Pure Appl. Geophys. 179 (2022), 3751–3765

https://doi.org/10.1007/s00024-022-03131-8

Pure and Applied Geophysics

Suspended Sediment Modeling Using Sequential Minimal Optimization Regression and Isotonic Regression Algorithms Integrated with an Iterative Classifier Optimizer

Mir Jafar Sadegh Safari,[1] Sarita Gajbhiye Meshram,[2] Khabat Khosravi,[3,4] and Adel Moatamed[5,6,7]

Abstract—Suspended sediment load modeling through advanced computational algorithms is of major importance and a challenging topic for developing highly accurate hydrological models. To model the suspended sediment load in the Rampur watershed station in the Mahanadi River Basin, Chhattisgarh State, India, unique integrated computational intelligence regression models with an optimizer are proposed in this study. For the first time in the literature, the isotonic regression (ISO) and sequential minimal optimization regression (SMOR) models and their hybrid versions with an iterative classifier optimizer (ICO) are applied for suspended sediment load modeling. The research is based on daily discharge and suspended sediment data collected over a 38-year period (1976–2014). Root mean square error (*RMSE*), relative root mean square error (*RRMSE*), coefficient of determination (R^2), and Nash–Sutcliffe efficiency (*NSE*) were employed to evaluate the performance of the standalone ISO and SMOR, as well as the proposed ICO–ISO and ICO–SMOR hybrid models. Ten different scenarios were considered for modeling to investigate the performance of the models using different input combinations. The proposed new models were found to be more reliable than standalone ISO and SMOR models. Results revealed that the performance of the hybrid model was mostly attributable to the basic algorithm for the model development, where both SMOR and ICO–SMOR models were superior to their ISO and ICO–ISO counterparts in terms of accurate computation. Overall, the ICO–SMOR models outperformed the other models in terms of accuracy, with *RMSE*, *RRMSE*, R^2, and *NSE* of 5495.1 tons/day, 2.77, 0.90, and 0.86, respectively. The current study's findings support the applicability of the proposed methodology for modeling of suspended sediment load and encourage the use of these methods in alternative hydrological modeling.

Keywords: Isotonic regression, iterative classifier optimizer, Mahanadi River, suspended sediment, sequential minimal optimization regression.

1. Introduction

Because of (a) the huge regional diversity of catchment characteristics and precipitation patterns, and (b) the number of variables involved in physical process modeling, runoff–sediment yield is one of the most difficult hydrological phenomena to comprehend. The amount of runoff and sediment yield produced by a particular rainfall is mostly determined by the rate, length, and distribution of the rainfall, as well as initial soil moisture, land use, catchment geomorphology, and other factors.

Runoff determination is essential to many tasks, including constructing flood control systems, preserving agricultural land, and storing and releasing water. Rainfall and runoff both cause sediment overflow, which reduces the holding capacity of rivers and other hydraulic systems. It has also been blamed for transporting contaminants such as dangerous materials, herbicides, and fertilizers. Since the 1930s, a number of models have been created for simulating rainfall–runoff, runoff–sediment yield, and rainfall–runoff–sediment yield processes in river catchment systems, and they are widely classified into regression, stochastic, conceptual or parametric, and (dynamic) system models.

[1] Department of Civil Engineering, Yaşar University, Izmir, Turkey. E-mail: jafar.safari@yasar.edu.tr

[2] WRAM Research Lab Pvt. Ltd., Nagpur, Maharashtra 440027, India. E-mail: gajbhiyesarita@gmail.com

[3] Department of Watershed Management Engineering, Ferdowsi University of Mashhad, Mashhad, Iran. E-mail: kkhosrav@fiu.edu; Khabat.Khosravi@gmail.com

[4] Department of Earth and Environment, Florida International University, Miami, USA.

[5] Department of Geography, College of Humanities, King Khalid University, Abha, Saudi Arabia. E-mail: amuatmed@kku.edu.sa

[6] Department of Geography, Faculty of Arts, Assiut University, Assiut, Egypt.

[7] Prince Sultan Bin Abdul-Aziz Center for Environment and Tourism Research and Studies, King Khalid University, Abha, Saudi Arabia.

Investigating the sediment load in rivers is important for addressing water scarcity and water quality problems (Sharafati et al., 2020a). With a major role in sediment transport in hydrological science, soil erosion poses a significant threat to sustainable farming and the climate. It has become an extreme issue because of an insufficient understanding of the bearing capacity of soil. Extensive soil erosion and related issues have deteriorated the soil and water resources of the world. The presence of multiple, frequently interrelated climatic and physiographic factors makes the phase of rainfall-sediment not only very complicated to understand but often extremely difficult to simulate (Abebe & Gebremariam, 2019; Christanto et al., 2019; Gudino-Elizondo et al., 2019; Meshram et al. 2020; Tuset et al., 2015). Because sediment load plays a crucial role in any decision-making process about water availability, precise simulation of sediment load is important for sustainable water supply and environmental systems. The use of data-driven modeling techniques to improve sediment yield rating curves has attracted considerable attention in recent years (Ampomah et al., 2020; Yadav et al., 2018). Multiple sediment prediction models have been developed by hydrology researchers, ranging from empirical, such as the Universal Soil Loss Equation (USLE)/Revised USLE (RUSLE) (Arekhi et al., 2012; Borrelli et al., 2017), to mathematical, such as kinematic/diffusion wave theory (Liu et al. 2004; Schneider, 2018) or linear/nonlinear programming optimization (Nicklow & Mays, 2000) and physical process-based models such as the Soil & Water Assessment Tool (SWAT) (Asres & Awulachew, 2010; Chandra et al., 2014; Dutta and Sen, 2018; Liu & Jiang, 2019) and Water Erosion Prediction Project (WEPP) (Ahmadi et al., 2020; Singh et al., 2017), and these and many others have contributed to a better understanding of sediment yield modeling, but they are often data-hungry. As a result, alternate approaches for forecasting runoff and sediment yield must be sought. One way of resolving such issues is to use artificial intelligence.

Linked to robustness of artificial intelligence (AI) algorithms, various types of AI can be applied for sediment transport modeling (Safari & Shirzad, 2019; Safari, 2020; Achite et al., 2021; Larson et al., 2021; Meshram et al., 2019; 2021; Harun et al., 2021; Mohammadei et al., 2021; Vaheddoost et al., 2022; Samadianfard et al., 2022; Essam et al., 2022). Effective sediment yield or load predictions have been made using AI algorithms such as support vector machine (SVM) (Buyukyildiz & Kumcu, 2017; Cimen, 2008; Misra et al., 2009; Samantaray et al., 2020), least-squares SVM (LSSVM) (Kisi, 2012; Kisi & Ozkan, 2017) and artificial neural networks (ANN) (Bouzeria et al., 2017; Jothiprakash & Garg, 2009; Talebizadeh et al., 2010). Despite the high prediction accuracy obtained by SVM, its value is diminished by the need to evaluate four kernel functions to decide best. It also needs a number of parameters to determine optimum values. Although the ANN is the most widely used AI method, it has certain flaws, such as low prediction power when the range of test data exceeds the range of training data and when the datasets are small (Khosravi et al., 2018). To tackle these issues, researchers combined the ANN model with a fuzzy logic and adaptive neuro-fuzzy inference method (ANFIS). Flood forecasting (Kim et al., 2019; Patel & Parekh, 2014; Ullaha & Choudhury, 2010), crop yield prediction (Naderloo et al., 2012), and water quality prediction (Naderloo et al., 2012) all used ANFIS algorithms (Tiwari et al., 2018). Yuan et al. (2018) applied the long short-term memory neural network–antlion optimizer (LSTM–ALO) model for monthly runoff forecasting. The simulation results by the LSTM–ALO were compared with those of the LSTM and LSTM–particle swarm optimization (LSTM–PSO). The comparisons show that the ALO could increase the accuracy of the LSTM model in forecasting monthly runoff with different model inputs. Sharafati et al. (2020b) predicted suspended sediment load (SSL) using gradient boost regression (GBR), AdaBoost regression (ABR) and random forest regression (RFR) models. Based on performance metrics and visualization, the RFR model shows a slight lead in prediction performance. Doroudi et al. (2021) predicted SSL using a new hybrid support vector regression–observer-teacher-learner-based optimization (SVR-OTLBO) model. The results indicated that the SVR-OTLBO model performed better than standalone SVR models. Adnan et al. (2021) predicted stream flow using a new hybrid extreme learning machine (ELM) model combined with hybrid PSO and grey wolf optimization (GWO). The PSO- and GWO-based ELM models also performed better than standalone ELM models, with an improvement in the root mean square error (RMSE) by 19.9 and 20.3%, respectively. Adnan et al. (2022) predicted sediment load using a fuzzy c-means-based neuro-fuzzy system using the hybrid particle swarm optimization-gravitational search algorithm (ANFIS–FCM–PSOGSA). Based on the results, ANFIS–FCM–PSOGSA was able to improve the prediction performance of the ANFIS–FCM–PSO (or ANFIS–FCM) models.

Although ANFIS is a powerful algorithm, it is hampered by internal parameters and the need to accurately hybridize ANFIS with a meta-heuristic method to tackle this problem. Meta-heuristic algorithms find the ideal mass on their own. Because hybrids are more adaptable and hence more robust with noisy data than standalone algorithms, they can more easily characterize the non-linearity of input and output variables. While this constraint is overcome, hybridization increases the model's complexity and time consumption by requiring a time-consuming search for the best meta-heuristic method among a large number of meta-heuristic models with various topologies. Researchers are still striving to develop AI methods that are simple, scalable, adaptable, and dependable.

A new form of AI algorithm has recently been developed to solve regression problems and reduce the AI disadvantages. In order to quantify hydrology, climatology, and hydraulics, new algorithms such as random forest (RF), pace regression (PR), isotonic regression (ISO), sequential minimal optimization regression (SMOR). and iterative classifier optimizer (ICO) have been used to quantify landslide susceptibility mapping (Pham et al., 2019), stratigraphic data modeling (Polucci et al., 2020), software development (Veni & Srinivasan, 2020), and environmental analysis (Hussian et al., 2005). The lack of hidden layers and transparency modeling in AI algorithms (i.e., RF, PR, ISO, SMOR, ICO, and others) allows better modeling performance than ANN and ANFIS.

The main objective of this study is to predict the suspended sediment load using two standalone algorithms (ISO and SMOR) and two hybrid algorithms (ICO–ISO and ICO–SMOR). Although standalone algorithms can predict suspended sediment load satisfactorily, and their predictive power has been demonstrated in applications to other hydrological phenomena, combining them with classifier algorithms can improve predictive accuracy and eliminate the inherent flaws of each model. To this end, the major novelties of this study are as follows:

i. While there are a variety of artificial intelligence techniques for suspended sediment load modeling, the accuracy of standalone models is not as high as that of hybrid models. Therefore, for the first time in the literature, this study recommends advanced and novel hybrid algorithms, ICO–ISO and ICO–SMOR, for suspended sediment load modeling.
ii. Because of the complexity and difficulty in suspended sediment load modeling, defining an appropriate scenario for model development is a challenging task. Therefore, this study applied ten different scenarios for suspended sediment load modeling.
iii. There was no recorded work for the ISO and SMOR models integrated with ICO for suspended sediment forecasting in the relevant literature.

2. *Materials and Methods*

2.1. *Study Area and Modeling Data*

The Rampur watershed originates from the Jonk River catchment (Mahanadi basin), India. The largest watershed area includes the Chhattisgarh state districts of Raipur and Mahasamund, and the minor area lies in the Odisha state districts of Nuapada and Bargarh. The watershed extends over an area of 3424 km^2 and ranges from 81.28° 16′ 00″ to 83.18° 42′ 00″ east longitude and 20.27° 53′ 00″ to 21.47° 49′ 00″ north latitude. In this area, the climate is mostly tropical wet and dry, and the average temperature varies between 15 and 35 °C. While temperature remains moderate throughout the year, months such as April and May can be exceptionally hot, where temperatures can often rise above 48 °C. Average annual rainfall in the area varies between 800 and 1200 mm. About 50% of the watershed area

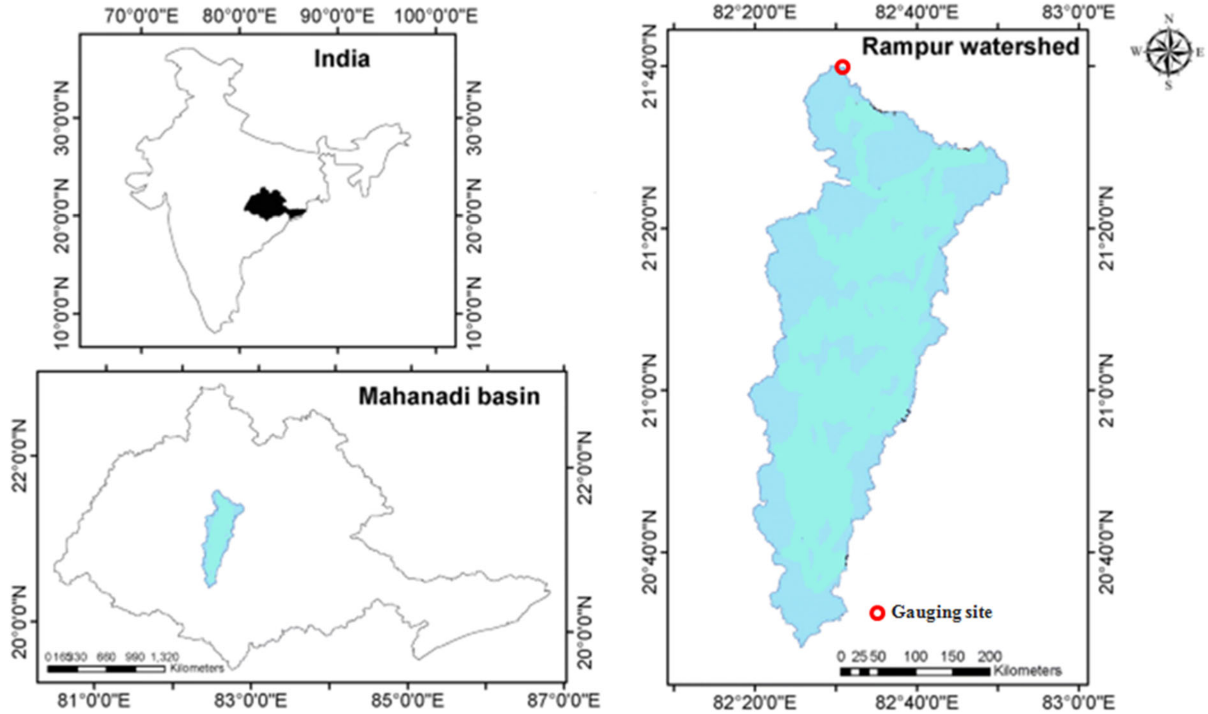


Figure 1
Index map of the Rampur watershed (study area)

comprises forest and agricultural land. The region's main cultivated crop is paddy. Watershed elevation varies from 205 to 875 m. The watershed is composed of three types of soil, i.e., clay, loam and sandy loam (Fig. 1).

For the period 1976 to 2014, daily rates of discharge (m^3/s) and suspended sediment load (tons/-day) from the Rampur station were used; 75% of the data were used for model development/calibration, while the remaining 25% were used to test and evaluate the model's performance. The time series of the whole dataset that was applied for the Rampur station is shown in Fig. 2. The statistical parameters for the results are listed in Table 1.

2.2. *Sequential Minimal Optimization Regression (SMOR)*

Sequential minimal optimization regression (SMOR) is an efficient algorithm for training the conventional support vector regression (SVR) method. Due to large size in the objective function for the optimization problem in SVR:

$$\text{Max } \mathcal{W}(\alpha) = \sum\nolimits_{i=1}^{n} \alpha_1 - \frac{1}{2}\sum\nolimits_{i=1}^{n} \alpha_i \alpha_j y_i y_j \mathcal{K}(\mathcal{X}_i, \mathcal{X}_j). \quad (1)$$

Subject to

$$\sum\nolimits_{i=1}^{n} \alpha_1 y_i = 0(2) 0 \le \alpha_i \le \mathcal{C}; \quad i = 1, 2, \ldots e, \quad (2)$$

where $\mathcal{K}(\mathcal{X}_i, \mathcal{X}_j)$ denotes the kernel function, the quadratic problem arising from SVRs cannot be efficiently handled using typical numerical quadratic issue (QP) methodologies, especially when the problem is large in size. Numerous algorithms are presented for resolving the dual function problem. Platt (1999) introduced a sequential minimum optimization technique for classification problems that iteratively selects a working set of size two and optimizes the objective function using analytical solutions to subproblems (Platt, 1999). The technique is continued iteratively until all training instances satisfy the Karush–Kuhn–Tucker (KKT) requirements. Smola and Schölkopf enhance the SMOR algorithm's capability to solve regression problems (Smola and Scholkopf, 2004).

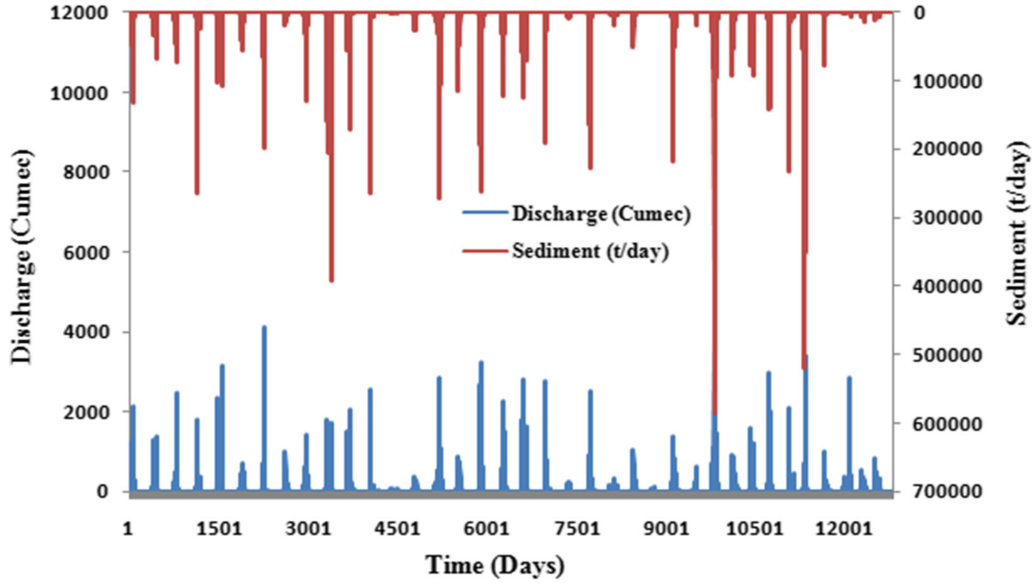


Figure 2
Time series of observed data (discharge and sediment) used for training and testing stages

Table 1

Statistics of the data

Dataset	Data type	Data no.	Mean	STD	CV	Max	Min
Training	Discharge (m^3/s)	8938	42.57039	173.2609	406.9987	4100	0
	Sediment load (tons/day)		1542.86	11,793.35	764.3823	391,611.8	0
Testing	Discharge (m^3/s)	3823	48.09995	169.4996	352.3905	3380	0
	Sediment load (tons/day)		1985.985	14,903.71	750.444	350,438.4	0

2.3. *Isotonic Regression (ISO)*

Isotonic regression is a popular nonparametric regression technique. We will quickly describe isotonic regression in the following section where the parameters have simple order relations. Consider p populations, with μ_i denoting an important scalar parameter for group $i = 1, 2, \ldots, p$ and $\mu = (\mu_1, \ldots, \mu_p)$. It is assumed that there is simple order among μ_i, such as $\mu_1 \geq \cdots \geq \mu_p$.

Let $(\hat{\mu}_1)$ denote an estimator of μ_i for $i = 1, 2, \ldots, p$ and $\hat{\mu} = (\widehat{\mu_1}, \ldots, \widehat{\mu_p})$. To satisfy $\mu_1 \geq \cdots \geq \mu_p$, the isotonic regression of $\widehat{\mu}$,, denoted by $\hat{\mu}^{\mathcal{IR}} = (\widehat{\mu_1}^{\mathcal{IR}}, \ldots, \widehat{\mu_p}^{\mathcal{IR}})$, was provided by Nagatsuka et al. (2012).

$$\hat{\mu}^{\mathcal{IR}} = \arg{}^{\min}_{\mu} w \sum\nolimits_{i=1}^{p} (\widehat{\mu_i} - \mu_i)^2 w_i,\ \mu_1 \geq \cdots \mu_p, \tag{3}$$

where w_i, $i = 1, 2, \ldots, p$ are given or suitably chosen weights. Usually, the weights w_i are chosen such that $w_1 = w_2 = \cdots = w_p$.

2.4. *Iterative Classifier Optimizer (ICO)*

The iterative classifier optimizer (ICO) employs cross-validation and optimizes the number of iterations for a given classifier; it is capable of handling missing, nominal, and binary classes and characteristics such as numeric, nominal, binary, and empty nominal (Omondi & Rajapakse, 2010). After constructing the model and comparing it with observed and measured values, the model's performance is

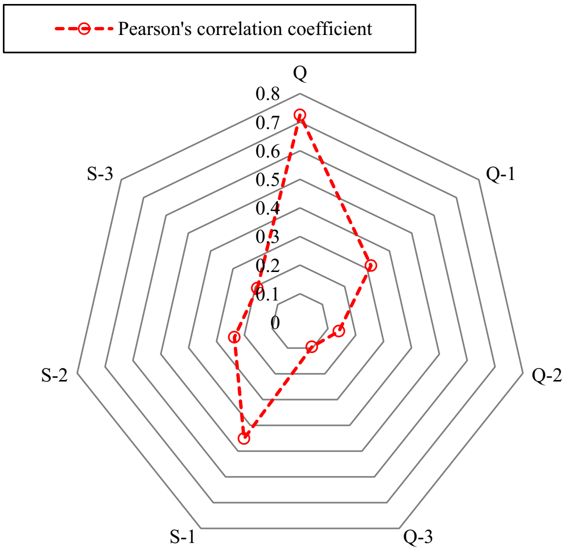


Figure 3
Pearson correlation coefficient (PCC) for studied scenarios

evaluated using the ICO algorithm's optimization technique. The information gained is then used to tune the model's outputs.

2.5. Determination of Input Parameters

Using correlation coefficients calculated for various time lags between input variables and suspended sediment load (*S*), the most efficient independent parameters for the computation of suspended sediment load were identified as shown in Fig. 3. In this study, as shown in Table 2, ten scenarios of input parameters were considered using (i) discharge data (*Q*); (ii) *Q*, *Q*-1; (iii) *Q*, *Q*-1, *Q*-2; (iv) *Q*, *Q*-1, *Q*-2, *Q*-3; (v) *S*-1; (vi) *S*-1, *S*-2; (vii) *S*-1, *S*-2, *S*-3; (viii) *Q*, *S*-1; (ix) *Q*, *Q*-1, *S*-1; and (x) *Q*, *Q*-1, *Q*-2, *Q*-3, *S*-1, *S*-2, *S*-3. According to Fig. 3, discharge influences suspended sediment load (*S*) most significantly, which is in agreement with results reported by Chiang and Tsai (2011) and Kisi et al. (2012). The above models were trained with different input combinations and then used to compute *S* in the Rampur watershed, Mahanadi River. The RMSE criterion is considered to determine the best input combination. This step used the default operator for each model (e.g., Kisi et al., 2012).

Table 2

Various considered scenarios

Combination no.	Scenarios
1	*Q*
2	*Q*, *Q*-1
3	*Q*, *Q*-1, *Q*-2
4	*Q*, *Q*-1, *Q*-2, *Q*-3
5	*S*-1
6	*S*-1, *S*-2
7	*S*-1, *S*-2, *S*-3
8	*Q*, *S*-1
9	*Q*, *S*-1, *Q*-1
10	*Q*, *Q*-1, *Q*-2, *Q*-3, *S*-1, *S*-2, *S*-3

2.6. Model Setup

Determination of the best model structure is an essential step in the model development procedure. It can be achieved by training the models using different input combinations together with determining the best hyperparameters. To search for the best model hyperparameters, the WEKA package was used. At the first step of the modeling procedure, the default values of the packages were applied, and subsequently, through a trial-and-error procedure, the best parameters were determined. The performance criterion for the *RMSE* was utilized to evaluate the developed model performance for best hyperparameter selection. Hyperparameters for the studied models are given in Table 3.

2.7. Performance Criteria

There were four statistical indexes used in this study to assess the accuracy of stand-alone ISO and SMOR as well as the hybrid ICO-SMOR and ICO-ISO models for modeling suspended sediment load, including root mean square error (*RMSE*), relative root mean square error (*RRMSE*), determination coefficient (R^2) and Nash–Sutcliffe efficiency (*NSE*). The formulation can be expressed as follows:

$$R^2 = \left(\frac{1}{n} \times \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(\sigma_x)(\sigma_y)} \right)^2 \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (x_i - y_i)^2}{n}} \quad (5)$$

Table 3

Optimal values of model hyperparameters

Model hyperparameters	Optimum value		
	ISO	ICO	SMOR
Batch size	100	100	100
C	–	–	1
Number of decimal places	–	2	2
Do not check capabilities	–	No	No
Debug	No	No	No
Number of folds	–	10	–
Filter type	–	–	Normalize training data
Kernel	–	–	Poly kernel
Regression optimizer	–	–	Regression SMOR improved
Look-ahead iteration	–	50	–
Number of decimal places	–	2	–
Pool size	–	1	–

$$RRMSE = \frac{RMSE}{\frac{1}{n}\sum_{i=1}^{n} y_i} \quad (6)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x}_i)^2} \quad (7)$$

where n is the number of data, x and y are observed and estimated values, and σ_x and σ_y are the standard deviation of the observed and estimated data. It should be mentioned that low value (closer to zero) for the *RRMSE* and *RMSE* while for *NSE* indicator as well as R^2, a high value (closer to the unity) signify that there is a good agreement between observed and modeled estimation data.

3. Results and Discussion

This study used daily discharge and suspended sediment load data from the Rampur watershed, Mahanadi River in India. As stated earlier, four models, i.e., ISO, SMOR, ICO–ISO, and ICO–SMOR, were developed for discharge and suspended sediment load modeling. An overview of the study is given in Fig. 4.

3.1. Statistical Analysis of Data

Initially, daily discharge and sediment load data were divided into two parts, calibration/training and testing, with 70% selected for calibration and the remaining 30% for testing the developed models. The statistical parameters for calibration/training and testing of discharge and sediment load datasets were calculated as shown in Table 1.

The coefficient of variation (CV) and standard deviation (STD) values for the discharge training datasets were higher than those for the testing, but the mean and standard deviation values for sediment training datasets were lower than those for the testing (Table 1). Furthermore, the maximum value of the discharge variable is higher during the training dataset. The maximum values for the sediment load training dataset were higher than those for the testing.

3.2. Model Performance and Validation

In this study, input selection using the Pearson correlation coefficient (*PCC*) was conducted on the entire dataset to determine the best input parameters with the best correlation with suspended sediment load. Results showed that daily runoff data had the highest correlation coefficient with runoff of the previous day (Fig. 3). The *PCC* approaches were used to choose the most important driving variable among the input variables (Chiang & Tsai, 2011; Kisi et al. 2012; Khosravi et al., 2018). Discharge has the greatest effect on suspended sediment load (*PCC* = 0.72), followed by *S*-1 (*PCC* = 0.45), *Q*-1 (*PCC* =

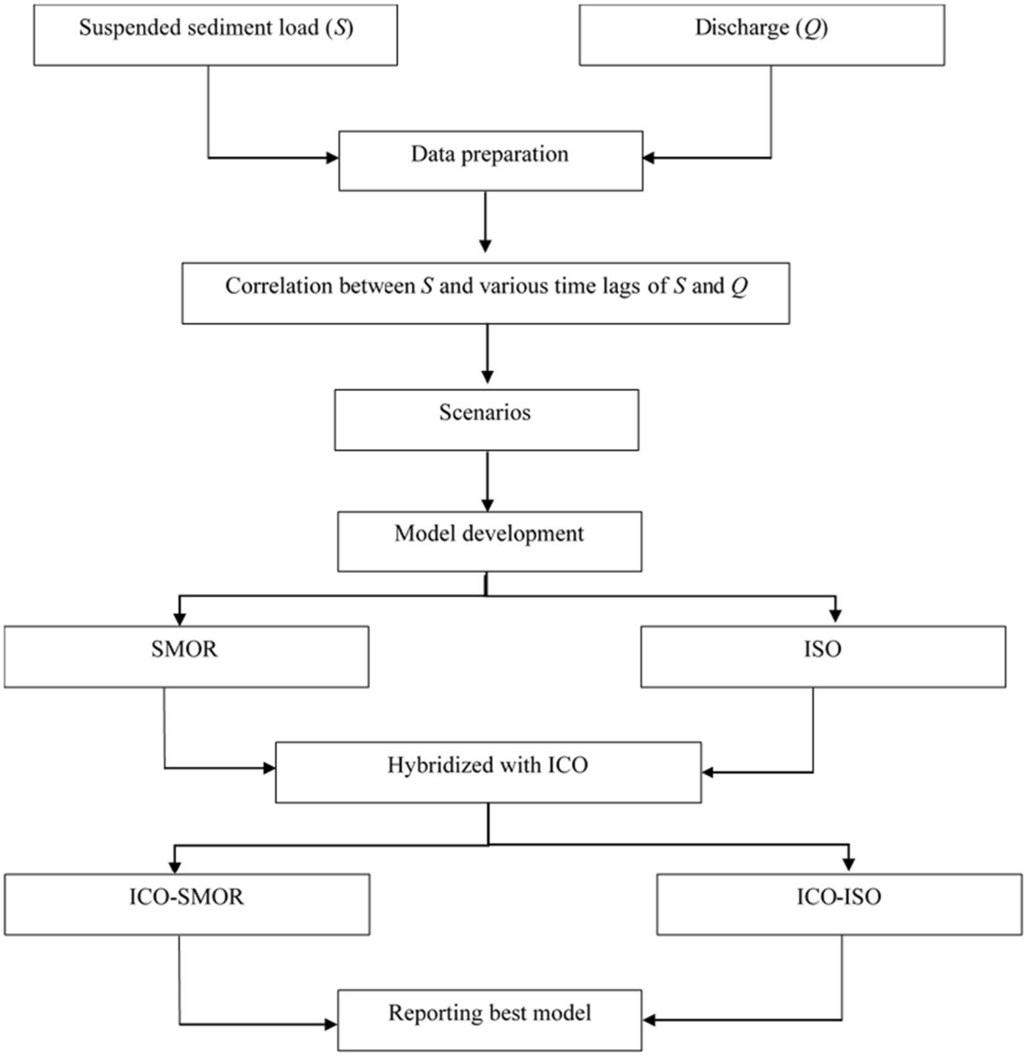


Figure 4
Flowchart of the modeling procedure

0.32), *S*-2 (*PCC* = 0.24), *S*-3 (*PCC* = 0.19), and *Q*-3 (*PCC* = 0.10), according to the *PCC* values in Fig. 3.

Table 2 shows ten different combinations that were developed and investigated based on particular *PCC* values. For each of the several sets of input parameters, all of the models created in this work (ISO, SMOR, ICO–ISO, and ICO–SMOR) use the same datasets. The model efficiency was assessed using the *RMSE* (as shown in Fig. 5) and various input parameters. As shown in Fig. 5, modeling of the suspended sediment load by incorporating only discharge and its lags up to 3 days yielded no significant improvement in the performance of the developed ISO, SMOR, ICO–ISO, and ICO–SMOR. The *RMSE* lines shown in Fig. 5 for four developed models tend to be a straight line from one input parameter of *Q* to the four input parameters of *Q*, *Q*-1, *Q*-2, and *Q*-3. This result indicates that for the current case study, suspended sediment load cannot be modeled considering only the discharge and its lags as input parameters. The accuracy of all the developed models remains almost the same when only lags of suspended sediment load parameters are considered for modeling the current suspended sediment load, although slight improvement and worsening are shown for ISO-based and SMOR-

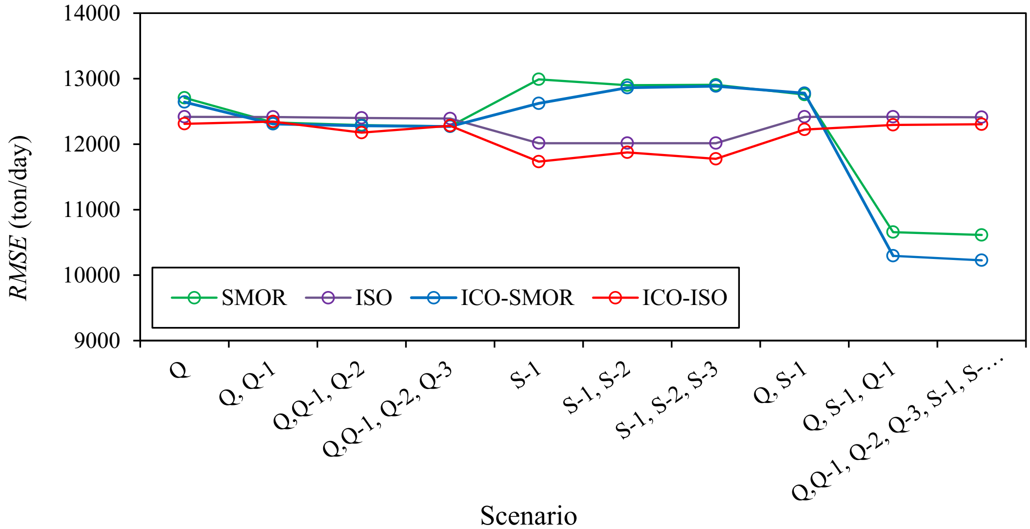


Figure 5
Performance of SMOR, ISO, ICO–SMOR, and ICO–ISO models for different scenarios in terms of *RMSE*

Table 4
Comparison of SMOR, ISO, ICO–SMOR, and ICO–ISO models in terms of RMSE, RRMSE, R^2 and NSE

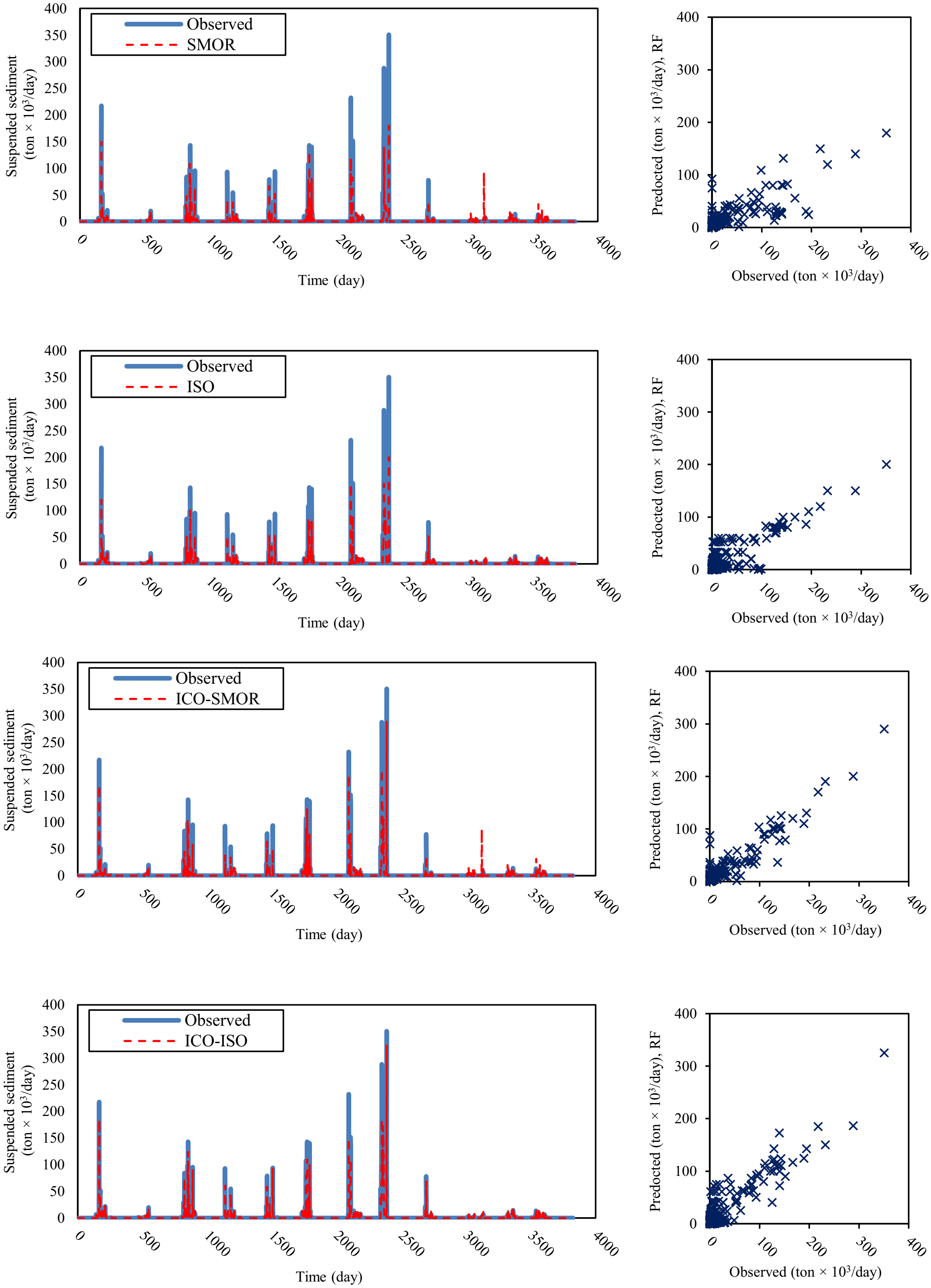
Model	*RMSE* (tons/day)	RRMSE	R^2	NSE
SMOR	9328.4	4.70	0.70	0.61
ISO	8037.6	4.05	0.77	0.71
ICO–SMOR	**5495.1**	**2.77**	**0.90**	**0.86**
ICO–ISO	5609.3	2.82	0.86	0.86

Bold values indicate based model as lowest RMSE, highest R^2 and NSE compared to other model

based models, respectively. Simultaneously incorporating the discharge and its lags as well as the suspended sediment load lags significantly enhanced the performance of the SMOR-based standalone SMOR and hybrid ICO–SMOR models. Two scenarios of *Q*, *S*-1, *Q*-1, and *Q*, *Q*-1, *Q*-2, *Q*-3, *S*-1, *S*-2, and *S*-3 provide better results for the SMOR and d ICO–SMOR models.

Among the applied artificial intelligence methods, best scenarios from each model (ISO, SMOR, ICO–ISO, and ICO–SMOR) were selected for discussion. It can be seen from Table 4 that ICO–SMOR has the highest values of R^2 and *NSE* of 0.90 and 0.86, respectively, and the lowest *RMSE* and *RRMSE* of 5495.1 tons/day and 2.77, respectively. The ICO–SMOR-based model outperformed the R^2 (*NSE*) accuracy of ICO–ISO, ISO, and SMOR by 4.44% (0%), 14.44% (17.44%), and 22.22% (29.07%), respectively; also, ICO–SMOR outperformed the *RMSE* (*RRMSE*) accuracy of ICO–ISO, ISO, and SMOR by 2.04% (1.77%), 31.63% (31.60%), and 41.09% (41.06), respectively.

Figure 6 shows the model performance for the estimated and observed suspended sediment by SMOR, ISO, ICO–SMOR, and ICO–ISO, and it can be seen that all four models underestimated the peak observed values, which is similar to the findings of Adnan et al. (2019) and Kişi (2004). When compared with the other models, the estimated values of ICO–SMOR are closer to the observed values with the least scattered estimated values and highest R^2. The closeness of the estimated suspended sediment to the observed one and highest R^2 values of the models are


Observed
SMOR
ISO
ICO-SMOR
ICO-ISO
Suspended sediment
(ton × 10³/day)
Time (day)
Predocted (ton × 10³/day), RF
Observed (ton × 10³/day)
0
50
100
150
200
250
300
350
400
500
1000
1500
2000
2500
3000
3500
4000

◀ Figure 6
Comparison of SMOR, ISO, ICO–SMOR, and ICO–ISO model performance in terms of observed and computed suspended sediment load

in the order ICO–SMOR > ICO–ISO > ISO > SMOR.

As an alternative visual model performance evaluation criterion, violin plots given in Fig. 7 are used. Violin plots have a feature in which the probability distribution of the developed model results can be compared with the corresponding observed values. It is seen in Fig. 7 that the ICO–SMOR and ICO–ISO models provide better performance than SMOR and ISO standalone models. It must be noted that in terms of probability distribution, ICO–ISO gives better results than ICO–SMOR, where its violin shape is mostly similar to the observed counterpart. Furthermore, the performance of the developed models is investigated using a Taylor diagram as shown in Fig. 8. The main advantage of a Taylor diagram is that it uses three statistical performance criteria simultaneously; therefore, reliable justification can be produced. The observed or reference point is located in the *X*-axis. A model that is closest to the reference point is considered as a best model. The outcomes obtained based on the Taylor diagram are in agreement with findings obtained in previous sections where ICO–SMOR outperforms its alternatives in suspended sediment load estimation.

As a complex hydrological problem, suspended sediment load causes serious uncertainties in the hydrological properties of a river. Suspended sediment load has a direct impact on the design of hydraulic structures, ecosystem, water quality, and pollution control. Therefore, accurate estimation of the suspended sediment load of a river is of importance in both theory and practice. Owing to the clarification above, developing robust artificial intelligence models may facilitate dealing with sediment load problems in the rivers. This study first investigates the importance of parameters involved in the modeling where discharge is found as the most

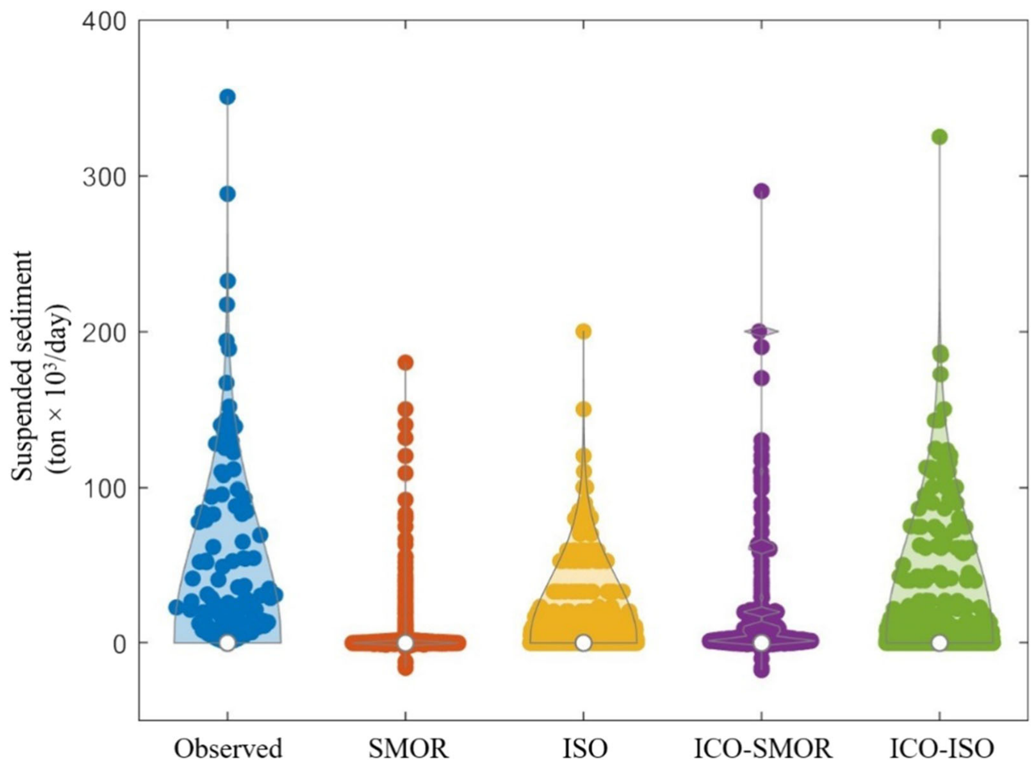


Figure 7
Violin plot for model performance evaluation

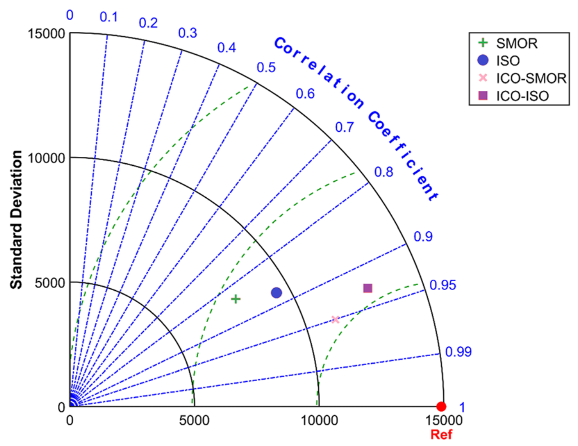


Figure 8
Taylor diagram graphical presentation

important parameter in suspended sediment load modeling. Applying novel types of AI techniques, standalone SMOR and ISO models are first developed and then hybridized with ICO to develop ICO–SMOR and ICO–ISO models. Results illustrate that hybrid ICO–SMOR and ICO–ISO models outperform standalone SMOR and ISO models in terms of different visual and mathematical performance indices. It is found that the performance of ICO–SMOR is better in terms of error, while the ICO–ISO model, by means of probability distribution, provides better results. Consequently, the models developed in this study can be applied for suspended sediment load calculation in rivers. It is worth mentioning that the efficiency of the developed scenarios and recommended modeling techniques must be further examined in regions with different climate conditions.

4. Conclusions

The main goal of this study was to model the suspended sediment load in the Rampur watershed (Mahanadi basin), India, by employing four artificial intelligent techniques, i.e., SMOR, ISO (standalone models) and ICO–SMOR, ICO–ISO (hybrid models). Statistical metrics and graphical evaluation were used to quantify the predictive accuracy of these models. The simulation is based on daily discharge and sediment load data from 1 to 2 years ahead of historical records. Different input combinations were tested on all of the models in order to choose the optimal scenario for further investigation. The hybrid models outperformed the standalone models in estimating the daily sediment load, and were ranked first (ICO–SMOR) and second (ICO–ISO), respectively, in a comparison of the models produced based on a number of statistical error measurement indicators. As a limitation, the models and scenarios developed in this study can be applied for rivers with similar climate conditions, but further evaluation is needed before their application can be recommended in regions with different climate conditions. In this study, suspended sediment load and discharge parameters were used for model development. Incorporating more hydrometeorological parameters for suspended sediment load modeling can be considered as a future research direction.

Acknowledgements

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University, Abha, Kingdom of Saudi Arabia for funding this work through small research groups under grant number RGP. 1/113/43. Special thanks to Mr. Behzad Shakouri from Urmia University for his help during the revision of the manuscript.

Funding

This research work was supported by the Deanship of Scientific Research at King Khalid University under Grant number RGP. 1/113/43.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflicts of interest The authors declare no conflicts of interest.

Code availability None.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication All authors agree to publish.

References

Abebe, T., & Gebremariam, B. (2019). Modeling runoff and sediment yield of Kesem dam watershed, Awash basin, Ethiopia. *SN Applied Science, 1*, 446. https://doi.org/10.1007/s42452-019-0347-1

Achite, M., Yaseen, Z. M., Heddam, S., Malik, A., & Kisi, O. (2021). Advanced machine learning models development for suspended sediment prediction: Comparative analysis study. *Geocarto International,* 1–25.

Adnan, R.M., Liang, Z., El-Shafie, A., Zounemat-Kermani, M., Kisi, O. (2019). Prediction of Suspended Sediment Load Using Data-Driven Models. Water, *11*, 2060. https://doi.org/10.3390/w11102060.

Adnan, R. M. R., Mostafa, R., Kisi, O., Yaseen, Z. M., Shahid, S., & Zounemat-Kermani, M. (2021). Improving streamflow prediction using a new hybrid ELM model combined with hybrid particle swarm optimization and grey wolf optimization. *Knowledge-Based Systems*, *230*, 107379. https://doi.org/10.1016/j.knosys.2021.107379.

Adnan, R. M. R., Yaseen, Z. M., Heddam, S., Shahid, S., Sadeghi-Niaraki, A., & Kisi, O. (2022). Predictability performance enhancement for suspended sediment in rivers: Inspection of newly developed hybrid adaptive neuro-fuzzy system model. *International Journal of Sediment Research, 37*(3), 383–398.

Ahmadi, M., Minaei, M., Ebrahimi, O., et al. (2020). Evaluation of WEPP and EPM for improved predictions of soil erosion in mountainous watersheds: A case study of Kangir River basin, Iran. *Modelling Earth System Environment, 6*, 2303–2315. https://doi.org/10.1007/s40808-020-00814-w

Ampomah, R., Hosseiny, H., Zhang, L., Smith, V., & Sample-Lord, K. (2020). A regression-based prediction model of suspended sediment yield in the Cuyahoga River in Ohio using historical satellite images and precipitation data. *Water, 12*, 881. https://doi.org/10.3390/w12030881

Arekhi, S., Niazi, Y., & Kalteh, A. M. (2012). Soil erosion and sediment yield modeling using RS and GIS techniques: A case study, Iran. *Arabian Journal of Geosciences, 5*(2), 285–296. https://doi.org/10.1007/s12517-010-0220-4

Asres, M. T., & Awulachew, S. B. (2010). SWAT based runoff and sediment yield modelling: A case study of the Gumera watershed in the Blue Nile basin. *Ecohydrology and Hydrobiology, 10*(2–4), 191–199. https://doi.org/10.2478/v10104-011-0020-9

Borrelli, P., Robinson, D. A., Fleischer, L. R., Lugato, E., Ballabio, C., Alewell, C., & Bagarello, V. (2017). An assessment of the global impact of 21st century land use change on soil erosion. *Nature Communications, 8*(1), 1–13. https://doi.org/10.1038/s41467-017-02142-7

Bouzeria, H., Ghenim, A. N., & Khanchoul, K. (2017). Using artificial neural network (ANN) for prediction of sediment loads, application to the Mellah catchment, northeast Algeria. *Journal of Water and Land Development, 33*(IV–VI), 47–55.

Buyukyildiz, M., & Kumcu, S. Y. (2017). An estimation of the suspended sediment load using adaptive network based fuzzy inference system, support vector machine and artificial neural network models. *Water Resources Management: an International Journal, Published for the European Water Resources Association (EWRA) Springer; European Water Resources Association (EWRA), 31*(4), 1343–1359.

Chandra, P., Patel, P. L., Porey, P. D., & Gupta, I. D. (2014). Estimation of sediment yield using SWAT model for Upper Tapi basin. *ISH Journal of Hydraulic Engineering*. https://doi.org/10.1080/09715010.2014.902170

Chiang, J. L., & Tsai, Y. S. (2011). Suspended sediment load estimate using support vector machines in Kaoping river basin. In *Consumer electronics, communications and networks (CECNet), international conference* (pp. 1750–1753).

Christanto, N., Setiawan, M. A., Nurkholis, A., Istikhomah, S., Anajib, D. W., & Purnomo, A. D. (2019). Rainfall-runoff and sediment yield modeling in volcanic catchment using SWAT, a case study in Opak Watershed. *IOP Conference Series: Earth and Environmental Science, 256*, 012015. https://doi.org/10.1088/1755-1315/256/1/012015

Cimen, M. (2008). Estimation of daily suspended sediments using support vector machines. *Journal Hydrological Sciences Journal, 53*(3), 656–666. https://doi.org/10.1623/hysj.53.3.656

Doroudi, S., Sharafati, A., & Mohajeri, S. H. (2021). Estimation of daily suspended sediment load using a novel hybrid support vector regression model incorporated with observer-teacher-learner-based optimization method. *Complexity*. https://doi.org/10.1155/2021/5540284

Dutta, S., & Sen, D. (2018). Application of SWAT model for predicting soil erosion and sediment yield. *Sustainable Water Resources Management, 4*, 447–468. https://doi.org/10.1007/s40899-017-0127-2

Essam, Y., Huang, Y. F., Birima, A. H., Ahmed, A. N., & El-Shafie, A. (2022). Predicting suspended sediment load in Peninsular Malaysia using support vector machine and deep learning algorithms. *Scientific Reports, 12*(1), 1–29.

Gudino-Elizondo, N., Biggs, T. W., Bingner, R. L., Langendoen, E. J., Kretzschmar, T., Taguas, E. V., Taniguchi-Quan, K. T., Liden, D., & Yuan, Y. (2019). Modelling runoff and sediment loads in a developing coastal watershed of the US-Mexico border. *Water, 11*, 1024. https://doi.org/10.3390/w11051024

Harun, M. A., Safari, M. J. S., Gul, E., & Ab Ghani, A. (2021). Regression models for sediment transport in tropical rivers. *Environmental Science and Pollution Research, 28*(38), 53097–53115.

Hussian, M., Grimvall, A., Burdakov, O., & Sysoev, O. (2005). Monotonic regression for the detection of temporal trends in environmental quality data. *MATCH Communication Mathematics Computational Chemistry, 54*, 535–550.

Jothiprakash, V., & Garg, V. (2009). Reservoir sedimentation estimation using artificial neural network. *Hydrologic Engineering, 14*(9), 1035–1040. https://doi.org/10.1061/ASCEHE.1943-5584.0000075

Kim, B., Choi, S. Y., & Han, K. Y. (2019). Integrated real-time flood forecasting and inundation analysis in small-medium streams. *Water, 11*, 919. https://doi.org/10.3390/w11050919

Kisi, O. (2012). Modeling discharge-suspended sediment relationship using least square support vector machine. *Journal of Hydrology, 456–457*, 110–120.

Kisi, O., Dailr, A.H., Cimen, M., & Shiri, J. (2012). Suspended sediment modeling using genetic programming and soft computing techniques. *Journal of Hydrology*, *450–451*, 48–58. https://doi.org/10.1016/j.jhydrol.2012.05.031.

Kisi O, Ozkan C (2017). A new approach for modeling sediment-discharge relationship: local weighted linear regression. *Water Resources Management: An International Journal, Published for the European Water Resources Association (EWRA), Springer; European Water Resources Association (EWRA), 31*(1):1–23.

Khosravi, K., Sartaj, M., Tsai Frank, T. C., Singh, V. P., Kazakis, N., Melesse, A. M., Prakash, I., Tien Bui, D., & Pham, B. H. (2018). A comparison study of DRASTIC methods with various objective methods for groundwater vulnerability assessment. *Science of the Total Environment, 642,* 1032–1049. https://doi.org/10.1016/j.scitotenv.2018.06.130.

Larson, M. D., SimicMilas, A., Vincent, R. K., & Evans, J. E. (2021). Landsat 8 monitoring of multi-depth suspended sediment concentrations in Lake Erie's Maumee River using machine learning. *International Journal of Remote Sensing, 42*(11), 4064–4086.

Liu, Y., & Jiang, H. (2019). Sediment yield modeling using SWAT model: Case of Changjiang River Basin. *IOP Conference Series: Earth and Environmental Science, 234*, 012031. https://doi.org/10.1088/1755-1315/234/1/012031

Liua, Q. Q., Chena, L., Lia, J. C., & Singh, V. P. (2004). Two-dimensional kinematic wave model of overland-flow. *Journal of Hydrology, 291*, 28–41.

Meshram, S. G., Ghorbani, M. A., Deo, R. C., Kashani, M. H., Meshram, C., & Karimi, V. (2019). New approach for sediment yield forecasting with a two-phase feed forward neuron network-particle swarm optimization model integrated with the gravitational search algorithm. *Water Resource Management*, *33*(7), 2335–2356. https://doi.org/10.1007/s11269-019-02265-0

Meshram, S.G., Singh, V.P., Kisi, O., Karimi, V., & Meshram, C. (2020). Application of artificial neural networks, support vector machine and multiple model- ANN to sediment yield prediction. *Water Resource Management*. https://doi.org/10.1007/s11269-020-02672-8

Meshram, S. G., Safari, M. J. S., Khosravi, K., & Meshram, C. (2021). Iterative classifier optimizer-based pace regression and random forest hybrid models for suspended sediment load prediction. *Environmental Science and Pollution Research, 28*(9), 11637–11649.

Misra, D., Oommen, T., Agarwal, A., Mishra, S. K., & Thompson, A. M. (2009). Application and analysis of support vector machine based simulation for runoff and sediment yield. *Biosystems Engineering, 103*, 527–535.

Mohammadi, B., Guan, Y., Moazenzadeh, R., & Safari, M. J. S. (2021). Implementation of hybrid particle swarm optimization-differential evolution algorithms coupled with multi-layer perceptron for suspended sediment load estimation. *CATENA, 198*, 105024.

Naderloo, L., Alimardani, R., Omid, M., Sarmadian, F., Javadi-kiaTorabi, M. Y., & Alimardani, F. (2012). Application of ANFIS to predict crop yield based on different energy inputs. *Measurement, 45*(6), 1406–1413.

Nagatsuka, H., Uchino, M., & Yamamoto, H. (2012). Parameter estimation of multivariate distributions under order restrictions of the parameters: An extension of isotonic regression. *Quality Technology & Quantitative Management, 9*(3), 283–293. https://doi.org/10.1080/16843703.2012.11673292.

Nicklow, J. W., Mays, L. W. (2000). Optimization of Multiple Reservoir Networks for Sedimentation Control. *Journal of Hydraulic Engineering*. https://doi.org/10.1061/(ASCE)0733-9429(2000)126:4(232)

Omondi, R. A., & Rajapakse, C. J. (2010). *FPGA implementations of neural networks* (1st edn.). Springer.

Patel, D., & Parekh, F. (2014). Flood forecasting using adaptive neuro-fuzzy inference system (ANFIS). *International Journal of Engineering Trends and Technology (IJETT), 12*(10), 510–514.

Pham, B. T., Prakash, I., Chen, W., Ly, H. B., Ho, L. S., Omidvar, E., Tran, V. P., & Bui, D. T. (2019). A novel intelligence approach of a sequential minimal optimization-based support vector machine for landslide susceptibility mapping. *Sustainability, 11*, 6323. https://doi.org/10.3390/su11226323

Platt, J. (1999). Fast training of support vector machines using sequential minimal optimization. In B. Schölkopf, C. J. C. Burges, & A. J. Smola (Eds.), *Advances in kernel methods—Support vector learning* (pp. 185–208). Cambridge, MA: MIT Press.

Polucci, D., Marchetti, M., & Fiori, S. (2020). A novel non-isotonic statistical bivariate regression method—Application to stratigraphic data modeling and interpolation. *Mathematics Computing Application, 25*(1), 15. https://doi.org/10.3390/mca25010015

Safari, M. J. S. (2020). Hybridization of multivariate adaptive regression splines and random forest models with an empirical equation for sediment deposition prediction in open channel flow. *Journal of Hydrology, 590*, 125392. https://doi.org/10.1016/j.jhydrol.2020.125392

Safari, M. J. S., & Shirzad, A. (2019). Self-cleansing design of sewers: Definition of the optimum deposited bed thickness. *Water Environment Research, 91*(5), 407–416. https://doi.org/10.1002/wer.1037

Samadianfard, S., Kargar, K., Shadkani, S., Hashemi, S., Abbaspour, A., & Safari, M. J. S. (2022). Hybrid models for suspended sediment prediction: Optimized random forest and multi-layer perceptron through genetic algorithm and stochastic gradient descent methods. *Neural Computing and Applications, 34*(4), 3033–3051.

Samantaray, S., Sahoo, A., & Ghose, D. K. (2020). Assessment of sediment load concentration using SVM, SVM-FFA and PSR-SVM-FFA in arid watershed, India: A case study. *KSCE Journal of Civil Engineering, 24*, 1944–1957. https://doi.org/10.1007/s12205-020-1889-x

Schneider, W. (2018). On basic equations and kinematic-wave theory of separation processes in suspensions with gravity, centrifugal and Coriolis forces. *Acta Mechanica, 229*, 779–794. https://doi.org/10.1007/s00707-017-1998-x

Sharafati, A., Haji SeyedAsadollah, S. B., Motta, D., & Yaseen, Z. M. (2020b). Application of newly developed ensemble machine learning models for daily suspended sediment load prediction and related uncertainty analysis. *Hydrological Sciences Journal*. https://doi.org/10.1080/02626667.2020.1786571

Sharafati, A., Pezeshki, E., Shahid, S., & Motta, D. (2020a). Quantification and uncertainty of the impact of climate change on river discharge and sediment yield in the Dehbar river basin in Iran. *Journal of Soils and Sediments, 20*(7), 2977–2996. https://doi.org/10.1007/s11368-020-02632-0

Singh, H. V., Panuska, J., & Thompson, A. M. (2017). Estimating sediment delivery ratios for grassed waterways using WEPP. *Land Degradation and Development*. https://doi.org/10.1002/ldr.2727

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*, 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Talebizadeh, M., Morid, S., Ayyoubzadeh, S. A., et al. (2010). Uncertainty analysis in sediment load modeling using ANN and SWAT model. *Water Resources Management, 24*, 1747–1761. https://doi.org/10.1007/s11269-009-9522-2

Tiwari, S., Babbar, R., & Kaur, G. (2018). Performance evaluation of two ANFIS Models for predicting water quality index of River Satluj (India). *Advance in Civil Engineering*. https://doi.org/10.1155/2018/8971079

Tuset, J., Vericat, D., & Batalla, R. J. (2015). Rainfall, runoff and sediment transport in a Mediterranean mountainous catchment. *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv.2015.07.075

Ullaha, N., & Choudhury, P. (2010). Flood forecasting in river system using ANFIS. *AIP Conference Proceedings, 1298*, 694. https://doi.org/10.1063/1.3516407

Vaheddoost, B., Vazifehkhah, S., & Safari, M. J. S. (2022). A stochastic approach for the assessment of suspended sediment concentration at the Upper Rhone River basin, Switzerland. *Environmental Science and Pollution Research, 29*(26), 39860–39876.

Veni, S., & Srinivasan, A. (2020). Comparison of linear regression and isotonic regression analysis implemented for project management in software development life cycle. *International Journal of Engineering Sciences and Research Technology, 6*(12).

Yadav, A., Chatterjee, S., & Equeenuddin, S. M. (2018). Suspended sediment yield estimation using genetic algorithm-based artificial intelligence models: Case study of Mahanadi River, India. *Hydrological Sciences Journal, 63*(8), 1162–1182. https://doi.org/10.1080/02626667.2018.1483581

Yuan, X., Chen, C., Lei, X., Yuan, Y., & Muhammad Adnan, R. (2018). Monthly runoff forecasting based on LSTM-ALO model. *Stochastic Environment Research Risk Assessment, 32*(8), 2199–2212. https://doi.org/10.1007/s00477-018-1560-y

(Received February 12, 2022, revised August 10, 2022, accepted August 11, 2022, Published online September 29, 2022)