Applied Soft Computing 35 (2015) 618–628

Contents lists available at ScienceDirect

Applied Soft Computing

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

Gene expression programming to predict the discharge coefficient in rectangular side weirs

Isa Ebtehaj [a,b], Hossein Bonakdari [a,b,*], Amir Hossein Zaji [a,b], Hamed Azimi [a,b], Ali Sharifi [c]

[a] Department of Civil Engineering, Razi University, Kermanshah, Iran
[b] Water and Wastewater Research Center, Razi University, Kermanshah, Iran
[c] Department of Statistics, Razi University, Kermanshah, Iran

ARTICLE INFO

Article history:
Received 10 August 2014
Received in revised form 5 March 2015
Accepted 2 July 2015
Available online 10 July 2015

Keywords:
Discharge coefficient
Gene expression programming (GEP)
Sensitivity analysis
Side weir

ABSTRACT

In this study, gene expression programming (GEP) is employed as a new method for estimating the side weir discharge coefficient. The accuracy of existing equations in evaluating the side weir discharge coefficient is first examined. Afterward, taking into consideration the dimensionless parameters that affect the estimation of this parameter and sensitivity analysis, five different models are presented. Coefficient determination (R^2), root mean square error (*RMSE*), mean absolute relative error (*MARE*), scatter index (*SI*) and *BIAS* are used for measuring the models' performance. Two sets of experimental data are applied to evaluate the models. According to the results obtained indicate that the model with Froude number (F_1), dimensionless weir length (b/B), ratio of weir length to depth of upstream flow (b/y_1), and ratio of weir height to its length (p/y_1) parameters of $R^2 = 0.947$, $MARE = 0.05$, $RMSE = 0.037$, $BIAS = 0.01$ and $SI = 0.067$, performed the best. Accordingly, this new equation proposed through GEP can be utilized for estimating the discharge coefficient in rectangular sharp-crested side weirs.

1. Introduction

Side weirs are used for regulating the extra flow within urban wastewater collection networks, drainage-irrigation systems, flood protection and environmental preservation projects. A side weir is commonly installed on the side of a main channel. The flow passing over a side weir is considered a spatially varied flow with decreasing discharge. Numerous analytical, theoretical and experimental research works have been conducted by different researchers on the hydraulic behavior of channels with side weirs. De Marchi [1] solved the equations governing the spatially varied flow in a rectangular channel with a side weir and presented an equation for calculating discharge per unit side weir length.

Researchers including Subramanya and Awasthy [2], Yu-Tech [3], Nandesamoorthy and Thomson [4], El-Khashab [5], Ranga Raju et al. [6], Hager [7], Cheong [8], Singh et al. [9], Swamee et al. [10], Jalili and Borghei [11], and Ghodsian [12] have worked in this area. In an experimentally study, Borghei et al. [13] formulated an equation to calculate the discharge coefficient of the flow passing over a rectangular side weir. Their discharge coefficient equation was presented as a function of the side weir's upstream Froude number. The flow passing through a side weir within a rectangular channel was studied experimentally by Ghodsian [14], who calculated the preliminary discharge coefficient for a supercritical flow regime in the main channel along a side weir. Yüksel [15] examined the effect of the specific energy variation on a lateral weir in a rectangular conduit. Khorchani and Blanpain [16], Ramamurthy et al. [17], Huagao et al. [18], Venutelli [19], Vatankhah and Bijankhan [20], Emiroglu et al. [21,22] and Castro-Orgaz and Hager [23] formulated different equations for calculating the side weir discharge coefficient. They obtained the side weir discharge coefficient as a function of the ratio of the head above the weir to the side weir's sill height and Froude number. Bagheri et al. [24] conducted an empirical study on the features of flow passing over a rectangular sharp-crested side weir on the side of a rectangular channel. They analyzed the laboratory results using the artificial neural network and found that the side weir's upstream Froude number is the most effective parameter affecting the discharge coefficient of a rectangular sharp-crested side weir. Emiroglu et al. [25] determined the discharge characteristics of a trapezoidal labyrinth side weir with one and two cycles in subcritical flow regime.

Soft computing, a powerful tool in modeling and solving complex nonlinear problems in different fields such as hydraulic engineering [26,27], river engineering [28], and sediment transport [29,30], has come to the attention of a number of researchers in

* Corresponding author at: Department of Civil Engineering, Razi University, Kermanshah, Iran. Tel.: +98 831 427 4537; fax: +98 831 428 3264.
E-mail address: bonakdari@yahoo.com (H. Bonakdari).

http://dx.doi.org/10.1016/j.asoc.2015.07.003
1568-4946/

recent years. By using neural networks and neuro-fuzzy technique, Kisi [31] estimated daily suspended sediment load. For flood prediction, Feng and Lu [32] used an artificial neural network model. Emiroglu et al. [21] estimated the discharge coefficient and flow capacity of triangular labyrinth side weirs located on a rectangular channel using the adaptive Neuro Fuzzy Inference System (ANFIS). They also presented an equation for calculating the discharge coefficient in this type of side weir. Bateni et al. [33] predicted aquifer hydraulic parameters using the Genetic Algorithm (GA) and Ant Colony Optimization (ACO) methods. By employing different neural network techniques such as Feed Forward Neural Networks (FFNN) and Radial Basis Neural Networks (RBNN), Bilhan et al. [34] predicted the discharge coefficient and flow capacity in a rectangular sharp-crested side weir located on the side of a straight channel. Emiroglu et al. [35] used Artificial Neural Network (ANN) to estimate the discharge coefficient of a triangular labyrinth side weir on the side of a rectangular channel under subcritical flow regime. Kisi et al. [36] estimated the features of the flow passing from the side weir of a triangular labyrinth using Radial Basis Neural Networks (RBNN) and Generalized Regression Neural Networks (GRNN) algorithms and GEP. Dursun et al. [37] obtained an equation for calculating the discharge coefficient in semi-elliptical side weirs located on a rectangular channel in subcritical flow condition using ANFIS modeling.

Recently, the GEP method has been applied to recognize the manner of nonlinear systems such as multisite land-use allocation [38], sediment transport in sewer pipe systems [39], fatigue modeling of composite materials [40], stage-discharge curve [41], fault diagnosis of centrifugal pumps [42], real parameter optimization [43] and the prediction of faulty modules [44].

In the present study, the discharge coefficient of a rectangular side weir is simulated using the GEP method. Modeling was done using two different datasets. Five different equations were presented with various input combinations in order to find the most appropriate combination for modeling rectangular sharp-crested side weirs located on the side of a straight rectangular channel in subcritical flow condition. Finally, the most appropriate GEP equation was compared with the equations reported in earlier studies.

2. Overview of gene expression programming (GEP)

The GEP method, an extension of genetic programing (GP), is an evolutionary artificial intelligence technique introduced by Ferreira [45]. GEP evolves computer programs with different shapes and lengths encoded in linear chromosomes with fixed size. Subsequently, the chromosomes' information is decoded into expression Trees – a process called translation. The encoding process is very straightforward, whereby the ET reads the chromosomes from left to right and from top to bottom. An example of chromosome translation to ET is presented in Fig. 1. Each gene starts in the first position. However, the termination point is not always the last gene position and Open Reading Frames (ORF) are used. ORFs were explained in detail by Ferreira [45].

ETs are complex computer programs that are commonly evolved for the purpose of solving a particular problem, and they are selected based on their being fit to resolve the problem. Varied genetic alterations discover the characteristics of the existing population and as a result it is fitted to the particular problem that was meant to be solved. This means there is a suitable solution to solve the intended problem provided there is sufficient time and accurate particulars [46,47].

According to Fig. 2, the GEP procedure comprises some major steps. At the beginning, chromosomes from the initial population are generated randomly. Then the chromosomes are expressed

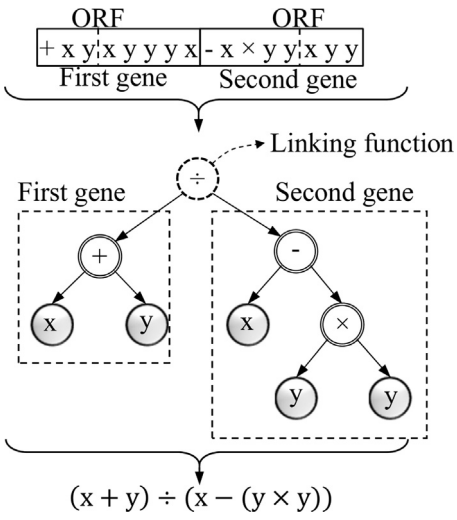

Fig. 1. GEP chromosome representation.

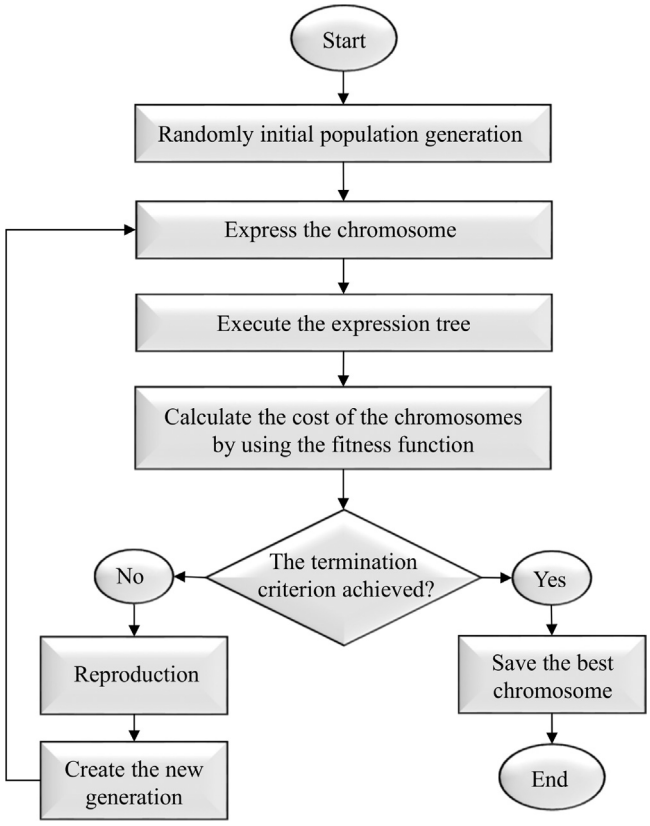

Fig. 2. Algorithm of gene expression programming.

and according to the considered fitness function, the cost of each chromosome is evaluated. The chromosomes are then selected keeping in view their cost to be produced again or to be modified. Providing new individuals through a similar process including genome expression, conformation of the environmental selection, and reproduction with modification comprises the next step. The mentioned process is repeated until the considered number of generations or acceptable model error is reached [48].

The chromosomes in each generation are optimized by genetic operators. The genetic operators applied in the GEP method are mutation, inversion, Insertion Sequence (IS) transposition, crossover, and gene transposition. Mutation can occur anywhere in

the chromosome. In the chromosome's head, each character could be replaced by functions or terminals. However, in the chromosome's tail the terminals are only allowed to be replaced with other terminals. Despite the mutation, inversion is only done in the chromosome's head. In the inversion process, a sequence is selected randomly and inverted. The insertion sequence (IS) transposition chooses a random sequence from a chromosome and copies it to any position of the chromosome's head despite the start position. In crossover, the parent chromosomes are combined with each other and two new children are produced. In gene transposition, an entire gene performs as a transposon and transposes itself to the chromosome's start position. After that, in order to prevent chromosome length change, the transposon gene is deleted from the original position. Modeling a GEP for each special problem requires five steps as presented in Section 5 in detail.

3. Data used

The data sets presented by Emiroglu et al. [22] and Bagheri et al. [24] were used in this study. Emiroglu et al. [22] conducted their experiments in the hydraulics laboratory at Firat University, Elazig, Turkey. In the experiment they used a 12 m long rectangular channel with width, depth and slope of 0.5 m, 0.5 m and 0.01 respectively. The main channel consisted of a smooth, horizontal, well-painted steel bed and glass lateral walls. A sluice gate was placed at the end of the main channel to control the depth of flow. A discharge collection channel was installed parallel to the main channel and had 0.7 m depth and 0.5 m width. The collection channel was made for the purpose of providing a circular free-surface condition in a circular shape to facilitate a free overflow condition. A Mitutoyo digital point gauge with ±0.01 mm sensitivity was installed 0.4 m away from the weir. The side weir was made of steel plates and the crests were completely sharp, aerated and installed at the same level as the side of the main channel. The water was controlled by a pipe and a sluice gate. In Emiroglu et al.'s [24] research, a discharge amount between 0.01 and 0.150 m^3/s was measured with an electromagnetic flow meter (±0.01 L/s sensitivity). Additionally, the discharge results from the electromagnetic flow meter were calibrated using a V-notched weir located at the beginning of the system. The discharge passing through the side weir was also calibrated using a standard rectangular weir installed downstream of the discharge collection channel.

Bagheri et al. [24] conducted experiments on rectangular sharp-crested weirs with different heights and widths within a horizontal rectangular channel 8 m long, 0.4 m wide, and 0.6 m deep. All experiments were carried out under subcritical flow condition. The free surface profiles were measured along the side weir's sill and along the channel's central axis using a point meter installed on a mobile carriage with ±0.5 mm accuracy. The upstream discharge (Q_1) was measured with a ±0.5% accuracy electromagnetic flow meter. The depth of flow and downstream discharge were controlled by a sluice gate that had been previously calibrated; the maximum amount of weir calibration error was approximately ±5%. The flow diverted from the weirs was calculated based on the difference between the upstream and downstream discharge. The data range employed in this study is presented in Table 1.

Table 1
Summary of experimental characteristics (Bagheri et al. [24] and Emiroglu et al. [22]).

Variable	B (m)	b (m)	p (m)	F_1	Q (L/s)
Bagheri et al. [24]	0.4	0.2–0.6	0.05–0.15	0.08–0.91	1.2–29.5
Emiroglu et al. [22]	0.5	0.15–1.50	0.12–0.20	0.08–0.92	10–150

4. Methodology

De Marchi [1] presented the change rate of discharge passing through a rectangular channel with a side weir (q) as follows:

$$q = -\frac{dQ}{ds} = \frac{2}{3} C_d \sqrt{2g}(y-p)^{1.5} \quad (1)$$

where Q is the main channel discharge, s is the measured distance along the length of the side weir from the upstream end, q is the discharge overflow per unit of weir length, g is the gravitational acceleration, p is the height of the side weir sill, y is the flow depth, C_d is the discharge coefficient and $(y-p)$ is the pressure head on the weir. The independent parameters that must be taken into consideration when estimating C_d are presented as the following equation while keeping in view the above equation (1) [2,5,9,13,49]:

$$C_d = f\left(F_1 = \frac{V_1}{\sqrt{gy_1}}, \frac{p}{y_1}, \frac{b}{y_1}, \frac{b}{B}, \psi, S_0\right) \quad (2)$$

where C_d is the discharge coefficient, F_1 is the Froude number, V_1 is the average velocity upstream of the side weir, g is the gravitational acceleration, p is the height of the weir sill, b is the weir width, B is the main channel width, ψ is the deviation angle of flow and S_0 is the channel slope.

Since for dimensionless weir length (b/B) the effect of deviation angle on the weir is considered [5], the deviation angle was not deemed a specific parameter in the previously suggested discharge coefficient equations for side weirs; therefore, the dimensionless

Table 2
Existing equations for side weir discharge coefficient in direct channels.

Reference	Category	Equation
Nandesamoorthy and Thomson [4]	1	$C_d = 0.432\left(\frac{2-F_1^2}{1+2F_1^2}\right)^{0.5}$
Yu-Tech [3]	1	$C_d = 0.623 - 0.222F_1$
Ranga Raju et al. [6]	1	$C_d = 0.81 - 0.6F_1$
Hager [7]	1	$C_d = 0.485\left(\frac{2+F_1^2}{2+3F_1^2}\right)^{0.5}$
Singh et al. [9]	2	$C_d = 0.33 - 0.18F_1 + 0.49\left(\frac{p}{y_1}\right)$
Jalili and Borghei [11]	2	$C_d = 0.71 - 0.41F_1 - 0.22\left(\frac{p}{y_1}\right)$
Borghei et al. [13]	3	$C_d = 0.7 - 0.48F_1 - 0.3\left(\frac{p}{y_1}\right) + 0.06\left(\frac{b}{B}\right)$
Emiroglu et al. [22]	3	$C_d = \left[0.836 + \left(-0.035 + 0.39\left(\frac{p}{y_1}\right)^{12.69} + 0.158\left(\frac{b}{B}\right)^{0.59} + 0.049\left(\frac{b}{y_1}\right)^{0.42} + 0.244F_1^{2.125}\right)^{3.018}\right]^{5.36}$

Table 3
GEP model parameters.

Parameter	Setting
Number of generations	20,000
Number of chromosomes	30
Number of genes	3
Mutation rate	0.03
Inversion rate	0.1
One-point recombination rate	0.3
Two-point recombination rate	0.3
Gene recombination rate	0.1
Gene transportation rate	0.1
Function set	×, −, /, + Logi2, 3Rt, Gau2, Pow
Linking function	Addition

parameters affecting discharge coefficient estimation are presented as follows:

$$C_d = f\left(F_1, \frac{p}{y_1}, \frac{b}{y_1}, \frac{b}{B}\right) \quad (3)$$

To investigate the effect of each parameter and with respect to the dimensionless parameters influencing side weir discharge coefficient estimation for direct channels, a group of five models is presented. In estimating the discharge coefficient, Model 1 involves all parameters presented in Eq. (3). However, the other models do not account for the dimensionless parameters in estimating the discharge coefficient. The five models are given below:

Model 1: $C_d = f\left(F_1, \frac{p}{y_1}, \frac{b}{y_1}, \frac{b}{B}\right)$
Model 2: $C_d = f\left(F_1, \frac{p}{y_1}, \frac{b}{y_1}\right)$
Model 3: $C_d = f(F_1, \frac{p}{y_1}, \frac{b}{B})$
Model 4: $C_d = f\left(F_1, \frac{b}{y_1}, \frac{b}{B}\right)$
Model 5: $C_d = f\left(\frac{p}{y_1}, \frac{b}{y_1}, \frac{b}{B}\right)$

The existing discharge coefficient estimation equations for rectangular side weirs in straight channels are presented in Table 2. These equations can be divided into three categories. The first category includes equations that only use the Froude number (F_1) parameter to estimate the discharge coefficient. The second category consists of equations that consider the P/y_1 parameter as well as the Froude number (F_1) parameter. The third group of equations considers the F_1 parameter for estimating the discharge coefficient (first and second category), the p/y_1 parameter (second category), as well as the dimensionless weir length (b/B) and b/y_1 parameters.

Table 4
Evaluating existing discharge coefficient equations using verification criteria.

Equation	R^2	*MARE*	*RMSE*	*SI*	*BIAS*
Nandesamoorthy and Thomson [4]	0.636	0.082	0.067	0.122	0.034
Yu-Tech [3]	0.655	0.064	0.060	0.109	0.011
Ranga Raju et al. [6]	0.655	0.119	0.074	0.135	−0.035
Hager [7]	0.637	0.164	0.115	0.209	0.096
Singh et al. [9]	0.190	0.153	0.122	0.221	0.025
Jalili and Borghei [11]	0.317	0.253	0.151	0.273	−0.124
Borghei et al. [13]	0.342	0.238	0.135	0.904	0.487
Emiroglu et al. [22]	0.611	0.237	0.164	0.297	0.130

The results of the comparison between previous equations and those proposed in this study are presented herein, in terms of criteria for coefficient of determination (R^2), root mean square of error (*RMSE*), mean absolute relative error (*MAPE*), scatter index (*SI*) and *BIAS*, as defined below:

$$R^2 = \left[\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \Big/ \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}\right]^2 \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \quad (5)$$

$$MARE = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i - y_i|}{x_i} \quad (6)$$

$$SI = \frac{RMSE}{\bar{x}} \quad (7)$$

$$BIAS = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) \quad (8)$$

where y_i and x_i are the modeled and actual C_d values, respectively, and $\bar{y}$ and $\bar{x}$ are the mean modeled and actual C_d values, respectively.

The afore-mentioned indexes represent the estimated values as prediction error average but provide no information on the prediction error distribution of the presented models. The *RMSE* index only indicates a model's ability to predict a value away from the

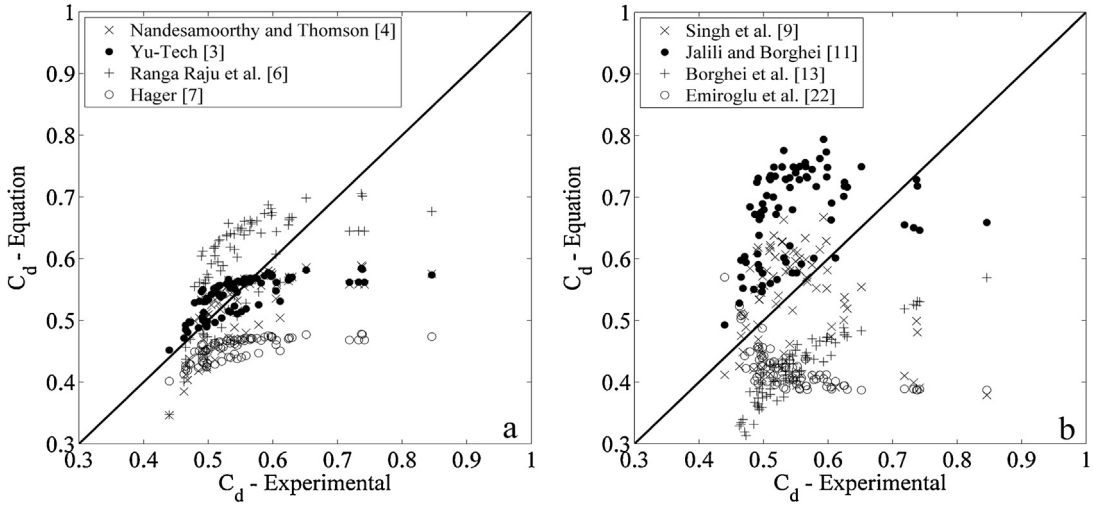

Fig. 3. Studying existing discharge coefficient equations for a rectangular weir using laboratory results [24].

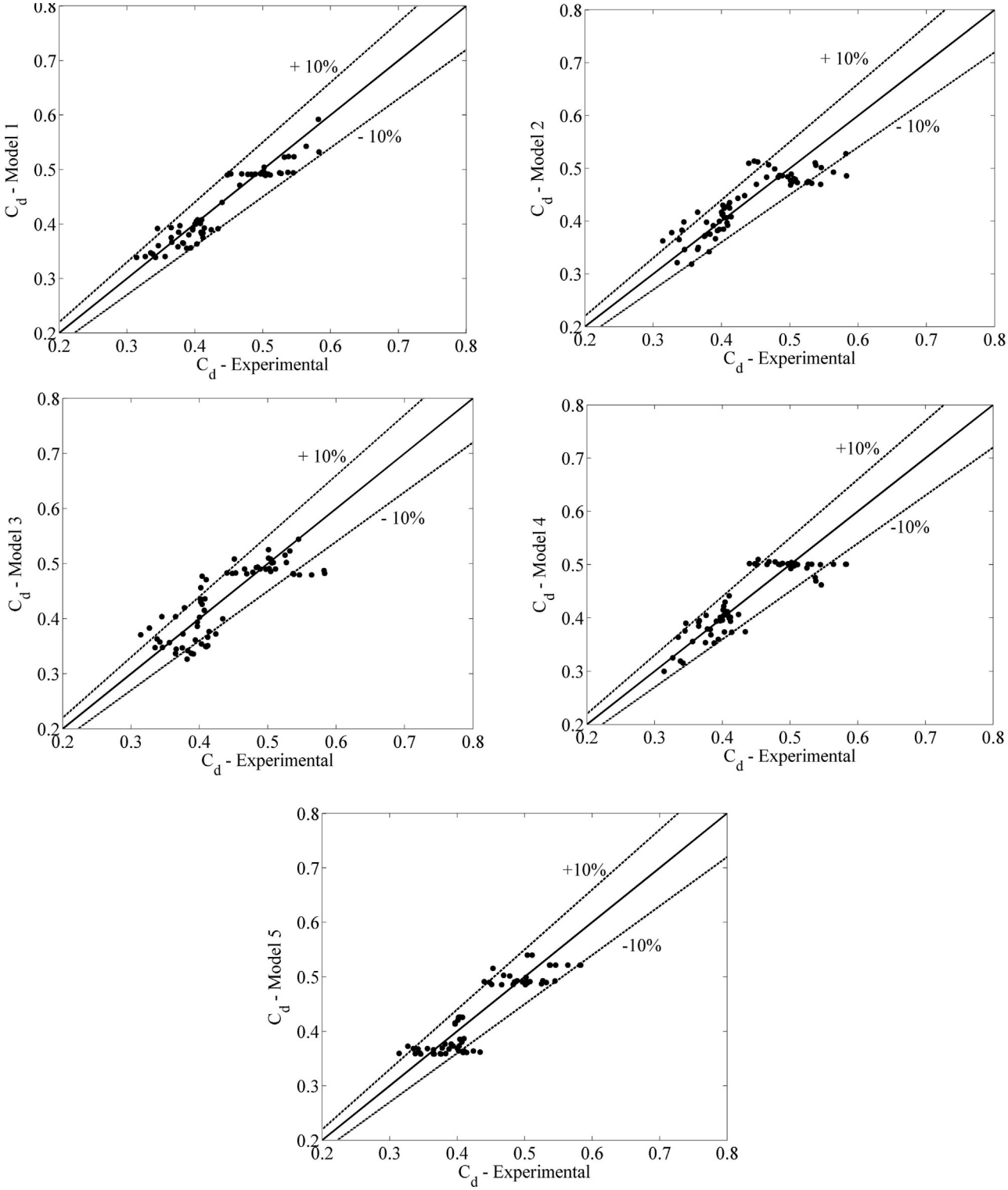

Fig. 4. Comparing the estimated results of the five models against the laboratory results of Emiroglu et al. [22] (training).

mean [50]. Therefore, the presented models must be evaluated using other indexes such as threshold statistics [51–54]. The TS_x index shows the error distribution in the amounts predicted by each model for x% of predictions. This parameter is determined for various amounts of mean absolute relative error. The amount of TS_x index for x% of the predictions is determined as follows:

$$TS_x = \frac{Y_x}{n} \times 100 \tag{9}$$

In the eqation above, Y_x represents the anticipated amounts for each *MARE* smaller than x% from the total data. The mathematical equations of these statistical indexes are provided in [54–57].

5. Derivation of densimetric Froude number based on GEP

In the present section, the GEP discharge coefficient modeling procedure is explained. The GEP models were trained with

Emiroglu et al.'s [22] experimental dataset. Following training, in order to examine the models' performance with the non-observed dataset, Bagheri et al.'s [24] empirical study was used.

Modeling a GEP requires five major steps. In the first step, the fitness function applied in the chromosome cost calculation should be determined. In this study, the following fitness function is applied:

$$f_i = \sum_{j=1}^{C_t} \left(M - \left|C_{(i,j)} - T_j\right|\right) \quad (10)$$

where M is the selection range, $C_{(i,j)}$ is the ith chromosome value for the jth fitness case, and T_j is the observed value for the jth fitness case. If the precision of $(|C_{(I,j)} - T|)$ is equal to or lower than 0.01, the precision value is considered zero. Here, M is equal to 100, therefore the maximum value of f is $f_i = f_{max} = C_t M = 1000$. Using the explained fitness function leads to the GEP model finding the optimized solution to the problem.

The second step in GEP modeling involves determining the mathematical functions that chromosomes are allowed to use in their programs and in the final equation. There is no definitive rule in choosing a mathematical function combination. In this paper, the four basic mathematical functions $\{\times, -, /, +\}$ and other more complex mathematical functions, e.g. log, exp, power, and cubic root are used.

The third step is determining the chromosomal architecture, which involves determining the head length and gene numbers. In the present study, the trial and error method is used to establish these two parameters. The GEP is run for various head length and gene number combinations. The results show that the GEP model with three genes and head length of five produces the most accurate result in modeling the rectangular side weir discharge coefficient. Increasing these two parameters has no effect on the model's performance. In addition, because an initial population in the 30–100 interval leads to acceptable results [47], an initial population of 50 was considered in the current study.

The fourth step of the GEP modeling procedure is to choose a proper linking function. Recent studies on GEP application in the hydraulic field indicate that adding the linking function leads to more accurate results [39,41]. For this reason, the additional linking function is used in the present work.

Finally, in the last step of GEP modeling, the genetic operators that GEP utilized in the reproduction process are determined. The genetic operators result in diversity of evolution generations. Despite the mutation rate, which is determined by trial and error method, other genetic operators are considered as the proposed initial values of GEP [47]. The genetic operator parameter values and other properties considered in the previous steps are presented in Table 3.

6. Applications and results

Regarding the parameters influencing C_d estimation, the equations for side weir discharge coefficient can be classified into three general categories. Fig. 3a presents the accuracy of discharge coefficient equations that only use the Froude number (F_1) parameter to estimate the discharge coefficient, and Fig. 3b studies the accuracy of the other equations shown in Table 2 that are based on Bagheri et al.'s [24] empirical results. Fig. 3a and b indicates the vast scattering of results obtained with existing discharge coefficient equations. According to the results, the dispersion in case (a) is less than case (b). This means that compared to other existing equations, those that only used the Froude number parameter for discharge coefficient estimation present better results. Fig. 3a demonstrates that the equations proposed by previous researchers follow a relatively similar process. As the amount of C_d increases in all a-state presented, the accuracy decreases. However, this is not true of the equation presented by Ranga Raju et al. [6], where low discharge coefficient values do not yield good estimation either. As per Fig. 3a, all equations generate high relative error results when the discharge coefficient is larger than 0.6, as at some points, the relative estimation error of these equations can even reach 40% (Hager's equation). In the majority of cases, the equation presented by Nandesamoorthy and Thomson [4] estimates lower values than reported by Bagheri et al. [24]. The results obtained with this model are not very accurate, as the estimated results had an approximately 15% relative error average. It should be noted that this model estimates results with an approximate relative error of 30% at some points, meaning that this model is not accurate in estimating the side weir discharge coefficient. Although the equation suggested by Yu-Tech [3] does not estimate the amount of discharge coefficient with acceptable error, like other models shown in Fig. 3a, it gives better results in comparison with other equations given in Fig. 3a with relative error of approximately 5% (Table 4).

Fig. 3b investigates the estimated discharge coefficient of a side weir using existing equations related to the laboratory results of Bagheri et al. [24]. The equations shown in this figure are different from state a, in that the equations in this figure consider other parameters such as dimensionless weir length (b/B), p/y_1 and b/y_1 in addition to the Froude number parameter. It is evident that the majority of amounts estimated by these equations do not conform to the amounts presented by Bagheri et al. [24]. Considering Table 4, which indicates the verification criterion results for the presented equations, Singh et al.'s [9] equation is more accurate than the two other equations, as the relative error of this equation is approximately 15% while this index is more than 20% for the two other equations. Also, the *RMSE* of Singh et al.'s equation [9] is less than that of the two other equations proposed by Jalili and Borghei [11] and Emiroglu et al. [21].

Considering the explanations on discharge coefficient estimation results from existing equations and using the statistical indexes in this study (the results for all equations are given in

Table 5
Quantitative study of the presented models' results in comparison with the laboratory results of Emiroglu et al. [22] using verification criteria (training).

Models	R^2	*MARE*	*RMSE*	*BIAS*	*SI*
Model 1	0.947	0.041	0.023	0.006	0.052
Model 2	0.868	0.062	0.035	0.004	0.079
Model 3	0.818	0.076	0.041	0.006	0.093
Model 4	0.876	0.056	0.033	0.004	0.076
Model 5	0.891	0.063	0.032	0.007	0.074

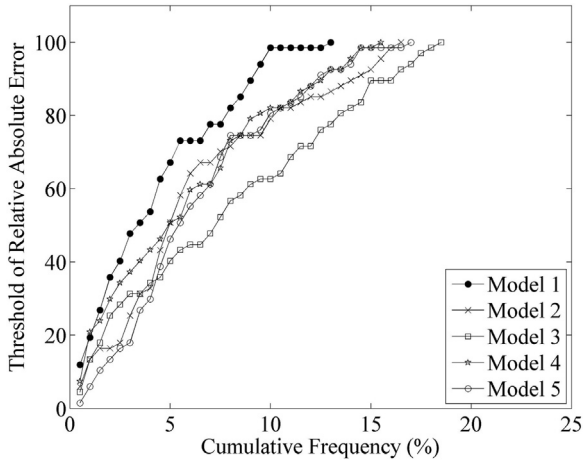

Fig. 5. GEP error distribution for all models (Emiroglu et al.'s [22] dataset).

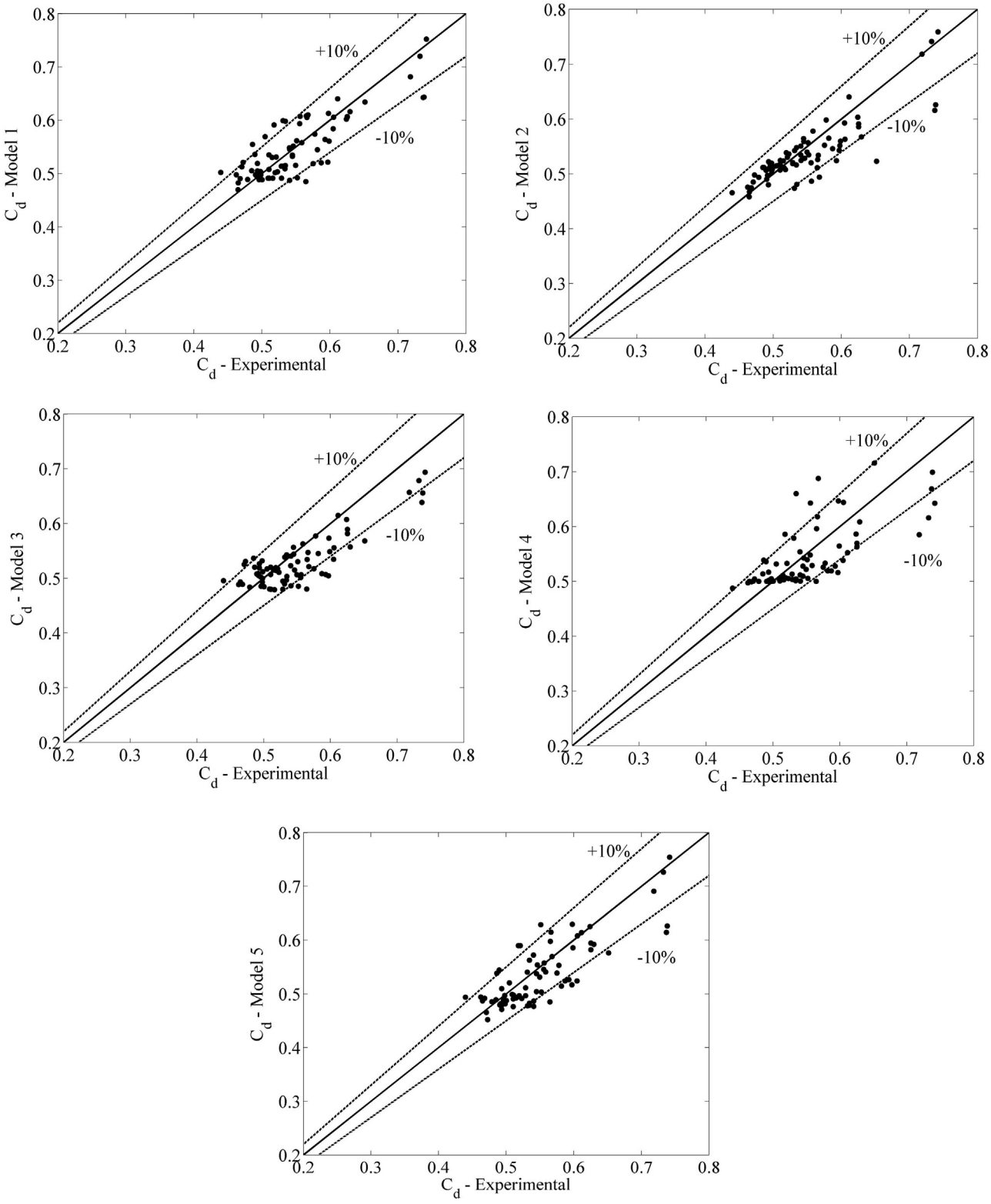

Fig. 6. Comparing the estimated results of the five models to the laboratory results of Bagheri et al. [24] (Testing).

Table 4), in case we want to select the best discharge coefficient estimation equation from amongst the existing equations in this field we would select Yu-Tech's equation [3], which has the least *MARE* and *RMSE* among other equations. It is worth mentioning that this equation's results have an approximate relative error up to 30% at some points; therefore, this equation cannot be used confidently for estimating the discharge coefficient.

Due to the low accuracy of the equations and because nearly all models estimate the discharge coefficient with relatively high error (Fig. 3), it will be estimated using 5 models presented through GEP for all states of different equations in this section. The experimental data of Emiroglu et al. [22] will be used to estimate the equations. Also, Bagheri et al.'s [24] laboratory data are used to evaluate the accuracy of the presented models using the set of data not applied

to estimate the models. Fig. 4 shows the estimated discharge coefficients using the laboratory results presented by Bagheri et al. [24]. Four dimensionless parameters, namely F_1, b/B, b/y_1 and p/y_1 were used for estimating the discharge coefficient in these models. Model 1 includes all parameters and Models 2–5 are used to study the effects of not using each of these parameters on estimating the discharge coefficient in order to analyze sensitivity. Fig. 4 demonstrates that Model 1 estimated the results fairly accurately, with all the discharge coefficient amounts estimated by this model having a relative error below 10%. In addition, as Table 5 reveals, the average relative error presented by this model is almost 4%. According to Fig. 5, nearly 75% of the discharge coefficients estimated using Model 1 have a relative error lower than 6%. Based on the given explanations so far, it can be said that Model 1, which uses all four presented dimensionless parameters in Eq. (3) to estimate the discharge coefficient, is fairly accurate in estimating C_d.

The dimensionless parameters F_1, b/y_1 and p/y_1 were considered for estimating the discharge coefficient using Model 2. Fig. 4 shows that the results obtained with Model 2 are mostly estimated with an error less than 10%. However, according to Fig. 5, almost 50% of the estimated amounts have less than 5% error, and this is approximately 65% for Model 1. Also, Table 5 demonstrates that the relative error average produced by this model is about 6%. Therefore, it is observed that not using the dimensionless weir length parameter (b/B) leads to a 2% increase in relative error.

The F_1, b/B and p/y_1 parameters were applied in Model 3 to estimate the discharge coefficient. Table 5 indicates that the amount of relative error in estimating C_d using this model is approximately 7%, which is the largest relative error compared with the other models. Approximately 50% of the estimated discharge coefficient amounts have relative error of approximately 5%, while it is 80% for Model 1 that also uses the b/y_1 parameter besides the parameters used by this model (Fig. 5). Therefore, it can be concluded that not considering the b/y_1 parameter leads to a 3% increase in the relative error average compared with Model 1.

Model 4 presented good results in estimating the discharge coefficient using F_1, b/B and b/y_1. Table 5 clarifies that the discharge coefficients estimated with this equation have mostly less than 10% error, while the relative error average of this parameter estimated by Model 4 is almost 5%. Additionally, based on Fig. 5, 50% of the discharge coefficient amounts estimated by this model have a relative error of approximately 5%, which is 75% for Model 1. Therefore, it can be stated that although Model 4 estimates the discharge coefficient relatively accurately, not using p/y_1 leads to a 1% increase in the relative error of Model 4 compared with Model 1.

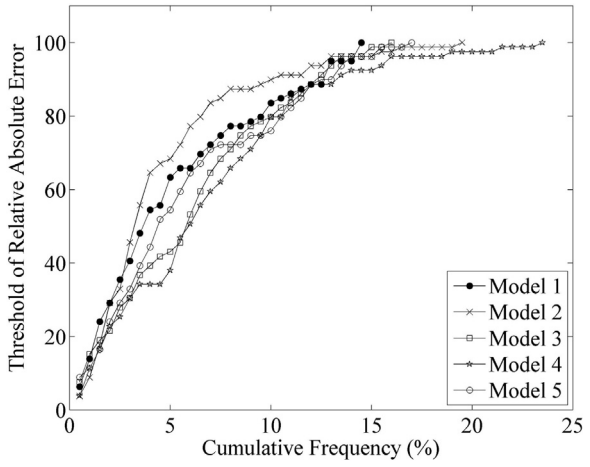

Fig. 7. GEP error distribution for all models (Bagheri et al.'s [24] dataset).

In comparison with the laboratory results presented by Emiroglu et al. [22], Model 5 mostly estimates the discharge coefficient with a maximum relative error of approximately 10%. It is evident that the relative error produced by this model is about 6% (Table 5), which is similar to that of Model 2. The parameters applied in this model to estimate the discharge coefficient include b/B, b/y_1 and p/y_1. Fig. 5 demonstrates that the results estimated by Model 5 have a relative error of roughly 6% for 50% of the data and 75% for Model 1.

Accordingly, it could be stated that in contrast to the laboratory results of Emiroglu et al. [22], Model 1 has the highest accuracy with the least amount of *MARE* and *RMSE* compared with other models. Not considering each of the parameters F_1, b/B, b/y_1 and p/y_1 leads to an increase in relative error as opposed to Model 1, which considers all dimensionless parameters simultaneously. In view of the presented results, compared with other models, the increase in relative error is less for Model 2 (b/B) and Model 4 (p/y_1). Furthermore, not using the b/y_1 parameter has the greatest effect on increasing the relative error, while the discharge coefficient cannot be estimated confidently by this model.

To investigate the accuracy of a model it is necessary to evaluate it using a dataset that has had no role in estimating the model as well as using data employed in estimating the model. Thus, the model accuracy is assessed under different conditions. While the laboratory results of Emiroglu et al. [22] were used in this study to present the 5 models, the empirical results of Bagheri et al. [24] were also used to investigate the accuracy of the models with the dataset that was not used to estimate the model itself.

Fig. 6 displays the estimation discharge coefficients as presented by the 5 models in this study against the laboratory results of Bagheri et al. [24]. Considering Fig. 5, almost all models presented approximately 80% of the estimated amount with a relative error less than 10%. It is observed that with an average relative error of approximately 5%, Model 1 produced results relatively similar to the state when the laboratory results of Emiroglu et al. [22] were used. As in the previous state (using Emiroglu et al.'s [22] data), Model 2 gave good results as well. As seen in Fig. 7, 90% of the estimated discharge coefficient results for this model had a relative error less than 10%. In cases where the amount estimated by Model 2 was more than 10%, the error amount even reached 19% while the largest estimation error was almost 14% for Model 1, which is similar to the state when Emiroglu et al.'s [22] data was used. Therefore, it can be stated that not using the b/B parameter increased the estimation error very little, although considering Table 6, the average error estimation presented by both models is nearly 4%. Model 3 estimated the discharge coefficient with a relative error similar to that of Model 1, save the fact that in this case the amounts with relative error more than 10% are less than the amounts presented by Bagheri et al. [24], which is in contrary to Model 1. The relative error of Model 3 is roughly 4% (Table 6), which is not very different from Models 1 and 2. The greatest error by Model 3 is approximately 16%, which means this model estimates results more accurately than Model 2 but less accurately than Model 1. According to Fig. 6, Model 4 presents significantly weaker results than the other models, as its greatest amount of

Table 6
Quantitative investigation of the presented models' results in comparison with the laboratory results of Bagheri et al. [24] using verification criteria (testing).

Models	R^2	*MARE*	*RMSE*	*BIAS*	*SI*
Model 1	0.870	0.051	0.037	0.010	0.067
Model 2	0.885	0.045	0.036	0.018	0.066
Model 3	0.870	0.059	0.042	0.008	0.076
Model 4	0.750	0.066	0.050	0.013	0.090
Model 5	0.853	0.057	0.041	0.551	0.075

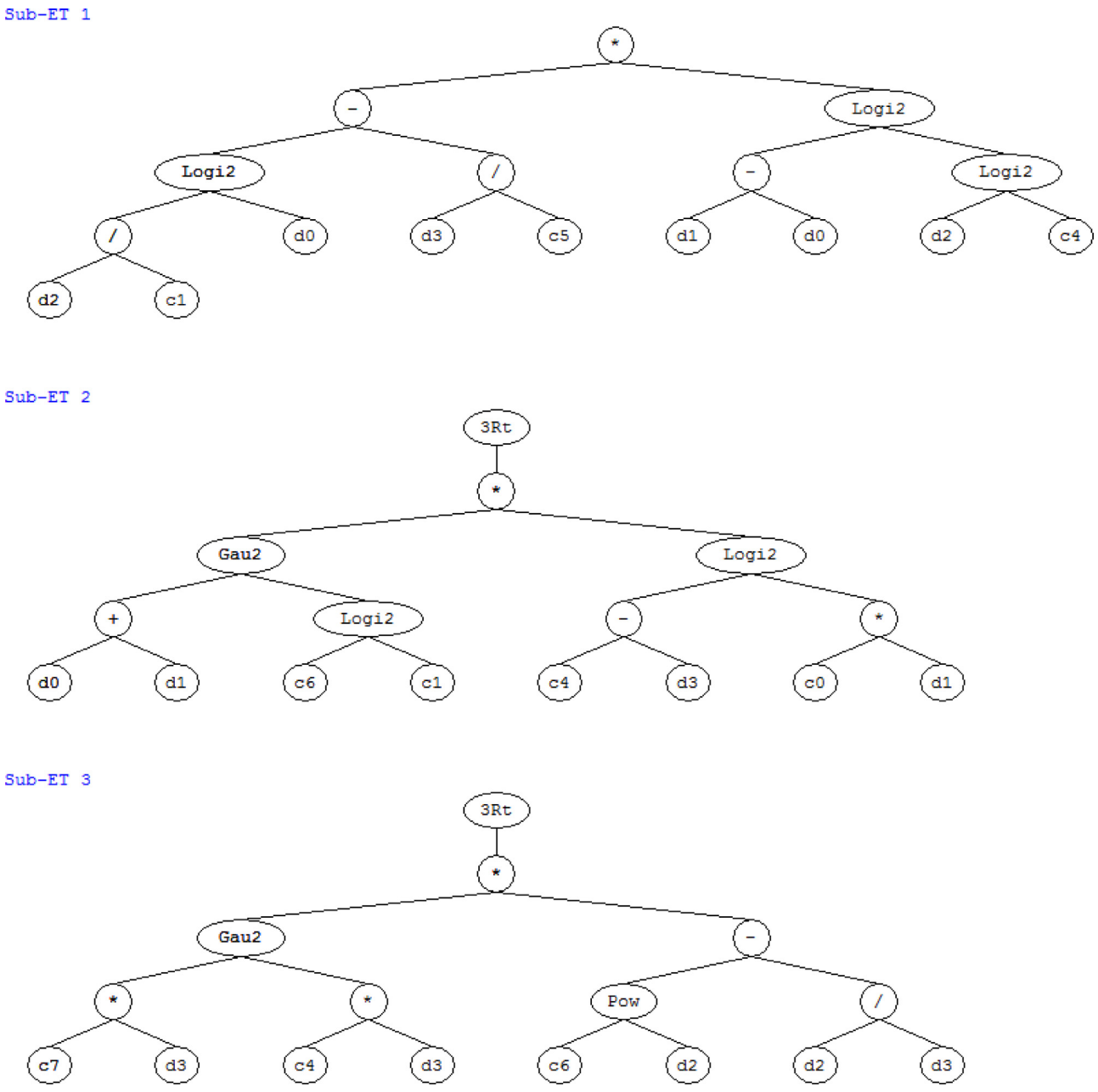

Fig. 8. Expression Tree for GEP formulation (Model 1).

relative error reached up to 23%. Therefore, it can be said that not using the b/y_1 parameter has a greater effect on estimating the discharge coefficient compared with excluding other parameters, although comparing the relative error of this model with that of other models (Table 6) indicates that it is not a significant amount. Likewise, Model 5 produced the same results as Model 3 for both conditions using different laboratory data (Emiroglu et al. [22] and Bagheri et al. [24]).

It can be concluded that all five presented models provide fairly good results for both states of including different datasets. The only difference between the estimations of different models is their relative error distribution. Accordingly, considering the given explanations, the fact that compared to other models the greatest relative error by Model 1 was the lowest for both sets of data, and the *RMSE* and *MARE* indexes presented for this model are good, it can be argued that compared to other models, Model 1 can serve as a replacement method.

Regarding the explanations for the different models presented in this study, Model 1 is more accurate than the rest. The equation that employs data from Model 1 is presented below using the expression tree provided for this model as given in Fig. 8. In this figure, 3RT and Gau2 (x, y) are cubic root and $\exp(-(x+y)^2)$, respectively. In addition, Logi2 is a condition statement that is shown in Eqs. (12)–(15). Other constants from Fig. 8 are presented in Table 7.

$$C_d = \left(\text{logit}(u) + \frac{p}{69.67y_1}\right) \times \text{logit}(v) + \left(\left[e^{-\left(F_1+\frac{b}{B}+1\right)^2}\right] \times [\text{logit}(s)]\right)^{\left(\frac{1}{3}\right)} + \left(\left[e^{-\left(\frac{11.76p}{y_1}\right)^2}\right] \times 9.28\left(\frac{b}{y_1}\right) - \frac{b/y_1}{p/y_1}\right)^{\left(\frac{1}{3}\right)} \quad (11)$$

Table 7
Values of parameters used in ET (Fig. 8).

$d0$	$d1$	$d2$	$d3$	$G1$			$G2$				$G3$		
				$C5$	$C4$	$C1$	$C6$	$C1$	$C4$	$C0$	$C7$	$C4$	$C6$
F_1	b/y_1	b/B	p/y_1	−69.67	−3.88	−10.25	5.47	79.85	−10.82	22.89	77.4	−65.64	9.28

where F_1 is the Froude number, b is the weir length, B is the length of the main channel, y_1 is the upstream flow depth and p is the weir height. The logit(u), logit(v) and logit(s) parameters are defined as follows.

$$\text{logit}(u) = \begin{cases} \dfrac{1}{1+\exp\left(\dfrac{\left|\frac{b}{79.85y_1}\right|}{709\times\left(\frac{b}{79.85y_1}+F_1\right)}\right)} & \text{if } \left|\dfrac{b}{79.85y_1}\right| > 709 \\ \dfrac{1}{1+e^{-\left(\frac{b}{79.85y_1}+F\right)}} & \text{if } \left|\dfrac{b}{79.85y_1}\right| \le 709 \end{cases} \quad (12)$$

$$\text{logit}(v) = \begin{cases} \dfrac{1}{1+\exp\left(\dfrac{\left|\frac{b}{B}-F_1+\text{logit}(x)\right|}{709\times\left(\frac{b}{B}-F_1+\text{logit}(x)\right)}\right)} & \text{if } \dfrac{b}{B}-F_1+\text{logit}(x) > 709 \\ \dfrac{1}{1+e^{-\left(\frac{b}{B}-F_1+\text{logit}(x)\right)}} & \text{if } \dfrac{b}{B}-F_1+\text{logit}(x) \le 709 \end{cases} \quad (13)$$

$$\text{logit}(x) = \begin{cases} \dfrac{1}{1+\exp\left(\dfrac{\left|\frac{b}{y_1}-3.88\right|}{709\times\frac{b}{y_1}-3.88}\right)} & \text{if } \left|\dfrac{b}{y_1}-3.88\right| > 709 \\ \dfrac{1}{1+e^{-\left(\frac{b}{y_1}-3.88\right)}} & \text{if } \left|\dfrac{b}{y_1}-3.88\right| \le 709 \end{cases} \quad (14)$$

$$\text{logit}(s) = \begin{cases} \dfrac{1}{1+\exp\left(\dfrac{\left|-10.82-\frac{p}{y_1+22.89\times F_1}\right|}{709\times\left(-10.82-\frac{p}{y_1+22.89\times F_1}\right)}\right)} & \text{if } \left|-10.82-\dfrac{p}{y_1+22.89\times F_1} > 709\right| \\ \dfrac{1}{e^{-\left(-10.82-\frac{p}{y_1+22.89\times F_1}\right)}} & \text{if } \left|-10.82-\dfrac{p}{y_1+22.89\times F_1} \le 709\right| \end{cases} \quad (15)$$

7. Conclusion

Gene expression programming (GEP) was used to present an equation for estimating the discharge coefficient in a side weir located on the side of a rectangular channel. The parameters affecting the discharge coefficient were presented by introducing the Froude number (F_1), dimensionless weir length (b/B), ratio of weir length to upstream flow depth (b/y_1) and the ratio of weir height to its length (p/y_1) dimensionless parameters. In order to analyze the sensitivity of each of the dimensionless parameters addressed, five different models were employed using the mentioned parameters. The study results indicate that not using the b/y_1 parameter has the greatest effect on increasing the relative error average of estimation, as this model cannot be used confidently to estimate the discharge coefficient. Most of the presented models estimated the discharge coefficient fairly well except for Model 3, which produced the greatest relative error of approximately 7% while using the laboratory results of Emiroglu et al. [22], indicating the GEP's capability in estimating the discharge coefficient. Therefore, it can be stated that Model 1, which produced good results with both different datasets, can be used confidently to estimate the discharge coefficient. It is also worth mentioning that GEP has high precision in discharge coefficient estimation and may be used as an alternative to existing methods.

Acknowledgments

The authors would like to express their appreciation to the anonymous reviewers for their valuable comments and suggestions to improve the manuscript and to Maya Binder for the final English editing.

References

[1] G. De Marchi, Saggio di teoria del funzionamento degli stramazzi laterali, L'Energia electrica Milan 11 (1934) 849–860 (in Italian).
[2] K. Subramanya, S.C. Awasthy, Spatially varied flow over side weirs, J. Hydraul. Div. 98 (1972) 1–10.
[3] L. Yu-Tech, Discussion of spatially varied flow over side weir, J. Hydraul. Eng. 98 (1972) 2046–2048.
[4] T. Nandesamoorthy, A. Thomson, Discussion of spatially varied flow over side weir, J. Hydraul. Div. 98 (1972) 2234–2235.
[5] A.M.M. El-Khashab, Hydraulics of flow over side weirs (Ph.D. thesis), University of Southampton, UK, 1975.
[6] K.G. Ranga Raju, B. Parasad, S.K. Gupta, Side weir in rectangular channel, J. Hydraul. Eng. 105 (1979) 547–554.
[7] W.H. Hager, Lateral outflow over side weirs, J. Hydraul. Eng. 113 (1987) 491–504.
[8] H. Cheong, Discharge coefficient of lateral diversion from trapezoidal channel, J. Irrig. Drain. Eng. 117 (1991) 461–475.
[9] R. Singh, D. Manivannan, T. Satyanarayana, Discharge coefficient of rectangular side weirs, J. Irrig. Drain. Eng. 120 (1994) 814–819.
[10] P.K. Swamee, S.K. Pathak, M. Mohan, S.K. Agrawal, M.S. Ali, Subcritical flow over rectangular side weir, J. Irrig. Drain. Eng. 120 (1994) 212–217.
[11] M.R. Jalili, S.M. Borghei, Discussion of discharge coefficient of rectangular side weir, by R. Singh, D. Manivannan and T. Satyanarayana, J. Irrig. Drain. Eng. 122 (1996) 132.
[12] M. Ghodsian, Elementary discharge coefficient for rectangular side weir, in: Proc. 4th Int. Conf. Civil Eng., Tehran, Iran, 1997.
[13] S. Borghei, M. Jalili, M. Ghodsian, Discharge coefficient for sharp crested side-weirs in subcritical flow, J. Hydraul. Eng. 125 (1999) 1051–1056.
[14] M. Ghodsian, Supercritical flow over a rectangular side weir, Can. J. Civil. Eng. 30 (2003) 596–600.
[15] E. Yüksel, Effect of specific energy variation on lateral overflows, Flow Meas. Instrum. 15 (2004) 259–269.
[16] M. Khorchani, O. Blanpain, Development of a discharge equation for side weirs using artificial neural networks, J. Hydroinform. 7 (2005) 31–39.
[17] A.S. Ramamurthy, J. Qu, D. Vo, Nonlinear PLS method for side weir flows, J. Irrig. Drain. Eng. 132 (2006) 486–489.
[18] T. Huagao, L. Wang, G. Ken, Design of side weirs in subcritical flow, in: ASCE Conf. Proc. Urban Drain. Model, Orlando, USA, 2007.
[19] M. Venutelli, Method of solution of no uniform flow with the presence of rectangular side weir, J. Irrig. Drain. Eng. 134 (2008) 840–846.
[20] A.R. Vatankhah, M. Bijankhan, Discussion of method of solution of non-uniform flow with the presence of rectangular side weir, J. Irrig. Drain. Eng. 135 (2009) 812–814.

[21] M.E. Emiroglu, O. Kisi, O. Bilhan, Predicting discharge capacity of triangular labyrinth side weir located on a straight channel by using an adaptive neuro-fuzzy technique, Adv. Eng. Softw. 41 (2010) 154–160.
[22] M.E. Emiroglu, H. Agaccioglu, N. Kaya, Discharging capacity of rectangular side weirs in straight open channels, Flow Meas. Instrum. 22 (2011) 319–330.
[23] O. Castro-Orgaz, W.H. Hager, Subcritical side-weir flow at high lateral discharge, J. Hydraul. Eng. 138 (2012) 777–787.
[24] S. Bagheri, A.R. Kabiri-Samani, H. Heidarpour, Discharge coefficient of rectangular sharp-crested side weirs. Part II: Dominguez's method, Flow Meas. Instrum. 35 (2013) 116–121.
[25] M.E. Emiroglu, M.C. Aydin, N. Kaya, Discharge characteristics of a trapezoidal labyrinth side weir with one and two cycles in subcritical flow, J. Irrig. Drain. Eng. 140 (2014) 1–13.
[26] H. Bonakdari, S. Baghalian, F. Nazari, M. Fazli, Numerical analysis and prediction of the velocity field in curved open channel using Artificial Neural Network and Genetic Algorithm, Eng. Appl. Comput. Fluid Mech. 5 (2011) 384–396.
[27] S. Baghalian, H. Bonakdari, F. Nazari, M. Fazli, Closed-form solution for flow field in curved channels in comparison with experimental and numerical analyses and Artificial Neural Network, Eng. Appl. Comput. Fluid Mech. 6 (2012) 514–526.
[28] C.H.F. Toro, S.G. Meire, J.F. Gálvez, F. Fdez-Riverola, A hybrid artificial intelligence model for river flow forecasting, Appl. Soft Comput. 13 (2013) 3449–3458.
[29] H.M. Azamathulla, A.A. Ghani, S.Y. Fei, ANFIS-based approach for predicting sediment transport in clean sewer, Appl. Soft Comput. 12 (2012) 1227–1230.
[30] I. Ebtehaj, H. Bonakdari, Evaluation of sediment transport in sewer using artificial neural network, Eng. Appl. Comput. Fluid Mech. 7 (2013) 382–392.
[31] O. Kisi, Suspended sediment estimation using neuro-fuzzy and neural network approaches, Hydrol. Sci. J. 50 (2005) 683–696.
[32] L.H. Feng, J. Lu, The practical research on flood forecasting based on artificial neural networks, Expert Syst. Appl. 37 (2010) 2974–2977.
[33] S.M. Bateni, M. Mortazavi-Naeini, B. Ataie-Ashtiani, D.S. Jeng, R. Khanbilvardi, Evaluation of methods for estimating aquifer hydraulic parameters, Appl. Soft Comput. 28 (2015) 541–549.
[34] O. Bilhan, M.E. Emiroglu, O. Kisi, Application of two different neural network techniques to lateral outflow over rectangular side weirs located on a straight channel, Adv. Eng. Softw. 41 (2010) 831–837.
[35] M.E. Emiroglu, O. Bilhan, O. Kisi, Neural networks for estimation of discharge capacity of triangular labyrinth side-weir located on a straight channel, Expert Syst. Appl. 38 (2011) 867–874.
[36] O. Kisi, M.E. Emiroglu, O. Bilhan, A. Guven, Prediction of lateral outflow over triangular labyrinth side weirs under subcritical conditions using soft computing approaches, Expert Syst. Appl. 39 (2012) 3454–3460.
[37] O.F. Dursun, N. Kaya, M. Firat, Estimating discharge coefficient of semi-elliptical side weir using ANFIS, J. Hydrol. 426/427 (2012) 55–62.
[38] K. Eldrandaly, A GEP-based spatial decision support system for multisite land use allocation, Appl. Soft Comput. 10 (2010) 694–702.
[39] A.A. Ghani, H. Md. Azamathulla, Gene-expression programming for sediment transport in sewer pipe systems, J. Pipeline Syst. Eng. Pract. 2 (2010) 102–106.
[40] M.A. Antoniou, E.F. Georgopoulos, K.A. Theofilatos, A.P. Vassilopoulos, S.D. Likothanassis, A gene expression programming environment for fatigue modeling of composite materials, in: Artificial Intelligence: Theories, Models and Applications, Springer, Berlin, Heidelberg, 2010, pp. 297–302.
[41] H.Md. Azamathulla, A.A. Ghani, C.S. Leow, C.K. Chang, N.A. Zakaria, Gene-expression programming for the development of a stage-discharge curve of the Pahang River, Water Resour. Manag. 25 (2011) 2901–2916.
[42] N.R. Sakthivel, B.B. Nair, V. Sugumaran, Soft computing approach to fault diagnosis of centrifugal pump, Appl. Soft Comput. 12 (2012) 1574–1581.
[43] K. Xu, Y. Liu, R. Tang, J. Zuo, J. Zhu, C. Tang, A novel method for real parameter optimization based on Gene Expression Programming, Appl. Soft Comput. 9 (2009) 725–737.
[44] R.R. Malhotra, Comparative analysis of statistical and machine learning methods for predicting faulty modules, Appl. Soft Comput. 21 (2014) 286–297.
[45] C. Ferreira, Gene expression programming in problem solving, in: R. Roy, M. Köppen, S. Ovaska, T. Furuhashi, F. Hoffmann (Eds.), Soft Computing and Industry: Recent Applications, Springer-Verlag, 2002.
[46] J.R. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection, MIT, Cambridge, USA, 1992.
[47] C. Ferreira, Gene expression programming: a new adaptive algorithm for solving problems, Complex Syst. 13 (2001) 87–129.
[48] C. Ferreira, Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence, 2nd ed. (Rev.), Springer-Verlag, Berlin, NY, USA, 2006.
[49] K.H.V. Durga Rao, C.R.S. Pillai, Study of flow over side weirs under subcritical conditions, Water Resour. Manag. 22 (2008) 131–143.
[50] K. Hsu, V.H. Gupta, S. Sorroshian, Artificial neural network modeling of the rainfall-runoff process, Water Resour. Res. 31 (1995) 2517–2530.
[51] A. Jain, A.K. Varshney, U.C. Joshi, Short-term water demand forecast modeling at IIT Kanpur using artificial neural networks, Water Resour. Manag. 15 (2001) 299–321.
[52] A. Jain, L.E. Ormsbee, Evaluation of short-term water demand forecast modeling techniques: conventional methods versus AI, J. Am. Water Works Assoc. 94 (2002) 64–72.
[53] M.P. Rajurkar, U.C. Kothyarib, U.C. Chaube, Modeling of the daily rainfall-runoff relationship with artificial neural network, J. Hydrol. 285 (2004) 96–113.
[54] M.F. Maghrebi, M. Givehchi, Using non-dimensional velocity curves for estimation of longitudinal dispersion coefficient, in: Proc. 7th Int. Symp. River Eng., Ahwaz, Iran, 2007.
[55] H.R. Maier, G.C. Dandy, The use of artificial neural networks for the prediction of water quality parameters, Water Resour. Res. 32 (1996) 1013–1022.
[56] O. Kisi, M.E. Karahan, Z. Sen, River suspended sediment modeling using a fuzzy logic approach, Hydrol. Process. 20 (2006) 4351–4362.
[57] R.P. Gopakumar, P. Mujumdar, A fuzzy dynamic wave routing model, Hydrol. Recess. 21 (2007) 458–467.