Prediction of Side Weir Discharge Coefficient by Genetic Programming Technique

Abbas Parsaie [1] ***and Amir Hamzeh Haghiabi*** [2]

[1] Ph.D. Candidate of Hydrostructures, Department of Water Engineering, Lorestan University, Khorram Abad, Iran.
E-Mail: Abbas_Parsaie@yahoo.com
[2] Associate Professor, Department of Water Engineering, Lorestan University, Khorramabad, Iran.
E-Mail: haghiabi@yahoo.com

ABSTRACT

Side weir has many possible uses in hydraulic engineering and has also been investigated as an important structure in hydrosystems. In this paper, the genetic programing technique was used to predict side weir discharge coefficient. The main parameters which are efficacious in the side weir discharge coefficient are: Froude number (Fr), ratio of side weir height to total upstream head ($\frac{P}{h_1}$), ratio of side weir length to total upstream head ($\frac{L}{h_1}$) and ratio of side weir length to channel width ($\frac{L}{b}$). Principal component analysis indicates that the most important parameters are: Fr_1 and $\frac{P}{h_1}$. The results show that the most accurate empirical formula is the Emiroglu formula with error indices (R^2 = 0.64 and RSME=0.1). The performance of GP was compared to the empirical formulae proposed to calculate Cd_{sw}. The results of the GP model indicate that the accuracy of the GP model with error indices (R^2 = 0.95 and RSME=0.09) is suitable. At the end, a formula is proposed to calculate Cd_{sw} based on the GP approach.

KEYWORDS: Hydraulic structures,Weir, GP, Discharge coefficient, Principal component analysis.

INTRODUCTION

Modeling of hydraulic structures is the main part of hydraulic engineering research (Ettema, 2000). Modeling of hydraulic structures has been usually conducted by physical laboratory model creation and numerical simulation. Investigators, by studying the physical hydraulic models, tried to define the various hydraulic characteristics of hydraulic structures. Results of the experimental studies are usually declared by an empirical formula or by presenting a graph. Empirical formulae are derived by classical regression. Studies on side weir hydraulic properties were first conducted by experimental approaches. The aims of these studies were defining the water surface profile along the side weir at sub-critical and super-critical flow conditions and defining the side weir discharge coefficient to improve the efficiency of side weir by proposing various shapes of the crest of side weir (El-Khashab and Smith, 1976; Uyumaz and Muslu, 1985; Hager, 1987; Uyumaz and Smith, 1991; Cheong, 1991; Swamee et al., 1994b; Swamee et al., 1994a; Singh et al., 1994; Swamee et al., 1994c; Jalili and Borghei, 1996; Borghei et al., 1999; Ghodsian, 2003; Coşar and Agaccioglu, 2004; Durga Rao and Pillai, 2008; Borghei and Parvaneh, 2011; Kaya et al., 2011; Emiroglu et al., 2011a; Emiroglu and Kaya, 2011; Rahimpour et al., 2011; Haddadi and Rahimpour, 2012; Vatankhah, 2012, 2013a; Vatankhah, 2013b). Figure (1) shows a diagram scheme of side weir at sub-critical flow condition.

Received on 27/8/2015.
Accepted for Publication on 15/11/2016.

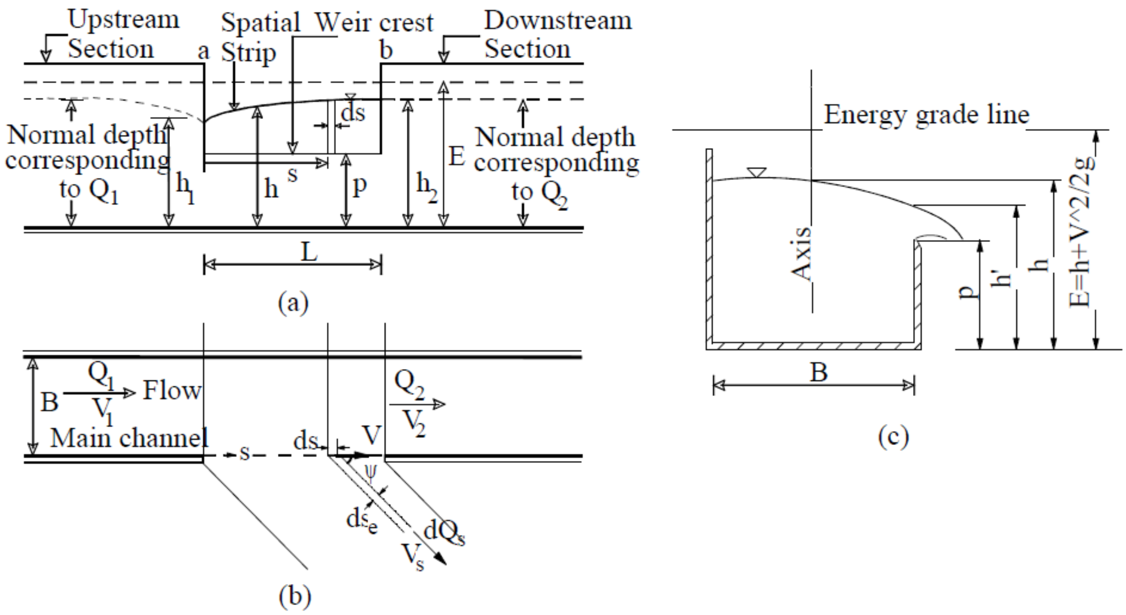


Figure (1): sketch of side weir at sub-critical flow condition

In Figure (1), L is the side weir length, B(b) is the channel width, P is the weir height, h_1 is the upstream flow depth, h_2 is the downstream flow depth, Q_0 is the upstream discharge, Q_s is the discharge passing through the side weir and Q_2 is the downstream discharge. Researchers who conducted experimental studies on side weir hydraulics tried to explain the effect of influential parameters, such as Froude number, weir length, flow angle, crest shape, among others, on the side weir discharge coefficient. Because of the high cost of physical modeling and laboratory equipment, investigators used numerical approaches for simulating the performance of hydraulic structures. In the field of numerical simulation, the governing hydraulic equations were solved by numerical approaches. Another way of numerical simulation is using the computational fluid dynamics (CFD) approach. Numerical simulation shows the flow pattern, pressure and velocity distribution (Parsaie and Haghiabi, 2014; Aydin, 2012; Aydin and Emiroglu, 2013). Recently, by advancing the artificial intelligence techniques, such as neural network (ANN) models in water engineering, predicting the hydraulic phenomena has been conducted with more accuracy. Developing ANN models is based on the data set. This means that the hydraulic characteristics of the phenomena should be measured in advance. In the field of ANN, the Adaptive Neuro Fuzzy Inference System (ANFIS) was used by Emiroglu and Kisi (2013) and the Multi-layer Perceptron (MLP) neural network was implemented by Kisi et al. (2012), Bilhan et al. (2010), Bilhan et al. (2011) and Emiroglu et al. (2011b). The results of all the ANN studies on the side weir indicate that the accuracy of ANN models is much higher than that of empirical formulae. The ANN model and the ANFIS model present a network instead of a formula, so the researcher hasn't more detailed information on the processes which are carried out inside the ANN model. By advancing the neural network models, today another type of neural networks is present which gives a formula in addition to the network. In this paper, the genetic programing technique is used to predict the discharge coefficient of side weir.

MATERIALS AND METHODS

Discharge coefficient of side weir as stated in the literature is a function of hydraulic and geometric characteristics. The main parameters which have influence on side weir are given in Eq. (1).

$$Cd_{sw} = f\ (v_1, L, b, h_1, P, \psi, s_0) \qquad (1)$$

By using dimensional analysis techniques, such as Buckingham π theory, researchers reduce the number of experiments. The dimensionless parameters derived from the dimensional analysis process are given in Eq. (2).

$$Cd_{sw} = f_2\left(Fr_1, \frac{L}{b}, \frac{L}{h_1}, \frac{P}{h_1}\right) \qquad (2)$$

where Fr_1 is the Froude number. Equation (2) is a basic formula for developing experimental studies and artificial intelligent techniques on side weir hydraulics. A summary of the most famous empirical formulae is given in Table (1). As mentioned in the literature, developing ANN models is based on the data set. So, for predicting Cd_{sw} by genetic programing techniques, about 477 data sets related to Eq. (2) were published in credible journals and the range of them is given in Table (2). Some of the resources that were used for data derivation are given as follows: (Emiroglu et al., 2011a; Singh et al., 1994; Borghei et al., 1999; Bagheri et al., 2014; Subramanya and Awasthy, 1972).

Table 1. Some empirical formulae to calculate the side weir discharge coefficient

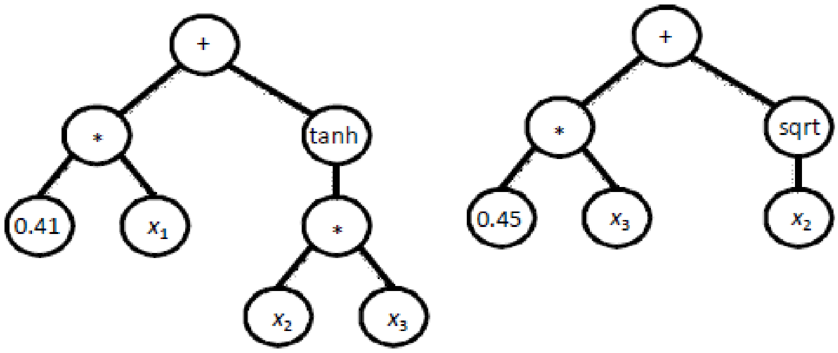
Row	Author(s)	Equation
1	Nandesamoorthy and Thomson (1972)	$C_d = 0.432\left(\frac{2-Fr_1^2}{1+2Fr_1^2}\right)^{0.5}$
2	Subramanya and Awasthy (1972)	$C_d = 0.864\left(\frac{1-Fr_1^2}{2+Fr_1^2}\right)^{0.5}$
3	Yu-Tech (1972)	$C_d = 0.623 - 0.222Fr_1$
4	Ranga Raju et al. (1979)	$C_d = 0.81 - 0.6Fr_1$
5	Hager (1987)	$C_d = 0.485\left(\frac{2-Fr_1^2}{2+3Fr_1^2}\right)^{0.5}$
6	Cheong (1991)	$C_d = 0.45 - 0.221Fr_1$
7	Singh et al. (1994)	$C_d = 0.33 - 0.18Fr_1 + 0.49\left(\frac{P}{h_1}\right)$
8	Jalili and Borghei (1996)	$C_d = 0.71 - 0.41Fr_1 + 0.22\left(\frac{P}{h_1}\right)$
9	Borghei et al. 1(999)	$C_d = 0.7 - 0.48Fr_1 + 0.3\left(\frac{P}{h_1}\right) + 0.06\left(\frac{L}{h_1}\right)$
10	Emiroglu et al. (2011)	

Table 2. Range of collected data related to the side weir discharge coefficient

Data range	Fr_1	P/h_1	L/b	L/h_1	Cd
Min.	0.09	0.03	0.21	0.19	0.09
Max.	0.84	2.28	3.00	10.71	1.75
Avg.	0.43	0.76	1.13	3.87	0.50
St. dev.	0.18	0.43	0.85	3.06	0.17

Genetic Programing Overview

Genetic programing (GP) technique is a machine learning approach which is used for modeling input-output complex non-linear systems that are based on the data set. Developing GP is based on the genetic algorithm (GA) concept. This means that the concepts which are used in GA are repeated in GP, such as genes, multi-genes, mutation… and so on. GP is also used to build an empirical formula from the input-output data set. It is also often known as symbolic regression. GP creates the formula that consists of input variables and several mathematical operators such as (+,−, / and *) and functions such as (ex, x, sin, cos, tan, lg, sqrt, ln, power). GP carries out this process by randomly generating a population of computer programs (represented by tree structures) and then mutating and crossing over the best performing trees to create a new population. This process is continued until the formula with most suitable accuracy is achieved. Unlike classical regression analysis by which the designer defines the structure of the empirical formula, GP automatically generates both the structure and the parameters of the empirical formula. An individual multi-gene is comprised of one or more genes and is named a GP tree. To improve the performance of fitness (e.g. to reduce a model's sum of squared errors on a data set), the genes are obtained incrementally. The final formula may be a weighted linear or non-linear combination of all genes. The optimal weights for the genes are automatically obtained by using the ordinary least squares to regress the genes against the output data. Figure (2) shows a pseudo-formula obtained by a GP technique. In this formula, y is the output, while the inputs are x_1, x_2 and x_3 (Brameier and Banzhaf, 2007).



$$y = d_0 + d_1(0.41x_1 + \tanh(x_2x_3) + d_2(0.45x_3 + \mathrm{sqrt}(x_2))$$

Figure (2): Sketch of formula generation by GP technique

Genetic Programing Development

Preparation of a formula based on the genetic programing technique as similar to other machine learning techniques, such as all neural network models, is based on the data set. This means that to present a suitable formula for modeling an event, the effective parameters on the event should be previously defined and measured. Modeling the side weir discharge coefficient based on the GP technique needs to define and measure the influential parameters. For this purpose, the dimensionless parameters which are obtained in the dimensional analysis section (Eq.2) are used. The coefficient values which appear in the gene process are defined by using least squares operation in Table (2) data set.

RESULTS AND DISCUSSION

The performance of empirical formulae was assessed by conducting a compression on the measured data the range of which is given in Table (2) and the empirical formula results. For this purpose, some error indices, such as correlation coefficient and root mean square error, were used to calculate the average error. The results of error indices' calculation are given in Table (3). As seen in Table (3), the Emiroglu et al. (2011) formula with a coefficient correlation of (0.64) and a root mean square error of (0.03) is the most accurate one among the empirical formulae. Figure (3) shows the results of the Emiroglu formula *versus* the measured data. Development of an equation based on the genetic programing technique is based on the researcher's experience; where the recommendations of researchers who conducted similar studies are useful. In addition, mathematical approaches, such as principal component analysis (PCA), also help investigators select the parameters which are more effective. Preparing a formula based on PCA analysis leads to develop an optimal formula which includes the most important parameters. In this paper, by giving attention to Table (1), Froude number repeats in all empirical formulae, which means that Froude number is one of the main influential parameters. The PCA technique was carried out on the collected data (Table 2) to derive the most important parameters which have influence on Cd_{sw} . The results of PCA are present in Table (4) and Figure (4). PCA results uphold the experiments of empirical researchers who considered Froude number and P/h_1 as important parameters.

Table 3.The performances of empirical formulas

Author(s)	R^2	RSME
Nandesamoorthy and Thomson (1972)	0.01	0.00
Subramanya and Awasthy (1972)	0.01	0.00
Yu-Tech (1972)	0.01	0.00
Ranga Raju et al. (1979)	0.01	0.00
Hager (1987)	0.01	0.01
Cheong (1991)	0.01	0.01
Singh et al. (1994)	0.07	0.01
Jalili and Borghei (1996)	0.06	0.01
Borghei et al. 1(999)	0.11	0.02
Emiroglu et al. (2011)	0.65	0.03

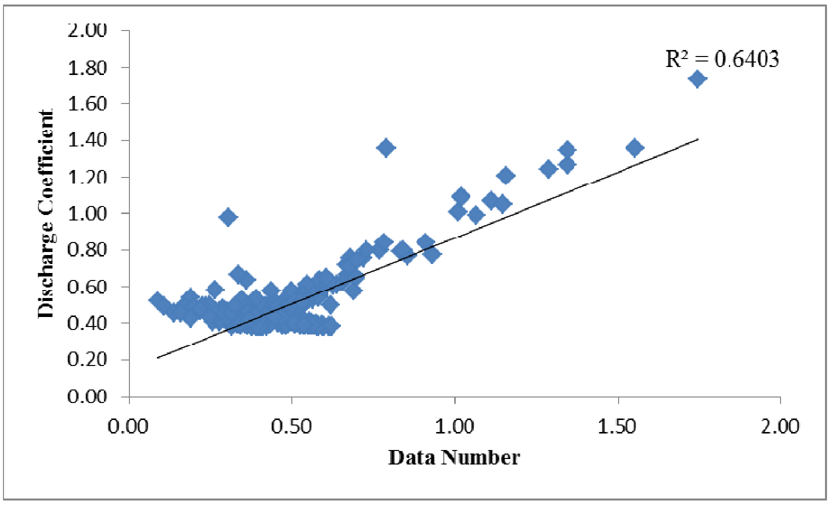


Figure (3): Results of the Emiroglu formula *versus* the measured data

Table 4. Results of principal component analysis

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	2.000	49.997	49.997	2.000	49.997	49.997	1.944
2	1.033	25.821	75.818	1.033	25.821	75.818	1.088
3	.692	17.289	93.107				
4	.276	6.893	100.000				

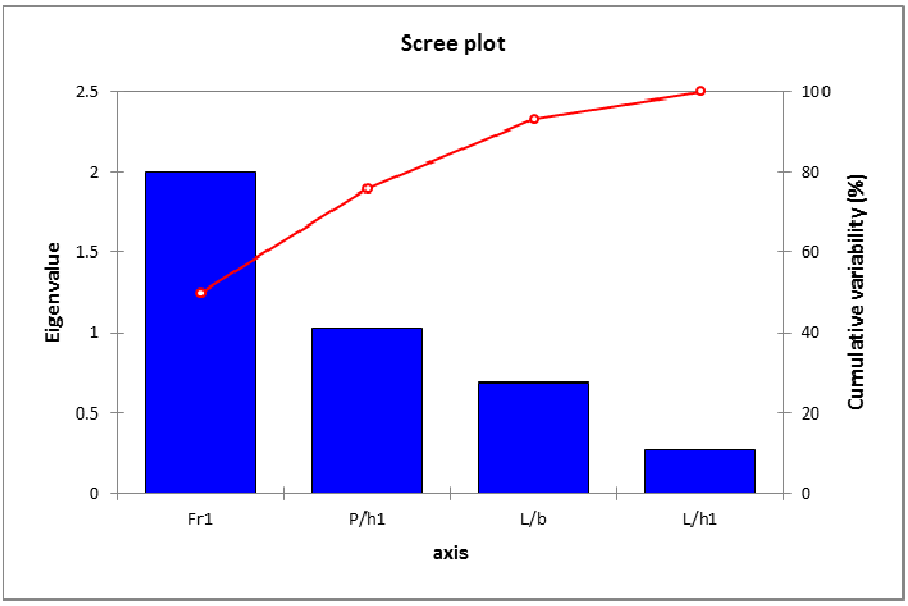


Figure (4): The influence of Cd_{sw} parameters

PCA results indicate that Froude number and ratio of P/h_1 are the most important parameters for the prediction of Cd_{sw}. As previously mentioned, developing the GP model, as similar to other intelligent models, is based on the data set. Therefore, all the collected data is divided into categories as training and testing categories. Training data sets included 70% (380 data sets) of all the data, while the rest (30% : 98 data sets) was considered as testing data sets. To assess the performance and illustrate the optimal equation based on genetic programing, several scenarios were considered and the performance of each scenario was assessed. Choosing a formula as an appropriate result of GP needs to consider accuracy and simplicity. Equation (3) was derived during the genetic programing development. As seen from Eq. (3), Froude number and P/h_1 are the most weighted parameters. The results of genetic programing uphold the PCA results. The performance of genetic programing during the training and testing stages is presented in Figures (5 and 6). As shown in Figures (5 and 6), the accuracy of the GP model is suitable with ($R^2 = 0.95$ and RMSE= 0.09). The main parameters which were set in the GP model development are given in Table (5).

$$C_d = 0.305 * Fr_1 \left(\frac{P}{h_1} + 3\frac{L}{b} - \frac{L}{h_1} - 10.06 \right) - 0.306 Fr_1^2 \left(\frac{P}{h_1} \right)^3 \frac{L}{b} \left(Fr_1 - \frac{L}{b} \right) + 0.594 \qquad (3)$$

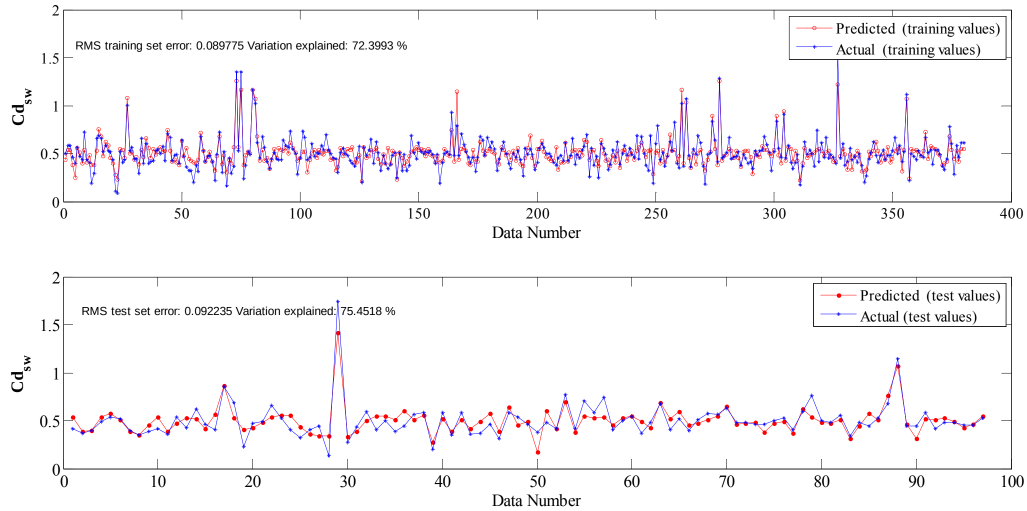


Figure (5): Performance of GP model during training and testing stages

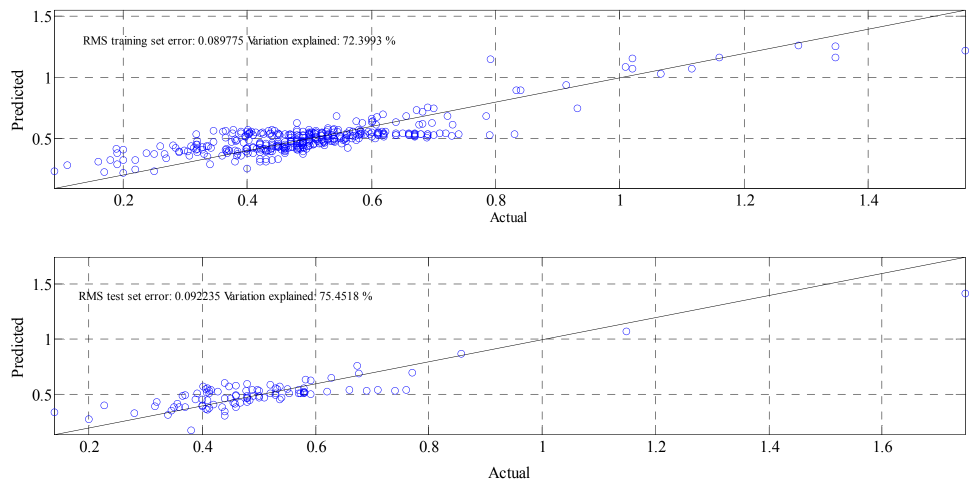


Figure (6): GP model results *versus* actual (measured) data during training and testing stages

Table 5. Setting of GP parameters during the development stage

Description of parameter	Setting of parameter
Function set	exp, cos, sin, minus, plus, divide, power
Population size	100
Maximum depth of trees	5
Generation number	100

CONCLUSION

In this paper, predicting the side weir discharge coefficient (Cd_{sw}) by an empirical formula and genetic programing (GP) technique was considered. Assessing the results of empirical formulae shows that the Emiroglu formula is the most accurate one among the empirical formulae. The principal component analysis (PCA) was implemented to derive the most important parameters which influence Cd_{sw}. The results of PCA indicated that the upstream Froude number and ratio of weir height to the total upstream head are the most important parameters. Modeling and predicting Cd_{sw} by GP was also considered. The performance of GP shows that the ability of GP for modeling Cd_{sw} is suitable, where the accuracy of the GP model was much higher than that of empirical formulae.

REFERENCES

Aydin, M.C. (2012). "CFD simulation of free-surface flow over triangular labyrinth side weir". Advances in Engineering Software, 45 (1), 159-166. doi:http://dx.doi.org/10.1016/j.advengsoft.2011.09.006.

Aydin, M.C., and Emiroglu, M.E. (2013). "Determination of capacity of labyrinth side weir by CFD". Flow Measurement and Instrumentation, 29 (0), 1-8. doi:http://dx.doi.org/10.1016/j.flowmeasinst.2012.09.008.

Bagheri, S., Kabiri-Samani, A.R., and Heidarpour, M. (2014). "Discharge coefficient of rectangular sharp-crested side weirs, Part I: Traditional weir equation". Flow Measurement and Instrumentation, 35 (0), 109-115. doi:http://dx.doi.org/10.1016/j.flowmeasinst.2013.11.005.

Bilhan, O., Emin Emiroglu, M., and Kisi, O. (2010). "Application of two different neural network techniques to lateral outflow over rectangular side weirs located on a straight channel". Advances in Engineering Software, 41 (6), 831-837. doi:http://dx.doi.org/10.1016/j.advengsoft.2010.03.001.

Bilhan, O., Emiroglu, M.E., and Kisi, O. (2011). "Use of artificial neural networks for prediction of discharge coefficient of triangular labyrinth side weir in curved channels". Advances in Engineering Software, 42 (4), 208-214. doi:http:// dx.doi.org/ 10.1016/ j. advengsoft. 2011.02.006.

Borghei, S., Jalili, M., and Ghodsian, M. (1999). "Discharge coefficient for sharp-crested side weir in subcritical flow". Journal of Hydraulic Engineering, 125 (10), 1051-1056. doi:10.1061/(ASCE)0733-9429(1999)125:10(1051).

Borghei, S.M., and Parvaneh, A. (2011). "Discharge characteristics of a modified oblique side weir in sub-critical flow". Flow Measurement and Instrumentation, 22 (5), 370-376. doi:http://dx.doi.org/10.1016/j.flowmeasinst.2011.04.009.

Brameier, M.F., and Banzhaf, W. (2007). "Linear genetic programming". Springer, US.

Cheong, H. (1991). "Discharge coefficient of lateral diversion from trapezoidal channel". Journal of Irrigation and Drainage Engineering, 117 (4), 461-475. doi:10.1061/(ASCE)0733-9437(1991)117:4(461).

Coşar, A., and Agaccioglu, H. (2004). “Discharge coefficient of a triangular side-weir located on a curved channel". Journal of Irrigation and Drainage Engineering, 130 (5), 410-423. doi:10.1061/(ASCE)0733-9437(2004)130:5(410).

Durga Rao, K.H.V., and Pillai, C.R.S. (2008). "Study of flow over side weirs under super-critical conditions". Water Resources Management, 22 (1), 131-143. doi:10.1007/s11269-007-9153-4.

El-Khashab, A., and Smith, K.V. (1976). "Experimental investigation of flow over side weirs". Journal of the Hydraulics Division, 102 (9), 1255-1268.

Emiroglu, M., and Kaya, N. (2011). "Discharge coefficient for trapezoidal labyrinth side weir in sub-critical flow". Water Resources Management, 25 (3), 1037-1058. doi:10.1007/s11269-010-9740-7.

Emiroglu, M., and Kisi, O. (2013). "Prediction of discharge coefficient for trapezoidal labyrinth side weir using a neuro-fuzzy approach". Water Resources Management, 27 (5), 1473-1488. doi:10.1007/s11269-012-0249-0.

Emiroglu, M.E., Agaccioglu, H., and Kaya, N. (2011a). "Discharging capacity of rectangular side weirs in straight open channels". Flow Measurement and Instrumentation, 22 (4), 319-330. doi:http://dx.doi.org/10.1016/j.flowmeasinst.2011.04.003.

Emiroglu, M.E., Bilhan, O., and Kisi, O. (2011b). "Neural networks for estimation of discharge capacity of triangular labyrinth side-weir located on a straight channel". Expert Systems with Applications, 38 (1), 867-874. doi:http://dx.doi.org/10.1016/j.eswa.2010.07.058.

Ettema, R. (2000). "Hydraulic modeling: concepts and practice". ASCE.

Ghodsian, M. (2003). "Super-critical flow over a rectangular side weir". Canadian Journal of Civil Engineering, 30 (3), 596-600. doi:10.1139/l03-004.

Haddadi, H., and Rahimpour, M. (2012). "A discharge coefficient for a trapezoidal broad-crested side weir in subcritical flow". Flow Measurement and Instrumentation, 26 (0), 63-67. doi:http://dx.doi.org/10.1016/j.flowmeasinst.2012.04.002.

Hager, W. (1987). "Lateral outflow over side weirs". Journal of Hydraulic Engineering, 113 (4), 491-504. doi:10.1061/(ASCE)0733-9429(1987)113:4(491).

Jalili, M., and Borghei, S. (1996). "Discussion: discharge coefficient of rectangular side weirs". Journal of Irrigation and Drainage Engineering, 122 (2), 132.

Kaya, N., Emiroglu, M.E., and Agaccioglu, H. (2011). "Discharge coefficient of a semi-elliptical side weir in sub-critical flow". Flow Measurement and Instrumentation, 22 (1), 25-32. doi:http://dx.doi.org/10.1016/j.flowmeasinst.2010.11.002.

Kisi, O., Emin Emiroglu, M., Bilhan, O., and Guven, A. (2012). "Prediction of lateral outflow over triangular labyrinth side weirs under sub-critical conditions using soft computing approaches". Expert Systems with Applications, 39 (3), 3454-3460. doi:http://dx.doi.org/10.1016/j.eswa.2011.09.035.

Parsaie, A., and Haghiabi, A. (2014). "Predicting the side weir discharge coefficient using the optimized neural network by genetic algorithm". Scientific Journal of Pure and Applied Sciences, 3 (3), 103-112.

Rahimpour, M., Keshavarz, Z., and Ahmadi, M. (2011). "Flow over trapezoidal side weir". Flow Measurement and Instrumentation 22 (6), 507-510. doi:http://dx.doi.org/10.1016/j.flowmeasinst.2011.08.004.

Singh, R., Manivannan, D., and Satyanarayana, T. (1994). "Discharge coefficient of rectangular side weirs". Journal of Irrigation and Drainage Engineering, 120 (4), 814-819. doi:10.1061/(ASCE)0733-9437(1994)120:4(814).

Subramanya, K., and Awasthy, S.C. (1972). "Spatially varied flow over side-weirs". Journal of the Hydraulics Division, 98 (1), 1-10.

Swamee, P., Pathak, S., and Ali, M. (1994a). "Side-weir analysis using elementary discharge coefficient". Journal of Irrigation and Drainage Engineering, 120 (4), 742-755. doi:10.1061/(ASCE)0733-9437(1994)120:4(742).

Swamee, P., Pathak, S., Mohan, M., Agrawal, S., and Ali, M. (1994b). "Sub-critical flow over rectangular side weir". Journal of Irrigation and Drainage Engineering, 120 (1), 212-217. doi:10.1061/(ASCE)0733-9437(1994)120:1(212).

Swamee, P.K., Pathak, S.K., Mohan, M., Agrawal, S.K., and Ali, M.S. (1994c). "Sub-critical flow over rectangular side weir". Journal of Irrigation and Drainage Engineering, 120 (1), 212-217

Uyumaz, A., and Muslu, Y. (1985). "Flow over side weirs in circular channels". Journal of Hydraulic Engineering, 111 (1), 144-160. doi:10.1061/(ASCE)0733-9429(1985)111:1(144).

Uyumaz, A., and Smith, R. (1991). "Design procedure for flow over side weirs". Journal of Irrigation and Drainage Engineering, 117 (1), 79-90. doi:10.1061/(ASCE)0733-9437(1991)117:1(79).

Vatankhah, A. (2012). "New solution method for water surface profile along a side weir in a circular channel". Journal of Irrigation and Drainage Engineering, 138 (10), 948-954. doi:10.1061/(ASCE)IR. 1943-4774. 0000483.

Vatankhah, A. (2013a). "Water surface profiles along a rectangular side weir in a u-shaped channel (analytical findings)". Journal of Hydrologic Engineering, 18 (5), 595-602. doi: 10.1061/ (ASCE) HE. 1943-5584. 0000657.

Vatankhah, A.R. (2013b). "Water surface profile along a side weir in a parabolic channel". Flow Measurement and Instrumentation, 32 (0), 90-95. doi:http://dx. doi.org/10.1016/j.flowmeasinst.2013.04.010.