Advances in Computational Design, *Vol. 6, No. 2 (2021) 135-151*
DOI: https://doi.org/10.12989/acd.2021.6.2.135

Discharge coefficient estimation for rectangular side weir using GEP and GMDH methods

Ajmal Hussain[1a], Ali Shariq*[1], Mohd Danish[2b] and Mujib A. Ansari[1c]

[1]Department of Civil Engineering, Zakir Hussain College of Engineering & Technology, Aligarh Muslim University, Aligarh-202002, India
[2]Civil Engineering Section, University Polytechnic, Aligarh Muslim University, Aligarh-202002, India

(Received June 18, 2020, Revised December 22, 2020, Accepted December 24, 2020)

Abstract. Flow through the rectangular side weir is a spatially varied type flow with decreasing discharge and used as a flow diversion structure. They are mainly used in the field of hydraulic, irrigation, and environmental engineering for diverting and controlling the flow of water in irrigation–drainage systems, drainage canal systems, and wastewater channels. In this study, gene expression programming and group method of data handling were used to estimate the coefficient of discharge for rectangular side weir under subcritical flow condition. Based on dimensional analysis, the coefficient of the discharge depends on the ratio of the crest height to length, ratio of the width of channel to crest length, ratio of the upstream depth in the channel to crest length and the approach Froude number. The performance of the proposed GMDH and GEP model is based on the coefficient of correlation (0.91), mean absolute percentage error (3.54), average absolute deviation (3.3), root mean square error (0.027) and the coefficient of correlation (0.905), mean absolute percentage error (4.12) average absolute deviation (3.9), root mean square error (0.029), respectively. Finally, the results reveal that GMDH model could provide more satisfactorily estimations as compared to those obtained by traditional regression and GEP models.

Keywords: rectangular side weir; coefficient of discharge; froude number; GMDH; GEP

1. Introduction

The side weirs may be of different shapes such as triangular, trapezoidal, rectangular or their combination according to application. They are generally used in river-control structures, reservoirs, dams, river-intake facilities, irrigation canals, and wastewater-treatment plants. The study of diversion of flow from the primary channel to the secondary channel, the main river to another river, or the main canal to sub-canal is important aspects for hydraulic engineering. The various hydraulic structures used to divert flow are weirs, spillway, sluice gate, and orifice. (Hussain *et al.* 2014, Hussain *et al.* 2016, Shariq *et al.* 2018, Ansari *et al.* 2019, Shariq *et al.* 2020). Spatially varied flow with decreasing discharge are observed in side weirs and side orifices

*Corresponding author, Ph.D. Student, E-mail: shariq.ali792@gmail.com
[a]Assistant Professor, E-mail: ajmalamin.iitr@gmail.com
[b]Assistant Professor, E-mail: mohd.danish999@gmail.com
[c]Professor, E-mail: mujibansari68@gmail.com

http://www.techno-press.org/?journal=acd&subpage=7 ISSN: 2383-8477 (Print), 2466-0523 (Online)

that are used for diverting water from irrigation or drainage systems, for controlling the water depth in a canal, and in flood schemes relief on the river.

In past studies, the extensive literature on side weirs is available due to its wide range of applications in environmental and hydraulic engineering. De Marchi (1934) provides the first theoretical approach on the hydraulics of rectangular side weir in a rectangular channel. Hydraulics and flow characteristics of rectangular side weir have been widely studied experimentally, theoretically and numerically for different shapes (rectangular, triangular, trapezoidal, and circular) of the channels by many researchers (De Marchi 1934, Emiroglu *et al.* 2011, Ranga Raju *et al.* 1979, Shariq *et al.* 2018, Shariq 2016, Vatankhah 2012, Hager 1987, Mohammed *et al.* 2013, Mohammed and Golijanek-Jędrzejczyk 2020).

Many Researcher's studies have formulated discharge coefficients equation for side weirs. The flow through the side weir in a rectangular channel has been the subject of many investigations (Subramanya and Awasthy 1972, Ranga Raju *et al.* 1979, Hager 1987). De Marchi (1934) provides the first theoretical approach for the discharge passed through the rectangular side weir in a rectangular channel. For developing a general expression, it is assumed that specific energy along the rectangular side weir is constant, uniform flow is maintained in the primary channel, and the edges of the rectangular side weir are sharp. One of the most common and fundamental bases for designing of side weirs is De Marchi's approach. Dominguez (1999) reported the following discharge equation for the rectangular side weir.

$$Q = \tfrac{4}{15} C_d L \sqrt{2g} \left[\frac{h_2^{2.5} - h_1^{2.5}}{(h_2 - h_1)} \right] \tag{1}$$

Where, Q is discharge passed through the rectangular side weir, g is the acceleration due to gravity, L is the crest length of the rectangular side weir, C_d is coefficient of discharge, and h is the head over the crest of rectangular side weir. The upstream and downstream sections of side weir are referred by the subscript 1 and 2, respectively. For developing a general expression, it is assumed that specific energy along the rectangular side weir is constant, uniform flow is maintained in the primary channel, and the edges of the rectangular side weir are sharp.

Kaveh *et al.* (2018a) adopted four soft computing-based techniques for Analysis of slope stability failures, Patient Rule-Induction Method (PRIM), M5 algorithm, Group Method of Data Handling (GMDH) and Multivariate Adaptive Regression Splines (MARS). Kaveh *et al.* (2018b) predicted shear strength of both FRP-reinforced concrete members with and without stirrups using the Group Method of Data Handling (GMDH) technique. Alkroosh and Sarker (2019) used gene expression programming (GEP) for predicting the compressive strength of fly ash geopolymer concrete. Kose and Kayadelen (2010) predicted the effects of infill walls on-base reactions and roof drift of reinforced concrete frames using adaptive neuro-fuzzy inference system (ANFIS) and gene expression programming (GEP). Khorrami and Derakhshani (2019) predict the ultimate bearing capacity of the shallow foundations using a combination of the M5-GP approach. Mohammed and Sharifi (2020) also provided the coefficient of discharge equation for obliged side weir using GEP method.

In recent past, various artificial intelligence techniques such as artificial neural networks (ANNs), adaptive neuro-fuzzy inference system (ANFIS), genetic programming, support vector machines (SVMs) were used extensively for solving various problems in different fields of civil engineering (Azmathulla *et al.* 2010, Ansari and Atthar 2013, Ansari *et al.* 2019, Ayaz and Mansoor 2018, Dutta *et al.* 2018, Alam *et al.* 2017, Ansari *et al.* 2018, Shao *et al.* 2014, Li *et al.*

2016, Saridemir 2016). Recently, the GMDH network is used in many fields to forecast and model the behaviours of unknown or complex systems based on different sets of multi-input-single-output data pairs (Amanifard *et al.* 2008). Moreover, in various researches such as energy conservation, economics and engineering geology, control engineering system identification, the GMDH approach is applied (Srinivasan 2008, Najafzadeh *et al.* 2013, Ansari 2014, Faisal *et al.* 2020, Rizvi *et al.* 2020).

The Gene Expression Programming technique is an extended form of genetic programming (GP), and it is an evolutionary artificial intelligence technique introduced by Ferreira. Gene Expression Programming evolves computer programs with various lengths and shapes encoded in linear chromosomes with a fixed size.

The present study aims to re-analyze the databases and to develop a GMDH and GEP model for the prediction of the coefficient of discharge of rectangular side weir. Few studies available in literature related to application of GMDH on side weir, an attempt has been made to developed a model to estimate a coefficient of discharge of side rectangular weir, which provide satisfactory results. The proposed equation obtained through the GMDH and GEP model is also compared with existing regression equations available in literature. Among all computational intelligence methods, the Group Method of Data Handling (GMDH) is known as a self-organized system with the capability of solving extremely complex nonlinear problems (Amanifard *et al.* 2008). This specific approach has been used because several studies related to application of GMDH methods have reported that it is one of the best approaches in dealing with problems related to water resources engineering.

2. Dimensional analysis

Dimensional analysis was performed to estimate the functional relationship for the coefficient of discharge for rectangular side weir. Coefficient of discharge of rectangular side weir can be expressed as a function of the upstream depth of flow (y_1), acceleration due to gravity (g), average flow velocity over the cross-section of the channel (V), the dynamic viscosity of water (μ), the density of water (ρ), a crest length of side weir (L), the width of the main channel (B), and crest height of side weir (P).

$$C_d = f\ P, L, B, g, V, y_1, \rho, \mu \tag{2}$$

$$C_d = f\left(y_1/L, P/L, F_1 = V/\sqrt{g y_1}, B/L \right) \tag{3}$$

3. Data collection

The data sets presented by Shariq *et al*. (2018), Azza and Al-Talib (2012), and Bagheri *et al*. (2014) have been used in this study. The experimental set-up of Shariq *et al*. (2018) consisted of a primary flume of length, width, and depth of 12.8 m, 0.29 m, and 0.39 m, respectively. A rectangular side weir was constructed on the right wall from the upstream end of the primary

Table 1 Range of experimental data for the present study

Parameters	Unit	Range of data
Q_1	*l/s*	7.1 – 44.6
Q2	*l/s*	0.4 – 29.07
B	*cm*	29 & 40
y_1	*cm*	9 – 32.1
L	*cm*	15 – 60.5
F_1	-	0.11-0.77

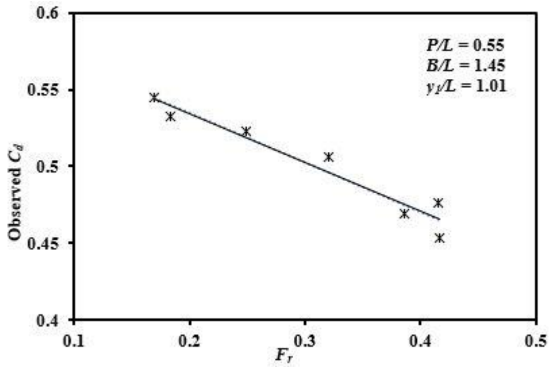


Fig. 1 Variation of C_d with Froude number

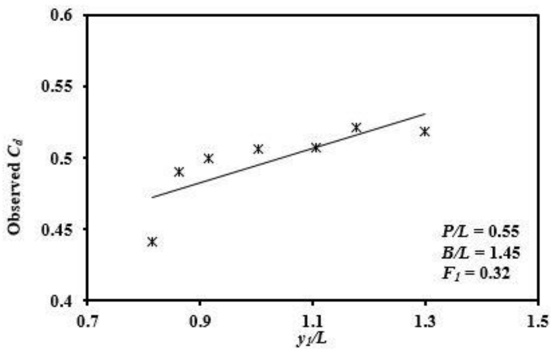


Fig. 2 Variation of C_d with y_1/L

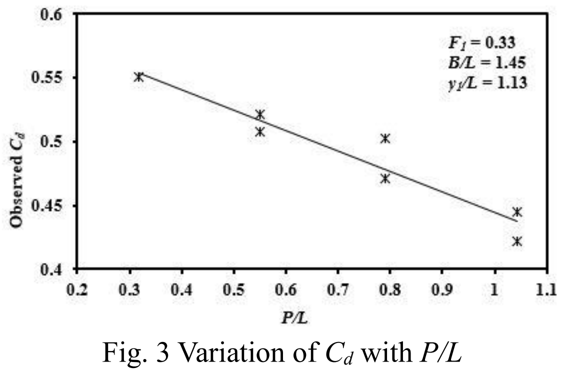


Fig. 3 Variation of C_d with P/L

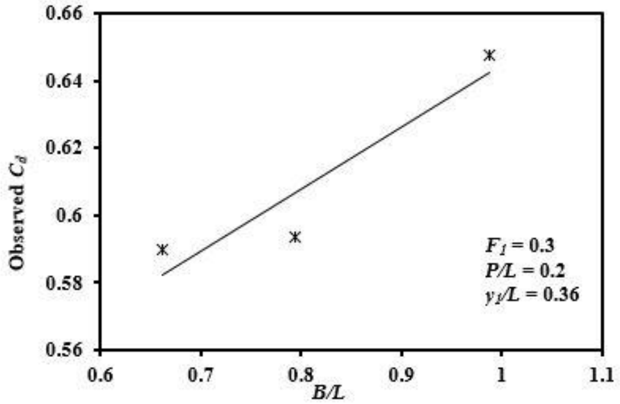


Fig. 4 Variation of C_d with *B/L*

Table 2 Available equation of C_d in literature

S.No.	Source	Discharge coefficient equations for rectangular side weirs
1.	Ghodsian (1997)	$C_d = (1-0.63F_1^{0.33})[0.611+0.075(y_1-P)/P]$
2.	Shariq *et al.* (2018)	$C_d = \left\{1.1308-1.5396\left(\frac{P}{L}\right)^{0.0394}-0.1492\left(F_1\right)^{0.8292}+0.0105\left(\frac{y_1}{L}\right)^{3.6295}+0.487\left(\frac{B}{L}\right)^{-0.0357}\right\}^{0.2322}$
3	Borghei *et al.* (1999)	$C_d = 0.55-0.47F_1$

channel at 8.20 m distance. Discharge over the rectangular side weir was passed into a secondary channel consisted of 4.18 m length, 0.2 m width, and 0.35 m depth and, then, moved to a return channel. The set-up of Bagheri *et al.* (2014) consisted of rectangular channels of length, height, and width are 8 m, 0.4 m, and 0.6 m, respectively. All the experiments conducted under subcritical flow conditions. The range of experimental data collected for the present study is shown in Table 1.

4. Analysis of data, results, and discussions

4.1 Effect of the dimensionless parameter on C_d

The effect of the dimensionless parameters y_1/L, F_1, P/L, and B/L on the observed coefficient of discharge, C_d was conducted. Thorough data analysis indicates that B/L, F_1, P/L, and y_1/L are the affecting dimensionless parameters for C_d. To show the variation of C_d against upstream Froude number, F_1 by keeping the other affecting parameters y_1/L, B/L, and P/L as constant, is shown in Fig. 1. It indicates that C_d decrease with the increase of F_1. In Fig. 2, the variation of C_d against y_1/L while keeping the affecting parameters F_1, B/L, and P/L as constant, indicates that C_d increases with the increase of y_1/L. Similarly, in Fig. 3 the variation of C_d against P/L, shows that C_d decreases with the increase in P/L when other affecting parameters such as y_1/L, B/L, and F_1 remain constant. The variation of C_d against B/L indicates that C_d increases with the increase of

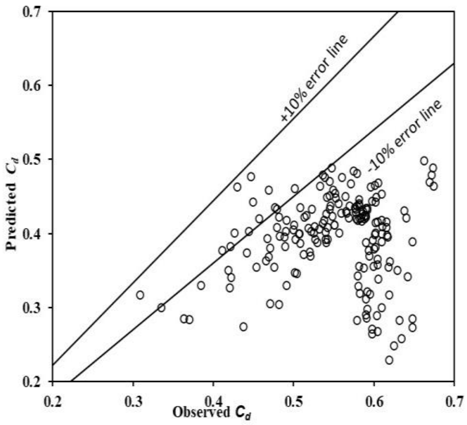


Fig. 5 Comparison between observed and predicted C_d for Bhorghei *et al.* (1999) model for all data sets

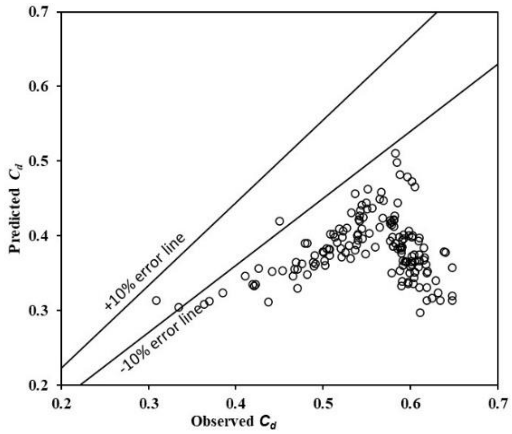


Fig. 6 Comparison between observed and predicted C_d for Ghodsian (1997) model for all data sets

B/L when other affecting parameters such as y_1/L, P/L, and F_1 remains constant, as shown in Fig. 4.

4.2 Accuracy of existing relationships for C_d

Extensive literature is available for the estimation of the coefficient of discharge. In order to verify the accuracy of the existing models, the entire available range of data was used. Table 1 shows the range of data for all the parameters used in the present investigation and Table 2 shows the models proposed by Borghei *et al.* (1999), Ghodsian (1997), and Shariq *et al.* (2018). These models were selected for comparison in the present study. The comparison between the observed C_d of rectangular side weir and those computed by the proposed available models are shown in Figs. 5-7, and the qualitative performance parameters are presented in Table 4. A close study of Figs. 5-7 reveals that none of the existing models was able to estimate the values of C_d of rectangular side weir for the range of data used in the present study.

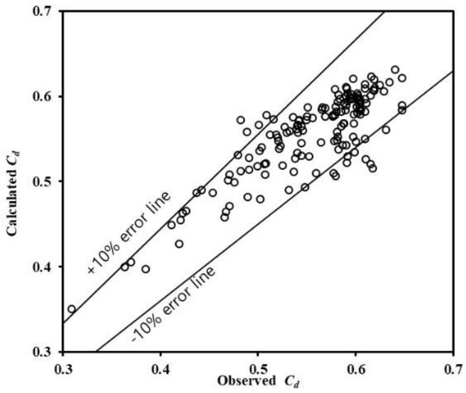


Fig. 7 Comparison between observed and predicted C_d for Shariq *et al.* (2018) model for all data sets

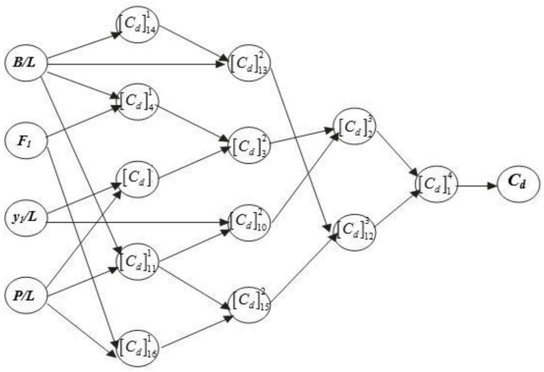


Fig. 8 Network Architecture of the GMDH model for predicting the coefficient of discharge

4.3 Proposed GMDH model for the coefficient of discharge of rectangular side weir

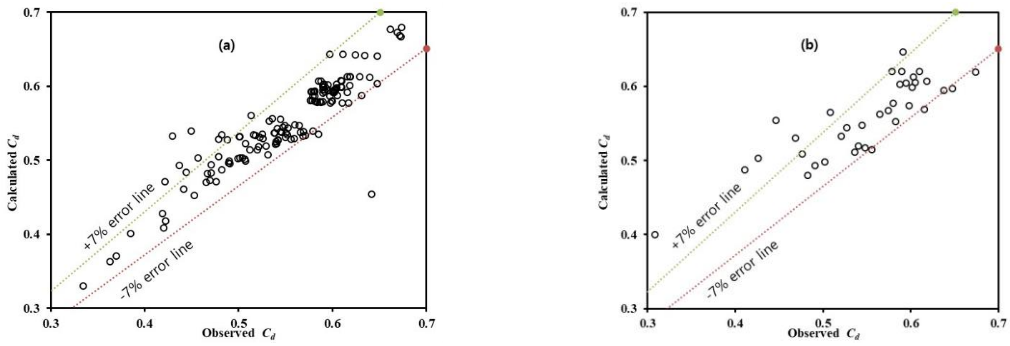
Group Method of Data Handling (GMDH) traditionally uses quadratic two-variable polynomial while developing the network. A modified form of GMDH network can be obtained by introducing several other types of polynomials and functions to enhance the performance of the model. In the present study, the GMDH network was modified by using two variable quadratic polynomial and one variable logarithmic function, as shown in Eqs. (4)-(5).

Quadratic: 2 variables $$\hat{y} = G(x_i x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \quad (4)$$

Log: 1 variable $$\hat{y} = G(x_i x_j) = a_0 + a_1 \log(x_i + a_2) \quad (5)$$

Besides, the results obtained by the GMDH model were compared with the regression models proposed by Borghei *et al.* (1999), Ghodsian (1997) and Shariq *et al.* (2018). The proposed GMDH network under consideration yielded a correlation coefficient of 0.91.

One of the critical properties of GMDH networks is that it provides analytical equations, which was obtained using a logarithmic function and quadratic polynomial. Analytical Eqs. (A1)-(A13)



(a) (b)

Fig. 9 Comparison between predicted and observed C_d using present GMDH model for training data sets

obtained by GMDH network for predicting C_d of rectangular side weir are presented in the Appendix.

In Eqs. (A1)-(A13), the subscript and superscript of each parameter represent the number of pertaining layers and neurons, respectively. The proposed structure of the GMDH network containing five selective neurons in the first layer, four selective neurons in the second layer, two selective neurons in the third and one selective neuron in the fourth respectively and a selective neuron in the output layer (5-4-2-1) for predicting the coefficient of discharge is presented in Fig. 8.

The predicted values of C_d have been plotted against its observed values for training and validation data sets, as shown in Fig. 9 for the GMDH model. It can be observed from Fig. 9 that most of the data lie within ±7% error band. Therefore, the GMDH model, along with corresponding logarithmic function with one variable and quadratic function with two variable polynomials (Eqs. (A1)-(A13)) is recommended for general use to predict C_d of rectangular side weir.

4.4 Proposed Gene Expression Programming model for the coefficient of discharge of rectangular side weir

Gene Expression Programming (GEP) is a procedure that mimics biological evolution to create a computer program to model some phenomena (Ferreira 2001, Azamathulla *et al.* 2011, Mohammed and Sharifi 2020). It is a system for encoding articulation that allows fast operation of an extensive range of mutations and cross-breeding methods while ensuring that the resulting expression will always be acceptable (Ferreira 2001, Ferreira 2006). It is associated with the principle of natural selection that is fit; healthier individuals should breed and yields generation at a rapid rate than unfit, sick individuals. Through this alternative process, each offspring becomes fitter and healthier.

The healthier individuals in each breed are unconditionally reproduced unchanged into the next breed. An expression tree is a better way to describe expression in a system because the tree can be complicated, and expression trees can be evaluated immediately (Ferreira 2001).

To identify the best combination of the model building parameter of GEP and determining the most favourable value of population size, gene head length, gene per chromosome, maximum

Table 3 GEP model parameters

Parameter	Setting
Population size	55
Number of genes per chromosome	05
Gene head length	12
Number of generations	10000
Generation without improvement	10000
Linking function	+
Fitness function	RRSE
Function set	+, -, ×, ÷, logistic 4
Chromosome length	66
Mutation rate	0.044
Inversion rate	0.1

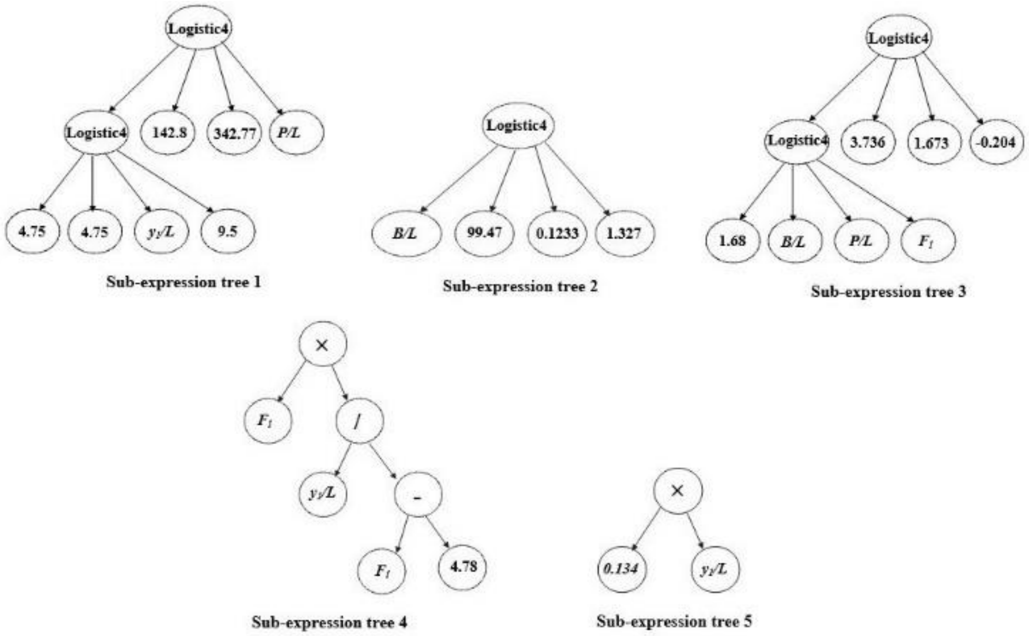


Fig. 10 Sub expression trees corresponding to each gene for the Eq. (6)

generation, and generations without improvement (GWI) was found by minimizing the variation between the estimated values and the desired output of GEP model. The GEP method has also been used for determining the C_d of the rectangular side weir. The performance of GEP models was deduced based on Mean Absolute Percentage Error (*MAPE*), Root Mean Squared Error *(RMSE*), Efficiency coefficient (E) & Average Absolute Deviation (*AAD*) and coefficient of correlation (*R*). The training of the GEP models was stopped when it achieved a satisfactory precision, or the maximum generation reached the recommended limit. Table 3 shows the parameters used in developing the GEP model.

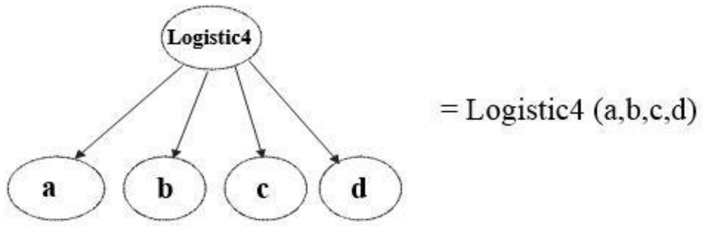
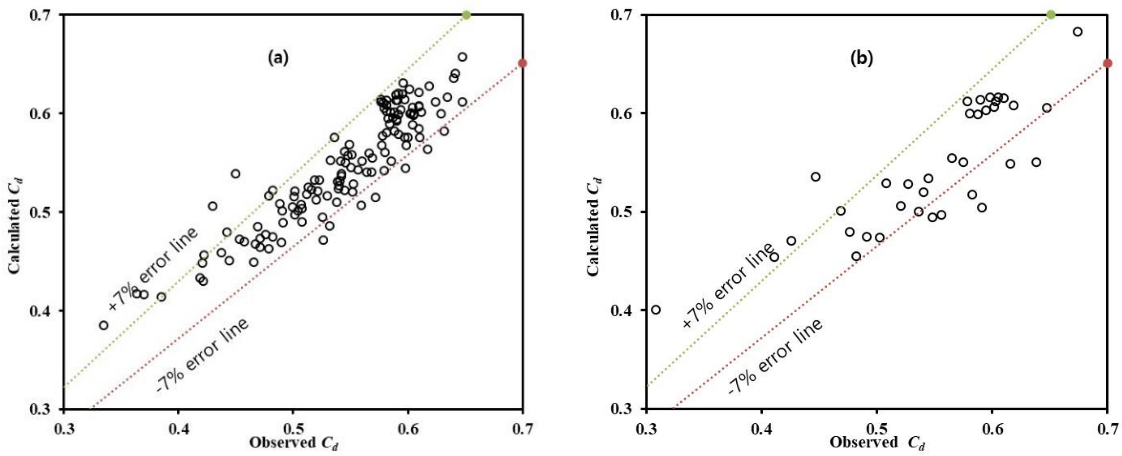


Fig. 11 The expression for the logistic function



(a) (b)

Fig. 12(a) Comparison between predicted and observed C_d using present GEP model for training data sets

The explicit formulation of the GEP model for C_d of rectangular side weir has been optimized as Eq. (7):

$$C_d = \frac{342.77}{1+\exp\left[142.8\left(\frac{y_1/L}{1+\exp\ 4.75\ 4.75-9.5} - P/L\right)\right]} + \frac{0.123}{1+\exp\ 99.47\ \ B/L-1.327}$$

$$+\frac{1.673}{1+\exp\left[3.736\left(\frac{P/L}{1+\exp\ \ B/L\times\ 1.68-F_1} - (-0.204)\right)\right]} + \left[F_1\times\left(\frac{y_1/L}{F_1-4.78}\right)\right] + (0.134\times y_1/L) \qquad (6)$$

From Eq. (6), it has been observed that there is sub-expression corresponding to each gene in the equation. The sub-expression trees of the gene are shown in Fig. 10. Logistic4 (a,b,c,d) is shown in Fig. 11 can be represented as Eq. (7).

The observed and predicted values of the C_d of rectangular side weir using a GEP model for the training and validation data are compared graphically, as shown in Fig. 12. It shows that the predicted C_d lies within ±7% of the observed values for training data as well as validation data, which is a better estimation of C_d for side rectangular sharp-crested weir. The qualitative performance of the present GEP model for all data sets has a mean absolute percentage error of 4.12 and the average absolute deviation of 3.9 with a coefficient of correlation of 0.905.

Table 4 Performance parameters of existing, GEP and GMDH models

		R	MAPE	AAD	RMSE	E
Ghodsian (1997)	Training	0.26	29.203	30.11	0.1792	-5.951
	Testing	0.28	29.403	30.15	0.1816	-6.640
	All	0.27	29.244	30.12	0.1797	-6.056
Shariq *et al.* (2018)	Training	0.87	4.95	4.79	0.0340	0.754
	Testing	0.85	4.65	4.65	0.0324	0.717
	All	0.87	4.89	4.73	0.0337	0.748
Borghei *et al.* (1999)	Training	0.076	27.847	28.876	0.1801	-6.538
	Testing	0.086	27.834	28.792	0.1804	-6.538
	All	0.081	27.844	28.859	0.1801	-6.090
GEP Model (Eq. (6))	Training	0.928	3.621	3.470	0.025	0.861
	Testing	0.832	6.260	5.815	0.042	0.689
	All	0.905	4.120	3.912	0.029	0.820
GMDH Model (Eq. (A13))	Training	0.912	3.158	3.071	0.028	0.827
	Testing	0.847	6.368	5.755	0.042	0.685
	All	0.91	3.454	3.301	0.027	0.832

Table 5 Comparison between existing relations, GEP and GMDH model

Source	Percentage of data having error less than			
	±4%	±8%	±12%	±16%
Ghodsian (1997)	0.55	3.29	3.38	9.34
Borghei *et al.* (1999)	1.64	3.83	9.87	18.11
Shariq *et al.* (2018)	53.84	78.56	92.29	99.98
GEP Model (Eq. (6))	63.18	87.45	96.24	100
GMDH Model (Eq. (A1)-(A13))	73.62	91.75	97.24	100

4.5 Comparison between GMDH, GEP model and available equations in literature

Tables 4 and 5 show the comparison between performance parameters and percentage error of GMDH, GEP model, and available equation of coefficient of discharge of rectangular side weir in literature. Both GMDH and GEP models predicted results satisfactorily as compared to the available equations of C_d for rectangular side weir. The qualitative performance of the present GEP has lowest *MAPE* (4.12), *AAD* (3.9), *RMSE* (0.029), *E* (0.820), highest *R* (0.905) and GMDH model has lowest *MAPE* (3.45), *AAD* (3.33), *RMSE* (0.027), *E* (0.832), highest *R* (0.91), respectively, which indicates that it has better performance as compared to other existing predictors. The percentage of data having error less than ±8% for Ghodsian (1997), Borghei *et al.* 1999, and Shariq *et al.* 2018 have been found 3.29%, 3.83%, and 78.56%, respectively, which were lesser as compared to present GEP and GMDH model. The proposed GEP and GMDH models provided results with a maximum error of ±12% for about 96.24% and 97.24 % of the total data, respectively, that shows the favourable performance of the present GEP and GMDH models.

5. Conclusions

In this study, the Group method of data handling (GMDH) and Gene expression programming (GEP) model have been used to estimate the coefficient of discharge for rectangular side weir.

• The variation of C_d with the upstream Froude number shows that C_d decreases with the increase of Froude number.

• The variation of C_d with P/L indicates that C_d decreases with the increase of P/L. The variation of C_d with y_1/L indicates that C_d is directly proportioned to y_1/L.

• Observed and calculated values of C_d of rectangular side weir using GMDH model for the test data are compared graphically. It shows that the computed C_d lies within ±7% of the observed values, which may be considered as a satisfactory estimation of the coefficient of discharge for rectangular side weir.

• The qualitative performance of the present GEP model for all data sets has Mean absolute percentage error (4.12) & average absolute deviation (3.9), root mean square error (0.029), efficiency coefficient (0.820), and coefficient of correlation (0.905).

• The qualitative performance of the present GMDH model indicates that it has the lowest *MAPE* (3.4), *AAD* (3.33), *RMSE* (0.027), *E* (0.832) and highest *R* (0.91) as compared to other existing predictors.

• Proposed GEP and GMDH model provides much better results as compared to the available models in the literature (Shariq *et al.* 2018, Bhorghei *et al.* 1999, Ghodsian 1997).

• The proposed GEP and GMDH models produced results with a maximum error of ±12% for about 96.24% and 97.24% of the total data, respectively, that shows the excellent performance of both the models.

References

Ahmad, F., Ansari, M.A., Hussain, A. and Jahangeer, J. (2020), "Model development for estimation of sediment removal efficiency of settling basins using group methods of data handling", *J. Irrigation Drainage Eng.*, **147**(2), 04020043. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001532.

Alam, J., Kim, D. and Choi, D. (2017), "Seismic probabilistic risk assessment of weir structures considering the earthquake hazard in the Korean Peninsula", *Earthq. Struct.*, **4**(13), 421-427. https://doi.org/10.12989/eas.2018.13.4.421.

Alkroosh, I.S. and Sarker, P.K. (2019), "Prediction of the compressive strength of fly ash geopolymer concrete using gene expression programming", *Comput. Concrete*, **24**(4), 295-302. https://doi.org/10.12989/cac.2019.24.4.295.

Amanifard, N., Nariman-Zadeh, N., Farahani, M.H. and Khalkhali, A. (2008), "Modeling of multiple short-length-scale stall cells in an axial compressor using evolved gmdh neural networks", *J Energy Convers Manage*, **49**(10), 2588-2594. https://doi.org/10.1016/j.enconman.2008.05.025.

Ansari, M.A. (2014), "Sediment removal efficiency computation in vortex settling chamber using artificial neural networks", *Water Energy Int.,* **71**(1), 54-67.

Ansari, M.A. and Athar M. (2013), "Artificial neural networks approach for estimation of sediment removal efficiency of vortex settling basins", *ISH J. Hydraulic Eng.*, **19**(1), 38-48. doi:10.1080/09715010.2012.758415.

Ansari, M.A., Ansari, S.A. and Alam, S. (2018), "Computation of scour depth below pipelines using artificial neural networks", *Water Energy Int.*, **61**(6), 55-62.

Ansari, M.A., Hussain, A., Shariq, A. and Alam, F. (2019), "Experimental and numerical study for the estimation of coefficient of discharge for side compound weir", *Canadian J. Civil Eng.,* **46**(10), 887-895.

https://doi.org/10.1139/cjce-2017-0689.

Ayaz, M. and Mansoor, T. (2018), "Discharge coefficient of oblique sharp crested weir for free and submerged flow using trained ann model", *Water Sci*, **32**(2), 192-212. https://doi.org/10.1016/j.wsj.2018.10.002.

Azamathulla, H.M., Ghani, A.A., Leo, C.S., Chang, C.K. and Zakaria, N.A. (2011), "Gene-expression programming for the development of a stage-discharge curve of the pahang river", *Water Resources Management*, **25**, 2901-2916.

Azamathulla, H.M., Ghani, A.A., Zakaria, N.A. and Guven, A. (2010), "Genetic programming to predict bridge pier scour", *J. Hydraulic Eng.*, **136**(3),165-169. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000133.

Azza, N. and Al-Talib, (2012), "Flow over oblique side weir", *J. Damascus Univ.*, **28**(1), 15-22.

Bagheri, S., Kabiri-Samani, A.R. and Heidarpour, M. (2014), "Discharge coefficient of rectangular sharp-crested side weirs part i: traditional weir equation", *Flow Measure. Instrumentation,* **35**, 109-115. https://doi.org/10.1016/j.flowmeasinst.2013.11.005.

Borghei, S.M., Jalili, M.R. and Ghodsian, M., (1999), "Discharge coefficient for sharp-crested side weirs in subcritical flow", *J. Hydraul Eng.,* **125**(10), 1051-1056. https://doi.org/10.1061/(ASCE)0733-9429(1999)125:10(1051).

Dutta, S., Samui, P. and Kim, D. (2018), "Comparison of machine learning techniques to predict compressive strength of concrete", *Comput. Concrete*, **21**(4), 463-470. https://doi.org/10.12989/cac.2018.21.4.463.

Emiroglu, M.E., Agaccioglu, H. and Kaya, N. (2011), "Discharging capacity of rectangular side weirs in straight open channels", *Flow Measure. Instrumentation*, **22,** 319-330. https://doi.org/10.1016/j.flowmeasinst.2011.04.003.

F.J. Domínguez, (1999) Editorial Universitaria, Santiago, Chile, (in Spanish).

Ferreira, C. (2001), "Gene expression programming: a new adaptive algorithm for solving problems", *J. Complex Syst.,* **13**(2), 87-129.

Ferreira, C. (2006), "Gene expression programming; mathematical modelling by an artificial intelligence", Springer, Heidelberg.

Ghodsian, M. (1997), "Elementary discharge coefficient for rectangular side weir", *Proceeding of 4th Int. Conf. on Civil Engineering*, Tehran.

Hager, W.H. (1987), "Lateral outflow of side weirs", *J. Hydraul. Eng.*, **113**(4), 491-504. https://doi.org/10.1061/(ASCE)0733-9429(1987)113:4(491).

Hussain, A., Ahmad, Z. and Ojha, C.S.P. (2016), "Flow through lateral circular orifice under free and submerged flow conditions", *Flow Measure. Instrumentation*, **36**(10), 32-35. https://doi.org/10.1016/j.flowmeasinst.2016.09.007.

Hussain, S., Hussain, A. and Ahmad, Z. (2014), "Discharge characteristics of orifice spillway under oblique approach flow", *Flow Measure. Instrumentation*, **39**, 9-18. https://doi.org/10.1016/j.flowmeasinst.2014.05.022.

Kaveh, A., Bakhshpoori, T. and Hamze-Ziabari, S.M. (2018b), "Gmdh-based prediction of shear strength of frp-rc beams with and without stirrups", *Comput. Concrete*, **22**(2), 197-207. https://doi.org/10.12989/cac.2018.22.2.197.

Kaveh, A., Hamze-Ziabari, S.M. and Bakhshpoori, T. (2018a) "Soft computing-based slope stability assessment: A comparative study", *Geomech. Eng.*, **14**(3), 257-269. http://dx.doi.org/10.12989/gae.2018.14.3.257.

Khorrami, R. and Derakhshani, A. (2019), "Estimation of ultimate bearing capacity of shallow foundations resting on cohesionless soils using a new hybrid m5\gp model", *Geomech. Eng.,* **19**(2), 127-139. https://doi.org/10.12989/gae.2019.19.2.127.

Kose, M.M. and Kayadelen, C. (2010), "Effects of infill walls on RC buildings under time history loading using genetic programming and neuro-fuzzy", *Struct. Eng. Mech.*, **47**(3), 401-419. https://doi.org/10.12989/sem.2013.47.3.401.

Li, S., Yu, S., Shangguan, Z. and Wang, Z. (2016), "Estimating model parameters of rockfill materials based on genetic algorithm and strain measurements", *Geomech. Eng.*, **10**(1), 37-48. http://dx.doi.org/10.12989/gae.2016.10.1.037.

Marchi, D. (1934), "G. Essay on the performance of lateral weirs", *L Energia Electrica Milano*, **11(**11), 849-860.

Mohammed, A.Y. and Golijanek-Jędrzejczyk, A. (2020), "Estimating the uncertainty of discharge coefficient predicted for oblique side weir using monte carlo method", *Flow Measure. Instrumentation*, **73**(4),1-6. https://doi.org/10.1016/j.flowmeasinst.2020.101727.

Mohammed, A.Y. and Sharifi, A. (2020), "Gene expression programming (gep) to predict coefficient of discharge for oblique side weir", *Appl. Water Sci.*, **10**, 145. https://doi.org/10.1007/s13201-020-01211-5.

Mohammed, A.Y., Al-Talib, A.N. and Basheer, T.A. (2013), "Simulation of flow over the side weir using simulink. Scientiairanica", **20**(4), 1094-1100.

Najafzadeh, M., Barani, G.A. and Hessami Kermani, M.R. (2013), "Abutment scour in clear-water and live-bed conditions by GMDH network", *Water Sci. Technol.*, **67**(5), 1121-1128. https://doi.org/10.2166/wst.2013.670.

Ranga Raju, K.G., Prasad, B. and Gupta, S.K. (1979), "Side weirs in rectangular channels", *J. Hydraul Div*. **105**(5) 547-554.

Rizvi, Z.H., Baqir Husain, S.M., Haider, H. and Wuttke, F. (2020), "Effective thermal conductivity of sands estimated by group method of data handling (gmdh), *Proc. of Materials Today*, **26**(2), 2103-2107.

Saridemir, M. (2016), "Empirical modeling of flexural and splitting tensile strengths of concrete containing fly ash by gep", *Comput. Concrete*, **17**(4), 489-498. https://doi.org/10.12989/cac.2016.17.4.489.

Shao, G., Jiang, L. and Chouw, N. (2014), "Experimental investigations of the seismic performance of bridge piers with rounded rectangular cross-sections", *Earthq. Struct.*, **7**(4), 463-484. https://doi.org/10.12989/eas.2014.7.4.463.

Shariq, A. (2016), *Flow characteristics of side weirs in open channel*, Master's Thesis, Aligarh Muslim University, Aligarh, India.

Shariq, A., Hussain, A. and Ahmad, Z. (2020), "Discharge equation for the gabion weir under through flow condition", *Flow Measurement Instrumentation*, **74**, 101769.

Shariq, A., Hussain, A. and Ansari, M.A. (2018), "Lateral flow through the sharp crested side rectangular weirs in open channels", *Flow Measure. Instrumentation*, **59,** 8-17. https://doi.org/10.1016/j.flowmeasinst.2017.11.007.

Srinivasan, D. (2008), "Energy demand prediction using gmdh networks", *Neuro Computing*, **72**(1-3), 625-629. https://doi.org/10.1016/j.neucom.2008.08.006.

Subramanya, K. and Awasthy, S.C. (1972), "Spatially varied flow over side weirs", *J. Hydraul Div. (ASCE)*, **98**(1), 1-10. https://doi.org/10.1061/JYCEAJ.0003188.

Vatankhah, A. (2012), "Analytical solution for water surface profile along a side weir in a triangular channel", *Flow Measure. Instrument.*, **23**(1), 76-79. https://doi.org/10.1016/j.flowmeasinst.2011.10.001.

CC

Appendix

$$\left[C_d\right]_4^1 = 0.635 - 0.228*B/L + 0.127\left(B/L\right)^2 + 0.713F_1 - 0.289F_1^2 - 0.526*B/L*F_1 \quad \text{(A1)}$$

$$\left[C_d\right]_7^1 = 0.707 - 0.178*y_1/L + 0.093*\left(y_1/L\right)^2 - 0.38*P/L + 0.079\left(P/L\right)^2 + 0.140*y_1/L*P/L \quad \text{(A2)}$$

$$\left[C_d\right]_3^2 = -3.157 + 8.822\left[C_d\right]_4^1 - 0.703\left(\left[C_d\right]_4^1\right)^2 + 3.135\left[C_d\right]_7^1 + 4.306\left(\left[C_d\right]_7^1\right)^2 - 13.053\left[C_d\right]_4^1\left[C_d\right]_7^1 \quad \text{(A3)}$$

$$\left[C_d\right]_{11}^1 = 0.9 - 0.406*B/L + 0.1328*\left(B/L\right)^2 - 0.452*P/L + 0.1317*\left(P/L\right)^2 + 0.166*B/L*P/L \quad \text{(A4)}$$

$$\left[C_d\right]_{10}^2 = -4.119 + 0.172y_1/L + 0.103\left(y_1/L\right)^2 + 15.558\left[C_d\right]_{11}^1 - 12.62\left(\left[C_d\right]_{11}^1\right)^2 - 0.621\left(y_1/L\right)\left[C_d\right]_{11}^1 \quad \text{(A5)}$$

$$\left[C_d\right]_2^3 = -0.267 + 7.907\left[C_d\right]_3^2 - 5.628\left(\left[C_d\right]_3^2\right)^2 - 5.975\left[C_d\right]_{10}^2 - 17.703\left(\left[C_d\right]_{10}^2\right)^2 - 24.149\left[C_d\right]_{10}^2\left[C_d\right]_{10}^2 \quad \text{(A6)}$$

$$\left[C_d\right]_{14}^1 = 0.534 - 0.0282*\log\left(B/L - 0.615\right) \quad \text{(A7)}$$

$$\left[C_d\right]_{13}^2 = -1665.113 - 1107.707*B/L + 18.856\left(B/L\right)^2 + 5929.096\left[C_d\right]_{14}^1 - 5627.62\left(\left[C_d\right]_{14}^1\right)^2 - 2138.981\left(B/L\right)\left(\left[C_d\right]_{14}^1\right) \quad \text{(A8)}$$

$$\left[C_d\right]_{16}^2 = 0.694 - 0.112F_1 - 0.085\left(F_1\right)^2 - 0.192P/L + 0.195\left(P/L\right)^2 - 0.588*F_1*P/L \quad \text{(A9)}$$

$$\left[C_d\right]_{15}^2 = -1.278 + 3.489\left[C_d\right]_{16}^1 - 1.57\left(\left[C_d\right]_{16}^1\right)^2 + 1.997\left[C_d\right]_{11}^1 - 0.078\left(\left[C_d\right]_{11}^1\right)^2 - 2.26\left[C_d\right]_{16}^1\left[C_d\right]_{11}^1 \quad \text{(A10)}$$

$$\left[C_d\right]_{12}^3 = -0.904 - 0.826\left[C_d\right]_{13}^2 - 5.795\left(\left[C_d\right]_{13}^2\right)^2 + 4.986\left[C_d\right]_{15}^2 - 0.237\left(\left[C_d\right]_{15}^2\right)^2 - 8.784\left[C_d\right]_{13}^2\left[C_d\right]_{15}^2 \quad \text{(A11)}$$

$$\left[C_d\right]_1^4 = 0.694 + 5.414\left[C_d\right]_2^3 - 1.837\left(\left[C_d\right]_2^3\right)^2 - 7.025\left[C_d\right]_{12}^3 - 13.133\left(\left[C_d\right]_{12}^3\right)^2 - 12.558\left[C_d\right]_2^3\left[C_d\right]_{12}^3 \quad \text{(A12)}$$

$$C_d = 0.548 + 0.983*\log\left(\left(\left[C_d\right]_1^4\right) + 0.4535\right) \quad \text{(A13)}$$

Appendix II: Performance indices

The qualitative performances of the available equations in terms of coefficient of correlation (R), Root mean square error (RMSE), Mean absolute percentage error (MAPE), and Average Absolute Deviation (AAD) are also calculated and defined below.

The coefficient of correlation describes the degree of co-linearity between simulated and measured data, which ranges from -1 to +1, and is an index of the degree of the linear relationship between observed and simulated data. If R = 0, no linear relationship exists. If R = ±1, a perfect positive or negative linear relationship exists. Its equation is

$$R = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(C_{d0}(i)-\overline{C}_{d0}\right)\left(C_{df}(i)-\overline{C}_{df}\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(C_{d0}(i)-\overline{C}_{d0}\right)^2}\times\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(C_{df}(i)-\overline{C}_{df}\right)^2}} \tag{8}$$

R and R^2 have widely been used for model evaluation, though they are oversensitive to high extreme values (outlier) and insensitive to additive and proportional differences between model predictions and measure data.

Mean Absolute Percentage Error (MAPE) is a measure of the accuracy in a fitted time series value in statistics and has been used for discharge prediction evaluation. It expresses the accuracy as a percentage and is defined as

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{C_{df}(i)-C_{do}(i)}{C_{do}(i)}\right|\times 100 = Mean(|e|) \tag{9}$$

where $C_{do}(i)$ and $C_{df}(i)$ are observed and predicted discharge, respectively. $\overline{C_{do}}$ & $\overline{C_{df}}$ denote their mean observed and predicted discharge respectively, and n is a number of data considered.

The average absolute deviation (AAD) or simply deviation of a data set is the average of an absolute deviation from a central point. In the general form, the central point can be the mean, median, mode, or the result of another measure of central tendency.

$$AAD = \frac{1}{n}\sum_{i=1}^{n}\left|X_i - n(X)\right| \tag{10}$$

Root Mean Squared Error (RMSE) is often used to measure the difference between values predicted by a model and those actually observed from the thing being modeled. RMSE is one of the commonly used error-index statistics and is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(C_{df}(i)-C_{do}(i)\right)^2}{n}} \tag{11}$$

The Nash-Sutcliffe model efficiency coefficient is used to assess the predictive power of hydrological models. It is the normalized statistic that determines the relative magnitude of the residual variance ("noise") compared to the measured data variance and indicates how well the plot of observed versus predicted data fits the 1:1 line. It is defined as

$$E = 1 - \frac{\sum_{i=1}^{n}\left(C_{do}(i) - C_{df}(i)\right)}{\sum_{i=1}^{n}\left(C_{do}(i) - \overline{C_{do}}\right)} \tag{12}$$

Nash-Sutcliffe efficiencies ranges between (-∞, 1]: E=1 correspond to a perfect match of predicted coefficient of discharge to the observed data; E=0 shows that the model are as accurate as the mean of the observed data; and -∞<E<0 occurs when the observed mean is a better than the model, which indicates unacceptable performance.