IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

# Improving total sediment load prediction using genetic programming technique (Case Study: Malaysia)

IOP Conf. Series: Materials Science and Engineering **736** (2020) 022108 doi:10.1088/1757-899X/736/2/022108

# Improving total sediment load prediction using genetic programming technique (Case Study: Malaysia)

**N A Ahmad Abdul Ghani[1*] , N A Kamal[2] and J Ariffin[2]**

[1]Faculty of Civil Engineering Technology, Universiti Malaysia Pahang, 26300 Gambang, Pahang, Malaysia
[2]Faculty of Civil Engineering, Universiti Teknologi Mara (Shah Alam), 40450 Shah Alam, Selangor, Malaysia

*nadiatul@ump.edu.my

**Abstract**. Predicted total sediment load is usually used to identify the intensity of a sedimentation process. Currently, the existing available models to predict total load are mostly developed based on data collected from flumes, channels and rivers located in western countries. These models known as sediment transport model may not be valid to predict total sediment load of rivers in the tropics due to significant differences in the hydrological and sediment characteristics conditions. A new technique called Genetic programming (GP) technique is used to develop a new model to improve the prediction of total sediment load for rivers in tropical Malaysia. The new model named Evolutionary Polynomial Regression (EPR) model is compared with other three available sediment transport models using the different techniques including, Regression Equation, Modified Graf and Multiple Regression. Statistical analyses based on 82 data sets show the sediment transport model using this new technique perform well compare to other models.

## 1. Introduction

The total load or total sediment load is defined as the combination of wash and bed material load. Under conditions when wash load is not present, the term - bed material load and total load - are used interchangeably. The total sediment load process in a river varies across its length and cross-section. The variation depends on the interaction of the hydraulics and sediment variables. The reliable estimation of total sediment load in the river can assist in the estimation of deposition which helps to predict the water surface elevations during floods.

Currently, there are a few models that are used to estimate total sediment load, including Engelund & Hansen [1], Graf [2], Ackers & White [3], Yang & Molinas [4], Van Rijn [5], Karim [6] and Nagy et al. [7]. However, most of these models are developed based on flume data from western countries, such as America and Western Europe and have not been widely used in other parts of the world [8]. Since the 1990s, some researchers have developed models based on Malaysian conditions [8,9,10] using the different various technique. Those models failed to achieve consistent success with the accuracy of predicting sediment loads and thus there is a need to using a new technique for a more accurate sediment model.

In this paper, a new sediment transport model named Evolutionary Polynomial Regression (EPR) model was developed exclusively for rivers with upstream in-stream mining activities. This new model was compared with three other models that used other technique; that include Regression Equation, Modified Graf and Multiple Regression.

An Evolutionary Polynomial Regression (EPR) model [11] is developed based on a set of 256 recorded data of total sediment load. EPR is a model developed by Giustolisi and Savic [12] that constructs symbolic models by integrating the best features of numerical regression [13] with genetic programming and symbolic regression [14]. EPR can be classified as a "grey box" technique that details out the conceptualization of physical phenomena [12]. Figure 1 shows a pictorial representation of this classification, where greater physical knowledge used during the development of the model would lead to a better physical interpretation of
the phenomena by the user.

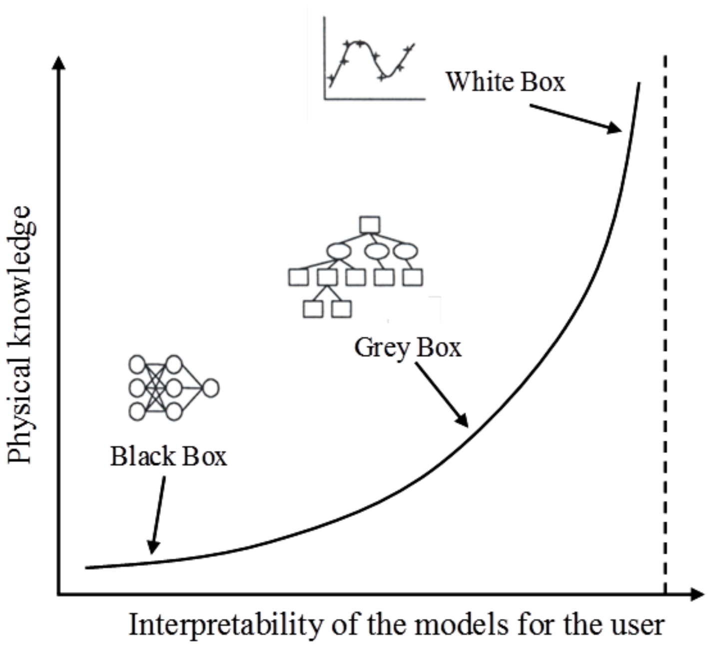


**Figure 1.** Graphical classification of EPR among modelling techniques [15].

EPR had been used successfully to solve problems in civil engineering, for instance, pipeline system [16,17], groundwater study [15], geotechnical and soil performance [18,19]. A detailed description of the technique is found in [11 and 12].

## 2. Model development

The first important step in the development of the EPR model is to identify the potential model inputs and outputs. Based on previous studies [20], seven input variables were selected in five parameter class are considered to be the most significant factors affecting sediment transport. The variables are, relative roughness on the bed *($R/d_{50}$)* inflow resistance parameter class, stream width ratio *($B/y_o$)*, in conveyance and shape class which are shear velocity ratio to fall velocity *($U^*/w_s$)*, fall velocity to shear velocity *($w_s/U^*$)* in sediment properties class is ratio of shear stress to average velocity *($U^*/V$)* and dimensionless unit stream power ($VS_o/ w_s$) in mobility class and the last variable is velocity head *($v^2/2g$)*. The output variable is the total sediment load *($Qt/Q$)*.

The initial step in developing the EPR model is selecting the related parameters to develop the model. This is carried out by a trial-and-error approach in which a number of EPR models are trained with the selected parameters until ultimately the optimum model is obtained. The data are randomly divided into two sets, i.e. a training set for model calibration and an independent validation set for model verification. In dividing the data into their sets, the training and validation sets are selected such that they are statistically consistent to represent the same statistical population [21]. Hence, out of 256 data cases, 174 cases (68%) are used for training and 82 cases (32%) for validation. Table 1 shows the range of the data used in model development.

**Table 1.** Range of data used in model development.

| Parameter | Range |
|---|---|
| Total sediment load, $T_j$ (kg/s) | 0.2846-44.0144 |
| Flow, $Q$ (m$^3$/s) | 0.737-87.792 |
| Velocity, $V$ (m/s) | 0.194-1.26 |
| Depth of water, $y_o$ (m) | 0.23-3.23 |
| Particle mean size, $d_{50}$ (m) | 0.0004-0.004 |
| Water surface slope, $S_o$ | 0.0003-0.0167 |
| Fall velocity, $W_s$ (m/s) | 0.051-14.821 |
| Hydraulic radius, $R$ (m) | 0.22-2.66 |

*2.1. Performance indicator*

The best model can be obtained by modeling the information in the data rather than modeling the data itself [22]. Data contains both noise and information. Over-fitting is a poor strategy and under-fitting also means getting a poor model that will not give sufficient information. A model that has a good balance between over-fitting and the under-fitting is required. The discrepancy ratio is the ratio between the predicted and measured sediment total load, and a model is considered to be suitable if its discrepancy ratio falls within the range of 0.5-2.0 [8].

*2.2. Model accuracy*

To examine its accuracy, the EPR model predictions are compared with those obtained from three available sediment transport models Regression Equation [9], Modified Graf [10] and Multiple Regression [8]. Summary of sediment parameters used for comparison is given in table 2.

**Table 2.** Summary of sediment parameters used in the available models.

| Model | Input parameters used |
|---|---|
| Regression Equation | $R/d_{50}, U^*/\omega_s, U^*/V, V^2/gy_o$ |
| Modified Graf | $(S_s-1)d_{50}/RS_o, C_vVR/\sqrt{g(S_s-1)d_{50}}$ |
| Multiple Regression | $VS_o/\omega_s, R/d_{50}, \sqrt{g(S_s-1)d_{50}^{\,3}}/VR$ |
| Genetic programming | $u*/V$, $R/d_{50}$ |

where; $\gamma_s$ = unit weight of sediment; $V$ is flow velocity; $d_{50}$ *is* median diameter of sediment load; $g$ =acceleration of gravity; $\gamma_w$ = unit weight of water; $\tau$ = mean bed shear stress; $S_s$ = specific gravity of sediment; $R$ is hydraulic radius; $C_v$ = volumetric sediment concentration; $U^*$ = shear velocity and $\omega_s$ = fall velocity.

## 3. Results and discussion

A statistical analysis carried out on 82 cases of the validation set for discrepancy ratio. The results of the comparison are given in table 3 and shown graphically in figures 2 - 5.

**Table 3**. Performance of the Total Sediment Load Model.

| Model | Performance measure |
| --- | --- |
| | Discrepancy ratio (%) |
| Regression Equation [9] | 70 |
| Modified Graf [10] | 23 |
| Multiple Regression [8] | 90 |
| Genetic Programming (GP) | 100 |

Table 3 shows that the EPR model using Genetic Programming technique outperforms the other methods in all the performance measures used. This model gives 100 % accuracy in differential ratio measurement. The graphical result in Figure 5 shows that all the predicted and measured sediment total loads are between the equality lines. The model developed using Multiple Regression Technique [8] can be considered to be of second best. The graphical results in Figure 4 also indicate that it has the least scattering around the line of equality between the predicted and measured sediment total loads.

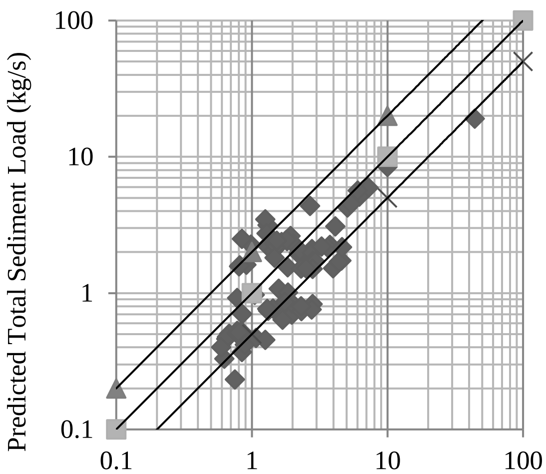


**Figure 2.** Performance of Regression Equation.

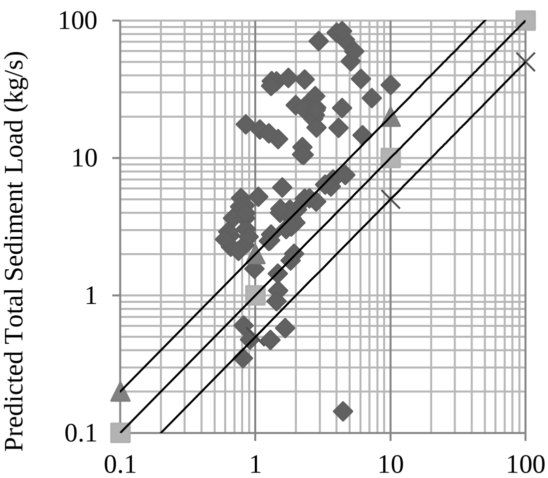


**Figure 3.** Performance of Modified Graf.

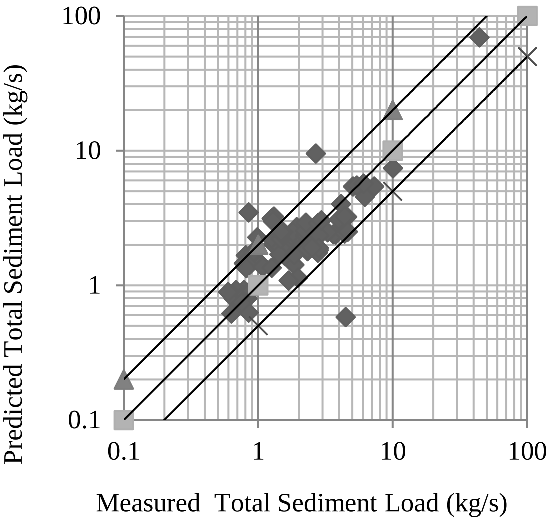


**Figure 4.** Performance of Multiple Regression.

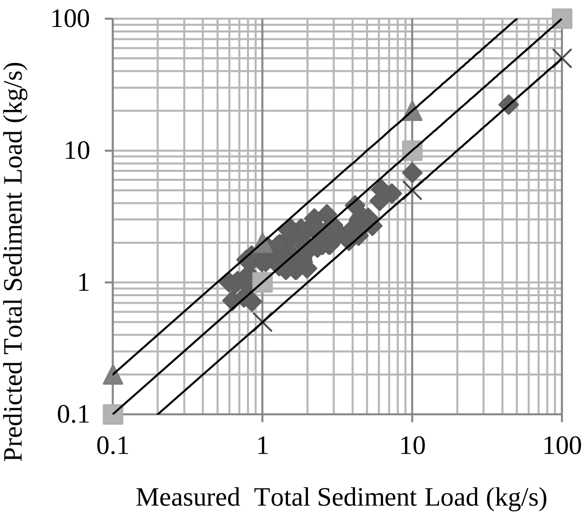


**Figure 5.** Performance of Genetic Programming.

## 4. Conclusion
The performance of the EPR model, to the testing set, shows less scattering around the line of equality between the measured and the predicted total sediment loads. For the EPR model, the discrepancy ratios are 100% respectively. Whereas the results values for other model is 70 % respectively for Regression Equation, 23% respectively for Modified Graf, and 90% respectively for Multiple Regression. These results indicate that the developed EPR model using Genetic Programming technique outperforms the other available methods. This is because the parameters used in the EPR model are based on rivers in Malaysia that appropriately fit the model characteristics and the sediment development process.

## Acknowledgement
The authors acknowledge the Department of Irrigation and Drainage (DID) Malaysia, the Ministry of Natural Resources & Environment, Malaysia for providing the data used in this study.

## References
[1] Engelund F and Hansen E A 1967 Monograph on Sediment Transport in Alluvial Streams. (Copenhagen, Teknisk Forlag: Denmark)
[2] Graf W H 1971 *Hydraulics of Sediment Transport* (New York: McGraw Hill)
[3] Ackers P and White W R 1973 *ASCE* **99** 2041–2060
[4] Yang C T and Molinas A 1982 *ASCE* **108** 774–793
[5] Van Rijin L C 1984 *J. Hydraul.Eng.* **110** 1431–1456
[6] Karim F 1998 *J. Hydraul. Eng.* **124** 597–604
[7] Nagy H M, Watanabe K and Hirano M 2002 *ASCE* **128** 588–595
[8] Sinnakaudan S K, Ab. Ghani A, Ahmad M S and Zakaria N A 2006 *J. Hydraul. Eng.* **132** 521–528
[9] Ariffin J 2004 Development of Sediment Transport Models for Rivers in Malaysia Using Regression Analysis and Artificial Neural Networks, PhD Thesis, Department of Civil Engineering, Universiti Sains Malaysia, Malaysia

IOP Conf. Series: Materials Science and Engineering **736** (2020) 022108 doi:10.1088/1757-899X/736/2/022108


[10] Chan C K, Ab. Ghani A, Zakaria N A, Abu Hasan Z and Abdullah R 2005 *Int. J. River Basin Manag.* **3** 203–208

[11] Abdul Ghani N A, Tolibon D A and Ariffin J 2019 *ASM Sci. J.* **12**

[12] Giustolisi O and Savic D A 2006 *J. Hydroinform.* **8** 207–222

[13] Draper N R and Smith H 1998 *Applied Regression Analysis* (New York: John Wiley and Sons)

[14] Koza J R 1992 *Genetic Programming: On the Programming of Computers by Means of Natural Selection* (MIT Press: Cambridge, MA)

[15] Giustolisi O, Doglioni A, Savic D A and Pierro F 2008 *Water Resour. Res.* **44**

[16] Berardi L, Giustolisi O, Kapelan Z and Savic D A 2008 *J. Hydroinform.* 10 113-126

[17] Kaydani H, Najafzadeh M Mohebbi A 2014 *J. Pipeline Syst. Eng.* **5**

[18] Yin Z Y, Jin Y F, Huang H W and Shen S L 2016 *Eng. Geol.* **210** 158-167

[19] Ghorbani A and Hasanzadehshooiili H 2018 *Soils Found.* **58** 34-49

[20] Ariffin J 2019 *Fluvial Geomorphology Anthropogenic Agents of Change and Their Implications* (Penerbit Press: Universiti Teknologi Mara)

[21] Shahin M A, Maier H R and Jaksa M B 2004 *J. Comp. Civil Eng.* **18** 105–114

[22] Shaqlaih A, White L and Zaman M 2011 Resilient modulus modeling with information theory approach *Int. J. Geomech.*