




# Reliable method of determining stable threshold channel shape using experimental and gene expression programming techniques

Azadeh Gholami<sup>1</sup> · Hossein Bonakdari<sup>1</sup>  · Mohammad Zeynoddin<sup>1</sup> · Isa Ebtehaj<sup>1</sup> · Bahram Gharabaghi<sup>2</sup> · Saeed Reza Khodashenas<sup>3</sup>

Received: 21 July 2017 / Accepted: 26 February 2018 / Published online: 5 March 2018  
© The Natural Computing Applications Forum 2018

## Abstract

The geometric dimensions and bank profile shape of channels with boundaries containing particles on the verge of motion (threshold channels) are significant factors in channel design. In this study, extensive experimental work was done at different flow velocities to propose a reliable method capable of estimating stable channel bank profile. The proposed method is based on gene expression programming (GEP). Laboratorial datasets obtained from Mikhailova et al. (Hydro Tech Constr 14:714–722, 1980), Ikeda (J Hydraul Div ASCE 107:389–406 1981), Diplas (J Hydraul Eng ASCE 116:707–728, 1990) and Hassanzadeh et al. (J Civil Environ Eng 43(4):59–68, 2014) were used to train, test, validate and examine the GEP model in various geometric and hydraulic conditions. The obtained results demonstrate that the proposed model can estimate bank profile characteristics with great accuracy (determination coefficient of 0.973 and mean absolute relative error of 0.147). Moreover, for practical calculations of channel dimensions, the model provides a specific mathematical relationship to solve problems with different discharge rates ( $Q$ ) and particles with various median grain sizes ( $D_{50}$ ). The proposed model's performance is compared with 8 relationships suggested previously by researchers (based on empirical and theoretical analyses) and a relationship obtained using a nonlinear regression model with different experimental data. The polynomial VDM and two exponential functions, i.e. IKM and DIM, are introduced as the superior existing models. According to the present study results, the proposed GEP model can predict the bank profile shape trend well and similar to the experimental datasets. Sensitivity analysis was also applied to assess the impact of each input variable ( $x$ ,  $Q$  and  $D_{50}$ ) on the presented relationship. According to the current study, the GEP model provides a suitable equation for predicting the bank profile shape of stable channels.

**Keywords** Gene expression programming · Bank profile shape · Threshold channel · Sensitivity analysis

## 1 Introduction

River science engineering requires determining the threshold conditions in natural rivers and irrigation channels for design purposes. The threshold or static equilibrium status of a channel is the final stage of channel

widening [48], when the channel's cross-sectional area reaches the smallest value. The geometric profile of a threshold channel's cross section is determined according to the properties of flow and materials on the channel bed. If the side slope is defined, it is possible to calculate the geometric shape of a channel with a specified central depth. Yu and Knight [57] defined two common methods of determining the geometric shape of a threshold channel. One of the methods is established on the channel widening procedure, whereby channel widening is simulated using numerical modelling of the bed shape change. This widening process occurs when the rate of sediment transport at any point on the wall is gradually removed and a threshold channel forms accordingly. The second method considers the equilibrium of the effective shear force on

✉ Hossein Bonakdari  
bonakdari@yahoo.com

<sup>1</sup> Department of Civil Engineering, Razi University, Kermanshah, Iran

<sup>2</sup> School of Engineering, University of Guelph, Guelph, ON N1G 2W1, Canada

<sup>3</sup> Water Engineering Department, Ferdowsi University of Mashhad, Mashhad, Iran

particles at any point on the bed, the shear stress value at the incipient motion of sediment and the flow discharge to calculate the threshold condition of materials on the channel's side slope. Glover and Florey [26] were the first who studied the notion of a stable channel profile. Stebbings [52], Mikhailova et al. [40] and Macky [37] subsequently carried out extensive experimental investigations on the bank profile shapes of stable channels. Parker [47] employed an evolutionary method to determine the distribution of mobile bed shear stress in channels with certain flow velocities. Ikeda et al. [31] presented a model for predicting the depth and width of rivers with gravel beds. They affirmed that the main factor affecting channels in stable state is lateral diffusion due to fluid turbulence. Moreover, the first effective parameters in determining stable channel geometry were introduced as the flow velocity, free-surface longitudinal slope and average size of particles on the bed. Cao and Knight [4] proposed a geometric model including flow continuity, frictional resistance, sediment transport, and bank profile-based entropy profile equations. Cao and Knight [4] stated that the bank profile shape is a polynomial curve. Mironenko et al. [41] deduced that the shape of a channel's transverse profile is a parabolic curve, while Diplas [6] and Pizzuto [49] suggested an exponential function shape. Diplas and Vigilar [7] used a numerical model and solved the momentum, fluid and energy balance equations for sediment particles in impending motion. They also presented some relationships for predicting channel stability. Their results suggested a polynomial of degree five for the predicted bank profile, which is quite different from conventional threshold bank shapes, such as cosine, parabolic and exponential. Vigilar and Diplas [55] suggested an exponential equation for bank profile calculation. The optimum existing channel profile shape is polynomial. Vigilar and Diplas [56] subsequently suggested the third-degree polynomial. Babaeyan-Koopaei and Valentine [3] evaluated stable alluvial channel profile geometry experimentally and compared the results with previously published equations. Apparently, the polynomial equation recommended by Diplas and Vigilar [7] leads to better estimation with the normal-depth method. The researchers suggested two hyperbolic functions for predicting channel cross-sectional shape. Dey [5] presented a modified model to calculate the cross-sectional dimensions of a threshold channel. Macky [38] proposed two models, namely a conventional and a diffusive model to describe the widening process in stable channels.

In recent decades, various soft computing methods have been applied for modelling complex phenomena. Many researchers have used soft computing to model a diversity of engineering cases, such as simulating open channel bends [19–24], predicting the discharge in flow diversion structures, such as weirs and side orifices [8, 34], and

scouring [45]. Based on the authors' knowledge, the only method ever used for predicting the bank profile shape of stable channels involves artificial neural networks (ANN). Khadangi et al. [32] employed ANN to model the geometry of alluvial channels and compared the results with semi-empirical equations. Madvar et al. [39] used a multi-layer perceptron neural network (MLPNN) to estimate stable channel dimensions. Taher-Shamsi et al. [53] employed ANN modelling to predict stable channel width and compared the results with regression equations. Shamsi et al. emphasized the high ANN model accuracy in predicting stable channel dimensions. One of the major setbacks of ANN models is the absence of a specific relationship for use in practical applications. Gholami et al. [25] evaluated GMDH performance in designing stable channel width, depth and slope for instance. Their results validated the high efficiency of the proposed GMDH model. However, they did not discuss bank profile shape in stable state nor assessed equations governing the stable bank profile. Shaghaghi et al. [51] compared a GMDH model with a particle swarm optimization evolutionary algorithm in estimating stable channel width. They declared that the generalized structure of the GMDH type (GS-GMDH) based on GA is more efficient than GMDH-PSO in making predictions.

As a subset of genetic algorithms, gene expression programming (GEP) is very popular with researchers (e.g. [2, 9, 17, 18, 27, 42, 46, 58]) because GEP facilitates the precise prediction of hydraulic phenomena and consumes less calculation time. In general, the major problem with some soft computing models (such as ANN and ANFIS) is the presence of complex rules that prevent these models from providing relationships for predicting variables regarding different inputs for use in various applications. The unique properties of the GEP model and its multi-genetic nature facilitate complex program improvement [15]. One of the noteworthy advantages of GEP over ANN is that GEP models provide an efficient and transparent modelling solution [28]. Moreover, GEP models have the powerful ability to generate equations for predicting variables [15, 43]. In a GEP model, the most important prediction set is selected and it does not require setting many parameters. These advantages render GEP a simple model for common uses [10, 50].

The main aim of the current study is to present a reliable GEP model for predicting the bank profile characteristics of stable channels. Therefore, a vast experimental study was carried out using discharge rates of 1.157, 2.18, 2.57 and 6 l/s to predict stable channel shape. In addition, extensive experimental data obtained by other researchers in various hydraulic and geometric conditions in different countries were applied to create the proposed models. A remarkable advantage of GEP is the possibility to extract a

simple mathematical relationship for problem solving. GEP modelling also facilitates determining an appropriate channel profile geometric shape by using different discharge rates and median grain sizes. The proposed GEP model results are compared with 8 theoretical and experimental relationships, and the superior model is introduced. Sensitivity analysis of the different input parameters for predicting stable channel shape is carried out as well.

## 2 Materials and methods

This section presents the various models used to predict stable channel shape, the experiments performed, the optimum GEP model and the model assessment criteria.

### 2.1 Overview of existing models

The shape of a channel in stable state is significant to bank profile geometric shape formation. Figure 1 illustrates two bank profile characteristics. In this figure,  $x$  and  $y$  (m) are the transverse and vertical distances from the bed centre, respectively,  $h$  is the flow depth at the centre line and  $T$  is the water surface width. The bank profile characteristics are dimensionless ( $y^* = y/h$ ,  $x^* = x/h$  and  $T^* = T/h$ ) and shown in Fig. 1.

Vigilar and Diplas [55] presented the following relationships to determine  $h$  and  $T^*$ :

$$h = \frac{(S_s - 1)d_{50}}{\delta_{cr}^*} \tau_c^* \tag{1}$$

$$T^* = \frac{T}{h} = -16.1814\mu^3 + 44.3206\mu^2 - 43.5548\mu + 21.1496 \tag{2}$$

where  $h$  = the depth at the bed centre in stable condition,  $T^*$  = dimensionless water surface width,  $\rho_s$  = sediment density,  $\rho$  = water density,  $d_{50}$  = mean sediment size (m),  $S_s = \rho_s/\rho$  = relative density,  $\tau_c^* = \tau_c/\rho g d_{50} R_s$  is the dimensionless critical shear stress,  $\tau_c$  = critical shear

stress,  $g$  = gravitational acceleration,  $R_s = (\rho_s - \rho)/\rho$  is the submerged specific gravity of sediment, and  $\delta_{cr}^*$  = the critical bed stress depth that is obtained as follows:

$$\delta_{cr}^* = -0.0223\mu^3 + .01481\mu^2 - 0.31403\mu + 1.031 \tag{3}$$

where  $\mu = \tan(\theta)$  is the submerged coefficient of the static friction of sediment particles and  $\theta$  is the angle of repose of sediment. Yu and Knight [57] presented an empirical equation for  $\mu$  as follows:

$$\mu = \tan \left[ \frac{0.302(\log 100d_{50})^5 + 0.126(\log 100d_{50})^4}{-1.811(\log 100d_{50})^3 - 0.57(\log 100d_{50})^2 + 5.952(\log 100d_{50}) + 37.52} \right] \tag{4}$$

$\tau_c^*$  has a dimensionless value that may be obtained from Van-Rijn’s [54] equations, where  $D^* = d_{50} \left[ \frac{(S_s - 1)g}{\nu^2} \right]^{1/3}$  is the dimensionless particle size and  $\nu$  is the kinematic viscosity of water. Table 1 lists the relationships examined in this study and proposed by Glover and Florey [26], Ikeda [30], Pizzuto [49], Diplas [6], Babaeyan-Koopaei and Valentine [3], Cao and Knight [4], Vigilar and Diplas [56] and Dey [5]. These relationships were achieved from numerical, experimental or field models based on analytical, mathematical, theoretical or regression relationships. The abovementioned relationships are expressed in the present study as follows: GFM (Glover and Florey’s Model), IKM (Ikeda’s Model), PIM (Pizzuto’s Model), DIM (Diplas’ Model), VDM (Vigilar and Diplas’ Model), CKM (Cao and Knight’s Model), BVM (Babaeyan-Koopaei and Valentine’s Model) and DEM (Dey’s Model).

### 2.2 Experimental model

In the present study, ten experimental datasets were used for model assessment. The laboratorial data obtained from Ikeda [30], Mikhailova et al. [40] (two sets), Diplas [6] and Hassanzadeh et al. [29] (two sets), and four datasets from the current laboratorial work were used to train, test and calibrate the models for stable channel bank profile estimation. The data applied in the models consist of several lateral points on the lateral profile of a channel’s mobile bed. All experiments were done in laboratory flumes. The sediment was coarse and the flow shallow to avoid ripple formation. The Froude number was larger than the dune formation range [6].

In addition, four sets of experiments were carried out to measure the bank profile parameters. The flume employed was 20 m long, 1.22 m wide and 0.6 m deep. The flume bed was filled with uniform sand with  $d_{50}$  of 0.53 mm and standard deviation ( $\sigma$ ) of 1.23. For the experiments, two trapezoidal and triangular-shaped cross sections were selected. Flow passed through the channel at four

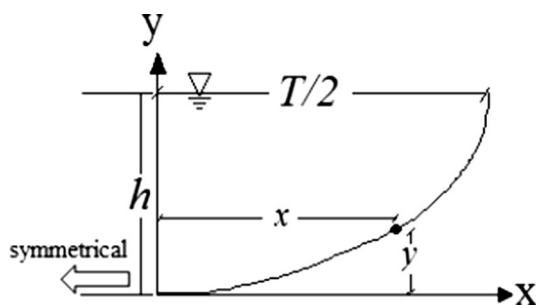


Fig. 1 Sketch of geometric characteristics

**Table 1** Models investigated in this study

Model	Equation	Parameters
GFM [26]	$y^* = 1 - \cos(\mu x^*)$	
IKM [30]	$y^* = \exp\left(\frac{x^* - \frac{T}{2}}{K}\right)$	$K = \frac{1}{h} \int_0^{T/2} y dy = \text{displacement thickness}$
CKM [4]	$y^* = (\mu x^*)^2 / 4$	$\mu = \text{coefficient of static friction } (= \tan \phi)$ $\phi = \text{angle of repose of sediment}$
DIM [6]	$y^* = \exp\left(x^* - \frac{T}{2}\right)$	
PIM [49]	$y^* = \exp\left[\mu\left(x^* - \frac{T}{2}\right)\right]$	
VDM [56]	$y^* = (a_3 x^{*3} + a_2 x^{*2} + a_1 x^* + a_0)$	$a_0, a_1, a_2$ and $a_3$ are a function of $\mu$ and $\delta_c^*$ is the dimensionless critical stress depth obtained by Vigilar and Diplas [55]
BVM (Babaeyan-Koopaei and Valentin [3])	$y^* = 3.5 \left[\tanh\left(-\frac{x^*}{4}\right)\right]^4$	
DEM [5]	$\sqrt{\beta^2 \mu^2 - 1} \frac{dy^*}{dx^*} = \mu \sqrt{[(1 - y^* + C m x^{*m-1})(1 - \beta \mu) + \beta \mu]^2 - 1}$	$C$ and $m$ are the coefficient and exponent, respectively, which are determined based on experimental data. $\beta = \frac{F_L}{F_D}$ , $F_L = \text{lift force}$ and $F_D = \text{drag force}$

velocities, each of which remained constant until stable channel state was attained. A magnetic flow metre was used to measure flow velocity. For each discharge rate (flow velocity), a 0.23% slope was fixed using a wooden former fixed on a carriage. Each test set continued until stable channel state was achieved at the plane and cross sections. All experiments were carried out for 10–12 h each until stable channel state was achieved. Finally, the geometry of the cross sections and bank profile was recorded. Two free water surface and channel bed levels were measured using a point gage at transverse distances of 8, 9 and 11 m from the channel entrance [33]. Table 2 summarizes the characteristics of the experimental sets conducted by Ikeda [30], Mikhailova et al. [40], Diplas [6] and Hassanzadeh et al. [29] as well as the current laboratory work. The laboratory flume used in the current study is shown in Fig. 2.

### 2.3 Gene expression programming (GEP) model

GEP is a linear-based genetic programming method presented by Ferreira [11] and developed from two evolutionary algorithms, namely the genetic algorithm (GA) and genetic programming (GP). GA presents the estimated results only in numerical form but GP solves the GA problem using a specified mathematical relationship and a computer program with a tree structure [36]. Ferreira [11] exploited the benefits of using both GA and GP simultaneously to overcome the shortcomings of the two individual methods. Because GEP uses different genetic operators in the chromosomes, it leads to genetic diversity. This method is also considered multigenetic, which leads to more various complex programs [11–13]. To obtain a solution for a specific problem, GEP employs fixed-length character sequences and presents the solution as parse

**Table 2** Experimental characteristic summary of the data used in the present study

Researchers	Number of experiments	$D_{50}$ (mm)	Discharge $Q$ (l/s)	Water surface width $T$ (cm)	Central water depth $h$ (cm)	Water surface slope $S$ ( $\times 10^{-3}$ )
Mikhailova et al. [40]	3	2	44–69	143–362	14–14.4	8–1.8
Ikeda [30]	34	1.3	1.06–25.2	24.8–166.8	1.63–4.03	1.8–5
Diplas [6]	7	1.9	4.9–12.52	54.8–79.8	2.1–4.9	3.7–9.1
Hassanzadeh et al. [29]	6	1.2–1.6	11.09–20.07	64–81.2	8.6–10.9	–
Present paper	3	0.53	1.16–6.2	19.6–43.4	3.7–8	2.3



**Fig. 2** Laboratorial model

trees, or expression trees (ETs) [1] of different shapes and sizes [44]. The main components of the GEP algorithms include the control, fitness function, function set, terminal set and terminal condition parameters [44]. Each gene in GEP includes a fixed-length set of symbols that can be any components of the terminal set  $\{Z, L, M, N, -1\}$  or function set  $\{+, -, \times, /, \sqrt{\quad}, \tan, \cos\}$  [1]. For instance, for a chromosome with two genes consisting of four functions  $\{\sqrt{\quad}, \times, /, +\}$  and three terminals  $(a, b, c)$ , the related algebraic expression is  $(a/b) + (\sqrt{a \times b})$ . The GEP modelling process is initialized with the random generation of a primary chromosome population. Each individual chromosome in the initial population is evaluated using the fitness function defined for the problem. Fitness functions may include mean square error (MSE), root mean square error (RMSE), root relative square error (RRSE) and relative square error (RSE) [2, 11, 35]. However, owing to the widespread use and good RRSE results in various studies [35], this fitness function is used in the current study. Chromosomes deemed to have superior fitness function values are selected for use in the next generation. Subsequently, the selected chromosomes are adapted using genetic operators. Several genetic operators used for chromosome modification are: mutation, inversion, three transportation operators and three recombination operators, which are defined as follows:

### 2.3.1 Mutation

Mutation can happen anywhere in the chromosome. The existing function at the chromosome head can be switched with a function or terminal, but at the tail, a terminal must be exchanged only with another terminal. Therefore, mutation helps maintain structurally organized chromosomes and structurally correct programs when using the new individuals generated by this operator. There are no imperatives regarding the number or type of mutations in a certain chromosome. Mutation is done at point 0 on gene 1, point 3 on gene 2 and point 1 on gene 3 (Fig. 3).

### 2.3.2 Inversion

This genetic operator is limited to the gene head and inverts a selected sequence in the gene head. To modify the gene, the terminal and initial points of the head portion are inverted through the random selection of chromosomes. With the inversion operator, each chromosome is changed only once. In Fig. 3, inversion is conducted at gene 2, whereby sequence *adjkb* is the inverse of sequence *bkjda*.

### 2.3.3 IS transposition

The insertion sequence (IS) transposition elements are selected randomly throughout the chromosomes. These

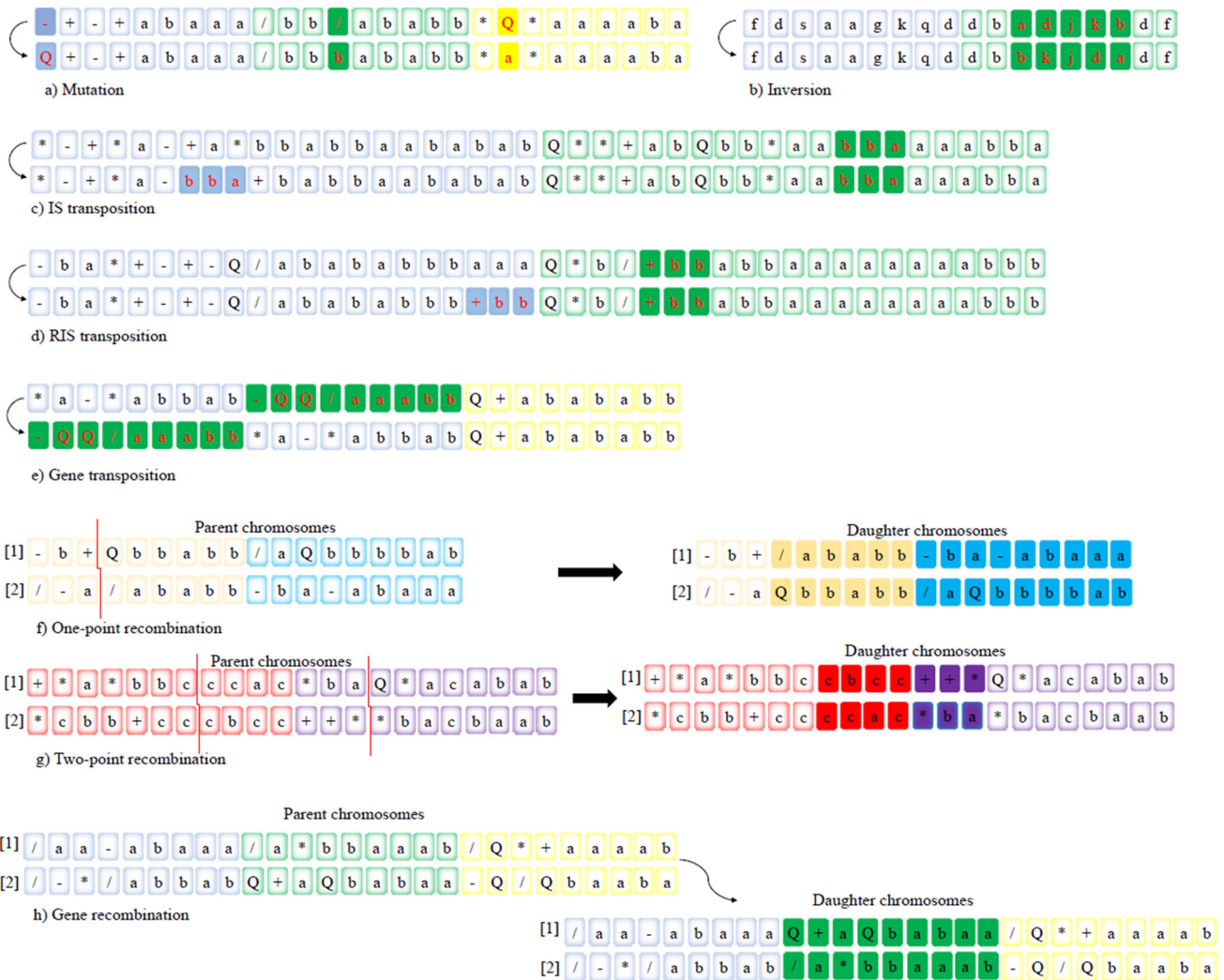


Fig. 3 GEP model genetic operators

elements are short genomic fragments with a terminal or function in the first position. IS transposition randomly selects a chromosome, a gene to be reclaimed, the start of an element, the target site and the transposon length. Considering a chromosome with two genes (Fig. 3), sequence *bba* was selected in the second gene for the IS component. Bond 6 in gene 1 was chosen to be cut and sequence *bba* was copied in the insertion place. With this operator, the sequence upstream of the insertion position remains unaltered, though the sequence downstream of the replicated IS component is eliminated at the head end, so in the examples, *a\*b* is eliminated. It should be noted that insertion does not change the chromosome structural organization and the newly generated individuals are suitable programs. Moreover, the transposition results in extreme expression tree changes, which means a great adjustment occurs upstream of the insertion area.

### 2.3.4 RIS transposition

The difference between IS and root insertion sequence (RIS) is at the starting point. The starting point for IS can be a function or terminal, whereas for RIS it is always a function. Therefore, the RIS components are selected between head sequences. After randomly selecting a point in the head, the gene is checked downstream to find a function. If no function is found, the RIS operator is ineffective. The results of RIS transportation copying in the gene root are presented in Fig. 3. Here, sequence *+bb* in gene 2 of a chromosome with two genes is considered a component of RIS transposition.

### 2.3.5 Gene transposition

With this operator, a complete gene operates as a transposon. The gene transposition operator transposes itself to

the beginning of the chromosome and deletes the gene at the place of origin, unlike IS and RIS transposition. Therefore, the chromosome length is maintained. Random selection is considered to select the chromosome that undergoes gene transposition and one of its genes except the first one is selected for transposition. An example of gene transposition is presented in Fig. 3.

### 2.3.6 One-point recombination

In one-point recombination, the crossover point is selected randomly and the parent chromosomes are coupled at similar points. The gene’s contribution following the crossover point is replaced among two chromosomes, generating two daughter chromosomes (Fig. 3).

### 2.3.7 Two-point recombination

In two-point recombination, two crossover points are selected randomly and two parent chromosomes are coupled. Two new chromosomes are produced by replacing the material among two crossover points. In this operation, the first gene of both parents is cut downstream of the termination point. Gene 2 related to chromosome 1 is split downstream of the termination point. Moreover, the second gene related to chromosome 2 is cleaved upstream of the termination point (Fig. 3). Two-point recombination is more powerful than one-point recombination and is fruitful for the evolution of more complex problems, chiefly when using multi genetic chromosomes with multiple genes.

### 2.3.8 Gene recombination

In this operation, there are gene replacements among two parent chromosomes, creating daughter chromosomes that include genes from the two parents. The displaced genes are selected casually and situated in identical locations as the parent chromosomes. The newly generated individuals consist of genes from the two previous parents. In this operation, similar genes replace each other and most times, the replaced genes differ from the previous ones and new materials are created. In gene recombination, new genes have different arrangements from existing genes, and this operator is not capable of producing new genes.

Chromosomes thus reproduce through this process, which continues until a stopping criterion is fulfilled. Figure 4 displays the proposed method flowchart.

## 2.4 Deriving stable channel bank profile characteristics for GEP modelling

To generate the initial population, a chromosome with three different genes is used. According to Ferreira [11],

the initial population number should be determined based on the 30–100 range, and in this study, the population number selected was 30. By considering other effective modelling parameters as fixed, the initial population was larger. However, even with a population increase no significant changes in modelling results were observed. The best value obtained for the preliminary population was 80. Then the fitness of every individual was estimated with the following equation:

$$f_i = 1000/(1 + E) \quad E_i = P_{ij} - O_j \tag{5}$$

where  $O_{ij}$  and  $P_{ij}$  are the observed value and value obtained using the  $i$ th individual chromosome for fitness case  $j$ . The optimum result was achieved when the observed and predicted values were equal ( $E_{ij} = 0$ ). Subsequently, the function and terminal set were selected. To select the function set,  $\{+, -, \times, /$  was used first. Trial and error was employed to improve modelling accuracy and additional functions were added separately. Function set  $\{+, -, \times, /, x^2, x^4, x^5, \sqrt{x}, \text{Ln}\}$  was ultimately considered and the terminal set was as follows:

$$T = \{y, Q, x\} \tag{6}$$

The head/tail length and number of genes were obtained by trial and error. Three genes were used in each chromosome and the head length was 7. As the maximum number of arguments was 3, the tail was 15 ( $t = 7 \times (3 - 1) + 1$ ). Owing to the good performance of addition as a linking function, this function was used in this study [1, 9, 16, 59]. In the next stage, the values of different genetic operators, such as mutation, crossover, were determined as presented in Table 3.

## 2.5 Goodness of fit of model performance

To assess the different models employed in comparison with experimental data, the following statistical indices were used: three absolute error indices, i.e. Mean Error (ME), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), the relative error index Mean Absolute Relative Error (MARE), and the statistical parameters coefficient of determination ( $R^2$ ), correlation coefficient ( $R$ ), BIAS and  $\rho$  (Eqs. 7–14).

$$ME = \frac{1}{n} \sum_{i=1}^n (O_i - P_i) \tag{7}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_i - P_i| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \tag{9}$$

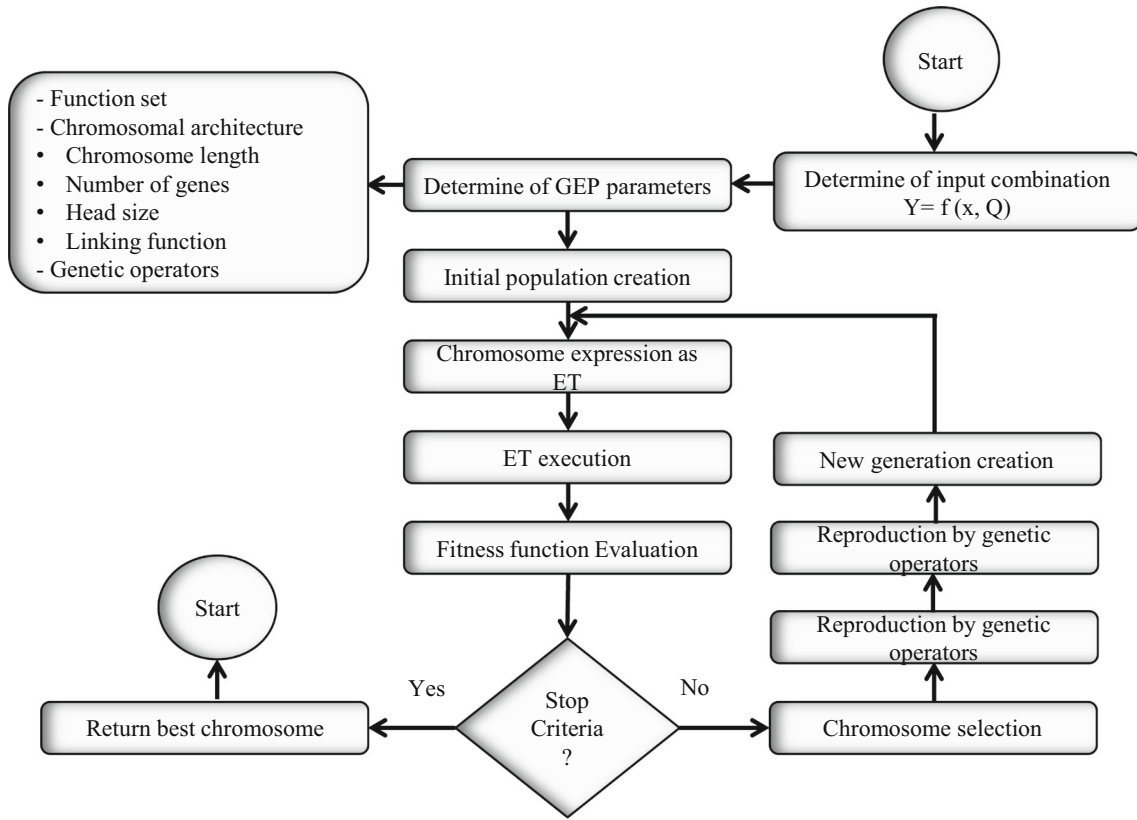


Fig. 4 GEP flowchart

Table 3 GEP model parameters

Parameter type	Parameter	Value
General setting	P1 number of generations	25,000
	P2 function set	+, −, ×, /, x <sup>2</sup> , x <sup>4</sup> , x <sup>5</sup> , √x, Ln
Chromosomal architecture	P3 chromosome length	80
	P4 number of genes	3
	P5 head size	7
	P6 linking function	Addition (+)
Genetic operators	P7 mutation rate	0.014
	P8 inversion	0.1
	P9 IS transposition	0.1
	P10 RIS transposition	0.1
	P11 on-point recombination	0.3
	P12 two-point recombination	0.3
	P13 gene recombination rate	0.1
	P14 gene transposition rate	0.1

$$MARE = \frac{1}{n} \sum_{i=1}^n \left( \frac{|O_i - P_i|}{O_i} \right) \tag{10}$$

$$Bias = \frac{\sum_{i=1}^N (P_i - O_i)}{n} \tag{12}$$

$$R = \frac{\sum_{i=1}^n (O_i - \bar{O}_i) \cdot (P_i - \bar{P}_i)}{\sqrt{\sum_{i=1}^n (O_i - \bar{O}_i)^2 \sum_{i=1}^n (P_i - \bar{P}_i)^2}} \tag{11}$$

$$SI = \frac{RMSE}{\frac{1}{n} \sum_{i=1}^n O_i} \tag{13}$$

$$\rho = \frac{SI}{1 + R} \tag{14}$$



where  $O_i$  is the observational value,  $P_i$  is the value predicted by the models,  $\overline{P}_i$  and  $\overline{O}_i$  are the means of the parameters and  $n$  is the number of parameters.  $R^2$  (coefficient of determination), which is the linear regression line between the observed values and the values estimated by the models, was used to determine the model performance. RMSE, MARE and MAE values closer to zero indicate little difference between the values predicted by the models and the experimental values. The SI values (dimensionless form of RMSE values) were used for better comparison. The BIAS index represents how the models performed in estimating variables as negative and positive. Negative and positive BIAS values indicate model underestimation and overestimation, respectively. The models (GEP, 8 relations and NLR) should be compared with the experimental model using another suitable criterion and not only  $R$ . This is because  $R$  does not change meaningfully by shifting a model's result values similarly, and the error functions only signify the error values and not the correlation in terms of RMSE and MARE. The  $\rho$  index was introduced by Gandomi and Roke [14].

### 3 Results and discussion

The first part of this section evaluates in detail the application of ten observational datasets to model learning. In the second part, the GEP and NLR models as well as available empirical relationships are verified, and stable channel shape is discussed. The third part addresses the bank profile shape suggested by the model introduced in this study. Finally, the model sensitivity to different input variables is analysed.

#### 3.1 Data processing

In the present study, ten experimental datasets were utilized for model training to predict threshold channel bank profile [30, 40] (two sets), [6], Hassanzadeh et al. [29] (two sets) and four sets from the current laboratorial work). Each dataset is related to a specific laboratorial setup with different hydraulic conditions and sand particles applied for stable channel formation. In each laboratorial dataset, the  $Q$  (discharge) and  $D_{50}$  (median grain size) values completely differ from each other. Hence, the models were trained in different discharge conditions, which is a significant advantage of the present study. The models were examined for different values of  $D_{50}$  and discharge ( $Q$ ) (hydraulic condition) as well as  $x$  (geometric parameter). Various profile shapes were obtained for each governing circumstance. Three variables, namely  $Q$ ,  $D_{50}$  and  $x$  (transverse distance from the channel centre), were used

as input variables and  $y$  (vertical boundary level of points located on a stable bank profile) served as output data from the introduced model. In total, there were 276 laboratorial data for  $Q$ ,  $D_{50}$  and  $x$  and 276 data corresponding to  $y$ . From the total, four laboratorial datasets were utilized from the current research (155 data) for training, while four laboratorial dataset (51 data) from Ikeda [30], Mikhailova et al. [40] (two sets) and Diplas [6] were used for testing, and two datasets (71 data) from Hassanzadeh et al. [29] were used for model calibration. Therefore, different ranges of hydraulic conditions ( $Q$  in a high range of 2–69 l/s and  $D_{50}$  in the range of 0.2–1.9 mm) and geometric variations ( $x$  in the 0–90.77 cm range and  $y$  in the 0–12.24 cm range) were utilized for model training so the extracted model results would be reliable.

#### 3.2 Verification of existing methods in predicting bank profile characteristics

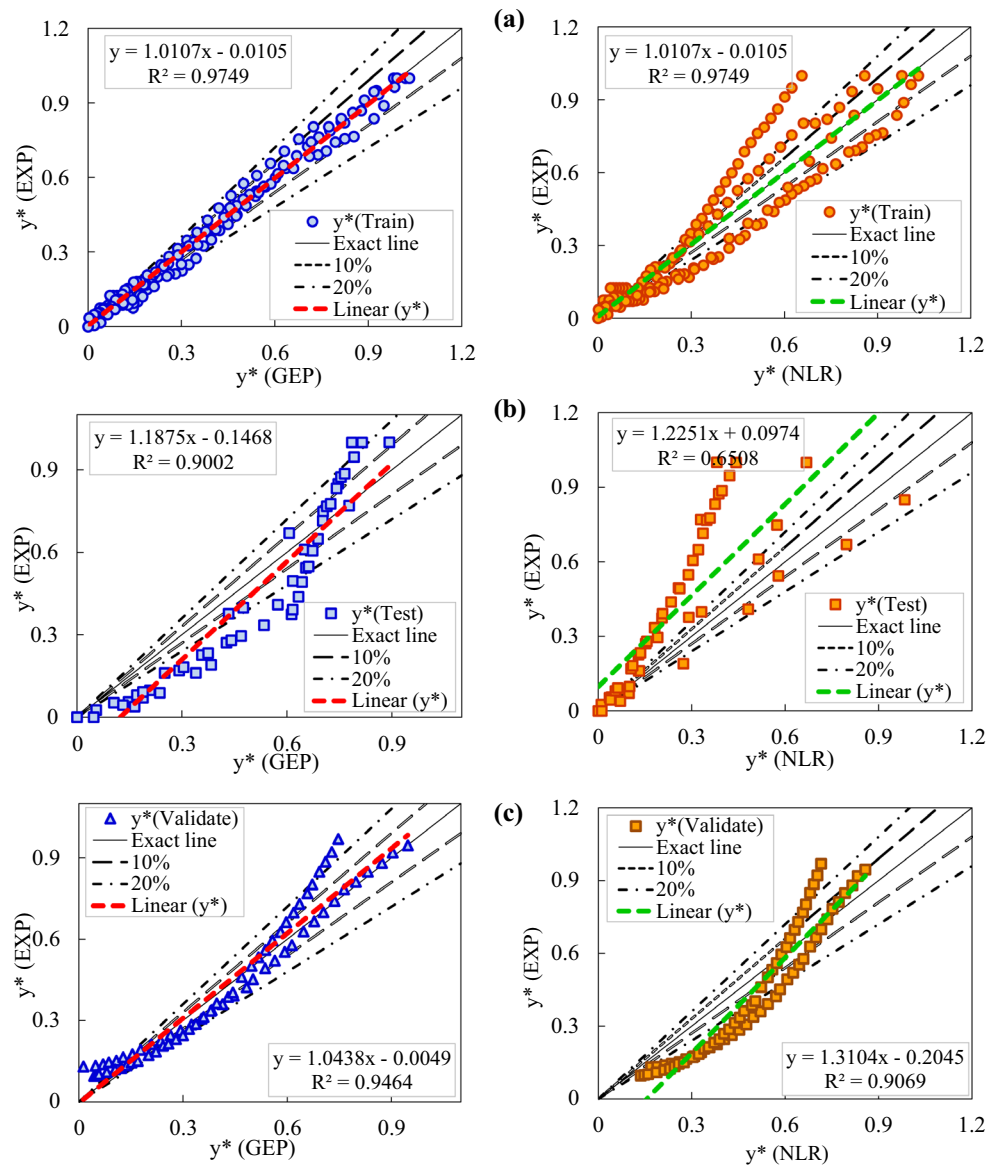
This section evaluates the performance of GEP, NLR and prior empirical relationships in predicting two threshold channel bank profile characteristics ( $x$ ,  $y$ ). The models were trained using ten experimental datasets with different hydraulic and geometric conditions. In the present study, in addition to artificial intelligence (AI), nonlinear regression (NLR) was also fitted to the experimental data to provide the following Eq. (15) for channel bank profile ( $y$ ) prediction, where  $Q$ ,  $D_{50}$  and  $x$  are considered input data:

$$y = 0.11628 \times Q^{-0.53536} \times D_{50}^{-0.08449} \times x^{1.53783} \quad (15)$$

Figure 5 presents scatter plots of  $y^*$  predicted by the GEP and NLR models in comparison with experimental values in training (current research experimental data), testing [30, 40] (two sets); [6]) and validation [29] (two sets). The figure shows the  $\pm 10$  and  $\pm 20\%$  error lines accordingly. Table 4 displays the error indices for the models with all ten datasets compared with experimental values.

According to Fig. 5, the GEP model data compressed around the exact line and the data correlation percentage is high, with acceptable  $R^2$  values approaching 0.9 in all stages. The  $R^2$  coefficient in training was 0.975, which is very close to 1 and indicates high GEP model prediction accuracy. Moreover, almost 90% of data are within the  $\pm 10\%$  error lines and all data are located within the  $\pm 20\%$  error lines, signifying low model prediction error. The GEP models were validated with high-precision experimental datasets [29] and were in good compliance with the experimental results ( $R^2 = 0.9464$ ). For the validation datasets, the fitted line almost coincides with the exact line. Moreover, few data exceed the  $+ 20\%$  error line. Hence, the GEP model results can be trusted in diverse practical cases. For the NLR model, data compression is not around

**Fig. 5** Comparison of  $y^*$  predicted by the GEP and NLR models with experimental results in **a** training, **b** testing and **c** validation



**Table 4** Comparison of GEP and NLR models with experimental results according to various statistical indices

Models	Stage	$R$	MARE	RMSE	MAE	BIAS	$\rho$
GEP	Training	0.993	0.188	0.034	0.026	0.0005	0.049
	Testing	0.938	0.800	0.133	0.159	0.055	0.164
	Validation	0.973	0.147	0.064	0.047	- 0.013	0.076
NLR	Training	0.976	0.312	0.102	0.076	- 0.0035	0.149
	Testing	0.807	0.950	0.272	0.306	- 0.068	0.360
	Validation	0.952	0.375	0.115	0.103	0.054	1.191

the exact line and most data are located above the fitted (dotted) line. If the fitted line is at the right of the exact line, it means the model overestimated and if it is at the left, the model underestimated. With the NLR model, the fitted line is much to the left of the exact line (especially in testing), which represents model underestimation. In

training, the GEP results were in acceptable compliance with the experimental results, with a high  $R$  value of 0.976. In the validation stage, the GEP model exhibited a high correlation value ( $R = 0.952$ ), but a few data exceed the  $\pm 20\%$  error lines. Also with the NLR model, most data are in the positive error range and only a small

percentage of data are scattered around the exact and  $\pm 10\%$  error lines. According to Table 4 (for training and testing), the GEP model had  $R$  of 0.993 and 0.973 in training and validation, respectively, which are very close to 1 (the same values as for  $R^2$ ), indicating 100% model compliance with the experimental results. Moreover, the low relative and absolute error index values in all stages signify the high GEP model accuracy (MARE = 0.147 and RMSE = 0.064 with the validation dataset). Also according to the table regarding the validation dataset, the lower relative error index value (MARE = 0.147) compared to the corresponding testing (0.80) and training (0.188) values represents superior model performance in the validation stage. The higher validation values for other indices (RMSE = 0.064,  $\rho = 0.076$ ) than training (RMSE = 0.034,  $\rho = 0.049$ ) were due to model weakness in predicting high  $y$  values (close to the free water surface values), which produced high RMSE and MAE values.

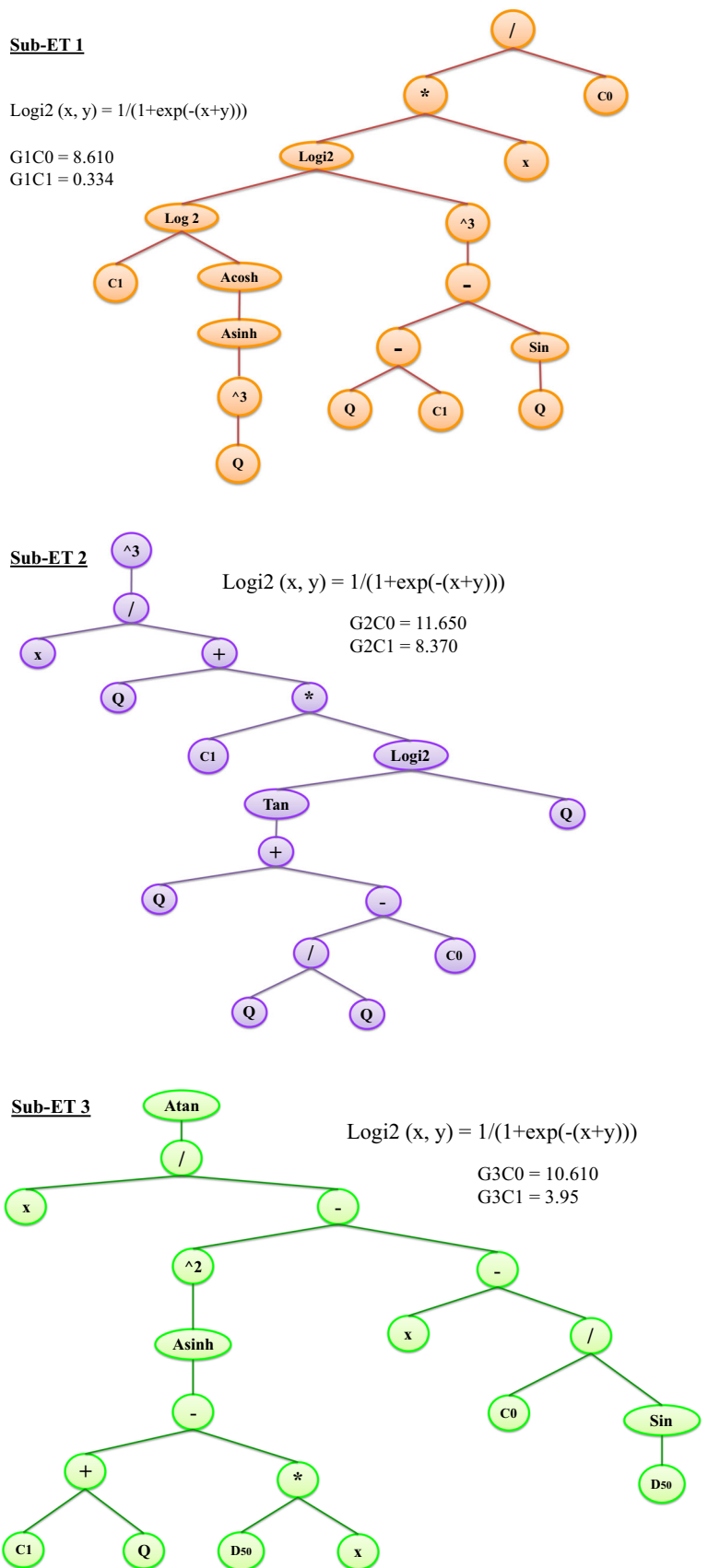
With NLR,  $R$  was 0.976 and 0.952 in training and validation, respectively, which are very close to 1, signifying high model accuracy. However, compared with the GEP model ( $R = 0.938$ ), this index value decreased by 16% in testing for the NLR model ( $R = 0.807$ ). Moreover, the increasing  $\rho$  value for NLR in both modes (by about 3 times especially in testing), indicates the superior GEP model performance. The error index values of NLR were about 2.5 times greater than GEP (MARE = 0.375 and RMSE = 0.115). The negative and positive BIAS index values indicate model underestimation and overestimation, respectively. Table 4 shows that for the NLR model, this index in training and testing was  $-0.0035$  and  $-0.068$ , respectively, which indicates model underestimation (Fig. 5). For the GEP model, the BIAS values very close to 0 indicate high model accuracy (BIAS = 0.0005 and  $-0.013$  in training and validation, respectively). The MAE index shows the absolute difference between the predictions and experimental results. This index value is almost 65% lower for the GEP model (the lowest MAE value of 0.026 in training) than for NLR (the lowest MAE value of 0.076 in training). That means the GEP model is more accurate than NLR and can predict stable channel bank profiles well. Besides, the high GEP model accuracy with different experimental data ranges and in all stages (especially in validation with the lowest relative error and high  $R$  coefficient) represents the applicability of GEP in most similar hydraulic and geometric circumstances. Figure 6 presents the genotype and phenotype [Expression Trees (ETs)] of a population individual related to the GEP model. In order to simplify these ETs, the pseudocode in Fig. 7 was used to predict bank profiles with different particle sizes and discharge rates.

### 3.3 Comparison of GEP and NLR models with existing methods in stable bank profile shape prediction

In the present section, the results obtained with the GEP model are compared with those of 8 existing models. Table 5 shows the relative and absolute MARE and RMSE error indices for the different models [GEP, NLR and 8 theoretical equations (Table 1)] in comparison with experimental data from the present work and Ikeda [30] and Diplas' [6] data for discharge rates of 1.157, 16.28 and 12.526 l/s. Figure 8 presents MARE index bar graphs for better comparison.

Figure 9 presents the threshold channel bank profiles modelled by GEP, NLR, GFM, CKM, BVM, IKM and GFM in comparison with the experimental values from the present work and Ikeda [30] and Diplas' [6] data. Figure 8 and Table 5 demonstrates that the GEP model error was lower than previous models (MARE = 0.021 and RMSE = 0.0343) in training. After the GEP model, DEM and CKM exhibited the lowest error index (MARE of 0.12 and 0.1095, respectively). As seen in Fig. 9, these models performed almost the same and predicted a smaller cross section than the experimental outcome. The GFM model also calculated a small cross section and performed inefficiently, because the GFM equation does not consider lift force, which creates a sharp bank slope. Lift force is the main parameter affecting channel formation and equilibrium shape. The GFM model suggested a cosine curve as the channel shape, which is not a stable shape [7]. With three experimental datasets, CKM and GFM predicted smaller cross sections similar to other models but with a slight difference in error values. A channel with clean water and non-cohesive sediment particles was analysed. Based on the authors' data, despite the drop in MARE error value (0.1095) due to the non-cohesive sediments in the experiments, the CKM model estimated a smaller cross section than the experimental cross section. Moreover, CKM suggested a parabolic curve according to Mironenko et al.'s [41] study, which is not a suitable shape. Concerning lift force and non-cohesive sediments, DEM estimated a small cross section. However, the results indicate that the DEM modelling process was similar to VDM and the current research data, and DEM underestimated the experimental data (MARE = 0.1200). Unlike Diplas and Ikeda's data, the DEM model had quite a large error and was not as efficient as other models. After DEM, the PIM and VDM models had greater errors based on the present research data and calculated larger channel cross sections. The two models produced approximately the same error values (MARE = 0.3227 and 0.265 for PIM and VDM, respectively). Apart from Diplas and Ikeda's data, the PIM

**Fig. 6** Expression trees for the GEP model



**Fig. 7** Output program of the optimum GEP model

```

Clear
clc;

Q=input ('Q = ');
D50=input ('D50 = ');
x=input ('x = ');

% GEP Model Formula for estimating vertical boundary elevation (y):

x1=(ln(0.334106)/ln(acosh(asinh((Q.^3)))));
y1=(((Q-0.334106)-SIN(Q)).^3);
AA==1/(1+exp(-(x1+y1)));
ET1=(x.*AA)/8.609863;
ET2=(x/(Q+8.367737.*((1/(1+exp(-((tan(Q-10.649445))+Q))))))).^3;
ET3=atan(x./(((asinh(3.948792+Q-(D50.*x)).^2)-(x-(10.610443/sin(D50))))));

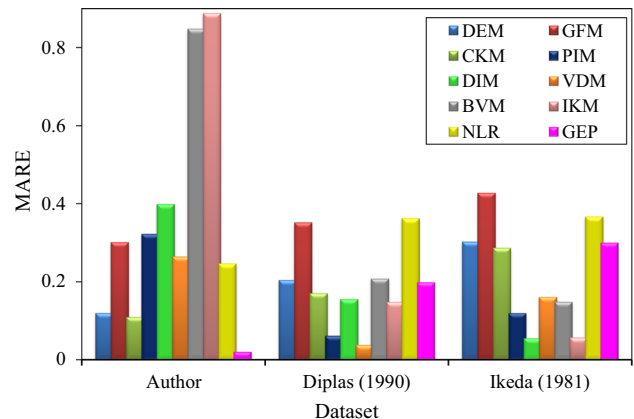
y=ET1+ET2+ET3;

disp('The vertical boundary elevation');
disp(y);
    
```

**Table 5** Evaluation of GEP, NLR and 8 previous models in comparison with experimental data from the present study, and Diplas [6] and Ikeda’s [30] data according to the relative and absolute MARE and RMSE error indices

EXP. data	Present work		Diplas [6]		Ikeda [30]	
	MARE	RMSE	MARE	RMSE	MARE	RMSE
<i>Models</i>						
DEM	0.120	0.13	0.2	0.5	0.3	0.7
GFM	0.30	0.310	0.35	0.92	0.43	1.03
CKM	0.109	0.119	0.17	0.47	0.29	0.82
PIM	0.32	0.337	0.06	0.22	0.12	0.23
DIM	0.40	0.434	0.16	0.36	0.06	0.1
VDM	0.265	0.290	0.04	0.13	0.16	0.35
BVM	0.850	1.03	0.21	0.39	0.15	0.27
IKM	0.890	1.070	0.15	0.31	0.06	0.11
NLR	0.247	0.164	0.36	0.3	0.37	0.3
GEP	0.021	0.034	0.2	0.12	0.3	0.09

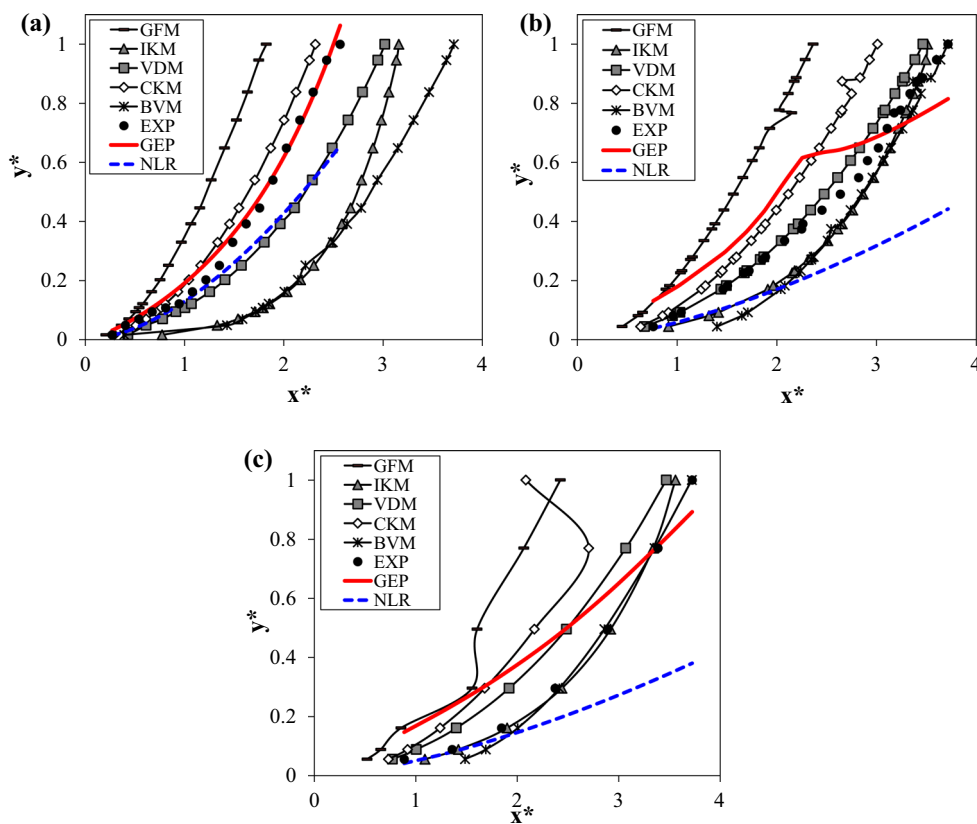
and VDM models performed the best with low error values. As seen in Fig. 9 b, the highest compliance with experimental data was achieved by the VDM model. Therefore, PIM and VDM are introduced as the best among the other models. The VDM model yielded a three-degree polynomial for a stable bank profile and is introduced as the best and most accurate model for stable channel profile



**Fig. 8** MARE error index of bank profile characteristics predicted by different models in comparison with the corresponding experimental values from the present work and Ikeda [30] and Diplas’ [6] data

prediction [33]. With the PIM and VDM models, the momentum equations for fluid and the force balance equations for sediment particles in impending motion were solved by a numerical method. The momentum diffusion term was also considered in the numerical models. The VDM and PIM models suggested polynomial and exponential curve channel profiles. The VDM model is the modified case of Pizzuto’s equation [49, 56]. Ikeda’s data fit well on two exponential curves of DIM and IKM, so these models are introduced as highly accurate based on Ikeda’s data. Also according to Diplas’ data, IKM and DIM are among the best performing models after VDM and

**Fig. 9** Bank profile shapes suggested by different models compared to the experimental model with **a** current study data, **b** Diplas' [6] data and **c** Ikeda's [30] data



PIM. With the present research data, the IKM and BVM models had maximum MARE and RMSE errors and predicted much larger cross sections, which are not good results. IKM introduced a displacement coefficient and proposed a cosine curve, whose coefficient calculation involves complex bank profile equations and trial and error. BVM proposed two hyperbolic functions for channel shape, and it was found that the equations obtained from VDM made better estimations than other equations (based on the cosine curve), signifying that hyperbolic equations perform similar to the polynomial VDM [33]. In other research studies, it seems that models which include momentum diffusion (polynomial curve) predict a more logical threshold channel form, although the error values are higher than models which neglect this term (cosine curve) [49]. The NLR model underwent a similar process to the experimental model, but it significantly differed in value prediction from the three experimental datasets (especially the current study dataset).

Therefore, it can be concluded that the VDM model with MARE of 0.265 that suggested a logical polynomial curve was the superior model with data from the current study. With Diplas and Ikeda's data, the VDM model (MARE = 0.04), and DIM and IKM (MARE = 0.06) produced the best performing relations.

With the data from the present study, the NLR model with MARE of 0.25 performed similar to VDM and proposed a similar shape as well. Compared to the other models, GEP with MARE of 0.034 was in high compliance with the experimental results and is introduced as the superior model with the data from the present study. Based on Diplas and Ikeda's experimental data, the GEP model predicted the profile shape almost according to the experimental data and other models, but it predicted  $y$  values with a slight difference from the experimental data. This difference is on account of the varying hydraulic and geometric circumstances in each laboratorial work; therefore, the proposed GEP model can be considered an appropriate equation for any experimental and real-case condition. With both datasets, the GEP models functioned similar to DEM [5] with MARE of 0.2 and 0.3. With Diplas' data, the RMSE for the DEM model (0.5) was greater than the GEP model (0.12). That is due to the weak performance of DEM with high  $y^*$  values, unlike the GEP model (Fig. 9b) with acceptable prediction values in this area. In Fig. 9 and according to all experimental data, CKM, and VDM and GEP proposed parabolic and polynomial curves for the channel, respectively. With the present study data, the BVM and IKM models proposed a cosine curve with a large difference from the experimental data, unlike Diplas's and Ikeda's data where IKM and

BVM were the closest to the experimental values. GFM predicted a small and unsuitable stable channel cross-sectional shape with all experimental data. The next section investigates the curved shape proposed by the GEP model at different discharge rates and with different experimental data.

### 3.4 Bank profile shape predicted by GEP in different hydraulic conditions

Figure 10 compares the bank profile shape calculated by the GEP model based on different experimental data and discharge rates with the experimental values. Table 6 presents the error indices for the bank profile characteristic predictions made by the GEP models compared to different experimental values at various discharge rates. Figure 10 shows regression line plots for each GEP model and the fitted equation for each discharge rate.

The proposed shape of a threshold channel is therefore the “polynomial type” due to the transverse slope at the channel centreline, which is zero and equal to the position angle of sediment at the water surface. It is worth noting that unlike the models presented in the prior section, the channel shape is not a cosine curve, parabolic, exponential or a hyperbolic function. An evaluation of the diagrams for different discharge rates based on different experimental datasets indicates that with increasing discharge, the threshold channel’s cross section is wider and deeper. With Mikhailova et al.’s [40] datasets, a wider cross section is clear at high discharge of 69 l/s rather than at 65 l/s. Moreover, sudden cross section widening occurred near the free water surface with these datasets. In this condition, the side bank slope reduced. In natural channels, during greater flood periods, the water surface slope is lower. As the width and bank adjustment rates occur much lower than the depth adjustment, the discharge increases, causing river water surface deepening. With Hassanzadeh et al.’s [29] data, the GEP model predicted the side bank slope reduction well with increasing discharge up to 17.96 l/s, but the difference between predicted values and the corresponding experimental data was slightly greater than the values predicted at 12.07 l/s. According to the error values in Table 6, with all experimental datasets and at different discharge rates, the error values decreased with increasing discharge and model efficiency increased except for Hassanzadeh et al.’s [29] dataset. With this dataset, the error values for two discharge rates were nearly equal, with slightly higher values at high discharge of 17.96 l/s. As mentioned previously, near the free water surface, the efficiency reduced slightly compared with the experimental values. With Mikhailova et al.’s [40] datasets, the RMSE value at 65 l/s (0.094) was less than at 69 l/s (0.122), unlike the MARE and  $R$  values. It can be said that the GEP

model was weak in predicting the  $y$  value near the free water surface and could not predict the sudden channel widening. Therefore, the RMSE value, which is the squared difference between the model and observational values increased at this water surface level. Based on Ikeda’s [30] data, the GEP model with higher MARE (0.8) than the experimental datasets exhibited low efficiency in predicting  $y$ . With Ikeda [30] and Diplas’ [6] datasets, despite the high difference between the model and experimental values, GEP predicted the bank profile shape trend similar to the observational values.

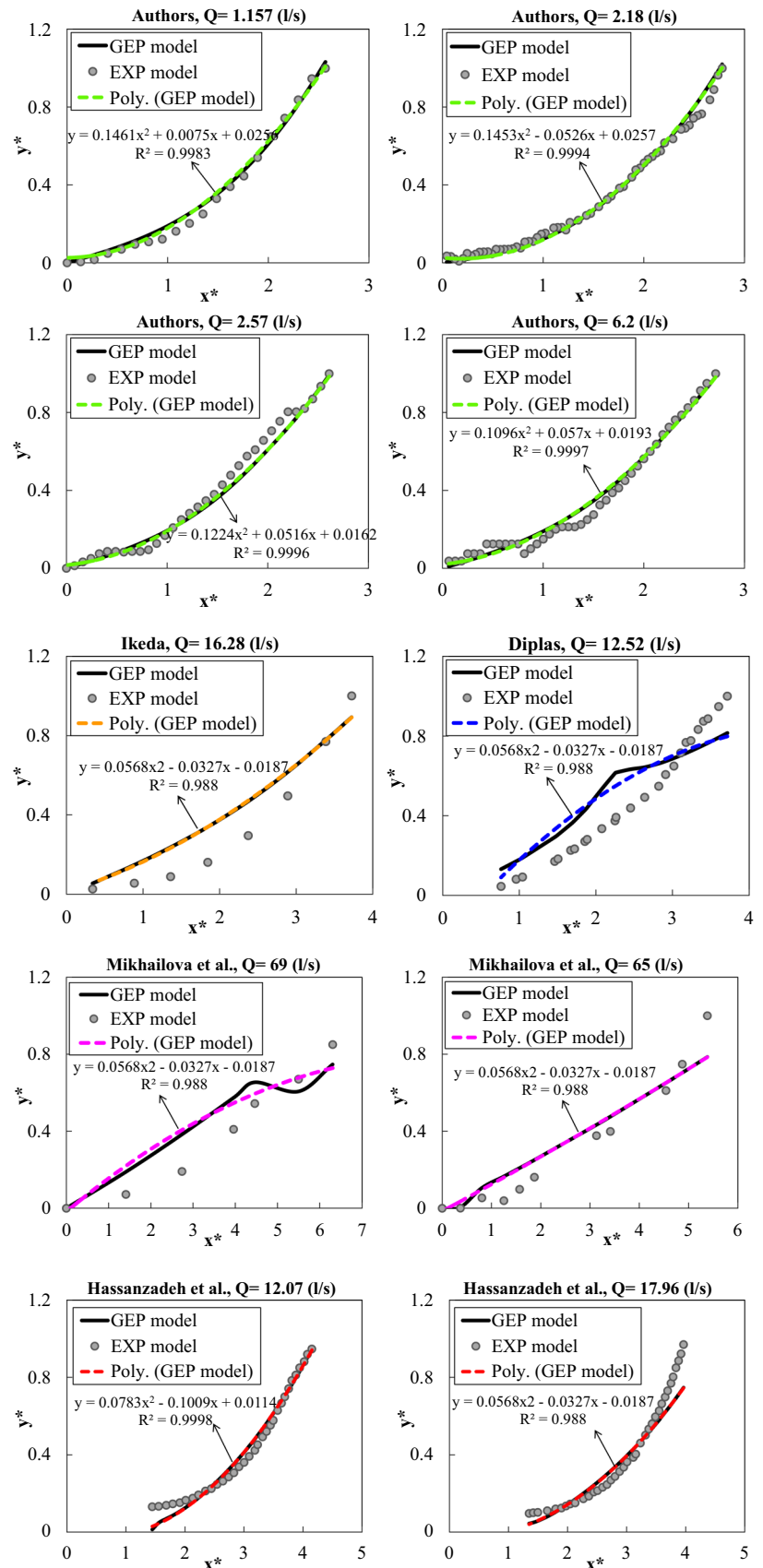
With the present study data, the statistical index values were better at the 1.157, 2.57 and 6.2 l/s discharge rates, because the experimental channel section in all three modes was the same (trapezoidal) and at 2.18 l/s the channel section was compound. It can also be concluded from Table 6 that with increasing discharge, the relative and absolute error indices decreased and the  $R$  coefficient value increased. Therefore, it can be said that with increasing discharge GEP model accuracy increased and it performed the best at 6.2 l/s with the lowest index errors ( $R = 0.994$ , MARE = 0.143, RMSE = 0.042). At this discharge rate, the absolute difference between the predicted and experimental data (MAE = 0.030) was very low (about 3%).

The proposed GEP model exhibited acceptable performance in all stages of training, testing and especially validation. Moreover, the GEP model was the most efficient with Hassanzadeh et al.’s [29] validation dataset compared to other datasets. Hence, the proposed GEP model is applicable in different flow hydraulic and geometric circumstances. The GEP model could predict the bank profile trend the same as the laboratory measurement despite the small differences in some cases. It can be seen that the best fitted line is a polynomial curve, which is in complete compliance with the GEP model for all datasets.

### 3.5 Sensitivity analysis

In the present study, the effect of the  $Q$ ,  $D_{50}$  and  $x$  parameters on estimating  $y$  (boundary elevation of a stable channel bank profile) was evaluated using sensitivity analysis. To analyse the  $y$  variable change trend according to the  $Q$ ,  $D_{50}$  and  $x$  input variables, partial derivative sensitivity analysis was applied to the equation extracted from the GEP model. With this method, the partial differences between the output variable and each input parameter were used to investigate the sensitivity results of input parameter  $x_i$ . It is clear that much higher partial derivatives indicated greater input variable effect on the results. If the derivative value was positive, an increase in input parameters led to an increase in the output variable. The sensitivity analysis of the GEP model is shown in Fig. 11 for bank profile

**Fig. 10** Stable channel bank profile predicted by GEP model and regression trend based on different laboratorial datasets at different discharge rates in comparison with experimental values

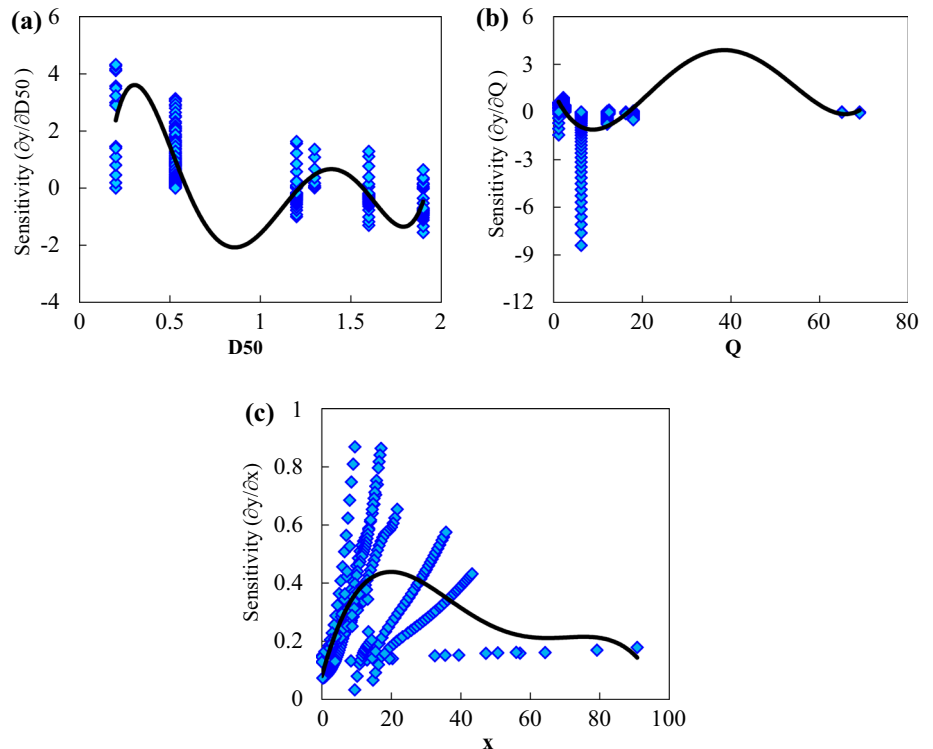




**Table 6** Different error indices for bank profile characteristics predicted by the GEP model compared to experimental results at various discharge rates

Researchers	Discharge (Q) (l/s)	R	MARE	RMSE	MAE	BIAS	$\rho$
Authors' data (training)	1.157	0.997	0.338	0.030	0.0265	0.0165	0.0021
	2.57	0.994	0.143	0.042	0.033	-0.021	0.0017
	6.2	0.992	0.202	0.037	0.030	0.0096	0.0012
	2.18 (compound)	0.998	0.157	0.028	0.02	0.00003	0.00075
Ikeda [30] (Testing)	16.28	0.97	0.80	0.122	0.107	0.080	0.021
Mikhailova et al. [40] (testing)	65	0.97	0.70	0.094	0.074	0.0264	0.014
	69	0.935	0.496	0.122	0.107	0.0595	0.020
Diplas [6] (testing)	12.526	0.936	0.387	0.137	0.125	0.062	0.0198
Hassanzadeh et al. [29] (validation)	12.07	0.988	0.146	0.043	0.034	0.0015	0.00145
	17.96	0.977	0.158	0.079	0.058	-0.0265	0.0026

**Fig. 11** Sensitivity analysis results for different input parameters: **a**  $D_{50}$ , **b**  $Q$  and **c**  $x$  in bank profile characteristic prediction by GEP with different experimental datasets



characteristic prediction based on Mikhailova et al. [40], Ikeda [30], Diplas [6], Hassanzadeh et al. [29] and the current study authors' experimental data.

Figure 11 indicates that at  $D_{50} < 0.6$  the sensitivity values are positive, signifying the strong effect of this parameter on  $y$  value prediction. With increasing  $D_{50}$  ( $D_{50} > 0.6$ ), the sensitivity values gradually tend towards negative, meaning that the parameter's effect on  $y$  prediction reduced in this range. In evaluating the effect of  $Q$ , it can be said that at low discharge (almost  $Q < 8$  l/s), the sensitivity values are negative and positive (at very low

discharge) and with increasing discharge ( $Q > 8$  l/s) the effect of this parameter rises. In evaluating the effect of parameter  $x$  on  $y$  value prediction, the sensitivity values in all  $x$  ranges are positive, affirming the strong effect of  $x$  on  $y$  prediction. At  $x < 20$  the sensitivity values increase to high values on the axis ( $\sim 1$ ), demonstrating the strong effect of this parameter in low  $x$  value ranges. With increasing  $x$  ( $x > 40$ ), the effect of this parameters on  $y$  value prediction is fixed.

## 4 Conclusions

In this study, we examined GEP model performance in estimating a vertical boundary level located in the bank profile of a threshold channel. The GEP model was trained, tested and validated based on ten sets of laboratorial data (276 data in total) with different hydraulic and geometric conditions. A vast experimental study was done in order to measure the stable bank profile shape in different hydraulic conditions. A robust and practical relationship was achieved from the proposed GEP model, while a nonlinear regression model was fitted on the same observational data. In addition, different relationships extracted by previous researchers based on regression, numerical and analytical methods were tested for comparison with the presented GEP model in order to assess the models' efficiency in terms of various observational data. Furthermore, sensitivity analysis was done to evaluate the impact of different parameters such  $Q$ ,  $d_{50}$  and  $x$  related to each point on GEP model performance. The results demonstrate that GEP model performance was acceptable with low error indices in all stages and especially in the validation stage. Therefore, the GEP results can be trusted in a range of hydraulic and geometric channel conditions. Based on the authors' current experimental data for GEP model training, the GEP model introduced performed the greatest compared with different prior models, with lower error indices. Based on all available experimental datasets, the GEP model generally predicted the bank profile shape trend (by suggesting a polynomial equation) similar to the experimental data despite the small error between predicted and experimental values. The GEP model suggested the polynomial shape for the cross section of a stable channel, which was also expressed by the VDM and PIM models. A great advantage of the proposed GEP model is that it presents a reliable equation suitable for different hydraulic and geometric conditions. The equation is capable of predicting the vertical boundary elevation of a bank profile and can be utilized as an alternative to previous traditional methods of practical predesigning. However, the presence of various functions and operators in the relationship proposed by the GEP model leads to higher GEP model complexity compared to previous analytical and regression-based methods. Therefore, we suggest using other computational intelligence method functions based on evolutionary algorithms to estimate the bank profile shape of stable threshold channels and compare the results with those from the present study.

## Compliance with ethical standards

**Conflict of interest** The authors declare no conflict of interest.

## References

- Alavi AH, Gandomi AH, Nejad HC, Mollahasani A, Rashed A (2013) Design equations for prediction of pressure meter soil deformation moduli utilizing expression programming systems. *Neural Comput Appl* 23:1771–1786
- Azamathulla HMd, Ahmad Z (2012) Gene-expression programming for transverse mixing coefficient. *J Hydrol* 434–435:142–148
- Babaeyan-Koopaei K, Valentine EM (1998) Bank profiles of self-formed straight stable channels. In: *Proceeding of the third international conference on hydroscience and engineering*, Cottbus/Berlin, Germany, August 31/September 3
- Cao S, Knight DW (1998) Design for hydraulic geometry of alluvial channels. *J Hydraul Eng ASCE* 124:484–492
- Dey S (2001) Bank profile of threshold channels: a simplified approach. *J Irrig Drain Eng* 127:184–187
- Diplas P (1990) Characteristics of self-formed straight channels. *J Hydraul Eng ASCE* 116:707–728
- Diplas P, Vigilar G (1992) Hydraulic geometry of threshold channels. *J Hydraul Eng ASCE* 118:597–614
- Dursun OF, Kaya N, Firat M (2012) Estimating discharge coefficient of semi-elliptical side weirs using ANFIS. *J Hydrol* 426:55–62
- Ebtehaj I, Bonakdari H, Zaji AH, Azimi H, Sharifi A (2015) Gene expression programming to predict the discharge coefficient in rectangular side weirs. *Appl Soft Comput* 35:618–628
- Ebtehaj I, Bonakdari H (2017) No-deposition sediment transport in sewers using gene expression programming. *J Soft Comput Civil Eng* 1:26–50
- Ferreira C (2001) Gene expression programming: a new adaptive algorithm for solving problems. *Complex Syst* 13:87–129
- Ferreira C (2002) Gene expression programming in problem solving. *Soft computing and industry*. Springer, London, pp 635–653
- Ferreira C (2006) Gene expression programming: mathematical modeling by an artificial intelligence, vol 21. Springer, Berlin, p 478
- Gandomi AH, Roke DA (2013) Intelligent formulation of structural engineering systems. In: *Seventh M.I.T. conference on computational fluid and solid mechanics-focus: multiphysics and multiscale*. Massachusetts Institute of Technology, Cambridge
- Gandomi AH, Tabatabaei SM, Moradian MH, Radfar A, Alavi AH (2011) A new prediction model for the load capacity of castellated steel beams. *J Constr Steel Res* 67(7):1096–1105
- Gharagheizi F, Ilani-Kashkouli P, Farahani N, Mohammadi AH (2012) Gene expression programming strategy for estimation of flash point temperature of non-electrolyte organic compounds. *Fluid Phase Equil* 329:71–77
- Ghani AA, Azamathulla HMD (2011) Gene expression programming for sediment transport in sewer pipe systems. *J Pipeline Syst Eng Pract* 2:102–106
- Ghani AA, Azamathulla HM (2014) Development of GEP-based functional relationship for sediment transport in tropical rivers. *Neural Comput Appl* 24(2):271–276
- Gholami A, Bonakdari H, Zaji AH, Akhtari AA (2015) Simulation of open channel bend characteristics using computational fluid dynamics and artificial neural networks. *Eng Appl Comput Fluid Mech* 9:355–361
- Gholami A, Bonakdari H, Zaji AH, Akhtari AA, Khodashenas SR (2015) Predicting the velocity field in a 90° open channel bend using a gene expression programming model. *Flow Meas Instrum* 46:189–192
- Gholami A, Bonakdari H, Zaji AH, Ajeel Fenjan S, Akhtari AA (2016) Design of modified structure multi-layer perceptron

- networks based on decision trees for the prediction of flow parameters in 90° open-channel bends. *Eng Appl Comput Fluid Mech* 10:194–209
22. Gholami A, Bonakdari H, Zaji AH, Michelson DG, Akhtari AA (2016) Improving the performance of multi-layer perceptron and radial basis function models with a decision tree model to predict flow variables in a sharp 90 bend. *Appl Soft Comput* 48:563–583
  23. Gholami A, Bonakdari H, Zaji AH, Fenjan SA, Akhtari AA (2017) New radial basis function network method based on decision trees to predict flow variables in a curved channel. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-017-2875-1>
  24. Gholami A, Bonakdari H, Ebtehaj I, Akhtari AA (2017) Design of an adaptive neuro-fuzzy computing technique for predicting flow variables in a 90° sharp bend. *J Hydroinf* 19:572
  25. Gholami A, Bonakdari H, Ebtehaj I, Shaghghi S, Khoshbin F (2017) Developing an expert group method of data handling system for predicting the geometry of a stable channel with a gravel bed. *Earth Surf Proc Land* 42(10):1460–1471
  26. Glover RE, Florey QL (1951) Stable channel profiles. Lab. Rep. 325 Hydraul, U.S. Bureau of Reclamation, Washington
  27. Guven A, Gunal M (2008) Genetic programming approach for prediction of local scour downstream of hydraulic structures. *J Irrig Drain Eng* 134:241–249
  28. Hashmi MZ, Shamseldin AY, Melville BW (2011) Statistical downscaling of watershed precipitation using gene expression programming (GEP). *Environ Model Softw* 26(12):1639–1646
  29. Hassanzadeh Y, Tabatabai MR, Imanshoar F, Jafari A (2014) Validation of river bank profiles in Sand-Bed Rivers. *J Civil Environ Eng* 43(4):59–68
  30. Ikeda S (1981) Self-formed straight channels in sandy beds. *J Hydraul Div ASCE* 107:389–406
  31. Ikeda S, Parker G, Kimura Y (1988) Stable width and depth of straight gravel rivers with heterogeneous bed materials. *Water Resour Res* 24:713–722
  32. Khadangi E, Madvar HR, Kiani H (2009) Application of artificial neural networks in establishing regime channel relationships. In: 2nd international conference on computer, control and communication, IEEEE, Carachi, Feb 17–18
  33. Khodashenas SR (2016) Threshold gravel channels bank profile: a comparison among 13 models. *Int J River Basin Manag* 14(3):337–344
  34. Khoshbin F, Bonakdari H, Ashraf Taleh SH, Ebtehaj I, Zaji AH, Azimi H (2015) ANFIS multi-objective optimization using genetic algorithm and SVD to modeling of discharge coefficient in rectangular sharp-crested side weirs. *Eng Optim* 48:1–16
  35. Kisi O, Shiri J (2012) River suspended sediment estimation by climatic variables implication: comparative study among soft computing techniques. *Comput Geosci* 43:73–82
  36. Koza JR (1992) Genetic programming: on the programming of computers by means of natural selection. MIT Press, Cambridge
  37. Macky GH (1999) Large flume experiments on the stable straight gravel bed channel. *Water Resour Res* 35:2601–2603
  38. Macky GH (2010) Diffusion of bed load and the development of stable straight gravel bed channels. In: 17th congress of the Asia and Pacific division of the international association of hydraulic engineering and research, APD, Auckland University, New Zealand, 241–246, Feb 2010
  39. Madvar HR, Ayyoubzadeh SA, Atani MGH (2011) Developing an expert system for predicting alluvial channel geometry using ANN. *Expert Syst Appl* 38:215–222
  40. Mikhailova NA, Shevchenko OB, Selyametov MM (1980) Laboratory of Investigation of the formation of stable channels. *Hydro Tech Constr* 14:714–722
  41. Mironenko AP, Willardson LS, Jenab SA (1984) Parabolic canal design and analysis. *J Irrig Drain Eng ASCE* 110:241–246
  42. Mohammadpour R, Ghani AA, Azamathulla HMD (2013) Estimation of dimension and time variation of local scour at short abutment. *Int J River Basin Manag* 11:121–135
  43. Mousavi SM, Aminian P, Gandomi AH, Alavi AH, Bolandi H (2012) A new predictive model for compressive strength of HPC using gene expression programming. *Adv Eng Softw* 45(1):105–114
  44. Mousavi SM, Mostafavi ES, Hosseinpour F (2014) Gene expression programming as a basis for new generation of electricity demand prediction models. *Comput Ind Eng* 74:120–128
  45. Najafzadeh M, Barani GhA, Kermani MRH (2013) GMDH network based back propagation algorithm to predict abutment scour in cohesive soils. *Ocean Eng* 59:100–106
  46. Onen F (2014) GEP prediction of scour around a side weir in curved channel. *J Environ Eng Landsc Manag* 22:161–170
  47. Parker G (1978) Self-formed straight rivers with equilibrium banks and mobile bed: Part 2: the gravel river. *J Fluid Mech* 89:127–146
  48. Pitlick J, Marr J, Pizzuto JE (2013) Width adjustment in experimental gravel-bed channels in response to overbank flows. *J Geophys Res* 118:553–570
  49. Pizzuto JE (1990) Numerical simulation of gravel river widening. *Water Resour Res* 26:1971–1980
  50. Rezaei H, Rahmati M, Modarress H (2017) Application of ANFIS and MLR models for prediction of methane adsorption on X and Y faujasite zeolites: effect of cations substitution. *Neural Comput Appl* 28(2):301–312
  51. Shaghghi S, Bonakdari H, Gholami A, Ebtehaj I, Zeinolabedini M (2017) Comparative analysis of GMDH neural network based on genetic algorithm and particle swarm optimization in stable channel design. *Appl Math Comput* 313:271–286
  52. Stebbings J (1963) The shape of self-formed model alluvial channels. In: ICE Proceedings, vol 25, pp 485–510, Thomas Telford. <http://dx.doi.org/10.1680/iicep.1963.10544>
  53. Taher-Shamsi A, Tabatabai MRM, Shirkhani R (2013) An evaluation model of artificial neural network to predict stable width in gravel bed rivers. *Int J Environ Sci Technol* 9:333–342
  54. Van-Rijn LC (1984) Sediment transport, Part 1: bed load transport. *J Hydraul Eng ASCE* 110:1431–1456
  55. Vigilar G, Diplas P (1997) Stable channels with mobile bed: formulation and numerical solution. *J Hydraul Eng ASCE* 123:189–199
  56. Vigilar G, Diplas P (1998) Stable channels with mobile bed: model verification and graphical solution. *J Hydraul Eng ASCE* 124:1097–1106
  57. Yu G, Knight DW (1998) Geometry of self-formed straight threshold channels in uniform material. *Proc Inst Civil Eng Water Marit Energy* 130:31–41
  58. Zakaria NA, Azamathulla HMD, Chang CK, Ghani AA (2010) Gene-expression programming for total bed material load estimation—a case study. *Sci Total Environ* 408:5078–5085
  59. Zhang Y, Pu Y, Zhang H, Su Y, Zhang L, Zhou J (2013) Using gene expression programming to infer gene regulatory networks from time-series data. *Comput Biol Chem* 47:198–206