Catena 196 (2021) 104917

Contents lists available at ScienceDirect

Catena

journal homepage: www.elsevier.com/locate/catena

Predictive models for the estimation of riverbank erosion rates

A. Saadon[a], J. Abdullah[b], N.S. Muhammad[c,*], J. Ariffin[b], P.Y. Julien[d]

[a] Faculty of Engineering and Built Environment, SEGi University, Petaling Jaya, Selangor, Malaysia
[b] Faculty of Civil Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[c] Department of Civil Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
[d] Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO 80523, USA

ARTICLE INFO

Keywords:
Predictive models
Bank erosion measurements
Riverbank erosion rates
Multiple regression
Bernam River

ABSTRACT

Riverbank erosion is a complex soil-water interaction process, highly dynamic and constantly changing. Consequently, the estimation of riverbank erosion rate requires an in-depth understanding between the riverbank properties and the hydraulic characteristics of the river. Given the complexity of predicting riverbank erosion rate and the limitation of existing analytical solutions, this study developed empirical models that include the analyses of basal erosion and bank failure using 358 erosion pin measurement collected from River Bernam in Selangor, Malaysia. Field measurements of the river and sediment data were performed following strict international standard protocols and instrumentation, such as gauging and wading technique, survey pole and SEBA F-1 current meter. Based on 50 years of record, the mean annual flood in River Bernam is 10 m^3/s. This study includes extensive analysis between each measured variable in representing the factors influencing riverbank erosion rates, development of empirical predictive models in quantifying riverbank erosion rates using statistical approach and improves predictive performance. Model parameterization was performed using sensitivity analysis and comparison of measured riverbank erosion rates with flow-induced variables yielded the strongest correlation, whereas other variables were found to be less significant. These variables were evaluated based on Pearson's correlation coefficient, trend of the scatter plots and degree of determination. Novel findings from the sensitivity analysis of this study are one of the substantial factors in deriving the most influential parameters which constitutes to the rate of bank erosion. Multiple linear, non-linear and logarithmic functions were employed in the development of new predictive model. This study concluded that the developed empirical equations using logarithmic-transformation is the best predictor. Logarithmic-transformation equations show the highest percentage of accuracy and degree of determination, i.e. up to 93.5% and 0.783, respectively, and the data are between the limit of discrepancy ratio $0.5 < DR < 2$. The influential factors include hydraulic characteristics of the flow, soil characteristics and bank geometry. On the other hand, parameters of the ratio of bankfull width to the mean particle diameter, bank angle, and channel reach yield very weak correlation with bank erosion rates.

1. Introduction

The rate of riverbank erosion is significantly affected by parameters related to the pulses of water against the bank, changes in water level, geometry of the bank and the characteristics of bank material. These parameters are highly dynamic and constantly changing, therefore, the estimation of the rate of bank erosion requires an in-depth understanding between the hydraulic characteristics of the flow affecting the bank, including the bank properties. This soil-water interaction is a complex process that may cause significant changes in river morphology. Thus, an increase in river erosion rate is expected, causing significant socio-economic impacts such as lost of ecological diversity (e.g. Roy et al., 2019), high cost of bank stabilization (e.g. Bernhardt et al., 2005) and damage to riparian land and infrastructure (e.g. Barman et al., 2019).

Several prediction models have been derived using multiple linear regression and logistic regression and these include models developed by Varouchakis et al. (2016); Saadon et al. (2016); Toriman et al. (2010); Richard et al. (2005); and Nanson and Hickin (1983). Saadon et al. (2016) derived an empirical equation quantifying riverbank erosion rates using time series data from field measurements. However, the developed empirical model only uses linear function in the regression analysis and the developed model performed at 60% accuracy. Varouchakis et al. (2016) uses a more advanced statistical approach, i.e.

* Corresponding author.
E-mail address: shazwani.muhammad@ukm.edu.my (N.S. Muhammad).

https://doi.org/10.1016/j.catena.2020.104917
Received 22 January 2020; Received in revised form 11 August 2020; Accepted 8 September 2020

logistic regression and locally weighted regression techniques in predicting area susceptible to erosion. Most of these researchers established solutions in predicting bank erosion spatially, hence, neglecting its occurrences temporally. Any advanced instruments, for instance, Photo-Electronic Erosion Pins (PEEP) which functions to directly measure the rates of erosion at a particular point of erosion are expensive (Lawler, 1993), as it involves advanced installation and maintenance. These type of pins also prone to be washed away by flood water when installed at the riverbank. Reliable and continuous data for bank erosion rates, flow-induced parameters, bank geometry, soil properties and bank resistance are necessary for a detailed assessment of bank erosion rates. The integration between flow-induced parameters and the bank properties will provide accurate predictive model in quantifying the bank erosion rates. Furthermore, existing developed models did not clearly explain the model parameterization analysis, especially with representation of multiple factors and variables over a large scale study areas. The existing prediction models have presented lack findings on the effects of correlation between variables, as the influential factors to the rate of bank erosion.

1.1. Riverbank erosion predictive models

Research on significant parameters influencing riverbank erosion rate have evolved since the 1980s. A list of prominent research related to this study is given in Table 1. Table 1 summarizes the review of the existing equations with the influential parameters carried out to identify the common variables deemed significant by the various researches. In general, these studies suggested that erosion parameters are associated with flow-induced forces, bed particle size, bank particle size, concentration of suspended sediment, channel width, channel gradient, planform geometry and vegetative protection. Ikeda et al. (1981) reasoned that one of the most important soil erodibility parameters is the hydraulic action which includes the effect of near-bank velocity acting on the bank. These flow-induced forces (i.e. drag force, resistance force, and lift force) acts on the bank surface (Duan, 2005; Duan and Julien, 2010). Due to substantial evidences denoting a relationship between the rates of erosion and hydraulic characteristics of the water, Hasegawa (1989) has derived a generalized universal bank erosion coefficient based on the channel shift data collected from several Japanese rivers. The derivative bank erosion coefficient has included the type of bank texture (sand or clay), however, it did not clearly investigate the integration between fluvial and bank properties. Several other studies conducted in regards to riverbank erosion focused on the lateral movements of the river planform, such as Engel and Rhoads (2017) and Roy et al. (2019). Previous studies (Richard et al., 2005; Randle, 2004; Constantine et al., 2009) applied statistical analyses to define the relationships between measured erosion rates and hydraulic characteristics, sediment, channel geometry variables, resistance of the bank, and soil properties. Duan (2005) developed an analytical solution to quantify the rate of bank erosion that includes the effects of hydraulic force, bank geometry, bank material properties and probability of bank failure.

Table 1
Summary of riverbank erosion and migration parameters published by previous researchers and this study.

Source	Parameters used in equations	Description
Ikeda et al. (1981)	ξ/u_b	Ratio of erosion rates to the near-bank velocity
Hasegawa (1989)	$\tau_{bo}/(\tau_{bo}-\tau_{bc})$	Ratio of boundary shear stress to critical shear stress
	u_b^2/gd_{50}	Ratio of near-bank velocity to the mean particle size
Randle (2004, 2006)	B/r_c	Bankfull width to channel radius curvature
	r_d/h_b	Vegetation root depth to the height of bank
	d_W/D	Average height of large woody debris to the hydraulic water depth
	h_b/D	Height of bank to the hydraulic water depth
	LWD	Fraction of area covered by trees or large woody debris (%)
	r_γ	Fraction of bank area covered by vegetation roots (%)
	d_c	Portion of bank sediment too coarse for incipient motion (%)
	C_s	Bed material sediment concentration (ppm)
	ξ/u_b	Ratio of erosion rates to the near-bank velocity
Duan (2005): Duan and Julien (2010)	$1-(\tau_{bc}/\tau_{bo})$	Ratio of boundary shear stress to critical shear stress
	C/C_*	Ratio of actual concentration of suspended load to the equilibrium concentration of suspended sediment
	d_{50}/k_s	Mean particle size to roughness height
	D/d_{50}	Hydraulic water depth to mean particle size
	B	Bank angle (degrees)
Constantine et al. (2009)	ξ/u_b	Ratio of erosion rates to the near-bank velocity
	u_b/H	Near-bank velocity to reach-averaged depth
	$\tau_{bo}/\rho u_b^2$	Friction factor
Posner and Duan (2012)	$\tau_{bo}/\rho u_b^2$	Friction factor
Varouchakis et al. (2016)	β	Bank slope (degrees)
	B	Cross section width (m)
This study	ξ/u_b	Ratio of bank erosion to near-bank velocity
	$\tau_{bo}/\rho u_b^2$	Ratio of boundary shear stress to near-bank velocity
	u_*/u_b	Ratio of shear velocity to near-bank velocity
	ω/u_b	Ratio of fall velocity to near-bank velocity
	$\tau_c/\rho u_b^2$	Ratio of critical shear stress to near-bank velocity
	gd_{50}/u_b^2	Ratio of mean size particle to near-bank velocity
	B/d_{50}	Ratio of channel width to mean particle size
	Y/d_{50}	Ratio of water depth to mean particle size
	β	Bank angle (degrees)
	S_0	Reach slope
	h_b/d_{50}	Ratio of bank height to mean particle size
	C	Equilibrium concentration of suspended sediment

Bank geometry characteristics with regards to soil type and pore water pressure are determinants of the rate of bank erosion while planform geometry of the channel provides information on lateral erosion rates. This can be useful in the prediction of river advances or retreat. The most studied parameter by several researchers (Hickin and Nanson, 1975, 1984; Randle, 2004, 2006; Varouchakis et al., 2016) is the ratio between channel radius of curvature and channel width (r_c/B). The effect of other bank geometry variables, such as bank angle, β and height of the bank, h_b (Duan, 2005; Randle, 2004, 2006; Varouchakis et al., 2016) suggests a strong correlation to erosion rates. A model simulating channel migration process was developed by Randle (2004) who has included function of sediment transport capacity, radius of channel curvature and the bank material properties acting to resist the erosion. The erosion resistance factors include vegetation, large woody debris, cohesion and armoring. Although this model may be used as a tool for predicting future channel alignment, the approach does not carefully consider the correlation between fluvial erosion and bank failure. One of the limitations of the model is that a land use plan is needed to assess the fraction of area covered by vegetation roots as the armoring towards bank erosion.

Particles on riverbanks and on channel bed are subjected to both kinetic energy of the flowing water and gravitational forces. These particles can be easily transported should the flow-induced parameters be larger than particle weight and resistance. Particle size has great influence on the rate of erosion (Randle, 2004, 2006; Duan, 2005, Duan and Julien, 2010). Factors that may affect resistance of the bank include the percentage of vegetative cover, vegetative root depth, and the fraction of large woody debris on the bank. Nevertheless, the riverbank face length or bank toe angle can have some effect on the bank resistance. Duan (2005) included the effect of bed roughness thickness, k_s as a means of armor to the channel bed. However, in extreme conditions with relatively high rainfall, small and large particles can be easily eroded and transported. Duan (2005) provided an analytical solution to quantify rate of erosion, which was derived using the relationship between fluvial erosion and bank failure processes. The model included various parameters such as hydraulic component of the flowing water, geometry of the bank, soil properties and bank failure likelihood. The bank failure likelihood was presented by the computation of bank failure and the critical bank height, H, using heuristic approach. Further work is needed to determine a direct method in quantifying riverbank erosion rates, specifically using fieldwork measurements data.

With the limitations mentioned, we represent riverbank erosion rates as a function of flow-induced parameter, bank geometry, soil properties, bank resistance and the effect of sediment supply. The new developed empirical model significantly improves prediction compared to the existing empirical solutions. This study aims to (1) provide analysis between each measured variables in representing the factors influencing riverbank erosion rates, (2) development of empirical predictive models in quantifying riverbank erosion rates using statistical approach, and (3) improves predictive performance. Model parameterization performed using sensitivity analysis and comparison of measured riverbank erosion rates with flow-induced variables yielded the strongest correlation, whereas other variables were less significant. Multiple linear, non-linear and logarithmic function were employed in the development of new predictive model. Model performance exhibited the newly predictive model achieved high accuracy by combining multiple factors, despite low correlation between variables.

2. Methodology

This section covers the methodology used in this study to meet the study objectives as highlighted in Fig. 1, namely, (i) field measurements, (ii) functional relationship and sensitivity analysis, (iii) regression models, and (iv) model performance.

2.1. Field measurements

Field measurements of riverbank erosion rates were obtained at Sungai (Sg.) Bernam (or River Bernam), Selangor. The Bernam river basin has an area of 2836.33 km^2. Sg. Bernam is located between the states of Perak and Selangor, demarcating the border between these two states of latitude 3°51.02 N and longitude 100°50.15 E. It flows from the upstream of Mount Triang Timur, located in the east of Mount Titiwangsa to the Straits of Malacca in the west, as shown in Fig. 2. The upper region of the catchment is dominated by vegetation cover such as tropical hill rainforests, oil palm trees and rubber trees. Other land covers are several small or medium sized urbanized areas, especially along the riverbanks and roadsides. Most downstream areas are swampy. Records of observed maximum flow, water level and rainfall for the respective station show that most maximum flow annually occurred between the month of April – May and September – December. Based on 50 years of record (i.e. 1965–2014), the mean annual flood in Sg. Bernam is 10 m^3/s. Field data collection was located along the main river, Sg. Bernam, located 800 m upstream from the flow station. The measured location consist of a 50 m reach length on the left and right banks. Both banks were divided into several intervals of 10 m each, as shown in Fig. 2. The bank height extending from 3.7 m to 3.8 m and inclined at an angle ranging between 60 and 70 degrees.

The reach was selected based on the physical observation of the eroded bank. The outer bank of the reach evidenced fluvial entrainment and mass failure of the bank. Cantilever failure observed at this section on the outer bank. The right bank (inner bank) is nearly vertical with bed varies of shallow ripples and sand point bar. Some points are exposed to weathering and fluvial entrainment. Basal cleanout observed along the section vertically and horizontally. Point 2 and 3 experienced cantilever failure on the outer bank, consist of a structure defined and cascading bank on the floodplain and exposed to crack development on the bank surface. Slab failure is observed on the inner bank at Points 2 and 3. Horizontal tension cracks observed within the bank. Point 3 (inner bank) evidenced rotational and slump failure. Major section of eroded bank materials can be observed at Point 4 (30–40 m). The outer bank height ranging from 3.6 m to 3.7 m. Rotational and slump failure observed on the upper bank and cantilever failure observed at the bank near water edge and hanging with bushy vegetation. Inner bank height extending from 3.4 m to 3.7 m. Slab and block failure are the most typical observed at this section. Fallen block of silt materials are visible at the bank foot.

Field measurements of the river and sediment data include riverbank erosion rate, hydraulics characteristics, bank geometry, soil characteristics and properties, bank resistance and sediment supply. Following this, boundary shear stress, shear velocity, fall velocity and critical shear stress were calculated using formula given in Julien (2012). A summary of variables for the different classes and equipment/method/formula used in this study is presented in Table 2. A total of 358 erosion pin measurements were collected in one year duration. Riverbank erosion rates were measured using erosion pins as presented by Lawler (1993). A schematic diagram on the arrangement of erosion pins for erosion measurements is shown in Fig. 3. There are six transect locations for the erosion pin points. More pins were driven at critical points i.e. high flow, river bend, etc. for better representation of the eroded profiles. Records of erosion depth measurements were taken at the point of erosion, bi-weekly during normal flow and after rainfall events. The values of erosion depths were translated into bank erosion rates by dividing the erosion depth with the number of days.

The gauging technique employed in this study was in accordance with the Drainage and Irrigation Department, Hydrological Procedure No. 15 (1976), reprinted (1995). Wading technique was employed for velocity measurement for depth of flow less than one-meter. This study adopted the measurement of near-bank velocity similarly to Hasegawa (1989), where the near-bank velocity was measured at 1 m distance from the eroded bank horizontally, rather than the stream-averaged

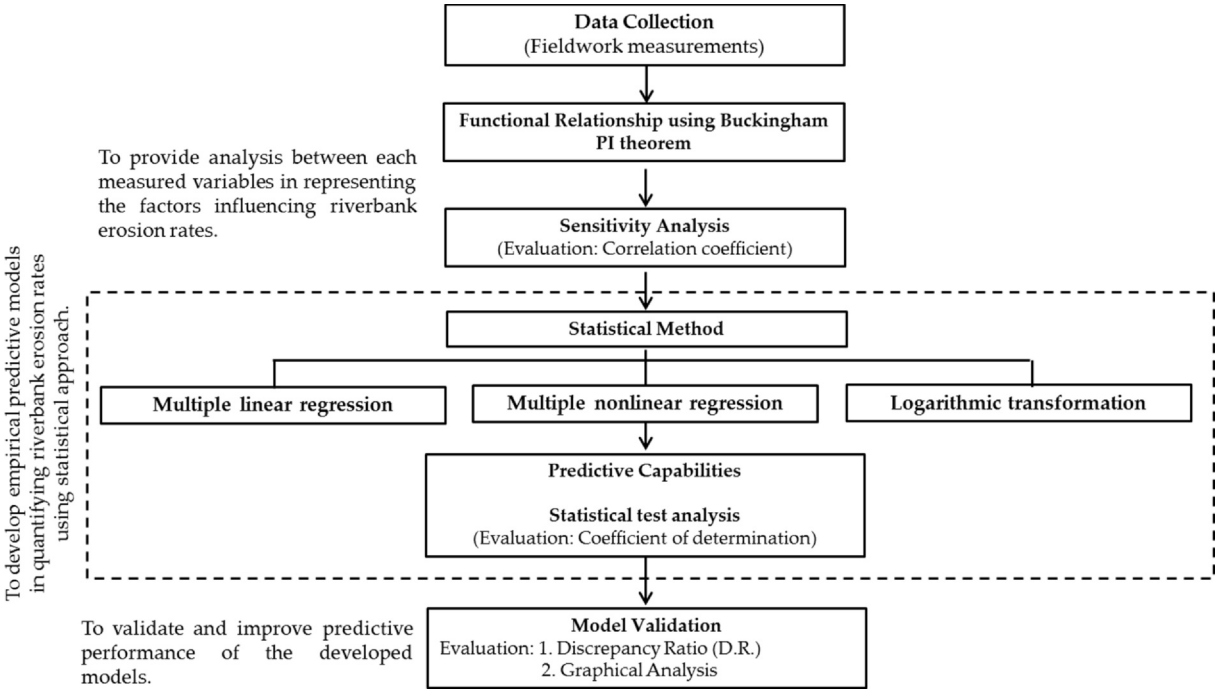

Fig. 1. Flow chart of the methodology.

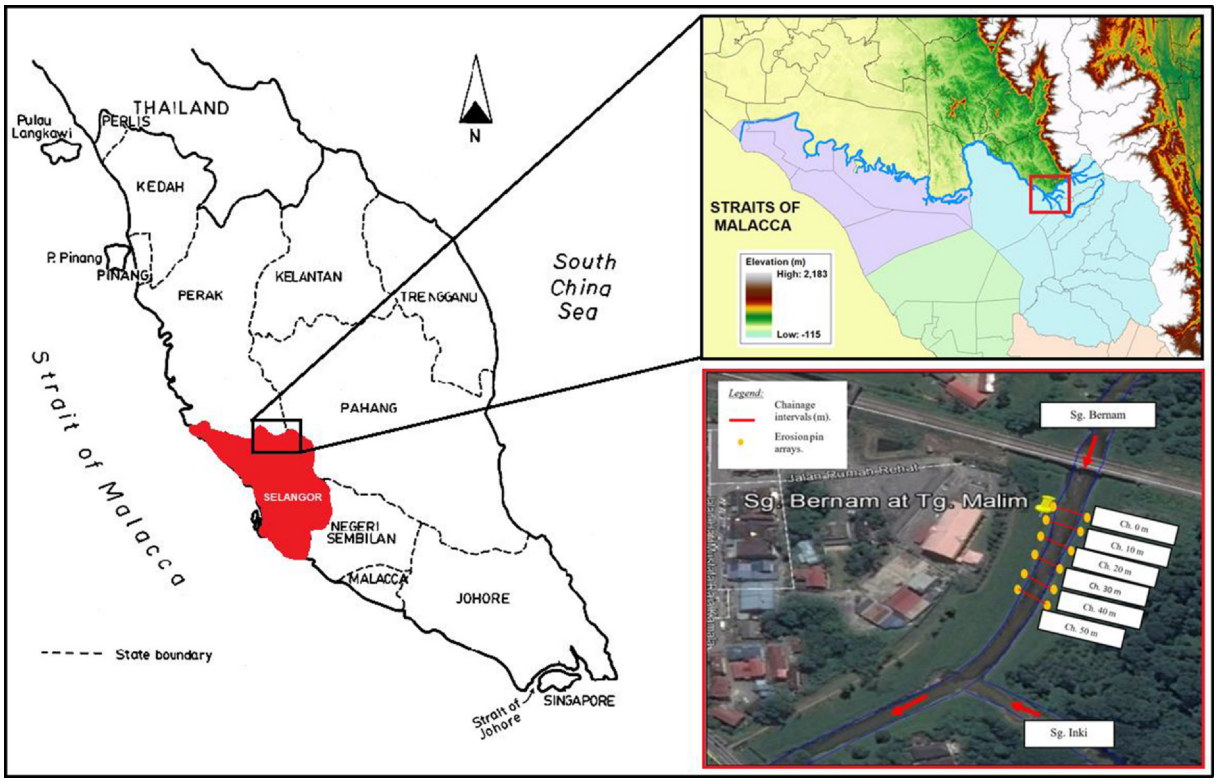

Fig. 2. Map and aerial view of the point of measurements.

velocity, *u,* and controls the phenomenon of bank erosion or accretion. Measurements of channel width were taken before measurements of near-bank velocity. The width was measured using survey measuring tape and tied onto a stick on both sides of the banks. The depth of flow at each vertical point was first determined to locate the point at which the velocity should be measured. This was done using a survey pole. For the depth (d) of flow less than 0.3 m, the velocity was measured using one-point measurement, at 0.6d from the water surface. It is considered as the average velocity for the respective vertical point. For water depth exceeding one-meter, the velocity measurement is taken using two-point measurements, at 0.2d and 0.8d from the water surface. Near-bank velocity was measured at 1 m horizontal distance from the eroded bank. In all operations, a standard universal SEBA F-1 current meter was used to measure velocity.

The near-bank velocity (u_b) can be a useful parameter affecting the riverbank erosion rate (ξ). The erosion rate is a measure of eroded

Table 2
Variables for the different classes and equipment/method/formula used in this study.

No	Classes	Variables	Symbol	Equipment/Method/Formula
1.	Riverbank erosion rate	Erosion rates in unit length over time	ξ	Conventional erosion pins drive perpendicular to the bank surface
2.	Hydraulic characteristics	Near-bank velocity	u_b	Vela-port velocity meter
		Discharge	Q	OTT Q-liner
		Boundary shear stress	τ_o	Calculated value*
		Shear velocity	u_*	Calculated value*
		Fall velocity	ω	Calculated value*
		Water depth	Y	Auto level
3.	Bank geometry	Height of bank	h_b	Auto level
		Bank angle	β	Slope meter
		Reach slope	S	Auto level and measuring tape
		Bankfull width	B	Measuring tape
4.	Soil characteristics and properties	Critical shear stress	τ_c	Calculated value*
5.	Bank resistance	Bank material	$d_{50(bank)}$	Hand Auger and sieve analysis
		Bed material	$d_{50(bed)}$	Van Veen grab sampler and sieve analysis
6.	Sediment supply	Equilibrium concentration of suspended sediment	C	Portable suspended load apparatus

Note: * Calculated value using formula given in Julien (2012).

material on the bank surface in unit length over time. These measurements are taken from the readings attached on the erosion pins driven perpendicular to the bank surface divided by number of days. Since the point on measurements include both right and left banks of the channel, the outer bank exhibit erosion and deposition occur along the inner bank. Accumulation of bank materials on the erosion pins caused difficulty in the deposition measurements. Most of the erosion pins were buried due to the bank material deposits, and new pins were installed for replacement. Hence, the deposition of bank materials were not included in the empirical model development in this study. Fig. 4 shows the relationship between the bank erosion rate and near-bank velocity for one year of fieldwork measurements. The considerable scatter in the data suggests that the bank erosion rate is extremely complex and may be a function of factors other than near-bank velocity. The bank erosion rate is maximized at 0.2 m/s $< u_b <$ 0.5 m/s, and the corresponding rate of bank erosion ranging between 1.5 m/year and 1.8 m/year. This

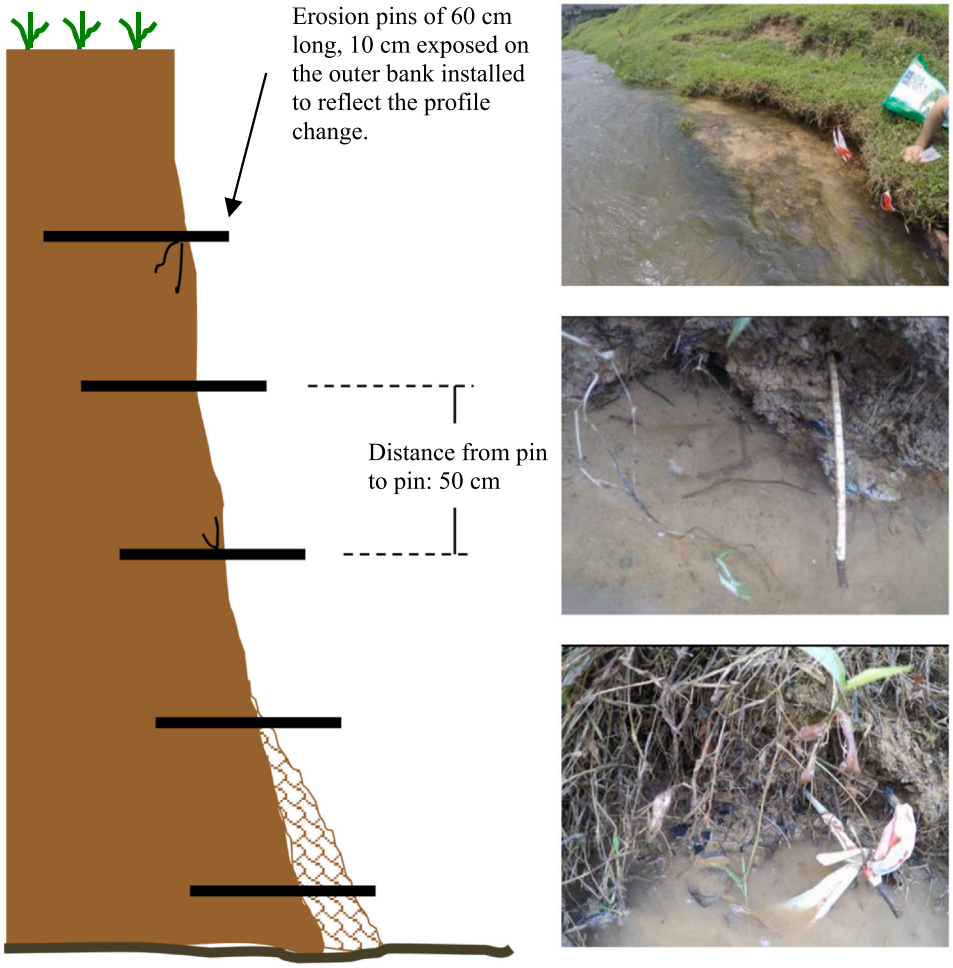

Fig. 3. Schematic diagram on the erosion pins arrangement and field measurements.

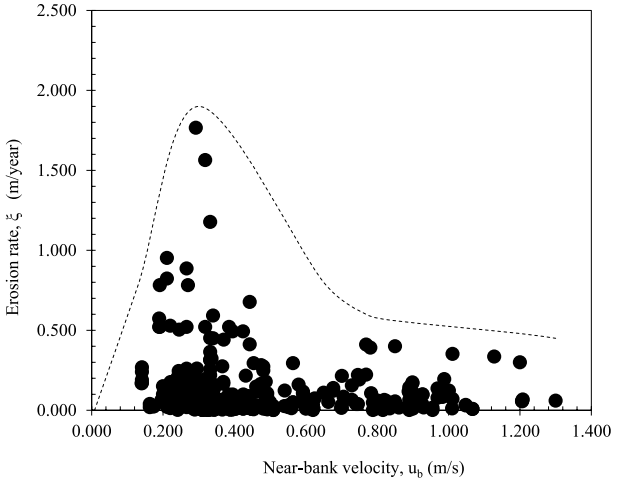

Fig. 4. Relationship between bank erosion rate and near-bank velocity.

shows that the rate of bank erosion is influenced by low near-bank velocity, and the occurrence of cracks on the bank face reduces the bank stability. The highest erosion rate of 1.8 m/year is estimated at near bank velocity of 0.4 m/s.

2.2. Functional relationship and sensitivity analysis

In order to incorporate variation of factors influencing the bank erosion rates, this study suggest a functional relationship consist of various parameters which have been measured during the fieldwork. The fundamental concept is to suggest a non-dimension relationship and identify dimensionless groups serve as a function to quantify bank erosion rate. Possible governing variables for riverbank erosion were identified using method of repeating variables, known as the Buckingham PI theorem. The derived variables are grouped into five (5) categories; riverbank erosion rate, hydraulic characteristics, bank properties and soil properties, and sediment supply, as shown in Table 3. The variables in the different categories include near-bank velocity, shear velocity and shear stress, bank geometry and soil properties. Near bank velocity and shear velocity describe the velocity near the boundary of a flow while shear stress describes the shear stress in between the fluid layers. Knowledge on flow and boundary shear stress near the banks of natural channels are essential in the prediction of bank and bed materials transport near the banks that lead to riverbank erosion. The bank geometry and relative roughness of the bank (soil properties) will affect the stress distribution of the bed and banks.

Table 3
Selected variables used in the dimensional analysis.

No	Classes	Variables	Symbol	Units	Dimensions
1.	Riverbank erosion rate	Erosion rates in unit length over time	ξ	m/s	LT^{-1}
2.	Hydraulic characteristics (flow resistance)	Near-bank velocity	u_b	m/s	LT^{-1}
		Boundary shear stress	τ_o	N/ms^2	$ML^{-1}T^{-2}$
		Shear velocity	u_*	m/s	LT^{-1}
		Fall velocity	ω	m/s	LT^{-1}
		Water depth	Y	m	L
3.	Bank properties and soil properties	Height of bank	h_b	m	L
		Bank angle	β	-	-
		Reach slope	S	-	-
		Bankfull width	B	M	L
		Critical shear stress	τ_c	N/ms^2	$ML^{-1}T^{-2}$
4.	Bank resistance	Bank material	$d_{50(bank)}$	mm	L
5.	Sediment supply	Equilibrium concentration of suspended sediment	C	–	–

Three variables were used as repeating variables, i.e. near-bank velocity (u_b), density (ρ_w), and mean particle diameter of bank material (d_{50}). The mean particle diameter ranging between 0.9 mm and 1.8 mm. These variables were selected based on the measured data during the fieldwork. The selection of repeating variables were made based on the following considerations: (i) the selected repeating variable represent all the primary dimensions in the problem (M, L and T); (ii) the selected repeating variables must not by themselves be able to form a dimensionless group, otherwise it would be impossible to generate the functional relationship; (iii) the selected repeating variables should be common variables to appear in each functional relationship; and (iv) the selected repeating variables should be simple variables over complex variables whenever possible.

The dimensionless group of variables derived using Buckingham PI theorem were group into parameter classes. These parameter classes constitutes the factors influencing riverbank erosion rates. Based on the parameter classes, sensitivity analysis was performed on all dimensionless group established hypothetically to evaluate the significance and correlation between each independent variables (IV) and dependent variable (DV). The IVs serve as the function to the DV, independently on its own. Sensitivity analysis is important prior to regression analysis. Scatter plots of DV against each IV were established for all dimensionless group parameters. These parameters were evaluated based on statistics correlation functions, namely correlation coefficient (r-squared) and Pearson's correlation. High calculated value of correlation coefficient (r-squared) between IV and DV is to unity. Perfect Pearson's correlation is defined when the calculated value is 1 or −1 and zero for no correlation (Richard et al., 2005). The highly correlated parameters (positive or negative) are identified as the influential parameters to the riverbank erosion rate.

2.3. Regression models

Predictive models quantifying riverbank erosion rates using statistical approach from 215 fieldwork measurements at Sg. Bernam, while the remaining 143 data were used to validate the model. Nonlinear model uses the power relationship, while the logarithmic transformation models uses the $\log_{10}$-transformed regression. The general linear regression theoretical derivation of the formula is as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \quad (1)$$

where, y is the dependent variables or responsive variable; x_1, x_2, x_3, … , x_k is the independent variables or predictors; β_0 is the intercept, the value of y when all x are zero; β_k is the population regression coefficients; and ε is the random error term.

In the case of assessing a nonlinear relationship, the power relationship in a nonlinear general form of a power functions between independent variables to the dependent variable (Eq. (2)).

$$y = (\beta_0)(x_1^{\beta_1})(x_2^{\beta_2})(x_3^{\beta_3})\cdots(x_i^{\beta_i}) \quad (2)$$

where, y is the dependent variable; x_1, x_2, and x_i are independent variables; β_0, β_1, β_2, β_3 and β_i and constants generated from power regression analysis.

Logarithmically transforming variables is a common solution when a nonlinear relationship exists between the IVs and DVs. There are four common categories of logarithmic models with logarithmic transformations, namely, linear model, linear-log model, log-linear model, and log-linear model. This study focuses on the derivation of logarithmic transformation technique using log-log model as in Eq. (3).

$$\log y = \beta_0 \times \log(x_1)^{\beta_1} \times \log(x_2)^{\beta_2} \times \log(x_3)^{\beta_3} \times \log(x_i)^{\beta_i} \quad (3)$$

where, log y is the dependent variables or responsive variable; log x_1, x_2, x_3, … , x_i is the independent variables or predictors; β_0 is the intercept, the value of y when all x are zero; and β_i is the population regression coefficients.

2.4. Model performance

Model performance and validation uses the remaining field data, i.e. 143 erosion pin measurements (40% of the total data), based on the optimal split proportions approach highlighted in the previous section. Prediction accuracy uses R^2, and discrepancy ratio (DR), which is the ratio of the predicted values to the measured values. Model performance includes graphical representation of scatter plot between predicted values and measured values (Azamathulla et al., 2010). The empirical model exhibit excellent prediction when the predicted values calculated similar prediction as the measured values. The values are deemed to be accurately acceptable if the data lie within 0.5 – 2.0 limit along the line of perfect agreement (Saadon et al., 2020, 2016; Sinnakaudan et al., 2010).

3. Results and discussion

This study uses the Statistical Package for the Social Sciences (SPSS) for outliers filtering, data splitting, analysis and derivation of empirical equations for the riverbank erosion rate. Three regression models were employed in the development of a new riverbank erosion equation, namely, (i) multiple liner regression, (ii) multiple nonlinear regression, and (iii) logarithmic-transformation regression. The mathematical models were selected based on the highest correlation coefficient obtained from the scatter plots of fieldwork data area. This section highlights the findings of the study based on (i) influential parameters for riverbank erosion using dimensional analysis and model parameterization, (ii) riverbank erosion estimates using statistical approach, and (iii) best regression models.

3.1. Influential parameters for riverbank erosion

Identifying the influential parameters for riverbank erosion denotes the most critical aspect in developing a new riverbank erosion equation. Dimensional analysis was performed for the derivation of a functional relationship corresponding to riverbank erosion rates (e.g. Azamathulla et al., 2010; Chavan and Kumar, 2018; Kuntjoro and Harijanto, 2018). Prior to that, model parameterization using sensitivity analysis was performed in order to assist in developing the numerical equations. The sensitivity analysis was utilized to observe a trend between dependent variable (DV) to the independent variables (IV).

Dimensional analysis utilized 15 variables and 3 fundamental quantities involved in the relationship, denotes as the *n*-variables. The significant parameters controlling the rate of riverbank erosion (ξ) along a channel can be divided into five (5) major categories; bank geometry, hydraulic, soil capacity (resistance to erosion), grain resistance and others. The parameters for bank geometry consist of channel width, B, water depth, y, bank height, h_b, bank angle, β, and channel slope, S_o. The parameters for hydraulic consist of streambank erosion rate, ξ, streambank depth-averaged velocity, u_b, and boundary shear stress, τ_o. The parameters for soil capacity (resistance to erosion) includes critical shear stress, τ_c and the grain resistance includes mean particle diameter, d_{50}, fall velocity, ω, shear velocity, u^*, and concentration of suspended load to equilibrium suspended concentration, C. Other variables include gravity acceleration, g, and water density, ρ_w. The rate of bank erosion (ξ) serves as the dependent variable. Twelve non-dimensional parameters were obtained using Buckingham PI theorem; one dependent non-dimensionless parameter to eleven independent non-dimensionless parameters. These parameters are shown in Eq. (4) below.

$$\frac{\xi}{u_b} = f\left(\frac{B}{d_{50}}, \frac{Y}{d_{50}}, \frac{h_b}{d_{50}}, \beta, S_0, \frac{\tau_0}{\rho_w {u_b}^2}, \frac{\tau_c}{\rho_w {u_b}^2}, \frac{\omega}{u_b}, \frac{u_*}{u_b}, C, \frac{gd_{50}}{{u_b}^2}\right) \quad (4)$$

These dimensionless variables are grouped into five (5) parameter classes shown in Table 4. These parameters were further evaluated on their significance in the model parametrization using sensitivity analysis.

Table 4
Selected variables used in the dimensional analysis.

No	Parameter class	Dimensionless group	Variable category
1.	Riverbank erosion rate	$\frac{\xi}{u_b}$	Dependent variable
2.	Hydraulic characteristics	$\frac{\tau_o}{\rho_w u_b^2}$, $\frac{\omega}{u_b}$, $\frac{u_*}{u_b}$, $\frac{gd_{50}}{u_b^2}$	Independent variable
3.	Bank properties and soil properties	β, S_o, $\frac{\tau_c}{\rho_w u_b^2}$	Independent variable
4.	Bank resistance	$\frac{B}{d_{50}}$, $\frac{Y}{d_{50}}$, $\frac{h_b}{d_{50}}$	Independent variable
5.	Sediment supply	C	Independent variable

Table 5
Correlation coefficient of the independent variables to the dependent variable.

Notation	Dimensionless parameter	Linear and nonlinear correlation model (*r*-squared)	Pearson's correlation model
IV1	$\frac{\tau_o}{\rho_w u_b^2}$	0.335	0.576
IV2	$\frac{gd_{50}}{u_b^2}$	0.390	0.483
IV3	$\frac{u_*}{u_b}$	0.353	0.576
IV4	$\frac{\omega}{u_b}$	0.412	0.521
IV5	$\frac{\tau_c}{\rho_w u_b^2}$	0.399	0.463
IV6	$\frac{B}{d_{50}}$	0.220	−0.056
IV7	$\frac{Y}{d_{50}}$	0.200	0.064
IV8	$\frac{h_b}{d_{50}}$	0.150	0.038
IV8	β	0.160	−0.025
IV9	S_o	0.210	−0.035
IV10	C	0.112	0.037

A sensitivity analysis is performed on all established dimensionless parameters. Each of the IVs was plotted against the DV. These trends provide indication on the significance of the variables for model development. The correlations were obtained from the scatter plots and Pearson's correlation using *r*-squared in the linear and nonlinear correlation models. Table 5 presents the correlation coefficient between the dependent variable to the independent variables used in this study.

Five of the independent parameters exhibit strong correlation with the dependent parameter. They are the ratio of boundary shear stress to the near-bank velocity (Fig. 5a), ratio of mean particle size to the near-bank velocity (Fig. 5b), ratio of shear velocity to the near-bank velocity (Fig. 5c), ratio of fall velocity to the near-bank velocity (Fig. 5d), and ratio of critical shear stress to the near-bank velocity (Fig. 5e). The highest correlation was obtained for ratio of fall velocity to the near-bank velocity in both the linear and nonlinear correlations while the ratio of shear velocity and boundary shear stress to the near-bank velocity showed high Pearson's correlation.

A strong positive correlation was also found in the plot of IV3 (shear velocity to the near-bank velocity) against DV (ratio of erosion rate to the near-bank velocity). Neglecting the denominator term of the parameter fraction increases the shear velocity symmetrically. It can be concluded that the variation of riverbank erosion rate is accounted for by the variation of the shear velocity of the flow. Similar trend can be observed for IV4 (ratio of fall velocity to the near-bank velocity). IV2, ratio of boundary shear stress to the near-bank velocity indicates a positive correlation to the DV with Pearson's correlation accuracy of 0.567. The boundary shear stress (τ_o) is the force per unit area exerted on the surface beneath the flow in the direction of flow. This is related

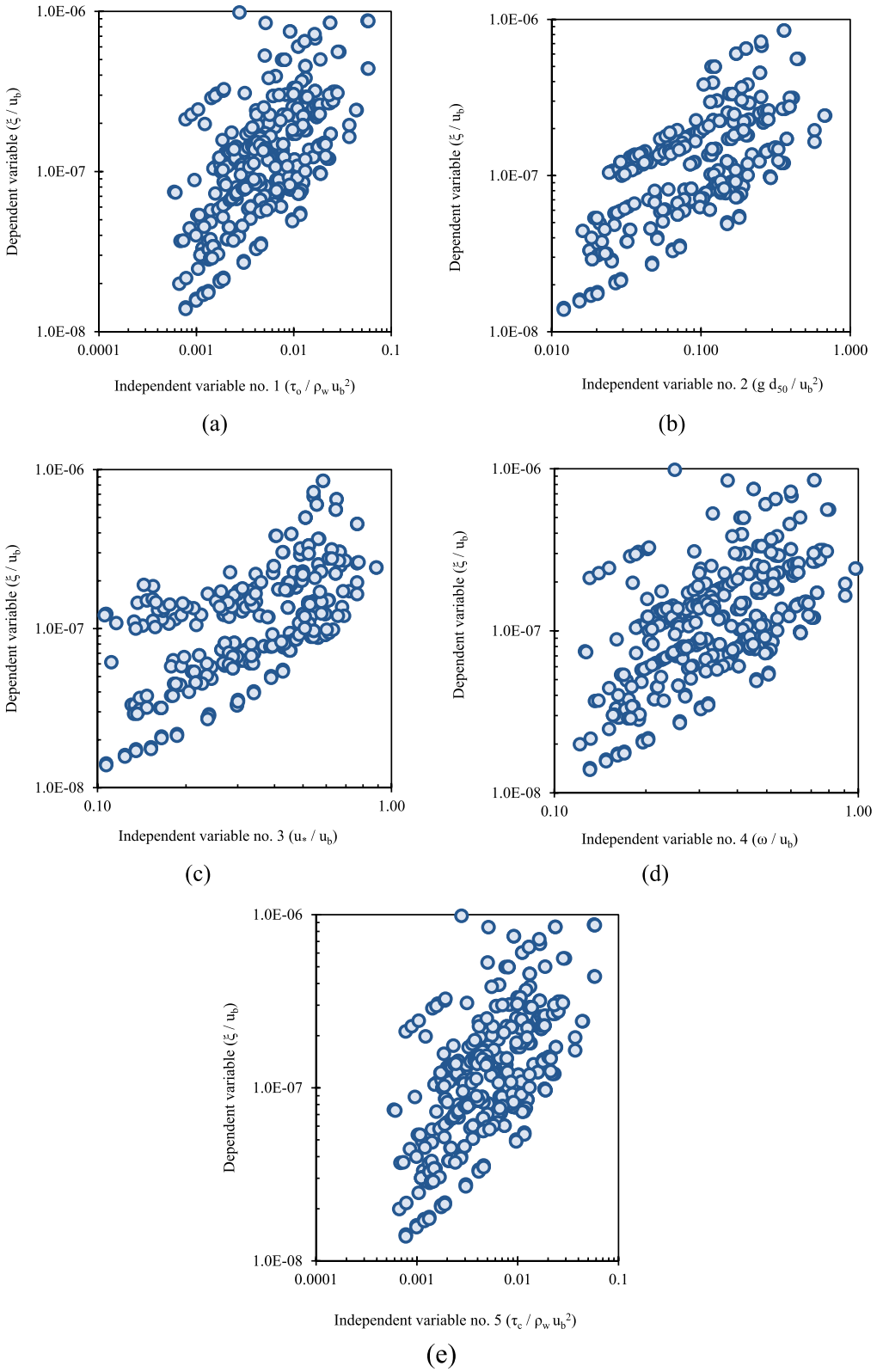

Fig. 5. Relationship between each independent variables (a) $\tau_o/\rho_w\ u_b^2$; (b) $g\ d_{50}/u_b^2$; (c) u_*/u_b; (d) ω/u_b; and (e) $\tau_c/\rho_w\ u_b^2$ to the dependent variable, ξ/u_b.

to the drag and lift forces exerted by turbulent flows on the particles. The boundary shear stress adopted in the variable is calculated based on the water depth and the mean particle size of the grain (Duan, 2005). From the scatter plot, the variation of riverbank erosion rate is accounted for by the variation of water depth and the particle size resting on the surface of the bank. Similar trend can be observed for IV5, ratio of critical shear stress to the near-bank velocity. Novel findings from the sensitivity analysis of this study are one of the substantial factors in deriving the most influential parameters which constitutes to the rate of bank erosion.

3.2. Riverbank erosion estimates

The proposed empirical model uses 358 data obtained from erosion pin measurement collected from extensive fieldwork on Sg. Bernam, Selangor, Malaysia. The optimal split proportions approach has been adopted in data splitting. This split sample approach is widely used to study design in high dimensional settings (Dobbin and Simon, 2011; Smith et al., 2019, Saadon et al., 2020). This approach is based on simulations designed to understand qualitatively the relationships among dataset characteristics, ideally data size ranging from 100 to 1,000,000. In this method, the data has been divided into a percentage of training data and testing data. Dobbin and Simon (2011) recommended an allocation between 1/2 and 3/4 of the data for training, and the remaining dedicated for the testing. Therefore, 60% (215 data) of the data were used for training and the remaining 40% (143 data) were used for testing the empirical model development. The optimal proportion was found to depend on the full set of data (n) and the classification accuracy – with higher accuracy and smaller n resulting in more assigned to the training set. The commonly used strategy of allocating 2/3 of the overall data for training was close to optimal for reasonable sized of data (n > 100).

Model development makes use of 215 erosion pin measurements at Sg. Bernam, while the remaining 143 data were used to validate the model. Analyses had used linear, nonlinear and logarithmic transformation for the various combinations of the IVs. Stepwise regression was used to identify the significance of the variables. It is a semi-automated process of building a model by successively adding or removing variables based solely on the t-statistics of their estimated coefficients. The regression analysis yields total of 180 empirical equations in a linear expression, nonlinear (power relationship) expression and log-log function. The analysis included all parameter classes that contribute to the local variability in the bank erosion rates; hydraulic characteristics, bank geometry, soil characteristics, bank resistance, and sediment. The model accuracy is measured using model degree of determination or r-squared, 8 equations represent the best predictor as highlighted in Table 6. Equation 12 using logarithmic-transformation regression depicts the best predictor with degree of determination of 0.783. The influential factor includes hydraulic characteristics of the flow, soil characteristics and bank geometry. The same influential factor can be evidenced in Eq. (11) with r-squared of 0.668 model accuracy. Multiple linear regression analysis showed that the hydraulic characteristics, bank geometry and sediment supply are representation of bank erosion. This includes the shear velocity, height of the bank, bank angle, channel reach, bankfull width, and equilibrium concentration of suspended sediment. All 3 linear equations yield model accuracy greater than 0.5. Power relationship represent the analysis for nonlinear multiple regression include hydraulic characteristics, soil characteristics, bank geometry and bank resistance as the influential factors to bank erosion. Eq. (8) achieved model accuracy at 0.527, Eq. (9) at 0.608 and Eq. (10) at 0.707. Further analysis includes the application of these 8 empirical equations to predict the rate of bank erosion for the remaining 143 numbers of data for Sg. Bernam.

3.3. Best regression models

The performance of the developed prediction model is assessed using Discrepancy Ratio (DR), a ratio between the measured values to the predicted values. These values are deemed accurate if the data lies between 0.5 and 2.0 limits. This approach has been widely used in previous studies e.g. Sinnakaudan et al. (2010), Ibrahim et al. (2017) and Saadon et al. (2020). All 8 best predictors are used to predict the rate of bank erosion using 143 data, initially excluded in the model development phase. These prediction values are compared with the measured values during the fieldwork investigation using erosion pins stated earlier in Section 2.1. Discrepancy ratio is calculated and the percentage of data lies between 0.5 and 2.0 limits is calculated for all 8 empirical models.

The model performance and validation for all 8 empirical model developed using regression analysis indicated that all developed models successfully predicted erosion rates at the percentage of more than 60%. Model validation confirms that developed empirical equations using logarithmic-transformation achieved the highest percentage, where 93.5% (Fig. 6h) and 90.2% (Fig. 6g) of the data lies between the limit of discrepancy ratio. This result synchronized with the result of model accuracy where Eqs. (11) and (12) achieved the highest accuracy at 66.8% and 78.3%, respectively. Model generated using nonlinear regression (power relationship) for Eqs. (8)–(10) (Fig. 6d, e and f) gave a better performance as compared to the model generated using linear function for Eqs. (5)–(7) (Fig. 6a, b and c). This result is favorable since all trends plotted in the sensitivity analysis between independent variables and dependent variable showed a strong positive correlation using power function. The scatter plots of measured data shows a uniform scatter plot which is increasing steadily with positive correlation between the DV and IVs.

Table 6
Results from the multiple regression analysis showing best predicted riverbank erosion rates along Sg. Bernam section.

Regression method	Eq. no.	Empirical model development of bank erosion rates		
		Equation	r-squared	DR (in %)
Multiple linear	5	$\frac{\xi}{u_b} = 1.239 \times 10^{-7} + 3.910 \times 10^{-7}\left(\frac{u_*}{u_b}\right) - 1.254 \times 10^{-11}\left(\frac{h_b}{d_{50}}\right) - 4.090 \times 10^{-10}(\beta) - 1.996 \times 10^{-5}(S_0) + 0.003(C)$	0.506	73.3%
	6	$\frac{\xi}{u_b} = 3.940 \times 10^{-8} + 3.290 \times 10^{-7}\left(\frac{u_*}{u_b}\right) - 3.110 \times 10^{-12}\left(\frac{B}{d_{50}}\right)$	0.707	77.8%
	7	$\frac{\xi}{u_b} = 1.263 \times 10^{-7} + 3.809 \times 10^{-7}\left(\frac{u_*}{u_b}\right) - 3.635 \times 10^{-12}\left(\frac{B}{d_{50}}\right) - 2.966 \times 10^{-10}(\beta) - 2.017 \times 10^{-5}(S_0) + 0.003(C)$	0.705	71.6%
Multiple nonlinear	8	$\frac{\xi}{u_b} = 1.593 \times 10^{-10}\left(\frac{\omega}{u_b}\right)^{12.914}\left(\frac{\tau_c}{\rho_w {u_b}^2}\right)^{-5.907}\left(\frac{B}{d_{50}}\right)^{-1.212}\left(\frac{Y}{d_{50}}\right)^{0.034}$	0.527	81.3%
	9	$\frac{\xi}{u_b} = 1.144 \times 10^{-8}\left(\frac{\tau_c}{\rho_w {u_b}^2}\right)^{0.574}\left(\frac{h_b}{d_{50}}\right)^{-0.102}(S_0)^{-0.808}$	0.608	78.4%
	10	$\frac{\xi}{u_b} = 1.328 \times 10^{-11}\left(\frac{\omega}{u_b}\right)^{113.19}\left(\frac{\tau_c}{\rho_w {u_b}^2}\right)^{-5.064}\left(\frac{h_b}{d_{50}}\right)^{-0.717}$	0.707	83.5%
Logarithmic- transformation	11	$\log\frac{\xi}{u_b} = -10.563\log\left[\left(\frac{\omega}{u_b}\right)^{5.855}\left(\frac{\tau_c}{\rho_w {u_b}^2}\right)^{-2.275}(S_0)^{-0.459}\right]$	0.668	90.2%
	12	$\log\frac{\xi}{u_b} = -8.996\log\left[\left(\frac{\tau_c}{\rho_w {u_b}^2}\right)^{1.325}\left(\frac{u_*}{u_b}\right)^{-0.061}(S_0)^{-1.063}\right]$	0.783	93.5%

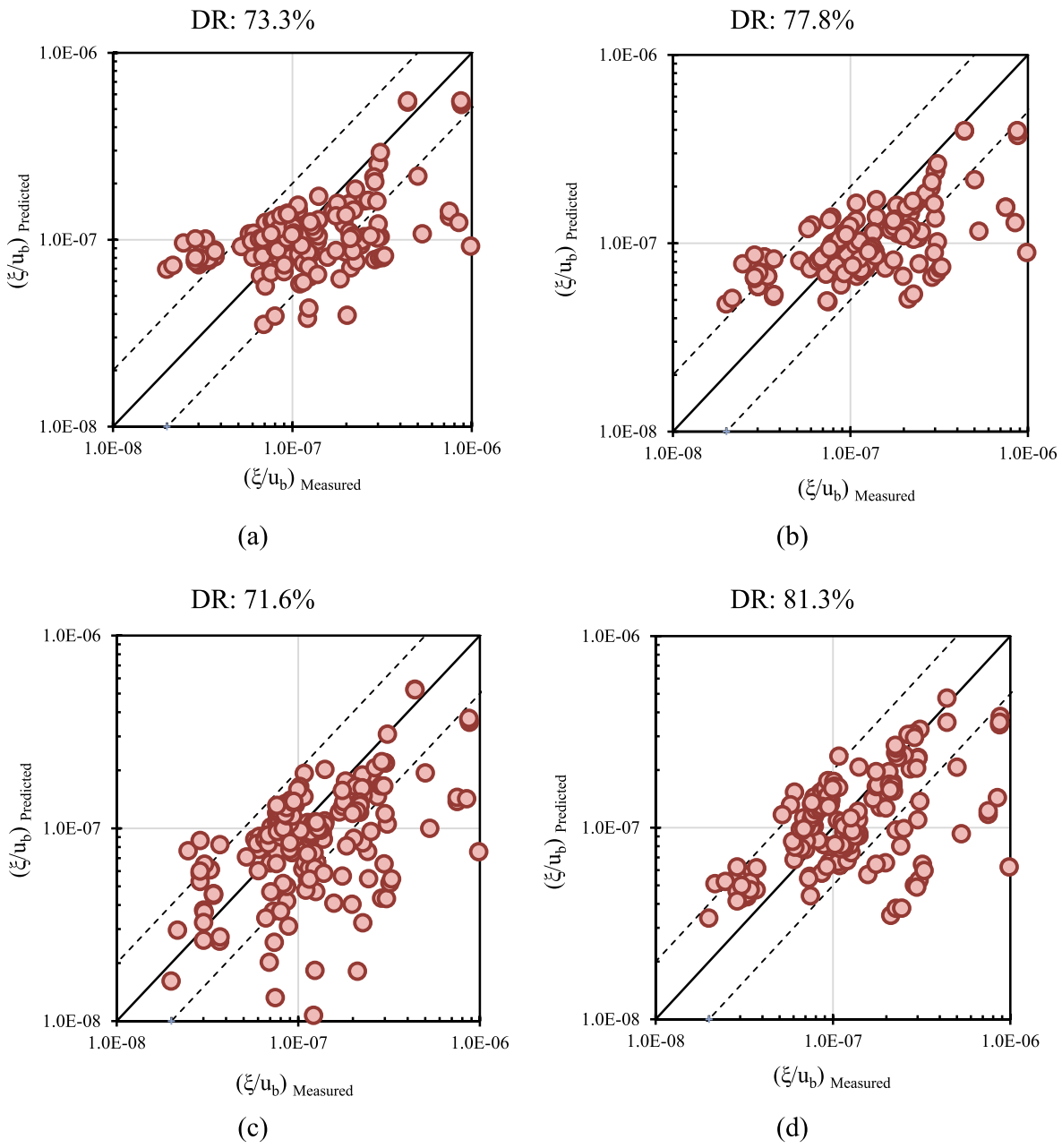

Fig. 6. Predicted erosion rates against measures erosion rate for (a) Eq. (5), (b) Eq. (6), (c) Eq. (7), (d) Eq. (8), (e) Eq. (9), (f) Eq. (10), (g) Eq. (11), (h) Eq. (12).

The graphical analysis of model performance and validation confirms that Eqs. (11) and (12) gives better trends than other equations, in which all data gathered coincides within the discrepancy limits. Eq. (12), however, showed more favorable trends along the perfect line of agreement, compared to Eq. (11). Table 6 (5th column) shows the summary of predictive models developed using multiple linear, multiple nonlinear, and logarithmic-transformation regression analysis.

4. Conclusions

Based on field measurements from River Bernam, this study has developed predictive models for the estimation of riverbank erosion rates using empirical regression equations. All developed equations yield model accuracy greater than 0.5. The important conclusions from this study are:

a) Parameters of the ratio of bankfull width to the mean particle diameter, bank angle, and channel reach yield very weak correlation with bank erosion rates.
b) The multiple linear regression analysis shows that the hydraulic characteristics, bank geometry and sediment supply significantly represent bank erosion. This includes the shear velocity, height of the bank, bank angle, channel reach, bankfull width, and equilibrium concentration of suspended sediment.
c) The nonlinear multiple regression shows that hydraulic characteristics, soil characteristics, bank geometry and bank resistance are the influential factors to bank erosion. Eq. (8) achieved a model accuracy at 0.527, Eq. (9) at 0.608 and Eq. (10) at 0.707.
d) After log-transformation, Eq. (12) gives the best predictor with a degree of determination of 0.783. The influential factor includes hydraulic characteristics of the flow, soil characteristics and bank geometry. The same influential factor can be evidenced in Eq. (11) with r-squared of 0.668 model accuracy.
e) Model validation confirms that the developed empirical equations

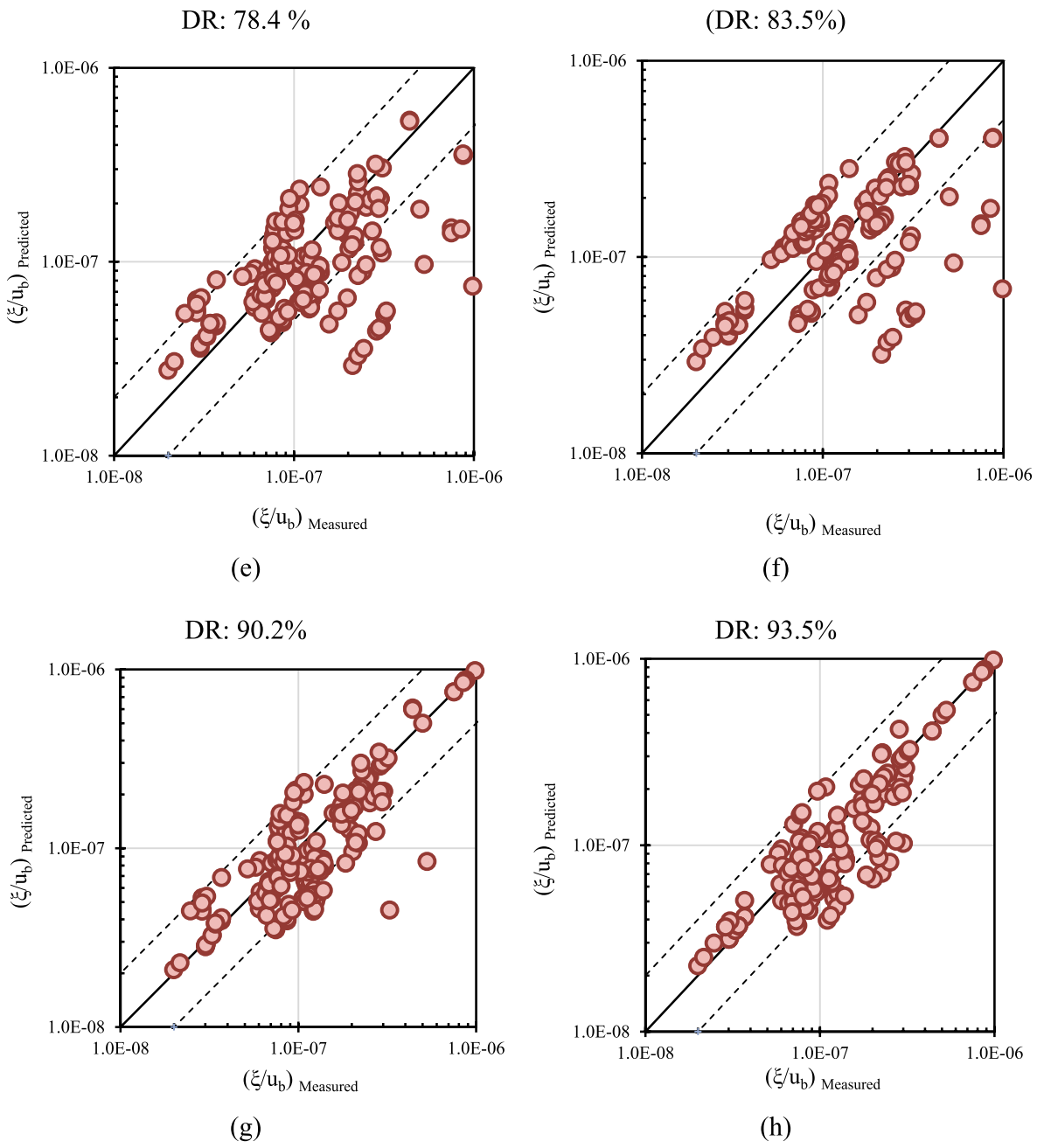

Fig. 6. (*continued*)

using logarithmic-transformation, namely Eqs. (11) and (12) give the highest percentage of accuracy, i.e. 90.2% and 93.5%, respectively. This is consistent with the model development results. Additionally, the graphical analysis of model performance and validation confirms that Eqs. (11) and (12) give better trends than other equations, where data are between the limit of discrepancy ratio $0.5 < DR < 2$.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors appreciate the support from the Ministry of Science, Technology and Innovation (MOSTI), through Science Fund (Project Number: 06-01-01-SF0773). Additional funding to the second author, provided by the Ministry of Education, through the Fundamental Research Grant Scheme (Grant No. FRGS/1/2018/TK10/UITM/02/5) is also acknowledged.

References

Azamathulla, H., Ghani, A.B., Zakaria, N.Z., Guven, A., 2010. Genetic programing to predict bridge pier scour. J. Hydraul. Eng. 136, 165–169.

Barman, K., Roy, S., Das, V.K., Debnath, K., 2019. Effect of clay fraction on turbulence characteristics of flow near an eroded bank. J. Hydrol. 571, 87–102.

Chavan, R., Kumar, B., 2018. Prediction of scour depth and dune morphology around circular bridge piers in seepage affected alluvial channels, Environ Fluid Mech, Springer.

Bernhardt, E.S, Palmer, M.A, Allan, J.D, Alexander, G., Barnas, K., Brooks, S., Carr, J., Clayton, S., Dahm, C., Follstad-Shah, J., Galat, D., Gloss, S., Goodwin, P., Hart, D., Hassett, B., Jenkinson, R., Katz, S., Kondolf, G.M., Lake, P.S., Lave, R., Meyer, J.L., O'Donnell, T.K., Pagano, L., Powell, B., Sudduth, E., 2005. Synthesizing U.S. River Restoration Efforts. Science 308 (5722), 636–637.

Constantine, C.R., Dunne, T., Hanson, G.J., 2009. Examining the physical meaning of the bank erosion coefficient used in meander migration modeling. Geomorphology 106 (3–4), 242–252.

Drainage and Irrigation Department, 1976. Hydrological Procedure no. 15, River discharge measurement by current meter, Drainage and Irrigation Department, Ministry of Agriculture, Malaysia.

Dobbin, K., Simon, R., 2011. Optimally splitting cases for training and testing high dimensional classifiers. BMC Med. Genomics 4 (31).

Duan, J.G., 2005. Analytical approach to calculate rate of bank erosion. J. Hydraul. Eng. 131 (11), 980–990.

Duan, J.G., Julien, P.Y., 2010. Numerical simulation of meandering evolution. J. Hydrol. 391 (1–2), 34–46.

Engel, F.R., Rhoads, B.L., 2017. Velocity profiles and the structures of turbulence at the outer bank of a compound meander bend. Geomorphology 295, 191–201.

Hasegawa, K., 1989. Universal bank erosion coefficient for meandering rivers. J. Hydraulic Eng., ASCE 115 (6), 744–765.

Hickin, E.J., Nanson, G.C., 1975. The character of channel migration on the Beatton River, northeast British Columbia, Canada. Geol. Soc. Am. Bull. 86, 487–494.

Hickin, E.J., Nanson, G.C., 1984. Lateral migration rates of river bends. J. Hydraul. Eng. 110, 1557–1567.

Ibrahim, S., Abdullah, J., Hashim, K., Ariffin, J, 2017. Establishment of Jet Index Ji For Soil Erodibility Coefficients Using Jet Erosion Device (Jed), International J. GEOMATE 12(34), 152-157.

Ikeda, S., Parker, G., Sawai, K., 1981. Bend theory of river meanders; part i, linear development. J. Fluid Mech 112, 363–377.

Julien, P.Y., 2012. Erosion and sedimentation, second ed. Cambridge University Press.

Kuntjoro, Harijanto, D., 2018. The movement of the regularly river meanders on constant discharge. Int. J. Civ. Eng. Technol. (IJCIET) 9, 619–629.

Lawler, D.M., 1993. The measurement of riverbank erosion and lateral channel change: a review. Earth Surf. Proc. Land. 18, 777–821.

Nanson, G.C., Hickin, E.J., 1983. Channel migration end incision on the Beatton River. ASCE, J. Hydraulic Eng. 109, 327–337.

Posner, A.J., Duan, J.G., 2012. Simulating river meandering processes using stochastic bank erosion coefficient. Geomorphology 163–164, 26–36.

Randle, T.J., 2004. Channel migration model for meandering rivers, A thesis submitted to the University of Colorado at Denver in partial fulfilment of the requirements for the degree of Master of Science, Civil Engineering.

Randle, T.J., 2006. Channel migration model for meandering rivers. In: Proceedings of the 8th Federal Interagency Sedimentation Conference, pp. 241–248.

Richard, G.A., Julien, P.Y., Baird, D.C., 2005. Statistical analysis of lateral migration of the Rio Grande, New Mexico. Geomorphology 71 (2005), 139–155.

Roy, S., Das, V.K., Debnath, K., 2019. Characteristics of intermittent turbulent structures for river bank undercut depth increment. Catena 172 (2019), 356–368.

Saadon, A., Abdullah, J., Mohamed, S., Ariffin, J. 2020, Development of river bank erosion rate predictor for natural channels using NARX-QR Factorization model, for Neural Computing and Application, ISSN 0941-0643, https://doi.org/10.1007/s00521-020-04835-5.

Saadon, A., Abdullah, J., Ariffin, J., 2016. Dimensional analysis relationships of streambank erosion rates, Jurnal Teknologi (Sciences and Engineering), 78:5-5 pp. 79-85.

Sinnakaudan, S.K., Sulaiman, M.S., Teoh, S.H., 2010. Total bed material load equation for high gradient rivers. J. Hydro-environ. Res. 4 (2010), 243–251.

Smith, H.G., Spiekermann, R., Dymond, J., Basher, L., 2019. Predicting spatial patterns in riverbank erosion for catchment sediment budgets. N. Z. J. Mar. Freshwater Res. https://doi.org/10.1080/00288330.2018.1561475.

Toriman, M.E., Jaafar, O., Idris, M., Mastura, S.S.A., Juahir, H., Aziz, N.A., Kamarudin, K.A., Jamil, N.R., 2010. Study of water level-discharge relationship using artificial neural network (ANN) in Sg, Gumum, Tasik Chini Pahang, Malaysia. Res. J. Appl. Sci. 5 (1), 20–26.

Varouchakis, E.A., Giannakis, G.V., Lilli, M.A., Ioannidao, E., Nikolaidis, N.P., Karatzas, G.P., 2016. Development of a statistical took for the estimation of riverbank erosion probability. SOIL J. 2, 1–11.