

Research Article

Application of MGGP, ANN, MHBMO, GRG, and Linear Regression for Developing Daily Sediment Rating Curves

Majid Niazkar ¹ and Mohammad Zakwan ²

¹Department of Civil and Environmental Engineering, Shiraz University, Shiraz, Iran

²Civil Engineering Department, MANUU, Hyderabad, India

Correspondence should be addressed to Majid Niazkar; mniazkar@shirazu.ac.ir

Received 5 September 2021; Revised 25 October 2021; Accepted 6 December 2021; Published 23 December 2021

Academic Editor: Danial Armaghani

Copyright © 2021 Majid Niazkar and Mohammad Zakwan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A data-driven relationship between sediment and discharge of a river is among the most erratic relationships in river engineering due to the existence of an inevitable scatter in sediment rating curves. Recently, Multigene Genetic Programming (MGGP), as a machine learning (ML) method, has been proposed to develop data-driven models for various phenomena in the field of hydrology and water resource engineering. The present study explores the capability of MGGP-based models to develop daily sediment ratings of two gauging sites with 30-year sediment-discharge data, which was utilized previously in the literature. The results obtained by MGGP were compared with those achieved by an empirical model and Artificial Neural Network (ANN). The coefficients of the empirical model were calibrated using linear and nonlinear regression models (Generalized Reduced Gradient (GRG) and the Modified Honey Bee Mating Optimization (MHBMO) algorithm). According to the comparative analysis, the mean absolute error (MAE) at the two gauging stations reduced from 516.54 to 519.23 obtained by nonlinear regression to 447.26 and 504.23 achieved by MGGP, respectively. Similarly, all other performance indices indicated the suitability and accuracy of MGGP in developing sediment ratings. Therefore, it was demonstrated that ML-based models, particularly MGGP-based models, outperformed the empirical models for estimating sediment loads.

1. Introduction

Climate change and the surge in water demands have created a need for better strategies in river engineering [1, 2]. In this regard, a close-to-reality estimation of sediment loads transported by a river plays a vital role in developing river management strategies [3]. Furthermore, knowledge of sediment transport is also important for various river engineering applications, including channel restoration, reservoir sedimentation [4], soil loss [5], predicting bed roughness coefficients [6], water quality analysis [7], and design of hydraulic structures. With the advent of computational techniques, river management practices can be addressed more precisely to a certain extent.

Conventionally, sediment loads in a river are predicted through rating curves obtained by curve fitting techniques or the application of sediment transport formulas in the

presence of requisite data [3, 8–11]. However, Walling [8] observed significant uncertainty in the sediment loads predicted by curve fitting methods. Similar observations were made by Ferguson [12], who asserted that the sediment load is undermined by sediment rating curves. Additionally, Sickingabula [13] suggested that sediment ratings are not only rather complex and not always single-valued but also influenced by antecedent moisture, catchment hydraulics, and meteorology of the region. Furthermore, Haddadchi and Hicks [14] recommended that the difference in the sediment load arising from various land cover conditions of a catchment is more pronounced in case of frequently occurring low magnitude events. Accounting for the complexity of sediment-discharge relationships, researchers have sought artificial intelligence (AI) and machine learning (ML) methods as an alternative to the simple conventional sediment rating.

Tawfik et al. [15] proposed the application of Artificial Neural Network (ANN) to develop ratings. Since then, various AI and ML approaches have been employed to improve the accuracy of discharge and sediment ratings. For instance, Aytek and Kisi [16] employed Genetic Programming (GP) to model the sediment ratings. Jain [17] used a compound neural network to develop the sediment-discharge relationship. Furthermore, Cimen [18] applied support vector machine to estimate suspended sediment loads. Cobaner et al. [19] proposed the application of adaptive neuro-fuzzy for predicting sediment loads. Joshi et al. [20] demonstrated the superiority of ANN over the conventional rating curve for estimating sediment loads. In addition, Zakwan et al. [21], Adnan et al. [1], and Gupta et al. [22] presented a review of applications of data-driven and AI techniques for developing discharge and sediment ratings. Sharafati et al. [23] proposed the application of random forest regression for developing sediment ratings. Also, Bajirao et al. [24] recommended preprocessing of sediment-discharge data using wavelet transform to improve the estimates of sediment ratings obtained by applying AI techniques. Aldahoul et al. [25] reported the superiority of the long short-term memory algorithm in predicting suspended sediment concentrations as compared to ANN and the regression technique. Recently, Adnan et al. [5] proposed the application of hybrid multivariate adaptive regression spline and the K-means clustering algorithm to estimate the sediment load and reported it as a more efficient method than the model tree and regression analysis. Basically, the studies that applied AI and ML methods for sediment estimations indicate that these models generally perform better than the conventional approaches and common regression techniques. Such applications have received attention because (1) the sediment transport is substantially a complicated phenomenon and cannot be modelled easily, (2) there is no generalized formula with a desirable accuracy for predicting sediment discharges, (3) available sediment transport equations may result into sediment estimations, which sometimes are an order of magnitudes different from one another, and (4) numerical models require extensive data acquisition for calibration and estimation phases [26]. Thus, with the advancement of AI and ML models, the problem of sediment estimation is revisited in light of improving sediment load predictions.

Recently, a new ML approach, known as Multigene Genetic Programming (MGGP), has been used for various phenomena in the field of hydrology and water resource engineering [27, 28]. Based on the performance of MGGP in previous applications and the need for developing sediment ratings with higher accuracy, the present study aims to investigate the capability of MGGP to model the sediment ratings and compare the performance of MGGP models with other prevalent techniques. The parameters of the empirical model of sediment ratings were calibrated by linear regression and two different optimization algorithms (the Modified Honey Bee Mating Optimization (MHBMO) algorithm and Generalized Reduced Gradient (GRG)). To the authors' knowledge and based on the literature review, it is the first time the MHBMO algorithm and MGGP are used

for this purpose. Finally, the comparative analysis was conducted for sediment load estimates obtained by linear and nonlinear regression models: ANN and MGGP.

2. Materials and Methods

2.1. Datasets. In the present study, the sediment-discharge data of two gauging sites was used. The data were utilized previously in [3] and contain 10957 daily discharge (Q) and sediment (S) readings for each site. The first data correspond to Botovo, while the second dataset corresponds to Donji Miholjac gauging site on Drava River Basin, Croatia. Both datasets were randomly divided into two parts: (1) training data (75% of each dataset) and (2) testing data (25% of each dataset). The former was used for calibrating and training different models, whereas the latter was exploited for testing the predicted sediment loads with the observed ones. The observed daily sediment-flow discharges for the two sites are depicted in Figure 1. As shown, the data represents a highly complex relationship between sediment and discharge values measured at two gauging sites. Moreover, Table 1 presents the characteristics of the training and testing data for both datasets.

Figure 2 depicts the location of two gauging sites on the lower Drava River. As shown, the Botovo gauging site with an average elevation of 121.55 m above the mean sea level is around 150 km upstream of the Donji Miholjac (88.50 m elevation). The variation in discharge and sediment load values at the two gauging sites is very small on the monthly or seasonal scale [3].

2.2. Methods. This study primarily assesses the application of MGGP for developing sediment rating curves, which can be exploited to estimate a sediment discharge when a flow discharge is available. According to the literature, various techniques have already been utilized for developing sediment ratings, while MGGP has not been used for this purpose. To be more specific, the performances of different approaches, including linear regression, nonlinear regression (MHBMO and GRG), and machine learning methods (ANN and MGGP), are compared for finding the best method that develops an accurate relation between sediment discharge and discharge values at typical measuring stations. For this comparative analysis, 30-year sediment-discharge data measured at two gauging sites were utilized as the case studies. The collected data were divided into train and test parts. In other words, the problem statement of this study consists of developing sediment rating curves for the train data and using them to estimate sediment loads for the test data.

In general, the sediment rating curve, which relates a sediment load with its corresponding discharge, can be formulated using the sediment rating equation as follows:

$$S = aQ^b, \quad (1)$$

where a and b are fixed coefficients, which are calibrated using different methods when an observed database is available. In this study, linear and two nonlinear regression

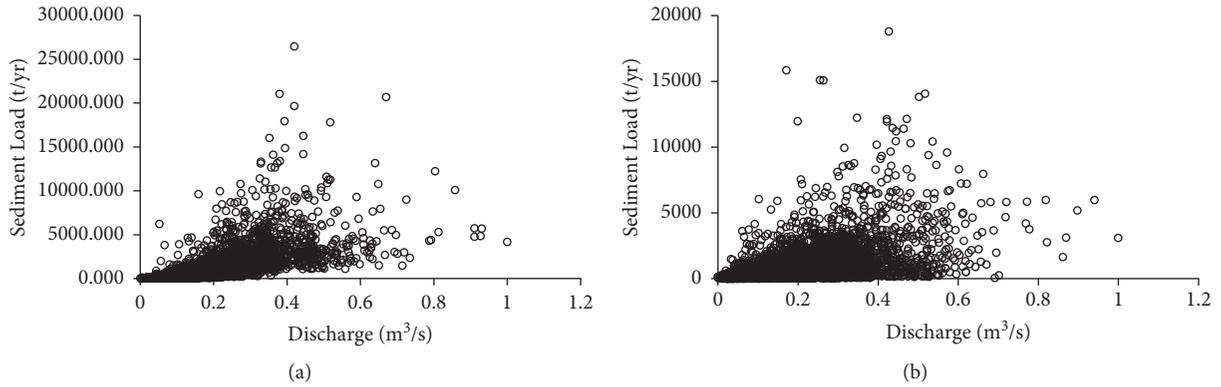


FIGURE 1: Sediment load versus discharge for (a) Botovo and (b) Donji Miholjac gauging stations.

TABLE 1: Sediment and flow discharges of the train and test data of the two stations.

Station	Type of data	Parameter	Number of data points	Maximum	Minimum	Mean
Botovo	Training data	Q (m ³ /s)	8218	2345.00	103.00	474.09
		S (t/year)	8218	26454.38	0.54	830.26
	Testing data	Q (m ³ /s)	2739	2217.00	134.00	476.51
		S (t/year)	2739	22481.71	1.87	832.31
Donji Miholjac	Training data	Q (m ³ /s)	8218	2166.00	175.00	509.39
		S (t/year)	8218	18774.72	0.18	726.09
	Testing data	Q (m ³ /s)	2739	2140.00	190.00	509.41
		S (t/year)	2739	17280.60	0.20	724.87

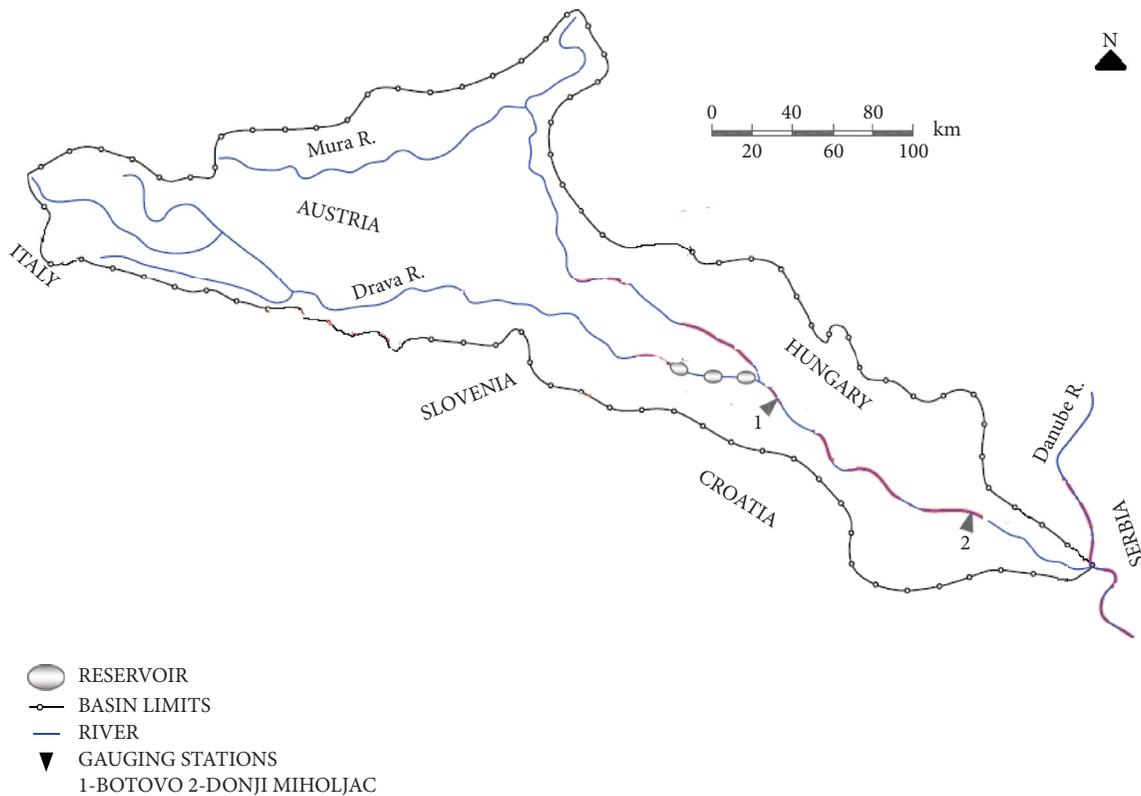


FIGURE 2: Sediment gauging sites on Lower Drava river [3].

methods were utilized for calibrating two parameters of the sediment rating curve, which are presented below.

2.2.1. Linear Regression Method. By applying the logarithm function to both sides of equation (1), the constant coefficients of equation (1) can be obtained by fitting a straight line to the logarithm sediment discharge-flow discharge plot:

$$\begin{aligned}\log(S) &= \log(aQ^b) \\ &= b \times \log(Q) + \log(a).\end{aligned}\quad (2)$$

2.2.2. Nonlinear Regression Methods. For calibrating a and b coefficients, two nonlinear regression approaches were exploited, which are presented as follows.

(1) Generalized Reduced Gradient. GRG is a gradient-based nonlinear optimization technique whose search direction is based on the steepest descent. In this optimization technique, the Quasi-Newton approach is generally applied to determine the optimal solution [29]. The dependence of search for an optimal solution on gradient involves chances of getting trapped in local optimal solution. Therefore, a multistart option has been used as a check. This first-order optimization algorithm has been widely used for developing discharge ratings [21, 27], flood routing [29], and modelling infiltration [28] and predicting a scour depth around bridge piers [30]. In the present work, the sediment rating shown in equation (1) is modelled with the help of the GRG technique.

(2) The MHBMO Algorithm. The Honey Bee Mating Optimization (HBMO) algorithm is a zero-order optimization algorithm, which was inspired by a mating process in a honey bee colony. In essence, the mating process is simulated as a brood-population generation, which occurs between drones and the queen of the colony. Basically, the queen is capable of storing a set of different drone's sperms in its spermatheca. This feature may provide the chance of producing better brood generation, which may be considered as an advantage of the HBMO algorithm in comparison with other evolutionary optimization algorithms. Nevertheless, the brood generation in the original version may get stuck in some local optima, which is inevitably counted as a convergence problem. However, the modified version (i.e., the MHBMO algorithm) attempts to improve the brood generation, while this improvement was carried out by considering three sperms, which are randomly selected from the queen's spermatheca to produce new improved drones. This characteristic made the MHBMO algorithm a powerful search-based algorithm, which has been used as an optimization tool for various applications in the field of water resources [30].

In this study, the MHBMO algorithm was employed to calibrate the coefficients of the sediment estimation model. For this purpose, the five main controlling parameters of the MHBMO algorithm were set as follows: number of initial population = 1000, number of broods = 500, number of

drones = 500, spermatheca size = 1500, and number of workers = 10. Finally, other MHBMO parameters were assumed similar to previous studies available in the literature [30, 31].

2.2.3. ML-Based Methods. In addition to the regression models, two ML methods were employed to develop sediment rating curves. These approaches are introduced below.

(1) Artificial Neural Network. In the parlance of AI methods, ANN is technically an estimation model that attempts to find a relation between an input dataset with an output dataset without considering the physical background of the data involved [32]. Generally, the input data is inserted into an input layer, whereas the output layer takes the output data. In addition to these two layers, one or more layers, so-called hidden layers, constitute the ANN architecture. Furthermore, each layer consists of neurons, which are allowed to have connections with neurons of the adjacent layers [30]. The data flow among the ANN input and output layers becomes possible with the aid of the neurons of the hidden layer(s). Such flow provides developing an estimation model when either a desirable accuracy is achieved or the maximum number of back-and-forth data flow is reached. The flexible structure of ANN has made it a widely-used AI method in the literature of water resources [31].

In this study, a feed-forward backpropagation ANN was exploited to predict sediment load. To be more specific, the dimensionless discharge and dimensionless sediment discharge for the daily sediment data are used as input and output variables. The dimensionless sediment loads estimated by ANN were turned into sediment discharge with dimensions in the comparative analysis. All controlling parameters of ANN were assumed similar to those used in previous studies [31].

(2) Multigene Genetic Programming. GP is a powerful ML method, whose development was merely to address the need of GA users. To be more precise, GP provides the opportunity of determining a prediction model with the aid of a tree-like structure and the GA search engine, while this application cannot be done with the exclusive use of GA and inevitably requires further coding on GA [32]. As a result, the fundamental steps of GP are those presumed in GA. As the classical GP has been paid quite considerable attention by researchers in different fields of research [33], a few modified variants of this ML method were developed. MGGP, as a modified version of GP, is capable of developing an estimation model between known input and output vectors. The main difference between GP and MGGP is the number of genes (trees) used in each individual (relation). To be more specific, the relation developed by MGGP is a weighted summation of one or more genes, whereas GP has only one tree in each individual. This feature suits MGGP for tackling problems whose input and output variables may have a complicated relationship [27].

In this study, a MATLAB code of MGGP adopted from the literature [34] was used. This MGGP code has already been utilized for other applications in water

resources [27]. The objective function of MGGP is to minimize the root mean square of errors between the predicted and measured dimensionless sediment loads for each scenario and dataset separately. The MGGP runs more than 50 times, which is a standard value established in the literature for this purpose [27], while the stopping criterion was 0.001. In addition, the controlling parameters of MGGP were set similar to previous studies in [28]. Although MGGP can seek for an estimation model without the need for specifying the exact type of equation for it, different embedded functions, such as the four basic mathematical operations, trigonometric functions, square function, and

exponential function, were allowed to participate in the process of developing the prediction model. Finally, normalized sediment loads estimated by the best MGGP-based model for each dataset and each scenario were tuned into sediment-discharge values with dimension for comparison purposes.

2.3. Performance Evaluation Criteria. The performance of prediction models for sediment ratings was compared based on several criteria, which are presented in (3) to (8) [32]:

$$\text{root mean square error (RMSE)} = \sqrt{\frac{\sum_{i=1}^N (S_{obs_i} - S_{est_i})^2}{N}}, \quad (3)$$

$$\text{mean absolute error (MAE)} = \frac{\sum_{i=1}^N |S_{obs_i} - S_{est_i}|}{N}, \quad (4)$$

$$\text{mean absolute relative error (MARE)} = \frac{1}{N} \sum_{i=1}^N \left| \frac{S_{obs_i} - S_{est_i}}{S_{obs_i}} \right|, \quad (5)$$

$$\text{maximum absolute relative error (MXARE)} = \max \left(\left| \frac{S_{obs_i} - S_{est_i}}{S_{obs_i}} \right| \right) \text{ for } i = 1, \dots, N, \quad (6)$$

$$\text{relative error (RE)} = \frac{S_{obs_i} - S_{est_i}}{S_{obs_i}} \text{ for } i = 1, \dots, N, \quad (7)$$

$$\text{determination coefficient}(R^2) = \left\{ \frac{\sum_{i=1}^N [(S_{obs_i} - \sum_{i=1}^N S_{obs_i}/N)(S_{est_i} - \sum_{i=1}^N S_{est_i}/N)]}{\sum_{i=1}^N [(S_{obs_i} - \sum_{i=1}^N S_{obs_i}/N)^2 (S_{obs_i} - \sum_{i=1}^N S_{est_i}/N)^2]} \right\}^2, \quad (8)$$

where S_{obs_i} is the i^{th} of the observed sediment load and S_{est_i} is the i^{th} of the estimated sediment load. Among the considered metrics, RMSE, MAE, MARE, MXARE, and RE should be as low as possible for the model with the highest accuracy, whereas the higher value R^2 is, the better a model performs. Moreover, RMSE, MAR, MARE, and R^2 are global indicators that represent the accuracy of each model for the whole dataset, whereas MXARE and RE are local indicators, which focus on evaluating the performance of each model for a data point [30].

3. Results and Discussion

Sediment load estimates are an essential input for many hydropower and river management projects. Generally, a large scatter exists in a sediment-discharge relationship in most rivers because, in reality, sediment loads in a river depend on some factors, such as moisture condition, land cover of watershed, distance of sediment sources, and channel slope. However, most of these data are seldom available. Likewise, in the present study, large scatters exist

in the sediment rating at each gauging site under investigation, as shown in Figure 1. To be more precise, addressing the scatter in the sediment-discharge relationship has remained a major concern for the hydrologists, while various ML approaches have been put to use to resolve this issue. In this regard, the present work also explores the possibility of improving the estimates of sediment loads by using two ML models (ANN and MGGP) over the commonly used sediment rating equation (equation (1)). For this purpose, the ML approaches were applied to the daily sediment and discharge data measured at the two gauging sites on the Drava River.

The conventional sediment ratings (equation (1)) were established using the linear regression, the MHBMO algorithm, and GRG technique. The parameters obtained by these techniques are reported in Table 2. It may be observed that the parameters obtained by GRG and MHBMO are the same, while the parameters obtained by linear regression are quite different at either of the gauging sites. In the linear regression approach, the sediment rating equation (equation (1)) is transformed into the linear equation using the

TABLE 2: Calibrated values of the coefficients of the sediment rating equation for two stations.

Method	Station			
	Botovo		Donji Miholjac	
	a	b	a	b
Linear regression	0.0118	2.6458	0.0467	2.0831
GRG	0.0207	1.7148	0.0276	1.6235
MHBMO	0.0207	1.7150	0.0276	1.6235

logarithm of sediment and discharge data as shown in equation (2). Then, a straight line is fitted to logarithm transformed data rather than actual data. On the contrary, GRG and MHBMO are nonlinear optimization techniques, which do not require the transformation of data into a logarithmic form. Hence, these techniques yield different parameters as compared to linear regression.

Before applying the ML methods, the observed data were normalized. For instance, each sediment load was normalized using $\bar{S} = (S_i - S_{obs_{min}}) / (S_{obs_{max}} - S_{obs_{min}})$, where \bar{S} ,

$S_{obs_{max}}$, and $S_{obs_{min}}$ are the normalized, maximum, and minimum sediment discharges and S_i is sediment discharge of the i^{th} observation.

ANN is a black-box model, which does not yield any functional form of relationship between the dependent and independent variables. On the contrary, the equations obtained from MGGP for estimating normalized sediment rates at Botovo and Donji stations are shown, respectively, as follows:

$$\bar{S} = 0.2739q + 0.02603 \sin[\exp(5.074q)] - 0.02552 \sin(11.65q) + 0.2848 \sin[\exp(q^2)] - 0.01196 \sin(14.66q) - 0.2525, \quad (9)$$

$$\bar{S} = 0.403q - 0.02104 \sin[\sin(9.432q)] - 0.01674 \cos[(7.056q)^2] - 0.2015 \sin[\sin(q)] - 0.005402 \cos[(9.432q)^2] - 0.01102 \sin[(q - 9.024)^2] + 0.019, \quad (10)$$

where q is the normalized discharge.

Figure 3 depicts estimated versus observed daily sediment loads for the test data of the first dataset (i.e., the Botovo station). As shown, most of data points are close to the identical line, which indicates that the estimated values are close to the measured ones. However, there are also some data points located far from the identical line, which implies that the developed sediment rating curves may perform inadequately for some data points. Likewise, Figure 4 illustrates the same pattern for the predicted versus observed sediment loads for the test part of the sediment data measured at the second station (i.e., Donji Miholjac). Based on Figures 3-4, a typical sediment rating curve may not perform accurately for the points that have the same discharge but different sediment loads, which may occur particularly at daily time scale measurements. This situation primarily occurs because a sediment rating curve attempts to find a relationship between discharge and sediment discharge, while other variables involved in the complicated process of sediment transport are commonly neglected in the current sediment ratings. Although this is counted as one of the bottlenecks of available sediment rating curves, it provides acceptable approximations of sediment loads for most of data points, as shown in Figures 3-4. Since addressing the aforementioned shortcoming is beyond the scope of this study, including a new variable in the development of sediment ratings in future studies for those data points that have the same sediment discharge but different discharge values is suggested. Obviously, this issue requires

further investigations in this field of research. Finally, as it is difficult to quantify how much each model performs in developing sediment-discharge relationships, the predefined metrics were utilized to conduct a scientific comparative analysis.

The performance of various modelling techniques used in the present study is shown in Figures 5–11. Figures 5–9 represent the RMSE, MAE, R^2 , MARE, and MXARE for the two gauging sites, respectively. Figures 10 and 11 depict the relative error plots for the two sites, respectively. The RMSE, MAE, MARE, and MXARE indices were a magnitude higher in case of linear regression as compared to the other approaches. Consequently, their inclusion in Figures 5–11 would have reduced the clarity of the comparative analysis of other approaches. Therefore, the results of the linear regression are excluded from Figures 5–11. The performance indices of the linear regression are presented in Table 3. It may be observed that the performance of the linear regression technique is the worst at either of the gauging sites. The mechanism of logarithmic transformation of data and retransformation of arithmetic scale in case of linear regression introduces a bias in the estimates [3, 21], leading to poor performance of the approach at both the gauging site.

According to Figure 5, ML-based models (ANN and MGGP) perform better than the nonlinear regression models (GRG and MHBMO) based on the RMSE values of the train data at both measuring stations. To be more specific, the best RMSE value was obtained by MGGP for the train and test data of the Botovo station, while ANN

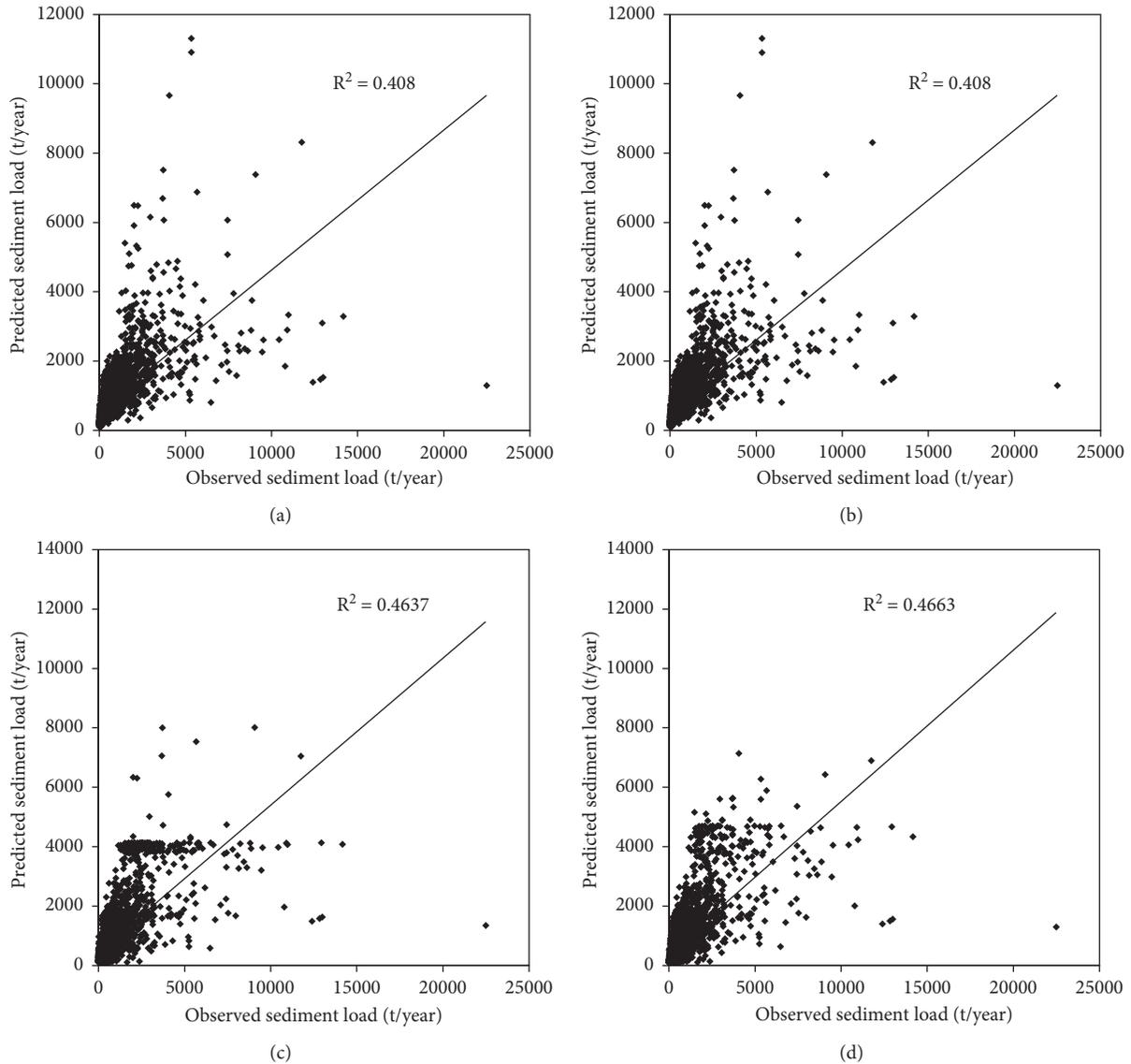


FIGURE 3: Estimated versus observed sediment loads for the test data of the Botovo station.

performs as the second best model for developing sediment rating curves at the first station, as shown in Figure 5(a). Moreover, Figure 5(b) depicts that even though ML-based models achieved better RMSE values for the train data, only MGGP performs better than MHBMO and GRG for the test data at the second station. In other words, MHBMO and GRG yield a better RMSE for the test of the Donji Miholjac station than ANN, where ANN results in a better RMSE than MHBMO and GRG. Therefore, the comparison in terms of RMSE indicates that ML-based models, in particular MGGP, obtained the best results for estimating sediment loads.

The comparison based on MAE, which is shown in Figure 6, obviously demonstrates the outperformance of the ML-based models over the nonlinear regression models for developing sediment rating curves. More precisely, Figure 6 reveals that the MAE values obtained by ANN and MGGP for both stations and both parts of data (i.e., the train and test data) are lower than those achieved by MHBMO and GRG.

Furthermore, Figure 5 indicates that MGGP yields better MAE values than ANN for all scenarios. Thus, the comparison based on the MAE metric clearly implies that applying ML-based models, in particular MGGP, enhances the accuracy of estimating sediment discharges.

Figure 7 compares the performances of different models in terms of R^2 . As shown in Figure 7(a), the ML-based models enhanced R^2 values of other models for both train and test data of the Botovo station. Additionally, Figure 7(a) shows that MGGP achieved slightly better values of R^2 than ANN for both parts of data at the first measuring station. Furthermore, Figure 7(b) indicates that ANN and MGGP result in the best R^2 values for the train and test data of the Donji Miholjac station, respectively. Hence, Figure 7 overall demonstrates that the ML-based models improve the precision of the estimation of daily sediment loads.

Comparing the performances of different models in terms of MARE, presented in Figure 8, obviously shows that

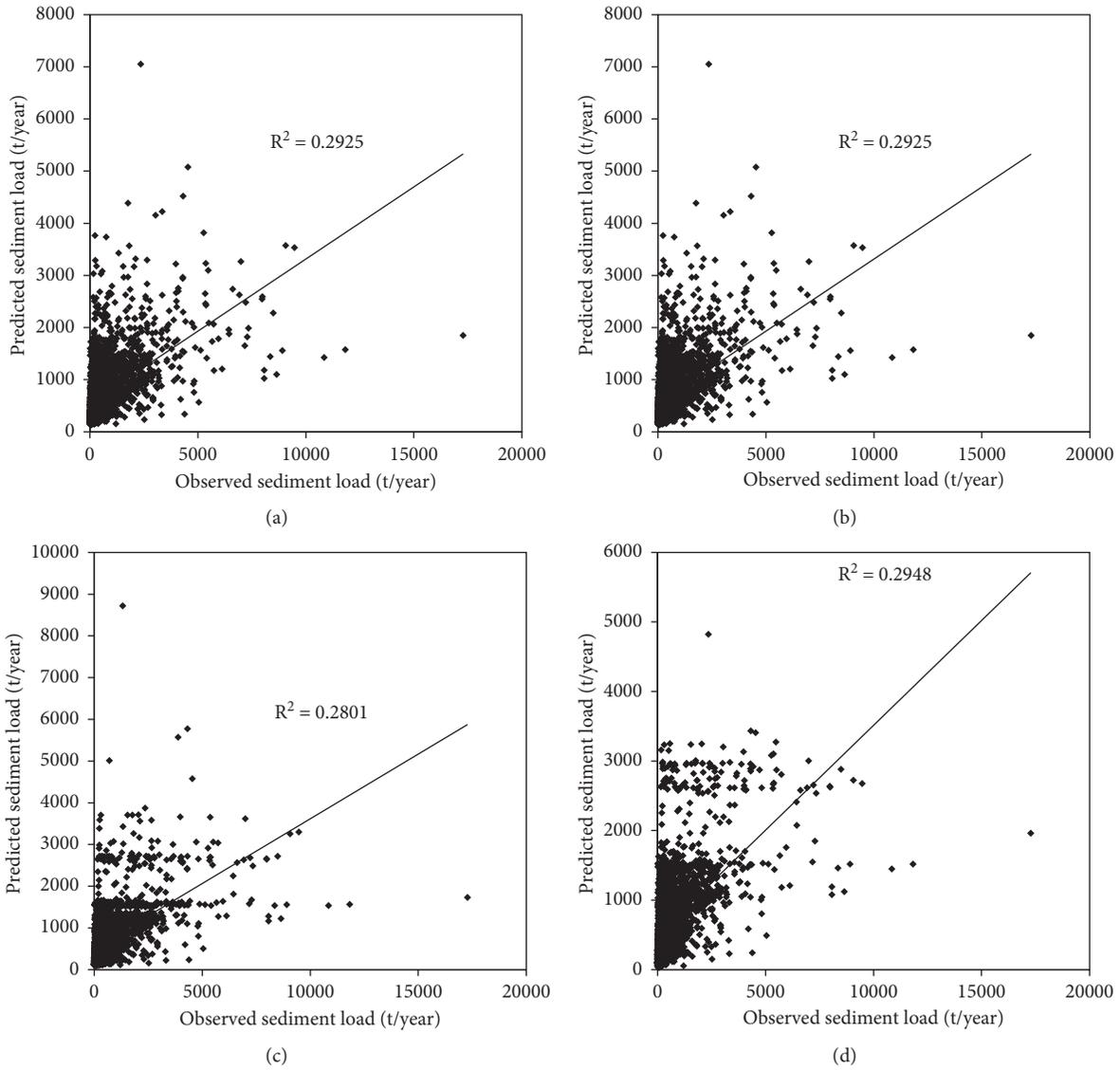


FIGURE 4: Estimated versus observed sediment loads for the test data of the Donji Miholjac station.

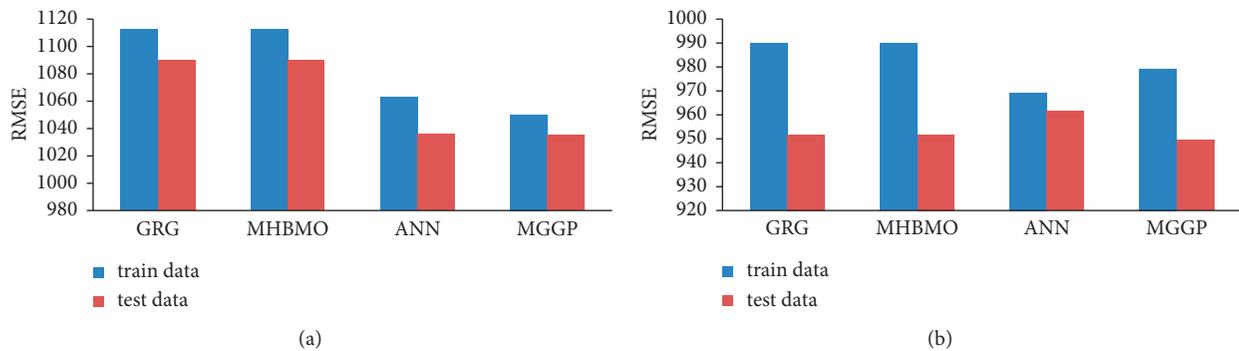


FIGURE 5: Comparison of RMSE for the (a) Botovo and (b) Donji Miholjac gauging stations.

both ML-based methods considered in this study considerably reduced the MARE criterion for both train and test data of both stations. Moreover, MGPP slightly leads to

better MARE values than ANN for all scenarios. The comparative analysis in terms of this metric, like those conducted based on RMSE, MAE, and R^2 , indicates that the

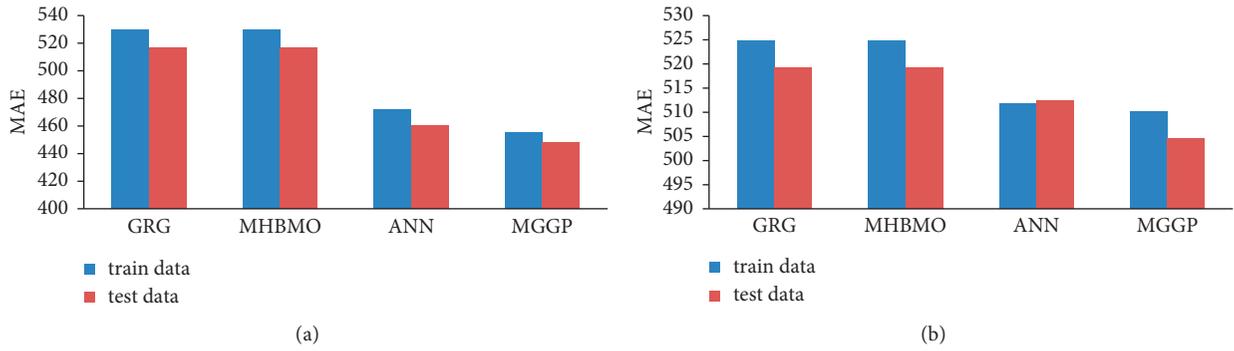


FIGURE 6: Comparison of MAE for the (a) Botovo and (b) Donji Miholjac gauging stations.

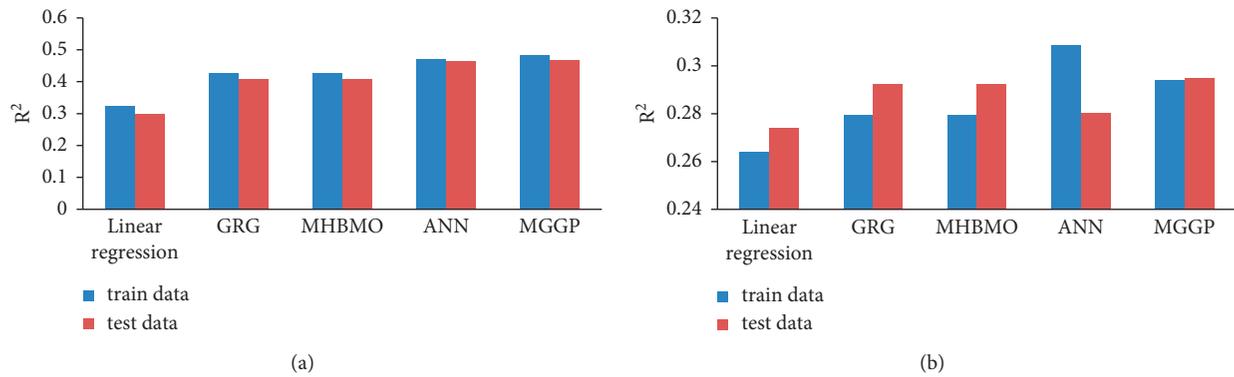


FIGURE 7: Comparison of R^2 for the (a) Botovo and (b) Donji Miholjac gauging stations.

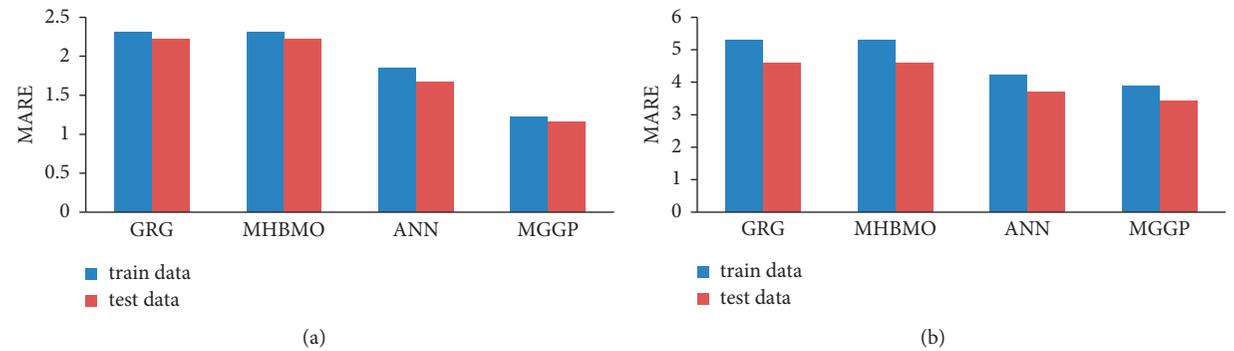


FIGURE 8: Comparison of MARE for the (a) Botovo and (b) Donji Miholjac gauging stations.

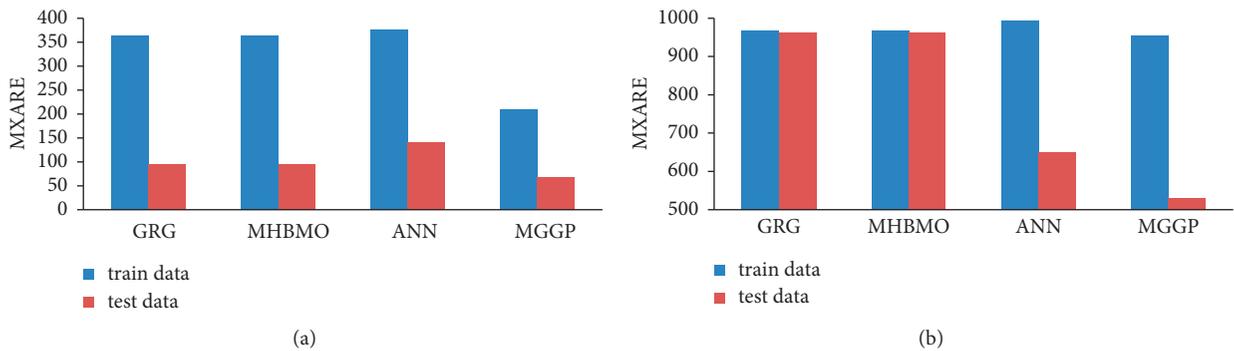


FIGURE 9: Comparison of MXARE for the (a) Botovo and (b) Donji Miholjac gauging stations.

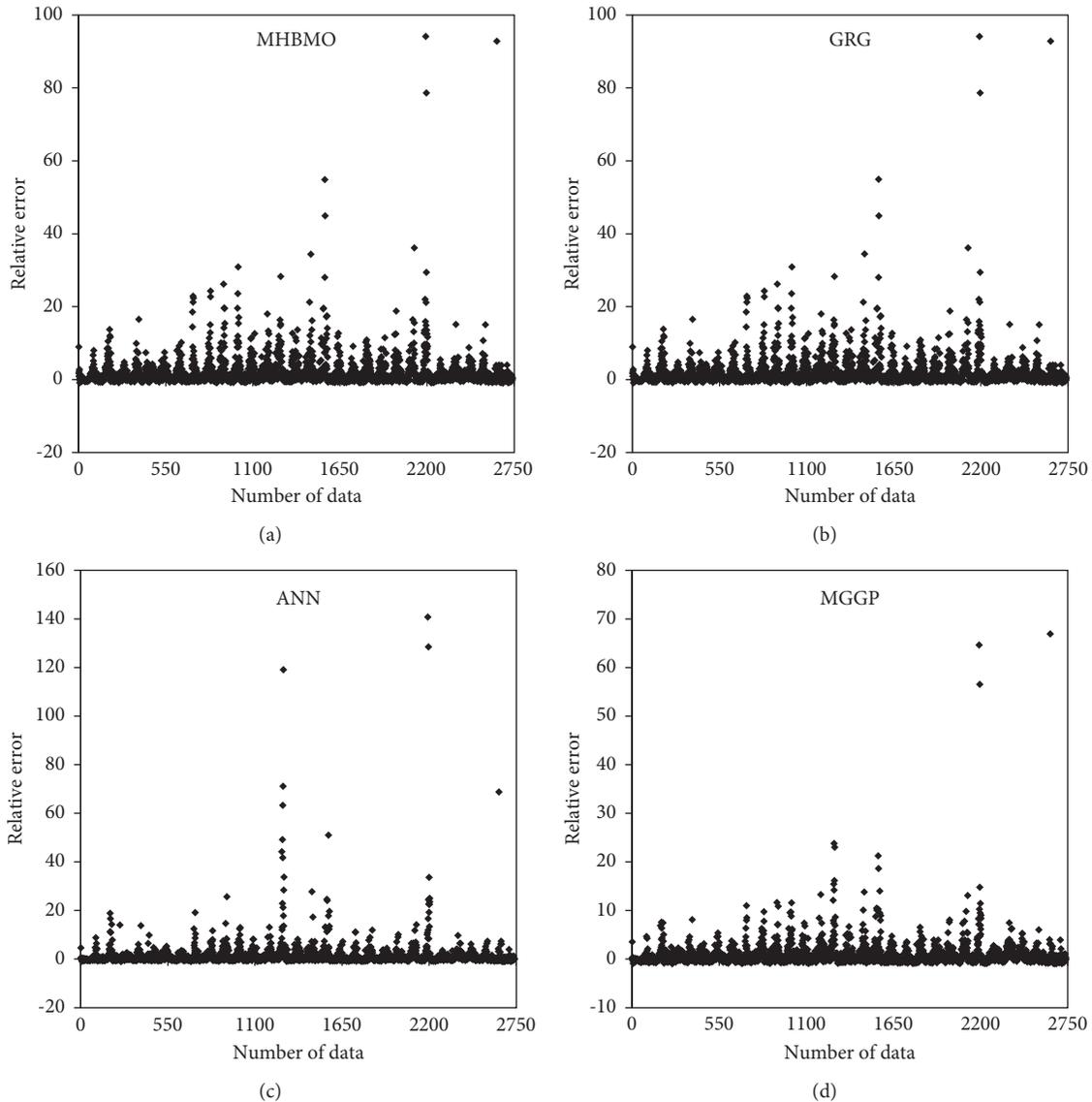


FIGURE 10: Relative errors of daily sediment loads of the test data at station 1 (Botovo).

application of ML methods can provide a sediment rating curve, which predicts daily sediment loads with higher precision. This finding is in agreement with previous studies conducted in the literature [16–18]. Also, the comparison carried out in this study indicates that MGGP, which is proposed in this study for developing sediment ratings for the first time, obtains more accurate sediment-discharge relationships than ANN based on Figures 5–8.

Figure 9 shows that MGGP achieved the lowest values of MXARE for both parts of the data of both measuring stations. Furthermore, the ANN performance is not better than the nonlinear regression models for the first gauging station, while it achieved a better MXARE than MHBMO and GRG for the test of the second station. Therefore, Figure 9 indicates that MGGP performs the best in terms of MXARE, which is a local performance evaluation index for all scenarios.

Figures 10 and 11 illustrate relative errors computed by different models for each data point of the test data for the first and second stations, respectively. As shown in Figure 10, most of the data points (except for three points) estimated by MGGP are located between -10 and 30 , which is the smallest bound compared to those of other models. In addition, Figure 11 demonstrates that the ranges of relative errors calculated by the ML-based models, particularly MGGP, are smaller than those obtained by MHBMO and GRG. Therefore, Figures 10 and 11 indicate the superiority of the ML-based models over the nonlinear regression models in terms of the range of relative errors.

According to Figures 5–11, the performances of ANN and MGGP were considerably better than those of MHBMO and GRG as they resulted in lower error and better coefficient of determination both during the calibration and validation at either of the gauging sites. Basically, MHBMO

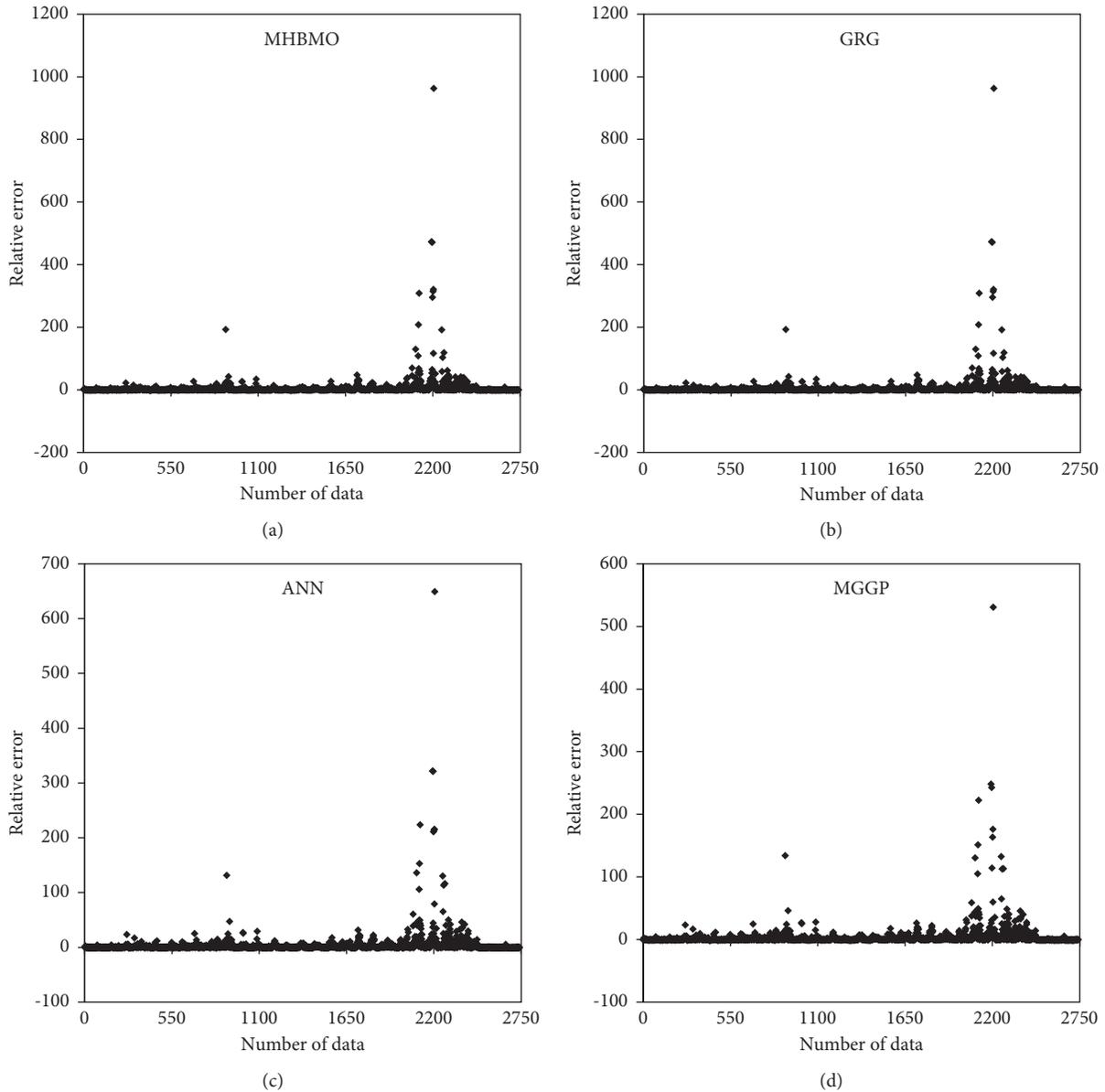


FIGURE 11: Relative errors of daily sediment loads of the test data at station 2 (Donji Miholjac).

TABLE 3: Performance indices obtained by linear regression.

Station	Type of data	RMSE	MAE	MARE	MXARE
Botovo	Training data	482134	225012	454	30211
	Testing data	504901	230654	445	15653
Donji Miholjac	Training data	35720	24119	157	28413
	Testing data	35359	24049	140	20067

and GRG fit the data into predetermined form of equation (equation (1)), whereas the expression in case of ANN and MGGP is not predetermined, leading to a better-fitted equation. Although the application of MGGP improves the estimates, it is at the cost of simplicity. Ergo, the expressions obtained by MGGP (equations (9) and (10)) are relatively

more complex than equation (1). According to Figures 5–11, the performance of MGGP was better than that of ANN based on all the criteria considered for the training and testing phases at the Botovo gauging site. On the contrary, ANN performed better than MGGP during the testing step at the Donji Miholjac site. However, during

testing at the second station, the performance of MGGP was better than ANN, as obviously shown in Figures 5–11, which may be due to overtraining of ANN during the training phase.

Overall, MGGP appears to be a reliable approach for developing the sediment ratings of rivers. Unlike ANN, it yields a functional form of expression for sediment rating, which is often necessary for solving many problems, such as the analytical estimate of effective discharge [3]. Further, MGGP does not require a predetermined structure of equation to fit the data, which makes it more appropriate to fit the haphazard and complex sediment ratings compared to the optimization techniques.

4. Conclusions

Sediment rating curves have various applications in hydropower projects and channel stability. Generally, a wide scatter exists in the daily sediment-discharge data of most rivers. In the absence of a generalized and adequate formula for sediment estimation, robust ML techniques can be counted as a suitable alternative in addressing the scatter in sediment ratings. In the present study, an attempt was made to develop sediment ratings for 30-year daily sediment data at two sites using linear regression, nonlinear regression (MHBMO and GRG), ANN, and MGGP models. Based on the achieved results, the conventional regression approaches provided poor estimates of sediment loads during both training and testing steps. On the contrary, MGGP-based models yielded the best predictions over other approaches for modelling the sediment load at the two gauging sites compared to linear regression (conventional method), nonlinear regression, and ANN. Finally, it is stipulated that applications of powerful ML methods, such as MGGP, may enhance the development of sediment rating curves. Further, the possibility of hybridizing MGGP with other techniques can be explored to improve the estimates of sediment ratings.

Abbreviations

AI:	Artificial intelligence
ANN:	Artificial neural network
GP:	Genetic Programming
GRG:	Generalized Reduced Gradient
HBMO:	Honey Bee Mating Optimization
MAE:	Mean Absolute Error
MARE:	Mean absolute relative error
MGGP:	Multigene Genetic Programming
MHBMO:	Modified Honey Bee Mating Optimization
ML:	Machine learning
MXARE:	Maximum absolute relative error
R^2 :	Determination coefficient
RE:	Relative error
RMSE:	Root mean square error.

Symbols

a :	Fixed coefficient
b :	Fixed coefficient
Q :	Flow discharge

q :	Normalized discharge
S :	Sediment discharge
S_{est} :	Estimated sediment load
S_{obs} :	Observed sediment load
$S_{obs,max}$:	Maximum sediment discharges
$S_{obs,min}$:	Minimum sediment discharges
\bar{S} :	Normalized sediment discharge
S_i :	Sediment discharge of the i^{th} observation.

Data Availability

The data can be provided upon request from the corresponding author. They were used in previous studies.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] R. M. Adnan, Z. Liang, A. E. Shafie, M. K. Zounemat, and O. Kisi, "Prediction of suspended sediment load using data-driven models," *Water*, vol. 11, no. 10, 2019.
- [2] T. Hırca, G. T. Eryılmaz, and M. Niazkar, "Applications of innovative polygonal trend analyses to precipitation series of Eastern Black Sea Basin, Turkey," *Theoretical and Applied Climatology*, 2021.
- [3] M. Zakwan, Q. B. Pham, and S. Zhu, "Effective discharge computation in the lower Drava river," *Hydrological Sciences Journal*, vol. 66, no. 1, 2021.
- [4] S. Zhang, D. Chen, L. Chen, X. Chen, and L. He, "Using sediment rating parameters to evaluate the changes in sediment transport regimes in the middle Yellow River basin, China," *Hydrological Processes*, vol. 33, no. 18, pp. 2481–2497, 2019.
- [5] R. M. Adnan, K. S. Parmar, S. Heddum, S. Shahid, and O. Kisi, "Suspended sediment modeling using a heuristic regression method hybridized with kmeans clustering," *Sustainability*, vol. 13, no. 9, 2021.
- [6] M. Niazkar, "An Excel VBA-based educational module for bed roughness predictors," *Computer Applications in Engineering Education*, vol. 29, no. 5, pp. 1051–1060, 2020.
- [7] K. H. Ahn and S. Steinschneider, "Time-varying, nonlinear suspended sediment rating curves to characterize trends in water quality: an application to the Upper Hudson and Mohawk Rivers, New York," *Hydrological Processes*, vol. 33, no. 13, pp. 1865–1882, 2019.
- [8] D. E. Walling, "Assessing the accuracy of suspended sediment rating curves for a small basin," *Water Resources Research*, vol. 13, no. 3, pp. 531–538, 1977.
- [9] M. B. Jansson, "Comparison of sediment rating curves developed on load and on concentration," *Hydrology Research*, vol. 28, no. 3, pp. 189–200, 1997.
- [10] S. H. Sadeghi and P. Saeidi, "Reliability of sediment rating curves for a deciduous forest watershed in Iran," *Hydrological Sciences Journal*, vol. 55, no. 5, pp. 821–831, 2010.
- [11] M. Delmas, O. Cerdan, B. Cheviron, and J. M. Mouchel, "River basin sediment flux assessments," *Hydrological Processes*, vol. 25, no. 10, pp. 1587–1596, 2011.
- [12] R. I. Ferguson, "River loads underestimated by rating curves," *Water Resources Research*, vol. 22, no. 1, pp. 74–76, 1986.
- [13] H. M. Sickingabula, "Factors controlling variations in suspended sediment concentration for single-valued sediment

- rating curves, Fraser River, British Columbia, Canada,” *Hydrological Processes*, vol. 12, no. 12, pp. 1869–1894, 1998.
- [14] A. Haddadchi and M. Hicks, “Understanding the effect of catchment characteristics on suspended sediment dynamics during flood events,” *Hydrological Processes*, vol. 34, no. 7, pp. 1558–1574, 2020.
- [15] M. Tawfik, A. Ibrahim, and H. Fahmy, “Hysteresis sensitive neural network for modeling rating curves,” *Journal of Computing in Civil Engineering*, vol. 11, no. 3, pp. 206–211, 1997.
- [16] A. Aytek and O. Kişi, “A genetic programming approach to suspended sediment modelling,” *Journal of Hydrology*, vol. 351, no. 3-4, pp. 288–298, 2008.
- [17] S. K. Jain, “Development of integrated discharge and sediment rating relation using a compound neural network,” *Journal of Hydrologic Engineering*, vol. 13, no. 3, pp. 124–131, 2008.
- [18] M. Çimen, “Estimation of daily suspended sediments using support vector machines,” *Hydrological Sciences Journal*, vol. 53, no. 3, pp. 656–666, 2008.
- [19] M. Cobaner, B. Unal, and O. Kisi, “Suspended sediment concentration estimation by an adaptive neuro-fuzzy and neural network approaches using hydro-meteorological data,” *Journal of Hydrology*, vol. 367, no. 1-2, pp. 52–61, 2009.
- [20] R. Joshi, K. Kumar, and V. P. S. Adhikari, “Modelling suspended sediment concentration using artificial neural networks for Gangotri glacier,” *Hydrological Processes*, vol. 30, no. 9, pp. 1354–1366, 2016.
- [21] M. Zakwan, M. Muzzammil, and J. Alam, “Application of data driven techniques in discharge rating curve-an overview. Aquademia,” *Water Environment and Technology*, vol. 1, no. 1, 2017.
- [22] D. Gupta, B. B. Hazarika, M. Berlin, U. M. Sharma, and K. Mishra, “Artificial intelligence for suspended sediment load prediction: a review,” *Environmental Earth Sciences*, vol. 80, no. 9, pp. 1–39, 2021.
- [23] A. Sharafati, S. B. Haji Seyed Asadollah, D. Motta, and Z. M. Yaseen, “Application of newly developed ensemble machine learning models for daily suspended sediment load prediction and related uncertainty analysis,” *Hydrological Sciences Journal*, vol. 65, no. 12, pp. 2022–2042, 2020.
- [24] T. S. Bajirao, P. Kumar, M. Kumar, A. Elbeltagi, and A. Kuriqi, “Superiority of hybrid soft computing models in daily suspended sediment estimation in highly dynamic rivers,” *Sustainability*, vol. 13, no. 2, 2021.
- [25] N. AlDahoul, Y. Essam, P. Kumar et al., “Suspended sediment load prediction using long short-term memory neural network,” *Scientific Reports*, vol. 11, no. 1, pp. 7826–7922, 2021.
- [26] M. H. Garcia, *Sediment Engineering: Processes, Management, Modelling and Practice*, American Society of Civil Engineers, New York, USA, 2008.
- [27] M. Niazkar and M. Zakwan, “Assessment of artificial intelligence models for developing single-value and loop rating curves,” *Complexity*, vol. 2021, Article ID 6627011, 21 pages, 2021.
- [28] M. Zakwan and M. Niazkar, “A comparative analysis of data-driven empirical and artificial intelligence models for estimating infiltration rates,” *Complexity*, vol. 2021, Article ID 9945218, 13 pages, 2021.
- [29] A. R. Nawaz, M. Zakwan, I. Khan, and Z. A. Rahim, “Comparative analysis of variants of Muskingum model,” *Water and Energy International*, vol. 63, no. 7, pp. 64–73, 2020.
- [30] M. Niazkar and S. H. Afzali, “Developing a new accuracy-improved model for estimating scour depth around piers using a hybrid method,” *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, vol. 43, no. 2, pp. 179–189, 2018.
- [31] M. Niazkar, “Assessment of artificial intelligence models for calculating optimum properties of lined channels,” *Journal of Hydroinformatics*, vol. 22, no. 5, pp. 1410–1423, 2020.
- [32] M. Niazkar, “Optimum design of straight circular channels incorporating constant and variable roughness scenarios: assessment of machine learning models,” *Mathematical Problems in Engineering*, vol. 2021, Article ID 9984934, 21 pages, 2021.
- [33] M. Niazkar, F. M. Hajizadeh, and G. T. Eryilmaz, “Assessment of artificial intelligence models for estimating lengths of gradually varied flow profiles,” *Complexity*, vol. 2021, Article ID 5547889, 11 pages, 2021.
- [34] D. Searson, “GPTIPS: Genetic Programming and Symbolic Regression for MATLAB,” Newcastle University Library, Newcastle Upon Tyne, UK, 2009.