Use of Evolutionary Polynomial Regression (EPR) for Prediction of Total Sediment Load of Malaysian Rivers

Nadiatul A. Abdul Ghani nadiatuladilah@yahoo.com
Faculty of Civil Engineering & Earth Resources,
University Malaysia Pahang,
Lebuhraya Tun Razak, Gambang,
Kuantan,26300 Pahang, Malaysia

Assc. Prof. Mohamed A. Shahin m.shahin@curtin.edu.au
Department of Civil Engineering,
Curtin University,
GPO Box U1987 Perth,
Western Australia 6845,Australia

Professor Hamid R. Nikraz h.nikraz@curtin.edu.au
Department of Civil Engineering,
Curtin University,
GPO Box U1987 Perth,
Western Australia 6845,Australia

Abstract

This study investigates the use of Evolutionary Polynomial Regression (EPR) for predicting the total sediment load of Malaysian rivers. EPR is a data-driven modelling hybrid technique, based on evolutionary computing, that has been recently used successfully in solving many problems in civil engineering. In order to apply the method for modelling the total sediment of Malaysian rivers, an extensive database obtained from the Department of Irrigation and Drainage (DID), Ministry of Natural Resources & Environment, Malaysia was sought, and unrestricted access was granted. A robustness study was performed in order to confirm the generalisation ability of the developed EPR model, and a sensitivity analysis was also conducted to determine the relative importance of model inputs. The results obtained from the EPR model were compared with those obtained from six other available sediment load prediction models. The performance of the EPR model demonstrates its predictive capability and generalisation ability to solve highly nonlinear problems of river engineering applications, such as sediment. Moreover, the EPR model produced reasonably improved results compared to those obtained from the other available sediment load methods.

Keywords: Evolutionary polynomial regression, sediment, rivers, Malaysia, prediction.

1. INTRODUCTION

Sedimentation is a process that changes the rivers shape and embankments in the form of altering the cross-section, longitudinal profile, course of flow and patterns of rivers. In order to sustain the cultural and economic developments along alluvial rivers, the principles of sediment transport should be carefully studied and solutions for its engineering and environmental problems need to be developed. Currently, there are a few models that can be used to identify the sedimentation process in the form of estimating the total sediment load. Some of the available models include Engelund & Hansen [1], Graf [2], Ackers & White [3], Yang & Molinas [4], Van Rijn [5], Karim [6] and Nagy et al. [7], among others. However, most of these models have been developed based on flume data from western countries, including America and Western Europe, and have not been widely used or evaluated in other parts of the world [8]. Since the 1990's, some Malaysian researchers have developed models based on the Malaysian conditions (e.g. [8]; [9]; [10]). However, these models failed to achieve consistent success in relation to accurate sediment prediction; hence, there is a need for more accurate sediment models.

In this paper, Evolutionary Polynomial Regression (EPR) was used to develop a more accurate model for predicting the total sediment load for rivers in Malaysia. EPR is an artificial intelligence technique that has the advantage of combining the genetic algorithms with traditional numerical regression [12]. The data used for model calibration and validation were collected from the Department of Irrigation and Drainage (DID), Ministry of Natural Resources & Environment, Malaysia (hereinafter referred to as the DID). The database comprises 338 data cases (from 1998 through to 2007) that represent ten different rivers across Malaysia for four river catchment areas, namely Kinta, Kerayong, Langat and Kulim (Figure 1). The first set of data was collected for Pari River in Taman Merdeka and Kerayong River in Kuala Lumpur from 1998 to 1999. The second set of data was undertaken at the Kinta River catchment, which consists of four rivers including Kinta River, Raia River, Pari River and Kampar River. The third set of data took place over the period 2000 to 2002, at the Langat River catchment area, comprising Langat River, Lui River and Semenyih River. The fourth and final set of data was completed at Kulim River in 2007.

The available data were divided into two sets: a training set for model calibration and an independent validation set for model verification. In order to test the performance of the developed model, consideration was given not only to the model predictive statistical accuracy in the training and validation set but also to the robustness and interpretive ability of the model. This was carried out by performing a parametric study to investigate the generalization ability (robustness) of the model and a sensitivity analysis to quantify the relative importance of the model inputs to the corresponding outputs (i.e. interpretive ability). Predictions from the developed EPR model were compared with those obtained from six other available models.

FIGURE 1: Map of river catchments of the study area. [13]

2. OVERVIEW OF EVOLUTIONARY POLYNOMIAL REGRESSION (EPR)

EPR is a data-driven hybrid regression technique, based on evolutionary computing, that was developed by Giustolisi and Savic [14]. EPR has been used successfully in solving several problems in civil engineering (e.g. [15]; [16]; [17]). It constructs symbolic models by integrating the soundest features of numerical regression [18] with genetic programming and symbolic regression [19]. This strategy provides the information in symbolic form expressions, as usually defined and referred to in the mathematical literature [20]. The following two steps roughly describe the underlying features of EPR, aimed to search for polynomial structures representing a system. In the first step, the selection of exponents for polynomial expressions is carried out, employing an evolutionary searching strategy by means of genetic algorithms [21]. In the second step, numerical regression using the least square method is conducted, aiming to compute the coefficients of the previously selected polynomial terms. The general form of expression in EPR can be presented as follows [14]:

$$y = \sum_{j=1}^{m} F(X, f(X), a_j) + a_o \tag{1}$$

where: y is the estimated vector of output of the process; m is the number of terms of the target expression; F is a function constructed by the process; X is the matrix of input variables; f is a function defined by the user; and a_j is a constant. A typical example of EPR pseudo-polynomial expression that belongs to the class of Eq. (1) is as follows [14]:

$$\hat{Y} = a_o + \sum_{j=i}^{m} a_j \ .(X_1)^{ES(j,1)} \text{..........} (X_k)^{ES(j,k)} .f\left[(X_1)^{ES(j,k+1)} \text{......} (X_k)^{ES(j,2k)}\right] \tag{2}$$

where: $\hat{Y}$ is the vector of target values; m is the length of the expression; a_j is the value of the constants; X_i is the vector(s) of the k candidate inputs; ES is the matrix of exponents; and f is a function selected by the user.

EPR is suitable for modelling physical phenomena, based on two features [15]: (i) the introduction of prior knowledge about the physical system/process – to be modelled at three different times, namely: before, during and after EPR modelling calibration; and (ii) the production of symbolic formulae, enabling data mining to discover patterns which describe the desired parameters. In the first EPR feature (i) above, before the construction of the EPR model, the modeller selects the relevant inputs and arranges them in a suitable format according to their physical meaning. During the EPR model construction, model structures are determined by following user-defined settings such as general polynomial structure, user-defined function types (e.g. natural logarithms, exponentials, tangential hyperbolics) and searching strategy parameters. The EPR starts from true polynomials and also allows for the development of non-polynomial expressions containing user-defined functions (e.g. natural logarithms). After EPR model calibration, an optimum model can be selected from among the series of returned models. The optimum model is selected based on the modeller's judgement, in addition to statistical performance indicators such as the coefficient of determination (CoD). A typical flow diagram of the EPR procedure is shown in Figure 2, and detailed description of the technique can be found in [14].

The EPR symbolic approach can be seen as opposite to those numerical regressions performed in Artificial Neural Networks. According to the classification of modelling techniques based on colour, whereby meaning is related to three levels of prior information required [22], EPR can be classified as a "grey box" technique (conceptualisation of physical phenomena), and Figure 3 shows a pictorial representation of this classification where the greater the physical knowledge used during the development of the model, the better the physical interpretation of the phenomena by the user. EPR is a technqique based on observed data; however, the mathematical structure it returns is symbolic and usually uncomplicated in its constitution [14].

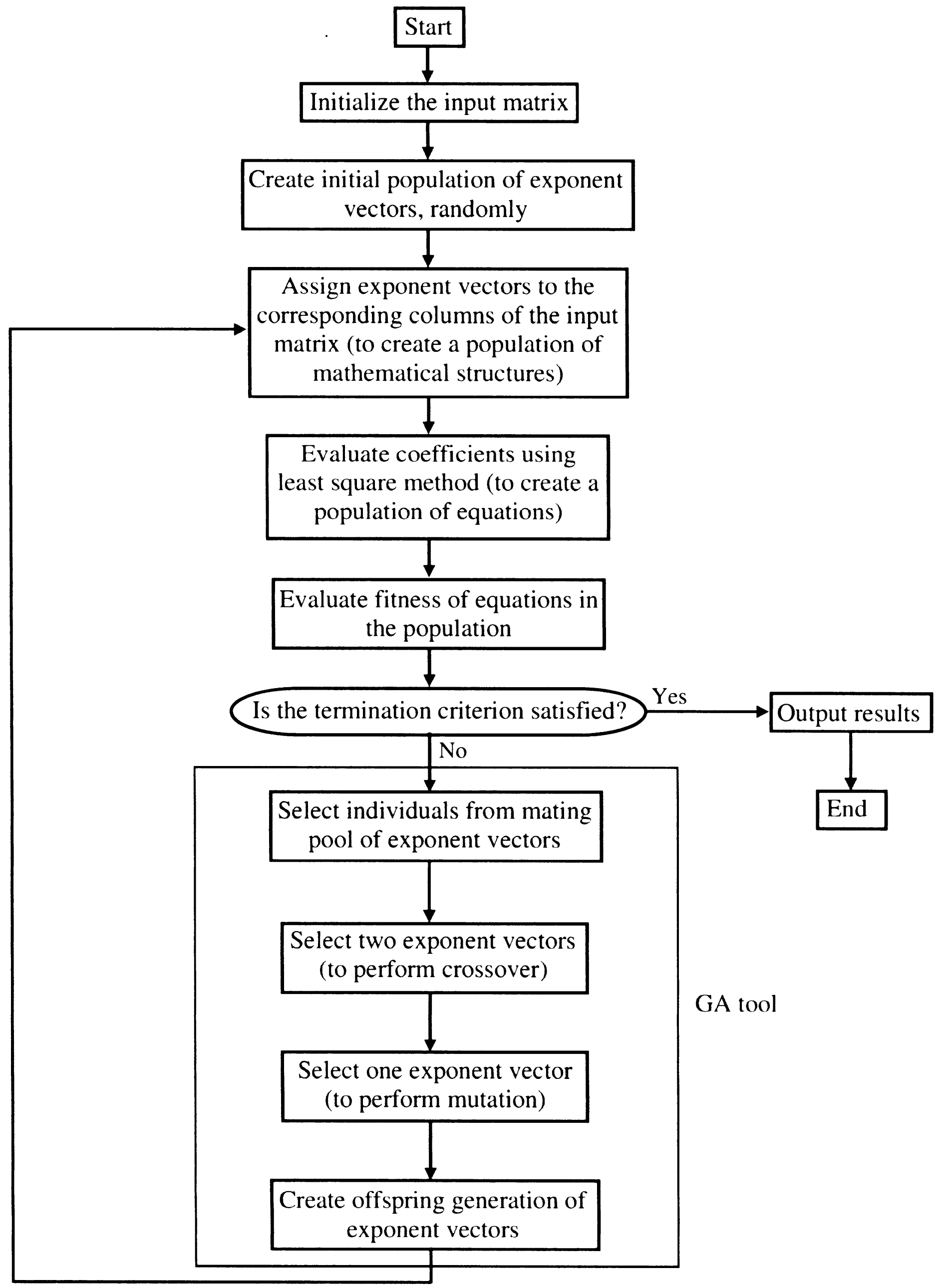


FIGURE 2: Typical flow diagram of EPR procedure. [31]

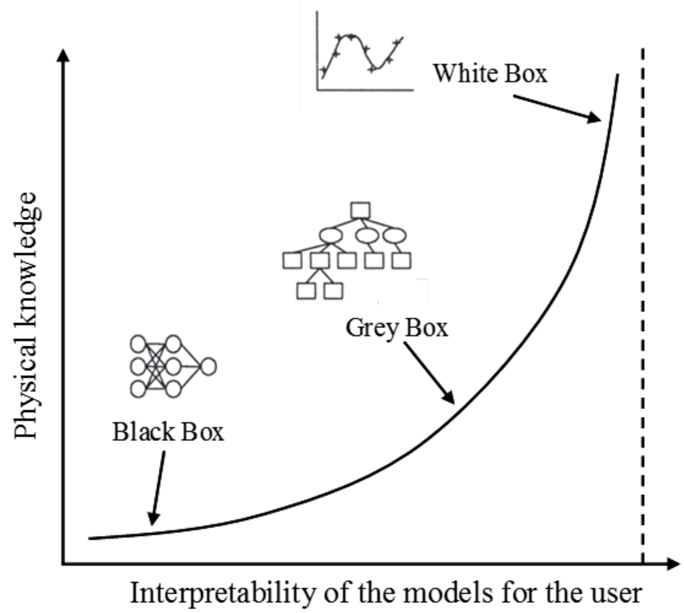


FIGURE 3: Graphical classification of EPR among modelling techniques. [17]

3. DEVELOPMENT OF SEDIMENT TRANSPORT MODEL USING EPR

In this study, the EPR model was developed based on a set of 338 data records collected from the DID, containing information on total sediment load. The collected data represent the sediment transport features of ten different rivers across Malaysia, as mentioned earlier. In modeling environmental phenomena, such as sediment, care has to be given to the data used. Incomplete sampled data always exist and analysis should provide new insights into the phenomena, give accurate forecasting of the output for a range of inputs. Another additional problem when dealing with environmental data is related to discontinuities, i.e. gaps often present in the data records, and reconstructing the information contained in the missing data, without influencing the construction of models, is needed [11]. The EPR model was developed using the available software package, EPR Toolbox Version 2 [23].

The first important step in the development of the EPR model was to identify the potential model inputs and corresponding outputs. Based on previous studies carried out by many researchers (e.g. [8]), for the purpose of this study, eight inputs were utilised, having deemed them to be the most significant factors affecting the sediment transport. These inputs include the hydraulic radius (R), flow depth (Y_o), flow velocity (V), median diameter of sediment load (d_{50}), stream width (B), water surface slope (S_o), fall velocity (ω_s) and flow discharge (Q). The only output is the total sediment load (T_j).

The next step taken in the development of the EPR model was the data division. In this study, the data were randomly divided into two sets: a training set for model calibration and an independent validation set for model verification. In dividing the data into their sets, the training and testing sets were selected to be statistically consistent, thus, represent the same statistical population, as recommended by Shahin et al. [24]. In total, 271 data cases (80%) of the available 338 data cases were used for training, and 67 data cases (20%) were used for validation. The statistics of the data cases used for the training and validation sets are given in Table 1, including the mean, standard deviation, minimum, maximum and range. It should be noted that the extreme values of the data cases were included in the training set.

Model variables & data sets	Statistical parameters				
	Mean	Standard Deviation	Minimum	Maximum	Range
Flow discharge, Q (m^3/s)					
Training set Testing set	7.28 7.96	6.62 7.28	0.74 1.19	47.90 35.91	47.16 34.72
Flow depth, y_o (m)					
Training set Testing set	0.57 0.60	0.27 0.30	0.22 0.24	1.87 1.61	1.65 1.37
Flow velocity, V (m/s)					
Training set Testing set	0.62 0.64	0.20 0.19	0.19 0.26	1.26 1.10	1.07 0.84
Median diameter of bed material, d_{50}					
Training set Testing set	0.0014 0.0016	0.0008 0.0010	0.0004 0.0005	0.0040 0.0039	0.0036 0.0034
Hydraulic radius, R (m)					
Training set Testing set	0.54 0.56	0.24 0.25	0.21 0.23	1.77 1.39	1.56 1.16
Stream width, B (m)					
Training set Testing set	17.85 17.92	3.70 3.89	13.50 13.80	28.00 28.00	14.50 14.20
Bed slope, S_o (m)					
Training set Testing set	0.0034 0.0033	0.0027 0.0027	0.0003 0.0010	0.01 0.01	0.01 0.01
Fall velocity, ω_s (m^2/s)					
Training set Testing set	0.22 0.23	0.29 0.26	0.04 0.06	1.74 1.34	1.69 1.28
Total Load, Tj (kg/s)					
Training set Testing set	2.76 3.08	3.57 3.62	0.11 0.18	28.52 17.85	28.41 17.66

TABLE 1: EPR input and output variables used and their statistics.

The following step in the development of the EPR model was selecting the related internal parameters for evolving the model. This was carried out by a trial-and-error approach in which a number of EPR models were trained, using the parameters given in Table 2, until the optimum model was obtained. A more detailed description of the modelling parameters used in Table 2 can be found in the EPR Toolbox manual [23].

Parameter	EPR setting
Regression type	Statistical
Polynomial structure	$Y = sum(a_i \times X_1 \times X_2 \times f(X_1) \times f(X_2)) + a_o$
Function type	Exponent
Term	[1:5]
Range of exponents	[0, 0.5, 1, 2]
Generation	10
Offset (a_o)	Yes
Constant estimation method	Least Square

TABLE 2: Internal parameters used in the EPR modeling.

3.1 Performance indicators

As mentioned earlier, the optimum EPR model was obtained by a trial-and-error approach in which a number of EPR models were trained with different internal modelling parameters, and three models were found to give the best results, as shown in Table 3. It can be seen that five performance measures that evaluate the relationship between the measured and predicted total loads were used, namely: the coefficient of correlation, *r*, coefficient of efficiency, *E*, root mean squared error, *RMSE*, discrepancy ratio, *DR*, and Akaike information criterion, *AIC*. The coefficient of correlation, *r*, is the performance measure that is widely used in civil engineering but sometimes can be biased in reflecting higher or lower values, leading to misleading model performance. The coefficient of efficiency, *E*, is an unbiased performance estimate and provides an assessment of the overall model performance, which can range from minus infinity to 1.0, with higher values indicating better agreement [25]. The *RMSE* has the advantage in that large errors receive much greater attention than small errors, as indicated by Shahin et al. [26]. The discrepancy ratio, *DR*, is the ratio between the predicted and measured total sediment loads, and a model is considered to be suitable if its discrepancy ratio falls within the range of 0.5–2.0, as indicated by Sinnakaudan et al. [8]. The *AIC* gives an estimate of the expected relative distance between the fitted model and the unknown true model. The smallest value of *AIC* is considered to be the most favourable amongst the set of candidate models [27].

Table 3 shows that the three best EPR models have *r*, *E*, *RMSE* and *DR* close to each other and that all three models have consistent performance in both the training and testing sets. However, based on the *AIC* results, Table 3 shows that Model-1 is superior to the other models and can be considered to be optimal.

Performance measurement	**Model-1**	**Model-2**	**Model-3**
Correlation coefficient, *r*			
Training	0.72	0.72	0.73
Validation	0.74	0.74	0.74
Coefficient of efficiency, *E*			
Training	0.52	0.52	0.52
Validation	0.55	0.55	0.55
RMSE			
Training	2.46	2.46	2.46
Validation	2.41	2.41	2.41
Discrepancy ratio, *DR*			
Training	0.68	0.69	0.69
Validation	0.64	0.66	0.66
AIC			
Training	0.00	4.10	4.00
Validation	0.00	5.20	5.20

TABLE 3: Performance results of the EPR models in the training and testing sets.

As can be seen in the following equations (i.e. Eqns. 3–5), Model-1 has only 6 input variables (Eqn. 3), whereas both Model-2 (Eqn. 4) and Model-3 (Eqn. 5) have 8 input variables each. It should be noted that the performance results of these models are considered to be acceptable in representing the sediment transport problem compared to those of most available methods, as will be seen in the next section. The symbolic formulae obtained from the EPR Models are as follows:

$$T_j = 226356.81\, V\, d_{50}^{2} + 18.37\, Q^{0.5}\, Y_o\, S_o^{0.5}\, e^{0.5V} + 0.000012\, Q\, d_{50}^{0.5}\, e^{0.5B} \quad (3)$$

$$T_j = 222250.88\, V\, d_{50}^{2} + 18.17\, Q^{0.5}\, Y_o\, S_o^{0.5}\, e^{0.5V} + 0.000012\, Q\, d_{50}^{0.5}\, e^{0.5B} + 1.23\, Q\, Y_o\, W_S^{2}\, R^2\, S_O\, e^{2Ws+2R} \quad (4)$$

$$T_j = 162.24\, B^2\, Y_o\, W_s^2\, R^2\, S_o^2 + 222624.92\, V\, d_{50}^{2} + 18.15\, Q^{0.5}\, Y_o\, S_o^{0.5}\, e^{0.5V} + 0.000012 Q d_{50}^{0.5} e^{0.5B} + 0.000023\, Q^2\, W_s\, R^2\, e^{2R} \quad (5)$$

where: T_j is the total sediment load, V is the flow velocity, d_{50} *is* the median diameter of sediment load, Q is the flow discharge, Y_o is the flow depth, S_o is the water surface slope, B is the stream width, R is the hydraulic radius and ω_s is the fall velocity.

3.2 Robustness study

In order to confirm the robustness of the EPR model to generalise within the range of the data used for model training, an additional validation approach was utilised, as proposed by Shahin et al. [26]. The approach consists of carrying out a parametric study, part of which includes investigating the response of the EPR model output to changes in its inputs. All input variables, except one, were fixed to the mean values used for training, and a set of synthetic data (between the minimum and maximum values used for model training), was generated for the input that was not set to a fixed value. The synthetic data set was generated by increasing its values in increments equal to 5% of the total range between the minimum and maximum values, and the model response was then examined. This process was repeated using another input variable until the model response has been tested for all input variables. The robustness of the model was tested by examining how well the trends of the total sediment loads, over the range of the inputs examined, are in agreement with the underlying physical meaning of sediment problem. The results of the robustness study are shown in Figure 4, which agree with hypothetical expectations based on the known physical behaviour of the total sediment load. Figures 4 (a□f) shows that the predicted total sediment load increases in a relatively consistent and smooth fashion, as the discharge, velocity, width, river depth, median diameter, slope, hydraulic radius and fall velocity increase.

3.3 Interpretive ability of EPR model

When evaluating the EPR model, consideration must be given not only to its predictive accuracy but also to the interpretive ability of the model. This can be made by carrying out a sensitivity analysis that quantifies the relative importance of model inputs to the corresponding outputs. In this study, the relative importance was determined using three different sensitivity measures, namely the range (r_a), gradient (g_a) and variance (v_a), as follows [28]:

$$r_a = \max(y_a) - \min(y_a) \quad (6)$$

$$g_a = \sum_{j=2}^{L} \left| y_{a,j} - y_{a,j-1} \right| / (L-1) \quad (7)$$

$$v_a = \sum_{j=2}^{L} (y_{a,j} - \bar{y}_a)^2 / (L-1) \quad (8)$$

For all of the above metrics, the higher the value the more relevant is the input. Thus, the relative importance (R_a) can be given as follows [29]:

$$R_a = s_a / \sum_{i=1}^{I} s_i \times 100(\%) \quad (9)$$

where: $y_{a,j}$ is the sensitivity response for $x_{a,j}$ and s is the sensitivity measure (i.e. *r, g* or *v*). Figure 5 shows the graphical representation of the relative importance measures in the form of bar charts. It can be seen from Figure 5 that the river depth, Y_o, seems to provide greater importance than the other input variables for almost all sensitivity measures used, while the flow velocity, V, and median diameter of sediment load, d_{50}, hold less importance than the other input variables.

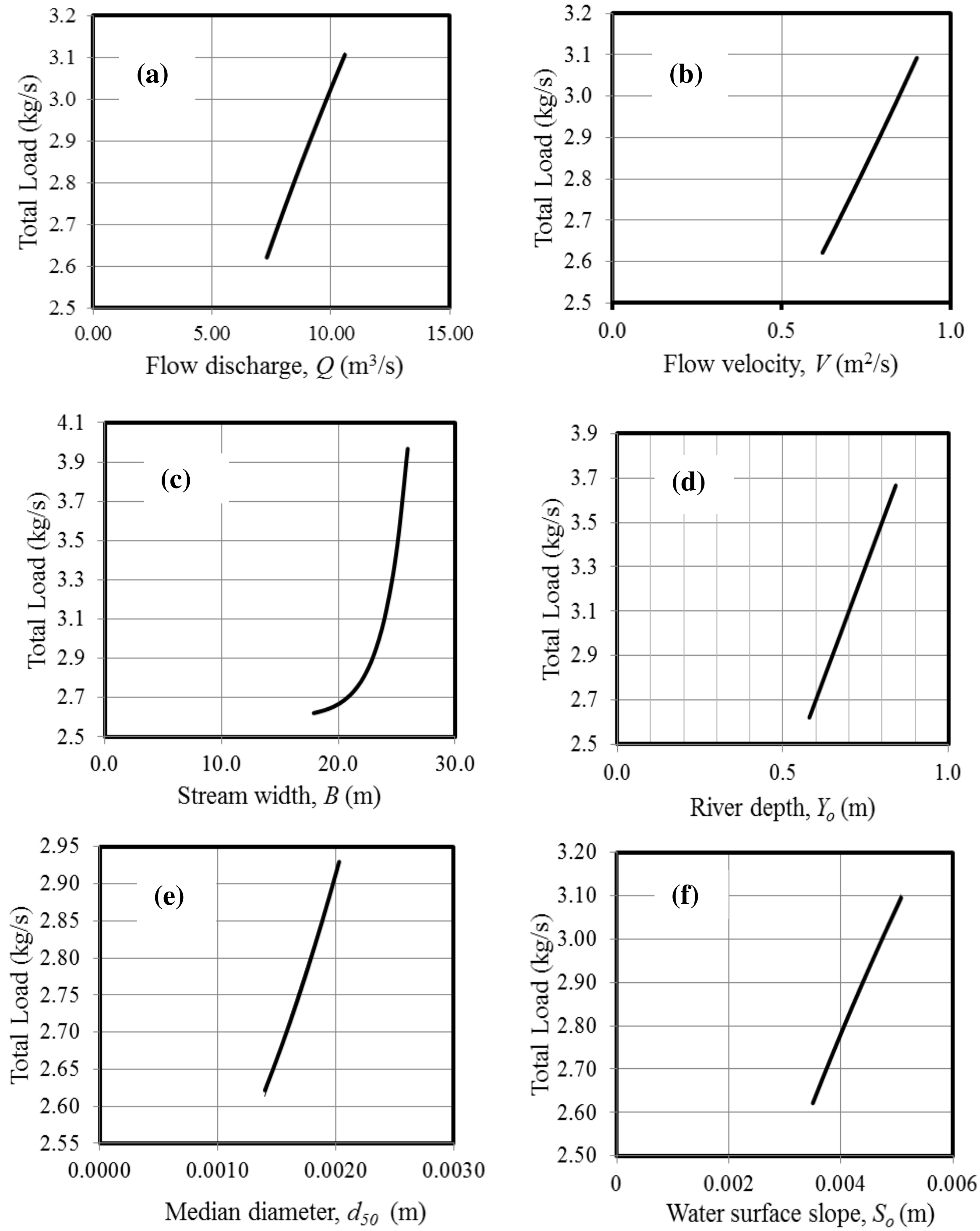


FIGURE 4: Robustness study showing the EPR model ability to generalise.

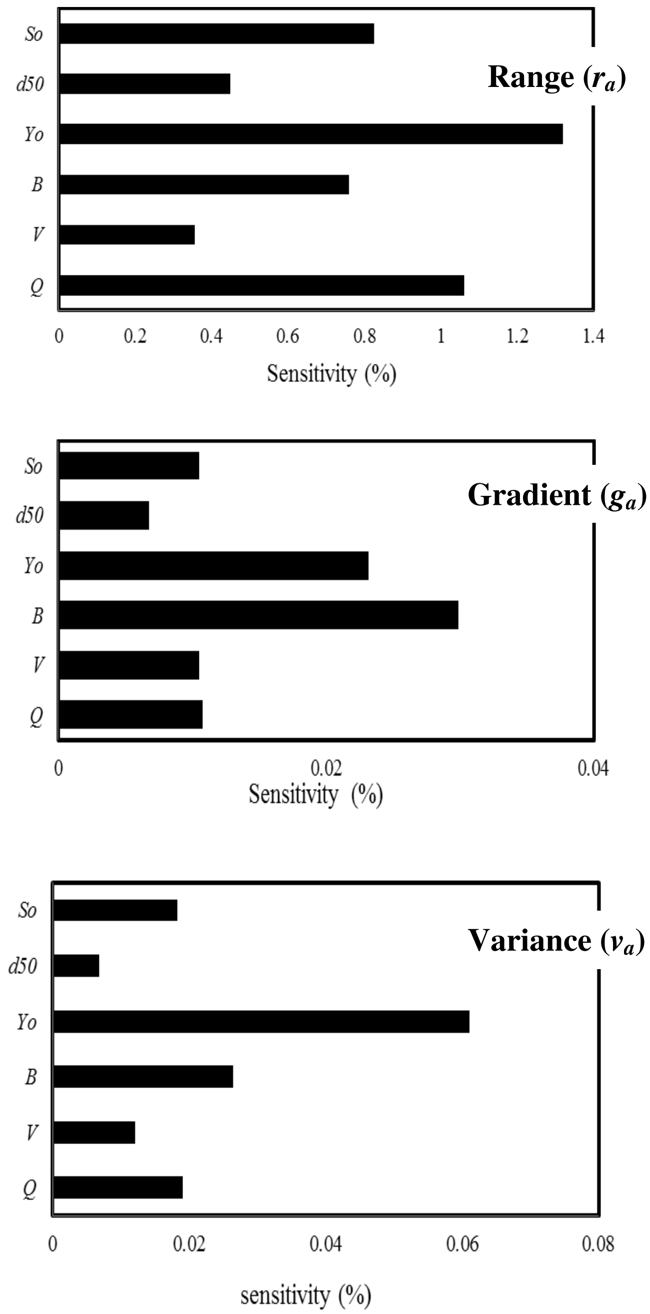


FIGURE 5: Sensitivity analysis showing the relative importance of the EPR model inputs.

3.4 Comparison of optimum EPR model with available models

In order to examine the accuracy of the developed EPR model against other available models, the EPR model predictions were compared with those obtained from six available sediment transport models, including Engelund & Hansen [1], Graf [2], Ariffin [9], Chan et al. [10], Sinnakaudan et al. [8], Zakaria et al. [30] and Aminuddin et al. [33]. A summary of the sediment parameters for other available methods used for comparison is given in Table 4. Statistical analyses, in relation to the 67 cases of the validation set, were carried out and the results are given numerically in Table 5 and represented graphically in Figure 6.

Model	Input parameters used
Engelund–Hansen [1]	$\gamma_s, V^2, \sqrt{d_{50}/g(\gamma_s/\gamma_w)}, \sqrt[1.5]{\tau/(\gamma_s-\gamma_w)d_{50}}$
Graf [2]	$(S_s-1)d_{50}/RS_o, C_vVR/\sqrt{g(S_s-1)d_{50}^3}$
Ariffin [9]	$R/d_{50}, U^*/\omega_s, U^*/V, V^2/gy_o$
Chan et al. [10]	$(S_s-1)d_{50}/RS_o, C_vVR/\sqrt{g(S_s-1)d_{50}}$
Sinnakaudan et al. [8]	$VS_o/\omega_s, R/d_{50}, \sqrt{g(S_s-1)d_{50}^3}/VR$
Zakaria et al. [30]	*Q, V, B, Yo, R, So, Ws,d50*
Ab. Ghani et al. [32]	*Q, V, B, Yo, A, P, So*

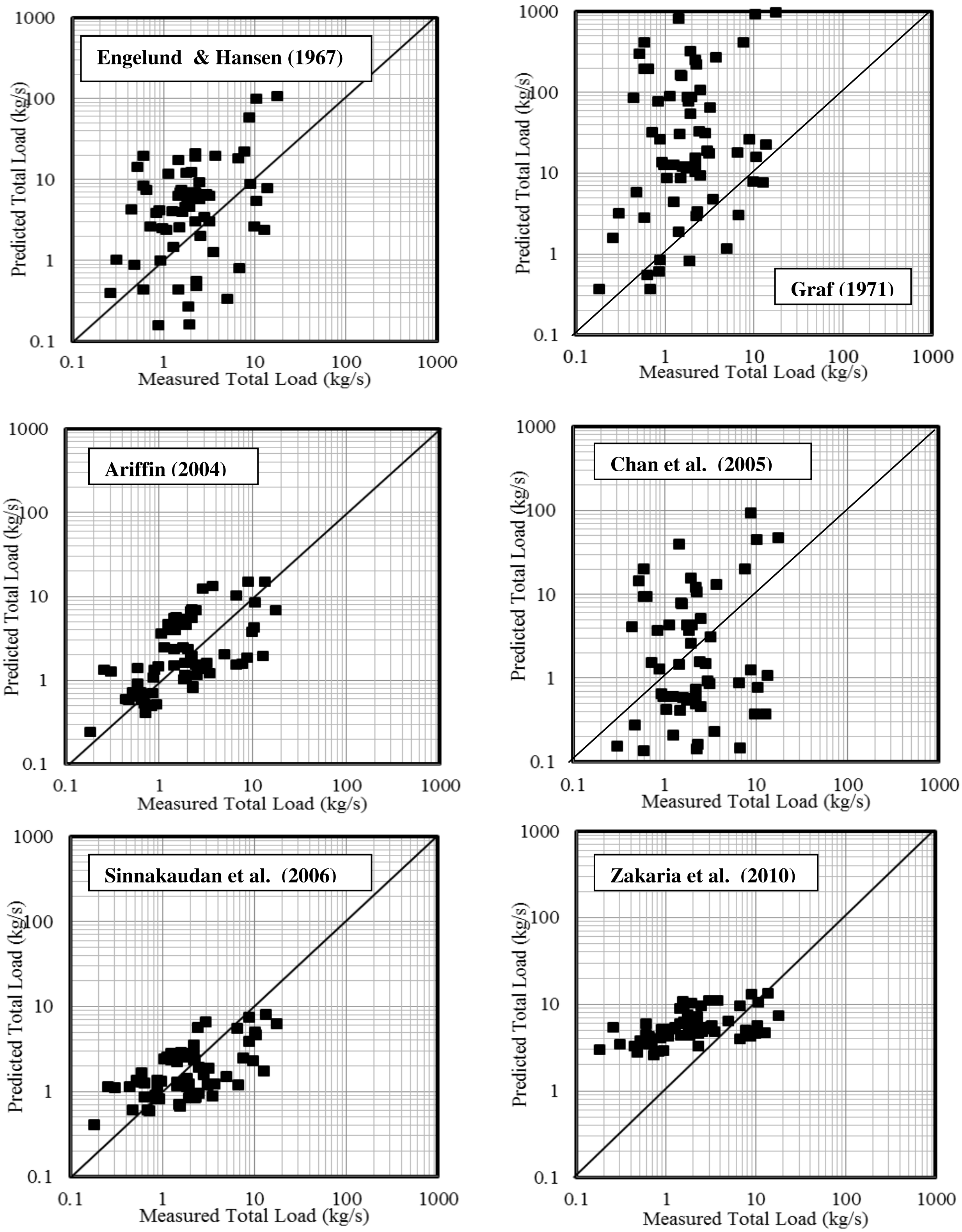
γ_s = unit weight of sediment; V = flow velocity; d_{50} = median diameter of sediment load; g = acceleration of gravity; γ_w = unit weight of water; τ = mean bed shear stress; S_s = specific gravity of sediment; R = hydraulic radius; C_v = volumetric sediment concentration; U^* = shear velocity, ω_s = fall velocity, Q = flow discharge; B = stream width, Y_o = flow depth, S_o = water surface slope; A = river cross sectional area, P = river perimeter.

TABLE 4: Summary of sediment parameters used in available methods.

It can be seen from Table 5 that the EPR model outperforms the other available methods in all performance measures used. It can also be seen that the model developed by Sinnakaudan et al. [8] comes second in order of best model performance. The graphical results also indicate that both the EPR model and Sinnakaudan et al. [8] have the least scattering around the line of equality between the predicted and measured sediment total loads, and this observation is confirmed numerically by the efficiency values, *E*, obtained in Table 5.

Model	Performance measure				
	R	***RMSE***	***E***	***DR***	***AIC***
Engelund & Hansen [1]	0.59	17.72	-23.28	0.21	94.8
Graf [2]	0.39	23.46	-8088.71	0.19	258.9
Ariffin [9]	0.47	3.63	-0.02	0.46	0.0
Chan et al. [10]	0.39	13.75	-13.62	0.15	75.1
Sinnakaudan et al. [8]	0.64	2.97	0.32	0.53	12.2
Zakaria et al. [30]	0.40	4.33	-0.45	0.24	39.2
Current study (EPR)	0.74	2.41	0.55	0.64	0.0

TABLE 5: Comparison of EPR model and other available methods (validation set – 67 data cases).


Engelund & Hansen (1967)
Graf (1971)
Ariffin (2004)
Chan et al. (2005)
Sinnakaudan et al. (2006)
Zakaria et al. (2010)
Predicted Total Load (kg/s)
Measured Total Load (kg/s)
0.1
1
10
100
1000

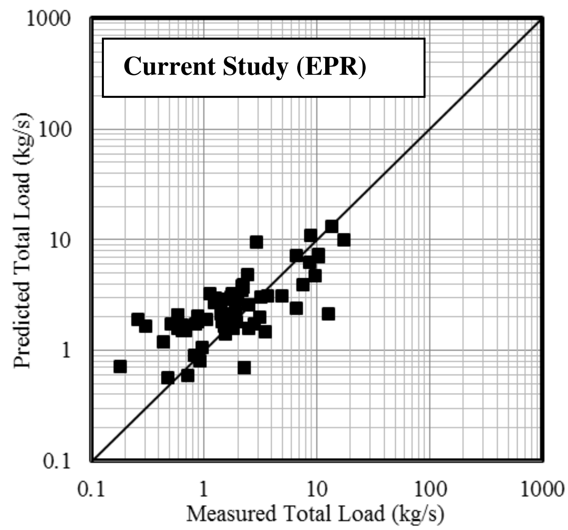


FIGURE 6: Predicted vs measured total sediment load for EPR and other methods.

4. CONCLUSIONS

This study investigated the use of the Evolutionary Polynomial Regression (EPR) technique in developing a new model for predicting sediment transport in Malaysian rivers. The data used for model calibration and validation involved 338 cases that were collected from the Department of Irrigation and Drainage (DID), Ministry of Natural Resources & Environment, Malaysia. The data were divided into 80% for model calibration (training) and 20% for model validation (testing). The EPR models were trained with eight input variables that thought to be significant including the hydraulic radius (R), flow depth (Y_o), flow velocity (V), median diameter of sediment load (d_{50}), stream width (B), water surface slope (S_o), fall velocity (ω_s) and flow discharge (Q). The only output is the total sediment load (T_j). Robustness study to investigate the generalisation ability of the developed EPR model was conducted, and a sensitivity analysis was also carried out to check the relative importance of model inputs to the corresponding output. Predictions from the developed EPR model were compared with those obtained from six available methods including: Engelund & Hansen [1], Graf [2], Ariffin [9], Chan et al. [10], Sinnakaudan et al. [8] and Zakaria et al. [30]. The statistical analyses used for comparison of performance of models included the coefficient of correlation, *r*, root mean squared error, *RMSE*, coefficient of efficiency, *E*, discrepancy ratio, *DR*, and Akaike information criterion, *AIC*.

The results indicate that the EPR model with six input variables (i.e. R, Y_o, d_{50}, B, S_o and Q) provided the best performance and was thus considered to be optimal. This optimum EPR model showed better performance, in relation to the validation set, than the other methods used for comparison with less scattering around the line of equality between the measured and predicted total sediment loads. For the EPR model: *r*, *RMSE*, *E*, *DR* and *AIC* were found to be equal to 0.74, 2.41, 0.55, 0.64 and 0.0, respectively. These measures were found to outperform those of the other available methods. The EPR model was also found to be robust in terms of its generalisation ability as its behaviour was found to be in agreement with the underlying physical meaning of sediment transport. The sensitivity analysis indicated that the river depth, Y_o, provided greater importance than the other input variables, while the flow velocity, V, and median diameter of sediment load, d_{50}, and hold less importance than the other input variables. The above results indicate a high potential for using the EPR model over available methods for predicting the total sediment load of Malaysian rivers.

5. REFERENCES

[1] Engelund F. and Hansen. A monograph on sediment transport in alluvial streams. Denmark: Copenhagen. Teknisk Forlag, 1967.

[2] Graf W.H. Hydraulics of sediment transport. New York: McGraw Hill, 1971.

[3] Ackers P. and White W.R. (1973) “Sediment transport: new approach and analysis.” Journal of the Hydraulics Division. ASCE, vol. 99(11), pp. 2041-2060, 1973.

[4] Yang C.T and Molinas A. “ Sediment transport and unit stream power function”, Journal of Hydraulic Engineering, ASCE, vol. 108(6), pp. 774-793, 1982.

[5] Van Rijn L.C. “Mathematical modelling of suspended sediment in non-uniform flows.” Journal of Hydraulic Engineering, ASCE, vol. 112(6), pp. 433-455, 1986.

[6] Karim F. “Bed material discharge prediction for non-uniform bed sediments.” Journal of Hydraulic Engineering, ASCE, vol. 124(6), pp. 597-604, 1998.

[7] Nagy H.M., Watanabe K. and Hirano M. “Prediction of sediment load concentration in rivers using artificial neural network model.” Journal of Hydraulic Engineering, ASCE, vol. 128(6), pp. 558-595, 2002.

[8] Sinnakaudan S.K., Ab.Ghani A., Ahmad M.S. and Zakaria N.A. “Multiple linear regression model for total bed material load prediction.” Journal of Hydraulic Engineering, ASCE, vol. 132(5), pp. 521–528, 2006.

[9] Ariffin J. “Development of sediment transport models for rivers in Malaysia using regression analysis and artificial neural networks.” PhD Thesis, Universiti Sains Malaysia, Malaysia, 2004.

[10] Chan C.K., Ab. Ghani, A., Zakaria N.A., Abu Hasan Z. and Abdullah R. “Sediment transport equation assessment for selected rivers in Malaysia.” International Journal of River Basin Management, vol. 3(3), pp. 203–208, 2005.

[11] Giustolisi O., Doglioni A., Savic D.A. and Webb, B.W. “A multi-model approach to analysis of environmental phenomena.” Environmental Modelling & Software Journal. vol, 22 pp. 674–682, 2007.

[12] Giustolisi, O., Savic, D.A., “Evolutionary Polynomial Regression (EPR): Development and Application.” Report 2003/1. School of Engineering, Computer Science and Mathemathics, Centre for Water Systems, University of Exeter, 2003.

[13] Azamathulla, H. Md., Chang, C.K., Ab. Ghani. A., Ariffin, J., Zakaria, N.A. and Abu Hassan, Z. “An ANFIS-based approach for predicting the bed load for moderately sized rivers.” Journal of Hydro-environmental Research”, vol. 3, pp. 35-44, 2009.

[14] Giustolisi, O. and Savic D.A. “A symbolic data driven technique based on Evolutionary Polynomial Regression.” Journal of Hydroinformatics, vol. 8(3), pp. 207–222, 2006.

[15] Savic D.A., Giutolisi O., Berardi L., Shepherd W., Djordjevic S. and Saul A. “Modelling sewer failure by evolutionary computing.” Proceeding of the Institution of Civil Engineers, Water Management, vol. 159(2), pp. 111–118, 2006.

[16] Berardi L., Giustolisi O., Kapelan Z. and Savic, D.A. “Development of pipe deterioration models for water distribution systems using EPR.” Journal of Hydro Informatics, vol. 10(2), pp. 113–126, 2008.

[17] Giustolisi O., Doglioni A., Savic D.A. and Pierro F. “An evolutionary multiobjective strategy for the effective management of groundwater resources.” Water Resources Research Journal, vol. 44(W01403), pp. 1–14, 2008.

[18] Draper N.R., and Smith H. Applied regression analysis. New York: John Wiley and Sons, 1998.

[19] Koza J.R. Genetic programming: on the programming of computers by means of natural selection. MIT Press, Massachusetts, 1992.

[20] Watson A., Parmee I. “System identification using genetic programming” Proceedings of ACEDC’96, University of Plymouth, United Kingdom, 1996.

[21] Goldberg D.E. Genetic algorithms in search, optimization and machine learning, Massachussets: Addison Wesley, 1989.

[22] Giustolisi, O. and Savic D.A. “A novel strategy to perform genetic programming: Evolutionary Polynomial Regression. “Sixth International Conference on Hydroinformatics, Singapore, 2004, pp. 787-794.

[23] Laucelli D., Berardi L. and Dogliono A. Evolutionary polynomial regression (EPR) – toolbox, Version 2.0 SA, Department of Civil and Environmental Engineering, Technical University of Bari, Italy, 2009.

[24] Shahin M.A., Maier H.R. and Jaksa M.B. “Data division for developing neural networks applied to geotechnical engineering.” Journal of Computing in Civil Engineering, ASCE, vol. 18(2), pp.105–114, 2004.

[25] Legates D.R. and McCabe Jr. G.J. “Evaluating the use of “Goodness-of-Fit” measures in hydrologic and hydroclimatic model validation.” Water Resources Research, vol. 35(1), pp. 233–241, 1999.

[26] Shahin M.A., Maier H.R. and Jaksa M.B. “Investigation into the robustness of artificial neural networks for a case study in civil engineering.” International Congress on Modelling and Simulation: Melbourne, 2004.

[27] Shaqlaih A., White L. and Zaman M. “Resilient modulus modeling with information theory approach.” International Journal of Geomechanics, in press.

[28] Kewley R., Embrechts M. and Breneman C.”Data strip mining for the virtual design of pharmaceuticals with neural networks.” IEEE Trans Neural Networks, vol. 11(3), pp. 668-679, 2000.

[29] Cortez P., Cerdeira A., Almeida F., Matos T., and Reis J. “Modeling wine preferences by data mining from physicochemical properties.” Decision Support Systems, vol. 47(4), pp. 547-553, 2009.

[30] Zakaria N.A, Azamathulla H.Md, Chang C.K. and Ab. Ghani A. “Gene expression programming for total bed material load estimation-a case study.” Journal of Science of the Total Environment, vol. 408(21), pp. 5078-5085, 2010.

[31] Rezania M., Faramarzi A. and Javadi A. "An evolutionary based approach for assessment of earthquake-induced soil liquefaction and lateral displacement." Engineering Applications of Artificial Intelligence, vol. 24(1), pp. 142-153, 2011.

[32] Ab. Ghani, A., Azamathulla, H.Md., Chang, C.K., Zakaria, N.A., Hassan, Z.A. " Prediction of total material load for rivers in Malaysia: A case study of Langat, Muda and Kurau Rivers." Environ Fluid Mech, vol. 11, pp. 307-318, 2011.