6 Proposing PSO based novel hybrid machine learning algorithms for stream flow prediction

In this study, we proposed PSO-RF, PSO-REPtree, and PSO-M5P algorithms to predict and forecast the stream-flow of the Padma river. Earlier studies have applied several stochastic models, like the average model and autoregressive moving average (ARMA), moving average model and autoregressive moving average (ARIMA) models to predict and forecast the highly nonlinear stream-flow behavior. These approaches performed well when the data placed within the range of former observations. When the data shows the extreme events, these models perform poorly. However, these problems have been solved by the advancement of the machine learning algorithms. Although the machine learning algorithms have several flaws which prevent models to perform accurately. Therefore, good optimization can overcome the flaws of the machine learning techniques. Therefore, in this paper, for the first time, PSO-RF, PSO-REPtree and PSO-M5P have been applied for stream-flow forecasting.

2.6. 1 Particle Swarm Intelligence

The PSO is a powerful meta-heuristic robust evolutionary algorithm for optimization based on the population behaviour and was first proposed by Eberhart and Kennedy (1995). The PSO theory was motivated by the social behavior of the fish and birds in groups for optimizing the shortest route to find the food (Roshanravan et al., 2019). Recently, the PSO algorithm has been successfully and extensively applied to resolve the non-linear problems in several fields like Geology (Gilani et al., 2020), flood susceptibility modelling (Bui et al., 2020), landslide susceptibility modelling (Sun et al., 2020), forest fire mapping(Zhang et al., 2020) because of the higher learning speed and takes less memory than the other optimization algorithm like genetic algorithm (Nguyen et al., 2019). The swarm of particles in the PSO algorithm always tries to find the potential answer to the problem which can be best positioned as per the best solution. The particles in PSO travel randomly along the search space. Based on its own and neighbor's knowledge, a swarm of particles dislocates in search space. The particles become skilled from each other within the group and travel in the direction of their best neighbor based on the obtained knowledge. In a nutshell, the PSO is based on the concept that each swarm of particles changes its location in the search space in order to get the best position or location that it has ever been and the best location nearer its neighbor.

2.6.2 Machine learning algorithms

Application of RF

RF is a modified bagging supervised artificial intelligence model, commonly used for prediction and classification. In recent times, this model is applied for forecasting time series datasets (Pal and Talukdar, 2020). The RF algorithm is a nonparametric ensemble classifier tool that works by using the algorithm of the flexible decision tree of Breiman (Breiman, 2001). The RF builds decision trees, in which each tree is constructed by utilizing the bootstrap training samples (Breiman, 2001). To build a better model, it is important to grow a large tree (Breiman, 2001). Accordingly, it is necessary to have a requisite number of selected predictor variables at all the nodes of the trees. The number of observations at the terminal nodes of the trees would be minimum. In this method, randomly selected training data from the actual dataset through the algorithm were applied to generate the model (Breiman, 2001; Youssef et al., 2016). The performance of different models is dependent significantly on the optimization of the model's parameters. However in the present, the RF model has been optimized by the PSO algorithm.

Application of REPtree

The REPTree algorithm is a speedy decision tree logic algorithm that employs the principle of computing information gain with entropy and reducing the variance error (Quinlan, 1987; Devasena, 2014). REPtree uses the regression tree algorithm and generates several trees in different calculation procedures from where it took the best one from all the produced trees (Devasena, 2014). REPTree is capable of generating the modeling procedure flexible and easy using creating the training datasets when the output is huge and it declines the complication of the tree interior structure (Mohamed et al., 2012). The pruning algorithm in this method has taken into account the backward over-fitting complexity and attempts to get the minimum version of best-precision tree logic using the post-pruning algorithm (Quinlan, 1987; Chen et al., 2019). It only selects values for numeric attributes once (Kalmegh, 2015). The performance of this model relies on information gain from entropy, reduction variance, and declined error pruning model (Srinivasan and Mekala, 2014).

2.6.3 *Application of M5P*

M5P tree, introduced by Quinlan (1992), consists of a conventional decision tree having the option of linear regression functions at the nodes (Suthar and Aggarwal, 2019). The divergence metric is known as the Standard Deviation Reduction (SDR), which is used to generate the decision tree. Moreover, a linear regression function is also applied to develop tree models. The process works through pruning, evacuation, and substitution of trees. Finally, a final tree model is constructed **(**Suthar and Aggarwal, 2019**).** A tree model generally used to predict the output of several input values after analyzing the provided data sets. This algorithm handles continuous class issues rather than discrete class issues and can deal with tasks with greater dimensionality. The M5P tree exhibits piecewise information of each linear function created to estimate nonlinear associations of each data set. M5P trees are easy but efficient and appropriate tools for modeling the tree patterns and association for large data sets (Quinlan, 1992).

2.6.3 Procedure for ensemble

In the present study, the PSO algorithm was applied to obtain the best structural parameters of the mentioned machine learning algorithms. The ensemble procedure for the proposed PSO-REPTree, PSO-RF, PSO-M5P could be as follows: parameters initialization of PSO algorithm→ training and testing of machine learning algorithm with the initialized parameters→ calculation of fitness function→ fitness value of each swarm of particle in reference to local and global best values→ updating the velocity and position of each swarm of particle accordingly→ reaching maximum number of iteration? if not reached, starting again from the second stage→ if it reached the maximum number of iteration, that would be the optimal parameters for the machine learning algorithms. The initialization of parameters of PSO itself were selected. The detailed initialized parameters and the optimized parameters for machine learning algorithms were provided in supplementary table 1.

After the algorithms produced better prediction results, we went for forecasting the stream-flow for incoming days. The flowchart adopted for this study is shown in Fig. 2.

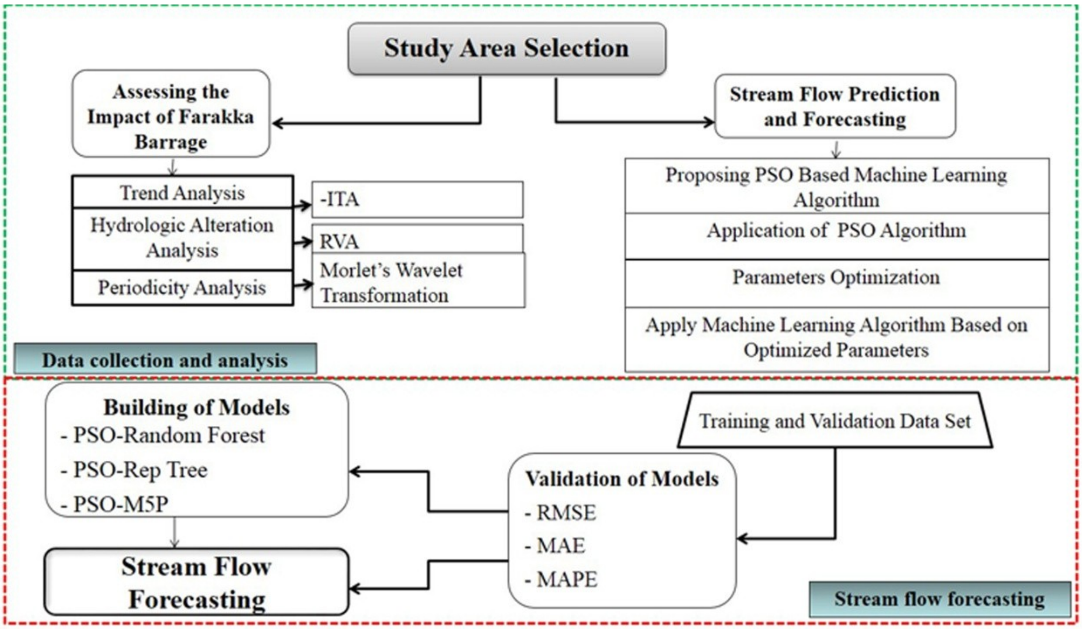

Figure 2: Flow chart of assessing the impacts of hydrological alteration due to Farakka barrage in the Padma River basin.

2.7 Validations

For assessing the precision of the models, different statistics have been established and used. The best known and most widely used techniques were used in the present study. In this work, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) measures were used to estimate performance of the different prediction models. The RMSE, MAE, and MAPE were calculated by using Eqs. 6, 7 and 8, respectively.

$$RSME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_{(Predicted\ flow)i} - Q_{(Observed\ flow)i})^2} \quad (6)$$

$$MAE = \frac{\sum_{i=1}^{n}\left|(P_{(Predicted\ flow)i} - Q_{(Observed\ flow)i}\right|}{n} \quad (7)$$

$$MAPE = \left(\frac{1}{n}\sum\frac{|Observed\ flow - Predicted\ flow|}{|Observed\ flow|}\right)\times 100 \quad (8)$$

3. Results and analysis

3.1 Trend Analysis

The outcome of innovative trend analysis is presented in Table 2, showing slope (D value) for average, maximum, and minimum discharge of the river. Here, the positive values indicate an increasing trend, and the decreasing trend was depicted by negative values. From Table 2, it was found that the average discharge of January, February, July, November, and December had positive D value, which indicates the mean discharge of the river has been increased. Whereas, March-June, and August-October showed negative D value, indicating the decreasing stream flows (Table 2). Besides, the graphical results of the ITA for average discharge for all IHA parameters are shown in supplementary figure1. Stream flow is very low in May and the discharge starts from 500 m^3/s, possibly reflecting dry seasonal withdrawals of waters by the FB.

However, for maximum discharge, January-March, July, August, and October-December showed positive values (Table 2). In July, December, and January the stream-flow increased significantly. The Negative D values were observed in April, May, June, and September (Table 2). The graphical form of ITA showed a clear increasing trend in January, May, July, and October to December (Fig. S2). In the case of minimum discharge, the D value showed a significant decreasing trend in April and May. Other IHA parameters, like March, June, and September also showed negative trends. July and January showed a slight positive trend for minimum discharge. Figure S3 showed the decreasing trend for the minimum discharge of March, April, May, June, and September.

Table 2: Slope (D) value of innovative trend analysis for average, maximum and minimum discharge in the Padma River basin

Month	Slope (D) value for average	Slope (D) value for maximum	Slope (D) value for minimum
January	0.878	1.23	0.237
February	0.339	0.92	0.024
March	-0.190	0.47	-0.780
April	-0.468	-0.083	-1.64
May	-1.174	-0.011	-1.475
June	-1.05	-0.137	-0.297
July	0.612	1.01	1.71
August	-0.016	0.041	0.125
September	-0.643	-0.552	-0.178
October	-0.025	0.340	0.683
November	0.344	0.81	0.420
December	1.23	2.28	0.89

3.2 Modeling of hydrological alteration

The monthly hydrological alteration of the post barrage period (1970 to 2018) for average flow has been severely altered (Supplementary table 2) from its natural condition, which indicates the river has been impacted from moderate to high condition of hydrological alteration, while comparing with the scale developed by Talukdar and Pal (2019). Consequently, the monthly alteration was evaluated through the DFA to obtain the magnitude of hydrological alteration in the basin.

In supplementary table 2, the high, low, and ecological thresholds were calculated by using the RVA based on pre-barrage discharge data. The ecological threshold values depicted the range of natural flow, which requires for sustaining a healthy ecosystem. Values greater than the higher threshold and smaller than the lower threshold are indicative of positive and negative hydrological alteration. In a sense, positive hydrological alteration can be considered as good, but in case of some species, it can cross their optimal required water and can be harmful for them to complete their survival cycle. The RVA analysis for maximum discharge illustrated that in February, March, April, and May, the water flow was lesser than the lower RVA-threshold (Supplementary table 2). However, in the case of minimum discharge, it was observed that the discharge was lower than the low RVA threshold in the entire year. This indicates that the observed discharge was lower than the expected discharge. The discharge of post-barrage conditions was lesser than the computed threshold of ecological water throughout the year indicating the altered hydrological condition. Subsequently, it is reasonable to state that the FB caused the hydrological alteration of the Padma river in Bangladesh.

To quantify the hydrological alteration in micro-scale caused by FB installations (considered all years), the heat maps were constructed (1976-2018) for maximum, minimum and average discharge conditions (Figure 3a-c). Results showed that the degree of alteration gradually increases since the construction of the FB, which expected to be continued in the future. The dendrogram showed a pattern with the recorded months according to the extent of the hydrological alteration in the river. Multilevel hierarchy showed a distinct difference in the degree of hydrological alteration on a monthly scale.

Figure 3a showed the magnitude of hydrological alteration considering the maximum discharge. The dark yellow color denotes the decrease of the flow or negative hydrological alteration, while green color denotes the increase of the water flow or positive hydrological alteration. The negative hydrological alteration gradually increased after the FB installation. Results showed the discharge increases in July, August, and October, indicating the change in flow due to monsoonal rainfall. The increasing pattern extended to September, November, December, and June only until 2007. The discharge decreased from January to May except year 2007, perhaps, this happened due to a residual effect of the large flood occurred in that year. In 2018, the decreasing rate was high. In fact, after 2007 the decreasing flow has been getting elevated (3a). The degree of hydrologic alteration for minimum discharge is shown in Figure 3b. It is clearly observed that the post-barrage period records very low discharge. It has been expected almost every year after the commissioning of the FB. This is a very concerning issue for the hydrological regime of Padma River, Bangladesh. Two recent consecutive years 2017 and 2018 showed a worsening scenario (Fig.3b). The heat map of the hydrological alteration for average discharge shown in Fig. 3c, which indicated decreasing discharge in the month of January, February, March, April, May, June, September, October. On the other hand, July and August had slightly upward discharge, however, from 2009 to 2018, the decreasing rate is high (Fig 3c).

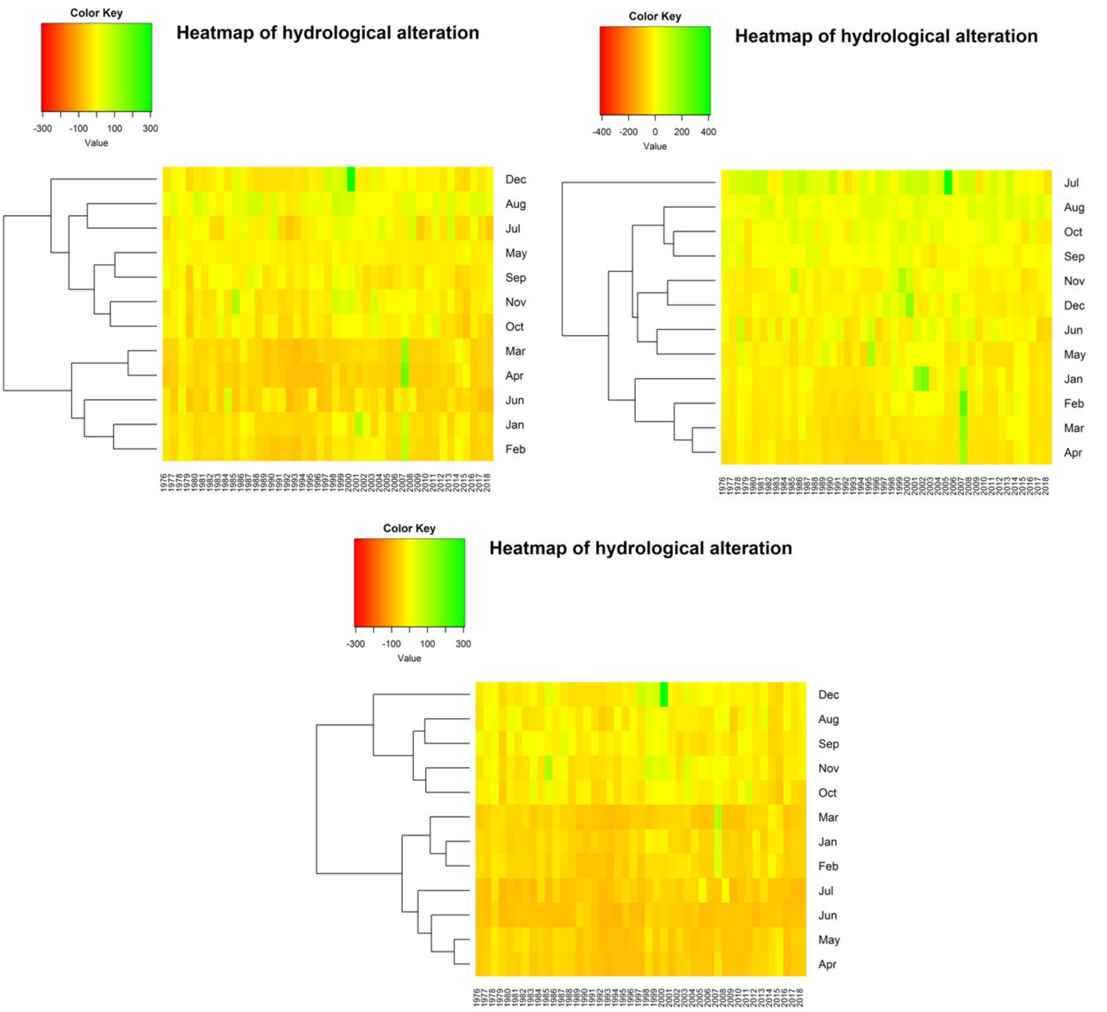

Figure 3: Degree of hydrological alteration for (a) Average, (b)Maximum, (c) Minimum discharge

3.3 Periodicity Analysis

Figures 4a-c showed the periodicity analysis of monthly average, maximum, and minimum discharge for the period of 1970 to 2018. The color pallet of the wavelet transform maps showed the distribution of power (absolute value squared) of the wavelet transform computed from the time series discharge datasets. The dark red color represented the stronger power, while light violet color expressed the weak power. The solid black line bands were the strong power significant at the level of 0.05. In case of average discharge, the strong wavelet power spectrum was found in the 1-month band for whole periods, which was significant at the level of 0.05, while a 0.5-month (15 days) band was observed in 1978, 1981, 1987, 1988, 1998-2000 (Fig. 4a). The relatively stronger power observed after 2003, indicating the disturbance of the normal periodic cycle of discharge. In case of maximum discharge, the strong wavelet power was found in the 1-month band for the whole time frame, while 0.5-month (15 days) band was observed in 1987 to 1989, 1998; 0.125 to the 0.5-month band was recorded in 2006 (Fig. 4b).

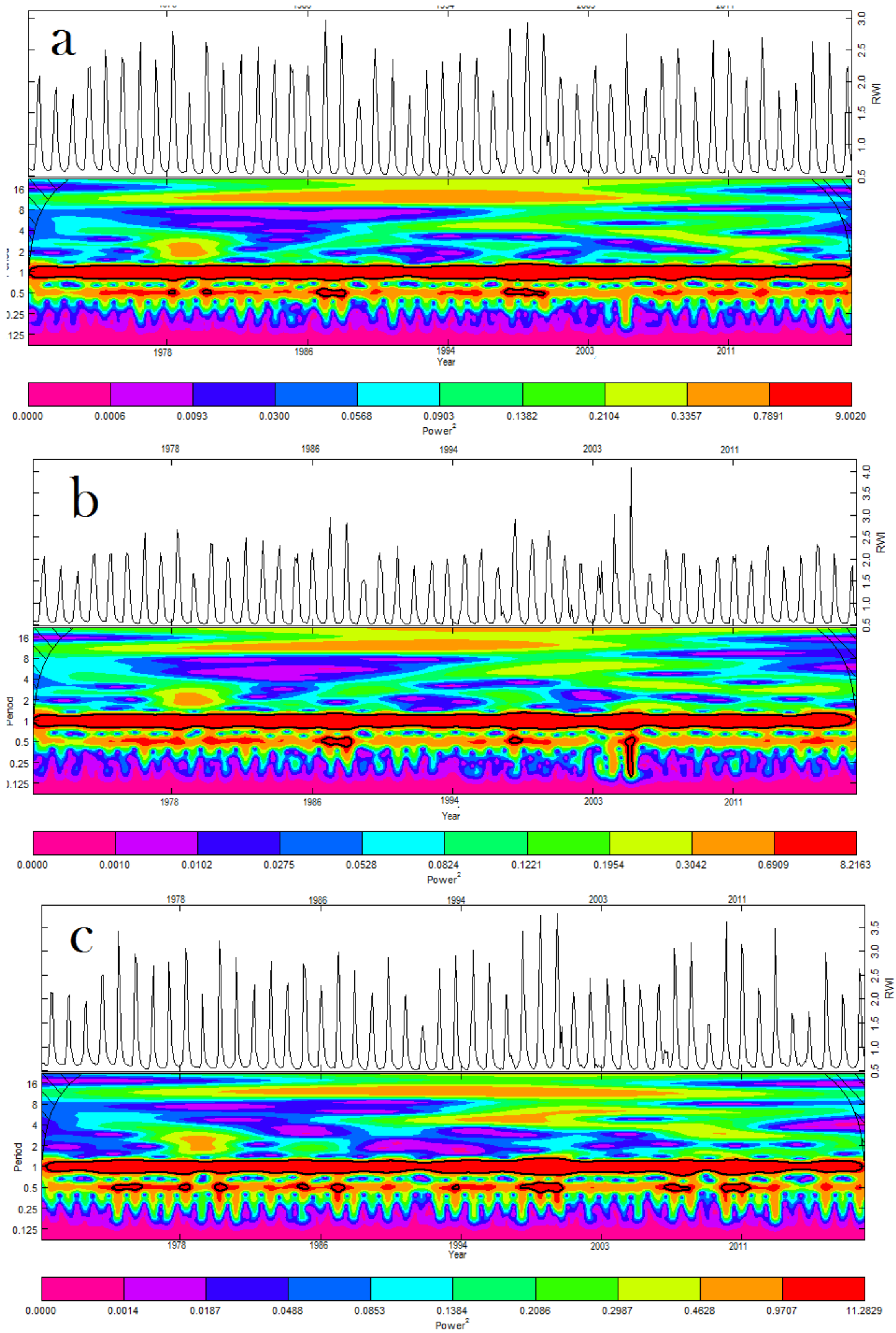

Figure 4: Periodicity analysis using morlet's wavelet transformation (a) average, (b) Maximum and (c) Minimum discharge

In case of minimum discharge, a strong wavelet power spectrum was found in 1-month band for the whole study period, while 0.5-month (15 days) band was found in 1974, 1975, 1978, 1980, 1985, 1994, 1998-2000, 1998, 2007, and 2011 (Fig. 4c). This strong wavelet power spectrum represents the variance of flow. From the analysis of average, maximum, and minimum, it is found that the highest power is in the band of 1 month. It means that all flow properties have been changed almost in the same direction and magnitude. While, the maximum concentration of strong power was observed after 2003 for average, maximum, and minimum discharge, which indicates that the nature process of periodic cycle has been disturbed significantly after 2003. Perhaps, the reason was the installation of FB in the downstream part.

Figure 5a-c showed the significance level of wavelet power against time. The findings showed that the significant power at the 95% level was centered in a 1-month band which provides an impartial and balanced estimation of the time series analysis.

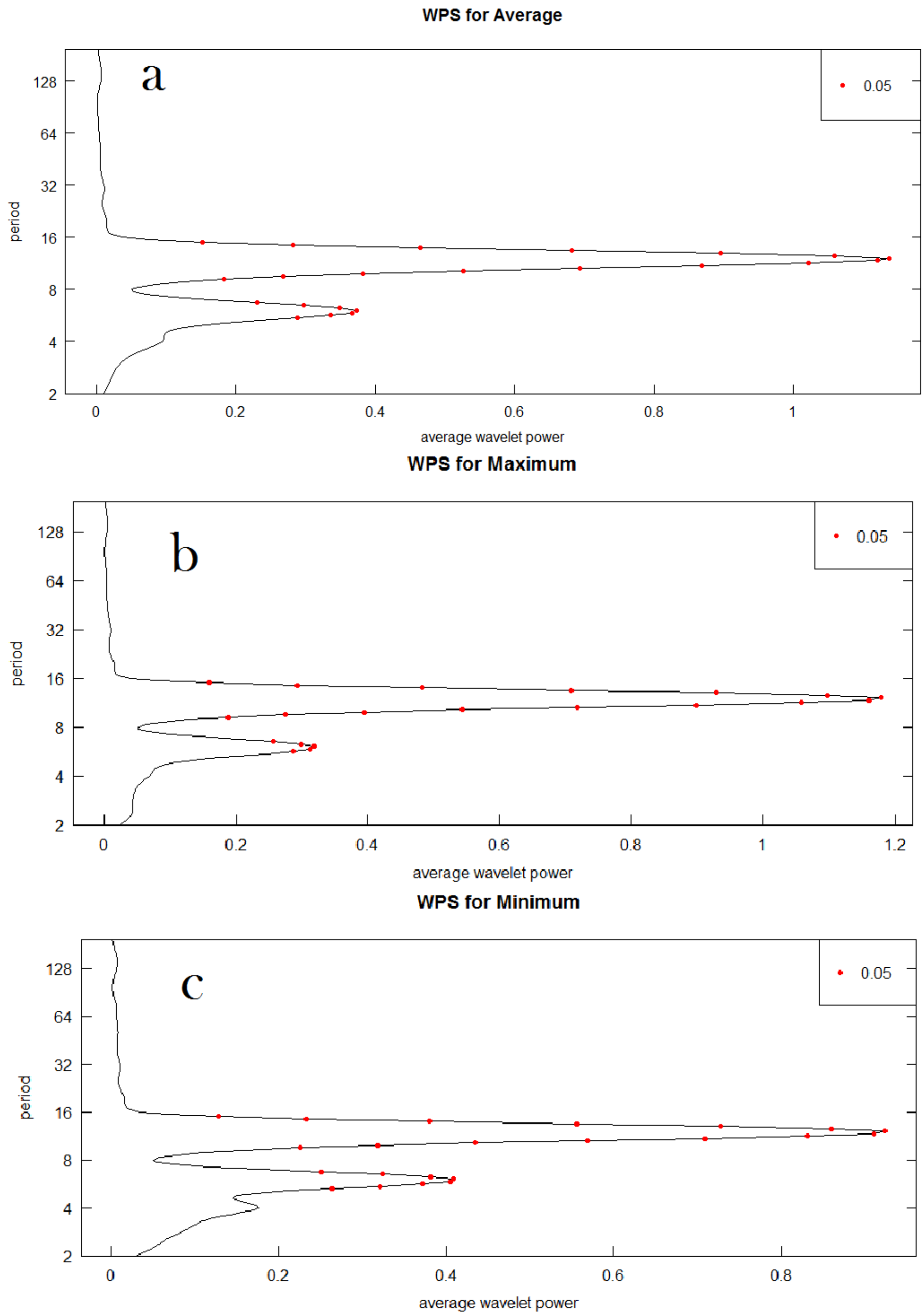

Figure 5: Global power spectrum at 0.05 significance level for (a) average, (b) Maximum and (c) Minimum discharge

3.4 Streamflow prediction by proposing novel hybrid machine learning algorithm

The average, maximum, and minimum discharge was predicted by using PSO-RF, PSO-REPTree, and PSO-M5P methods from 1970 to 2018 (Figures 6-8). The discharge for all conditions were predicted since 1976-2018 based on the lagged input parameters. For evaluation of the all predicted models, the error measures, like RMSE, MAE, and MAPE were computed between observed and simulated flow. The error measures showed the closeness of observed and predicted stream-flow, if the closeness or adjacency between the flows was reported, then the error would be less and vice-versa (figure 6-8). Figures 6, 7 and 8 showed the significance of the predicted values. The predicted values and the observed values were almost similar in the training dataset and closeness is more. Therefore, less error was found for all prediction models. To find out the best streamflow prediction models, the error measures were compared with each other. Lower values of RMSE, MAE, and MAPE indicate a better fit. Table 4 showed that among three models, PSO-REPtree appeared to be the best fitted for average discharge prediction, PSO-RF has the best fitted for maximum discharge and PSO-M5P was the best for minimum discharge prediction (Table 4).

Table 4: Performance measures between observed and simulated flow datasets for different models in this study

		PSO-RF	PSO-M5P	PSO-REPTree
RMSE	Average	0.21	0.25	0.14
	Maximum	0.3	0.36	0.41
	Minimum	0.26	0.18	0.24
MAE	Average	0.23	0.27	0.21
	Maximum	0.29	0.42	0.35
	Minimum	0.17	0.26	0.22

MAPE	Average	0.32	0.36	0.24
	Maximum	0.53	0.46	0.51
	Minimum	0.29	0.21	0.25

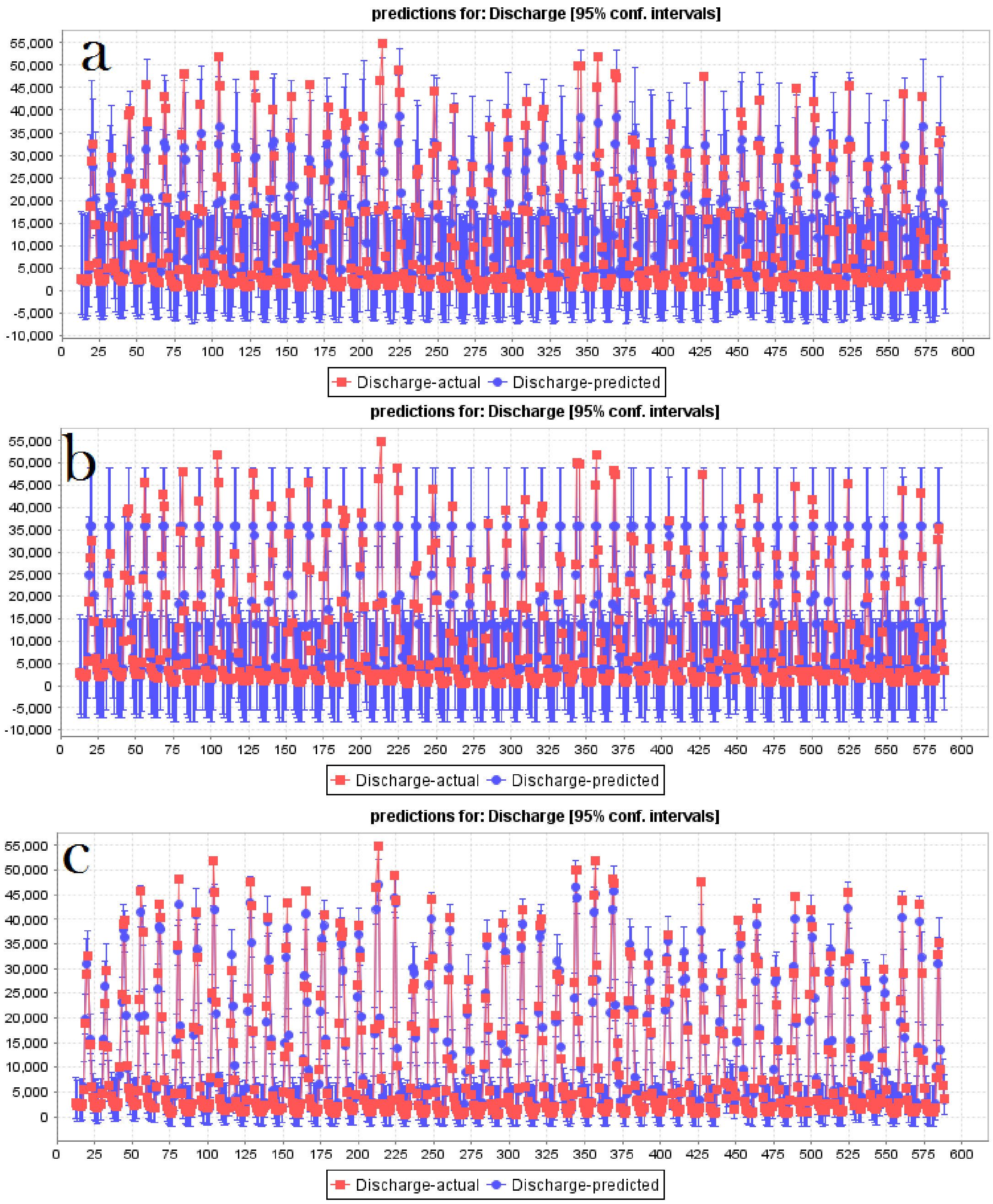

Figure 6: Overlaying the actual average discharge and predicted average discharge by (a) PSO-M5P, (b) PSO-REPTree, (c) PSO-RF model (N.B. Blue line represents the 95% confidence interval of the predicted data)

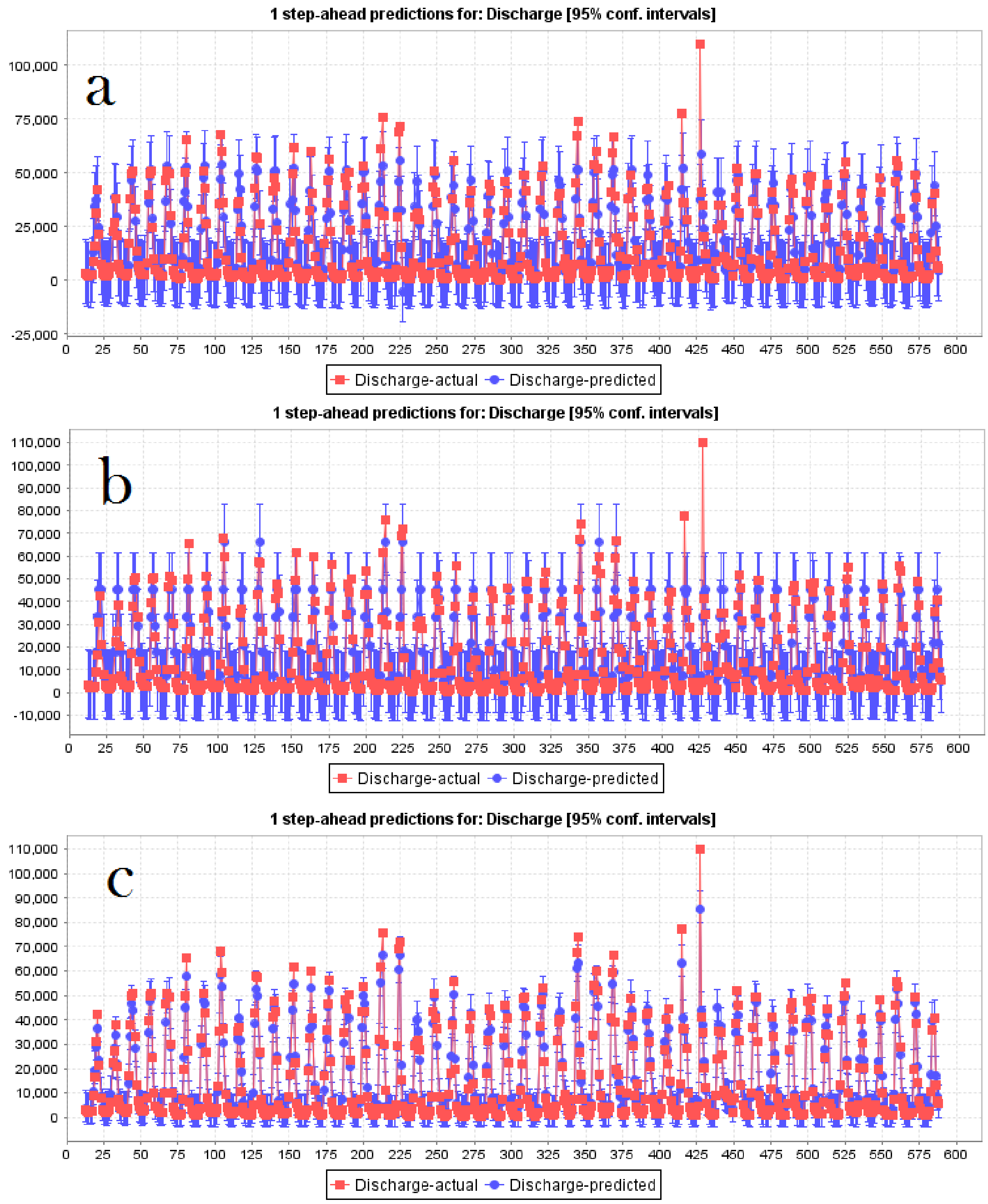

Figure 7: Overlaying the actual maximum discharge and predicted maximum discharge by (a) PSO-M5P, (b) PSO-REPTree, (c) PSO-RF model (N.B. Blue line represents the 95% confidence interval of the predicted data)

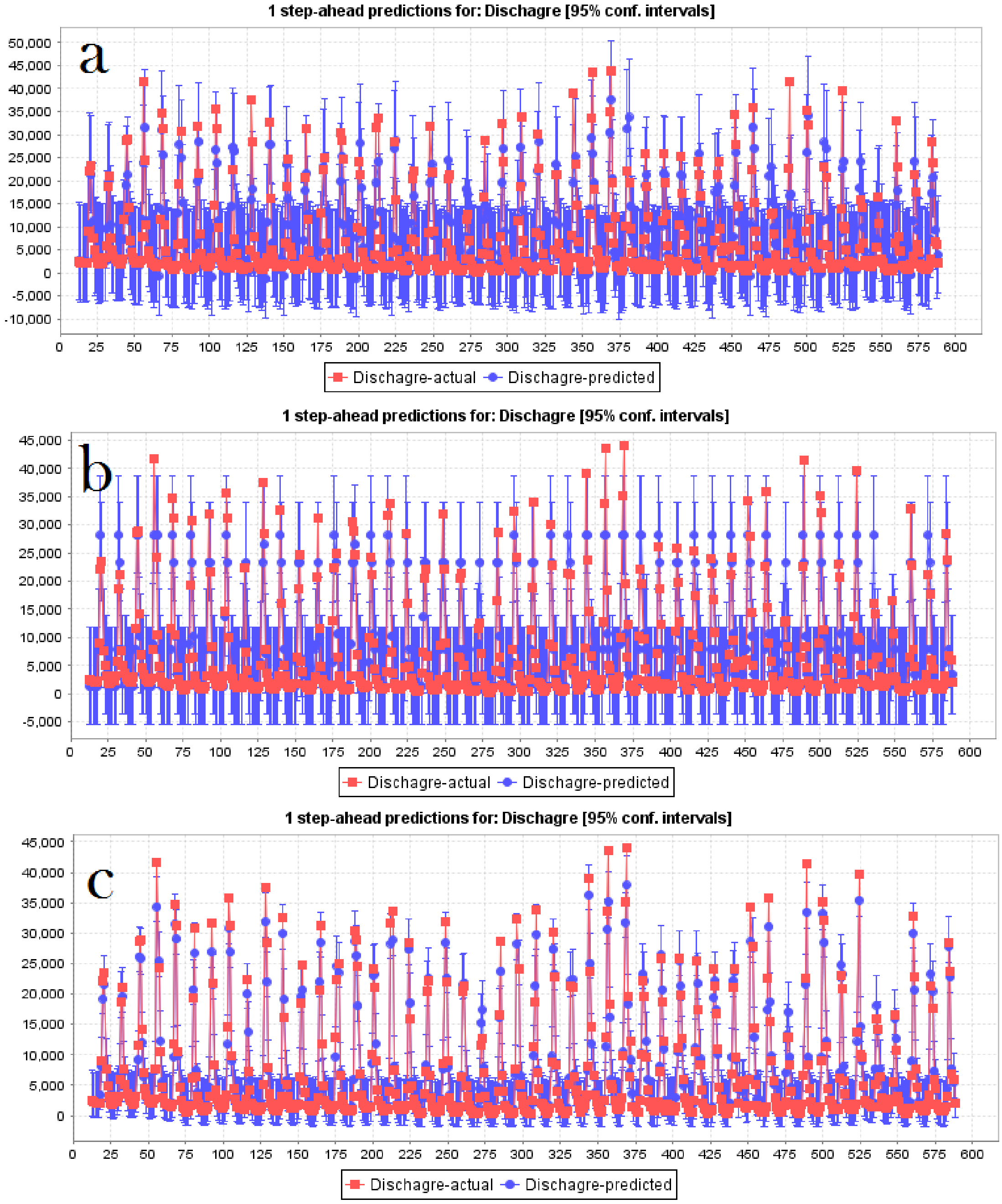

Figure 8: Overlaying the actual minimum discharge and predicted minimum discharge by (a) PSO-M5P, (b) PSO-REPTree, (c) PSO-RF model (N.B. Blue line represents the 95% confidence interval of the predicted data)

3.4.1 Average discharge forecasting

Figure 9a showed the discharge of water from 1970-2018 and the predicted discharge up to 2030. The actual discharge showed a gradual decrease in the flow. The predicted discharge up to 2030 showed the gradual decreasing trend. The highest value among the predicted discharges is 37260.7601 m^3s^{-1} which was less than 55000 m^3s^{-1} in the observed year. So, the impact of the barrage is clear.

Figure 9(b) showed the average discharge of observed and predicted years. Here the prediction is done by the PSO-REPTree method which denotes that from 2019 to 2030 the discharge is decreasing. After the year 1990 discharge rate is decreasing in most of the months. Most of them are in the range of 40000 m^3s^{-1}. The predicted data shows that in future the highest discharge will be 35810.22 m^3s^{-1}which is less than the actual discharge of the previous years.

Figure 9(c) showed average discharge prediction using the PSO-M5P method. Indicating that average discharge would be decreasing in the future. The highest average discharge is predicted to be about 30566.1274 m^3s^{-1}.

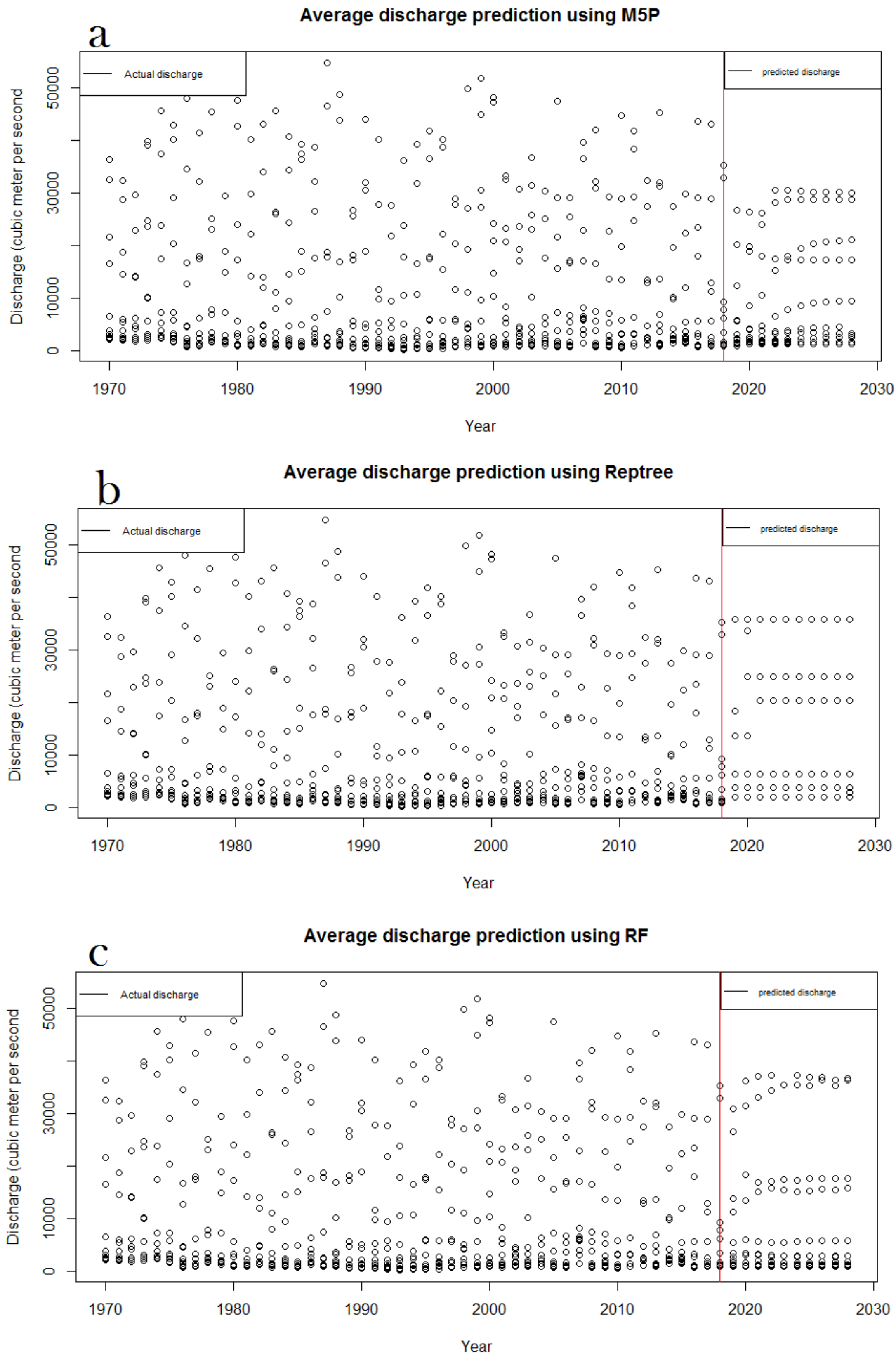

Figure 9: Average discharge prediction up to 2030 using (a) PSO-M5P, (b) PSO-REPTree, (c) PSO-RF

3.4.2 Maximum discharge forecasting

The maximum discharge of the Padma River is predicted using PSO-RF, PSO-REPTree and PSO-M5P methods. Figs. 10 (a-c) shows that discharge is decreasing in future. The Figure 10(a) depicted that the highest discharge will be around 40000 m^3s^{-1} in the predicted year until 2030. The highest discharge will be 41768.32 m^3s^{-1} in the year 2022. This is lower than the maximum discharge of the observed years from 1970 to 2018 as the maximum discharge is up to 50000 cubic meter per second. The maximum discharge is lowered after year 2000 with two or three months exception.

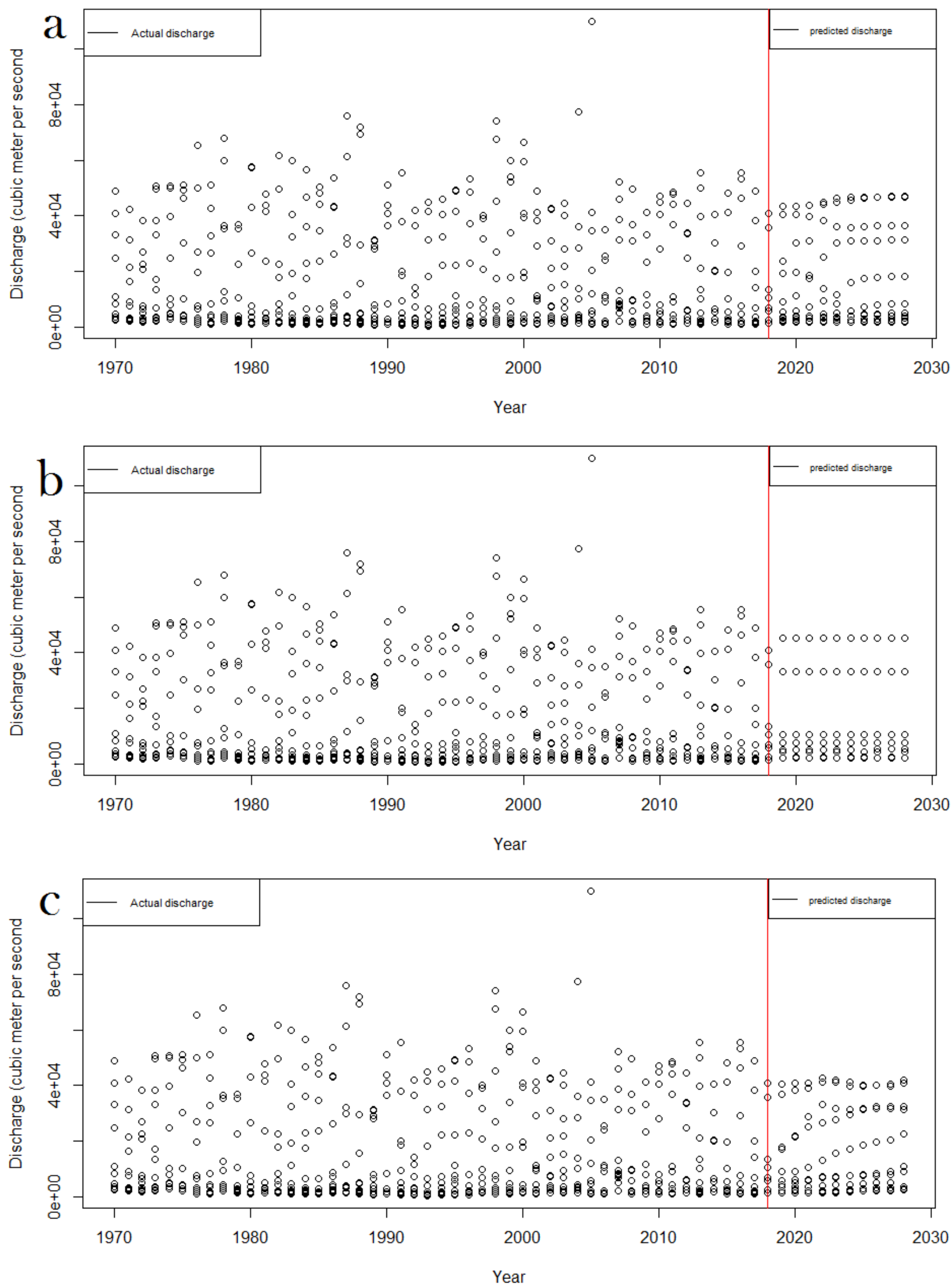

Figure 10: Maximum discharge prediction up to 2030 using (a) PSO-M5P, (b) PSO-REPTree, (c) PSO-RF

Figure 10(b) shows the maximum discharge using PSO-REPTree and its prediction up to 2030, illustrating the gradual decreasing trend of the water flow. The method predicted that the discharge will be around 45128.11 m^3s^{-1} in future. Figure 10(c) represents the maximum discharge prediction using the PSO-M5P method. This similar result is like maximum discharge. The maximum discharge could be around 45000 m^3s^{-1}.

3.4.3 Minimum discharge forecasting

The minimum discharge and its prediction up to 2030 is found by using PSO-RF, PSO-REPTree and PSO-M5P methods. The minimum discharge was lower than 40,000 from 1970 to 2018. But the predicted values show a huge decrease in the minimum discharge from 2019 to 2030 which would be very alarming. The prediction gives that the highest minimum discharge will be around 23982.73 m^3s^{-1}. The maximum predicted values are near 20000 m^3s^{-1} (Figure 11a).

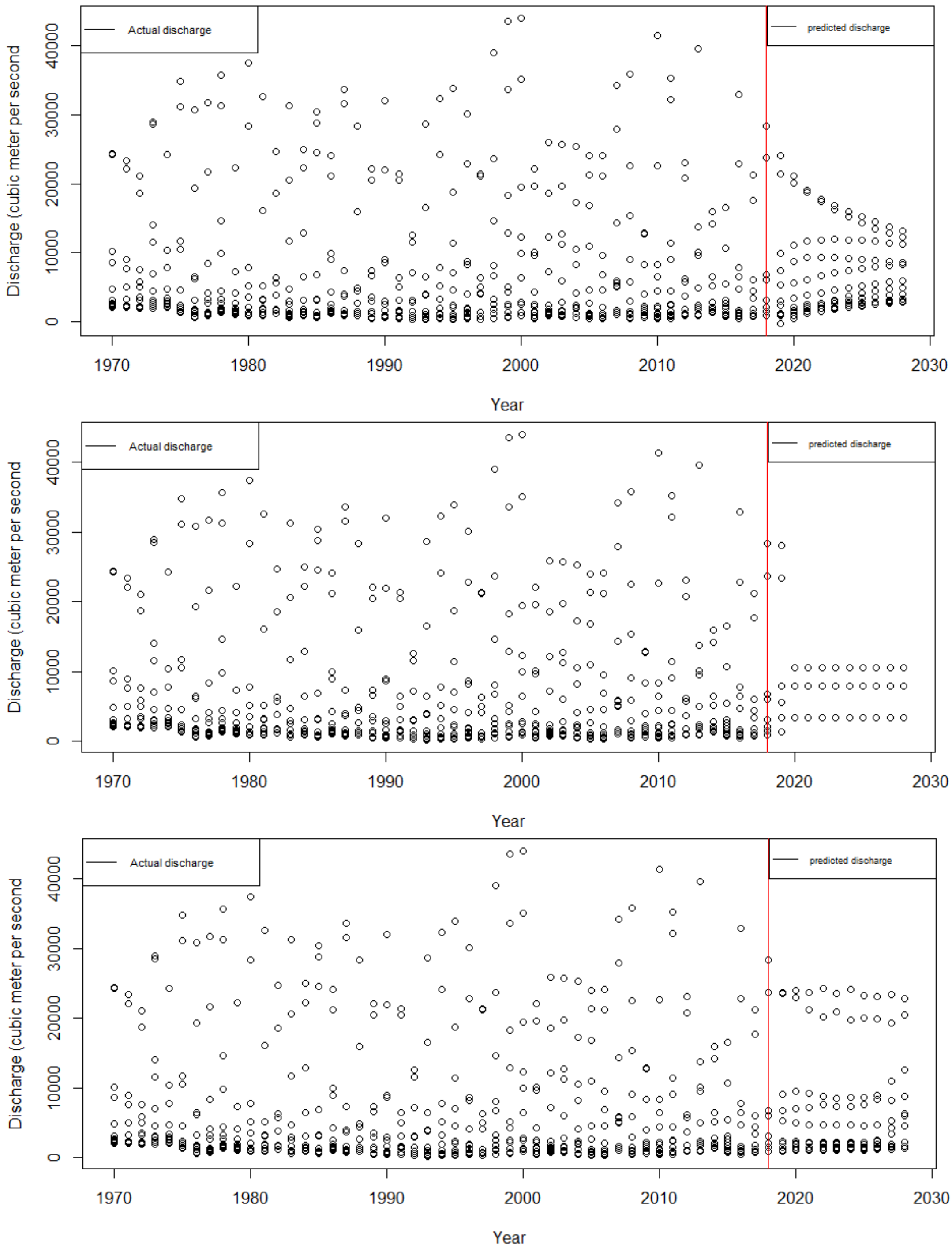

Figure 11: Minimum discharge prediction up to 2030 using (a) PSO-M5P, (b) PSO-REPTree, (c) PSO-RF

Figure 11(b) shows prediction of minimum discharge from 2019 to 2030. This illustrates the discharge will be very low and it will be around 10000 m^3s^{-1}. Figure 11(c) shows the prediction of the minimum discharge. It denotes that the discharge is decreasing gradually up to 2030. The predicted value shows that it will be 13237.72 m^3s^{-1} in 2030.

4. Discussion

The findings from the analysis indicated the impacts of the FB on the river flow. The river flow has been significantly obstructed by the FB over the entire downstream of the Padma River in Bangladesh. Although people of the upstream of the dam are getting the benefit of it, it becomes a serious burden on the people living in the downstream of the dam. The hydrology of the river has transformed due to the installation of the barrage and consequently, the wetland areas of the Padma river have been transformed. Due to the withdrawal of the Ganges water by FB, Bangladesh has been experiencing severe environmental degradation due to low flow in the Padma river.

The analyses identified that the water flow obstruction by FB is responsible for numerous changes in the hydrological regime during the dry season in the Padma river in Bangladesh (Mirza, 1997; Islam et al., 2016).

The present study clearly depicted the progressive rise of the degree of hydrological alteration through a heat map (Figs. 4 and 5). Such a degree of alteration is caused by increasing eco-deficit in the river (Pal and Talukdar, 2018). Supplementary table 1 showed that after the barrage commissioned, most of the months in each year represented the highest discharge restricted below the lower threshold or very close to the lower threshold flow of the river. Therefore, there are very obvious threats for the natural environment of the river in terms of its resources, ecosystem balance, biodiversity, hydrodynamics, livelihoods, and other socio-economic components of the river basin in Bangladesh. The result found in this study is similar to the work of Smakhtin et al. (2006), Gain, and Giupponi (2014). From the periodicity analysis, a significant change was found in the streamflow. Sanz et al. (2005), Richter et al. (1996), Olden and Poff (2003) reported the possible impacts of hydrological alteration on the biogeochemical cycle as well as biotic species diversity lived in the aquatic systems.

Pal and Talukdar (2020) rightly documented that climate change or anthropogenic control may be responsible for such changes. Talukdar and Pal (2017), accounted for the construction of the Komardanga dam over the Dhepa river, a major contributing tributary of river Punarbhaba and diversion of water through the canal system is the major reason. Islam et al., (2014) claimed that the construction of Teesta Barrage on the Teesta River is the major human interference. Pal (2015) studied the impact of the Massanjore dam on the Mayurakshi river and accounted for that dam was the main reason for declining the flow in the downstream segment of the river. These studies support the results of this study.

Three alternative methods (PSO-RF, PSO-REPTree, and PSO-M5P) are used for discharge prediction and PSO-REPTree, PSO-RF, PSO-M5P appear as the best fit for average, maximum, and minimum discharge prediction respectively. The predicted results showed that the average highest discharge will be 35810.22 m^3s^{-1} in future, which was less than 40000 m^3s^{-1} in the observed years. By using PSO-RF, the maximum highest discharge will be 41768.31 m^3s^{-1} in the year 2022. By using PSO-M5P, the minimum highest discharge will be 13237.71 m^3s^{-1} in the year 2028, but in observed years, the highest minimum discharge was up to 40000 m^3s^{-1}.

5. Conclusion

Trend pattern, periodicity and prediction are vital statistical features of hydrological alteration of the stream-flow. The underneath conclusions were drawn from this paper:

- For assessing the change of the water discharge and its impacts in Padma River due to FB, ITA, hydrological alteration, and periodicity analysis have been done. In most of the month on average, maximum, minimum discharge; the ITA shows a negative decreasing trend. In the dry season (January-May) the trends are almost negative.
- The RVA analysis for average discharge illustrates that in all the months the discharge is lower than the natural flows. It is found that hydrological alteration has been gradually increased over time and expected to be further altered in the future period. The variations are significant in yearly average discharges and water levels but changes are more significant in yearly minimum discharges and water levels.
- The forecasting of the discharge up to 2030 reveals the decreasing possibility of the streamflow.

Conflicts of Interest- None

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

Adnan, R.M., Yuan, X., Kisi, O., Anam, R., 2017. Improving accuracy of river flow forecasting using LSSVR with gravitational search algorithm. Adv Meteorol 2017, 23. https://doi.org/10.1155/2017/ 2391621

Akhter, S., Eibek, K.U., Islam, S., Islam, A.R.M.T., Shen, S., Chu, R., 2019. Predicting spatiotemporal changes of channel morphology in the reach of Teesta River, Bangladesh using GIS and ARIMA modeling, Quaternary International, 513, 80-94, https://doi.org/10.1016/j.quaint.2019.01.022.

Arévalo-Mejía, R., Leblois, E., Salinas-Tapia, H., et al., 2020. A baseline assessment of hydrological alteration degree for the Mexican catchments at gauged rivers (2016), Sci Total Environ, https://doi.org/10.1016/ j.scitotenv.2020.139041

Blöschl, G., 2016. Predictions in ungauged basins–where do we stand? Proc. IAHS, 373 : 57-60

Breiman, L., 2001. Random forests. Mach Learn 45, 5–32.

Chen, W., Pradhan, B., Li, S., Shahabi, H., Rizeei, H.M., Hou, E. and Wang, S., 2019. Novel hybrid integration approach of bagging-based fisher's linear discriminant function for groundwater potential analysis. Natural Resources Research, 28(4), 1239-1258. https://doi.org/10.1007/s11053-019-09465-w

Devasena, C.L., 2014. Comparative analysis of random forest, REP tree and J48 classifiers for credit risk prediction. International Journal of Computer Applications, 0975-8887.

Dewan, A, Corner, R., Saleem, A., Rahman, M.M., Haider, M.R., Rahman, M.M., Sarker, M. H., 2017. Assessing channel changes of the Ganges-Padma River system in Bangladesh using Landsat and hydrological data, Geomorphology, 276, 257-279, doi: 10.1016/j.geomorph.2016.10.017

Dunne, T., Leopold, L.B., 1978. Water in environmental planning. W. H. Freeman and Co., San Francisco.

Gain, A. K., Giupponi, C., 2014. Impact of the Farakka Dam on Thresholds of the Hydrologic Flow Regime in the Lower Ganges River Basin (Bangladesh). Water 6, 2501–2518.

.

GOUPiLLAUD. P. et al., 1984. Cycle-Octave and Related Transforms in Seismic Signal Analysis. Geoexploration 23, 85-102.

Graf, W. L., 2006. Downstream hydrologic and geomorphic effects of large dams on American rivers. Geomorphology 79, 336–360.

Granata, F., Saroli, M., de Marinis, G., Gargano R., 2018. Machine Learning Models for Spring Discharge Forecasting, Geofluids Volume 2018, Article ID 8328167, 13 pages https://doi.org/10.1155/2018/8328167

Hossain, M.Y., 2010. Morphometric relationships of length-weight and length-length of four Cyprinid small indigenous fish species from the Padma River (NW Bangladesh). Turkish Journal of Fisheries and Aquatic Sciences, 10(1), 131-134

Islam, A.R.M.T., Sein, Z.M.M., Ongoma, V. et al., 2016. Geomorphological and Land Use Mapping: A Case Study of Ishwardi Under Pabna District, Bangladesh, Advances in Research, 4(6): 378- 387. doi: 10.9734/AIR/2015/14149.

Islam, A.R.M.T., 2016. Assessment of fluvial channel dynamics of padma river in northwestern Bangladesh. Universal Journal of Geoscience 4, 41–49. https://doi.org/10. 13189/ujg.2016.040204.

Islam, M.S., Islam, A.R.M.T., Rahman, F., Ahmed, F., Haque, M.N., 2014. Geomorphology and land use mapping of northern part of Rangpur district, Bangladesh. Journal of Geosciences and Geomatics 2 (4), 145–150. https://doi.org/10.12691/jgg-2-4-2.

Kalmegh, S., 2015. Analysis of WEKA Data Mining Algorithm REPTree, SimpleCart and RandomTree for Classification of Indian News, IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 (2).

Kişi, O., Santos, C.A.G., Silva, R.M.D., Kermani, M.Z., 2018. Trend analysis of monthly streamflows using Şen's innovative trend method. GEOFIZIKA 35 (1), 53–68. DOI: 10.15233/gfz.2018.35.3.

Liu, W., Cao, S., Chen, Y., 2016. Seismic Time-Frequency Analysis via Empirical Wavelet Transform. IEEE Geosci. Remote Sensing Lett., 13(1), 28-32.

McCartney, M., 2009. Living with dams: Managing the environmental impacts. Water Policy 11, 121–139.

Mirza, M.M.Q., 1997. Hydrological changes in the Ganges system in Bangladesh in the post-Farakka period, Hydrological Sciences Journal, 42:5, 613-631, DOI: 10.1080/02626669709492062.

Mirza, M.M.Q., 1998. Diversion of the Ganges Water at Farakka and its effects on salinity in Bangladesh. Environ. Manag. 22(5), 711–722.

Mohamed, W.N.H.W., Salleh, M.N.M., Omar, A.H., 2012. A comparative study of reduced error pruning method in decision tree algorithms. Control System, Computing and Engineering (ICCSCE), 2012. IEEE International Conference on. IEEE, pp. 392–397.

Nawfee, S.M., Dewan, A., Rashid, T., 2018. Integrating subsurface stratigraphic records with satellite images to investigate channel change and bar evolution: a case study of the Padma River, Bangladesh. Environmental Earth Sciences 77, 1–14. https://doi. org/10.1007/s12665-018-7264-2

Olden, J.D., Poff, N.L., 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Res. Appl. 19, 101–121.

Pal, S. and Talukdar, S., 2020. Modelling seasonal flow regime and environmental flow in Punarbhaba river of India and Bangladesh. Journal of Cleaner Production, 252, 119724.

Pal, S., 2015. Impact of Massanjore Dam on hydro-geomorphological modification of Mayurakshi River, Eastern India, Environment Development and Sustainability. Springer Science 17.

Pal, S., Talukdar, S., 2018. Assessing the role of hydrological modifcations on land use/land cover dynamics in Punarbhaba river basin of Indo-Bangladesh. Environment, Development and Sustainability. https://doi.org/10.1007/s10668-018-0205-0.

Poff, N.L., Allan, J.D., Bain, M.B., Karr, J.R., Prestegaard, K.L., Brian, D., Sparks, R.E., Stromberg, J.C., Richter, B.D., 1997. The natural flow regime—A paradigm for river conservation and restoration. BioScience 47, 769–784.

Quinlan, J., 1987. Simplifying decision trees, International Journal of Man Machine Studies, 27(3), 221–234.

Quinlan, J.R., 1992. Learning with continuous classes. Proceedings AI'92, 5th Australian Joint Conference on Artificial Intelligence, Adams & Sterling (eds.), World Scientific, Singapore, 343–348.

Razavi, T., Coulibaly, P., 2012. Streamflow prediction in ungauged basins: review of regionalization methods, J. Hydrol. Eng., 18, 958-975

Richter, B.D., Baumgartner, J.V., Powell, J., Braun, D.P., 1996. A method for assessing hydrologic alteration within ecosystems. Conserv. Biol. 10, 1163–1174.

Richter, B.D., Mathews, R., Harrison, D.L., Wigington, R., 2003. Ecologically sustainable water management: managing river flows for ecological integrity. Ecol. Appl. 13 (1), 206–224, http://dx.doi.org/10.1890/ 1051-0761

Sanz, D.B., García del Jalón, D., Gutiérrez Teira, B., VizcaínoMartínez, P., 2005. Basin influence on natural variability of rivers in semi-arid environments. Int. J. River Basin Manag. 3, 247–259.

Sarker, M.H. and Thorne, C.R., 2006. Morphological response of the Brahmaputra–Padma- Lower Meghna river system to the Assam earthquake of 1950. Braided rivers: process, deposits, ecology and management, 21, pp.289-310.

Şen, Z., 2012. Innovative trend analysis methodology. J. Hydrol. Eng. 17, 1042–1046. DOI: 10.1061/(ASCE)HE.1943-5584.0000556.

Sepahvand, A, Singh, B., Sihag, P., Nazari, S.A, Ahmadi, H., Fiz, N.S., 2019. Assessment of the various soft computing techniques to predict sodium absorption ratio (SAR). ISH J Hydraul Eng. https ://doi.org/10.1080/09715 010.2019.15951 85

Sihag, P., Karimi, S.M., Angelaki, A., 2019. Random forest, M5P and regression analysis to estimate the field unsaturated hydraulic conductivity, Applied Water Science, 9, 129 https://doi.org/10.1007/s13201-019-1007-8

Smakhtin, V.U., Shilpakar, R.L., Hughes, D.A., 2006. Hydrology-based assessment of environmental flows: An example from Nepal. Hydrol. Sci. J. 51, 207–222.

Talukdar, S. and Pal, S., 2019. Effects of damming on the hydrological regime of Punarbhaba river basin wetlands. Ecological Engineering, *135*, 61-74.

Talukdar, S. and Pal, S., 2020. Modeling flood plain wetland transformation in consequences of flow alteration in Punarbhaba river in India and Bangladesh. Journal of Cleaner Production, 120767.

Talukdar, S., Pal, S., 2017 Impact of dam on flow regime and flood plain modification in Punarbhaba river basin of Indo-Bangladesh barind tract. Water Conserv. Sci. Eng. https://doi.org/10.1007/s4110 1-017-0025-3.

Tien Bui, D., Shirzadi, A., Amini, A., et a., 2020. A Hybrid Intelligence Approach to Enhance the Prediction Accuracy of Local Scour Depth at Complex Bridge Piers, Sustainability, 12, 1063; doi:10.3390/su12031063

Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., Al-Katheeri, M.M., 2016. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir region, Saudi Arabia. Landslides 13, 839–856