



Article Runoff and Sediment Yield Processes in a Tropical Eastern Indian River Basin: A Multiple Machine Learning Approach

Alireza Moghaddam Nia ¹, Debasmita Misra ^{2,*}, Mahsa Hasanpour Kashani ³, Mohsen Ghafari ⁴, Madhumita Sahoo ⁵, Marzieh Ghodsi ⁶, Mohammad Tahmoures ¹, Somayeh Taheri ¹ and Maryam Sadat Jaafarzadeh ¹

- ¹ Faculty of Natural Resources, University of Tehran, Karaj 3158777871, Iran; a.moghaddamnia@ut.ac.ir (A.M.N.); m.s.jaafarzadeh@ut.ac.ir (M.S.J.); tahmoures@ut.ac.ir (M.T.); stahery@ut.ac.ir (S.T.)
- ² Department of Civil, Geological and Environmental Engineering, College of Engineering and Mines, University of Alaska Fairbanks, P.O. Box 755800, Fairbanks, AK 99775, USA
- ³ Department of Water Engineering, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil 676W+5CX, Iran; m.hkashani@uma.ac.ir
- ⁴ Department of Range and Watershed Management, Faculty of Water and Soil, University of Zabol, Zabol 3585698613, Iran; m.ghafari@frw.ir
- ⁵ Department of Mining and Geological Engineering, College of Engineering and Mines, University of Alaska Fairbanks, P.O. Box 755800, Fairbanks, AK 99775, USA; sahoomadhu1989@gmail.com
- ⁶ Faculty of Geography, University of Tehran, Tehran 1417853933, Iran; mghodsi@ut.ac.ir
- Correspondence: dmisra@alaska.edu; Tel.: +1-907-474-5339

Abstract: Tropical Indian river basins are well-known for high and low discharges with high peaks of flood during the summer and the rest of the year, respectively. A high intensity of rainfall due to cyclonic and monsoon winds have caused the tropical Indian rivers to witness more runoff. These rivers are also known for carrying a significant amount of sediment load. The complex and non-linear nature of the sediment yield and runoff processes and the variability of these processes depend on precipitation patterns and river basin characteristics. There are a number of other elements that make it difficult to forecast with great precision. The present study attempts to model rainfall-runoff-sediment yield with the help of five machine learning (ML) algorithms-support vector regression (SVR), artificial neural network (ANN) with Elman network, artificial neural network with multilayer perceptron network, adaptive neuro-fuzzy inference system (ANFIS), and local linear regression, which are useful in river basins with scarce hydrological data. Daily, weekly, and monthly runoff and sediment yield (SY) time series of Vamsadhara river basin, India for a period from 1 June to 31 October for the years 1984 to 1995 were simulated using models based on these multiple machine learning algorithms. Simulated results were tested and compared by means of three evaluation criteria, namely Pearson correlation coefficient, Nash-Sutcliffe efficiency, and the difference of slope. The results suggested that daily and weekly predictions of runoff based on all the models can be successfully employed together with precipitation observations to predict future sediment yield in the study basin. The models prepared in the present study can be helpful in providing essential insight to the erosion-deposition dynamics of the river basin.

Keywords: tropical Indian river; simulating runoff and sediment yield; SVM; MLP; Elman; ANFIS; LLR

1. Introduction

The erosion–deposition dynamics in a river basin at both spatial and temporal scales play a vital role in shaping the hydrological processes occurred in the river basin [1–3]. Tropical Indian rivers are characterized by high and very low discharge during the summer and other months, respectively [4]. These rivers are also known for carrying sediment load to the oceans [5–8]. Reservoirs and channels associated with tropical rivers experience a



Citation: Moghaddam Nia, A.; Misra, D.; Kashani, M.H.; Ghafari, M.; Sahoo, M.; Ghodsi, M.; Tahmoures, M.; Taheri, S.; Jaafarzadeh, M.S. Runoff and Sediment Yield Processes in a Tropical Eastern Indian River Basin: A Multiple Machine Learning Approach. *Land* **2023**, *12*, 1565. https://doi.org/10.3390/ land12081565

Academic Editor: Carlos Rogério Mello

Received: 20 May 2023 Revised: 17 July 2023 Accepted: 27 July 2023 Published: 7 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

2 of 15

degraded surface water quality. Therefore, runoff and sediment yield simulation in shortterm and long-term periods in tropical rivers is of crucial importance in water resource management projects. The complex, nonlinear, and variability nature of the sediment yield (SY) and runoff processes depend on precipitation patterns and river basin characteristics. There are a number of other elements that make it difficult to forecast with great precision. In recent decades, different types of models have been developed by hydrologic modelers and professionals. The physically based models simulate the runoff and SY, if enough data exist. The lack of sufficient input data restricts the using of these models in low-monitoring areas. Hence, a number of models have been developed, which enable us to simulate SY and runoff with the existing dataset. Artificial neural networks and soft computing techniques have proven to be decent alternatives and have been employed by researchers for modeling rainfall–runoff–sediment yield processes [1,3,5,9–28]. Data-scarce river basins can take advantage of these popular soft-computing techniques for modeling purposes.

Machine learning models have the ability to effectively capture chaotic and hidden time-series patterns. Precipitation dynamics combined with landscape attributes control the runoff and sediment yield. Tropical Indian rivers are known to receive their precipitation from monsoon winds. Tropical cyclones also bring highly intensive flooding rainfalls. These basins cause high runoff irrespective of basin area [29,30]. These high intensity rainfall events result in high runoff and soil erosion. Sediment yield is, thus, very much related to the rainfall pattern and intensity [31]. In the present study, a tropical eastern Indian river basin, Vamsadhara river basin, was chosen as a case study and the rainfall-runoff-sediment yield dynamics were modeled based on the existing dataset. A comparative analysis among five ML algorithms, including SVR, ANN with Elman network (Elman), ANN with multilayer perceptron network (MLP), ANFIS and local linear regression (LLR), was made to identify the efficiency of these methods in capturing the complex runoff–SY process. The primary objective of this article is to evaluate performances of SVM, ANNs (MLP and Elman), ANFIS, and LLR techniques and compare them in terms of statistical criteria, emulating runoff and sediment yield at daily, weekly, and monthly timescales. Ref. [23]'s dataset was applied to a basin runoff and SY simulation to compare the performance of SVM with other machine learning algorithms as stated before. The tropical eastern Indian river basin is an important river for prospective river-linking projects in India. Devising a model based on historical data can help policymakers and decisionmakers to have a good understanding of infrastructure development in the basin area from an erosion-deposition perspective.

2. Materials and Methods

2.1. Description of Methods

The relationships among rainfall, runoff, and sediment yield at different temporal scales were analyzed using multiple machine learning algorithms. The methods employed in this study are as follows:

- (1) Support Vector Regression (SVR): Support vector machines (SVMs) have been well described by [32–36]. Learning in SVMs is very robust from the point of view of the precision of the computations [37,38]. Ref. [23] stated that, in SVR, employing a nonlinear mapping function (kernel function) results in having to work in a higher dimensional feature space, consequently overcoming the dimensionality issue. [39,40] proved that the Gaussian radial basis kernel function (GRKF) is superior and more common than kernel function.
- (2) Artificial Neural Networks (ANNs): A learning process makes neural networks able to conduct information processing. This process is a link to weights adaptation, which causes producing an approximate output(s) by the network. Indeed, an ANN learning process will improve the system's correct response to an input via the development of a strengthened current matrix of nodal weights [41]. We used two different networks for this study:

MLP Networks: They are the most common and popular of all ANNs. MLP networks have 3 layers or more: an input layer, which is used to present data to the network; an output layer, which is used to produce an appropriate response(s) to the given input; and one or more intermediate layers, which are used to act as a collection of feature detectors (Figure 1). MLP networks have been successfully used earlier in hydrological studies in different river basins across the world [11,42–48].



Figure 1. Illustration of an MLP network.

Elman Networks: They are a type of recurrent neural network with connections back to a specific copy layer from its hidden layer (Figure 2). This means that the network's learned function can be based on the current inputs as well as a history of the network's past state(s) and outputs. In other words, an Elman network is a finite state machine that learns which states to remember (i.e., what is relevant). Because the particular copy layer is considered as just another set of inputs, normal back-propagation learning techniques may be utilized (something that recurrent networks cannot perform). A duplicate of the concealed layer units is created to a copy layer at each time step [49]. Elman networks are an extension of two-layer sigmoid linear architecture [50,51]. Hence, they inherit the capability to fit any input/output function with a finite number of discontinuities. They can also fit temporal patterns, although a complicated function may need many neurons in the recurrent layer. In addition, the recurrent model takes significantly longer to train than the MLP model because of its more complicated design [52]. The use of Elman network in hydrological studies is, however, very limited [53,54].

- (3) Adaptive Neuro-Fuzzy Inference System (ANFIS): [55] proposed ANFIS, a universal approximator capable for approximating any actual continuous function on a compact set to any degree of precision. ANFIS is made up of nodes that have defined tasks or obligations, which are organized into layers with specified functions. To select a set of parameters, it employs a hybrid learning method that combines back-propagation gradient descent error digestion and the least-squares technique. It may be used to build a collection of fuzzy "If–Then" rules with suitable membership functions in order to produce the input–output pairings that have been pre-specified [56–58]. Earlier, ANFIS has been used in several hydrological studies successfully [59–63]. The Gaussian membership function is adopted for the present study. For details of theoretical background, interested readers can refer to [55].
- (4) Local Linear Regression (LLR): Like neural network models, LLR is a non-parametric method that does not require to be trained. Many difficulties with low-dimensional forecasting and smoothing have been solved using the LLR method. To create an initial forecast, the LLR technique simply requires three data points, after which it uses all newly updated data as it is able to make future forecasts [64]. LLR uses a set

of P_{max} closest points to a query point to perform linear regression, which results in the creation of a linear model in the query point's neighborhood [65]. The kd-tree (k-dimensional tree) technique, a rapid closest neighbor search algorithm detailed in [66], is used to calculate the P_{max} nearest points to the query point. The linear matrix problem is solved when given a neighborhood of P_{max} points:

$$Xm = y$$
 (1)

where $X = p_{\text{max}} \times d$ matrix of the P_{max} , y = column vector of length P_{max} , and m = column vector of parameters that must be determined to provide the optimal mapping from X to Y, such that:



Figure 2. Illustration of an Elman network.

The number of linearly independent rows in the matrix X determines whether or not the solution for m exists or is unique [64,65].

If the matrix *X* is square and non-singular, then the unique solution to Equation (1) is $m = X^{-1}y$. If *X* is not square or singular, we modify Equation (1) and attempt to find a vector *m*, which minimizes:

$$\left| Xm - y \right|^2 \tag{3}$$

Ref. [67] demonstrated that a pseudo-inverse matrix is the only answer to this problem [64,65,67]. The LLR technique has the benefit of enabling us to carry out statistical modeling locally with a small quantity of sample data. In locations of high data density in the input space, LLR can provide extremely accurate predictions. Ref. [66] can be consulted for more information on LLR.

2.2. Model Performance Indices

Three performance indices—Pearson's correlation coefficient (*r*), Nash–Sutcliffe efficiency (E), and the difference of slope (S_{Diff})—were used in the present study for evaluating the performance of the models. Nash–Sutcliffe efficiency indicator has been frequently used to assess the effectiveness of hydrologic models. E value can vary between $-\infty$ and 1, where an efficiency of 1 indicates that the observed and estimated values are in complete agreement. E = 0 shows that all estimated values are equal to the observed values' mean. The mean of the observed values is a better predictor than the predicted values when E is negative. The efficiency coefficient outperforms correlation-based sensitivity requirements for observed and calculated means and variances, but is too sensitive to outliers [68,69]. The idea behind S_{Diff} is that, although r indicates a model's variational accountability and E represents its efficiency, when observed and projected values are compared, there is no corresponding metric for the degree of predictability of a best-fit model to the 1:1 line. As a consequence, S_{Diff} was defined as the distance between the slope of a best-fit line on a scatter plot of expected vs. observed data for a particular model and the 1:1 line on a scatter plot of predicted vs. observed data. An S_{Diff} of 0% means that the scatter plot's best-fit line is parallel to the 1:1 line, guaranteeing that the best-fit linear model is fully predictable. If S_{Diff} is 100 percent, the best-fit line is the average line with a zero slope. The scatter plot's best-fit linear model would exaggerate the low observed values while underestimating the high ones if S_{Diff} was between 0% and 100%. According to a negative S_{Diff} score, the best-fit linear model of the scatter plot would underestimate the low observed values and overestimate the high ones. The mathematical representation of the performance indices are as follows:

$$r = \frac{\sum_{i=1}^{n} \left[(Q_{pi} - Q_{pavg}) (Q_{oi} - Q_{oavg}) \right]}{\sqrt{\left[\left(\sum_{i=1}^{n} (Q_{pi} - Q_{pavg})^2 \right) \left(\sum_{i=1}^{n} (Q_{oi} - Q_{oavg})^2 \right) \right]}}$$
(4)

$$E = 1 - \frac{\sum_{i=1}^{n} (Q_{pi} - Q_{oi})^2}{\sum_{i=1}^{n} (Q_{oi} - Q_{oavg})}$$
(5)

where, Q_{oi} and Q_{pi} are the observed and predicted (or simulated) values, respectively. Q_{oavg} and Q_{pavg} are the mean values of the observed and predicted values, respectively.

2.3. Study Area and Data Used

The present study was conducted in the Vamsadhara river basin, which is located between the Mahanadi and Godavari river basins in eastern India (Figure 3). The region lies between 18°15′ and 19°55′ north latitudes, and 83°20′ and 84°20′ east longitudes. The south–west monsoon and the rare cyclones generated owing to the depression in the Bay of Bengal impact the precipitation in the basin from June to October.

There are two types of climates in this river basin. The coastal area has a semi-arid climate, while the upper reaches have a dry sub-humid climate. The Vamsadhara river rises near Lanjigarh in Odisha and flows for 254 km before joining the BoB at Kalingapatnam in Andhra Pradesh. It has a basin area of 10,450 square kilometers. The average rainfall amount in the basin is 940.2 mm near the coast, 1551.6 mm in the northeast, and 1250.2 mm in the northwest [70]. For the Vamsadhara river basin, the annual average maximum temperature is about 32 °C and minimum temperature is about 21 °C. The elevation ranges in the Vamsadhara river basin from 10 m above MSL in the south near the coast to 1545 m in the northwest (hills near Bissam Cuttack). The reservoirs of Badnalla and Harabhangi are located within the Kashinagar gauge station, while the Gotta barrage is located outside the gauge station. The annual river flow and sediment load are about 5943.56 m³/s and 7511.54 kg/s, respectively. Clayey loam soils cover the majority of the Vamsadhara river basin. Over this basin, major land is occupied by forests (52%) and agricultural lands (30%).



Figure 3. Location of Vamsadhara river basin in India and location of rain gauge sites with Thiessen polygons within the river basin.

The study basin includes six rain gauge stations (Kutragada, Gudari, Kashinagar, Gotta Barrage, Mohana, and Mahandragarh) and the weighted rainfall for the study region was estimated using the Thiessen polygon method, as shown in Figure 3. Daily rainfall, runoff, and SY data were obtained for a period from 1 June to 31 October for the years 1984 to 1995. The precipitation from monsoon winds or cyclonic depressions is concentrated during these months in the study basin. The rainfall, runoff, and SY time series of the monsoon period (June–October) for the years of 1984–1987 was employed to train the SVM, ANN, ANFIS, and LLR models, and the time series of 1988–1989 and 1992–1995 were considered for their testing and validation, respectively. The input variables with different time lags selected for the SVM, ANN, ANFIS, and LLR models based on the correlation coefficient method for predicting runoff and SY (daily, weekly, and monthly) are given in Table 1.

Table 1. Input variables used for SVM, ANN, ANFIS, LLR, and MRPRT models (Ri = total rainfall in mm; Qu = runoff in m^3/s ; Sy = sediment yield in kg/s; t = current day; I/P = input; and O/P = output).

		Input – Output Variables with Lag Times						
Time Scale	Variable	Runoff				Sediment Yield		
		t	t – 1	t – 2	t – 3	t	t – 1	
	Ri	I/P	I/P	-	-	I/P	I/P	
Daily	Qu	O/P	I/P	I/P	I/P	I/P	-	
-	Sy	-	-	-	-	O/P	I/P	
Weekly	Ri	I/P	I/P	-	-	I/P	-	
	Qu	O/P	I/P	-	-	I/P	-	
-	Sy	-	-	-	-	O/P	-	

Time Scale	Variable	Input – Output Variables with Lag Times							
		Runoff				Sediment Yield			
	-	t	t-1	t – 2	t – 3	t	t-1		
	Ri	I/P	I/P	I/P	-	I/P	-		
Monthly	Qu	O/P	-	-	-	I/P	-		
	Sy	-	-	-	-	O/P	-		

Table 1. Cont.

3. Results and Discussion

All the datasets were normalized between 0 and 1 prior to feeding to the models. According to Table 1, daily values of total rainfall with 0 to 1 lag days (Ri(t) and Ri(t - 1)), daily values of runoff with 1 to 3 lag times (Qu(t - 1), Qu(t - 2), Qu(t - 3)), daily values of sediment yield with 0 to 1 lag time (Sy(t) and Sy(t - 1)), weekly values of total rainfall with 0 to 1 lag week (Ri(t) and Ri(t - 1)), weekly values of runoff with 1 lag week (Qu(t - 1)), current week sediment values (Sy(t)), monthly values of total rainfall with 0 to 2 lag months (Ri(t), Ri(t - 1), Ri(t - 2)), and monthly values of runoff (Qu(t)) and sediment (Sy(t)) on the current month were used as the inputs of the models for predicting current runoff and sediment (as the output variables).

Four years of data (1984–1987) were used to train the models and two years of data (1988–1989) were utilized to test the models' performance. The data over the next four years (1992–1995) were forecasted by the models. The performance metrics (r, E, and S_{Diff}) were used to compare the observed and estimated data from 1992 to 1995. All the predictions and calculations were conducted in MATLAB 2016a software.

Runoff prediction: The validation and testing results obtained for runoff at the three different time scales—daily, weekly, and monthly—for all the models are given in Table 2.

		Testing (1988–1989)		Validation (1992–1995) *			
Rentin Wouchs			r (%)	E (%)	r (%)	E (%)	Sdiff (%)
	SVM		86.1	72.6	92.10	80.02	33.79
	ANTNI	MLP	86.3	73.3	90.1	72.7	40.15
Daily	AININS	Elman	87.66	76.13	92.99	83.94	25.53
	ANFIS		89.64	79.99	94.62	88.78	14.35
	LLR		83.87	68.80	94.40	89.01	9.12
	SVM		80.67	62.03	91.16	67.55	47.0
	ANNs	MLP	79.6	60.3	87.4	54.6	56.95
Weekly		Elman	77.57	39.79	90.35	79.84	27.15
	ANFIS		87.10	75.55	90.99	77.05	28.96
	LLR		71.39	47.51	94.44	88.67	16.79
	SVM		81.47	24.31	86.33	14.13	60.09
	ANNs	MLP	77.4	26.0	79.3	-4.2	74.98
Monthly		Elman	73.82	41.05	75.53	45.98	42.61
	ANFIS		88.75	78.93	80.00	39.24	11.6
	LLR		90.98	67.74	86.62	69.18	33.77

Table 2. Comparison of the performance of SVM, ANN, ANFIS, and LLR models for daily, weekly, and monthly runoff prediction in testing (1988–1989) and validation (1992–1995) periods.

* The best performance index for a particular model for the validation data has been shown in bold.

At daily scale, the LLR model outperforms the other models based on the E (89.01%) and sdiff (9.12%) criteria. However, the ANFIS model shows a better performance than the LLR according to the r value (94.62%). Among the intelligent models, the ANFIS model performs successfully based on all the criteria. The Elman model follows the ANFIS as a second best intelligent model.

At weekly scale, the LLR model outperforms the other models based on the r (94.44%), E (88.67%), and sdiff (16.79%) criteria. Among the intelligent models, the Elman model performs accurately based on the E (79.84%) and sdiff (27.15%) criteria. The ANFIS model follows the Elman as a second best intelligent model.

At monthly scale, the LLR model outperforms the other models based on the r (86.62%) and E (69.18%) criteria. However, the ANFIS model shows a better performance than the LLR according to the sdiff value (11.6%). Among the intelligent models, the ANFIS model performs successfully based on the r (80%) and sdiff criteria. The Elman model follows the ANFIS as a second best intelligent model.

In general, the ANFIS and Elman techniques showed high ability in daily, weekly, and monthly runoff estimating. However, the non-intelligent LLR model showed notable accuracy in its estimations at different time scales.

From daily to monthly scales, the MLP model revealed a 10.8% reduction in performance based on the r criterion. A similar trend can be seen for the E performance metric of 76.9% for the MLP. The Elman model provided a reduced performance of 17.46% from daily to monthly predictions based on the r measure. A similar decreasing trend of 37.96% is observed according to the E performance measure for the Elman neural network model. ANFIS model showed a reducing performance of 14.62% and 49.54% in the r and E values, respectively, from daily to monthly scales. The LLR model performance (r criterion) for weekly runoff predictions was slightly better than daily runoff values, but the LLR provided a reduced performance of 7.82% between weekly and monthly predictions. A decreasing trend is observed from daily through monthly predictions based on the reduced E measure (about 19.83%).

In general, all of the applied intelligent and non-intelligent models showed a decreasing performance for runoff predictions from daily to monthly scales.

Figures 4–6 show the scatter plots of the models for runoff prediction at daily, weekly, and monthly scales, respectively. According to these figures, the LLR model is closer to the line 1:1 than the other models at daily and weekly scales. At monthly scale (Figure 6), the ANFIS model is close to line 1:1. The ANFIS and Elman models are closer to line 1:1 than the other intelligent models.



Figure 4. Scatter plot of the observed vs. estimated daily runoff (1992–1995) using SVM, ANN, ANFIS, and LLR methods.



Figure 5. Scatter plot of the observed vs. estimated weekly runoff (1992–1995) using SVM, ANN, ANFIS, and LLR methods.



Figure 6. Scatter plot of the observed vs. estimated monthly runoff (1992–1995) using SVM, ANN, ANFIS, and LLR methods.

Sediment yield prediction: The same models were applied in order to predict daily, weekly, and monthly sediment yields in the study basin. Testing and validation results are presented in Table 3.

Sediment Yield Models		Testing (1988–1989)		Validation (1992–1995) *			
			r (%)	E (%)	r (%)	E (%)	Sdiff (%)
	SVM		80.02	62.94	87.87	75.68	30.46
Daily	ANNs	MLP	79.30	62.80	83.20	68.00	21.18
		Elman	77.63	53.01	87.99	36.30	-29.61
	ANFIS		88.10	76.87	87.17	73.09	12.11
	LLR		93.20	86.25	82.18	66.78	27.03
	SVM		78.43	60.26	88.08	74.62	34.84
	ANNs	MLP	80.20	64.10	75.10	51.80	45.90
Weekly		Elman	79.75	57.98	90.03	43.69	-21.97
	ANFIS		83.12	67.19	83.13	67.55	79.15
	LLR		82.05	66.10	89.88	78.99	12.43
Monthly	SVM		81.32	62.44	87.66	74.49	24.76
	ANNs	MLP	89.40	79.10	74.10	53.70	52.06
		Elman	89.13	79.31	88.99	76.25	9.01
	ANFIS		99.94	99.81	85.47	-102.74	-62.28
	LLR		92.10	70.19	52.77	16.84	75.61

Table 3. Comparison of the performance of SVM, ANN, ANFIS, and LLR models for daily, weekly, and monthly sediment yield prediction in testing (1988–1989) and validation (1992–1995) periods.

* The best performance index for a particular model for the validation data are emboldened.

At validation stage and daily scale, the Elman model with an r value of 87.99%, SVM model with an E value of 75.68%, and the ANFIS model with an sdiff value of 12.11% outperform the other models. In general, the ANFIS model performs more precisely than the others based on all the criteria.

At weekly scale, the LLR model outperforms the other models based on the E (78.99%) and sdiff (12.43%) criteria values. However, the Elman model performs well based on the r measure (90.03%).

At monthly scale, the intelligent Elman model estimates sediment yield with a higher accuracy than the other models based on the r (88.99%), E (76.25%), and sdiff (9.01%) criteria. The SVM model follows the Elman as the second best model based on all the criteria.

From daily to monthly scales, the MLP model showed a 9.1% reduction in performance based on the r criteria. A similar pattern can be seen with the E and sdiff performance metrics for the MLP model. As the nonlinearity in the time series reduces, the MLP model's performance accuracy considerably reduces (from daily to monthly scales). The Elman model provided an increased r value of 1% from daily to monthly time scales. The E performance measure in the Elman neural network model shows a similar pattern with an increase of 39.95%. From daily to monthly scales, the ANFIS model showed a 1.7% reduction in performance (r criterion). A similar tendency can be seen for the E performance metric, while the E value of the ANFIS model for monthly predictions was considerably reduced, down to a negative value, which is not in accordance with Ref. [71]. The LLR model provided an increased r value of 7.7% between daily and weekly predictions, but a decreased value of 37.11% between weekly and monthly predictions. A similar trend with an increased E value of the LLR model significantly decreased down to 62.15% from weekly to monthly scale.

In general, the performance accuracy of all the models in sediment yield prediction reduces when the time scale is increased.

Figures 7–9 illustrate the scatter plots of the applied models at the validation stage for sediment yield prediction at daily, weekly, and monthly scales, respectively. According to these models, the ANFIS, LLR, and Elman models are closer to line 1:1 than the other models at daily, weekly, and monthly scales, respectively. The MLP, Elman, and SVM



models are the second best models based on the short distance from line 1:1, at daily, weekly, and monthly scales, respectively.

Figure 7. Scatter plot of the observed vs. estimated daily sediment yield (1992–1995) using SVM, ANN, ANFIS, and LLR methods.



Figure 8. Scatter plot of the observed vs. estimated weekly sediment yield (1992–1995) using SVM, ANN, ANFIS, and LLR methods.



Figure 9. Scatter plot of the observed vs. estimated monthly sediment yield (1992–1995) using SVM, ANN, ANFIS, and LLR methods.

In general, it can be concluded that the intelligent models have high ability in sediment yield, estimating at different time scales.

For both runoff and sediment yield prediction, the intelligent Elman model showed less error in its estimations at different time scales. The SVM and MLP models were more suitable for sediment yield prediction than the runoff prediction. The ANFIS and LLR models were more suitable for runoff prediction than the sediment yield prediction. Since the intelligent models are more sensitive toward their input variables, input data length, and their training parameter values, considering unsuitable variables or values can affect the models' performance and increase errors more than the non-intelligent models would, such as LLR. The intelligent models were more successful than the LLR model in sediment yield prediction because of the high nonlinearity degree of the rainfall–runoff–sediment process. By increasing the time scales and reducing the nonlinearity degree of the processes and also input data length, the accuracy of all the applied models decreased. The findings of this study are in accordance with most of the previous research [46,72].

4. Conclusions

The present study attempted to establish the effectiveness of machine learning algorithms, SVM, MLP, Elman, and ANFIS, to capture variability in rainfall–runoff–sediment yield dynamics in a tropical Indian river basin. Runoff and sediment yield were estimated on three time scales, daily, weekly, and monthly, with different input combinations and lag times. The performance of the intelligent models were compared with the non-intelligent LLR model using three performance criteria: r, E, and sdiff. The results showed that the ANFIS and LLR models were more suitable for runoff prediction than the sediment yield prediction. The SVM and MLP models were more suitable for sediment yield prediction than the runoff prediction. For both runoff and sediment yield prediction, the intelligent Elman model showed high ability at different time scales. The intelligent models were more successful than the LLR model in sediment yield prediction. By increasing the time scales, the accuracy of all the applied models decreased. For increasing the models accuracy, it is recommended to use other hydrological data such as temperature, evaporation, and different lag times. Moreover, it is recommended to apply different optimization algorithms such as genetic algorithm for the optimal determination of the intelligent models' parameters, improving their performance. Other intelligent models such as random forest are also recommended to apply and evaluate their efficiency in simulating the studied processes.

The models applied in the present study can be helpful in providing essential insight to the erosion–deposition dynamics of the Vamsadhara river basin. This study established the robustness of the multiple machine learning algorithms in capturing the variability of hydrological events in a tropical river basin.

Author Contributions: Conceptualization, D.M., A.M.N. and M.S.; methodology, D.M., A.M.N., M.H.K., M.G. (Mohsen Ghafari) and M.S.; software, D.M., A.M.N., M.H.K., M.G. (Mohsen Ghafari) and M.S.; data curation, M.H.K., M.G. (Mohsen Ghafari) and M.S.; writing—original draft preparation, D.M., A.M.N., M.H.K., M.G. (Mohsen Ghafari), M.S., M.S.J., M.T., S.T. and M.G. (Marzieh Ghodsi); writing—review and editing, D.M., A.M.N., M.H.K., M.G., M.S., M.S.J., M.T., S.T. and M.G. (Marzieh Ghodsi); visualization, D.M., A.M.N., M.H.K., M.G. (Mohsen Ghafari), M.S., M.S.J., M.S.J., M.T., S.T. and M.G. (Marzieh Ghodsi); project administration, D.M. and A.M.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be made available on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Bharti, B.; Pandey, A.; Tripathi, S.K.; Kumar, D. Modelling of runoff and sediment yield using ANN, LS-SVR, REPTree and M5 models. *Hydrol. Res.* 2017, 48, 1489–1507.
- Bhat, S.A.; Dar, M.U.D.; Meena, R.S. Soil erosion and management strategies. In Sustainable Management of Soil and Environment; Springer: Singapore, 2019; pp. 73–122.
- Bhat, S.A.; Hamid, I.; Dar, M.U.D.; Rasool, D.; Pandit, B.A.; Khan, S. Soil erosion modeling using RUSLE & GIS on micro watershed of J&K. J. Pharmacogn. Phytochem. 2017, 6, 838–842.
- 4. Latrubesse, E.M.; Stevaux, J.C.; Sinha, R. Tropical rivers. *Geomorphology* 2005, 70, 187–206.
- 5. Li, L.; Ni, J.; Chang, F.; Yue, Y.; Frolova, N.; Magritsky, D.; Borthwick, A.G.L.; Ciais, P.; Wang, Y.; Zheng, C.; et al. Global trends in water and sediment fluxes of the world's large rivers. *Sci. Bull.* **2020**, *65*, 62–69.
- Li, Z.; Xu, X.; Zhu, J.; Zhong, F.; Xu, C.; Wang, K. Can precipitation extremes explain variability in runoff and sediment yield across heterogeneous karst watersheds? J. Hydrol. 2020, 596, 125698.
- Milliman, J.D.; Farnsworth, K.L. River Discharge to the Coastal Ocean: A Global Synthesis; Cambridge University Press: Cambridge, UK, 2013.
- Straume, E.O.; Gaina, C.; Medvedev, S.; Hochmuth, K.; Gohl, K.; Whittaker, J.M.; Abdul Fattah, R.; Doornenbal, J.C.; Hopper, J.R. GlobSed: Updated total sediment thickness in the world's oceans. *Geochem. Geophys. Geosyst.* 2019, 20, 1756–1772.
- Afan, H.A.; El-Shafie, A.; Yaseen, Z.M.; Hameed, M.M.; Mohtar, W.H.M.W.; Hussain, A. ANN based sediment prediction model utilizing different input scenarios. *Water Resour. Manag.* 2015, 29, 1231–1245.
- 10. Agarwal, A.; Mishra, S.K.; Ram, S.; Singh, J.K. Simulation of runoff and sediment yield using artificial neural networks. *Biosyst. Eng.* **2006**, *94*, 597–613.
- 11. Agarwal, A.; Singh, R.D.; Mishra, S.K.; Bhunya, P.K. ANN-based sediment yield models for Vamsadhara river basin (India). *Water SA* 2005, *31*, 85–100.
- 12. Alp, M.; Cigizoglu, H.K. Suspended sediment load simulation by two artificial neural network methods using hydrometeorological data. *Environ. Model. Softw.* 2007, 22, 2–13.
- 13. Choubin, B.; Darabi, H.; Rahmati, O.; Sajedi-Hosseini, F.; Kløve, B. River suspended sediment modelling using the CART model: A comparative study of machine learning techniques. *Sci. Total Environ.* **2018**, *615*, 272–281.
- 14. Doroudi, S.; Sharafati, A.; Mohajeri, S.H. Satellite-based estimation of daily suspended sediment load using hybrid intelligent models. *Hydrol. Sci. J.* **2023**, *68*, 307–324.
- 15. Karami, H.; DadrasAjirlou, Y.; Jun, C.; Bateni, S.M.; Band, S.S.; Mosavi, A.; Moslehpour, M.; Chau, K.-W. A novel approach for estimation of sediment load in Dam reservoir with hybrid intelligent algorithms. *Front. Environ. Sci.* **2022**, *10*, 165.
- 16. Kakaei-Lafdani, E.; Nia, A.M.; Ahmadi, A. Daily suspended sediment load prediction using artificial neural networks and support vector machines. *J. Hydrol.* **2013**, *478*, 50–62.
- 17. Kisi, Ö. Constructing neural network sediment estimation models using a data-driven algorithm. *Math. Comput. Simul.* **2008**, *79*, 94–103.

- 18. Kisi, O.; Dailr, A.H.; Cimen, M.; Shiri, J. Suspended sediment modeling using genetic programming and soft computing techniques. *J. Hydrol.* **2012**, *450*, 48–58.
- 19. Kumar, D.; Pandey, A.; Sharma, N.; Flügel, W.A. Daily suspended sediment simulation using machine learning approach. *Catena* **2016**, *138*, 77–90.
- Kumar, R.; Kumar, M.; Shah, A.I.; Bhat, S.A.; Wani, M.A.; Ram, D. Modelling of soil loss using USLE through Remote Sensing and Geographical Information System in micro-watershed of Kashmir valley, India. J. Soil Water Conserv. 2016, 15, 40–45.
- Makarewicz, A.N. Quantifying Intermittent Streamflow with Low-Cost Sensors and Fully Integrated Hydrological Modelling of Upstream Runoff. Ph.D. Thesis, University of Adelaide, Adelaide, Australia, 2020.
- Makarynskyy, O.; Makarynska, D.; Rayson, M.; Langtry, S. Combining deterministic modelling with artificial neural networks for suspended sediment estimates. *Appl. Soft Comput.* 2015, 35, 247–256.
- 23. Misra, D.; Oommen, T.; Agarwal, A.; Mishra, S.K.; Thompson, A.M. Application and analysis of support vector machine-based simulation for runoff and sediment yield. *Biosyst. Eng.* **2009**, *103*, 527–535.
- Mustafa, M.R.; Isa, M.H.; Rezaur, R.B. A comparison of artificial neural networks for prediction of suspended sediment discharge in river-a case study in Malaysia. World Acad. Sci. Eng. Technol. 2011, 81, 372–376.
- Partal, T.; Cigizoglu, H.K. Estimation and forecasting of daily suspended sediment data using wavelet–neural networks. J. Hydrol. 2008, 358, 317–331.
- Peña-Angulo, D.; Nadal-Romero, E.; González-Hidalgo, J.C.; Albaladejo, J.; Andreu, V.; Bagarello, V.; Zorn, M. Spatial variability of the relationships of runoff and sediment yield with weather types throughout the Mediterranean basin. *J. Hydrol.* 2019, 571, 390–405.
- 27. Yilmaz, B.; Aras, E.; Nacar, S.; Kankal, M. Estimating suspended sediment load with multivariate adaptive regression spline, teaching-learning based optimization, and artificial bee colony models. *Sci. Total Environ.* **2018**, *639*, 826–840.
- Zhu, Y.M.; Lu, X.X.; Zhou, Y. Suspended sediment flux modeling with artificial neural network: An example of the Longchuanjiang River in the Upper Yangtze Catchment, China. *Geomorphology* 2007, 84, 111–125.
- 29. Panigrahi, S.; Behera, H.S. A study on leading machine learning techniques for high order fuzzy time series forecasting. *Eng. Appl. Artif. Intell.* **2020**, *87*, 103245.
- 30. Syvitski, J.P.; Cohen, S.; Kettner, A.J.; Brakenridge, G.R. How important and different are tropical rivers?—An overview. *Geomorphology* **2014**, 227, 5–17.
- 31. Puntenney-Desmond, K.C.; Bladon, K.D.; Silins, U. Runoff and sediment production from harvested hillslopes and the riparian area during high intensity rainfall events. *J. Hydrol.* **2020**, *582*, 124452.
- 32. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction; Springer: New York,* NY, USA, 2003; 533p, ISBN 0-387-95284-5.
- Kecman, V. Learning and Soft Computing, Support Vector Machines, Neural Network and Fuzzy Logic Models; MIT Press: Cambridge, MA, USA, 2000; 608p, ISBN 0-262-11255-8.
- Meyer, D.; Wien, F.T. Support vector machines. In *The Interface to Libsvm in Package e1071*; 2015; Volume 28, p. 597. Available online: https://cran.r-project.org/web/packages/e1071/e1071.pdf (accessed on 26 July 2023).
- 35. Vapnik, V.N. The Nature of Statistical Learning Theory; Springer: New York, NY, USA, 1995.
- 36. Vapnik, V.N. Statistical Learning Theory; John Wiley and Sons: New York, NY, USA, 1998.
- Anguita, D.; Boni, A.; Ridella, S. Learning algorithm for nonlinear support vector machines suited for digital VLSI. *Electron. Lett.* 1999, 35, 1349–1350.
- 38. Zhang, X.D. Support vector machines. In A Matrix Algebra Approach to Artificial Intelligence; Springer: Singapore, 2020; pp. 617–679.
- Deborah, S.A.; Mirnalinee, T.T.; Rajendram, S.M. Emotion Analysis on Text Using Multiple Kernel Gaussian. *Neural Process. Lett.* 2021, 53, 1187–1203.
- 40. Keerthi, S.S.; Lin, C.-J. Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel. *Neural Comput.* 2003, 15, 1667–1689.
- 41. Smith, J.; Eli, R.N. Neural-network models of rainfall-runoff process. J. Water Resour. Plan. Manag. 1995, 121, 499–508.
- 42. Hussain, D.; Khan, A.A. Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan. *Earth Sci. Inform.* **2020**, *13*, 939–949.
- 43. Nourani, V.; Kalantari, O. Integrated artificial neural network for spatiotemporal modeling of rainfall–runoff–sediment processes. *Environ. Eng. Sci.* 2010, 27, 411–422.
- Nourani, V.; Kisi, Ö.; Komasi, M. Two hybrid artificial intelligence approaches for modeling rainfall–runoff process. *J. Hydrol.* 2011, 402, 41–59.
- 45. Rodrigues, J.A.M.; de Oliveira Andrade, A.C.; Viola, M.R.; Ferreira, D.D.; Mello, C.R.D.; Thebaldi, M.S. Hydrological modeling in a basin of the Brazilian Cerrado biome. *Rev. Ambiente Água* **2021**, *16*. [CrossRef]
- 46. Sharma, N.; Zakaullah, M.; Tiwari, H.; Kumar, D. Runoff and sediment yield modeling using ANN and support vector machines: A case study from Nepal watershed. *Model. Earth Syst. Environ.* **2015**, *1*, 23.
- 47. Singh, A.; Malik, A.; Kumar, A.; Kisi, O. Rainfall-runoff modeling in hilly watershed using heuristic approaches with gamma test. *Arab. J. Geosci.* 2018, *11*, 261.
- Tebabal, S. Analysis and Prediction of Weather Condition Factors Using Machine Learning Techniques. Ph.D. Thesis, Bahir Dar University, Bahir Dar, Ethiopia, 2020.

- 49. Hasanpour, K.M. Flood estimation at ungauged sites using a new hybrid model. J. Appl. Sci. 2008, 9, 1744–1749.
- 50. Haykin, S. Neural Networks: A Comprehensive Foundation; Prentice-Hall: Hoboken, NJ, USA, 1999.
- 51. Zurada, J.M. Introduction to Artificial Neural Systems; West Publishing Company: Saint Paul, MN, USA, 1992.
- 52. Hayati, M. Short term load forecasting using artificial neural networks for the west of Iran. J. Appl. Sci. 2007, 12, 1582–1588.
- 53. Li, C.; Zhu, L.; He, Z.; Gao, H.; Yang, Y.; Yao, D.; Qu, X. Runoff prediction method based on adaptive Elman neural network. *Water* **2019**, *11*, 1113.
- 54. Sari, V.; dos Reis Castro, N.M.; Pedrollo, O.C. Estimate of suspended sediment concentration from monitored data of turbidity and water level using artificial neural networks. *Water Resour. Manag.* **2017**, *31*, 4909–4923.
- 55. Jang, J.S. ANFIS: Adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man Cybern. 1993, 23, 665–685.
- 56. Al-Mukhtar, M.; Al-Yaseen, F. Modeling water quality parameters using data-driven models, a case study Abu-Ziriq marsh in south of Iraq. *Hydrology* **2019**, *6*, 24.
- 57. Keskin, M.E.; Taylan, D.; Terzi, O. Adaptive neural-based fuzzy inference system (ANFIS) approach for modelling hydrological time series. *Hydrol. Sci. J.* 2006, *51*, 588–598.
- 58. Walia, N.; Singh, H.; Sharma, A. ANFIS: Adaptive neuro-fuzzy inference system-a survey. Int. J. Comput. Appl. 2015, 123, 32–38.
- 59. Hosseini, S.M.; Mahjouri, N. Integrating support vector regression and a geomorphologic artificial neural network for daily rainfall-runoff modeling. *Appl. Soft Comput.* **2016**, *38*, 329–345.
- Meshram, S.G.; Ghorbani, M.A.; Deo, R.C.; Kashani, M.H.; Meshram, C.; Karimi, V. New Approach for Sediment Yield Forecasting with a Two-Phase Feedforward Neuron Network-Particle Swarm Optimization Model Integrated with the Gravitational Search Algorithm. *Water Resour. Manag.* 2019, 33, 2335–2356.
- 61. Nourani, V.; Komasi, M. A geomorphology-based ANFIS model for multi-station modeling of rainfall–runoff process. *J. Hydrol.* **2013**, *490*, 41–55.
- 62. Sudhishri, S.; Kumar, A.; Singh, J.K. Comparative evaluation of neural network and regression-based models to simulate runoff and sediment yield in an outer Himalayan watershed. *J. Agric. Sci. Technol.* **2016**, *18*, 681–694.
- 63. Vafakhah, M. Comparison of cokriging and adaptive neuro-fuzzy inference system models for suspended sediment load forecasting. *Arab. J. Geosci.* 2013, *6*, 3003–3018.
- Remesan, R.; Shamim, M.A.; Han, D. Model data selection using gamma test for daily solar radiation estimation. *Hydrol. Process.* 2008, 22, 4301–4309.
- 65. Shamim, M.A.; Hassan, M.; Ahmad, S.; Zeeshan, M. A comparison of artificial neural networks (ANN) and local linear regression (LLR) techniques for predicting monthly reservoir levels. *KSCE J. Civ. Eng.* **2016**, *20*, 971–977.
- 66. Kemp, S.E. Gamma Test Analysis Tools for Non-Linear Time Series; Department of Computer & Mathematical Science, Faculty of Advanced Technology, University of Glamorgan: Wales, UK, 2006.
- 67. Penrose, R. A generalized inverse for matrices. Proc. Camb. Philos. Soc. 1955, 51, 406-413.
- 68. Legates, D.R.; McCabe, G.J. Evaluating the use of "Goodness of Fit" measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **1999**, *35*, 233–241.
- 69. Nash, J.E.; Sutcliffe, J.V. River Flow Forecasting Through Conceptual Models. J. Hydrol. 1970, 10, 282–290.
- 70. Eadara, A.; Kannam, H. Slope studies of Vamsadhara River basin: A Quantitative Approach. *Int. J. Eng. Innov. Technol.* 2013, *3*, 184–188.
- Samet, K.; Hoseini, K.; Karami, H.; Mohammadi, M. Comparison between soft computing methods for prediction of sediment load in rivers: Maku dam case study. *Iran. J. Sci. Technol. Trans. Civ. Eng.* 2019, 43, 93–103.
- Isazadeh, M.; Ahmadzadeh, H. Assessment of Normalization of Monthly Runoff Probabilistic Distribution impact on SVM and ANN Models Performance in Monthly River Flows Simulation (A Case Study: ZarrinehRud River Basin). J. Watershed Manag. Res. 2018, 8, 22–33.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.