GP approach for critical submergence of intakes in open channel flows

H. Md. Azamathulla and Z. Ahmad

ABSTRACT

This technical paper presents the genetic programming (GP) approach to predict the critical submergence for horizontal intakes in open channel flow for different bottom clearances. Laboratory data from the literature for the critical submergence for a wide range of flow conditions were used for the development and testing of the proposed method. Froude number, Reynolds number, Weber number and ratio of intake velocity and channel velocity were considered dominant parameters affecting the critical submergence. The proposed GP approach produced satisfactory results compared to the existing predictors.

Key words | critical submergence, genetic programming, intakes, open channel

H. Md. Azamathulla (corresponding author)
River Engineering and Urban Drainage Research Centre (REDAC), Universiti Sains Malaysia,
Engineering Campus, Seri Ampangan,
14300 Nibong Tebal,
Pulau Pinang,
Malaysia
E-mail: *redacazamath@eng.usm.my*; *mdazmath@gmail.com*

Z. Ahmad
Department of Civil Engineering,
Indian Institute of Technology Roorkee,
Roorkee-247667, (Uttarakhand),
India

NOTATIONS

b	Width of the channel
c	Bottom clearance
d_i	Diameter of intake
D	Depth of flow in the channel
$\mathbf{F}$	Froude number
g	Acceleration due to gravity
Q_i	Intake discharge
$\mathbf{R}$	Reynolds number
S_C	Critical submergence
U_∞	Velocity in the channel
U_i	Velocity in intake
$\mathbf{W}$	Weber number
Γ	Circulation
ρ	Mass density
σ	Surface tension
μ	Dynamic viscosity
δ	Average absolute deviation
R	Coefficient of correlation
RMSE	Root mean square error

INTRODUCTION

Intakes are economical, easier to operate, and draw less sediment when they are located near the water surface. However, if the water depth above an intake is shallow, strong vortices are formed which may lead to air entrainment. Vortices have been observed frequently at many installations such as the Hirfanli Dam in Turkey, the Harspranget Dam in Sweden, the Kariba Dam in Zimbabwe and so on (Ahmad *et al.* 2008). Such vortices not only cause appreciable loss in the efficiency of the hydraulic machinery and corrosion in the water conducting system, but also produce vibrations and noise. Air-entraining is more severe in tropical climates where the water demand is high and the reservoir level is low. Thus, a sufficient cover of water is required at the intake to avoid the formation of these vortices (Ahmad *et al.* 2008). The critical submergence for intakes has been investigated extensively in the laboratory and field (Ahmad *et al.* 2008). Denny (1956), who investigated critical submergence and air-entraining vortices in pump sumps, found that an entry of 1% air volume into an intake caused a 15% reduction in the pump efficiency. Also, several studies yielded empirical expressions to estimate the critical submergence for the different types of intake (Reddy & Pickford 1972; Swaroop 1973; Odgaard 1986; Kocabas and Yildrim 2002; Durai *et al.* 2007). Generally, the existing predictors consider Froude number, Reynolds number, Weber number,

doi: 10.2166/hydro.2012.089

circulation, geometry of intake, flow patterns near intake, etc. as the significant parameters. Yildirim and Kocabas (1995, 1998, 2002) and Yildirim *et al.* (2000) computed the critical submergence for intakes in open channel flow and still water reservoirs using potential theory and dimensional analysis. Furthermore, the critical submergence for a rectangular intake was studied by Yildirim (2004). Recently, Ahmad *et al.* (2008) proposed the following predictors for the critical submergence S_C for the horizontal intakes for different invert levels of the intake in open channel flow.

$$\frac{S_C}{d_i} = 0.36\mathbf{F}^{0.80}\left(\frac{U_\infty}{\sqrt{gd_i}}\right)^{-0.90} \text{for } c = 0 \tag{1}$$

$$\frac{S_C}{d_i} = 0.27\mathbf{F}^{0.039}\left(\frac{U_i}{U_\infty}\right)^{1.02} \text{for } c = d_i/2 \tag{2}$$

where d_i = diameter of intake, U_i = velocity in intake, g = acceleration due to gravity, U_∞ = velocity in the open channel, $\mathbf{F}$ = Froude number = $U_i/\sqrt{gd_i}$, and c = bottom clearance.

Due to the complexity of vortex formation near the intake, the empirical formulae obtained through the regression approach are not as accurate for the estimation of critical submergence. In the past few years, soft computing and data mining techniques such as genetic programming (GP), adaptive neuro fuzzy techniques and artificial neural networks (ANN) have widely been used in many engineering problems. GP and ANN models were used to estimate bridge pier scour (Azamathulla *et al.* 2010), scour below spillways (Azamathulla *et al.* 2008), sediment transport (Singh *et al.* 2007), evapo-transpiration (Guven *et al.* 2008) and the critical submergence of a vertical intake (Kocabas *et al.* 2008). Ayoubloo *et al.* (2011) used classification and regression tree (CART) and linear regression analysis for the prediction of critical submergence for horizontal intakes in open channel flow with different clearance bottoms. The above applications indicate that GP and ANNs are powerful tools for solving complex hydraulic problems. This study aims to investigate the performance of the GP approach in the prediction of critical submergence for horizontal intakes in open channel flow and to compare the results with those obtained from well known empirical formulae.

GENETIC PROGRAMMING

GP follows the process of natural evolution in which the species survive as per the principle of 'survival of the fittest'. It is similar to the widely known genetic algorithm (GA), but unlike GA its solution is a computer program or an equation instead of the set of numbers. In GP, a random population of individuals, that is equations or computer programs, is created, the fitness of individuals is evaluated and then the 'parents' are selected out of these individuals (Koza 1990, 1992). The principle of Darwinian natural selection is used to select and reproduce 'fitter' programs. GP creates computer programs of equal or unequal length, that consist of variables (terminal), and several sets of mathematical operators (function) as the solution. The functional set of the system can be composed of arithmetic operations (+, − , /, *) and function calls (such as {e^x, x, sin, cos, tan, log, *sqrt*, ln, *power*}). Each function implicitly includes an assignment to a variable, which facilitates the use of multiple program outputs in GP, whereas in tree-based GP those side effects need to be incorporated explicitly (Brameier & Banzhaf 2001).

The present GP utilizes a two-point string crossover. A segment of random position and random length is selected in both parents and is exchanged between them. If one of the resulting offspring exceeds the maximum length, the crossover is abandoned and is restarted by exchanging equalized segments (Brameier & Banzhaf 2001). An operand or an operator of an instruction is changed by mutation into another symbol over the same set.

The fitness of a GP individual f may be computed by using the equation

$$f = \sum_{j=1}^{N} \left(\left|X_j - Y_j\right|\right) \tag{3}$$

where X_j is the value returned by a chromosome for the fitness case j, Y_j is the expected value for the fitness case j, and N is the number of fitness cases.

The best individual (program) of a trained GP can be converted into a functional representation by successive replacements of the variables starting with the last effective instruction (Oltean & Groşan 2003).

GP has been extensively used in the field of hydraulic engineering to model complex problems (Babovic & Keijzer 2000; Keijzer & Babovic 2002). Davidson *et al.* (1999) determined empirical relationships for the friction in turbulent pipe flow and the additional resistance to the flow induced by flexible vegetation. Giustolisi (2004) determined the Chezy resistance coefficient in corrugated channels, and Kizhisseri *et al.* (2005) proposed a correlation between the temporal pattern of flow field and sediment transport by utilizing numerical model results and field data. Guven & Gunal (2008a, b) predicted local scour downstream of grade-control hydraulic structures. Guven (2009) applied linear GP technique to predict the flow rate. Recently, Sharifi *et al.* (2011) used GP to predict end-depth relationship in open channel flow.

Description of collected data and dimensional analysis

Laboratory and field data have been collected to develop the predictive models to estimate the critical submergence for a horizontal intake and to evaluate the performance of the existing models. Ahmad *et al.* (2008) experimentally studied the critical submergence for horizontal intakes in open channel flow (Figure 1). Experiments were performed in a concrete flume with 10 m length, 0.37 m width and 0.6 m depth. The intake was located in the lateral direction at a distance of 5 m from the upstream of the flume, horizontally. The diameters of the intake pipes, d_i were 4.25, 6.25 and 10.16 mm, and bottom clearance $c = 0$ and $c = d_i/2$. The experiments were performed for each intake pipe and for different intake discharges Q_i, velocity in the flume, velocity in the intake pipe, U_i, and the corresponding critical submergence S_C were measured. A total 324 data for three intakes pipes for $c = 0$ (162 data sets) and $c = d_i/2$ (162 data sets) were collected by Ahmad *et al.* (2008) and the same were used in the present study. The ranges of various parameters included in the present study are summarized in Table 1.

The factors affecting critical submergence are d_i, U_i, U_∞, c, width of the channel b, circulation Γ, mass density ρ, dynamic viscosity μ, surface tension σ, and acceleration due to gravity g. The functional relationship for the critical submergence S_C can be written as

$$S_C = f(d_i, b, U_i, U_\infty, \Gamma, c, \rho, \mu, \sigma, g) \tag{4}$$

These dimensional parameters are frequently limited in the laboratory ranges. Therefore, to cover the wide range of input data for the applicability of the models in the field, non-dimensional parameters are defined below using ρ, d_i, and V_i as repeating variables

$$\frac{S_C}{d_i} = f_1\left(\frac{c}{d_i}, \frac{b}{d_i}, \frac{U_\infty}{U_i}, \mathbf{F} = \frac{U_i}{\sqrt{gd_i}}, \mathbf{R} = \frac{\rho U_i d_i}{\mu}, \mathbf{W} = \frac{\rho U_i^2 d_i}{\sigma}, \frac{\Gamma}{U_i d_i}\right) \tag{5}$$

where $\mathbf{R}$ = intake Reynolds number and $\mathbf{W}$ = Weber number.

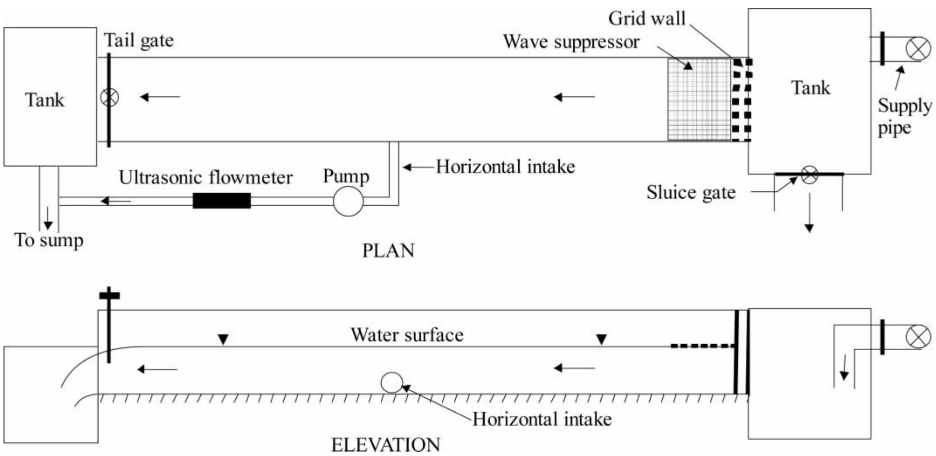

Figure 1 | Experimental setup (Ahmad *et al.* 2008).

Table 1 | Range of data used in the present study (Ahmad *et al.* 2008)

Parameters	Range
Intake pipe diameter, d_i	42.5, 6.25, 10.16 cm
Velocity of flow in intake, U_i	0.37 to 4.30 m/s
Velocity of flow in flume, U_∞	0.09 to 0.56 m/s
Bottom clearance, c	0 to 5.08 cm
Critical submergence, S_C	0.86 to 25.13 cm
Intake discharge, Q_i	0.0028 to 0.0064 m^3/s
Depth of flow, D	7.65 to 33.45 cm
Reynolds number, **R**	37600 to 183000
Froude number, **F**	0.37 to 6.66
Weber number, **W**	190 to 10800

Table 2 | Parameters of the optimized GP model

Parameter	Description of parameter	Setting of parameter
p_1	Function set	$+, -, *, /, \sqrt{}$, power
p_2	Population size	250
p_3	Mutation frequency %	96
p_4	Crossover frequency %	50
p_5	Number of replication	10
p_6	Block mutation rate %	30
p_7	Instruction mutation rate %	30
p_8	Instruction data mutation rate %	40
p_9	Homologous crossover %	4
p_{10}	Program size	Initial 64, maximum 256

Past studies revealed that the normalized critical submergence could be presented as the following non-dimensional parameters (Yildrim & Kocabas 1995, 1998, 2002; Yildrim *et al.* 2000; Ahmad *et al.* 2008)

$$\frac{S_C}{d_i} = f\left(\mathbf{F}, \mathbf{R}, \mathbf{W}, \frac{U_i}{U_\infty}\right) \tag{6}$$

The functional relationship for S_C/d_i has been developed using the available data, and by invoking the GP.

GP modeling for critical submergence

The scenarios considered in building the GP model include inputs (**F**, **R**,**W**, U_i/U_∞) and output (S_C/d_i). From the collected data sets used in this study, around 70% (114 data sets) were used for training (chosen randomly until the best training performance with low root mean square error (RMSE) was obtained), while the remaining 30% (48 data sets) were used for testing or validating the GP models developed individually for $c = 0$ and $d_i/2$.

In this study, four basic arithmetic operators (+, –, *, /) and three mathematical functions ($\sqrt{}$, x^2, power) were utilized. The number of generations used to obtain the optimal solution was 5,000. First, the maximum size of each program was specified as 256, starting with 64 instructions for the initial program. The functional set and operational parameters used in the present GP modeling are listed in Table 2.

The simplified analytic form of the proposed GP models for $c = 0$ and $c = d_i/2$ may be expressed, respectively, as

$$\frac{S_C}{d_i} = \left(\mathbf{W} + e^{-3.80-\mathbf{W}} - (\mathbf{F}+\mathbf{R})^{1/3}\right)^2 + \left(\left(\frac{\mathbf{F}}{U_i/U_\infty}\right)^{0.5} - U_i/U_\infty\right) \times (\ln(U_i/U_\infty) + 9.99\mathbf{R}) + \left(\mathbf{R} + 2.145(U_i/U_\infty)^{1/3} + (\mathbf{F}/\mathbf{R})^{1/3}\right) \tag{7}$$

$$\frac{S_C}{d_i} = \left(\mathbf{W} + e^{-2.60-\mathbf{W}} - (\mathbf{F}+\mathbf{R})^{1/3}\right)^2 + \left(\left(\frac{\mathbf{F}}{U_i/U_\infty}\right)^{0.5} - U_i/U_\infty\right) \times (\ln(U_i/U_\infty) + 7.45\mathbf{R}) + \left(\mathbf{R} + 1.675(U_i/U_\infty)^{1/3} + (\mathbf{F}/\mathbf{R})^{1/3}\right) \tag{8}$$

Training and testing of the proposed GP model

The performance of the proposed GP model is validated in terms of the common statistical evaluation parameters, i.e. R (coefficient of correlation), RMSE and δ (average absolute deviation) which are expressed as

$$R = \left(\frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}\right) \tag{9}$$

$$\mathrm{RMSE} = \left(\frac{\sum (X-Y)^2}{n}\right)^{1/2} \tag{10}$$

$$\delta = \frac{\sum |Y - X|}{\sum X} \times 100 \quad (11)$$

where $x = (X - \overline{X})$, $y = (Y - \overline{Y})$, $X =$ observed values, $\overline{X} =$ mean of X, Y=predicted value, $\overline{Y} =$ mean of Y, and $n =$ number of samples.

First, an attempt was made to assess the significance or influence of each input parameter on S_C/d_i, i.e. sensitivity analysis. Table 3 compares the performance of the various GP models, with one of the independent parameters removed in each case. It may be noted that deletion of any independent parameter from the input set, i.e. Equation (6) yields larger RMSE and lower R values. Therefore, it is concluded that the four independent parameters, **F, R, W** and U_i/U_∞ have significant influence on S_C/d_i – thus the first model of Table 3 is recommended. The GP models, i.e. Equations (7) and (8), resulted in highly nonlinear relationship between S_C/d_i and input parameters with high accuracy and relatively low error.

RESULTS AND DISCUSSION

The proposed GP models were tested with the unused data (about 30% of the total data), and the results were compared with the predicted value of critical submergence using Equations (1) and (2) proposed by Ahmad *et al.* (2008). For a quantitative comparison of the observed and predicted S_C/d_i, the statistical parameters such as R, RMSE and δ were computed and compared. The analysis of the data reveals that S_C/d_i increases with increase in **F**, U_i/U_∞, **R** and **W** for both $c = 0$ and $d_i/2$. However, the effects of **F** and U_i/U_∞ on S_C/d_i dominate over the other parameters.

The values of S_C/d_i obtained by the predictors proposed by Swaroop (1973) and Reddy and Pickford (1972) have also been checked with the data collected in the present study, and found to be much greater than the observed ones. Figures 2 and 3 show the results of the proposed GP and the predictors of Ahmad *et al.* (2008), in terms of scatter plots between predicted and observed S_C/d_i for $c = 0$ and $c = d_i/2$, respectively. The statistical parameters computed for the unused data for $c = 0$ and $c = d_i/2$ are given in Table 4. Both Figures 2 and 3, and Table 4 depict that the prediction of critical submergence by GP is better than the equations proposed by Ahmad *et al.* (2008) even though GP overestimates the value of S_C/d_i for a low value of S_C/d_i. The equations of Ahmad

Table 3 | Sensitivity analysis for independent parameters for the testing set for $c = 0$

Model	R	RMSE	δ
$S_c/d_i = f((\mathbf{F}, \mathbf{R}, \mathbf{W}, U_i/U_\infty)$	0.99	0.035	3.7
$S_c/d_i = f((\mathbf{F}, \mathbf{R}, \mathbf{W})$	0.94	11.65	18.45
$S_c/d_i = f((\mathbf{F}, \mathbf{R}, U_i/U_\infty)$	0.92	17.74	21.36
$S_c/d_i = f((\mathbf{F}, \mathbf{W}, U_i/U_\infty)$	0.88	23.472	26.71

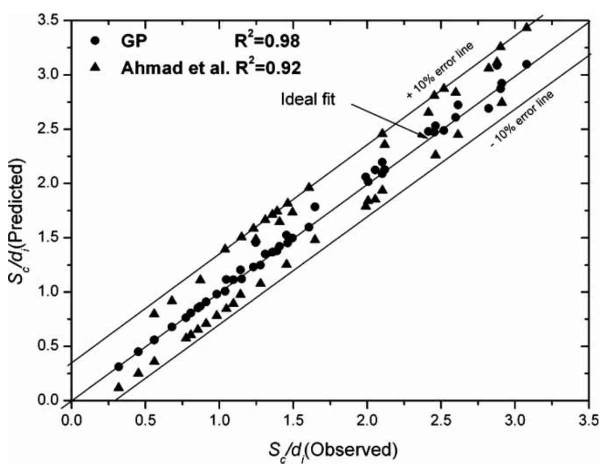

Figure 2 | Comparison of predicted S_C/d_i by different methods with observed one for $c = 0$ (testing).

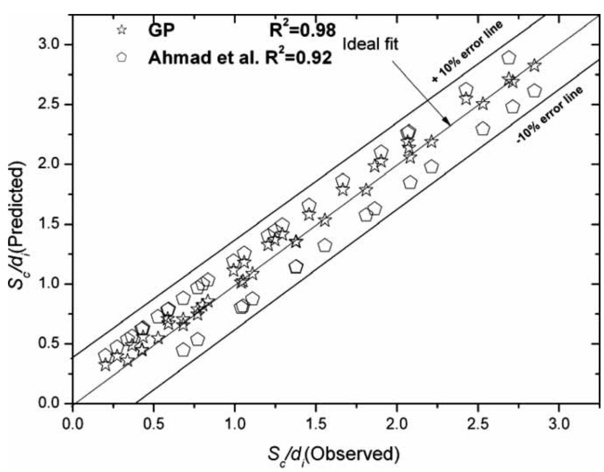

Figure 3 | Comparison of predicted S_C/d_i by different methods with observed one for $c = d_i/2$ (testing).

Table 4 | Statistical evaluation parameters for various models to predict S_C/d_i for the training and testing data

	for $c=0$			for $c=d_i/2$		
Model	**R**	**RMSE**	δ	**R**	**RMSE**	δ
(a) Training data						
Ahmad *et al.* (2008)	0.97	0.045	7.34	0.97	0.111	11.56
Proposed GP	0.99	0.024	1.84	0.99	0.015	5.75
(b) Testing data						
Ahmad *et al.* (2008)	0.96	0.067	8.41	0.96	0.144	13.85
Proposed GP	0.99	0.035	3.70	0.99	0.036	9.59

et al. (2008) are considered to be the best empirical equation which produces good correlation ($R=0.96$) for both $c=0$ and $d_i/2$. However they show higher RMSE and δ (8.41 and 13.85, respectively) relative to the present models for both the bottom clearances. The proposed GP models revealed a high generalization capacity with $R=0.99$ and RMSE=0.035, $\delta=3.7$ for $c=0$ and $R=0.99$ and RMSE = 0.036, $\delta=9.59$ for $c=d_i/2$. Figures 2 and 3 reveal that the proposed predictors produce satisfactory prediction of S_C/d_i with a discrepancy of +10% and for $c=0$ and $c=d_i/2$, compared to the observed ones.

CONCLUSION

The analysis of data revealed that S_C/d_i increased with increase in **F**, U_i/U_∞, **R** and **W** for both $c=0$ and $d_i/2$. However, the effects of **F** and U_i/U_∞ on S_C/d_i dominated over the other parameters. The trained GP model for the prediction of critical submergence in open channel flow for different bottom clearances was tested using the unused data and the results were compared with the observed ones. Critical submergence was also calculated using the existing equations and the results were compared with those observed. The prediction of critical submergence by GP was better than that by the existing equations even though GP overestimated the value of S_C/d_i for a low value of S_C/d_i. GP produced the highest correlation coefficient, low RMSE and average absolute deviations for both bottom clearances. The present GP predictions were in satisfactory agreement with the observed ones.

REFERENCES

Ahmad, Z., Rao, K. V. & Mittal, M. K. 2008 Critical submergence for horizontal intakes in open channel flows. *Dam Engineering* **XIX** (2), 70–90.

Ayoubloo, M. K., Azamathulla, H. Md., Jabbari, E. & Zanganeh, M. 2011 Predictive model-based for the critical submergence of horizontal intakes in open channel flows with different clearance bottoms using CART, ANN and linear regression approaches. *Expert Systems with Applications* **38**, 10114–10123.

Azamathulla, H. Md., Deo, M. C. & Deolalikar, P. B. 2008 Alternative neural networks to estimate the scour below spillways. *Advances in Engineering Software* **39**, 689–698.

Azamathulla, H. Md., Ghani, A. A., Zakaria, N. A. & Aytac, G. 2010 Genetic programming to predict bridge pier scour. *ASCE Journal of Hydraulic Engineering* **136**, 165–169.

Babovic, V. & Keijzer, M. 2000 Genetic programming as a model induction engine. *Journal of Hydroinformatics* **2**, 35–60.

Brameier, M. & Banzhaf, W. 2001 A comparison of linear genetic programming and neural networks in medical data mining. *IEEE Transactions on Evolutionary Computation* **5**, 17–26.

Davidson, J. W., Savic, D. A. & Walters, G. A. 1999 Method for identification of explicit polynomial formulae for the friction in turbulent pipe flow. *Journal of Hydroinformatics* **1**, 115–126.

Denny, D. F. 1956 An experiment study of air-entraining vortices at pump sumps. *Proceedings of the Institution of Mechanical Engineers* **170**, 106–116.

Durai, E. S. R., Ahmad, Z. & Mittal, M. K. 2007 Critical submergence at vertical pipe intakes. *Dam Engineering* **XVIII** (1), 17–33.

Giustolisi, O. 2004 Using genetic programming to determine Chèzy resistance coefficient in corrugated channels. *Journal of Hydroinformatics* **6**, 157–173.

Guven, A. 2009 Linear genetic programming for time-series modelling of daily flow rate. *Journal of Earth System Science* **118**, 137–146.

Guven, A. & Gunal, M. 2008a Prediction of scour downstream of grade-control structures using neural networks. *Journal of Hydraulic Engineering* **134**, 1656–1660.

Guven, A. & Gunal, M. 2008b Genetic programming approach for prediction of local scour downstream of hydraulic structures. *Journal of Irrigation and Drainage Engineering* **134**, 241–249.

Guven, A., Aytek, A., Yuce, M. I. & Aksoy, H. 2008 Genetic programming-based empirical model for daily reference evapotranspiration model. *CLEAN, Soil Air Water* **36**, 905–912.

Keijzer, M. & Babovic, V. 2002 Declarative and preferential bias in GP-based scientific discovery. *Genetic Programming and Evolvable Machines* **1**, 41–79.

Kizhisseri, A. S., Simmonds, D., Rafiq, Y. & Borthwick, M. 2005 An evolutionary computation approach to sediment transport modeling. In *Fifth International Conference on Coastal Dynamics*, April 4–8, Barcelona, Spain.

Kocabas, F. & Yildirim, N. 2002 Effect of circulation on critical submergence for an intake. *Journal of Hydraulic Research, IAHR* **40**, 741–752.

Kocabas, F., Unal, S. & Unal, B. 2008 A neural network approach for prediction of critical submergence of an intake in still water and open channel flow for permeable and impermeable bottom. *Computers and Fluids* **37**, 1040–1046.

Koza, J. R. 1990 Genetic Programming: A Paradigm for Genetically Breeding Populations of Computer Programs to Solve Problems. Technical Report, STAN-CS-90–1314, Dept. of Computer Science, Stanford University, CA, USA.

Koza, J. R. 1992 *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, USA.

Odgaard, A. J. 1986 Free surface air core vortex. *Journal of Hydraulic Engineering, ASCE* **112**, 610–620.

Oltean, M. & Groşan, C. 2003 A comparison of several linear genetic programming techniques. *Complex Systems* **14**, 1–29.

Reddy, Y. R. & Pickford, J. A. 1972 Vortices at intakes in conventional sump. *Water Power* **24**, 108–109.

Sharifi, S., Sterling, M. & Knight, D.W. 2011 Prediction of end-depth ratio in open channels using genetic programming. *Journal of Hydroinformatics* **13**, 36–48.

Singh, A. K., Deo, M. C. & Sanil Kumar, V. 2007 Neural network – genetic programming for sediment transport. *Journal of Maritime Engineering, The Institution of Civil Engineers, London* **160** (MA3), 113–119.

Swaroop, R. 1973 Vortex formation at intakes. M.E. Dissertation, Civil Engineering Department, University of Roorkee (now IIT Roorkee), Roorkee, India.

Yildirim, N. 2004 Critical submergence for a rectangular intake. *Journal of Engineering Mechanics, ASCE* **130**, 1195–1210.

Yildirim, N. & Kocabas, F. 1995 Critical submergence for intakes in open channel flow. *Journal of Hydraulic Engineering, ASCE* **121**, 900–905.

Yildirim, N. & Kocabas, F. 1998 Critical submergence for intakes in still water reservoir. *Journal of Hydraulic Engineering, ASCE* **124**, 103–104.

Yildirim, N. & Kocabas, F. 2002 Prediction of critical submergence for an intake pipe. *Journal of Hydraulic Research, IAHR* **124**, 507–518.

Yildirim, N., Kocabas, F. & Gulcan, S. C. 2000 Flow boundary effects on critical submergence of an intake pipe. *Journal of Hydraulic Engineering, ASCE* **126**, 288–297.

First received 3 July 2011; accepted in revised form 22 January 2012. Available online 9 April 2012