Environmental Science and Pollution Research
https://doi.org/10.1007/s11356-021-14479-0

RESEARCH ARTICLE

Regression models for sediment transport in tropical rivers

Mohd Afiq Harun[1] · Mir Jafar Sadegh Safari[2] · Enes Gul[3] · Aminuddin Ab Ghani[1]

Received: 15 February 2021 / Accepted: 14 May 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

The investigation of sediment transport in tropical rivers is essential for planning effective integrated river basin management to predict the changes in rivers. The characteristics of rivers and sediment in the tropical region are different compared to those of the rivers in Europe and the USA, where the median sediment size tends to be much more refined. The origins of the rivers are mainly tropical forests. Due to the complexity of determining sediment transport, many sediment transport equations were recommended in the literature. However, the accuracy of the prediction results remains low, particularly for the tropical rivers. The majority of the existing equations were developed using multiple non-linear regression (MNLR). Machine learning has recently been the method of choice to increase model prediction accuracy in complex hydrological problems. Compared to the conventional MNLR method, machine learning algorithms have advanced and can produce a useful prediction model. In this research, three machine learning models, namely evolutionary polynomial regression (EPR), multi-gene genetic programming (MGGP) and M5 tree model (M5P), were implemented to model sediment transport for rivers in Malaysia. The formulated variables for the prediction model were originated from the revised equations reported in the relevant literature for Malaysian rivers. Among the three machine learning models, in terms of different statistical measurement criteria, EPR gives the best prediction model, followed by MGGP and M5P. Machine learning is excellent at improving the prediction distribution of high data values but lacks accuracy compared to observations of lower data values. These results indicate that further study needs to be done to improve the machine learning model's accuracy to predict sediment transport.

Keywords Machine learning · Sediment transport · Total bed material load · Tropical rivers · Malaysia rivers

Introduction

Sediment transport is a vital element related to river engineering problems. It is important because many issues related to rivers are dependent on sediment mobility. Failure to manage the sedimentation process creates problems such as the reduction of river capacity, flooding, riverbank erosion, riverbed degradation, structure and infrastructure losses, navigation issues and water quality deterioration (van Vuren et al. 2015; Speed et al. 2016; Harun et al. 2020). The input and output of the sediment and the regular disturbance that happened along the river have to be adequately managed to provide a sustainable ecosystem (Templeton and Jay 2013; Frings and Ten Brinke 2017). An estimation of the total bed material load is essential to determine a stable channel design, solving sedimentation problems, predicting scour and floodplain management and preparing hydraulic structure design (Chang 1985; Sinnakaudan et al. 2003; Chang et al. 2005; DID 2009a). There are two main components in total bed material load: suspended load and bed load (Subhasish 2011; Haddadchi et al. 2013; Sulaiman et al. 2017a). The total material load can be estimated by applying direct and indirect approaches (Subhasish 2011). The direct method considers the combination of both bed load and suspended load, whereas the indirect approach separates the bed load and suspended load transport. Many of the total material load equations were derived based on laboratory set up data, which simplified the description of the sedimentation complex phenomenon (Sinnakaudan et al. 2006; Chang et al. 2012; Ab Ghani and Azamathulla 2014).

The uncertainty in the river watershed area presents a challenge in predicting the total bed material load precisely due to

Responsible Editor: Marcus Schulz

✉ Mir Jafar Sadegh Safari
jafar.safari@yasar.edu.tr

[1] River Engineering and Urban Drainage Research Centre (REDAC), Universiti Sains Malaysia, Engineering Campus, 14300, Nibong Tebal, Penang, Malaysia

[2] Department of Civil Engineering, Yaşar University, Izmir, Turkey

[3] Department of Civil Engineering, Inonu University, Malatya, Turkey

Published online: 22 May 2021

the different river data profile and the need to consider sediment properties and characteristics (Molinas and Wu 2001; Syvitski et al. 2014). According to Nagy et al. (2002), sediment transport equations were developed by various theoretical concepts. Bagnold (1996) and Engelund and Hansen (1967) used the power concept to simulate sediment transport processes. Later, Ackers and White (1973) applied stream power and represented sediment transport in the form of dimensionless analysis. Yang (1976) also introduced a sediment transport function based on the analytic power model by stream power per unit weight of the fluid. Laursen (1958), on the other hand, used the functional relationship to establish a connection between sediment discharge and flow condition. Shen and Hung (1972) have used regression analysis based on laboratory results to develop a sediment transport equation. Indeed, Brownlie (1981) also used the same method to develop an equation for sediment transport. This is followed by multiple non-linear regression (MNLR) analysis by Karim and Kennedy (1981) and Karim (1998). All the developed equations have the same downside—the range of the data and the characteristics of the sediments were different from one equation to another. The equations developed by Ackers and White (1973), Engelund and Hansen (1967) and Yang (1976) used data from a flume experiment, where water depth was less than 0.5 m.

For sediment prediction in Malaysian rivers, Saleh et al. (2017) and Sinnakaudan et al. (2006) reported that, for the tropical region, particularly for Malaysian rivers, the equations were less suitable because the hydraulic characteristics and sediment properties were different from the rivers investigated to for developing the existing equation. The same trend also applied to the neighbouring tropical country of Indonesia, where the reported discrepancy ratio (*DR*) was found to be below 28% (Gunawan et al. 2019). Inspired by the MNLR technique of predicting sediment transport in pipes by Ab Ghani (1993), Ariffin (2004) and Sinnakaudan et al. (2006) used the same approach to produce equations that suited the characteristics of rivers in Malaysia. Harun and Ab Ghani (2020) and Harun et al. (2020) later improved the MNLR equation by introducing a revised version of both Ariffin's (2004) and Sinnakaudan et al.'s (2006) equations, which are shown in Table 1.

The findings of Harun and Ab Ghani (2020) and Harun et al. (2020) suggested that there is a lack of accuracy in predicting the total material load when applying the conventional method (MNLR), particularly with higher data ranges, which results in low accuracy rates and low R^2 (coefficient of determination) and *MAE* (mean absolute error) values. The data from the three rivers adopted in this study were analysed by using the commonly used sediment transport equation; the results are shown in Table 2. The R^2 of all equations is less than 0.7 and the *MAEs* are in the range of 2.784–11.955, indicating that improvements are needed in order to increase model prediction accuracy.

Of late, the use of machine learning to predict sediment transport is gradually becoming the method of choice. Relevant literature (Shaghaghi et al. 2018a; Ebtehaj et al. 2019) revealed that machine learning techniques could produce better model prediction because they are more complex and can evolve to suit the model better, unlike the traditional regression method. Often, researchers used single and hybrid methods as an approach to improve the accuracy of the predictions (Yahaya 2019). Single methods, such as MNLR, artificial neural networks (ANN) and gene expression programming (GEP), were utilised to develop sediment transport models (Chang et al. 2012; Ab Ghani and Azamathulla 2014; Ara Rahman and Chakrabarty 2020), whereas hybrid models combined the methods to get the most appropriate model for the model predictions (Ab Ghani et al. 2010; Ab Ghani and Azamathulla 2014). According to Yahaya (2019), the performance of the hybrid methods is better than that of the single methods in most cases. In water resources engineering, the application of hybrid methods is widely used to predict stable channel dimensions, flow discharge, sediment transport modelling, scour depth and rainfall forecasting (Tayfur et al. 2003, 2013; Tayfur and Guldal 2006; Ulke et al. 2009; Nourani et al. 2012, 2016, 2019; Balouchi et al. 2015; Safari and Danandeh Mehr 2018; Shaghaghi et al. 2018b; Danandeh Mehr et al. 2019; Sharghi et al. 2019; Khosravi et al. 2020; Shiri et al. 2020). The studies conducted by Ara Rahman and Chakrabarty (2020), Sahraei et al. (2018) and Kitsikoudis et al. (2015) show that machine learning can successfully be applied in predicting sediment transport in rivers with high prediction accuracy. However, according to Rajaee and Jafari (2020), more precaution should be considered because machine learning is often influenced by extremely low- and high-value data. Among others, evolutionary polynomial regression (EPR), multi-gene genetic programming (MGGP) and the M5 tree model (M5P) are becoming the emerging machine learning tools used to develop model prediction. According to Bonakdari et al. (2020) and Ahmad Abdul Ghani et al. (2019), EPR is a robust prediction modelling method because the model can give high accuracy results with fewer errors.

The purpose of this study is to implement machine learning in sediment transport prediction modelling to enhance the existing total material load formulae accuracy. Revised equation parameters adopted by Harun et al. (2020) were used as the basis to generate the prediction model. Three machine learning algorithms—EPR, MGGP and M5P—were applied to investigate the effectiveness of the respective algorithms towards the total bed material load estimation.

Table 1 Total bed material load equations for Malaysian rivers

Reference	Equation
Ariffin (2004)	$C_v = 1.156 \times 10^{-5}(\frac{R}{d_{50}})^{0.716}(\frac{U^*}{\omega_s})^{-0.975}(\frac{U^*}{V})^{0.507}(\frac{V^2}{gy})^{0.524}$ (1)
Sinnakaudan et al. (2006)	$C_v = 1.811 \times 10^{-4}(\frac{VS_o}{\omega_s})^{0.293}(\frac{R}{d_{50}})^{1.390}(\frac{\sqrt{g(Ss-1)d_{50}3}}{VR})$ (2)
Harun et al. (2020)	$C_v = 4.032 \times 10^{-2}(\frac{U^*}{V})^{2.178}(\frac{V^2}{gy})^{0.795}$ (revised Ariffin 2004) (3)
	$C_v = 6.237 \times 10^{-3}(\frac{VS_o}{\omega_s})^{0.712}(\frac{R}{d_{50}})^{1.068}(\frac{\sqrt{g(S_s-1)d_{50}{}^3}}{VR})$ (revised Sinnakaudan et al. 2006) (4)

C_v is the sediment concentration by volume, S_s is the specific gravity of the sediment, R is the hydraulic radius, d_{50} is the median size of bed material, U^* is the shear velocity, S_o is the bed slope, w_s is the fall velocity of the bed material, V is the average flow velocity, g is the standard gravity, and y is the average depth of the water

Material and methods

Study area

This research uses data from the Malaysian Department of Irrigation and Drainage (DID 2009a). Three different rivers were investigated in this study, and they were separated by differences in length and hydraulic characteristics. Muda River, Langat River and Kurau River stretch about 180 km, 120 km and 92 km, respectively. Sediment samplings were carried out at a cross section for each site. According to Molinas and Wu (2001), rivers can be categorised by observing their flow depths. A wide river has a flow depth of more than 4 m, a medium river has a flow depth between 1.5 and 4 m and a small river has a flow depth of less than 1.5 m. Muda River represented the wide category river, followed by Langat River (medium river) and Kurau River (small river). Each river consisted of six sampling locations. Figure 1 depicts the location of the rivers within Peninsular Malaysia.

Muda River Basin

The Muda River Basin is located in the northern region of Peninsular Malaysia. The river originates in the hilly area in the district of Sik and closes at Thailand's border. It is the largest river in Kedah State and is essential in providing water to the three states of Kedah, Perlis and Pulau Pinang (Sim et al. 2015). The river flows from the northeast to the southwest before turning westward, forming a natural boundary with Pulau Pinang state before rushing to the sea. The drainage area is approximately 4210 km^2. The upper and middle reaches of the river are entirely located within the state of Kedah; meanwhile, its downstream stretch, with a length of about 30 km from the sea, is shared between Kedah and Pulau Pinang states. The length of the river is about 180 km long, with a slope of ½,300 (or 0.00043) from its estuary to the upper reaches. In terms of river width, it is typically around 10 m upstream, 100 m mid-stream and widest at its estuary, averaging 300 m (DID 2009b). The locations of the sampling points are shown in Fig. 2.

Table 2 Summary of performance of the revised equations and the current commonly used equations

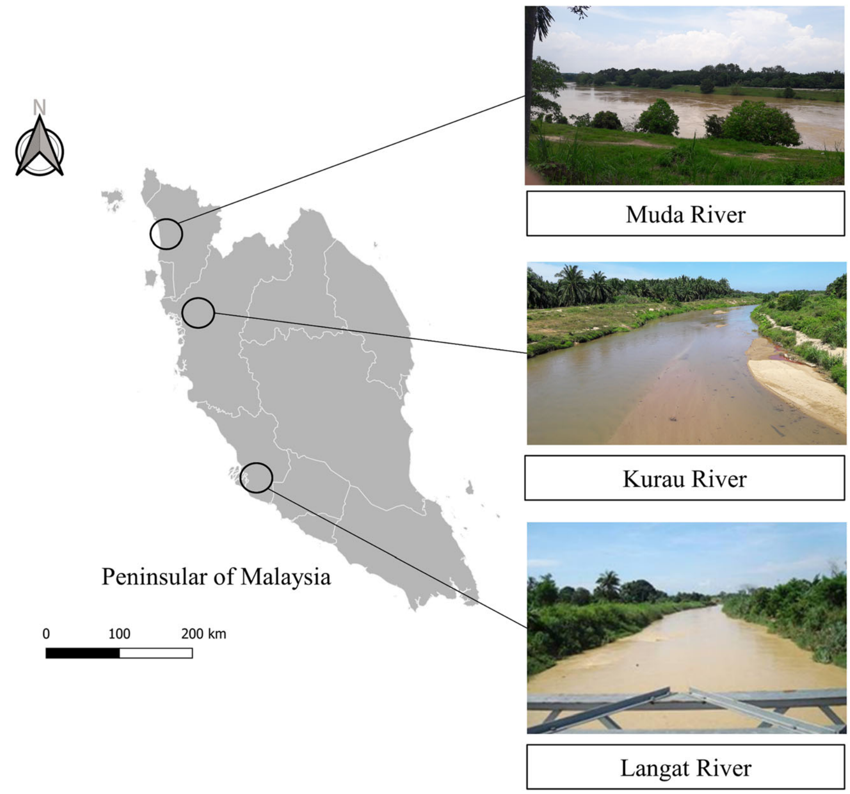
Equation	R^2	MAE
Revised Ariffin (2004)	0.616	2.854
Revised Sinnakaudan et al. (2006)	0.465	2.784
Ariffin (2004)	0.021	11.955
Sinnakaudan et al. (2006)	0.260	6.577
Engelund and Hansen (1967)	0.295	4.996
Yang (1979)	0.355	3.650

Langat River Basin

The Langat River Basin covers three different states—Selangor, Negeri Sembilan and the Wilayah Persekutuan. There are four main rivers in the Selangor state, and the Langat River makes up one of them. Langat River is a medium-sized river about 180 km long. The average annual flow is 35 m^3/s, and the mean annual flood is 300 m^3/s. The basin occupies the east of Titiwangsa Range and flows towards the sea (Straits of Malacca). A diverse topography was observed within the river basin, ranging from the hilly areas in the northeast, undulating in the middle hill and gentle in the southwest area (Fig. 3).

The river flows from the highland of Negeri Sembilan and then runs through Selangor and Wilayah Persekutuan before finally discharging into the Straits of Malacca. The Langat River Basin has a total catchment area of 2396 km^2. The basin in the Selangor state has an area totalling 1900 km^2. In contrast, the basin areas within the Federal Territories of Putrajaya and Kuala Lumpur are only 41 km^2 and 5 km^2, respectively. Negeri Sembilan state covers

Fig. 1 Study area location



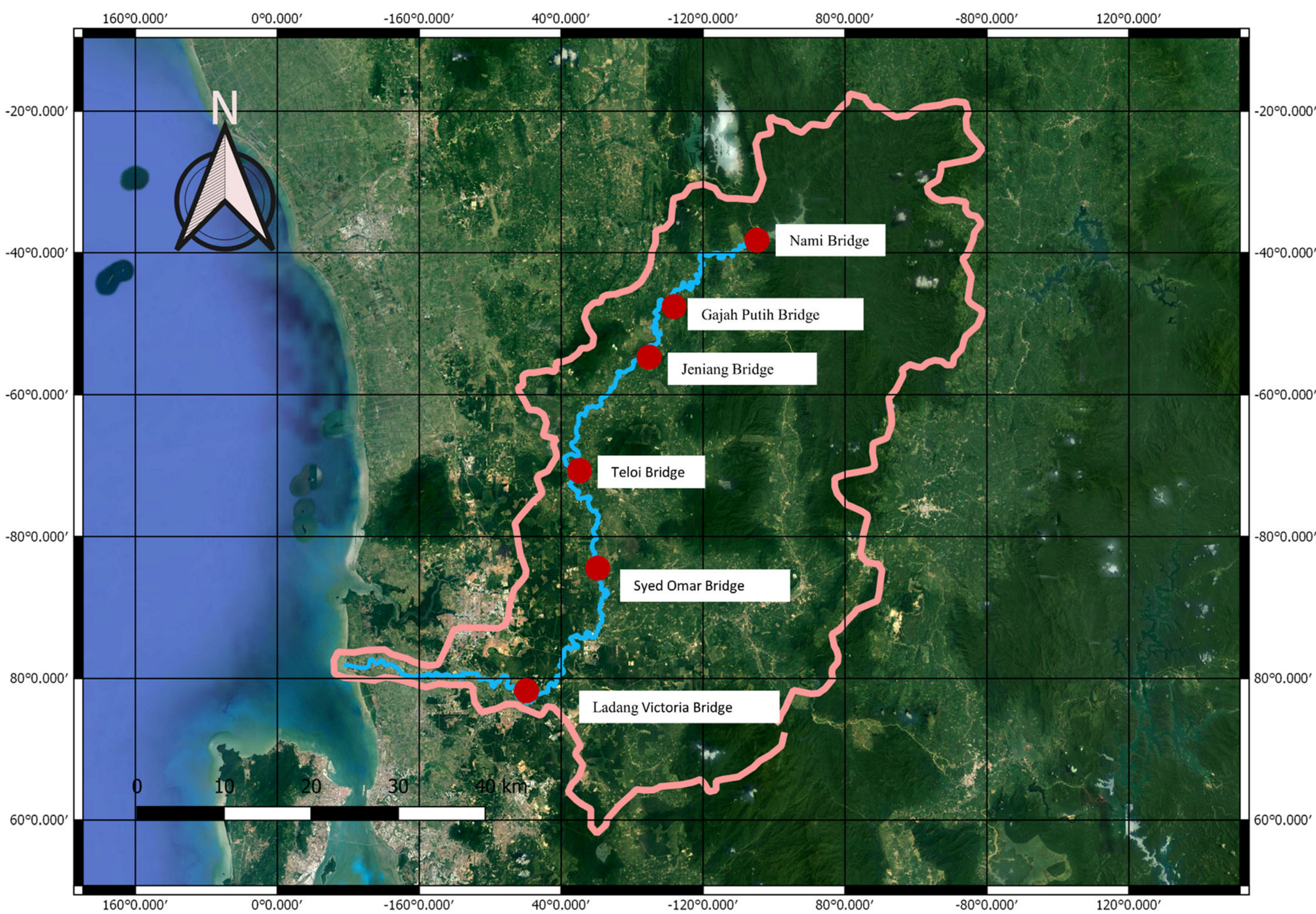


Fig. 2 Location of sampling points at the Muda River

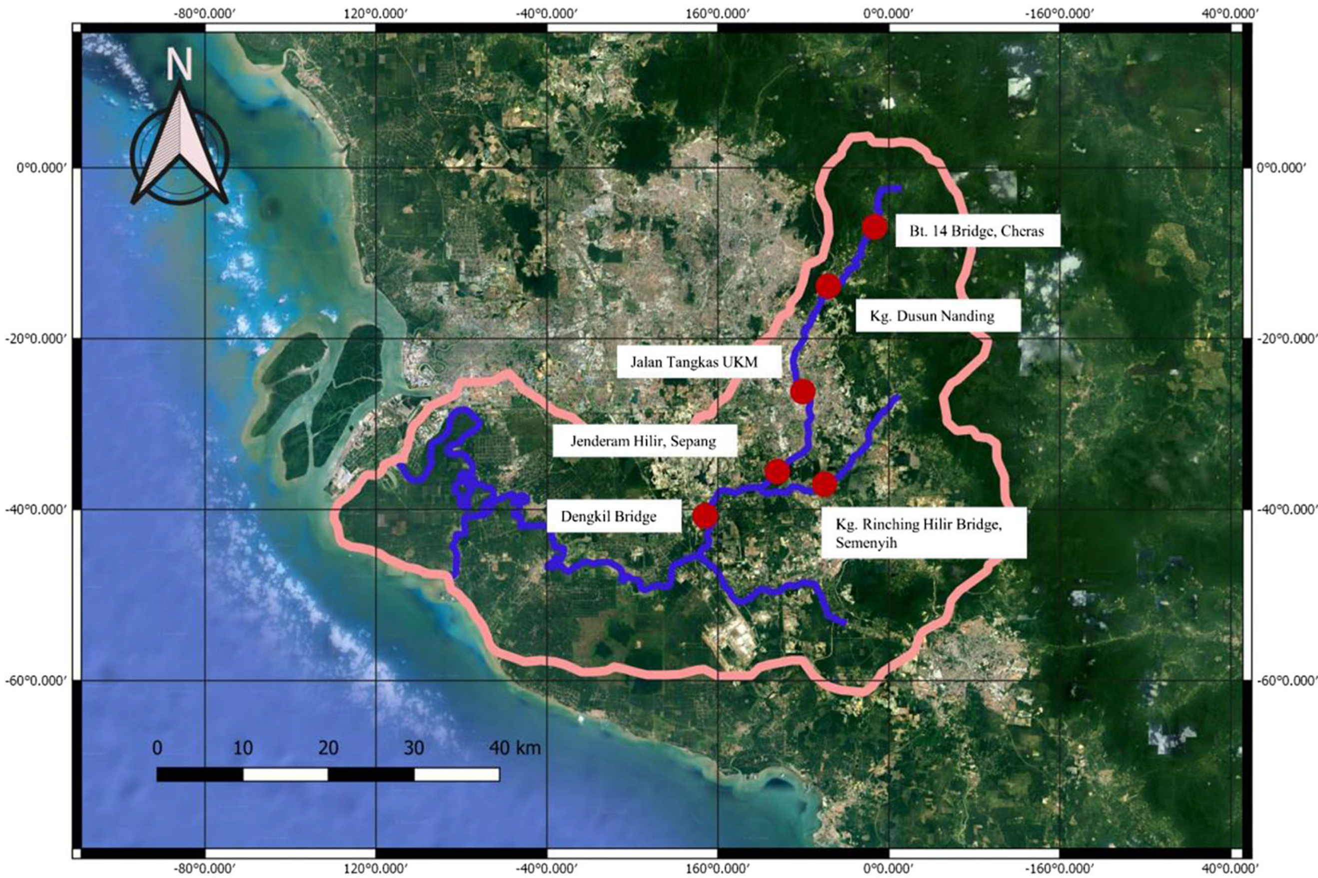


Fig. 3 Location of sampling points at the Langat River

the remaining 450 km^2 of the basin area (DID 2009c). Sampling locations are shown in Fig. 3.

Kurau River Basin

The basin sits on the floodplains of Perak state and closes at the sea. The basin can be considered as a small river basin with a drainage area of approximately 682 km^2 comprised of floodplains and swamps. It has a low-lying flat land characteristic. Bintang and Main Range make up the origin of the Kurau River, where the topography is found to be steep highland. Into the mid-stream, medium to undulating terrain was observed. As the river nears the sea, the topography changes quickly into flat and broad floodplains. Moderate elevation heights were observed at the river headwaters, ranging from 900–1200 m. In terms of slope, the upper reach and the lower reach range from 0.25–5%. The average velocity ranges between 0.45 and 0.636 m/s, and the highest sediment load recorded was 0.878 kg/s (Saleh et al. 2017). The Kurau River basin and the sampling points can be observed in Fig. 4.

A dam was constructed at the river mid-section (Bukit Merah reservoir) to serve as the primary irrigation source for paddy plantation. In the upstream of the reservoir, there are two major river systems that are drained into the reservoir—the Kurau River and the Merah River. Kurau River and Merah River land areas are occupied with tree crop agriculture, mainly palm oil farms. The river is located in the district of Larut, Matang and Selama (upper part) and flows toward the Kerian district (downstream part). The Kurau River basin is mostly rural, and many riverine villages were built along the river (DID 2009d).

This study was conducted using a much more comprehensive data range than the studies performed by Ariffin (2004) and Sinnakaudan et al. (2006). Tables 3 and 4 list the data range difference between the current study and past studies.

Field data collection

The collection of the data was done by referring to the guideline produced by Ab Ghani et al. (2003). The guideline consists of two different parts—field data collection and sediment analysis. Field data collection comprises flow measurement and river surveys. River surveys focused on measuring a river's cross section and bed elevation using electronic distance metre (EDM); meanwhile, data collection includes water surface slope, flow discharge, bed load and suspended load. The

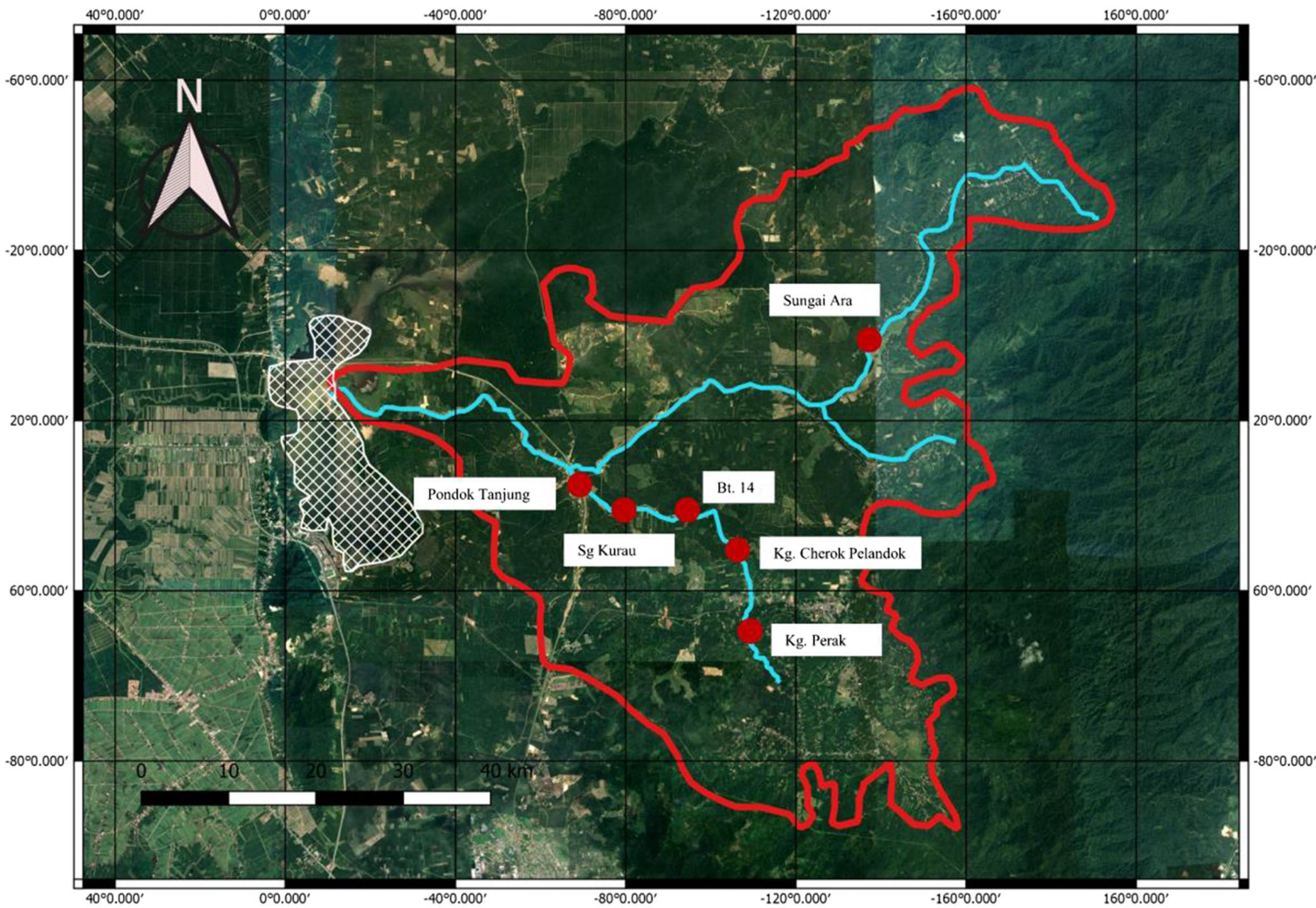


Fig. 4 Location of sampling points at the Kurau River

type of equipment used is dependent on the flow of the river. At the selected cross section, thalweg and bed level were measured to determine the stability of the rivers. The stability of a river can be observed by comparing the bed height and the river's thalweg with the suggested height proposed by the empirical and analytical method (Copeland 1994; Julien and Wargadalam 1995; Lee and Julien 2007; Jang et al. 2016; Harun et al. 2020). The wading technique was adopted for a low-flow river; meanwhile, for a high-flow river, the measurements were done by suspension from the bridge. An electromagnetic current metre was used in the wading method. For the high-flow river, the flow was measured by using Neyrflux Type 80 Universal current metre. Bed material was obtained by using a Van Veen sampler. Bed load was collected using the wading type Helley Smith and suspended type Helley Smith. The DH-48 sampler was used to manage the suspended

Table 3 Range of river data for the study conducted by Sinnakaudan et al. (2006) and Ariffin (2004)

Variable/parameter	Sinnakaudan et al. (2006)	Ariffin (2004)
Number of data	346	165
Discharge Q (m^3/s)	0.74–87.79	0.74–87.79
Average velocity V (m/s)	0.19–1.42	0.19–1.18
River width B	13.50–30.00	13.80–33.00
Flow depth Y_o	0.22–3.23	0.228–3.25
Area (m^2)	3.42–96.83	3.4–96.80
Hydraulic radius R (m)	0.22–2.66	0.22–2.66
Bed slope S_o	0.0004–0.0167	0.0004–0.0167
Sediment bed material, d_{50} (mm)	0.37–4.00	0.542–2.288
Total load (kg/s)	0.10–118.95	0.06–118.95

Table 4 Range of river data for the present study

Variable/parameter	The present study (Muda River)	The present study (Langat River)	The present study (Kurau River)
Number of data	76	60	78
Discharge Q (m^3/s)	2.59–343.71	2.75–120.76	0.63–28.94
Average velocity V (m/s)	0.14–1.45	0.23–1.01	0.27–1.12
River width B	9.0–90.00	16.4–37.60	6.30–26.00
Flow depth Y_o	0.73–6.90	0.64–5.77	0.36–1.91
Area (m^2)	6.12–278.34	8.17–153.57	1.43–33.45
Hydraulic radius R (m)	0.55–3.90	0.45–3.68	0.177–1.349
Bed slope S_o	0.00008–0.000235	0.00065–0.00185	0.00050–0.000210
Sediment bed material, d_{50} (mm)	0.29–2.10	0.31–3.00	0.41–1.90
Total load (kg/s)	0.099–15.644	0.525–99.398	0.089–2.970

load in the low-flow river, and a DH-59 sampler was used in the high-flow river. The final amount of total material load can be determined by summarising both the bed load and suspended load. Figure 5 showed the process of collecting the bed load and suspended load by using a suspended type Helley Smith and DH-48 sampler.

Revised total bed material load equations using MLR

According to Haddadchi et al. (2013) and Sulaiman et al. (2017b), total bed material load is the combination product of suspended loads and bed loads. The total bed material load (Q_t) was derived from the following relation:

Fig. 5 Data collection of bed load and suspended load by using suspension from the bridge

$$Q_t = CvQ\rho_s \tag{5}$$

where C_v is the sediment concentration by volume (dimensionless form), Q is the discharge and ρ_s is the density of the sediment. Sediment transport was influenced by the combination of significant parameter groups, namely mobility, sediment, conveyance and shape and flow resistance. Further explanation can be found in Harun et al. (2020). Since all the test cases were expressed in a single power-law equation, the possible regression analysis was analysed by applying statistical analysis software, SPSS. This study adopted $\ln C_v$ as the dependent variable; meanwhile, $\ln U^*/V$ and $\ln V^2/2gy$ were adopted as independent variables for Ariffin's (2004) equation. For the revised Sinnakaudan et al. (2006) equation, the dependent variable is *log phi* and the independent variables are $\log R/d_{50}$ and $\log VS_o/W_s$. The model was analysed further to find outliers using the standardised residual. This was later confirmed with influential outlier checking so that the outliers did not change the accuracy of the regression model dramatically.

EPR

EPR can be considered a data processing tool driven by the hybrid regression technique (Giustolisi and Savic 2006, 2009). This method uses a single genetic algorithm to concentrate on the formula symbol space to provide a few alternative models for prediction purposes (Giustolisi and Savic 2006, 2009). It is a non-linear stepwise regression that involves non-linear functions among variables but is linear to the regression parameters (Zahiri and Najafzadeh 2018).

EPR has a unique general structure that combines additive terms multiplied by many coefficients that can be described as follows:

$$\widehat{Y} = a_o + \sum_{j=1}^{m} a_j (X_1)^{ES(j,1)} \ldots\ldots (X_k)^{ES(j,k)} . f\left((X_1)^{ES(j,k+1)} \ldots\ldots (X_k)^{ES(j,2k)} \right) \quad (6)$$

where m can be defined as the maximum number of additive terms, X_1 and $\widehat{Y}$ are model input and output variables, function f is the exponents of the variables and ES can be chosen by the user beforehand (Giustolisi and Savic 2006, 2009). Ultimate regression expressions are linear to the coefficient a_j and often estimated using classical numeral regression (Giustolisi and Savic 2006, 2009).

MGGP

Originating from the GP, MGGP enhances the fitness of solutions by combining low depth GP to the monolithic GP (Safari and Danandeh Mehr 2018; Danandeh Mehr et al. 2019; Danandeh Mehr and Safari 2020). Danandeh Mehr et al. (2018) explained that the smaller tree application in MGGP is more straightforward compared to the monolithic GP. The summation of weighted outputs of two or more GP trees in a multi-gene programme produces the output variable; meanwhile, the bias depends on the stochastic term. The pseudo-linear MGGP model is represented by the output variable $\widehat{Y}$, which combines three genes. Each gene represents the function of a given input variable x_1 and x_2. Figure 6 shows an example of how MGGP operates. In this example, each multi-gene consists of three genes. Equation (5) describes the MGGP mathematical expression, where d_o is the bias term, d_1 and d_2 represent the gene weight and C_1 is the constant value.

$$\begin{aligned} \widehat{Y} &= d_o + d_1 (x_1 \times \cos x_2 + x_2 \times \sin x_1) \\ &\quad + d_2 (x_1 \times x_2 \sin x_1) + d_3 (C_1 \times x_1 + x_1 + x_2) \end{aligned} \quad (7)$$

Linear regression was applied in the MGGP to suit the non-linear condition of the physical system (Danandeh Mehr et al. 2018). Danandeh Mehr et al. (2018) also explained that any data pre-processing technique that can enhance the accuracy of the results could be used to optimise the gene weight.

M5P model tree

M5P is a linear tree-based model introduced by Quinlan (1992). An M5P decision tree is convenient because multivariate linear models can be operated within the model, and, indeed, it can be managed very flexibly (Balouchi et al. 2015; Khosravi et al. 2020). The main steps involved in developing M5P are constructing the tree, pruning the tree and smoothing the tree. In growing the trees, the best model was achieved by maximising the standard deviation reduction (SDR). SDR is explained in Eq. (5), where E is defined as the set of cases, E_i is ith subset of cases splitting the tree, SDE is the standard deviation of E and $SD(E_i)$ is the standard deviation of E_i

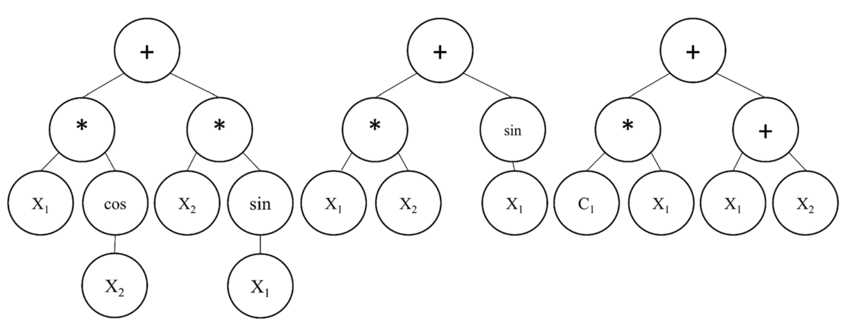
$$SDR = SDE - \sum_i \frac{|E_i|}{|E|} xSD(E_i) \quad (8)$$

The overfitting problem, in which the model is excellent in the dataset but does not perform well in the testing dataset, can be solved through the pruning step. In this step, subtrees were eliminated to maximise the results, and the attribute was reduced to minimise the error. Next, the smoothing step will continue to take place by adjusting the discontinuity at the leaves of the pruned tree (Khosravi et al. 2020). More details can be found by referring to the research done by Shaghaghi et al. (2018b) and Kargar et al. (2020).

The goodness of fit of model performance

Evaluation of the developed models was done based on several indices, which are coefficient of determination (R^2), Nash-Sutcliffe coefficient of Efficiency (NSE), root mean square error ($RMSE$), mean absolute error (MAE) and discrepancy ratio (DR). R^2 represents the correlation between measured and modelled values. Root mean square error ($RMSE$) represents the data unit squared for root mean error. Meanwhile, MAE shows the absolute error of the measured and modelled value. MAE makes use of absolute value to help reduce the bias towards the large event of prediction and

Fig. 6 Example of three genes of MGGP



observation data (Bennett et al. 2013). *NSE* is used to describe how much the modelling differs from the observed data. The *NSE* value of unity is the perfect result. Less than zero means underestimation of the model, and closer to the unity represents high accuracy of the predicted model (Bonakdari et al. 2020; Danandeh Mehr and Safari 2020). Discrepancy ratio (*DR*) is the comparison between the computed and measured total bed material load. The acceptable range of *DR* is 0.5–2.0 (Julien and Wargadalam 1995; Molinas and Wu 2001; Wu et al. 2008; Harun et al. 2020). Relationships for the computation of R^2, *NSE*, *MAE*, RMSE and *DR* can be written as follows

$$R2 = \left[\frac{\sum_{i=1}^{N} \left(O_i - \overline{O}_i\right)\left(P_i - \overline{P}_i\right)}{\sum_{i=1}^{N} \left(O_i - \overline{O}_i\right)^2 \sum_{i=1}^{N} \left(P_i - \overline{P}_i\right)^2 \Big)\Big)} \right]^2 \tag{9}$$

$$NSE = 1 - \frac{\sum_{i=1}^{N} \left(O_i - \overline{P}_i\right)^2}{\sum_{i=1}^{N} \left(O_i - \overline{O}_i\right)^2} \tag{10}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |O_i - P_i| \tag{11}$$

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^{N} (O_i - P_i)^2 \right]^{0.5} \tag{12}$$

$$DR = \frac{P_i}{O_i} \tag{13}$$

where O_i and P_i are observed and predicted values; meanwhile, O_i and P_i are the mean observed and predicted values, respectively.

Results and discussion

Revised total bed material load equation

Results from the MNLR, as discussed by Harun et al. (2020), showed that, for the revised version of Ariffin's (2004) equation, only two parameters were significant for sediment concentration computation. The values of R^2 and *MAE* turned out to be 0.616 and 2.526, respectively. Results from the regression indicate that the slope for variable $\frac{U^*}{V}$ is 2.178 and for the variable $\frac{V^2}{gy}$, the slope is 0.795. The intercept coefficient was found to be − 3.211. The revised version of Ariffin's (2004) equation can be re-written as follows:

$$C_v = 4.032 \times 10{-}2 \left(\frac{U^*}{V}\right) 2.178 \left(\frac{V^2}{gy}\right) 0.795 \tag{14}$$

Sinnakaudan et al.'s (2006) equation was revised by using the original parameters. The R^2 value is 0.482, and the *MSE* value is 2.784. Slope coefficient for $\frac{VS_o}{\omega_s}$ is 0.712 and for $\frac{R}{d_{50}}$, the slope coefficient is 1.068. This model intercepts at − 2.205. The yielded equation by implementing Sinnakaudan et al. (2006) parameter relationship can be described as Eq. (15).

$$C_v = 6.237 \times 10{-}3 \left(\frac{VS_o}{\omega_s}\right) 0.712 \left(\frac{R}{d_{50}}\right) 1.068 \left(\frac{\sqrt{g(Ss-1)d_{50}3}}{VR}\right) \tag{15}$$

The parameters for the inputs of EPR, MGGP and M5P machine learning models are selected based on the past research done by Harun et al. (2020). The parameters of the revised equation after applying MNLR were in the form of Q_t = f(Q, ρ_s, $\frac{U^*}{V}$, $\frac{V^2}{gy}$) (revised version of Ariffin's equation (2004)) and Q_t = f(Q, ρ_s, $\frac{VS_o}{\omega_s}$, $\frac{R}{d_{50}}$, $\frac{\sqrt{g(Ss-1)d_{50}3}}{VR}$) (revised version of Sinnakaudan et al.'s equation (2006)). Sensitivity analysis was done to test the significance of the parameters used. The parameters that are not significant to the prediction model are omitted. As a result, the parameters for the revised Sinnakaudan et al. (2006) equation are reduced to four parameters in the form of Q_t = f(Q, $\frac{VS_o}{\omega_s}$, $\frac{R}{d_{50}}$, $\frac{\sqrt{g(Ss-1)d_{50}3}}{VR}$), and parameters for the revised Ariffin (2004) equation are reduced to three parameters in the form of Q_t = f(Q, $\frac{U^*}{V}$, $\frac{V^2}{gy}$). It was observed that parameter ρ_s does not contribute a significant improvement to the prediction model, as the correlation coefficient (*R*) tends to be low. *MSE* tends to be higher in both revised equations. This study utilised 214 data in total. The dataset is split into two different parts: training and testing. The data used for training and testing were chosen by adopting the Kennard-Stone algorithm. The training process employs 70% of the data, and the testing process uses the remaining 30% of the data.

EPR

Equations (16) and (17) respectively are the yielded results for the revised Ariffin (2004) and revised Sinnakaudan et al. (2006) equations using EPR. The values of β_i, x_1, x_2, x_3 and x_4 are shown in Table 5 and Table 6. The equation is further analysed in the training and testing dataset.

$$Q_t = \sum_{i=1}^{13} P_i P_i = \beta_i \times Q^{x_1} \times \left(\frac{u^*}{V}\right)^{x_2} \times \left(\frac{V^2}{gy}\right)^{x_3} \tag{16}$$

Table 5 Value of β_i, x_1, x_2, x_3 and x_4 for revised Ariffin (2004)

	(β_i)	(x_1)	(x_2)	(x_3)
P1	1.03E-06	4	1	0
P2	2.95E+00	3	2	2
P3	4.49E-04	4	0	3
P4	− 1.39E-03	4	1	2
P5	− 8.01E+03	0	0	7
P6	4.56E+03	2	3	2
P7	5.06E+01	2	0	5
P8	− 3.82E-08	5	2	0
P9	− 5.79E-05	3	1	0
P10	− 4.10E+02	2	2	2
P11	1.39E-11	6	0	1
P12	− 2.30E-01	2	1	1
P13	7.54E+03	0	5	1

$$Q_t = \sum_{i=1}^{13} P_i P_i = \beta_i \times Q^{x_1} \times \left(\frac{R}{d_{50}}\right) x_2 \times \left(\frac{VS_0}{w_s}\right) x_3 \times \left(\frac{\sqrt{g(S_s-1){d_{50}}^3}}{VR}\right) x_4 \tag{17}$$

Figure 7 showed the performances of EPR models (training and testing) for both revised Ariffin (2004) and Sinnakaudan et al. (2006) parameters. For the revised Ariffin (2004) equation, the R^2 for training and testing are 0.949 and 0.892, respectively. Meanwhile, for *RMSE*, the training and testing are 2.564 and 4.596, respectively. As for the revised Sinnakaudan et al. (2006) equation, the R^2 for the training stage is 0.946, and for the testing stage, the value is 0.806.

Table 6 Value of β_i, x_1, x_2, x_3 and x_4 for revised Sinnakaudan et al. (2006)

	(β_i)	(x_1)	(x_2)	(x_3)
P1	− 0.004537924	4	0	2
P2	0.001396894	0	3	4
P3	1.80E-14	7	0	0
P4	5.33E+15	0	0	5
P5	807.2764724	4	0	2
P6	− 5.33E-10	6	0	1
P7	17428117874	1	1	0
P8	− 436934172.4	1	0	6
P9	− 3.32E+15	1	0	0
P10	2.02E-13	5	1	0
P11	− 7.34E-12	6	0	0
P12	− 51610.82939	3	0	1
P13	− 1.38064E+11	1	0	5

In terms of *RMSE*, the value is 2.912 (training) and 6.646 (testing).

MGGP

MGGP, on the other hand, depicts the following relationships for the revised Ariffin (2004) and Sinnakaudan et al.'s (2006) equations, respectively:

$$\begin{aligned} Q_t = {} & 25.8\frac{u^*}{V}\left(e^{e^{\frac{u^*}{V}}} - 0.869\frac{V^2}{gy}log\left(\frac{V^2}{gy}\right)\right) \\ & -200\frac{u^*}{V} - 4.2log(Q) - 6.15log\left(\frac{V^2}{gy}\right) - 0.787Q\frac{V^2}{gy} \\ & -0.135Q + 1311Q\left(\frac{u^*}{V}\right)^2\frac{V^2}{gy} - 1.96 \end{aligned} \tag{18}$$

$$\begin{aligned} Q_t = {} & 0.953\frac{\sqrt{g(S_s-1){d_{50}}^3}}{VR}\left(Q\left(Q+\frac{R}{d_{50}}\right) + Q^{\frac{\sqrt{g(S_s-1){d_{50}}^3}}{VR}}\right) \\ & -10.7Q\frac{VS_0}{w_s} - 0.0724Qlog\left(\frac{\sqrt{g(S_s-1){d_{50}}^3}}{VR}\right) - 0.00157Q^2 \\ & + +1.16Q^2 log\left(\left(\frac{\sqrt{g(S_s-1){d_{50}}^3}}{VR}\right)^{\frac{\sqrt{g(S_s-1){d_{50}}^3}}{VR}}\right) \\ & -2000Q^2\frac{VS_0}{w_s}\frac{\sqrt{g(S_s-1){d_{50}}^3}}{VR}log\left(\frac{\sqrt{g(S_s-1){d_{50}}^3}}{VR}\right) \end{aligned} \tag{19}$$

Results from the modelling by using MGGP show the moderate R^2 value for both revised equations. Revised Ariffin (2004) R^2 value for training is 0.796, and for the testing stage, the value is 0.781. *RMSE* for both training and testing stages are found as 10.578 and 12.727, respectively. The R^2 value for the revised Sinnakaudan et al. (2006) equation is slightly higher compared to the revised Ariffin (2004) equation's, which is 0.815 for training and 0.740 for testing stages. However, the *RMSE* is observed to be slightly higher in the revised Sinnakaudan et al. (2006), whereas the values for testing and training are found as 10.689 and 12.383. More details can be found in Fig. 8.

M5P

Summaries for the M5P regression tree for the revised Ariffin (2004) equation and the revised Sinnakaudan et al. (2006) are shown in Fig. 9 and Fig. 10. M5P gives mixed predictions for both the revised equations. Figure 11 gives an outlook on the predicted and observed values of both revised equations. The R^2 values for training were observed to be higher compared to the values of the testing stage. The R^2 values are 0.939

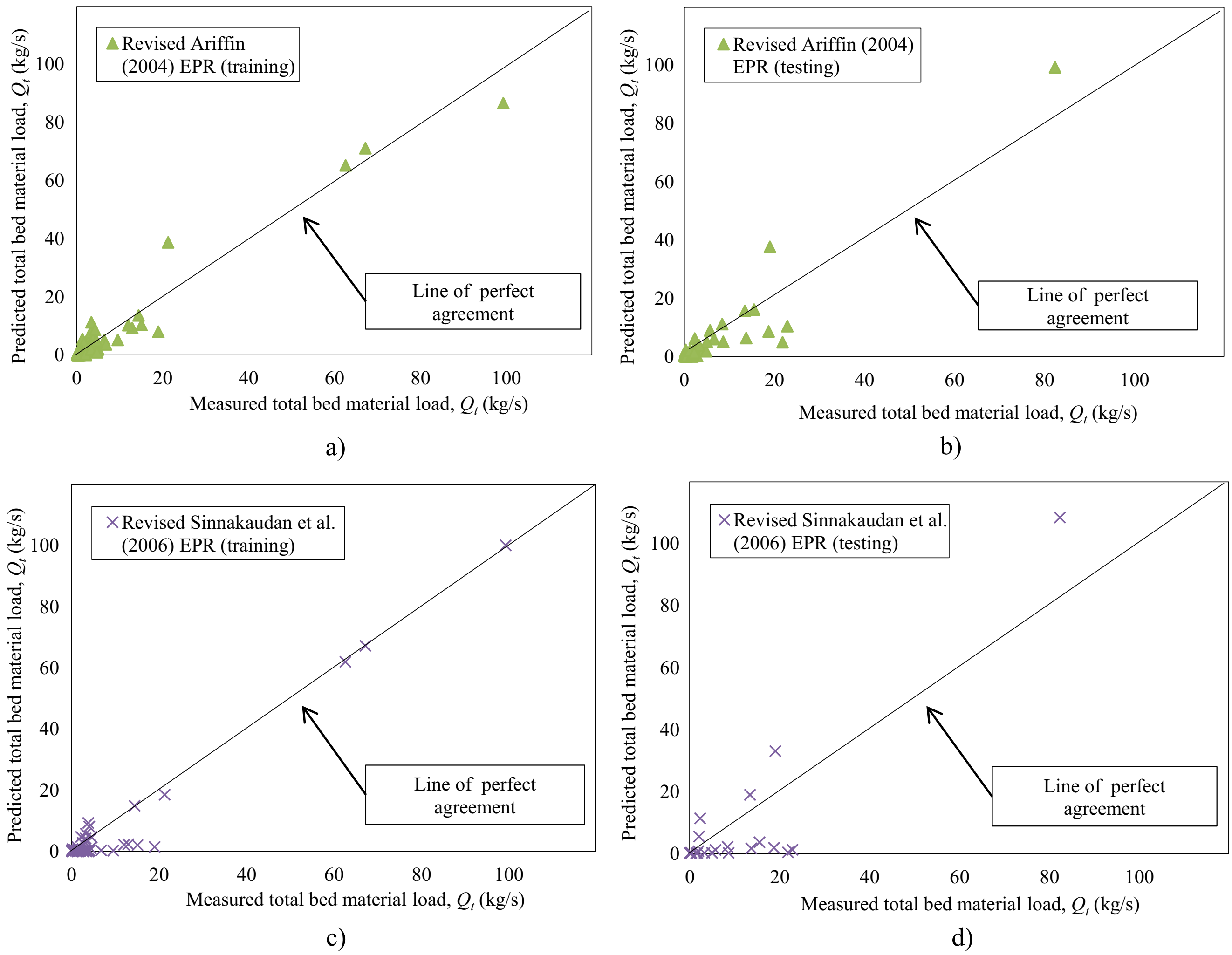


Fig. 7 Training and testing results of the revised Ariffin (2004) (**a**, **b**) and Sinnakaudan et al. (2006) (**c**, **d**) by using EPR

(training) and 0.553 (testing) for the revised Ariffin (2004) equation. *RMSE* values for training and testing stages were 11.388 and 14.108, respectively. The revised Sinnakaudan et al. (2006) equation, in turn, produced R^2 values of 0.718 (training) and 0.443 (testing). Compared to the revised Ariffin (2004) equation, *RMSE* for the revised Sinnakaudan et al. (2006) equation is 12.383 for training and 11.405 for testing.

Prediction modelling summary

The two revised equations using EPR were compared with the existing revised equations, as well as with the revised models that used MGGP and M5P machine learning algorithms. The results were also compared with existing results regarding non-tropical rivers introduced by Ackers and White (1973) and Karim (1998) respectively given as follows:

$$C_s = 10^6 c \frac{\rho_s d_{50}}{\rho \ R}\left(\frac{V}{U*}\right)^n \left(\frac{F_{gr}}{A_{aw}} - 1\right)^m \tag{20}$$

$$q_t = \Phi_t\left((S_s - 1) g d_{50}^3\right)^{0.5} \tag{21}$$

where C_s is defined as sediment concentration by weight, ρ_s is the soil mass density, ρ is the water mass density, F_{gr} is mobility numbers, q_t is the total load per unit time and width, ϕ_t is the total load transport intensity, c and A_{aw} are the coefficients and n and m are the exponents depending on the dimensionless grain size D_{gr} defined as

$$D_{gr} = d_{50}\left(\frac{(S_s - 1) g}{v^2}\right)^{1/3} \tag{22}$$

where   is the fluid kinematic viscosity. Table 7 listed the coefficient and exponents for the Ackers and White (1973) equation.

F_{gr} and $U'*$ is calculated by the following relation:

$$F_{gr} = \frac{U^{*n} U'^{*1-n}}{\sqrt{(S_s - 1) g d_{50}}} \tag{23}$$

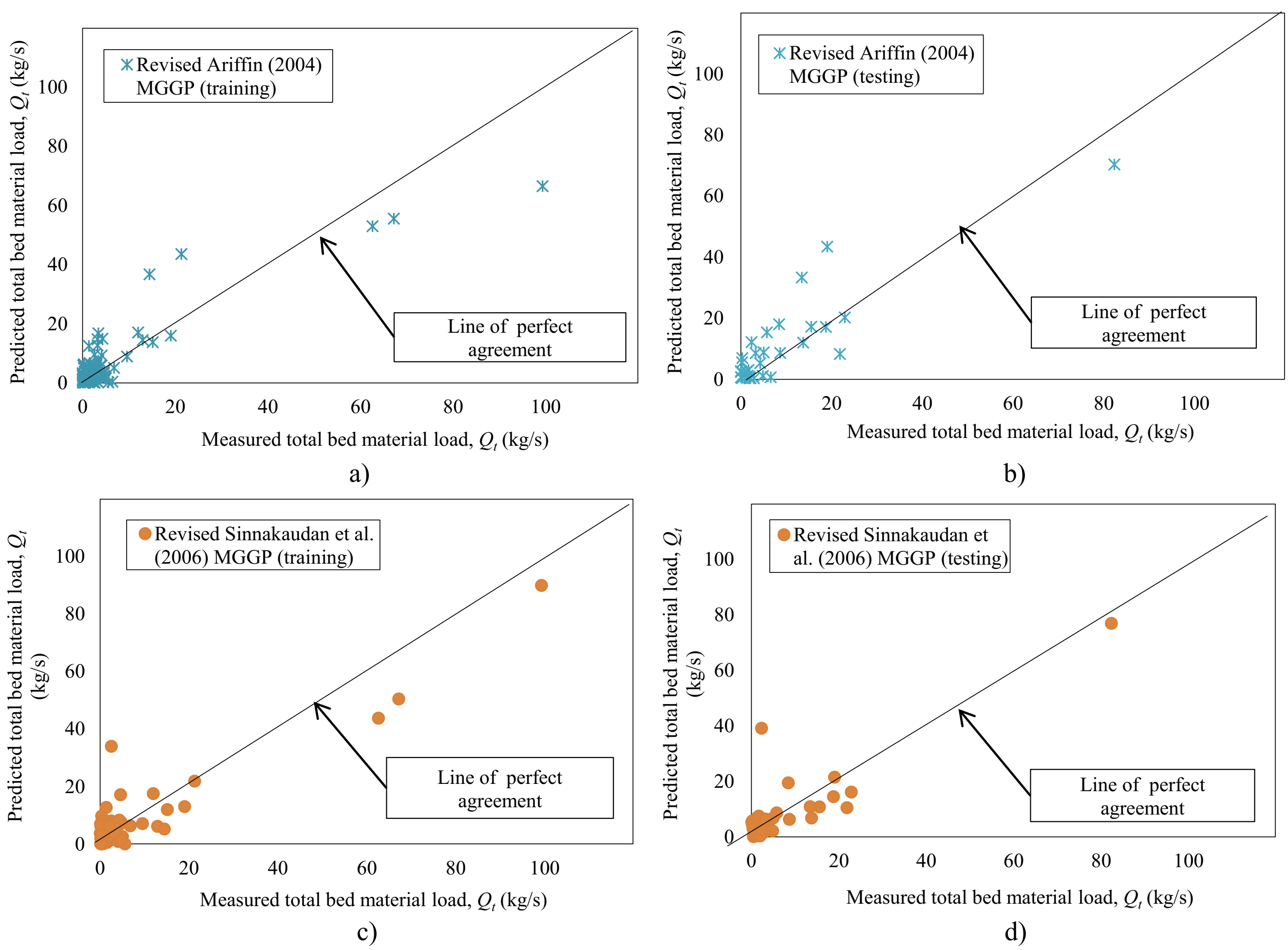


Fig. 8 Training and testing results of the revised Ariffin (2004) (**a**, **b**) and Sinnakaudan et al. (2006) (**c**, **d**) by using MGGP

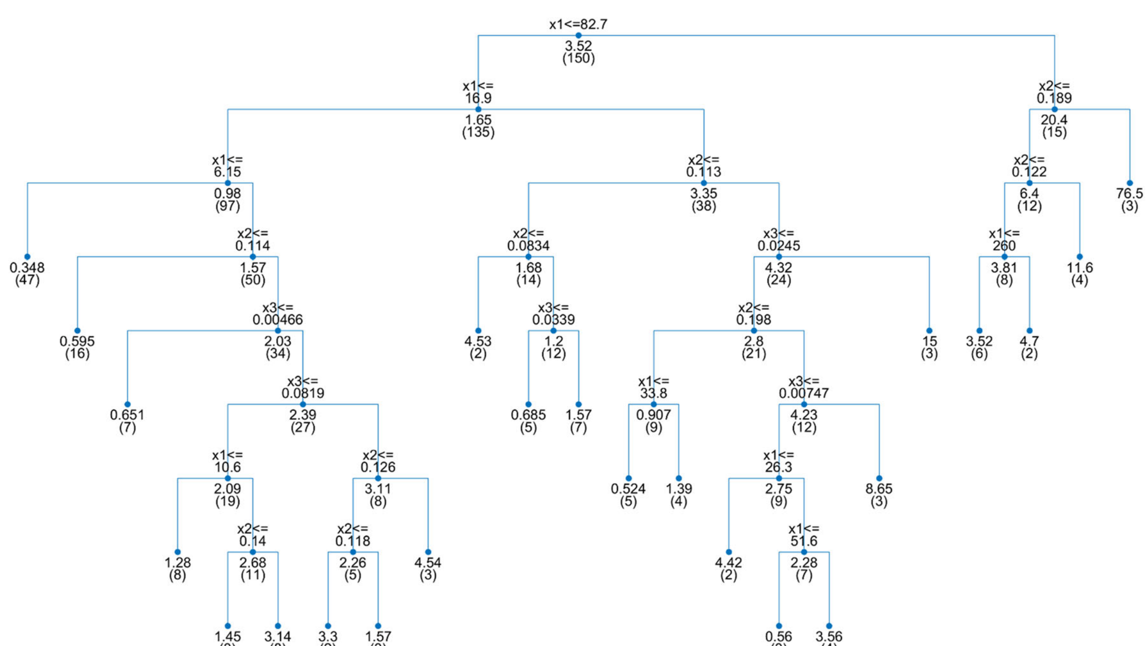


Fig. 9 M5P regression tree for revised Ariffin (2004)

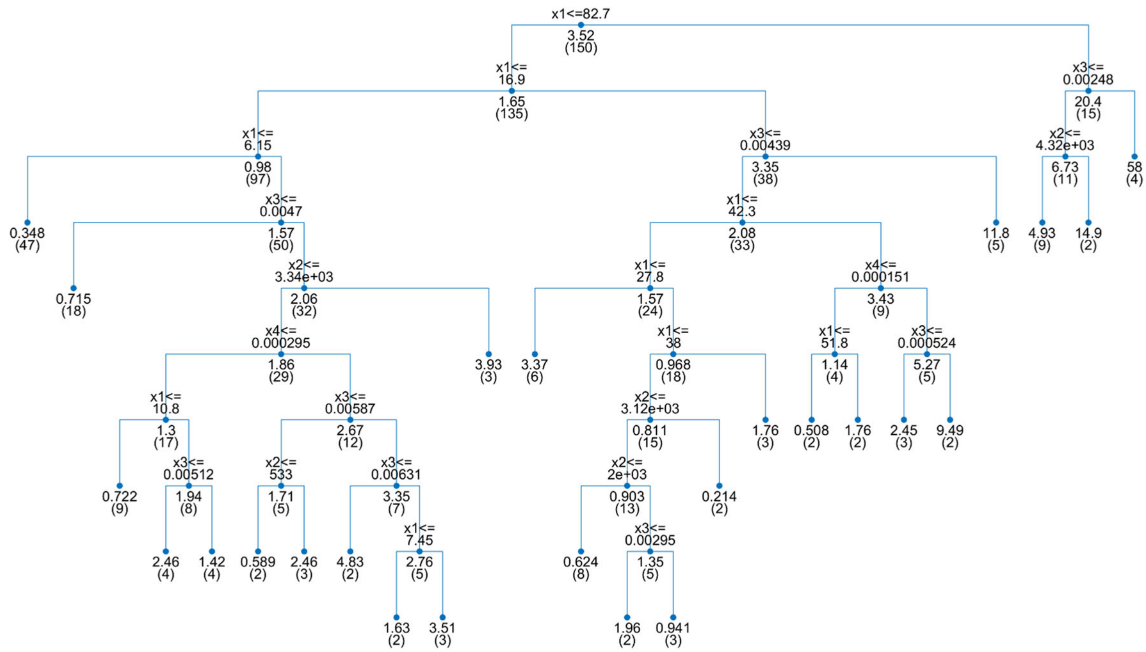


Fig. 10 M5P regression tree for revised Sinnakaudan et al. (2006)

$$U'* = \frac{V}{\sqrt{32\log\left(10\ \frac{R}{d_{50}}\right)}} \tag{24}$$

Meanwhile, for Karim (1998) equation, ϕ_t and F_d can be expressed as follows:

$$\phi_t = 1.39 \text{ x } 10^{-3} F_d{}^{2.97}\left(\frac{U^*}{\omega_s}\right)^{1.47} \tag{25}$$

$$F_d = \left(\frac{V}{\sqrt{\left(\left(S_s - 1\right)gd_{50}\right.}}\right) \tag{26}$$

All the machine learning programmes use the same parameters, as discussed in the "Revised total bed material load equation" section. The overall machine learning performance is summarised in Table 8. Indeed, all machine learning models are able to increase prediction accuracy with low error in comparison to the existing revised equations. The revised model using EPR was found to produce better prediction results in contrast to the MGGP and M5P models. The revised Ariffin (2004) EPR model has the highest R^2 and *NSE* values, which are 0.922 and 0.913, respectively, followed by the revised Sinnakaudan et al. (2006) EPR (R^2 = 0.884, *NSE* = 0.848), revised Ariffin (2004) MGGP (R^2 = 0.787, *NSE* = 0.784), revised Sinnakaudan et al. (2006) MGGP (R^2 = 0.787, *NSE* = 0.784), revised Ariffin (2004) M5P (R^2 = 0.786, *NSE* = 0.762), revised Sinnakaudan et al. (2006) M5P (R^2 = 0.622, *NSE* = 0.615), Karim (1998) (R^2 = 0.051, *NSE* = − 0.133) and Ackers and White (1973) (R^2 = 0.003, *NSE* = − 1.100). Among all the revised models, the revised Ariffin (2004) EPR model has the lowest *RMSE* (3.305) and *MAE* (1.552). Interestingly, Ackers and White's (1973) equation has the highest *RMSE* (16.254) and *MAE* (4.923). All machine learning seems to be able to increase the accuracy of the model. However, in terms of *DR*, only the revised Ariffin (2004) M5P and the revised Sinnakaudan et al. (2006) M5P give better *DR* prediction results than the revised MNLR results. Figures 12, 13, 14 explain the results in terms of *DR* for each respective machine learning programme. From Table 7, the revised Sinnakaudan et al. (2006) M5P turned out to have the highest *DR* of 73.36%, followed by the revised Ariffin (2004) M5P with 72.43%, revised Ariffin (2004) with 66.36%, revised Sinnakaudan et al. (2006) with 64.49%, revised Ariffin (2004) EPR with 34.58%, revised Sinnakaudan et al. (2006) MGGP with 31.31%, revised Ariffin (2004) MGGP with 21.03% and revised Sinnakaudan et al. (2006) EPR with 14.49%. It is also important to note that, even though the *DR* for M5P is considerably good (exceeding 73%), the data did not distribute well and is rather flattening at the lower total bed material load rate.

The results from the non-tropical equations from Ackers and White (1973) and Karim (1998), on the other hand, suggested that the equation is not suitable to be used in the tropical region. Although Ackers and White (1973) use a much more comprehensive range of sediment bed material (0.04–4.00 mm), the prediction accuracy is low and the equation only manages to achieve R^2 and *NSE* values of 0.003 and − 1.100, respectively.

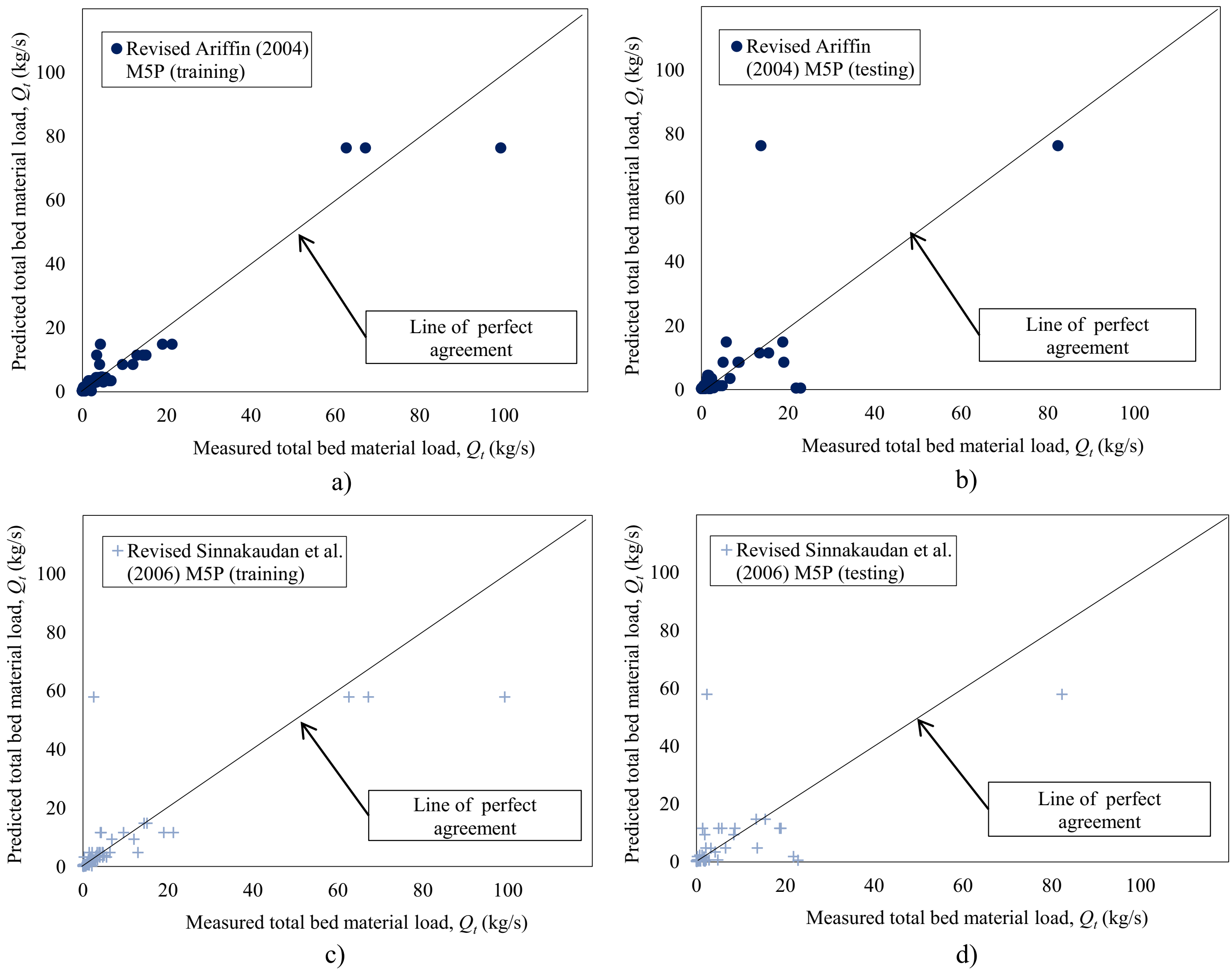


Fig. 11 Training and testing results of the revised Ariffin (2004) (**a**) and Sinnakaudan et al. (2006) (**b**) by using M5P

From the summary, EPR was found to improve the prediction distribution value the most by producing higher R^2 and *NSE* values and lower *RMSE* and *MAE* values, followed by MGGP and M5P. EPR is able to predict better results in both revised Ariffin (2004) and Sinnakaudan et al. (2006) equations, resulting in a better prediction model compared to those produced by MGGP and M5P. More importantly, despite the lack of accuracy in model prediction in terms of R^2 and *NSE* values using the M5P programme, in terms of the *DR*, M5P shows better prediction accuracy and gives better prediction results compared to the revised equations. Rajaee and Jafari (2020) suggest that machine learning is very sensitive. This research shows that machine learning is better at predicting total bed material load at a high value than at a lower value.

Table 7 Coefficient and exponents for Ackers and White (1973)

Parameter	$D_{gr} > 60$	$60 \geq D_{gr} > 1$
c	0.025	Log (c) = 2.86 log (D_{gr}) − [log (D_{gr})]^2 − 3.53
A_{aw}	0.17	$0.23/D_{gr}^{0.5} + 0.14$
n	0	1–0.56 log (D_{gr})
m	1.5	$9.66/D_{gr} + 1.34$

Sediment rating curve

The sediment rating curve is significant in giving general information about the relation between a river's flow rate and sediment yield. When the data is limited, the sediment rating curve can be a useful tool in predicting a river's sediment yield. The sediment rating curve can also be derived from the expected sediment transport prediction (Asselman 2000; Mohammadi et al. 2021). The predicted equation's data fitness can be measured by plotting the derived sediment prediction results to the present sediment rating curve. Figure 15 shows the derived sediment rating curve using the revised equation by using MNLR and machine learning programmes.

Table 8 Summary of performance of the models

Model	R^2	*NSE*	*RMSE*	*MAE*	*DR* (0.5–2.0) %
Revised Ariffin (2004)	0.616	0.228	9.462	2.526	66.36
Revised Sinnakaudan et al. (2006)	0.482	0.221	9.902	2.784	64.49
Revised Ariffin (2004) EPR	0.922	0.913	3.305	1.552	34.58
Revised Sinnakaudan et al. (2006) EPR	0.884	0.848	4.377	2.137	14.49
Revised Ariffin (2004) MGGP	0.787	0.784	5.217	3.054	21.03
Revised Sinnakaudan et al. (2006) MGGP	0.787	0.784	5.207	3.011	31.31
Revised Ariffin (2004) M5P	0.786	0.762	5.467	1.561	72.43
Revised Sinnakaudan et al. (2006) M5P	0.622	0.615	6.961	1.994	73.36
Ackers and White (1973)	0.003	− 1.100	16.254	4.923	21.03
Karim (1998)	0.051	− 0.133	11.938	3.823	38.32

From Fig. 15, the revised Ariffin (2004) and Sinnakaudan et al. (2006) equations using MNLR show better prediction results compared to the revised equations using machine learning. The revised equations, particularly the revised Ariffin (2004) equation, show smaller differences compared to the data from DID (2009a). As for the machine learning programme, the low prediction accuracy of sediment yield was observed at the low river discharge. The results are aligned with the findings earlier, whereby the machine learning algorithm is better at predicting higher rates of total bed material load.

Limitation of the proposed model

The current study focuses on developing a new prediction model for sediment transport with a median bed material between 0.29 and 3.00 mm. This study is limited by its number of samples (214 river data) and the river's lack of data with higher river discharge and sediment of 343.71 m^3/s and 15.64 kg/s. The application of machine learning in this study only focuses on EPR, MGGP and M5P. For better total material load model prediction, different machine learning algorithms can be further explored to increase the model prediction efficiency, especially for lower volume river discharge.

Conclusions

This study emphasises the great potential of machine learning in increasing sediment transport prediction accuracy, particularly for rivers in the tropical region. The findings suggest that machine learning can enhance the model prediction distribution data more than the conventional method, MNLR, but is lacking in terms of *DR*. Three types of machine learning algorithms were investigated in this study: EPR, MGGP and M5P. As a representation of the tropical region, 214 river data from three different Malaysian rivers were used in this study. Overall, compared to equations using MGGP and M5P, the revised equations using EPR gave better predictions of the total bed material load in terms of data distribution. EPR is able to improve the data prediction distribution of the revised Ariffin (2014) and revised Sinnakaudan et al. (2006) models, followed by

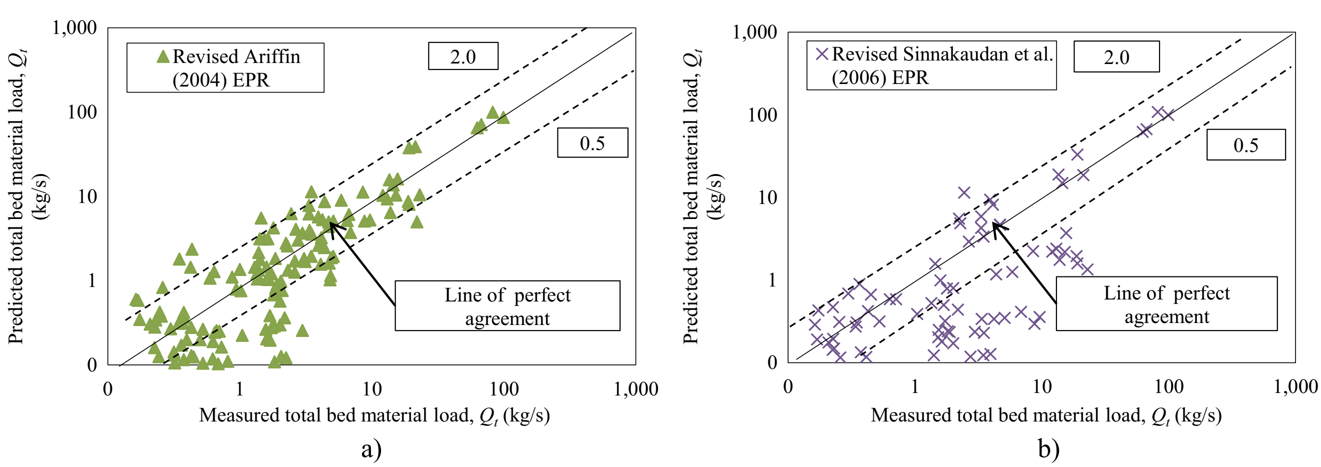


Fig. 12 Comparison results between measured and predicted total bed material load for the revised Ariffin (2004) (**a**) and Sinnakaudan et al. (2006) (**b**) by using EPR

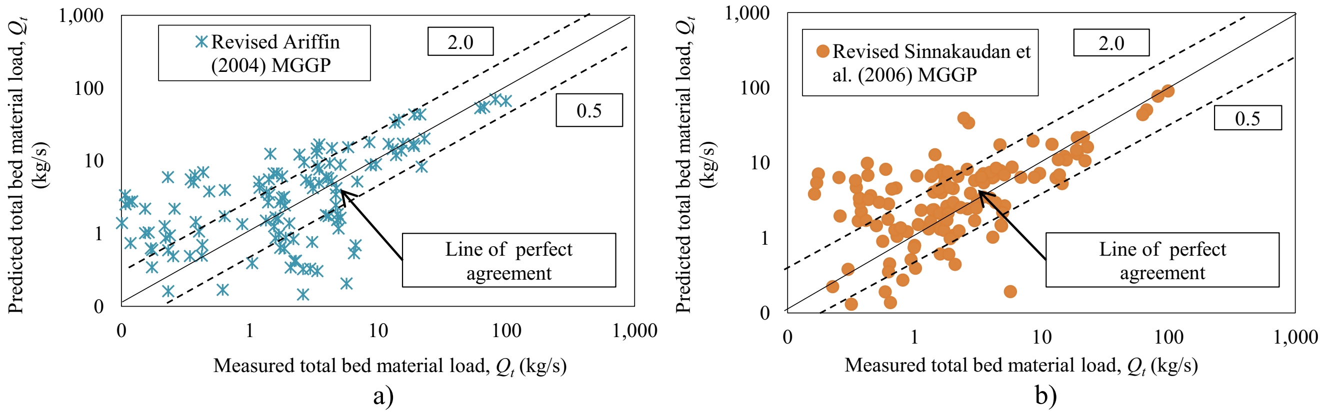


Fig. 13 Comparison results between measured and predicted total bed material load for the revised Ariffin (2004) (**a**) and Sinnakaudan et al. (2006) (**b**) by using MGGP

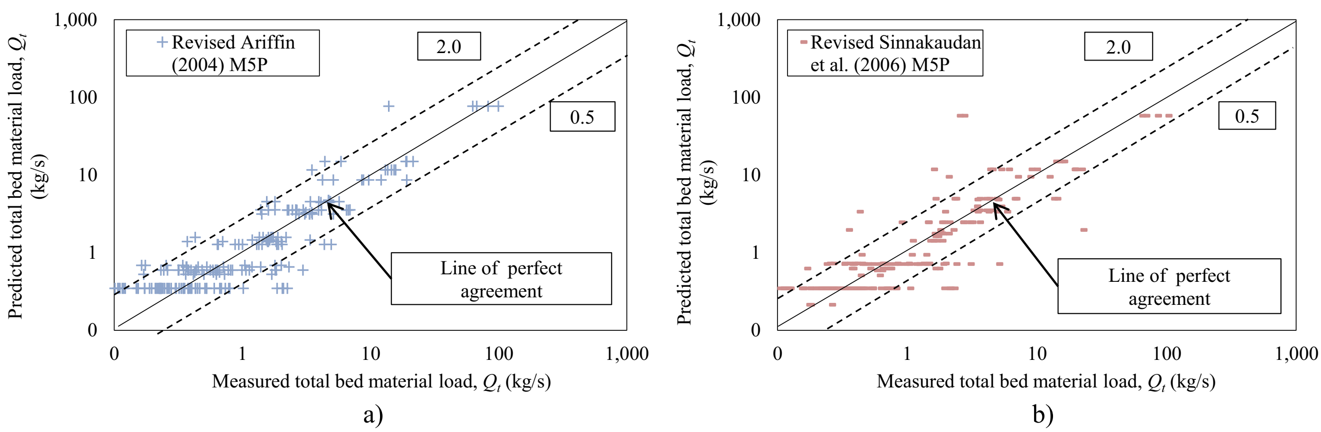
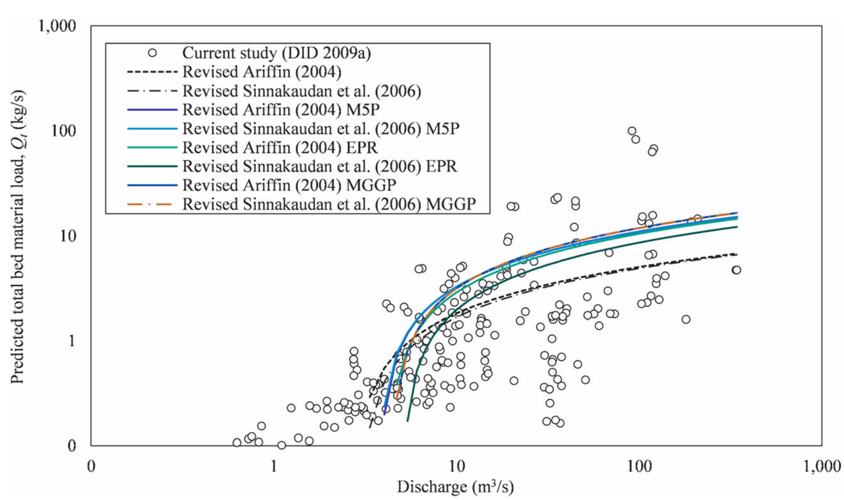


Fig. 14 Comparison results between measured and predicted total bed material load for the revised Ariffin (2004) (**a**) and Sinnakaudan et al. (2006) (**b**) by using M5P

Fig. 15 Derived sediment rating curves using the previous revised Ariffin (2004) and Sinnakaudan et al. (2006) equations and those derived from this present study



MGGP and M5P. The results showed that, among all the model predictions, the new revised Ariffin (2004) EPR model produced the lowest amount of errors ($RMSE$ = 3.305, MAE = 1.552) and had excellent prediction accuracy (R^2 = 0.922, NSE = 0.913). However, the improvement is found to be limited, particularly at lower river discharge. Machine learning was observed to be affected by the range of data and preferred to focus more on high prediction data. The improvement is less significant compared to the proposed revised equations reported in the literature. The DR of the EPR and MGGP revised equations is low compared to the proposed revised equations using MNLR. Even though M5P can give a better DR prediction ratio, the data is not well distributed at lower river discharges. The current study was limited by the river's hydraulic and sediment characteristics. Median sediment bed material (d_{50} (mm)) and streamflow range are within 0.29–3.00 mm and 0.63–343 m^3/s, respectively. Further research should be conducted to investigate a broader range of data with a different river profile to improve model prediction accuracy, particularly for low values of total bed material load.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11356-021-14479-0.

Acknowledgements The authors would like to express special thanks for the support provided by REDAC, USM. Acknowledgement also goes to the Public Service Department of Malaysia for the scholarship provided to the first author under the *Hadiah Latihan Persekutuan* (HLP) programme.

Author contributions The author contributions are listed as follows: conceptualisation: Mohd Afiq Harun, Aminuddin Ab Ghani; data curation: Mohd Afiq Harun, Aminuddin Ab Ghani; formal analysis: Mohd Afiq Harun, Mir Jafar Sadegh Safari; investigation: Mohd Afiq Harun, Enes Gul; methodology: Mohd Afiq Harun, Enes Gul; resources: Mohd Afiq Harun, Aminuddin Ab Ghani; software: Enes Gul; supervision: Mir Jafar Sadegh Safari, Aminuddin Ab Ghani; validation: Mohd Afiq Harun, Mir Jafar Sadegh Safari, Aminuddin Ab Ghani; visualisation: Mohd Afiq Harun, Mir Jafar Sadegh Safari, Aminuddin Ab Ghani; writing—original draft: Mohd Afiq Harun; writing—review and editing; Mir Jafar Sadegh Safari, Aminuddin Ab Ghani.

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

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

Ab Ghani A (1993) Sediment transport in sewers. Ph.D. thesis. University of Newcastle upon Tyne, UK.

Ab Ghani A, Azamathulla HM (2014) Development of GEP-based functional relationship for sediment transport in tropical rivers. Neural Comput Appl 24:271–276. https://doi.org/10.1007/s00521-012-1222-9

Ab Ghani A, Zakaria NA, Abdullah R et al (2003) Guidelines for field data collection and analysis of river sediment. Department of Drainage and Irrigation Malaysia, Kuala Lumpur

Ab Ghani A, Azamathulla HM, Chang CK et al (2010) Prediction of total bed material load for rivers in Malaysia: a case study of Langat, Muda and Kurau Rivers. Environ Fluid Mech 11:307–318. https://doi.org/10.1007/s10652-010-9177-9

Ackers P, White WR (1973) Sediment transport: new approach and analysis. J Hydraul Eng ASCE 99:2041–2060

Ahmad Abdul Ghani NA, Tholibon DA, Ariffin J (2019) Robustness analysis of model parameters for sediment transport equation development. ASM Sci J 12:1–17. https://doi.org/10.32802/asmscj.2019.268

Ara Rahman S, Chakrabarty D (2020) Sediment transport modelling in an alluvial river with artificial neural network. J Hydrol 588:125056. https://doi.org/10.1016/j.jhydrol.2020.125056

Ariffin J (2004) Development of sediment transport models for rivers in Malaysia using regression analysis and artificial neural network. Ph.D thesis. Universiti Sains Malaysia, Penang

Asselman NEM (2000) Fitting and interpretation of sediment rating curves. J Hydrol 234:228–248. https://doi.org/10.1016/S0022-1694(00)00253-5

Bagnold RA (1996) An approach to the sediment transport problem from general physics. U.S. Geological Survey Professional Paper No. 422-J

Balouchi B, Reza M, Adamowski J (2015) Development of expert systems for the prediction of scour depth under live-bed conditions at river confluences: application of different types of ANNs and the M5P model tree. Appl Soft Comput J 34:51–59. https://doi.org/10.1016/j.asoc.2015.04.040

Bennett ND, Croke BFW, Guariso G, Guillaume JHA, Hamilton SH, Jakeman AJ, Marsili-Libelli S, Newham LTH, Norton JP, Perrin C, Pierce SA, Robson B, Seppelt R, Voinov AA, Fath BD, Andreassian V (2013) Characterising performance of environmental models. Environ Model Softw 40:1–20. https://doi.org/10.1016/j.envsoft.2012.09.011

Bonakdari H, Gholami A, Sattar AMA, Gharabaghi B (2020) Development of robust evolutionary polynomial regression network in the estimation of stable alluvial channel dimensions. Geomorphology 350:106895. https://doi.org/10.1016/j.geomorph.2019.106895

Brownlie WR (1981) Prediction of flow depth and sediment discharge in open channels. Report No. KH-R-43A, California Institute of Technology, Pasadena, Calif

Chang HH (1985) Design of stable alluvial canals in a system. J Irrig Drain Eng 111:36–43. https://doi.org/10.1061/(ASCE)0733-9437(1985)111:1(36)

Chang CK, Abdullah R, Ghani A et al (2005) Sediment transport equation assessment for selected rivers in Malaysia. Int J River Basin Manag 3:203–208. https://doi.org/10.1080/15715124.2005.9635259

Chang CK, Azamathulla HM, Zakaria NA, Ghani AA (2012) Appraisal of soft computing techniques in prediction of total bed material load in tropical rivers. J Earth Syst Sci 121:125–133. https://doi.org/10.1007/s12040-012-0138-1

Copeland RR (1994) Application of Channel Stability Methods Case Studies Technical Report No. HL-94-11. Waterways Experiment Station, Vicksburg

Danandeh Mehr A, Safari MJS (2020) Application of soft computing techniques for particle Froude number estimation in sewer pipes. J Pipeline Syst Eng Pract 11:1–8. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000449

Danandeh Mehr A, Nourani V, Kahya E, Hrnjica B, Sattar AMA, Yaseen ZM (2018) Genetic programming in water resources engineering : a state-of-the-art review. J Hydrol 566:643–667. https://doi.org/10.1016/j.jhydrol.2018.09.043

Danandeh Mehr A, Jabarnejad M, Nourani V (2019) Pareto-optimal MPSA-MGGP: a new gene-annealing model for monthly rainfall forecasting. J Hydrol 571:406–415. https://doi.org/10.1016/j.jhydrol.2019.02.003

DID (2009a) Study on river sand mining capacity in Malaysia, main report. Department of Irrigation and Drainage Malaysia, Kuala Lumpur

DID (2009b) Study on river sand mining capacity in Malaysia, volume I - Sungai Muda. Malaysian Department of Irrigation and Drainage, Kuala Lumpur

DID (2009c) Study on river sand mining capacity in Malaysia, volume II - Sungai Langat. Malaysian Department of Irrigation and Drainage, Kuala Lumpur

DID (2009d) Study on river sand mining capacity in Malaysia, volume III - Sungai Kurau. Malaysian Department of Irrigation and Drainage, Kuala Lumpur

Ebtehaj I, Bonakdari H, Safari MJS, Gharabaghi B, Zaji AH, Riahi Madavar H, Sheikh Khozani Z, Es-haghi MS, Shishegaran A, Danandeh Mehr A (2019) Combination of sensitivity and uncertainty analyses for sediment transport modeling in sewer pipes. Int J Sediment Res 35:157–170. https://doi.org/10.1016/j.ijsrc.2019.08.005

Engelund, F, and Hansen E (1967) A monograph on sediment transport in alluvial streams. In: Teknisk Forlag. Copenhagen, Denmark

Frings RM, Ten Brinke WBM (2017) Ten reasons to set up sediment budgets for river management. Int J River Basin Manag 16:35–40. https://doi.org/10.1080/15715124.2017.1345916

Giustolisi O, Savic D (2006) Symbolic data-driven technique based on evolutionary polynomial regression. J Hydroinf 8:207–222. https://doi.org/10.2166/hydro.2006.020b

Giustolisi O, Savic D (2009) Advances in data-driven analyses and modelling using EPR-MOGA. J Hydroinf 11:225–236. https://doi.org/10.2166/hydro.2009.017

Gunawan TA, Daud A, Haki H, Sarino (2019) The estimation of total sediments load in river tributary for sustainable resources management. IOP Conf Ser Earth Environ Sci 248:0–11. https://doi.org/10.1088/1755-1315/248/1/012079

Haddadchi A, Omid MH, Sdehghani AA (2013) Total load transport in gravel bed and sand bed rivers case study: Chelichay watershed. Int J Sediment Res 28:46–57. https://doi.org/10.1016/S1001-6279(13)60017-7

Harun MA, Ab Ghani A (2020) Revised equations of total bed material load for rivers in Malaysia. In: Mohd Sidek L (ed) WRDM. Springer Singapore, Singapore, pp 332–340

Harun MA, Ab Ghani A, Mohammadpour R, Chan NW (2020) Stable channel analysis with sediment transport for rivers in Malaysia: a case study of the Muda, Kurau, and Langat rivers. Int J Sediment Res 35:455–466. https://doi.org/10.1016/j.ijsrc.2020.03.008

Jang EK, Ji U, Kim KH, Yeo WK (2016) Stable channel design with different sediment transport equations and geomorphologic constraints in Cheongmi stream. KSCE J Civ Eng 20:2041–2049. https://doi.org/10.1007/s12205-015-0126-5

Julien PY, Wargadalam J (1995) Alluvial channel geometry: theory and applications. J Hydraul Eng 121:312–325

Kargar K, Samadianfard S, Parsa J, Nabipour N, Shamshirband S, Mosavi A, Chau KW (2020) Estimating longitudinal dispersion coefficient in natural streams using empirical models and machine learning algorithms. Eng Appl Comput Fluid Mech 14:311–322. https://doi.org/10.1080/19942060.2020.1712260

Karim F (1998) Bed Material discharge prediction for nonuniform bed sediments. J Hydraul Eng 124:597–604. https://doi.org/10.1061/(asce)0733-9429(1999)125:9(985)

Karim MF, Kennedy JF (1981) Computer-based predictors for sediment discharge and friction factors of alluvial streams. Iowa Institute of Hydraulic Research, Report No 242, Iowa City

Khosravi K, Cooper JR, Daggupati P, Thai Pham B, Tien Bui D (2020) Bedload transport rate prediction : application of novel hybrid data mining techniques. J Hydrol 585:124774. https://doi.org/10.1016/j.jhydrol.2020.124774

Kitsikoudis V, Sidiropoulos E, Hrissanthou V (2015) Assessment of sediment transport approaches for sand-bed rivers by means of machine learning. Hydrol Sci J 60:1566–1586. https://doi.org/10.1080/02626667.2014.909599

Laursen EM (1958) The total sediment load of streams. J Hydraul Div, Am Soc Civ Eng 54:1–36

Lee J, Julien PY (2007) Downstream hydraulic geometry of alluvial channels. J Hydraul Eng 132:1347–1352

Mohammadi B, Safari MJS, Kargar K (2021) Implementation of hybrid particle swarm optimization-differential evolution algorithms coupled with multi-layer perceptron for suspended sediment load estimation. CATENA. 198:105024. https://doi.org/10.1016/j.catena.2020.105024

Molinas A, Wu B (2001) Transport of sediment in large sand-bed rivers. J Hydraul Res 39:135–146

Nagy HM, Watanabe K, Hirano M (2002) Prediction of sediment load concentration in rivers using artificial neural network model. J Hydraul Eng 128:588–595. https://doi.org/10.1061/(asce)0733-9429(2002)128:6(588)

Nourani V, Kalantari O, Baghanam AH (2012) Two semidistributed ANN-based models for estimation of suspended sediment load. J Hydrol Eng 17:368–1380

Nourani V, Alizadeh F, Roushangar K (2016) Evaluation of a two-stage SVM and Spatial Statistics methods for modeling monthly river suspended sediment load. Water Resour Manag 30:393–407

Nourani V, Molajou A, Najafi ADTH (2019) A wavelet based data mining technique for suspended sediment load modeling. Water Resour Manag 33:1769–1784

Quinlan JR (1992) Learning with continuous classes. In: 5th Australian joint conference on artificial intelligence. pp 343–348

Rajaee T, Jafari H (2020) Two decades on the artificial intelligence models advancement for modeling river sediment concentration: State-of-the-art. J Hydrol 588:125011. https://doi.org/10.1016/j.jhydrol.2020.125011

Safari MJS, Danandeh Mehr A (2018) Multigene genetic programming for sediment transport modeling in sewers for conditions of non-deposition with a bed deposit. Int J Sediment Res 33:262–270. https://doi.org/10.1016/j.ijsrc.2018.04.007

Sahraei S, Alizadeh MR, Talebbeydokhti N, Dehghani M (2018) Bed material load estimation in channels using machine learning and meta-heuristic methods. J Hydroinf 20:1–17. https://doi.org/10.2166/hydro.2017.129

Saleh A, Abustan I, Mohd Remy Rozainy, MAZ, Sabtu N (2017) Assessment of total bed material equations on selected Malaysia rivers. In: AIP Conference Proceedings. pp 070002(1–7)

Shaghaghi S, Bonakdari H, Gholami A, Kisi O, Shiri J, Binns AD, Gharabaghi B (2018a) Stable alluvial channel design using evolutionary neural networks. J Hydrol 566:770–782. https://doi.org/10.1016/J.JHYDROL.2018.09.057

Shaghaghi S, Bonakdari H, Gholami A, Kisi O, Binns A, Gharabaghi B (2018b) Predicting the geometry of regime rivers using M5 model tree, multivariate adaptive regression splines and least square support vector regression methods. Int J River Basin Manag 17:333–352. https://doi.org/10.1080/15715124.2018.1546731

Sharghi E, Nourani V, Najafi H, Gokcekus H (2019) Conjunction of a newly proposed emotional ANN (EANN) and wavelet transform for suspended sediment load modeling. Water Supply 19:1726–1734. https://doi.org/10.2166/ws.2019.044

Shen HW, Hung CS (1972) An engineering approach to total bed material load by regression analysis. In: Proc., Sedimentation Symposium. pp 14(14.1-14.7)

Shiri N, Shiri J, Nourani V, Karimi S (2020) Coupling wavelet transform with multivariate adaptive regression spline for simulating suspended sediment load: Independent testing approach. ISH J Hydraul Eng 00:1–10. https://doi.org/10.1080/09715010.2020.1801528

Sim LM, Chan NW, Ao M (2015) Stakeholders' participation in sustainable water resource management: a case study of Muda River Basin. In: 6th International Academic Consortium for Sustainable Cities (IACSC) Symposium

Sinnakaudan SK, Ab Ghani A, Ahmad MSS, Zakaria NA (2003) Flood risk mapping for Pari River incorporating sediment transport. Environ Model Softw 18:119–130. https://doi.org/10.1016/S1364-8152(02)00068-3

Sinnakaudan SK, Ghani AA, Ahmad MSS, Zakaria NA (2006) Multiple linear regression model for total bed material load prediction. J Hydraul Eng 132:521–528. https://doi.org/10.1061/(ASCE)0733-9429(2006)132:5(521)

Speed R, Li Y, Tickner D, et al (2016) River restoration: A strategic approach to planning and management. United Nations Education, Scientific and Cultural Organization, Paris. http://www.agu.org/pubs/crossref/2005/2005WR003985.shtm. Accessed 23 Oct 2019

Subhasish D (2011) Fluvial hyrodynamics: hydrodynamic and sediment transport phenomena. GeoPlanet: Earth and Planetary Sciences. Springer, Berlin

Sulaiman MS, Sinnakaudan SK, Azhari NN, Abidin RZ (2017a) Behavioral of sediment transport at lowland and mountainous rivers: a special reference to selected Malaysian rivers. Environ Earth Sci 76:300. https://doi.org/10.1007/s12665-017-6620-y

Sulaiman MS, Sinnakaudan SK, Ng SF, Strom K (2017b) Occurrence of bed load transport in the presence of stable clast. Int J Sediment Res 32:195–209. https://doi.org/10.1016/j.ijsrc.2017.02.005

Syvitski JPM, Cohen S, Kettner AJ, Brakenridge GR (2014) How important and different are tropical rivers? An overview. Geomorphology 227:5–17. https://doi.org/10.1016/j.geomorph.2014.02.029

Tayfur G, Guldal V (2006) Artificial neural networks for estimating daily total suspended sediment in natural streams. Hydrol Res 37:69–79

Tayfur G, Ozdemir S, Singh VP (2003) Fuzzy logic algorithm for runoff-induced sediment transport from bare soil surfaces. Adv Water Resour 26:1249–1256. https://doi.org/10.1016/j.advwatres.2003.08.005

Tayfur G, Karimi Y, Singh VP (2013) Principle component analysis in conjuction with data driven methods for sediment load prediction. Water Resour Manag 27:2541–2554

Templeton WJ, Jay DA (2013) Lower Columbia river sand supply and removal: estimates of two sand budget components. J Waterw Port Coast Ocean Eng 139:383–392. https://doi.org/10.1061/(ASCE)WW.1943-5460.0000188

Ulke A, Tayfur G, Ozkul S (2009) Predicting suspended sediment loads and missing data for Gediz River, Turkey. J Hydrol Eng 14(9):954–965

van Vuren S, Paarlberg A, Havinga H (2015) The aftermath of "Room for the River" and restoration works: coping with excessive maintenance dredging. J Hydro-Environ Res 9:172–186. https://doi.org/10.1016/j.jher.2015.02.001

Wu B, Van Maren DS, Li L (2008) Predictability of sediment transport in The Yellow River using selected transport formulas. Int J Sediment Res 23:283–298. https://doi.org/10.1016/S1001-6279(09)60001-9

Yahaya AS (2019) Application of statistical techniques in environmental modelling. AIP Conf Proc 2129:020074-1–020074–17. https://doi.org/10.1063/1.5118082

Yang CT (1976) Minimum unit stream power and fluvial hydraulics. J Hydraul Div 102:919–934

Yang CT (1979) Unit stream power equations for total load. J Hydrol 40:123–138

Zahiri A, Najafzadeh M (2018) Optimized expressions to evaluate the flow discharge in main channels and floodplains using evolutionary computing and model classification. Int J River Basin Manag 16:123–132. https://doi.org/10.1080/15715124.2017.1372448

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.