

Prediction of Boundary Shear Stress Distribution in Converging Compound Channel Using Gene Expression Programming

Bandita Naik (✉ banditanaik@methodist.edu.in)

Methodist College of Engineering and Technology

Vijay Kaushik

Delhi Technological University

Munendra Kumar

Delhi Technological University

Research Article

Keywords: Converging compound channel, Boundary shear force, Flow parameters, Gene expression programming (GEP), Error analysis

Posted Date: January 14th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1015017/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Prediction of Boundary Shear Stress Distribution in Converging Compound Channel Using Gene Expression Programming

Bandita Naik¹, Vijay Kaushik², Munendra Kumar³

¹Associate Professor, Department of Civil Engineering, Methodist College of Engineering, Hyderabad, India

Email: banditanaik@methodist.edu.in (author for correspondence)

²Research Scholar, Department of Civil Engineering, Delhi Technological University, Delhi, India.

Email: vijaykaushik_2k20phdce01@dtu.ac.in

³Professor, Department of Civil Engineering, Delhi Technological University, Delhi, India.

Email: munendrakumar@dtu.ac.in

Abstract

The computation of the boundary shear stress distribution in an open channel flow is required for a variety of applications, including the flow resistance relationship and the construction of stable channels. The river breaches the main channel and spills across the floodplain during overbank flow conditions on both sides. Due to the momentum shift between the primary channel and adjacent floodplains, the flow structure in such compound channels becomes complicated. This has a profound impact on the shear stress distribution in the floodplain and main channel subsections. In addition, agriculture and development activities have occurred in floodplain parts of a river system. As a consequence, the geometry of the floodplain changes over the length of the flow, resulting in a converging compound channel. Traditional formulas, which rely heavily on empirical approaches, are ineffective in predicting shear force distribution with high precision. As a result, innovative and precise approaches are still in great demand. The boundary shear force carried by floodplains is estimated by gene expression programming (GEP) in this paper. In terms of non-dimensional geometric and flow variables, a novel equation is constructed to forecast boundary shear force distribution. The proposed GEP-based method is found to be best when compared to conventional methods. The findings indicate that the predicted percentage shear force carried by floodplains determined using GEP is in good agreement with the experimental data compared to the conventional formulas ($R^2 = 0.96$ and $RMSE = 3.395$ for the training data and $R^2 = 0.95$ and $RMSE = 4.022$ for the testing data).

Keywords: Converging compound channel, Boundary shear force, Flow parameters, Gene expression programming (GEP), Error analysis.

1. Introduction

One or more adjacent floodplains usually surround the main channel in several natural rivers. Because these channels comprise more than one of the basic elementary configurations, they are referred to as compound channels. Numerous hydraulic phenomena, such as channel roughness, sedimentation, catchment disintegration, bed morphology, land subsidence, and overbank flow conditions, involve boundary shear stress as an essential parameter. Smooth and rough compound channels have different boundary shear stress patterns depending on the channel geometry and secondary flow cell arrangement. The hydraulic radius and the longitudinal pressure gradient can be used to determine the average shear stress over a compound channel. The high velocity gradient and the resistance caused by the fluid viscosity in these channels keep the energy lost. The boundary and local shear stresses are difficult to determine since they depend on the properties such as velocity field, secondary cell structures, cross-sectional geometry, and boundary roughness. In addition, it is difficult to identify the boundary shear stress in an open channel flow, which further complicates the situation. (Khuntia et al. 2018)

Various investigators conducted several experiments on this issue to understand it better (Cruff 1965; Gosh and Roy 1970; Myers 1978; Knight and Macdonald 1979; Knight 1981; Noutsopoulos and Hadjipanous 1982; Knight et al. 1984). In two-stage and single-stage channels with movable bed conditions, a number of studies have investigated the wall shear stress distribution and flow resistance (Knight and Patel 1985; Nezu et al. 1993; Rhodes and Knight 1994; Ackerman and Hoover 2001; Tayfur 2002). Yang and Lim (1998) developed an analytical approach to calculate wall shear force distributions in compound channels. Artificial neural networks were used by Khuntia et al. (2018) to forecast the distribution of boundary shear stress in a two-stage straight channel. Sellin (1964) has examined the momentum shift phenomenon in laboratory flumes. Consequently, many scientists believed that the non-

uniformity in the boundary shear stress patterns along the section perimeter was generated by momentum transfer (Ghosh and Jena 1971; Knight and Hamed 1984; Patra et al. 2004). A model for boundary shear stress distribution in a homogeneous compound channel was developed by Knight and Hamed (1984) using a width ratio ($\alpha = \text{flood plain width (B)}/\text{main channel width (b)}$) value up to 4. The work was carried out by Khatua and Patra (2007) based on the data collected during their experiments. They came up with a width ratio model that is suitable for channels with a width ratio of 5. With $6.67 \leq \alpha \leq 11.96$, Mohanty and Khatua (2014) have created another novel channel system model. In laboratory flumes, the geometries of prismatic and meandering compound channels were thoroughly studied. However, when prismatic compound channel data was compared to non-prismatic compound channel data, substantial errors in percent S_{fp} estimation were detected (Bousmar and Zech 1999; Bousmar et al. 2004; Proust et al. 2006). In non-prismatic compound channel flow models, the extra momentum shift should be taken into consideration. Boundary shear stress distribution is strongly influenced by the geometry of the cross-section and the two-phase flow structure. As a result, new non-prismatic compound section models are required. To produce a unique expression for $\%S_{fp}$, experiments were performed on two-stage channel with converging floodplains.

In order to assess the connections between dependent and independent components, it is very difficult to construct any model for wall shear stress using mathematical, analytic, or numerical methods. Moreover, these models become noticeably ponderous and laborious; hence, an easily implementable method such as GEP (Gene Expression Programming) can be used to estimate boundary shear stress. It not only reduces the effort of experimenting in a short period but also eliminates rigorous computations. As a result, compound channels are increasingly being estimated using machine learning algorithms such as support vector machine (SVM), gene expression programming (GEP), artificial neural networks (ANN), fuzzy neural networks

(FNN), M5 tree decision model (MT), multi adaptive regression splines (MARS), genetic programming (GP), and supervised learning algorithms (Najafzadeh and Zahiri 2015; Parsaie et al. 2017; Sahu et al. 2011; Seckin 2004; Unal et al. 2010; Zahiri and Azamathulla 2014). In comparison to other soft computing approaches, gene-expression programming (GEP) has the benefit of being able to generate simpler equations without assuming a previous form of the existing connection. GEP is a multigene, one-of-a-kind coding language that allows for modifying more complex equations that are broken down into multiple sub-equations. It also employs generations of genes, gene choices based on fitness, and the introduction of genetic diversity by one or more genetic operators.

The GEP's ability to produce mathematical correlations makes it different from other soft computing techniques such as ANN and SVM (Babovic and Keijzer 2002; Cousin and Savic 1997; Drecourt 1999; Karimi et al. 2015; Savic et al. 1999; Whigham and Crapper 1999, 2001). However, there is less research on river engineering by GEP technique (Azamathulla et al. 2013; Giustolisi 2004; Guven and Aytok 2009; Guven and Gunal 2008; Harris et al. 2003; Pradhan and Khatua 2017b). In a study conducted by Das et al. (2019), a model equation has been developed for estimating discharge in diverging and converging compound channels, which support the use of GEP. Mohanta and Patra (2021) have established a model equation for estimating discharge in meandering compound channels, confirming the adoption of GEP over the traditional channel division method. The present study uses the GEP method, an evolution process that generates mathematical expressions, decision trees, polynomial constructions, and logical expressions. This study offers a model equation for forecasting shear force in compound channels with converging floodplains. Researchers use current experimental results for converging compound channels with different convergence angles in addition to the data sets of other researchers for estimating the overall shear force carried by floodplains. The method considers variables such as the width ratio (α), relative depth (β),

converging angle (θ), and relative distance (X_r). Prior shear force models are evaluated to determine the efficacy of the suggested equation and their performance.

2. Methodology

2.1 Experimental Setup

A series of experiments were conducted in a concrete flume containing three converging compound channels. Perspex sheet measuring 15 m in length, 0.90 m in width, and 0.5 m in depth was used to build these components. The channel's width ratio was 1.8 and its aspect ratio was 5. The channel's converging angles were 12.38° , 9° , and 5° , respectively while keeping the geometry constant. The non-prismatic compound channel has converging lengths of 0.84, 1.26, and 2.28 m, respectively. There was a 0.0011 longitudinal bed slope in the non-prismatic compound channels, and subcritical flow conditions were attained in several locations. Manning's n value was calculated using in-bank flow data from the same surface materials used in the floodplains and main channel. The flowing portion of the experiment was turbulent. This system recirculates the water supply by pumping it from an underground sump to a reservoir located in the experimental channel. The rectangular notch has been surrounded by adjustable vertical gates and a flow strengtheners. The flume's removable tailgates help maintain a consistent flow over the test reach. The water that flows from the channel is collected by a volumetric tank. It then goes back to the underground sump. Figure 1 shows the cross section of compound channel with geometrical parameters.

Figure 2 illustrates the experimental setup from the top. The plan view of several tests with cross-sectional dimensions of both Naik and Khatua (2016), and Rezaei (2006) channels is shown in Figure 3. In the test section, Figure 4 depicts a typical grid showing velocity measurement points along with horizontal and vertical directions. As part of the compound channel design, each point on the channel's plan could be accessed for measuring. A moveable

bridge could be used to collect the measurements. The width-ratio and aspect ratio of the channel are the important parameters in the study. A micro-pitot tube with a diameter of 4.77 mm was used to measure the flow grid's velocity. The order of maximum velocity for a given flow path was determined using a flow detector that has a minimum count of 0.1° . Use of circular scale and pointer configuration on flow direction sensor to measure pitot tube leg angle in relation to channel longitudinal direction. When combining the longitudinal velocity plot with the volumetric tank collection, the total discharge computed was within ± 3 percent of the actual data. This study used velocity data and a semi-log plot to predict channel bed and wall shear stresses. The boundary shear stresses were estimated using manometer measurements of the head differences of Preston tube methods utilising Patel's (1965) relationship. Shear values were corrected by comparing them to the equivalent values calculated using the energy gradient technique. As a result, the findings were constantly within $\pm 3\%$ of the value. Pitot tube shear stress computation is superior to ADV in terms of accuracy, according to lab data analysis. For one thing, measuring velocity at the boundary with ADV is never trustworthy. In addition, ADV has certain limits when it comes to measuring the velocity near the bed. It can penetrate up to five centimetres below the top edge. Consequently, the micro-ADV down probe was unable to reach a distance of 5 cm from the free surface. A Pitot tube was used to measure the transitory drop in the concentration of the compound. The validity of this technique was tested using the energy gradient methodology, which confirmed that it was valid. (Naik and Khatua 2016)

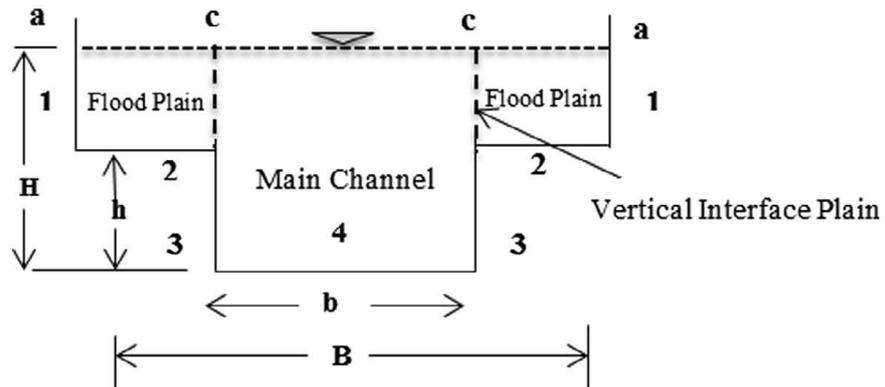


Fig. 1 Cross section of a compound channel with geometrical parameters (Naik and Khatua 2016)

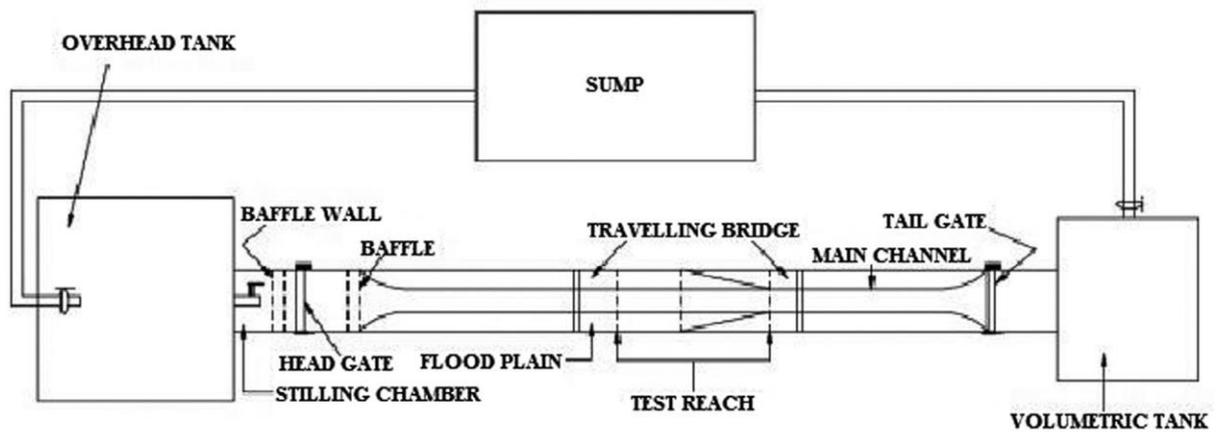


Fig. 2 Plan view of experimental setup (Naik and Khatua 2016)

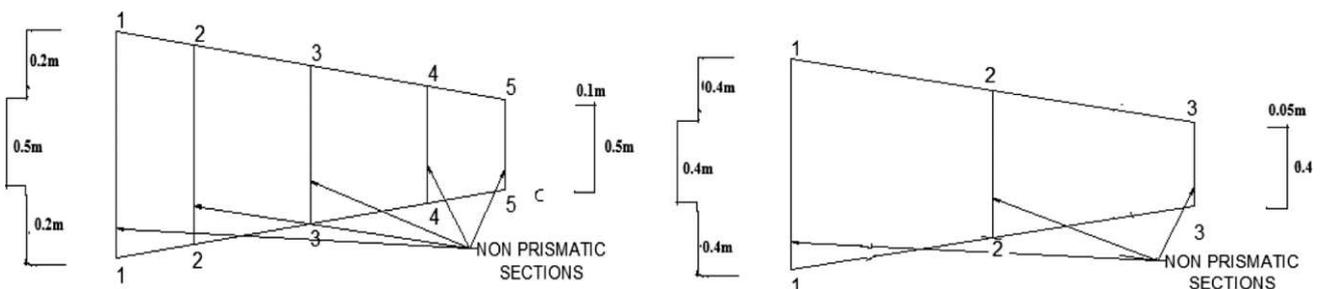


Fig. 3 Plan view of different test reaches with cross-sectional dimensions of both Naik & Rezaei channels (Naik and Khatua 2016)

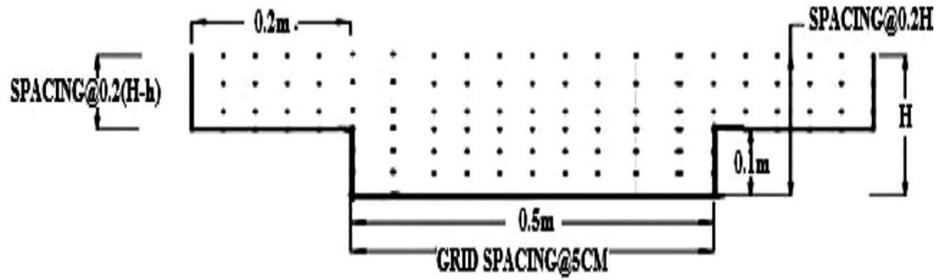


Fig. 4 Typical grid showing the arrangement of velocity measurement points at the test section (Naik and Khatua 2016)

2.2 Theoretical Background

Many researchers have simulated boundary shear stress in compound open channels. For compound sections with various characteristics, the formulae for the percentage boundary shear force that is carried out by floodplain are shown below.

1. A percent S_{fp} equation was established by Knight and Demetriou (1983) for compound channels with α up to 4 and expressed as

$$\%S_{fp} = 48(\alpha - 0.8)^{0.289}(2\beta)^m \quad (1)$$

2. In order to account for channels with non-homogeneous roughness, Knight and Hamed (1984) developed

$$\%S_{fp} = 48(\alpha - 0.8)^{0.289}(2\beta)^m \{1 + 1.02\sqrt{\beta \log \gamma}\} \quad (2)$$

The exponent m can be evaluated from the relation

$$m = \frac{1}{[0.75e^{0.38\alpha}]} \quad (3)$$

γ is the ratio of roughness between floodplain and main channel.

3. One of the models proposed by Khatua and Patra (2007) is valid up to $\alpha = 5.25$ and is defined as

$$\%S_{fp} = 1.23(\beta)^{0.1833}(38\ln\alpha + 3.6262)\{1 + 1.02\sqrt{\beta \log \gamma}\} \quad (4)$$

4. The preceding expression was modified by Khatua et al. (2012) for α up to 6.67

$$\%S_{fp} = 4.1045 \left(\frac{100\beta(\alpha-1)}{1+\beta(\alpha-1)} \right)^{0.6917} \quad (5)$$

5. Mohanty and Khatua (2014) established a novel relation for $\alpha \leq 12$

$$\%S_{fp} = 3.3254 \left(\frac{100 \{ \beta \alpha \delta - \beta (\delta + 2s) \}}{\beta \alpha \delta + (1 - \beta)(\delta + s)} \right)^{0.7467} \{ 1 + 1.02 \sqrt{\beta \log \gamma} \} \quad (6)$$

6. Devi et al. (2016) created a model for ranges of width ratio $3 \leq \alpha \leq 12$, relative flow depth $0.10 \leq \beta \leq 0.50$ and aspect ratio of main channel $3 \leq \delta \leq 10$

$$\%S_{fp} = 3.576 \left\{ \frac{100 \beta \left(\alpha - 1 - \frac{2.5s}{\delta} + \frac{0.5s}{\delta^*} \right)}{\left(1 + \frac{s}{\delta^*} \right) + \beta \left(\alpha - 1 - \frac{2s}{\delta} \right)} \right\} \quad (7)$$

7. The equation was derived by Naik and Khatua (2016) for converging compound channels with the following width ratio, $\alpha = 1.8$ and aspect ratio, $\delta = 5$.

$$\%S_{fp} = 18.505 + 62.140(\beta)^{0.631} - 24.42(X_r) + 1.38(\theta) \quad (8)$$

2.3 Development of model for boundary shear stress estimation

In open channel flow, the boundary shear per unit length (S_F) is usually considered to be uniform. That's why $S_F = \rho g A S$, where g represents the acceleration due to gravity, ρ denotes density of water and S is the bed slope for a particular channel. In terms of flow depth, the only variable that changes is flow area (A). Therefore, shear force is dependent on flow area. For instance, $\%A_{fp} = 100A_{fp}/A$ represents the proportion of the compound channel's area filled by floodplain subsections generated by vertical interfaces (Figure 3), where A_{fp} is the equivalent area by floodplain and A is the total area of the compound channel. As a result, a functional link has been established between $\%S_{fp}$ and $\%A_{fp}$. Curve fitting between $\%A_{fp}$ and $\%S_{fp}$ yielded the most significant regression coefficient, resulting in the new equation. For compound channels with converging flood plains, we attempted to establish an equation of $\%S_{fp}$ with α and β . In previous studies, different investigators, i.e., Equations (1) to (8), show that $\%S_{fp} = F(\alpha, \beta, \delta)$ for the prismatic compound channel, where F is the functional symbol.

However, when all equations are compared to compound open channels with narrowing sections, substantial errors are found. As a result, the change in $\%S_{fp}$ concerning several non-dimensional characteristics of a non-prismatic compound channel has been investigated here.

The percentage of shear force carried by floodplain ($\%S_{fp}$) for non-prismatic sections was calculated using a diverse set of experimental data, including three different types of converging compound channels from Naik and Khatua (2016), and three series of converging compound channel data from Rezaei (2006) (details of the datasets are given in Table 1). Both the main channel and the flood plain subsections of these compound channels exhibit uniform roughness. For all of these smooth surfaces, Manning's n values are set to 0.01. In a compound channel with narrow floodplains, the boundary shear distribution changes from section to section.

In order to estimate the boundary shear stress distribution of such compound channels, two additional factors (θ and X_r) are taken into account in addition to the two (α and β) already mentioned. By taking into account all factors, the author attempts to develop an equation for the boundary shear force carried by floodplains in a converging compound channel using a Gene-expression programming (GEP) approach with the help of GeneXproTools 5.0 (2014), where the generation of models is chosen based on the fitness of the training and testing datasets. As part of GEP, selected models are replicated by utilising one or more genetic operators, such as mutation or recombination. Brief theoretical overviews of GEP are described by Azamathulla et al. (2013) and Mallick et al. (2020). It is assumed that the converging compound channel's boundary shear force is influenced by dimensionless factors such as width ratio, relative flow depth, converging angle, and relative distance. A variety of parameters are selected to allow the resulting model equations to be applied to a variety of compounds. To simplify the equation for a straight compound channel, assume the converging angle as zero and simplify the equation for the straight compound channel.

The required dimensionless equation can be written as

$$\%S_{fp} = F(\alpha, \beta, \theta, X_r) \quad (9)$$

where θ = Converging angle, X_r = Relative distance (x/L), and L = Non-prismatic length.

Table 1 Details of parameters of experimental channels used in the present analysis

Verified Test Channel	Type of Channel	Angle of Convergent (θ)	Longitudinal Slope (S)	Cross-sectional Geometry	Total Channel Width (B) (m)	Main Channel Width (b) (m)	Main Channel Depth (h) (m)	Converging Length (X_r) (m)
Rezaei (2006)	Converging	11.31°	0.002	Rectangular	1.2	0.398	0.05	2
Rezaei (2006)	Converging	3.81°	0.002	Rectangular	1.2	0.398	0.05	6
Rezaei (2006)	Converging	1.91°	0.002	Rectangular	1.2	0.398	0.05	6
Naik and Khatua (2016)	Converging	5°	0.0011	Rectangular	0.9	0.5	0.1	2.28
Naik and Khatua (2016)	Converging	9°	0.0011	Rectangular	0.9	0.5	0.1	1.26
Naik and Khatua (2016)	Converging	12.38°	0.0011	Rectangular	0.9	0.5	0.1	0.84

2.4 Development of GEP model

The relationship (Eq. 9) indicates the percentage boundary shear force carried by floodplains as a function of geometric and hydraulic variables. In this study, the modeling procedure uses $\%S_{fp}$ as the target value and the four independent factors (α , β , θ , X_r) as input variables discussed in Eq. (9). The model is constructed using four fundamental arithmetic operators (+, -, \times , /) and a fundamental mathematical function (e^x , \ln , x^2 , average, cube root, maximum of two, inverse, tangent). There are 112 data sets used in the modeling process, some of which are shown in Table 1. The data are randomly distributed for the two different phases of the modeling process. For the current study, 50% of the data is used for training, while the other 50% is used for testing. In this study, RMSE was the fitness function (E_i) and the fitness (f_i) was computed by Eq. 10 that yields the target value's total sum of errors. Starting with just one gene and two head sizes, the first model was constructed. The set of genes and heads were then increased one by one throughout each run, and the outcomes of the training and testing datasets were recorded. For head lengths of more than eight and more than three genes, the performance of the training and testing data phase did not improve significantly.

As a consequence, eight were chosen as head length for inclusion in the GEP model, with three genes per chromosome. Addition was used to bind three genes together. Testing revealed that after 4,50,000 generations, there had been no visible change in the fitness function value and

coefficient of determination of training and testing data, indicating that generations may have come to an end. Table 2 lists the key characteristics that influence the effectiveness of GEP modeling and are utilized in the construction of a model for predicting percent S_{fp} . After a lot of trial and error, the final model of the GEP was found to be an algebraic equation between output and input variables. GeneXproTools (2014), a powerful soft computing software suite, was employed in this research.

GEP is modelled in the following way in terms of analytical form as

$$\%S_{fp} = \left[10.25 + \tan \left[\text{Max} \left((\beta + 2.62), (\theta - 7.23) \right) + \sqrt[3]{3.39} \right] \right] + \left[(X_r + X_r) \times \left(\frac{X_r - 2.62}{\alpha} \right) \times \left(\frac{7.28}{\alpha} \right) \right] + \left[\left(\frac{\beta + \alpha}{2} \times \frac{X_r + 7.77}{2} \right) - \left(\frac{\alpha - 3.95}{2} + \frac{\alpha + 0.196}{2} \right) \right]^2 \quad (10)$$

Further the Eq. (10) is simplified as

$$\%S_{fp} = 10.25 + \tan \left[\text{Max} \left((\beta + 2.62), (\theta - 7.23) \right) + 1.50 \right] + \left[\frac{14.56X_r^2 - 38.14X_r}{\alpha^2} \right] + \left[\left(\frac{(\alpha + \beta) \times (X_r + 7.77)}{4} \right) - \left(\frac{\alpha - 3.95}{2} + \frac{\alpha + 0.196}{2} \right) \right]^2 \quad (11)$$

In Fig. 5, the GEP model for predicting $\%S_{fp}$ is represented by an expression tree (ET) representation as in Eq. (10). Thus, in Fig. 5, d0 denotes the α , d1 denotes the β , d2 denotes the θ , d3 denotes the X_r , and G1c0, G1c3, G1c7, G1c8 represent the numerical constants employed in the model's first gene. Similarly, G2c5, G2c7 and G3c0, G3c3, and G3c5 are the constants used in the model's second and third genes.

Table 2 Functional set and Operational Parameters Used in GEP Model

Description of Parameter	Parameter Setting
Function Set	$+, -, \times, /, e^x, \ln, x^2, \text{average}, \sqrt[3]{}, \text{max}(x,y), \text{inverse}, \tan(x)$
Number of Chromosomes	30
Head Size	8
Number of Genes	3

Gene Size	26
Linking Function	Addition
Fitness Function	RMSE
Program Size	42
Literals	12
Number of Generations	450000
Constants per Gene	10
Data Type	Floating-point
Mutation	0.00138
Inversion	0.00546
Gene recombination rate	0.00277
One-point recombination rate	0.00277
Two-point recombination rate	0.00277
Gene transposition rate	0.00277
Insertion sequence (IS) transposition rate	0.00546
Root insertion sequence (RIS) transposition rate	0.00546

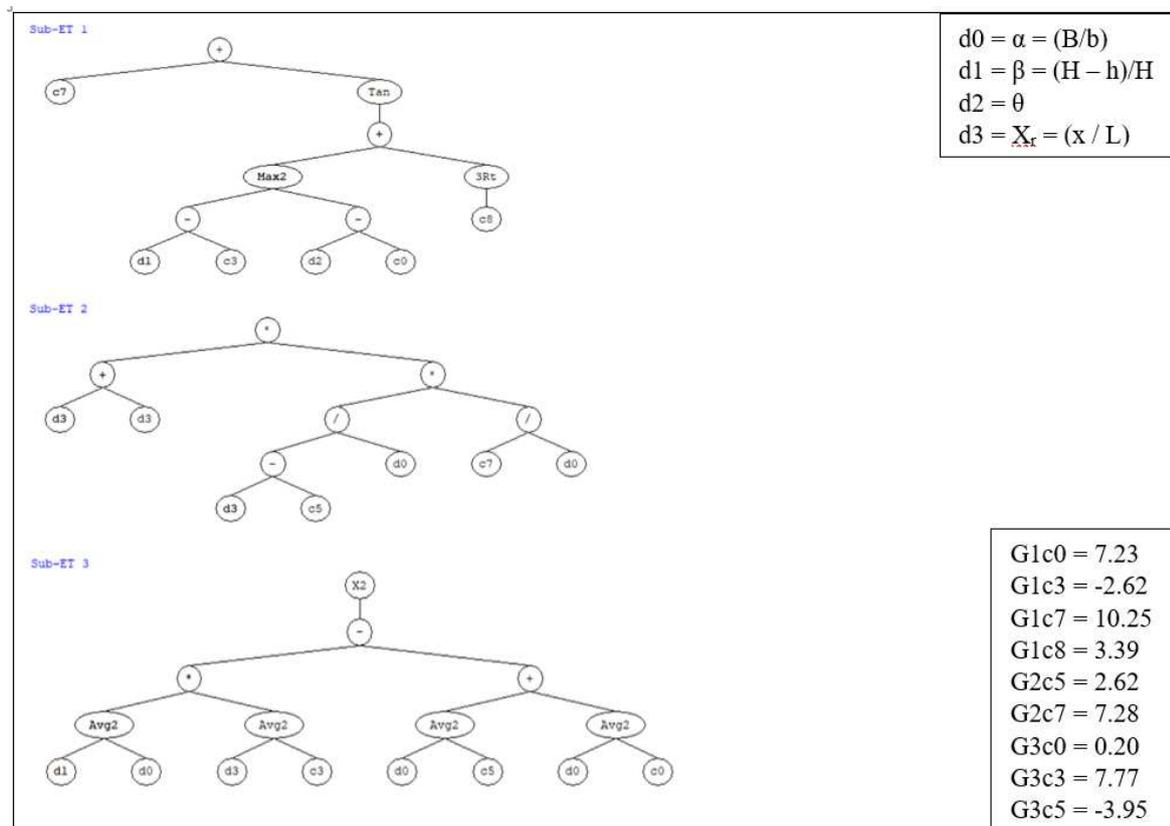


Fig. 5 Expression Tree (ET) for GEP formulation

3. Results and discussion

Figure 6 shows the findings of boundary shear stress distributions for different cross-sections of relative depth 0.15, 0.20, and 0.30 for the converging floodplain angle of 12.38° . The boundary shear stress distributions are generally symmetric in all sections, and steadily grow from sec-1 to sec-5, as seen in these figures. The boundary shear stress values rise as the flow depth increases. The boundary shear stress value is found to be maximum in the centre of the main channel and progressively falls towards the interface between the main channel and the floodplain in all sections. The boundary shear stress drops abruptly near the interface, then gradually declines until it reaches a minimum at both ends of floodplains. This is due to momentum transfer and increased interaction between the main channel and the floodplains. Figure 7 depicts the relationship between percent S_{fp} and relative flow depths at various converging angles for various cross sections. From the figure, it can be observed that the percent S_{fp} increases as the relative flow depth increases from section to section. Figure 8 shows the comparison for the training and testing data phases of the GEP method (Eq. 11) for the observed and predicted $\%S_{fp}$. The investigators utilized Eq. (11) to show performance matrices of the predicted $\%S_{fp}$ value to the observed value, as shown in Figure 9, for each data used in the training and testing data phases separately. For converging compound channels, the GEP method reveals a highly nonlinear relationship between $\%S_{fp}$ and the input parameters (α , β , θ , X_r) with high precision and relatively low residuals as shown in figure 10. For the training data ($R^2 = 0.96$ and $RMSE = 3.395$) and the testing data ($R^2 = 0.95$ and $RMSE = 4.022$), the GEP model produced the lowest errors. The GEP model can reflect the underlying phenomena of percentage shear force carried by floodplains, as seen in Figure 8. It says that selecting an input vector based on linear measurements between the variables of interest, which is typical, may nevertheless include some spurious features that contribute to the model's complexity.

Various types of error analysis, such as the coefficient of determination (R^2), mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean squared error (RMSE), are analyzed using the following equations for further testing of the accuracy of the developed model by GEP approach (Mohanta et al. 2018)

$$R^2 = \frac{\sum_{i=1}^N (a_i - \bar{a})^2 (p_i - \bar{p})^2}{\sum_{i=1}^N (a_i - \bar{a})^2 \sum_{i=1}^N (p_i - \bar{p})^2} \quad (11)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |p_i - a_i| \quad (12)$$

$$MAPE(\%) = \frac{1}{N} \sum_{i=1}^N \left(\frac{|p_i - a_i|}{a_i} \times 100 \right) \quad (13)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - a_i)^2} \quad (14)$$

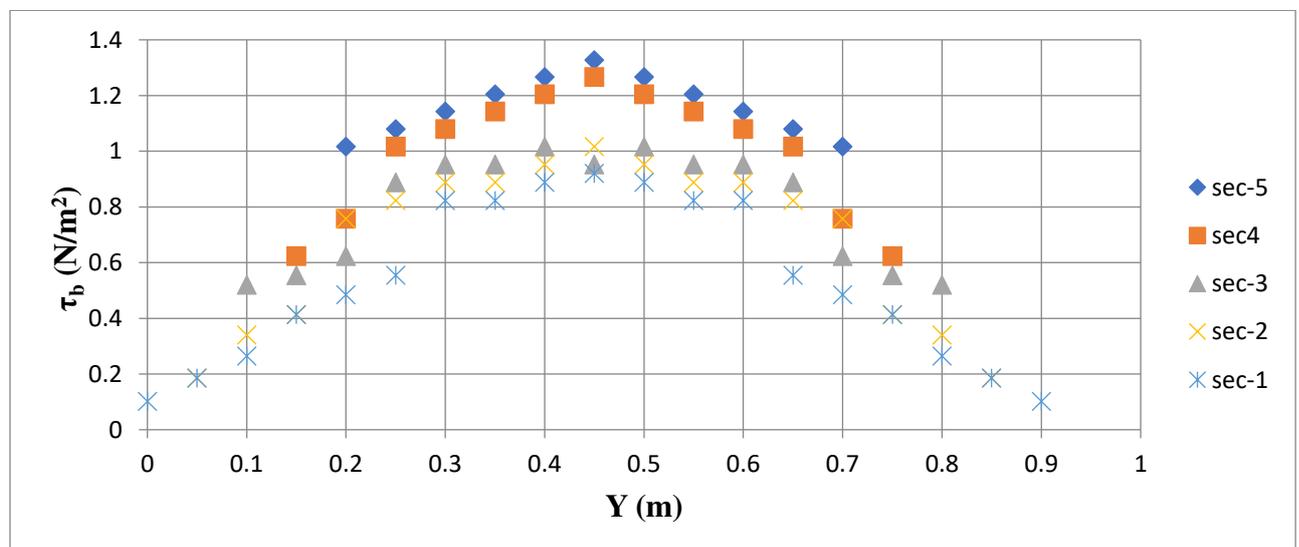
where a and p are the actual and predicted values, respectively, \bar{a} and \bar{p} are the mean of actual and predicted values, respectively and N is the number of datasets.

For both the training and testing datasets, performance assessments shown in table 3 in terms of R^2 (coefficient of determination), MAE (mean absolute error), RMSE (root mean square error), and MAPE (mean absolute percentage error) for the created GEP model were calculated. The model's adequacy is demonstrated by the R^2 value for the two datasets (training = 0.96 and testing = 0.95) in the current method. The RMSE values for the training and testing data sets were 3.395 and 4.022, respectively, according to the GEP model. MAE and MAPE values for training data were 2.70 and 10.45 percent, respectively, while MAE and MAPE values for testing data were 2.99 and 13.79 percent, indicating the predicted model's performance and accuracy. As a result, the proposed GEP is a reliable approach for predicting $\%S_{fp}$ with a high generalization capacity and does not exhibit overtraining.

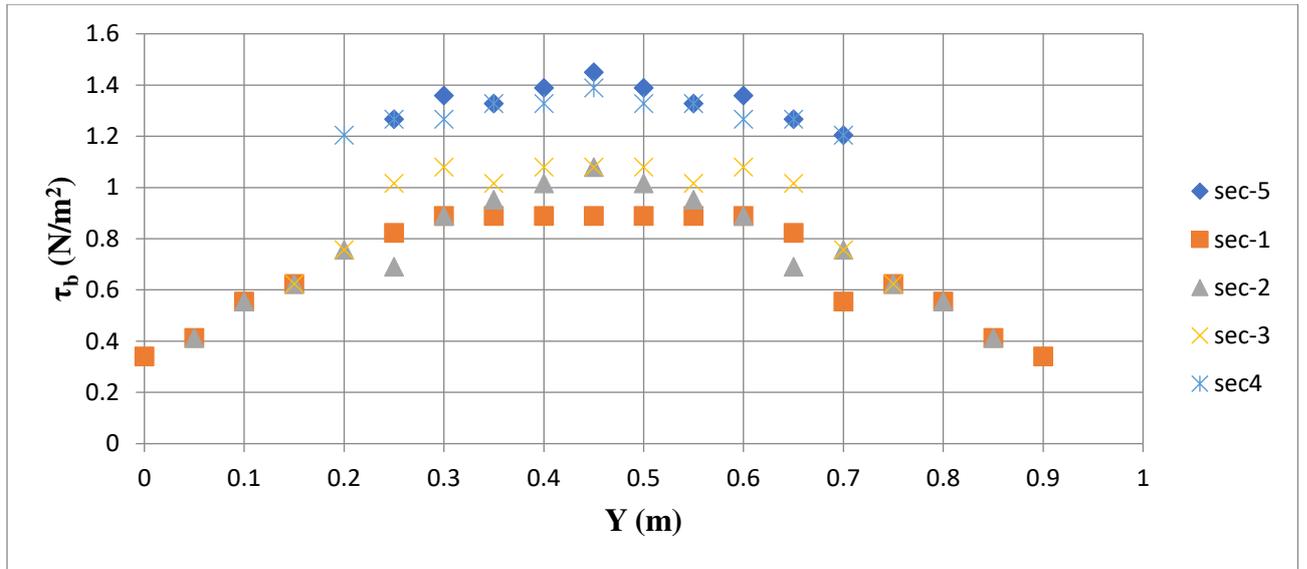
Figure 11 depicts the difference between the estimated ($\%S_{fp}$) values for compound channels with converging flood plains and the actual values for all six types of channels. When comparing the current model to the prior models for both Naik and Khatua (2016) and Rezaei

(2006) channels, the error in the estimation of ($\%S_{fp}$) is lower using the GEP approach, as shown in figure 12. Different conventional methods are evaluated using the novel equation for the flow scenarios examined in Naik and Khatua (2016) and Rezaei (2006) channels. Naik and Khatua (2016), Khatua et al. (2012), Knight and Hamed (1984), and Mohanty et al. (2014) are among the techniques evaluated using the novel equation. Table 4 compares statistical error analysis of several approaches, with the new GEP method appearing to be the best.

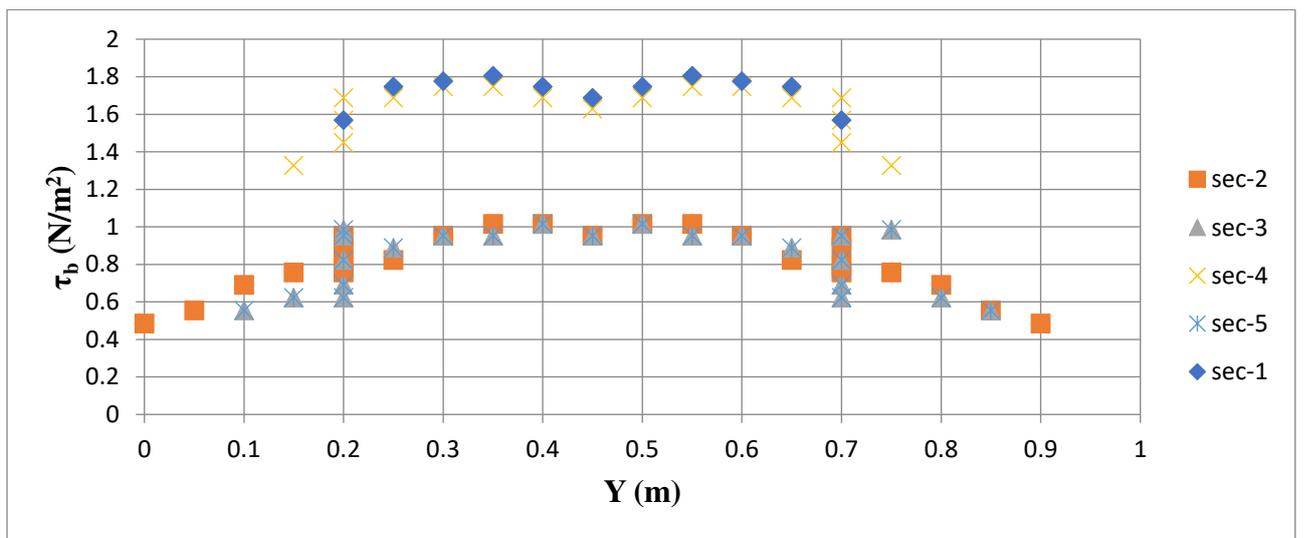
The suggested GEP model has a substantial benefit over traditional regression-based models (standard equations). It can project data into a high-dimensional feature space, where many approaches (discussed in the preceding section) may be utilized to identify relationships in the data. However, because the mapping is so broad, the relationships are very general. To decrease related uncertainties, physical-based, process-oriented riverine hydraulics research is required to develop techniques for predicting boundary shear stresses, in-channel and overbank roughness coefficients, and other variables impacting energy losses, velocity, and discharge. These issues have considerable influence on water resources and related studies, such as flood inundation studies, flood hazard warnings, hydraulic structure design and safety in channels and floodplains, sediment transport, geomorphic, and other studies that require reliable riverine hydraulic data.



(a)

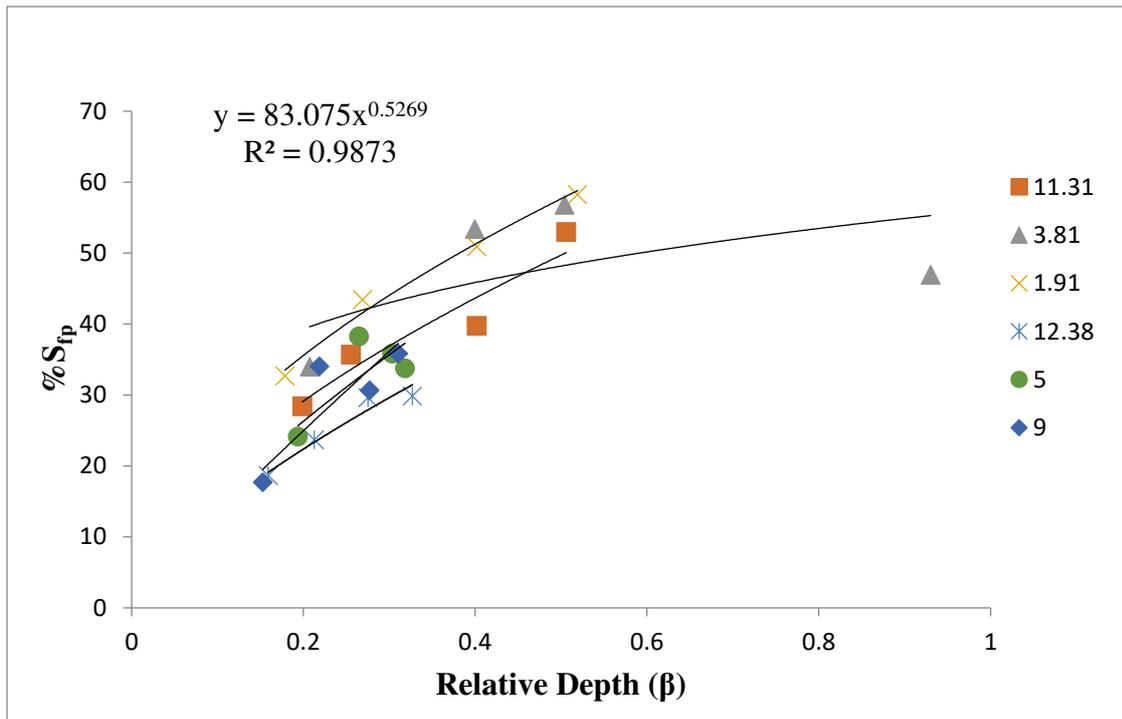


(b)

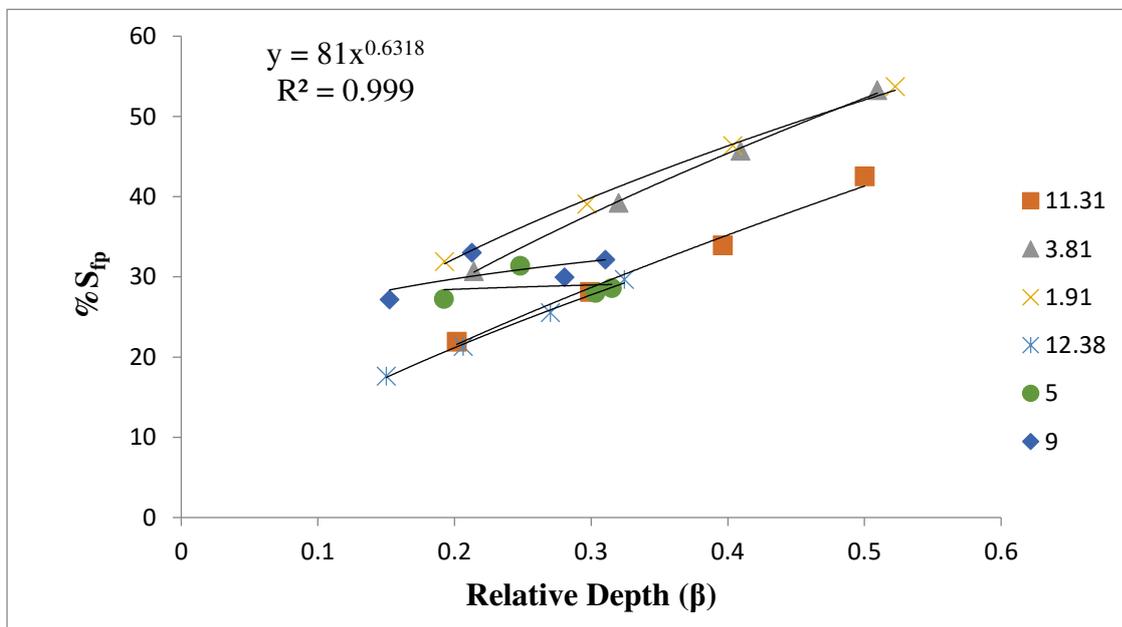


(c)

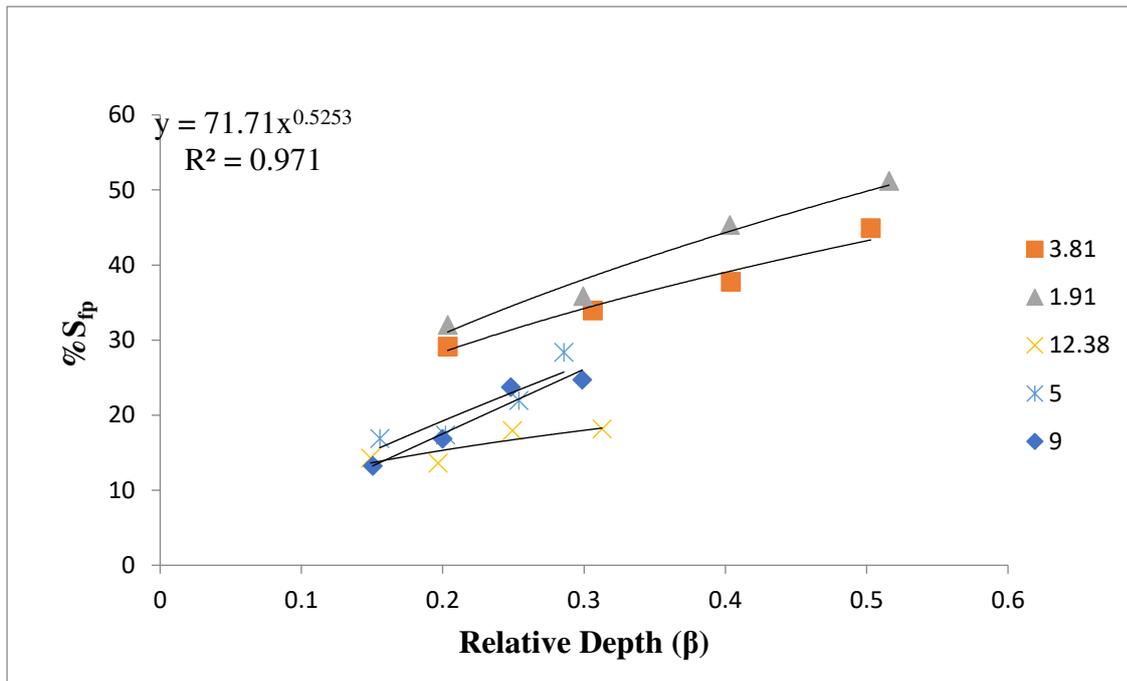
Fig. 6 Variation of boundary shear stress along the width of channel at converging angle of 12.38° for relative depth (a) 0.15 (b) 0.20 (c) 0.30



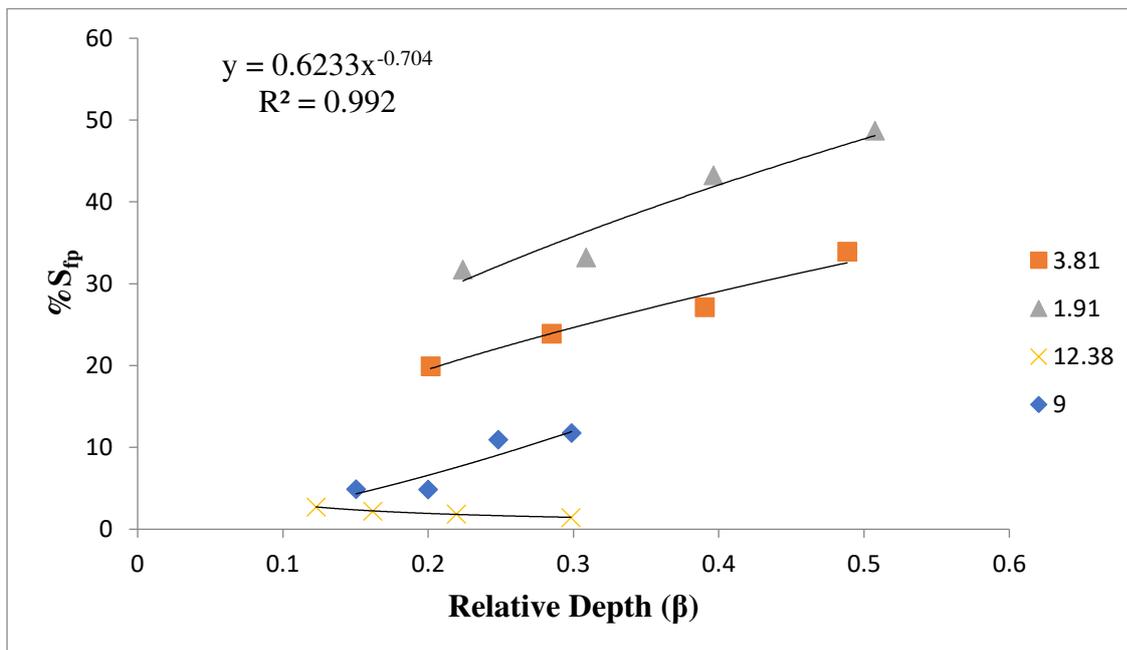
(a)



(b)

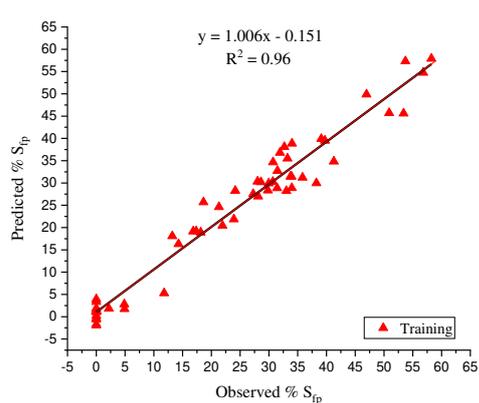


(c)

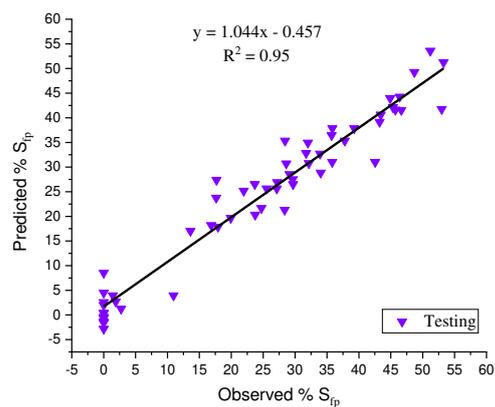


(d)

Fig. 7 Variation of percentage floodplain shear force with relative depth at various converging angles for different nonprismatic sections (a) section 1 (b) section 2 (c) section 3 (d) section 4

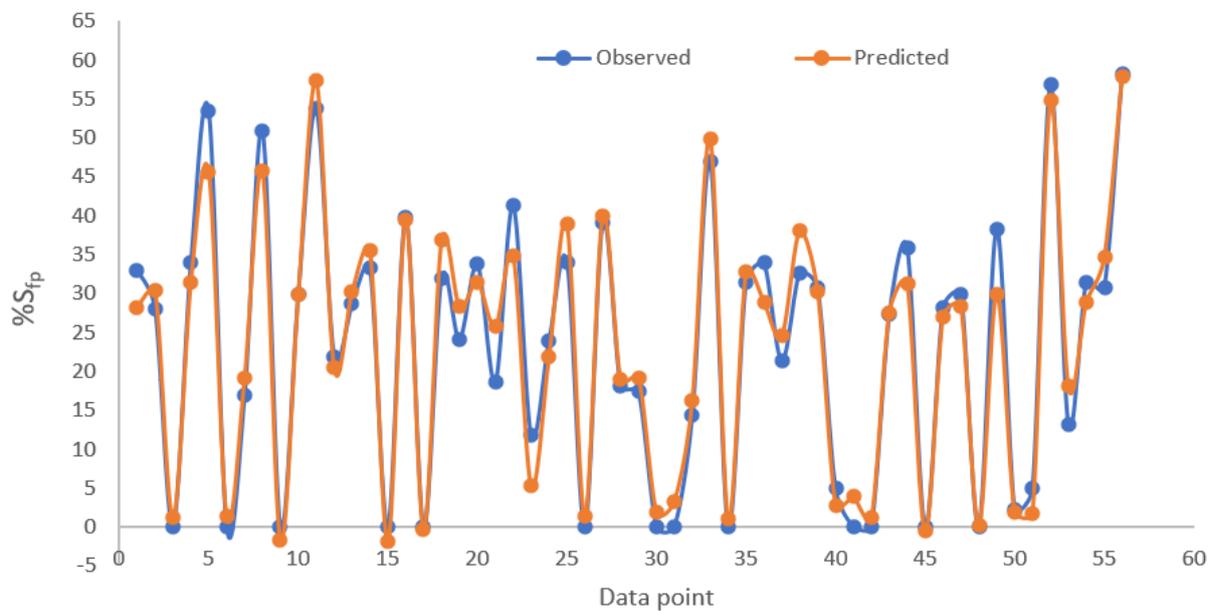


(a)

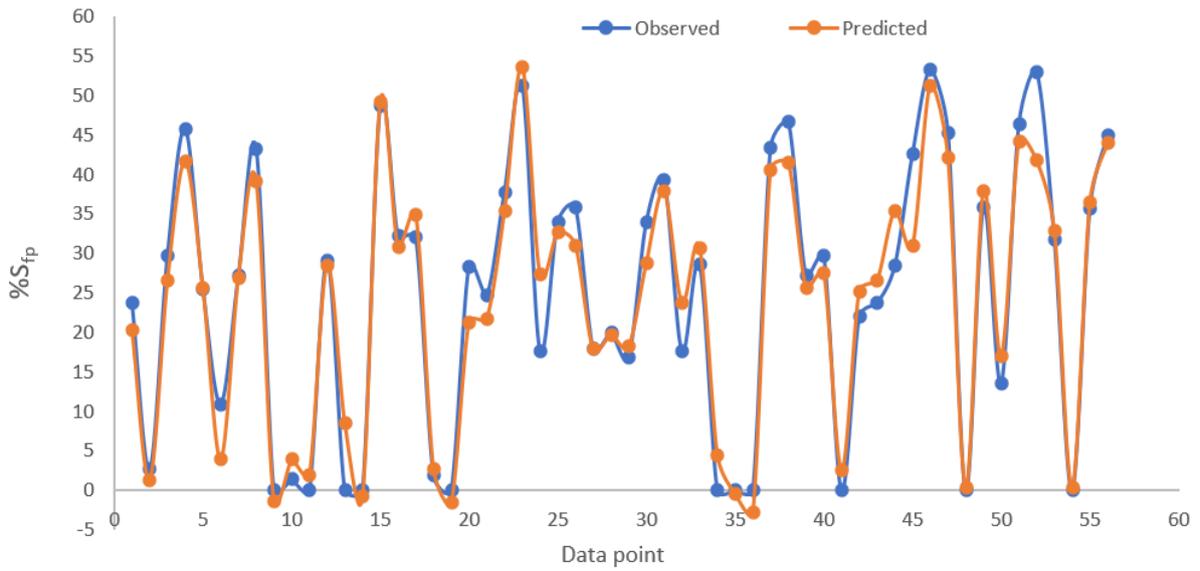


(b)

Fig. 8 Comparison between observed and predicted % S_{fp} for the different data phase (a) Training phase (b) Testing phase

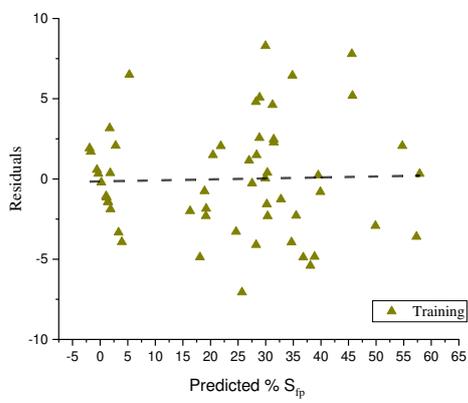


(a)

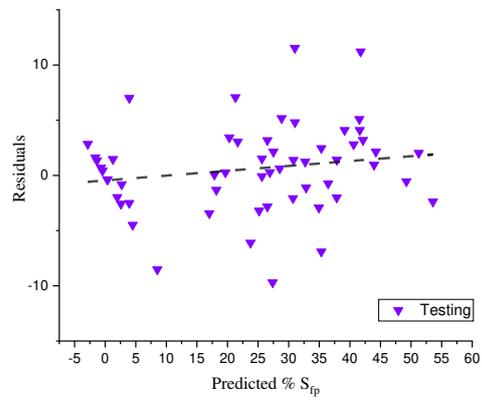


(b)

Fig. 9 Performance matrices for the prediction of the %S_{fp} for the different data phase (a) Training phase (b) Testing phase



(a)



(b)

Fig. 10 Residual plots for predicted %S_{fp} for the different data phase (a) Training phase (b) Testing phase

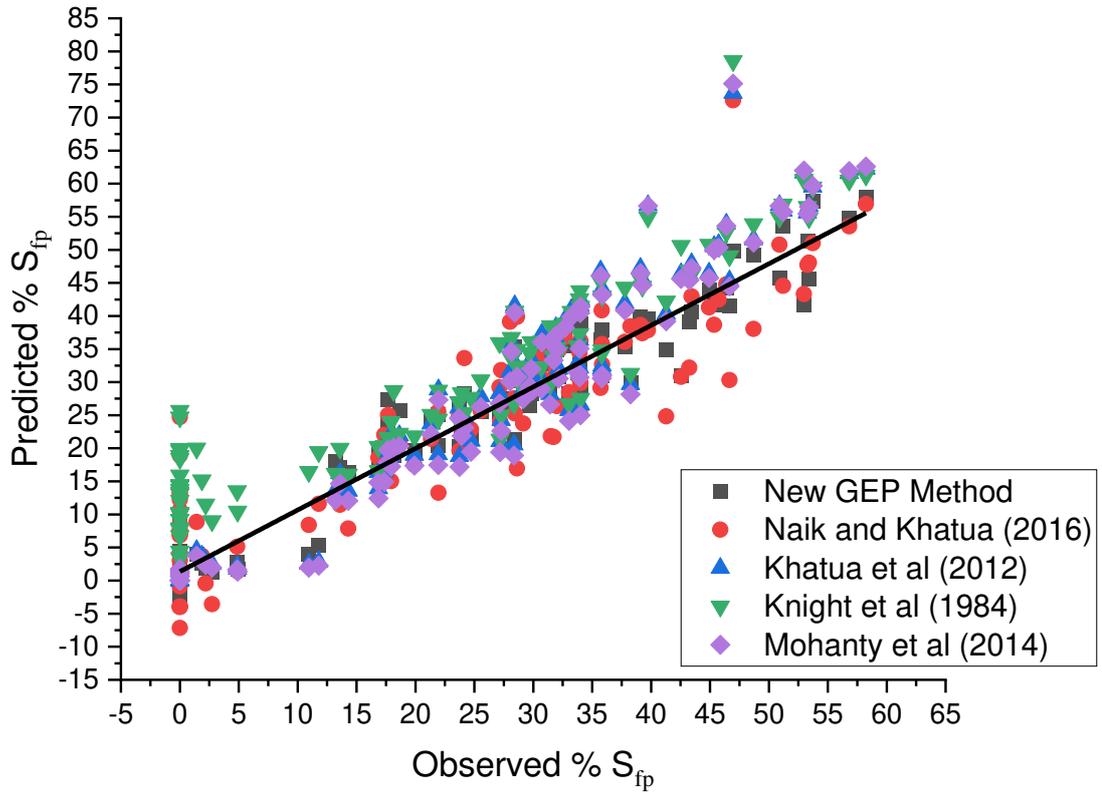


Fig. 11 Comparison of predicted value of $\%S_{fp}$ for various models

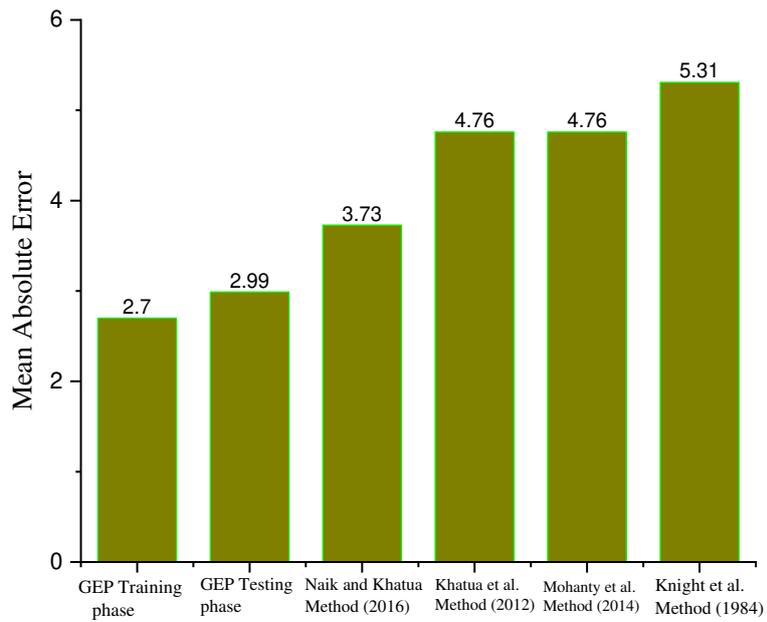


Fig. 12 Mean absolute error for different methods

Table 3 Performance evaluation of predicted %S_{fp} by GEP model for training and testing dataset

Dataset	R ²	Correlation Coefficient	MAE	RMSE	MAPE(%)
Training	0.96	0.98	2.70	3.395	10.45
Testing	0.95	0.97	2.99	4.022	13.79

Table 4 Statistical error analysis of different methods

Statistical parameters	New GEP Method		Naik and Khatua Method	Khatua et al. Method	Knight et al. method	Mohanty et al. method
	Training	Testing				
MSE	11.53	16.18	28.78	35.62	45.33	39.95
RMSE	3.395	4.022	5.36	5.97	6.73	6.32
MAE	2.70	2.99	3.73	4.76	5.31	4.76
MAPE(%)	10.45	13.79	13.43	17.20	21.63	17.47

4. Conclusions

Gene expression programming (GEP) is a computer programming language that leverages a fixed-length gene expression representation to encapsulate computer programs and rapidly discover succinct and understandable solutions. In this study, the GEP is utilized to estimate the percentage shear force carried by the floodplains in converging compound channels. The GEP model can accurately forecast boundary shear stress with a short run time, which is dependent on the number of generations. The author conducts experimental investigations using a novel set of converging compound channels, including a converging compound channel with varied angles of converging floodplains ranging from 1.91° to 12.38° and varying width ratios ranging from 1.0 to 3.0. Using 112 high-quality data from experimental compound channels with converging floodplains, the GEP model was constructed to determine the exact

values of the percentage shear force. The following are the conclusions derived as a result of this study:

The proposed model appears to be influenced by parameters such as width ratio, relative flow depth, converging angle, and relative distance. The GEP approach's suggested model is shown to be highly suitable for all of these kinds of channel systems, covering various laboratory models. In comparison to other flow parameters such as converging angle and relative distance, the relative flow depth and width ratio was found to be more appropriate in computing percentage shear force. In terms of R^2 , MAE, RMSE, and MAPE for different series of datasets, the established GEP model show superior outcomes in comparison to other methods as Naik and Khatua [48], Khatua et al. [46], Knight and Hamed [17], Mohanty et al. [54]. The GEP approach's compatibility is determined by the model's mean percentage of error and the approaches' suitability. Within a close range of non-dimensional parameters examined in this study, the findings clearly demonstrate the effectiveness of the GEP model and its potential utility for real applications. The results of this investigation show that the GEP model is more beneficial in any circumstance with no limits.

Acknowledgement

The authors would like to convey their heartfelt gratitude to the anonymous editor and reviewers for their time and effort in reviewing and providing helpful feedback on the article.

Notation

The following symbols are used in this paper:

A = total area of the compound channel;

A_{fp} = corresponding area by floodplain;

B = total width of compound channel;

b = total width of the main channel;

H = bank full depth;

h = total height of the main channel;

L = converging length;

S_0 = longitudinal bed slope;

$\%S_{fp}$ = percentage shear force carried by the floodplains;

S_F = total shear force of the compound channel;

X_r = relative distance(x/L);

x = distance between two consecutive sections;

α = width ratio (B/b);

β = relative flow depth [$(H - h)/H$];

δ = main channel aspect ratio (b/h);

δ^* = flow aspect ratio (b/H);

γ = roughness ratio;

θ = converging angle;

ρ = density of water;

g = acceleration due to gravity;

n = Manning's roughness coefficient;

s = side slope of main channel;

Declarations:

Ethical Approval: All procedures involving human participants in this investigation were carried out in conformity with ethical guidelines.

Consent to participate: The authors are voluntarily participating in this research.

Consent to publish: The authors have willingly decided to have their work published in the journal.

Authors Contributions: Bandita Naik carried out the experimental study, Vijay Kaushik performed the genetic modelling of the experimental data and prepared the manuscript, and Munendra Kumar assisted with the findings and preparation of the manuscript.

Funding: Not Applicable

Competing Interests: The authors reported no potential conflicts of interest.

Availability of data and materials: All the relevant data is incorporated in the manuscript.

References

- [1] Khuntia JR, Devi K, Khatua KK (2018) Boundary Shear Stress Distribution in Straight Compound Channel Flow Using Artificial Neural Network. *Journal of Hydrologic Engineering*, 23(5), 04018014. doi:10.1061/(asce)he.1943-5584.0001651
- [2] Cruff RW (1965) Cross-channel transfer of linear momentum in smooth rectangular channels. No. 1592-B, USGPO, Washington, DC.
- [3] Ghosh SN, Roy N (1970) Boundary shear distribution in open channel flow. *J. Hydraul. Div.*, 96(4), 967–994.
- [4] Myers WRC (1978) Momentum transfer in a compound channel. *J. Hydraul. Res.*, 16(2), 139–150.
- [5] Knight DW, Macdonald JA (1979) Open channel flow with varying bed roughness. *J. Hydraul. Div.*, 105(9), 1167–1183.
- [6] Knight DW (1981) Boundary shear in smooth and rough channels. *J. Hydraul. Div.*, 107(7), 839–851.
- [7] Noutsopoulos GC, Hadjipanous PA (1982) Boundary shear in smooth and rough channels-discussion. *J. Hydraul. Div.*, 108(6), 809–812.
- [8] Knight DW, Demetriou JD, Hamed ME (1984) Boundary shear in smooth rectangular channels. *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(1984)110:4(405), 405–422.

- [9] Knight DW, Patel HS (1985) Boundary shear in smooth rectangular ducts. *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(1985)111:1(29), 29–47.
- [10] Nezu I, Tominaga A, Nakagawa H (1993) Field measurements of secondary currents in straight rivers. *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(1993)119:5(598), 598–614.
- [11] Rhodes DG, Knight DW (1994) Distribution of shear force on boundary of smooth rectangular duct. *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(1994)120:7(787), 787–807.
- [12] Ackerman JD, Hoover TM (2001) Measurement of local bed shear stress in streams using a Preston-static tube. *Limnol. Oceanogr.*, 46(8), 2080–2087.
- [13] Tayfur G (2002) Artificial neural networks for sheet sediment transport. *Hydrol. Sci. J.*, 47(6), 879–892.
- [14] Yang SQ, Lim SY (1998) Boundary shear stress distributions in smooth rectangular open channel flows. *Proc. Inst. Civ. Eng. Water Marit. Energy*, 130(3), 163–173.
- [15] Sellin RHJ (1964) A laboratory investigation into the interaction between the flow in the channel of a river and that over its flood plain. *La Houille Blanche*, 7, 793–802.
- [16] Ghosh S, Jena SB (1971) Boundary shear stress distribution in open channel compound. *Proc. Inst. Civil Eng.*, 49, 417–430.
- [17] Knight DW, Hamed ME (1984) Boundary shear in symmetrical compound channels. *J. Hydraul. Eng.*, 110(10), 1412–1430.
- [18] Patra KC, Kar SK, Bhattacharya AK (2004) Flow and velocity distribution in meandering compound channels. *J. Hydraul. Eng.*, 130(5), 398–411.
- [19] Khatua KK, Patra KC (2007) Boundary shear stress distribution in compound open channel flow. *ISH J. Hydraul. Eng.*, 13(3), 39–54.

- [20] Mohanty PK, Khatua KK (2014) Estimation of discharge and its distribution in compound channels. *Journal of Hydrodynamics, Ser.B*, 26(1), 144–154.
- [21] Bousmar D, Zech Y (1999) Momentum transfer for practical flow computation in compound channels. *J. Hydraul. Eng.*, 125(7), 696–706.
- [22] Bousmar D, Wilkin N, Jacquemart JH, Zech Y (2004) Overbank flow in symmetrically narrowing floodplains. *J. Hydraul. Eng.*, 130(4), 305–312.
- [23] Proust S, Riviere N, Bousmar D, Paquier A, Zech Y (2006) Flow in compound channel with abrupt floodplain contraction. *J. Hydraul. Eng.*, 132(9), 958–970.
- [24] Najafzadeh M, Zahiri A (2015) Neuro-fuzzy GMDH-based evolutionary algorithms to predict flow discharge in straight compound channels. *J Hydrol Eng* 20(12):04015035.
- [25] Parsaie A, Yonesi H, Najafian S (2017) Prediction of flow discharge in compound open channels using adaptive neuro fuzzy inference system method. *Flow Meas Instrum* 54:288–297.
- [26] Sahu M, Khatua KK, Mahapatra SS (2011) A neural network approach for prediction of discharge in straight compound open channel flow. *Flow Meas Instrum* 22(5):438–446.
- [27] Seckin G (2004) A comparison of one-dimensional methods for estimating discharge capacity of straight compound channels. *Can J Civ Eng* 31(4):619–631.
- [28] Unal B, Mamak M, Seckin G, Cobaner M (2010) Comparison of an ANN approach with 1-D and 2-D methods for estimating discharge capacity of straight compound channels. *Adv Eng Softw* 41(2):120–129.
- [29] Zahiri A, Azamathulla HM (2014) Comparison between linear genetic programming and M5 tree models to predict flow discharge in compound channels. *Neural Comput Appl* 24(2):413–420.

- [30] Babovic V, Keijzer M (2002) Rainfall runoff modelling based on genetic programming. *Hydrol Res* 33(5):331–346.
- [31] Cousin N, Savic DA (1997) A rainfall-runoff model using genetic programming. Centre for Systems and Control Engineering, Rep. No. 97, 3.
- [32] Drecourt JP (1999) Application of neural networks and genetic programming to rainfall runoff modeling. *Water Resour Manage* 13(3):219–231.
- [33] Karimi S, Shiri J, Kisi O, Shiri AA (2015) Short-term and long-term streamflow prediction by using “wavelet-gene expression” programming approach. *ISH J Hydraul Eng* 22(2):148–162.
- [34] Savic DA, Walters GA, Davidson JW (1999) A genetic programming approach to rainfall-runoff modelling. *Water Resour Manage* 13(3):219–231.
- [35] Whigham PA, Crapper PF (1999) Time series modelling using genetic programming: an application to rainfall-runoff models. *Adv genet programm* 3:89–104.
- [36] Whigham PA, Crapper PF (2001) Modelling rainfall-runoff using genetic programming. *Math Comput Model* 33(6–7):707–721.
- [37] Azamathulla HM, Ahmad Z, Ghani AA (2013) An expert system for predicting Manning’s roughness coefficient in open channels by using gene expression programming. *Neural Comput Appl* 23(5):1343–1349.
- [38] Giustolisi O (2004) Using genetic programming to determine Chezy resistance coefficient in corrugated channels. *J Hydroinf* 6(3):157.
- [39] Guven A, Aytek A (2009) New approach for stage-discharge relationship: gene-expression programming. *J Hydrol Eng* 14(8):812–820.
- [40] Guven A, Gunal M (2008) Genetic programming approach for prediction of local scour downstream of hydraulic structures. *J Irrig Drain Eng* 134(2):241–249.

- [41] Harris EL, Babovic V, Falconer RA (2003) Velocity predictions in compound channels with vegetated floodplains using genetic programming. *Int J River Basin Manage* 1(2):117–123.
- [42] Pradhan A, Khatua KK (2017b) Gene expression programming to predict Manning's n in meandering flows. *Can J Civ Eng* 45(4):304–313.
- [43] Das BS, Devi K, Khatua KK (2019) Prediction of discharge in converging and diverging compound channel by gene expression programming. *J Hydraul Eng.* DOI:10.1080/09715010.2018.1558116
- [44] Mohanta A, Patra KC (2021) Gene-expression programming for calculating discharge in meandering compound channels. *Sustain. Water Resour. Manag.* 7, 33. <https://doi.org/10.1007/s40899-021-00504-0>
- [45] Knight DW, Demetriou JD (1983) Flood plain and main channel flow interaction. *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(1983) 109:8(1073), 1073–1092.
- [46] Khatua KK, Patra KC, Mohanty PK (2012) Stage-discharge prediction for straight and smooth compound channels with wide floodplains. *J. Hydraul. Eng.*, 10.1061/(ASCE)HY.1943-7900.0000491, 93–99.
- [47] Devi K, Khatua KK, Das BS (2016) Apparent shear in an asymmetric compound channel. *River Flow 2016: Iowa City, CRC Press, Boca Raton, FL*, 48–56.
- [48] Naik B, Khatua KK (2016) Boundary shear stress distribution for a converging compound channel. *ISH Journal of Hydraulic Engineering*, 22(2), 212–219. doi:10.1080/09715010.2016.1165633
- [49] Rezaei B (2006) Overbank flow in compound channels with prismatic and non-prismatic floodplains. PhD thesis, University of Birmingham, UK.
- [50] Patel VC (1965) Calibration of the Preston tube and limitations on its use in pressure gradients. *J. Fluid Mech.*, 231, 85–208. Cambridge University Press.

[51] Gepsoft G (2014) Version 5.0.

[52] Mallick M, Mohanta A, Kumar A, Patra KC (2020) Gene-expression programming for the assessment of surface mean pressure coefficient on building surfaces. *Build Simul* 13:401–418.

[53] Mohanta A, Patra KC, Sahoo BB (2018) Anticipate Manning's coefficient in meandering compound channels. *Hydrology* 5(3):47.

[54] Mohanty PK, Khatua KK, Dash SS (2014) Flow prediction in two stage wide compound channels. *ISH J. Hydraul. Eng.*