ORIGINAL ARTICLE

Gene Expression Programming (GEP) to predict coefficient of discharge for oblique side weir

Ahmed Y. Mohammed[1](http://orcid.org/0000-0002-2274-9202) · Ali Sharif²

Received: 29 February 2020 / Accepted: 28 April 2020 / Published online: 23 May 2020 © The Author(s) 2020

Abstract

In irrigation and drainage structures, side weir is widely used for fow diversion from main to branch channels. Side weir is also used as a measuring device for discharge measurements, so discharge coefficient was mainly studied in many previous studies. Skew side weir was not taking a good highlight in previous studies and literature, so the present work discharge coefficient calculation for the skew side weir was adopted and studied. Multiple Linear Regression (MLR) and Gene Expression Programming (GEP) tools were used in the present study and compared with observed values of C_d . The mean absolute error for C_d observed and calculated in MLR and GEP was not exceeded 5%. The C_d values for skew side weir ranged from (0.65) to (0.85) , while its values for straight vertical side from previous literature weir ranged from (0.45) to (0.65) ; this mean skew side weir can be used for increase in discharge diversion to the branch channel at the same water levels by 27%. The Akaike information criteria (AIC) with (AICs), root-mean-square error (RMSE), mean absolute relative error (MARE) and scatter index (SI) are used in this study for measuring the GEP model performance. From results, the GEP model has AIC=−216.51, AICs=−918.51, RMSE=0.004653, MARE=0.005234, R^2 =0.994 and SI=0.006231 performed the best. According to previous results, the new equation presented through GEP can be adopted for discharge coefficient calculation in skew side weir.

Keywords Skew side weir · Coefficient of discharge · Gene Expression Programming (GEP) · MLR · Open channel hydraulic · RMSE · AIC · MARE · SI

List of symbols

 \boxtimes Ahmed Y. Mohammed ahmedymaltaee@gmail.com; a.altaee@uomosul.edu.iq Ali Sharif a.sharif67@yahoo.com

¹ Department of Dams and Water Resources Engineering, University of Mosul, Mosul, Iraq

² Monetary and Banking Research Institution, Tehran, Iran

k The number of estimated parameters

Introduction

Side weirs are an overflow weir installed on the side of the main channel, which allows flow when water rises above the crest. This type of flow is considered as a spatially varied flow. Side weirs are usually used as a control structure and as head regulators in irrigation structure. Many studies deal with side weir hydraulics; some of these studies deal with sharp crested side weirs such as El-Khashab and Smith [\(1976](#page-8-0)), Uyumaz and Smith ([1991\)](#page-8-1) Swamee et al. ([1994](#page-8-2)), Hager [\(1987](#page-8-3)), Masoud ([2003](#page-8-4)), Singh et al. ([1994\)](#page-8-5), Rao and Pillai [\(2008\)](#page-8-6), Delkash and Babak ([2014](#page-8-7)) and other investigated deals with inclined and oblique side weir such as Mwafaq and Ahmed ([2011\)](#page-8-8). Honar and Javan [\(2007\)](#page-8-9) and Amir et al. [\(2016\)](#page-8-10). The numerical analysis on inclined side weir was investigated by Ahmed [\(2011\)](#page-8-11), Ahmed et al. [\(2013\)](#page-8-12) and Ahmed [\(2015](#page-8-13)). A powerful tool is recently used to solve complex nonlinear and multi-linear regression equations in hydraulic engineering such as artifcial neural network (ANN), genetic programming (GP) and statistic's analysis using Monte Carlo method, Kisi et al. [\(2012](#page-8-14)), Ahmed [\(2018\)](#page-8-15), Hayawi et al. [\(2019](#page-8-16)) and Ahmed and Anna ([2020](#page-8-17)). In the recent years GEP and learning machine were used to model of nonlinear problems of predicting discharge coefficient in side weir such as Isa et al. (2015) (2015) and Reza et al. [\(2020\)](#page-8-19). The aim of this study is to estimate MLR equation and compare equation modeled from GEP for coefficient of discharge calculation from skew side weir and then compare these values with values of C_d estimated from the rectangular side weir.

Experimental methodology

According to Al-Talib ([2012](#page-8-20)), the experimental works were achieved in rectangular laboratory channel 10 m long, 0.3 m wide and 0.45 m depth, while the side channel dimensions were 0.15 m wide, 0.3 m depth and 2 m long. The discharge was measured using standard sharp crested weir $(0.15 * 0.3 * 0.01)$ m dimensions at main channel; the side weirs were fxed at the entrance of the side channel by different angles starting from (90°) (perpendicular to the side channel) decreasing to (30°). Five diferent angles were taken (90 $^{\circ}$, 75 $^{\circ}$, 60 $^{\circ}$, 45 $^{\circ}$ and 30 $^{\circ}$) inclined to the left of flow direction. Figure [1.](#page-1-0)

Fig. 1 Sketch of side channel with skew side weir installed

Five discharges were taken ranging from $(7.3 \text{ to } 16.5 \text{ L/s})$; the actual discharge for the main channel was calculated from the equation.

$$
Q_{\text{act.}} = 0.58 \times H_{1.5} \tag{1}
$$

where Q_{act} = actual discharge and *H* = head over standard weir.

Equation ([1\)](#page-1-1) can be calculated by trial and error from volumetric calculations. The actual discharge was measured by closed side channel and measured depth of water over the standard weir at the end of main channel, then from Eq. [1](#page-1-1) found Q_1 , then open side channel and measured water depth over the standard weir at the end of the main channel once again and from Eq. [1](#page-1-1) found the discharge again, but discharge measured in this case (when side channel open) was Q_2 . Actual side channel discharge Q_3 used Eq. [2.](#page-1-2)

$$
Q_3 = Q_1 - Q_2 \tag{2}
$$

where Q_1 = actual discharge in main channel when side channel is closed, Q_2 = actual discharge in main channel when side channel is opened, and Q_3 =actual side channel discharge after subtracting Q_1 and Q_2 .

Theoretical methodology

The general flow through the side weir derived depends on head of water over the side weir as well as the velocity of fow through it, according to specifc energy assumption (De Marchi [1934\)](#page-8-21).

$$
E = y + \frac{V^2}{2g} \tag{3}
$$

where $E =$ specific energy, $y =$ head of water, $v =$ flow velocity and $g =$ gravity acceleration.

Depending on *Q*, Eq. [3](#page-2-0) can be written as discharge,

$$
Q = b \times y \sqrt{2g(E - P)}
$$
(4)

where $(b \times y)$ = cross-sectional area, *P* = weir height.

Depending on De Marchi, Eq. [4](#page-2-1) can be written as,

$$
q = -\frac{dQ}{dS} = \frac{2}{3}C_d\sqrt{2g(E - P)^{3/2}}
$$
 (5)

where $q =$ discharge per unit length, $S =$ longitudinal slope and C_d =coefficient of discharge.

Equation [5](#page-2-2) satisfes rectangular channel and side weir perpendicular to channel bed, so in skew side weir it is not perpendicular to channel bed; the angle for inclined side weir must have taken, and then, Eq. [5](#page-2-2) must change depending on these angles Fig. [1.](#page-1-0)

Dimensional analysis

Dimensional analysis is important to study the effects of the angle skew side weir to calculate coefficient of discharge (C_d) from standard side weir in rectangular channel. The parameters involved is calculated C_d in skew side weir as:

$$
C_{d} = f_{1}(v_{1}, y_{1}, P, L, b, g, \theta)
$$
\n(6)

where C_d = coefficient of discharge, v = flow velocity, y_1 =flow depth, *P*=weir height, *L*=weir length, *b*=channel width and g = acceleration due to gravity, θ = side weir angle.

By using Buckingham Pi theorem, parameters on Eq. [6](#page-2-3) can be used to develop a non-dimensional equation below:

$$
C_{\mathbf{d}} = \emptyset \left(F_r, \frac{P}{y_1}, \frac{L}{b}, \theta \right) \tag{7}
$$

where F_r = Froude number $\left(\frac{v}{\sqrt{gy}}\right)$.

Modeling of skew side weir using MLR

There are many applications for involved regression analysis. These applications deal with linear and nonlinear analysis, depending on variables that involve in the problem. In order to obtain a general equation for skew side weir, several trials with several equation models

examined using (Statistical Package Social Sciences SPSS user guide).

According to Ahmed ([2015](#page-8-13)), from Eq. [7](#page-2-4) using several models of SPSS, Eq. [8](#page-2-5) can be developed as MLR with a coefficient of determination R^2 (0.958)

$$
C_{d} = C_{1} + C_{2}Fr + C_{3}P/y + C_{4}L/b + C_{5}\theta
$$
\n(8)

where $C_1 - C_5$ = constants, and θ in radian.

GEP modeling for side weir

Gene Expression Programming was an artifcial procedure to solve genotype system. This way was invented by Ferreira [\(2001,](#page-8-22) [2006](#page-8-23)), and GEP was similar to (GA) genetic algorithms and (GP) genetic programming; GA deals with individuals as a linear string of length fxed (chromosomes), while GP deals with individuals as nonlinear entities for diferent parse tree structure. In GEP, the individuals deal with encoded linear strings (chromosomes) which are expressed as nonlinear entities. In GEP, there are two important players: the tree structure (ETS) and chromosomes. The decoding of the process information is called translation that implies obviously a type of code and rules. The genetic code of GEP was simple; a relation between the symbol of the chromosomes and the node is represented in the tree. The rules of GEP determine nodes in the trees and then the type of the interaction in sub-ETS. GEP programming depends on two principal languages: the genetic language and expression trees language. This bilingual notation in GEP is named as Karva. Figure [2](#page-3-0) shows the expression tree (ET) for an example of mathematical expressions $(xb + \sqrt{c + d})$ Mohd et al. [\(2015\)](#page-8-24) and Khalid and Negm [\(2008](#page-8-25)), This ET is encoded in Karva language, and then, the expression is called K-expression. Each gene starts at the frst left position, then scans all symbols in all directions every time when a symbol is fnally added to the K-expression, and then, the K-expression mentioned above can be written as $(+x\sqrt{ab + cd})$.

Figure [3](#page-3-1) shown the steps of GEP, include some steps at the begin with the randomly generate of the chromosome from initial population. Then, these chromosomes were expressed and excluded the tree expression to evaluate ftness. The individual is then selected with respect to their ftness to reproduce with the modifcation; these individuals are subject to the same development. This process was repeated several times until a good solution is found. (Ferreira [2004\)](#page-8-26) The basis of GEP is established on the structure of GEP gene. The simple structure of genes allows the encoding of thinkable program and allows their dynamic evolution due to these multilateral structural arrangements; a powerful set of genetics worker can be implemented to search efficiently solution Ferreira [2002.](#page-8-27)

Fig. 3 GEP fow chart

The equation obtained from GEP is given as:

$$
C_{d} = \left(A \tan \left(A \tan \left(A \tan \left(\tan \left(\tan \left(1101236493.83311 + \frac{P}{y}\right) + \cos \left(\exp \left\{\frac{L}{b}\right\}\right)\right)\right)\right)\right)^{4} + \frac{A \coth \left(\left(\left\{\theta \times 8173.64699339967\right\} - 5974.47259811262\right) \times \left(F_{r} - \frac{P}{y}\right)\right)}{\frac{\log(\theta)}{3.22324468279425} + \frac{P}{y} - \theta + \sqrt[3]{F_{r}}}\right)
$$

+
$$
A \coth \left(\left\{\left\{Csc \left(\frac{L_{f}}{P_{f y}}\right) \times \left(\frac{P}{y} + \frac{L}{b}\right)\right\} + \left\{\left(\frac{6049.86489822289 - \frac{L}{b}}{b}\right) - \frac{8224.16588065491}{L_{f y}}\right\}\right\} \times A \tanh \left(F_{r}\right)\right) (9)
$$

+
$$
+ \left\{A \csch \left(\exp \left\{\theta\right\} - A \tanh \left(\frac{\log \left(\exp \left\{\frac{L_{f y}}{560.990764280841}\right\}\right)\right)\right\}
$$

+
$$
\left\{F_{r}^{4} \times \left(\tan \left(\sin \left(\frac{6378.3895352031 \times \frac{P}{y} + \theta}{\frac{P}{y}}\right) \times \left(\sqrt{\frac{P}{y}} - \frac{6377.3895352031}{F_{r}}\right)\right)\right)\right\}
$$

The corresponding expression tree for the above equation is given in Fig. [4.](#page-4-0)

1 3

Fig. 4 Expression tree according to GEP equation

Fig. 4 (continued)

Table 1 GEP model parameters setting

Parameter	Setting
Chromosomes number	36
Genes number	5
Head size	12
Tail size	13
De size	13
Gene size	38
Linking function	Addition

Table 2 Statistics obtained from GEP run

Results and discussion

The genetic operation parameters setting is presented in Table [1](#page-5-0), while Table [2](#page-5-1) represents the statistics obtained from GEP after testing more than 1000 equation models and

running more than 11 h. The comparison between the results of the GEP and MLR presented in this study as well as MLR for previous studies illustrated in Table [3](#page-6-0) is presented in terms of coefficient of determination (R^2) , root-mean-square error (RMSE), Akaike information criteria (AIC) (AICs—which AIC with a correction for small sample sizes), mean absolute relative error (MARE) and scatter index (SI). These values are presented in Table [4](#page-6-1), and the equations are defned below:

$$
R^{2} = \left[\sum_{i=1}^{n} (x_{i} - \bar{x})(y_{i} - \bar{y}) / \sqrt{\sum_{i=1}^{n} (x_{i} - \bar{x})^{2} \sum_{i=1}^{n} (y_{i} - \bar{y})^{2}}\right]^{2}
$$
(10)

RMSE =
$$
\sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2}
$$
 (11)

$$
AIC = n \times \log \left(\frac{\sum_{i=1}^{n} (Fr_{Expi} - Fr_{Eqi})^2}{n} \right) + 2 \times k \tag{12}
$$

$$
AICs = AIC + \frac{2k^2 + 2k}{n - k - 1}
$$
 (13)

$$
MARE = \frac{1}{n} \sum_{i=1}^{n} \frac{|x_i - y_i|}{x_i}
$$
 (14)

$$
SI = \frac{RMSE}{\bar{x}}\tag{15}
$$

Table 3 Equations for calculating side weir discharge coefficient in previous lecturer

Table 4 Present work statistics comparison with previous studies

Fig. 6 Comparison between C_d calculated using MLR and C_d observed

Fig. 7 Comparison between C_d calculated using GEP and C_d observed

where x_i and y_i are the actual and modeled C_d values, respectively, \bar{x} and \bar{y} are the mean actual and modeled C_d values, respectively. *k* is the number of estimated parameters.

Results in Table [4](#page-6-1) represent statistics comparison for the present work with previous studies, and it may be seen that the GEP model refers to highest value of R^2 (0.994) and the lowest value of MARE and RMSE (0.00523 and 0.00465), respectively, as well as the AIC refers to the best value (-216.51) compared with all others equations, and all that indicate that the execution of GEP is the best with respect to other previous equations; overall, all values refer to a good agreement of equation for the present work compared with MLR according to Ahmed [\(2015\)](#page-8-13) and all other previous equations.

Figure [5](#page-6-2) shows a comparison between the discharge coefficients estimated using MLR models and GEP models,

respectively, while Figs. [6](#page-6-3) and [7](#page-7-0) show the discharge coefficient estimated using an MLR model and GEP model, respectively. Compared with observed coefficient of discharge, Figs. [5,](#page-6-2) [6](#page-6-3) and [7](#page-7-0) show agreement between compared coefficients of discharge computed from different models and observed values having a relative error below 5%.

Figure [8](#page-7-1) presents the discharge coefficient estimated from equations shown in Table [3](#page-6-0) as well as that value estimated from the present model with the observed value. According to these results, all values estimated from previous equation range from 0.45 to 0.75 for vertical side weir, while in the present work, values range between 0.65 and 0.85 for the skew side weir.

These values increased when the side angle increased; this means the discharge coefficient for skew side weir is greater than its values for vertical side weirs and these values for skew side weir increased when the side angle increased.

Conclusion

In the present study, a Gene Expression Programming (GEP) was used to predict an equation to calculate coefficient of discharge in skew side weir in a rectangular channel; this equation was compared with equation predicted from Multiple Linear Regression (MLR) which estimated from statistical tools. The two methods give a good result compared with the observed one with absolute error not exceeding 5% for both methods with correlation coefficient 0.9 and 0.996 for MLR and GEP, respectively, as well as the root-meansquare errors (RMSE) 0.0123 and 0.0046 for MLR and GEP, respectively. The results presented in this method compared with others equation calculated show that the accuracy of modeling and ftting of GEP is better than other methods.

Fig. 8 Comparison of C_d calculated using GEP and that values calculated using equations in Table [3](#page-6-0) from previous studies

This conclusion is made by considering the fact that the AIC for the number of parameters that ftted in model where its value (-216.51) is the best value compared with other equations, as well as the best value of the present work model for $(AICs = -918.51, MARE = 0.005234$ and $SI = 0.006231$) compared with other values of equations is presented in this study. The values of C_d for skew side weir were greater than its values for straight vertical. Finally, the results refer to using GEP that gives more accuracy than MLR and other previous literature equations in discharge coefficient calculation and may be used as an improved alternative technique.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahmed YM (2011) Theoretical analysis of fow over the side weir using Runge Kutta method. Int J Eng 9(2):47–50
- Ahmed YM (2015) Numerical analysis of flow over side weir. J King Saud Univ Sci 27:37–42. [https://doi.org/10.1016/j.jksue](https://doi.org/10.1016/j.jksues.2013.03.004) [s.2013.03.004](https://doi.org/10.1016/j.jksues.2013.03.004)
- Ahmed YM (2018) Artifcial neural network (ANN) model for end depth computations. J Civ Environ Eng 8(2):1–5. [https://doi.](https://doi.org/10.4172/2165-784X.1000316) [org/10.4172/2165-784X.1000316](https://doi.org/10.4172/2165-784X.1000316)
- Ahmed YM, Anna GJ (2020) Estimating the uncertainty of discharge coefficient predicted for oblique side weir using Monte Carlo method. Flow Meas Instrum 73(4):1–6. [https://doi.org/10.1016/j.](https://doi.org/10.1016/j.flowmeasinst.2020.101727) [fowmeasinst.2020.101727](https://doi.org/10.1016/j.flowmeasinst.2020.101727)
- Ahmed YM, Azza NA-T, Talal AB (2013) Simulation of fow over the side weir using simulink. Sci Iran 20(4):1094–1100
- Ali S, Ajmal H, Mujib A (2018) Lateral fow through the sharp crested side rectangular weirs in open channels. Flow Meas Instrum 59(3):8–17. [https://doi.org/10.1016/j.fowmeasinst.2017.11.007](https://doi.org/10.1016/j.flowmeasinst.2017.11.007)
- Al-Talib AN (2012) Flow over oblique side weir. Damascus Univ J 28(1):15–22
- Amir HZ, Hossein B, Shahboddin S (2016) Support vector regression for modified oblique side weirs discharge coefficient prediction. Flow Meas Instrum 51(10):1–7
- Borghei S, Jalili M, Ghodsian M (1999) Discharge coefficient for sharpcrested side weir in subcritical fow. J Hydraul Eng 125:1051– 1056. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1999\)125](https://doi.org/10.1061/(ASCE)0733-9429(1999)125)
- Borghei S, Jalili M, Ghodsian M (2003) Discharge coefficient for sharp-crested side weirs in subcritical flow. Water Marit Eng 156(2):185–191
- Cheong HF (1991) Discharge coefficient of lateral diversion from trapezoidal channel. J Irrig Drain Eng 117(4):321–333
- De Marchi G (1934) Saggio di teoria del funzionamento degli stramazzi laterali. L'Energ Elettr 11:849–860
- Delkash M, Babak EB (2014) An examination of rectangular side weir discharge coefficient equation under subcritical condition. Int J Hydraul Eng 3(1):24–34
- EL-Khashab A, Smith KVH (1976) Experimented investigation of fow over side weir. J Hydraul Eng 102(9):1255–1268
- Ferreira C (2001) Gene expression programming: a new adaptive algorithm for solving problems. Complex Syst 13:87–129
- Ferreira C (2002) "Gene expression programming in problem solving. In: Roy R, Köppen M, Ovaska S, Furuhashi T, Hofmann F (eds) Soft computing and industry. Springer, Berlin, pp 635–653
- Ferreira C (2004) Gene expression programming and the evolution of computer programs. In: de Castro LN, Zuben FJ (eds) Recent developments in biologically inspired computing. Idea Group Publishing, Hershey, pp 82–103
- Ferreira C (2006) Gene expression programming: mathematical modeling by an artifcial intelligence, 2nd edn (rev). Springer, Berlin
- Hager WH (1987) Lateral outfow over side weirs. J Hydraul Eng 113(4):491–504
- Hayawi HA, Al-Talib AN, Kattab NI (2019) Triangular side weir discharge coefficient calculation and comparison using ANN. INAE Lett 4(1):1–5. <https://doi.org/10.1007/s41403-019-00066-w>
- Honar T, Javan M (2007) Discharge coefficient in oblique side weirs. Iran Agric Res 25(2):27–36
- Isa E, Hossein B, Amir Hossein B, Hamed A, Ali Sh (2015) Gene expression programming to predict the discharge coefficient in rectangular side weir. Appl Soft Comput 35:618–628
- Jalili M, Borghei S (1996) Discussion: discharge coefficient of rectangular side weirs. J Irrig Drain Eng 122:132
- Khalid E, Negm A (2008) Performance evaluation of gene expression programming for hydraulic data mining. Int Arab J Inf Technol 5(2):126–131
- Kisi O, Emiroglu M, Guven A (2012) Prediction of lateral outfow over triangular labyrinth side weirs under subcritical conditions using soft computing approaches. Expert Syst Appl 39(3):3454–3460
- Masoud G (2003) Supercritical fow over a rectangular side weir. Can J Civ Eng 30(3):596–600
- Mohd M, Javed A, Mohd D (2015) Application of gene expression programming in food frequency analysis. J Indian Water Resour Soc 35(2):1–6
- Mwafaq YM, Ahmed YM (2011) Discharge coefficient for an inclined side weir crest using a constant energy approach. Flow Meas Instrum 22(6):495–499
- Ranga Raju KG, Prasad B, Grupta SK (1979) Side weir in rectangular channels. J Hydraul Div Proc 105(HY5):547–554
- Rao KH, Pillai GR (2008) Study of flow over side weirs under supercritical conditions. Water Resour Manag 22:131–143
- Reza G, Majeid H, Saeid K, Saeid S (2020) Simulation of discharge coefficient of side weirs placed on convergent canals using modern self-adaptive extreme learning machine. Appl Water Sci 10(1):1– 11.<https://doi.org/10.1007/s13201-019-1136-0>
- Singh R, Manivannana D, Stayanarayana T (1994) Discharge coefficient of rectangular side weirs. J Irrig Drain Eng 120(4):814–819
- Subramanya K, Awasthy SC (1972) Spatially varied fow over side weirs. J Hydraul Div Proc 98(HY1):1–10
- Swamee PK, Pathak SK, Ali MS (1994) Side weir analysis using elementary discharge coefficient. J Irrig Drain Eng 120(4):742-755
- Uyumaz A, Smith RH (1991) Design Procedure for fow over side weir. J Irrig Drain Eng 117(1):79–90

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

