Prediction of end-depth ratio in open channels using genetic programming

S. Sharifi, M. Sterling and D. W. Knight

ABSTRACT

In this paper, genetic programming (GP) is used as an effective model induction tool to solve a classic problem in open channel flow: the free overfall. By applying GP to experimental data of circular channels with a flat bed and employing a model selection procedure, a reliable expression in the form of $h_c = Ah_e e^{B\sqrt{S_0}}$ is found for calculating the critical depth (h_c) and end-depth ratio (EDR). Further effort is made to verify the applicability and superiority of this expression for channels with other cross sections. This global expression not only outperforms other expressions in estimating the critical depth, it is also dimensionally correct (unlike some other applications of GP) and can be used for channels with any cross-section and any flow regime.

Key words | data modelling, evolutionary algorithms, free overfall, genetic programming, open channel flow

S. Sharifi (corresponding author)
M. Sterling
D. W. Knight
Department of Civil Engineering,
The University of Birmingham,
Birmingham B15 2TT,
UK
Tel.: +44 (0)121-414 5152
E-mail: *sxs650@bham.ac.uk*

NOTATIONS

A	Area (m^2)
b	Channel width (m)
d	Diameter (m)
g	Gravitational acceleration (ms^{-2})
h	Flow depth (m)
l	Characteristic length parameter (m)
m	Channel side slope
n	Manning friction
Q	Discharge (m^3s^{-1})
S_0	Channel bed slope
S_c	Critical slope
T	Top width of flow (m)

SUBSCRIPTS

e	End section
c	Critical depth section
n	Normalized

doi: 10.2166/hydro.2010.087

ACRONYMS

CoD	coefficient of determination
EDR	End-depth ratio (h_e/h_c)
GP	Genetic Programming
MRSS	mean root of sum of squared residuals
RMSE	root mean square of errors

INTRODUCTION

A free overfall is a situation where the bottom of a channel drops suddenly, causing the flow to separate and form a free nappe. The depth of water at the section where the overfall occurs is known as the end depth (h_e) or brink depth (see Figure 1). Aside from its close relation to the broad-crested weir, the free overfall can form the starting point in computations of the surface profile in gradually varied flows and hence has a distinct importance in hydraulic engineering (Chaudhry 1993). In addition, based on various experiments on prismatic channels, the end depth bears a unique

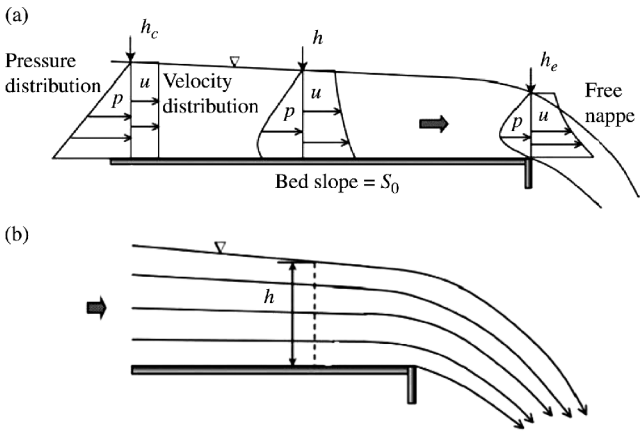

Figure 1 | (a) Schematic view of a typical free overfall and the hydraulic aspects; (b) Streamline pattern of a free overfall (Dey 2002*b*).

relationship with critical depth (h_c). As there exists a unique stage-discharge relationship at the critical depth, this relation enables the free overfall to be used as a simple flow measuring device (Gupta *et al.* 1993; Sterling & Knight 2001).

Van Leer (1922, 1924 cited in USBR 2001) was perhaps the first who used the free overfall principle to measure flow in pipes flowing partially full. Ledoux (1924) and Rouse (1936) also realized that the end depth of flow in a rectangular channel could be used as a simple flow measuring device that requires no calibration. Since then, because of its importance and also relatively simple laboratory setup, a large number of theoretical and experimental studies have been carried out to understand the hydraulics of the end-depth problem and to determine the end-depth ratio (EDR $= h_e/h_c$) in a wide range of channels.

The Boussinesq approach (Jaeger 1957; Dey 2002*a*,*b*), Energy approach (Anderson 1967; Hager 1983), Momentum approach (Delleur *et al.* 1956; Diskin 1961; Rajaratnam & Muralidhar 1964*a*,*b*, 1970; Keller & Fong 1989; Murty Bhallamudi 1994; Dey 1998, 2001*b*, 2003; Dey & Ravi Kumar 2002), Weir approach (Rouse 1936; Ferro 1999; Dey 2001*a*,*c*, 2002*b*), Free Vortex approach (Ali & Sykes 1972), Potential flow approach (Southwell & Vaisey 1946; Marchi 1993) and empirical methods (Gupta *et al.* 1993; Pagliara 1995; Davis *et al.* 1998; Sterling & Knight 2001) are among the various different classical approaches which have been well practised in dealing with the free overfall phenomena in different channels. A comprehensive review on these classical studies is available in Dey (2002*b*).

Even though the above approaches yield a number of promising solutions various inadequacies mainly relating to the assumed distributions of velocity and/or pressure, have led researchers to apply numerical techniques to solve this problem. Knowing that a complete solution of the free overfall requires an integration of the turbulent Reynolds equations, Finnie & Jeppson (1991), Mohapatra *et al.* (2001), Guo (2005), Guo *et al.* (2006) and Ramamurthy *et al.* (2006) applied turbulence modelling to find a complete solution of the free overfall in rectangular and trapezoidal channels. Furthermore, various machine learning techniques have been applied to tackle this problem. For instance, Raikar *et al.* (2004) used a four-layer Artificial Neural Network (ANN) model to analyze the experimental data and to determine the EDR for a smooth inverted semicircular channel in all flow regimes. Ozturk (2005) used the same technique and investigated the EDR in rectangular channels with different roughnesses. Most recently, Pal & Goel (2006, 2007) applied a support vector machine-based modelling technique to determine the end-depth ratio and discharge of a free overfall occurring over inverted smooth semi-circular channels, circular channels with flat base and also trapezoidal channels with different bed slopes.

This study extends the earlier work, and shows the application of another evolutionary machine learning algorithm called genetic programming (GP) (Koza 1992; Babovic & Abbott 1997*a*,*b*) as an effective model induction tool that is able to develop a conceptual transparent model of the physical process of the free overfall. However, unlike the previous work, this outlines an expression which appears to have a form independent of cross-sectional shape. The paper is organized as follows: firstly, a brief background is given on the hydraulic aspects of the free overfall. Then the basics of GP and its elements are explained. In the methodology section, after introducing the experimental data, the process of applying GP to the data and also the procedure of selecting the "best" model is discussed. Finally, the entire procedure is repeated for rectangular and trapezoidal channels and the performance of the "best" fit model is measured for other cross-sections.

BACKGROUND

Free overfall

Figure 1(a) shows a schematic view of the pressure and velocity distributions along a channel with a free overfall. At the brink section, the pressure above and below the falling nappe is atmospheric and therefore the pressure distribution at this section differs from the hydrostatic pressure distribution and has a mean pressure considerably less than the hydrostatic.

Figure 1(b) shows the variation of streamline curvature, being finite at the free surface and zero at the channel bed. The strong vertical component of acceleration due to the gravity affects the curvature of the free nappe in the vicinity of the brink section. As the free surface profile should be continuous, this effect is extended a short distance upstream of the brink section, causing an acceleration of the flow. This guarantees that the depth of flow at the brink section is less than critical depth. As a result, at sections upstream from the brink, the water surface curvature gradually decreases until a control section is reached where the vertical component of acceleration is weak and the pressure is hydrostatic (Sterling & Knight 2001; Dey 2002*b*).

In channels with a mild slope, the flow upstream of the brink is subcritical, becoming supercritical just before the brink section. Therefore at a short distance upstream of the brink; there is a section where the pressure distribution is hydrostatic, the specific energy attains a minimum value and the depth of flow is critical. When the slope is steep and the approaching flow is supercritical, a critical section does not exist upstream the end section (Sterling & Knight 2001; Dey 2002*b*). Furthermore, in supercritical conditions, every single disturbance creates cross-waves leading to difficulties in determining the depth of flow which makes accurate physical measurements difficult.

Overview of genetic programming

Genetic Programming (GP) (Koza 1990; Koza 1992) is a collection of evolutionary computation techniques based on the principle of Darwin's theory of evolution, that allow computers to solve problems automatically by evolving computer programs (Poli *et al.* 2008). GP was first introduced by Koza (1990) as a powerful tool for tackling problems in various fields of artificial intelligence. GP follows a similar procedure to genetic algorithm (GA) and is summarized below.

The standard GP starts with an initial population of randomly generated symbolic expressions (also known as parse trees) composed of functions and terminals appropriate to the problem domain. These functions may be standard arithmetic operations (such as addition, subtraction, multiplication, and division), standard mathematical functions (such as SIN, EXP, etc.) standard programming operations (such as If-Then-Else, Do-Until, etc.), Boolean functions (such as, AND, OR, XOR, NOT, etc.) and various domain-specific functions. A fitness function is then used to measure the performance of each individual symbolic expression in the particular problem environment. Predictably, the majority of the initial random symbolic expressions have exceedingly poor fitnesses but nonetheless, some individuals are more fit than others. Then, a sexual genetic reproduction process is preformed on pairs of expressions, which are selected in proportion to their fitness, and offsprings are created. The resulting offsprings are composed of sub-expressions from their parents and form the new generation which replaces the old population of parents. The fitness function is again used to measure the fitness of each individual in the new population and the process is repeated. Repeating this algorithm will gradually produce populations which, over a period of generations, reach a high average fitness in dealing with their environment. Figure 2 shows the flowchart of the computational procedure of GP. For an in-depth explanation of GP and its elements the reader is referred to Koza *et al.* (1992), Babovic & Keijzer (2000), Koza (2003) and Poli *et al.* (2008).

In the context of model induction, a distinct advantage of GP over similar techniques such as artificial neural networks is that it can provide answers in the form of a symbolic expression (Keijzer & Babovic 1999; Babovic & Keijzer 2000). This mathematical representation often provides a benefit in modelling and has resulted in the successful application of GP to a wide range of practical problems over the last two decades (Poli *et al.* 2008).

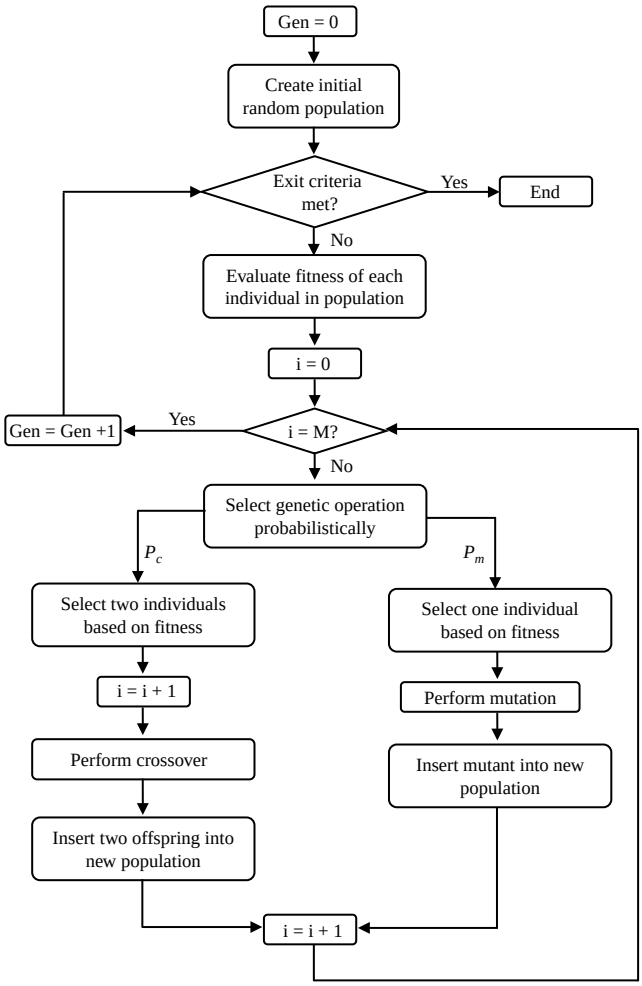

Figure 2 | Computational procedure of Genetic Programming (Koza 1992).

PROCEDURE

Data set used

The main data set used in this research was the laboratory data sets from the study by Sterling (1998). The experiments were undertaken in a 21.26 m long tilting channel with a working cross section of 610 mm wide by 365 mm deep supported on a hydraulic jack that enabled change in the bed slope (S_0). The experimental channel consisted of eight, 2 m long plastic PVC pipe sections with an internal diameter (D) of 244 mm with a wall thickness of 3 mm. A 110 mm wide slot was cut in the crown of the pipe sections to provide access in the pipe and make the measurements possible. A flat horizontal bed constructed from 9 mm thick PVC was added to the base of the circular pipe. Five series of experiments were carried out with five different bed thicknesses (t). For each test the brink depth at the centerline was measured by means of a pointer gauge to an accuracy of ± 0.1 mm and the discharge (Q) was measured via a calibrated orifice plate. The critical depth (h_c) was determined from $Q^2/g = A^3/T$ and the critical slope (S_c) was calculated using the average value of Manning's n (see Sterling 1998 for further details). Figure 3 shows the geometry of the channels and Table 1 shows a summary of the experimental data.

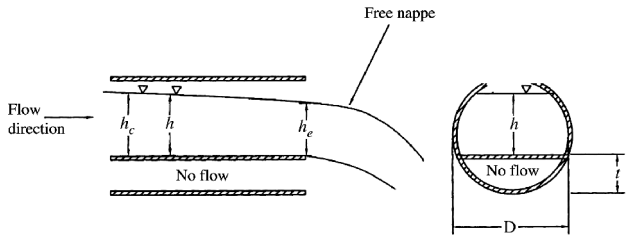

Figure 3 | Geometry of circular channels with flat bed (Pal & Goel 2006).

Data processing

The entire data set consisted of a sub total of 322 individual experiments within the 5 test series. Prior to the application of the GP, the data did not go under any integration, cleaning or discretization process. The only procedure carried out on the data was splitting the total data set into three disjoint subsets: training (66%), testing (23%) and validation data (11%), by means of uniform random sampling. The training data would be used as inputs for the GP modelling process, the testing data for model selection and the validation data for model validation.

Table 1 | Range of Sterling's (1998) experimental data

Series	Bed thickness (m)	Brink depth range (m)	Slope	Discharge range ($l\,s^{-1}$)
1	0.000	0.0197–0.1426	0.0	1.5–61.6
2	0.061	0.0217–0.0927	0.1%	4.8–38.8
3	0.081	0.0080–0.0689	0.4%	1.0–26.8
4	0.123	0.0074–0.0657	0.9%	1.1–21.7
5	0.162	0.0091–0.0310	1.6%	1.0–7.7

Table 2 | Parameters used in the genetic programming application

Parameter	Value
Function set	Plus, minus, times, power, exp, mydivide, mylog, mysqrt
Terminal set	$h_b, t, D, S_0, S_c, \sqrt{S_0}, \sqrt{S_c}$, random number
Population size	75
Tree Initialization method	Ramped-half-and-half
Tree size restriction	10 nodes
Genetic operators	Subtree Cross-over and Mutation
Operator probabilities	Variable (minimum equal to 0.20)
Fitness function	Sum of squared distance
Selection method	Lexical tournament
Number of generations	100

Model training

In this research, the genetic programming lab (GPLAB 2009) v.3 toolbox for Matlab (available from http://gplab.sourceforge.net) was used with slight modifications to evolve a relationship between the critical depth, and the end depth and other characteristics of the channel. A sensitivity analysis was first performed in order to obtain a robust algorithm parameter set (Table 2). Then, 50 independent runs of the GP algorithm were performed on the training data to limit the effect of randomness on the results.

An initial evaluation of the "best" expressions of each run revealed that $\sqrt{S_0}$ is repeated in a large number of expressions and hence is one of the principal factors. In order to increase the efficiency of the algorithm, the square root of the bed slope and critical slope were added to the terminal set and the algorithm was run another 50 times.

Model selection

In the GP process, reaching optimum coefficient values for a symbolic expression requires the initial population to evolve through many generations and as such increases

Table 3 | Selected expressions and the value of MRSS, RMSE and CoD for training and test data

		Training data			Test data		
No.	Expression	MRSS ($\times 10^{-4}$)	RMSE	CoD	MRSS ($\times 10^{-4}$)	RMSE	CoD
1	$h_c = h_e(e^{0.2471+S_0+1.5338\sqrt{S_0}})$	3.494	0.0749	0.9883	5.406	0.0726	0.9960
2	$h_c = h_e(e^{0.24+S_c+1.6706\sqrt{S_0}})$	3.564	0.0758	0.9878	5.508	0.0737	0.9958
3	$h_c = h_e e^{(0.1468+\sqrt{S_0})/0.6008}$	3.566	0.0758	0.9878	5.511	0.0735	0.9958
4	$h_c = 1.2769 h_e e^{1.6643\sqrt{S_0}}$	3.566	0.0758	0.9878	5.511	0.0735	0.9958
5	$h_c = h_e^{(1-\sqrt{S_c})}\sqrt{S_c}^{-\sqrt{S_0}}$	3.613	0.0992	0.9875	5.608	0.1003	0.9955
6	$h_c = h_e(e^{D+\sqrt{S_0}} + \sqrt{S_0})$	3.634	0.0767	0.9873	5.588	0.0743	0.9956
7	$h_c = h_e/(e^{(h_b)^2} - \sqrt{S_0} - 0.23602)$	3.670	0.0754	0.9873	5.940	0.0745	0.9953
8	$h_c = e^D h_e e^{2\sqrt{S_0}}$	3.706	0.0803	0.9890	5.293	0.0765	0.9963
9	$h_c = h_e/(e^{h_e - \sqrt{S_0}} - h_e - 0.23332)$	3.711	0.0758	0.9867	5.835	0.0748	0.9953
10	$h_c = h_e/(0.7680 - \sqrt{S_0})$	3.714	0.0754	0.9867	5.829	0.0742	0.9953
11	$h_c = h_e(e^{D+S_c+2\sqrt{S_0}})$	3.816	0.0815	0.9890	5.414	0.0779	0.9963
12	$h_c = e^D h_e D^{-\sqrt{S_0}}$	4.045	0.0784	0.9860	6.543	0.0771	0.9951
13	$h_c = h_e e^{\sqrt{\sqrt{S_0}+\sqrt{S_c}}}$	4.048	0.0802	0.9843	6.364	0.0803	0.9945
14	$h_c = 1.3039 h_e e^{(\sqrt{S_0}^{0.8782})}$	4.114	0.0790	0.9837	6.491	0.0782	0.9942
15	$h_c = (h_e/\sqrt{S_c}^{\sqrt{S_0}})^{0.94582}$	4.164	0.0895	0.9878	6.339	0.0847	0.9958
16	$h_c = h_e e^{(\sqrt{S_0}+0.28897)}$	4.374	0.0800	0.9816	6.985	0.0803	0.9934
17	$h_c = 1.2544 h_e e^{(\sqrt{S_0}+\sqrt{S_c})}$	4.405	0.0813	0.9813	7.084	0.0827	0.9932
18	$h_c = e^D h_e e^{(\sqrt{S_0}+\sqrt{S_c})}$	4.418	0.0831	0.9813	6.889	0.0849	0.9932
19	$h_c = e^D h_e(\sqrt{S_c} + e^{\sqrt{S_0}})$	4.462	0.0827	0.9806	7.042	0.0843	0.9929
20	$h_c = 1.37936 h_e S_c^{-S_0}$	4.501	0.0794	0.9806	7.275	0.0809	0.9929

Table 4 | Normalized values of RMSE and CoD for the remaining expressions

No.	Expression	Normalized Total RMSE	CoD	$CoD_n - RMSE_n$
1	$h_c = h_e(e^{0.2471+S_0+1.5338\sqrt{S_0}})$	0.0000	0.6785	0.6785
3	$h_c = h_e e^{(0.1468+\sqrt{S_0})/0.6008}$	0.1959	0.4794	0.2835
4	$h_c = 1.2769 h_e e^{1.6643\sqrt{S_0}}$	0.1959	0.4794	0.2834
2	$h_c = h_e(e^{0.24+S_c+1.6706\sqrt{S_0}})$	0.2210	0.4881	0.2671
8	$h_c = e^D h_e e^{2\sqrt{S_0}}$	1.0000	1.0000	0.0000
6	$h_c = h_e(e^{D+\sqrt{S_0}} + \sqrt{S_0})$	0.3820	0.2776	−0.1044
10	$h_c = h_e/(0.7680 - \sqrt{S_0})$	0.2294	0.0000	−0.2294

the computation time and memory requirements. To overcome this problem and to save computation time, a least square calculation was first performed on the training data to find the optimum coefficients for each of the expressions.

Evaluating what is meant by the concept of "best" can be both subjective and controversial. In the present context, the "best" model was evaluated by analyzing the complexity of the model and its goodness of fit. To find this "best" model among the set of selected expressions, a combination of a subjective and an objective selection methodology was applied. In the subjective selection, the best 100 generated models were examined and the top 20 expressions were selected based on two criteria:

1. *Complexity level*. The number and composition of functions and terminals along with the dimensions of the models were investigated to select simpler and dimensionally more correct models.
2. *Performance level*. The mean root of sum of squared errors (MRSS) values were compared to select models with higher performance.

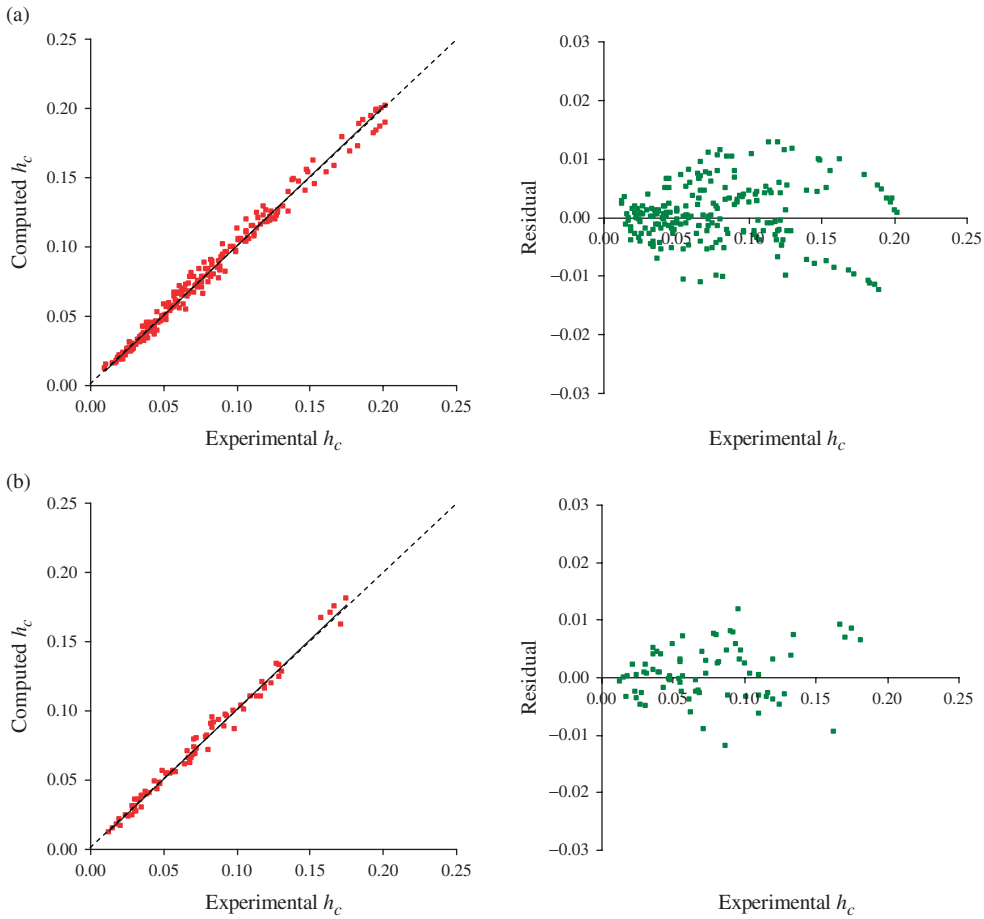

Figure 4 | Performance of $h_c = h_e(e^{0.2471+S_0+1.5338\sqrt{S_0}})$ on (a) training and (b) test data.

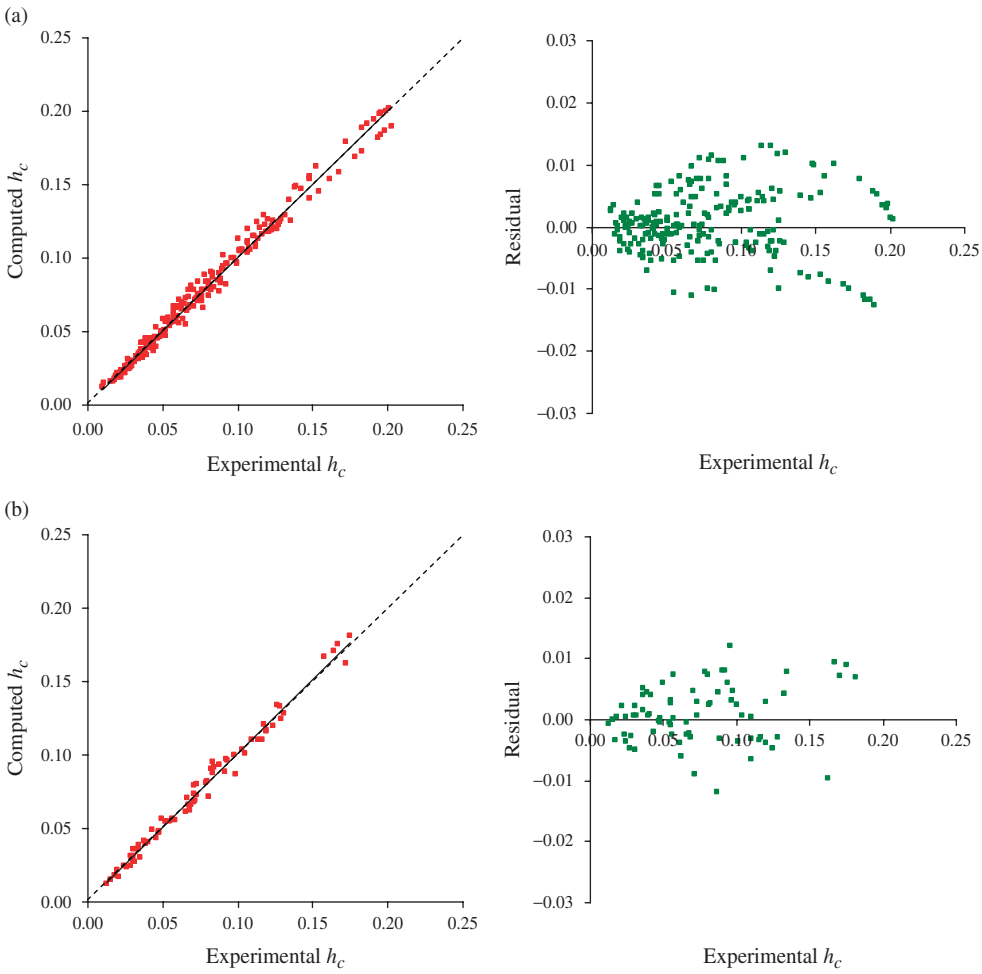

Figure 5 | Performance of $h_c = 1.2769 h_e e^{1.6643\sqrt{S_0}}$ on (a) training and (b) test data.

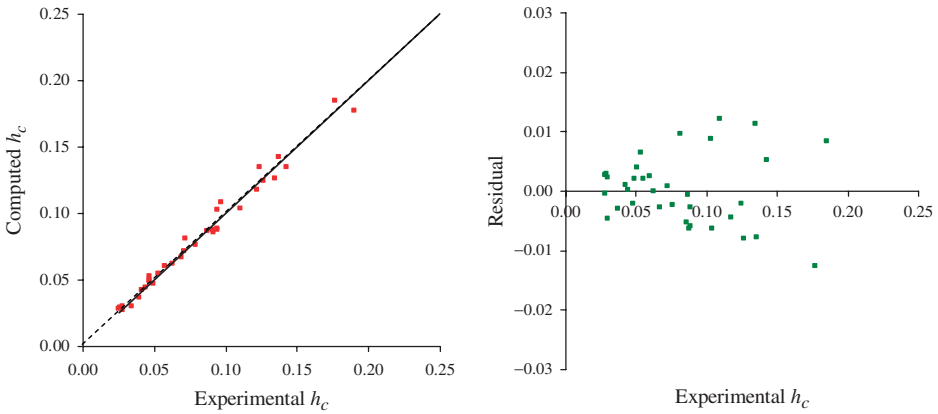

Figure 6 | Performance of $h_c = 1.2769 h_e e^{1.6643\sqrt{S_0}}$ on the validation data set.

Table 5 | Summary of additional data sets

Cross-section	Researcher(s)	Slope range S_0 (m/m)	Length parameter (m)	End depth range h_e (m)	Discharge range Q (ls^{-1})
Rectangular	Rajaratnam & Muralidhar (1968*a*,*b*)	−0.0101 ~ 0.0288	*b*: 0.46	0.0363–0.0957	31.43–67.32
	Rajaratnam & Muralidhar (1976)	0 ~ 0.0136	*b*: 0.46	0.0183–0.1751	12.84–108.09
	Davis *et al.* (1998)	0.003 ~ 0.02	*b*: 0.295	0.005–0.0365	0.44–14.01
	Ferro (1999)	0	*b*: 0.05 ~ 0.3	0.0167–0.0784	2.10–30.50
	Turan (2002)	0.0017 ~ 0.04	*b*: 1.00	0.0101–0.0581	12.33–77.97
	Fırat (2004)	0.0003 ~ 0.0394	*b*: 1.00	0.0038–0.0545	1.61–84.12
	Kutlu (2005)	0.00063 ~ 0.0387	*b*: 1.00	0.0046–0.0503	1.75–61.36
Trapezoidal	Diskin (1961)	0.0015	*b*: 0.125 (*m*: 1.5)	0.0399–0.1052	8.722–50.206
			b: 0.167 (*m*: 2.0)		
	Rajaratnam & Muralidhar (1970)	0.000 ~ 0.0673	*b*: 0.0127 ~ 0.1016 (*m*: 0.17 ~ 1.00)	0.0076–0.1676	0.878–44.288
	Keller & Fong (1989)	0.00067	*b*: 0.150 (*m*: 1.00)	0.0350–0.0779	6.130–23.190
	Yuen (1989)	0.000 ~ 0.02743	*b*: 0.150 (*m*: 1.00)	0.0144–0.0440	1.715–22.200
			b: 0.450 (*m*: 1.00)	0.0098–0.0425	
	Pagliara & Viti (1995)	0.000 ~ 0.0210	*b*: 0.280 (*m*: 1.00)	0.0100–0.1150	1.740–80.500
			b: 0.300 (*m*: 1.00)		
Circular	Smith (1962)	0.0000	*d*: 0.1532	0.0210–0.1125	1.048–18.916
	Rajaratnam & Muralidhar (1964*b*)	0.0000 ~ 0.0553	*d*: 0.2032	0.0192–0.1040	1.019–28.147
Inverted semi-circular	Dey *et al.* (2004)	0.0000 ~ 0.0270	*d*: 0.043 ~ 0.128	0.0029–0.0225	0.037–2.179
Δ-shaped	Dey & Ravi Kumar (2002)	0.0000	*b*: 0.12–0.18	0.0087–0.0740	0.787–17.824
U-shaped	Dey (2003)	0.0000	*d*: 0.07–0.13	0.0221–0.2287	0.729–0.73665
Triangular	Ahmad (2006)	0.0000 ~ 0.03333	*b*: 0.50	0.0370–0.1350	0.970–14.570
		−0.0200 ~ −0.0100	*b*: 0.50	0.073–0.1360	3.390–14.520

Table 6 | The best expressions found for rectangular and trapezoidal channels

			Training data			Test data		
Cross-section	**No.**	**Expression**	**MRSS (× 10^{-4})**	**RMSE**	**CoD**	**MRSS (× 10^{-4})**	**RMSE**	**CoD**
Rectangular	1	$h_c = 1.354h_e e^{2.048\sqrt{S_0}}$	2.927	0.0890	0.9738	5.203	0.0989	0.9803
	2	$h_c = 1.479h_e + h_e\sqrt{S_0}$	4.290	0.1230	0.9509	8.351	0.1573	0.9520
	3	$h_c = 1.474h_e + 0.150S_0$	4.403	0.1242	0.9443	8.506	0.1386	0.9465
	4	$h_c = 0.475h_e + h_e e^{\sqrt{S_0}}$	4.208	0.1207	0.9525	8.182	0.1545	0.9538
	5	$h_c = \sqrt{2h_e^2 + h_e S_0}$	3.769	0.1303	0.9580	7.164	0.1238	0.9637
Trapezoidal	1	$h_c = 1.354h_e e^{1.403\sqrt{S_0}}$	6.556	0.0813	0.9796	8.371	0.0785	0.9843
	2	$h_c = 1.431h_e + 0.280\sqrt{S_0}$	5.875	0.0956	0.9720	18.902	0.0810	0.9740
	3	$h_c = h_e + \sqrt{S_c}(\sqrt{h_e} + h_e \times S_0/S_e)$	5.661	0.0898	0.9776	9.352	0.0868	0.9714
	4	$h_c = 1.419h_e + 0.132S_0^{0.75}$	5.770	0.0916	0.9726	8.529	0.0781	0.9755
	5	$h_c = 1.384h_e + \sqrt{S_0.S_c}$	5.016	0.1030	0.9792	6.702	0.0803	0.9849

The issue of dimensional correctness is important and is the subject of significant debate. Indeed in some cases it is desirable to include measurement units in order to aid with the evolving of the equations (see Keijzer & Babovic (2002) for further details). This approach was not adopted below since it was discovered that a standard method (i.e. one which assumes dimensional correctness) worked sufficiently well for the purposes of the current paper. However, it is acknowledged that the approach outlined in Keijzer & Babovic (2002) may have results in a more efficient regression.

In order to gain a more complete picture of model performance, in the objective selection stage, two other fitness measures that are commonly reported in the literature, namely the root mean square of errors (RMSE) and coefficient of determination (CoD), were calculated for each expression on the training and testing data sets (Table 3). The RMSE describes the average difference between experimental data and model predictions, while CoD is a measure how much of the original uncertainty in the data is explained by the regression model (Weisberg 1980). It is suggested that a simultaneous assessment of these performance measures can provide a better insight on how thoroughly the model represents the system and hence a three step elimination strategy was followed to find the "best" model:

1. Firstly, the expressions were sorted on the RMSE of the testing data and the 10 worst were detected. The expressions were ranked then on the CoD of testing data and the 10 worst were found. Any expression placed in any worst set was eliminated.
2. Secondly, the total RMSE and CoD of the remaining expressions were calculated and normalized between 0.0 and 1.0. Then, the normalized RMSE was subtracted from the normalized CoD and the expressions were sorted on this value to find the fittest models (Table 4).
3. Finally, the computed values of h_c from the top 5 remaining expressions were plotted against the measured values for both training and test data and the deviation from the 45 degree line was inspected. The residual distribution of each expression was also plotted to investigate the degree of bias (see Figures 4 and 5). With the help of visual inspection and also judgment on the expression structure, a final elimination process was preformed to select the "best" expression. However, it is acknowledged that this approach is somewhat subjective.

The plots of the computed values of h_c against its measured values and the residual distributions (e.g. Figures 4 and 5) reveal that all top 5 expressions perform very closely on both sets of training and testing data and no one has any specific priority. Expressions 1, 3 and 4 express the critical depth as a function of end depth and bed slope and its only expression 2 that includes the critical slope. As measuring the critical slope accurately is a difficult and requires the knowledge of the critical depth, expression 2

was discarded. Furthermore, although having the best overall performance, expression 1 was also discarded because of its relative complexity. As a result, expressions 3 and 4 were found as the most suitable models for calculating the critical depth. A closer look to the expressions reveals that equations 3, 4 and 8 have the same structure and are only different in their coefficients.

Model validation

In order to validate the selected model, the fitness of the model was tested on a set of unseen data, namely the validation data. Figure 6 shows the performance of expression 4 on the validation data. The CoD for this expression is 0.996 and the RMSE is 0.0685. This shows that the model is a reliable model and can be used for other cases.

RESULTS AND DISCUSSION

The application of GP to the experimental data of circular channels with a flat bed resulted in finding an equation in the form of $h_c = Ah_e e^{B\sqrt{S_0}}$. According to this expression, the critical depth is only a function of the end depth and square root of the bed slope. Now an important question arises: can this equation be used for channels with other cross-sections? In order to investigate the answer to this question and verify the superiority of the found equation for channels with other cross-sections, a broad search was carried out on the literature and all the available and useful data sets were extracted (Table 5).

Rectangular and trapezoidal channels

The same modelling procedure was repeated for rectangular and trapezoidal channels where sufficient data for

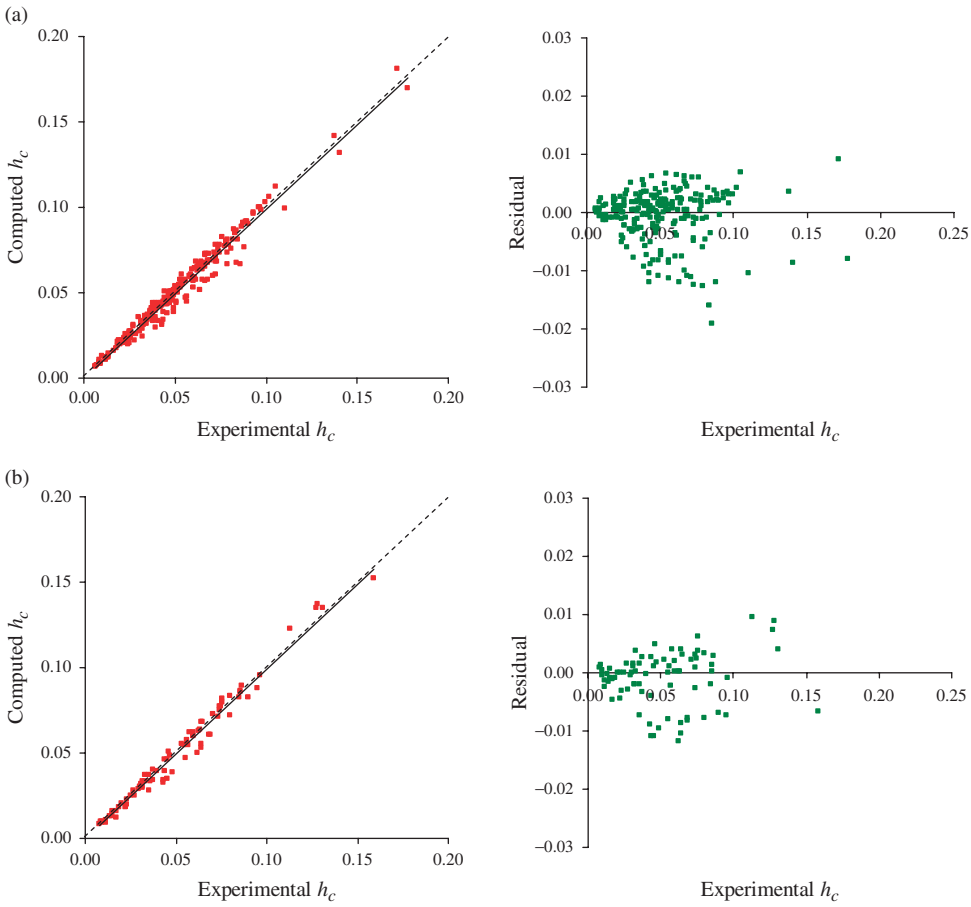

Figure 7 | Performance of $h_c = 1.354h_e e^{2.048\sqrt{S_0}}$ for rectangular channels. (a) training and (b) test data.

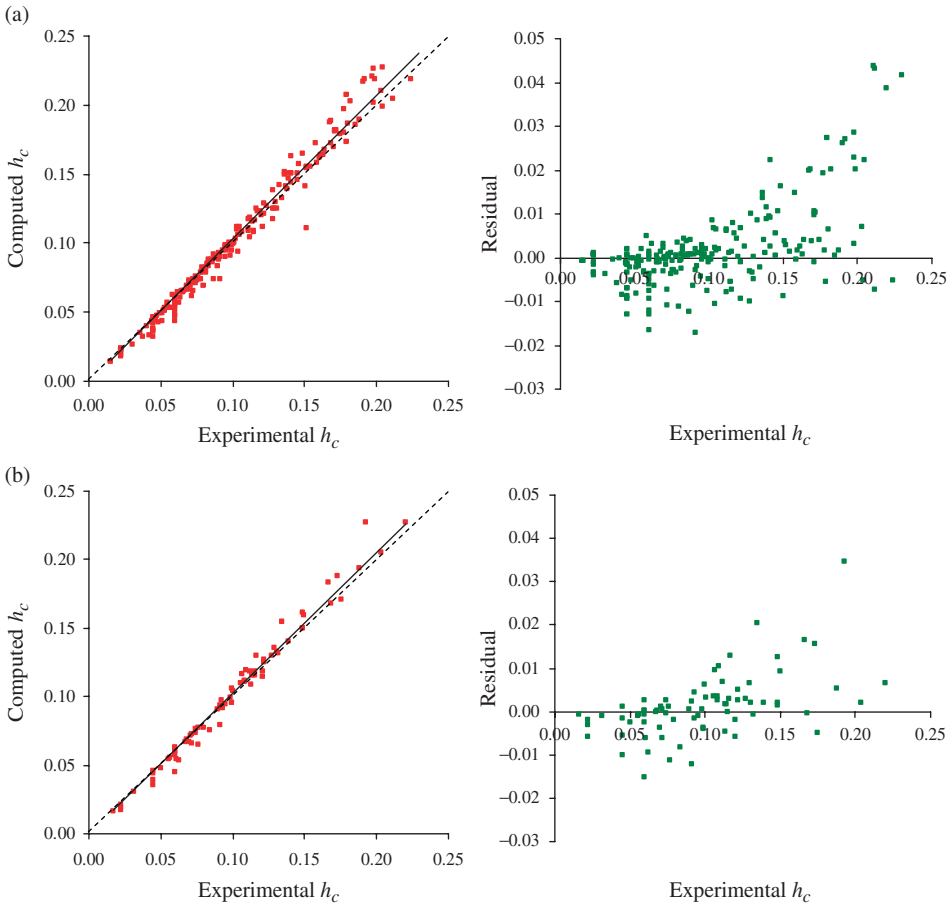

Figure 8 | Performance of $h_c = 1.354 h_e e^{1.403\sqrt{S_0}}$ for trapezoidal channels. (a) training and (b) test data.

modelling were available. Table 6 shows the five best expressions found after applying the GP algorithm and selection procedure. It is observed that again, an expression with the form of $h_c = Ah_e e^{B\sqrt{S_0}}$ is found as the "best" expression. Figures 7 and 8 illustrate the performance of the found equation for both channels.

Channels with other cross-sections

For the remaining cross-sections, first, the A and B coefficients within $h_c = Ah_e e^{B\sqrt{S_0}}$ were found by means of minimizing the sum of squared differences between the measured and calculated values of the critical depth. Then, the expression was tested on the available data and its

Table 7 | Performance of $h_c = Ah_e e^{B\sqrt{S_0}}$ on other cross-sections

Researchers	Cross-section	A	B	CoD	RMSE
Smith (1962)	Circular	1.2929	2.4584	0.9839	0.0491
Rajaratnam & Muralidhar (1964*a*,*b*)	Circular	1.3722	1.7063	0.9928	0.0057
Dey *et al.* (2004)	Inverted semi-circular	1.4129	1.2909	0.9839	0.0449
Dey & Ravi Kumar (2002)	Δ-shaped	1.4289	2.4584	0.9983	0.0160
Dey (2003)	U-shaped	1.3793	2.4584	0.9975	0.0318
Ahmad (2006)	Triangular	1.1878 1.1883	1.1517 0.3501	0.9819 0.9984	0.0412 0.0083

performance was measured. Table 7 illustrates the value of different goodness of fit criteria for each cross-section type. The CoD and RMSE values indicate that this model fits the data very well and show its applicability to any kind of cross-section.

CONCLUSIONS

In the present study, genetic programming has been used as a powerful model induction tool for solving a classic problem in open channel flow. Applying this tool and following a two-stage model selection procedure, a global transparent model in the form of $h_c = Ah_e e^{B\sqrt{S_0}}$ (or EDR $= h_e/h_c = (1/A)e^{-B\sqrt{S_0}}$) has been found as the most suitable expression for predicting critical depth and EDR in a wide range of channels. There are some interesting facts about this expression:

1. The expression appears to be universal and as such can be applied to channels with different geometries and flow regimes (subcritical and supercritical).
2. The expression is dimensionally correct.
3. Its overall performance is better than any other proposed empirical relationship.

Further work is required to find the probable physics behind this equation and also interpret its coefficients. Examining other channels, with different cross-sections and types of boundary roughness, should give an insight into how to select the appropriate values for A and B.

REFERENCES

Ahmad, Z. 2006 Flow measurements with triangular free overfall. *J. Inst. Eng.* **87**, 35–40.

Ali, K. H. M. & Sykes, A. 1972 Free-vortex theory applied to free overfall. *J. Hydraul. Div. ASCE* **98** (5), 973–979.

Anderson, M. V. 1967 Non-uniform flow in front of a free overfall. *Acta Polytech. Scand.* **42**, 1–24.

Babovic, V. & Abbott, M. B. 1997*a* The evolution of equation from hydraulic data, Part I: theory. *J. Hydraul. Res.* **35** (3), 397–410.

Babovic, V. & Abbott, M. B. 1997*b* The evolution of equation from hydraulic data, Part II: applications. *J. Hydraul. Res.* **35** (3), 411–430.

Babovic, V. & Keijzer, M. 2000 Genetic programming as a model induction engine. *J. Hydroinform.* **2** (1), 35–60.

Chaudhry, M. H. 1993 *Open-Channel Flow*. Prentice-Hall, Inc., Upper Saddle River, NJ.

Davis, A. C., Ellet, B. G. S. & Jacob, R. P. 1998 Flow measurement in sloping channels with rectangular free overfall. *J. Hydraul. Eng., ASCE* **124** (7), 760–763.

Delleur, J. W., Dooge, J. C. I. & Gent, K. W. 1956 Influence of slope and roughness on the free overfall. *J. Hydraul. Div., ASCE* **82** (4), 30–35.

Dey, S. 1998 End depth in circular channels. *J. Hydraul. Eng., ASCE* **124** (8), 856–863.

Dey, S. 2001*a* EDR in circular channels. *J. Irrig. Drain. Eng., ASCE* **127** (2), 110–112.

Dey, S. 2001*b* Flow metering by end-depth method in elliptic channels. *Dam Eng.* **12** (1), 5–19.

Dey, S. 2001*c* Flow measurement by the end-depth method in inverted semicircular channels. *Flow Meas. Instrum.* **12** (4), 253–258.

Dey, S. 2002*a* Free overfall in circular channels with flat base: a method of open channel flow measurement. *Flow Meas. Instrum.* **13**, 209–221.

Dey, S. 2002*b* Free overfall in open channels: state-of-the-art review. *Flow Meas. Instrum.* **13**, 247–264.

Dey, S. 2003 Overfall in U-shaped channels. *J. Eng. Mech.* **129** (3), 358–362.

Dey, S. & Ravi Kumar, B. 2002 Hydraulics of free overfall in Δ-shaped channels Sadhana. *Proc. Indian Acad. Sci.* **27** (June), 353–363.

Dey, S., Kumar, D. N. & Singh, D. R. 2004 End-depth in inverted semicircular channels: experimental and theoretical studies. *Nord. Hydrol.* **35** (1), 73–79.

Diskin, M. H. 1961 The end depth at a drop in trapezoidal channels. *J. Hydraul. Div., ASCE* **87** (4), 11–32.

Ferro, V. 1999 Theoretical end-depth-discharge relationship for free overfall. *J. Irrig. Drain. Eng., ASCE* **125** (1), 40–44.

Finnie, J. I. & Jeppson, R. W. 1991 Solving turbulent flows using finite elements. *J. Hydraul. Eng., ACSE* **117** (11), 1513–1530.

Fırat, C. E. 2004 *Effect of Roughness on Flow Measurements in Sloping Rectangular Free Overfall*. Ms Thesis, Middle East Technical University, Turkey.

GPLAB 2009 A Genetic Programming toolbox for MATLAB. http://gplab.sourceforge.net [Accessed April 20th 2009].

Guo, Y. 2005 Numerical modeling of free overfall. *J. Hydraul. Eng., ASCE* **131** (22), 134–138.

Guo, Y., Zhang, L. & Zhang, J. 2006 Numerical simulation of free overfall in a rough channel. In: *Proceedings of European Conference on Computational Fluid Dynamics*, September 5–8, 2006, Egmond Aan Zee, The Netherlands.

Gupta, R. D., Jamil, J. & Johsin, J. 1993 Discharge prediction in smooth trapezoidal free overfall – (positive, zero and negative slopes). *J. Irrig. Drain. Eng., ASCE* **119** (2), 215–224.

Hager, W. H. 1983 Hydraulics of the plane overfall. *J. Hydraul. Eng., ASCE* **109** (2), 1683–1697.

Jaeger, C. 1957 *Engineering Fluid Mechanics*. St. Martin's Press, Inc., New York, NY.

Keller, R. J. & Fong, S. S. 1989 Flow measurements with trapezoidal free overfall. *J. Irrig. Drain. Eng., ASCE* **115** (1), 125–136.

Keijzer, M. & Babovic, V. 1999 Dimensionally aware genetic programming. In *Proceedings of the Genetic and Evolutionary Computation Conference* (ed. W. Banzhaf, J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela & R. E. Smith), Vol. 2, pp. 1069–1076. Morgan Kaufmann, Orlando, Florida, USA.

Keijzer, M. & Babovic, V. 2002 Declarative and preferential bias in GP-based scientific discovery. *Genet. Program. Evol. Mach.* **3** (1), 41–79.

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.

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

Koza, J. R., Keane, M. A., Streeter, M. J., Mydlowec, W., Yu, J. & Lanza, G. 2003 *Genetic Programming IV: Routine Human-Competitive Machine Intelligence*. Kluwer Academic Publishers, Boston, MA.

Kutlu, I. 2005 *Scrutinizing Rectangular Free overfall*. Ms Thesis, Middle East Technical University, Turkey.

Ledoux, J. W. 1924 Open End Flume Water Meter based on Exponential Equation. *Engineering News Record*, September 25.

Marchi, E. 1993 On the free overfall. *J. Hydraul. Res.* **31** (6), 777–790.

Mohapatra, P. K., Bhallamudi, S. M. & Eswaran, V. 2001 Numerical study of flows with multiple free surfaces. *Int. J. Numer. Methods Fluids* **36**, 165–184.

Murty Bhallamudi, S. 1994 End depth in trapezoidal and exponential channels. *J. Hydraul. Res.* **32** (2), 219–232.

Ozturk, H. U. 2005 *Discharge predictions using ANN in sloping rectangular channels with free overfall*. MSc Thesis, Middle east technical University, Turkey.

Pagliara, S. 1995 Discussion on 'End depth in trapezoidal and exponential channels' by S. Murty Bhallamudi. *J. Hydraul. Res.* **33** (2), 283–286.

Pagliara, S. & Viti, C. 1995 Discussion on 'Discharge prediction in smooth trapezoidal free overfall' by R. D. Gupta, M. Jamil and M. Moshin. *J. Irrig. Drain. Eng., ASCE* **121** (1), 128–130.

Pal, M. & Goel, A. 2006 Prediction of the End depth ratio and discharge in semi circular and circular shaped channels using support vector machines. *Flow Meas. Instrum.* **17**, 50–57.

Pal, M. & Goel, A. 2007 Estimation of discharge and end depth in trapezoidal channel by support vector machines. *Water Resour. Manage.* **21**, 1763–1780.

Poli, R., Langdon, W. B. & McPhee, N. F. 2008 *A Field Guide to Genetic Programming*. Lulu.com Publisher, Raleigh, NC.

Raikar, R. V., Nagesh Kumar, D. & Dey, S. 2004 End depth computation in inverted semi circular channels using ANNs. *Flow Meas. Instrum.* **15**, 285–293.

Rajaratnam, N. & Muralidhar, D. 1964*a* End depth for exponential channels. *J. Irrig. Drain. Div., ASCE* **90** (1), 17–36.

Rajaratnam, N. & Muralidhar, D. 1964*b* End depth for circular channels. *J. Hydraul. Div., ASCE* **90** (2), 99–119.

Rajaratnam, N. & Muralidhar, D. 1968*a* Characteristics of rectangular free overfall. *J. Hydraul. Res.* **6** (3), 233–258.

Rajaratnam, N. & Muralidhar, D. 1968*b* The rectangular free overfall. *J. Hydraul. Div., ASCE* **94** (3), 849–850.

Rajaratnam, N. & Muralidhar, D. 1970 The trapezoidal free overfall. *J. Hydraul. Res.* **8** (4), 419–447.

Rajaratnam, N. & Muralidhar, D. 1976 Roughness effect of rectangular free overfall. *J. Hydraul. Div.* **102** (5), 559–614.

Ramamurthy, A. S., Qu, J. & Vo, D. 2006 VOF model for simulation of a free overfall in trapezoidal channels. *J. Irrig. Drain. Eng., ASCE* **132** (4), 425–428.

Rouse, H. 1936 Discharge characteristics of the free overfall. *Civil Eng., ASCE* **6** (4), 257–260.

Smith, C. D. 1962 Brink depth for a circular channel. *J. Hydraul. Div., ASCE* **88** (6), 125–134.

Southwell, R. V. & Vaisey, G. 1946 Relaxation methods applied to engineering problems. *Philos. Trans.* **240** (815), 117–161.

Sterling, M. 1998 *A Study of Boundary Shear Stress, Flow Resistance and the Free Overfall in Open Channels with a Circular Cross-section*. PhD Thesis, University of Birmingham, UK.

Sterling, M. & Knight, D. W. 2001 The free overfall as a flow measuring device in a circular channel. *Proc. Inst. Civil Eng., Water Maritime Eng.* **148** (4), 235–243.

Turan, C. 2002 *Experimental Studies on Free Overfall*. MSc Thesis, Middle East Technical University, Ankara, Turkey.

USBR 2001 *Water Measurement Manual, U.S. Bureau of Reclamation*. US Government Printing Office, Washington, DC.

Van Leer, B. R. 1922 The California Pipe Method of Water Measurement, *Engineering News Record*, August 3.

Van Leer, B. R. 1924 The California Pipe Method of Water Measurement, *Engineering News Record*, August 21.

Weisberg, S. 1980 *Applied Linear Regression*. Wiley, New York.

Yuen, K. W. H. 1989 *A Study of Boundary Shear Stress, Flow Resistance and Momentum Transfer in Open Channels with Simple and Compound Trapezoidal Cross Sections*. PhD Thesis, University of Birmingham, UK.

First received 4 December 2008; accepted in revised form 2 September 2009. Available online 18 March 2010