

Modeling scour depth downstream of grade-control structures using data driven and empirical approaches

Kiyoumars Roushangar, Samira Akhgar, Ali Erfan and Jalal Shiri

ABSTRACT

Local scour occurs in the immediate vicinity of structures as a result of impinging on a bed with a high velocity flow. Prediction of scour depth has an important role in control structure management and water resource engineering issues, so a study of new heuristic expressions governing it is necessary. The present study aims to investigate different methods' capabilities to estimate scour depth downstream of grade-control structures using field measurements from the literature.

Accordingly, data driven feed forward neural network and gene expression programming techniques were selected for the investigation. Additionally, the optimum data driven based scour depth models were compared with the corresponding physical–empirical based formulas. Three data categories corresponding to (a) scouring downstream of a ski-jump bucket, (b) a sharp-crested weir, and (c) an inclined slope controlled structure (as grade-control structures) were applied as reference patterns for developing and validating the applied models. A sensitivity analysis was also performed to identify the most influential parameters on scouring. The obtained results indicated that the applied methods have promising performance in estimating the scour depth downstream of spillways and control structures. Nevertheless, the applied data driven approaches show higher accuracy than the corresponding traditional formulas.

Key words | gene expression programming, grade-control structures, local scour, neural networks

Kiyoumars Roushangar (corresponding author)

Samira Akhgar

Ali Erfan

Department of Civil Engineering,

University of Tabriz,

Tabriz,

Iran

E-mail: kroshangar@yahoo.com

Jalal Shiri

Department of Water Engineering, Faculty of

Agriculture,

University of Tabriz,

Tabriz,

Iran

INTRODUCTION

When a high velocity jet impinges a bed (which consists of sand, gravel, or weak and jointed rock), deep scour holes might be formed downstream of spillways and control structures. So, construction of a control structure downstream of spillways is necessary to prevent severe erosion of the bed and formation of deep scour holes. An important common feature in all local scouring problems is the existence of strong secondary flows, along with a vortex system induced by a hydraulic jump, that govern the scouring downstream of a spillway (Guan *et al.* 2014). Study of the scour phenomenon formed downstream of spillways and control structures is necessary because it crucially endangers the safety of the main structures as well as the adjacent structures, such as training walls and guide bunds. Various

factors might govern the scour depth, including flow discharge, height of fall, bucket radius, lip angle, type of rock and the degree of its homogeneity, the time factor of the scouring process, and the mode of spillway operation, which makes it difficult to study this process (Bormann & Julien 1991; Azamathulla *et al.* 2005). The literature review by the authors showed that there are two varieties of studies focusing on the scour depth simulation: (I) studies focusing on the assessment of scour depth estimation formulas and (II) studies focusing on the temporal variations of scouring. Among group (I), Mason & Arumugam (1985), Chee & Yuen (1985), Yen (1987), and D'Agostino & Ferro (2004) proposed physical-based semi-empirical models for estimating scour depth, which will be discussed in the Methods section.

On the other side and in the context of heuristic model applications (group II), data driven approaches (e.g. artificial neural networks (ANNs) and gene expression programming (GEP)) have been applied across wide aspects of water resources engineering in recent years. Among others, ANNs have been applied for predicting the friction factor of open channel flow (Yuhong & Wenxin 2009), prediction of water resource variables in river systems (Maier *et al.* 2010), estimating river suspended sediment (e.g. Lin & Namin 2005; Hamidi & Kayaalp 2008; Yang *et al.* 2009), predicting daily and monthly suspended sediment load (Nourani & Andalib 2015), and modeling dissolved oxygen (DO) concentration in lakes (Bertone *et al.* 2015).

Further, GEP has been applied for modeling river suspended sediment loads (Kisi & Shiri 2012; Kisi *et al.* 2012), modeling energy dissipation over spillways (Roushangar *et al.* 2014a), modeling river total bed material load discharge (Roushangar *et al.* 2014b), modeling the friction factor in alluvial channels (Roushangar *et al.* 2014c), modeling solid load discharge of an alluvial river (Roushangar & Alizadeh 2015), as well as estimating sediment transport in sewer pipes (Bonakdari & Ebtehaj 2015), all of which have confirmed the GEP capabilities for mapping the nonlinear behavior of the input-target data set.

Babovic & Abbott (1997) applied genetic programming (GP) for the evolution of equations in hydraulic issues. Azamathulla *et al.* (2005) estimated scour depth downstream of a ski-jump bucket using neural networks. Azamathulla *et al.* (2006) applied neural networks to estimate scour below spillways. Azamathulla *et al.* (2008) applied GP to predict ski-jump bucket spillway scour. Babovic (2000) studied sediment transport utilizing GP and evolutionary algorithms. Azamathulla & Ghani (2011) predicted the scour depth at culvert outlets by using an adaptive neuro-fuzzy inference system (ANFIS). Sheikhalipour & Hassanpour (2013) estimated suspended sediment loads using GEP. Azamathulla & Zakaria (2007) predicted the scour below flip-bucket spillways using ANFIS-based approaches. Dargahi (2003) evaluated the scour development downstream of spillways. Azamathulla *et al.* (2010) predicted sediment concentration for Malaysian rivers utilizing the GEP approach. Guven (2011) employed GEP in estimating scour geometry downstream of hydraulic structures. GEP was also utilized by the following: Azamathulla & Haque (2012) in predicting

scour depth at culvert outlets; Azamathulla (2012) in predicting scour depth downstream of sills; and Guven & Gunal (2008) for predicting local scour downstream of hydraulic structures. Liao *et al.* (2012) used a probabilistic approach to create a preliminary inspection evaluation form for scour potential at bridge sites.

The present literature review shows that most studies have been focused on ANN and ANFIS applications for modeling scour depth (Azamathulla & Ghani 2011; Azamathulla 2012).

The present study aims at investigating the physically based traditional and data driven model (including ANN and GEP) capabilities for scour depth simulation, as well as comparison of the applied models. Therefore, modeling scour depth downstream of an inclined slope controlled structure using heuristic models has been carried out here for the first time. A sensitivity analysis was also performed to describe the most influential input parameters of the data driven models.

METHODOLOGICAL STRUCTURE

Data used

In this paper, three kinds of data corresponding to scouring downstream of (i) a ski-jump bucket (Figure 1(a)), (ii) a sharp-crested weir (Figure 1(b)), and (iii) an inclined slope controlled structure (Figure 1(c)), captured from published literature, are used.

Ski-jump bucket

Data corresponding to scouring downstream of a ski-jump bucket include 95 patterns that have been gathered at the Central Water and Power Research Station, India (Azamathulla *et al.* 2005).

Sharp-crested weir

The experimental data set used for simulating scour depth downstream of a sharp-crested weir includes 225 patterns captured from D'Agostino & Ferro (2004). The experiments carried out by Veronese (1937) include discharges varying from 0.001–0.083 (m^2s^{-1}) and maximum scour depths of

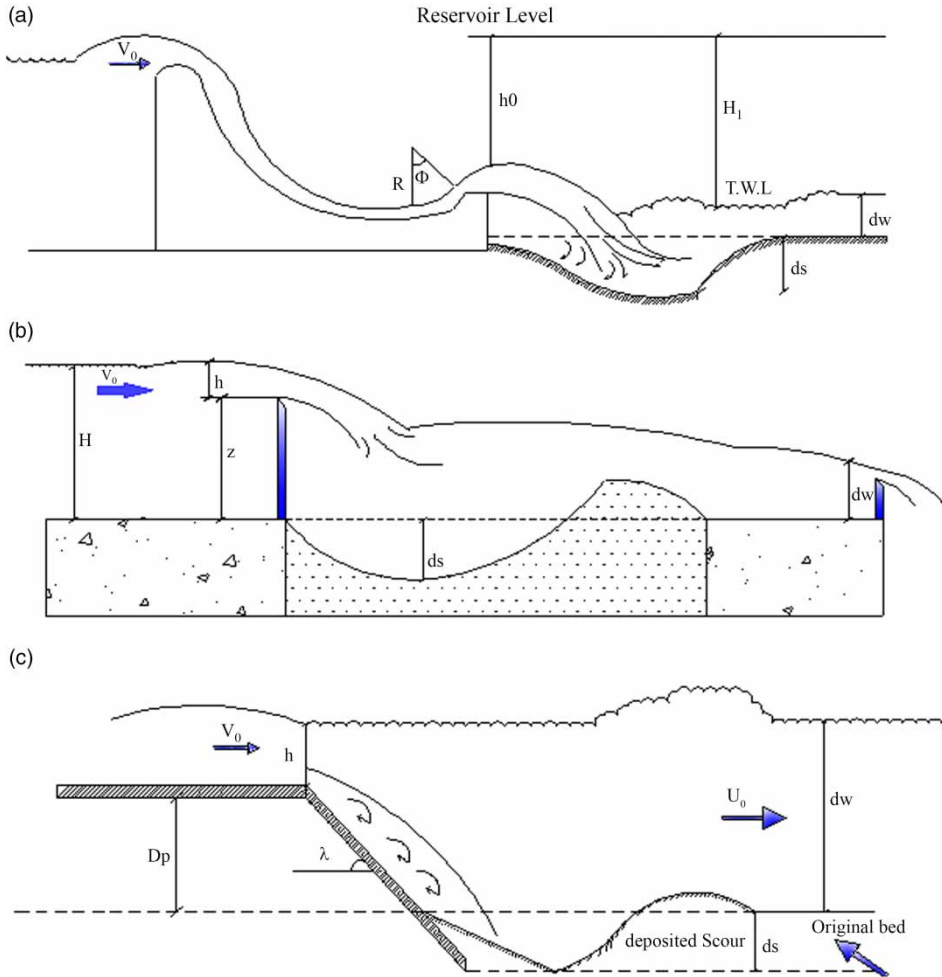


Figure 1 | Schematic presentation of various grade control structures: (a) ski-jump bucket, (b) sharp-crested weir, and (c) inclined slope.

0.055–0.22 m. The ratio between the weir width (b) and the channel width (B) was established to be unity ($b = B = 0.5$ m), and the water level difference between the upstream and downstream was set to be $H = 1$ m. [Mossa \(1998\)](#) carried out some experimental runs by modeling a sharp-crested weir. The experimental channel was 0.3 m in width, accompanied by a ratio of b/B as unity. The experimental discharges varied from 0.0045 to 0.0148 ($\text{m}^2 \text{s}^{-1}$), with corresponding maximum scour depths of 0.0352 and 0.145 m. [D'Agostino \(1994\)](#) carried out some experiments (114 run) in a channel of 0.5 m width. The runs were carried out through adopting two values of the ratio b/B (0.3 and 0.6), two values of the weir height (0.41 and 0.71 m) and two different gravel mixtures to reproduce the alluvial bed. The experimental discharges ranged from 0.0167 to 0.167 ($\text{m}^2 \text{s}^{-1}$),

producing maximum scour depths of 0.045 and 0.285 m, respectively.

Inclined slope

The experimental data set used for simulating scour depth downstream of slope control includes 88 patterns that have been carried out by [Bormann & Julien \(1991\)](#). These experiments have been conducted in a large outdoor flume (the ratio between weir channel width was unity, $b = B = 0.91$ m) at the Colorado State University Engineering Research Center. The elevation of the grade-control structure model was set at 2.13 m above the flume floor, with three downstream angles of the structure ($\lambda = 1.57, 0.79$ and 0.32 radian).

GEP

GEP was developed by Ferreira (2001) using fundamental principles of the genetic algorithms (GA) and GP (Koza 1992). GEP mimics biological evolution to create a computer program for simulating different phenomena. In GEP, a mathematical function is described as a chromosome with multi genes, and developed using the data presented to it. GEP performs the symbolic regression using most of the genetic operators of GA.

ANNs

ANNs are learning systems, consisting of a number of interconnected simple processing elements called neurons or nodes with the attractive attribute of information processing characteristics such as nonlinearity, parallelism, noise tolerance, learning and a generalization capability (Haykin 1998).

Feed forward neural networks (FFNN) with a back-propagation (BP) algorithm are commonly used ANNs in engineering applications. It has been proved that three-layered BP network models are satisfactory in forecasting and simulating issues (e.g. Hornik 1989; ASCE 2000; Nourani et al. 2008). The standard BP algorithm is a gradient descent algorithm, in which the network weights are modified through a negative direction of the gradient of performance function.

Semi-empirical formulas

In order to assess the performance of data driven models with respect to traditional formulas, the results of the GEP and FFNN models were compared with the corresponding selected semi-empirical formulas.

Ski-jump bucket formulas

1. Veronese formula (Bureau of Indian Standards 1985):

$$d_s = 1.90q^{0.54}H_1^{0.225} \quad (1)$$

where d_s is the equilibrium depth of scour below the tail water level, q is the unit discharge and H_1 is the head between the upper (reservoir) water level and the tail water level.

2. Azamathulla et al. (2005) suggested estimating the maximum scour depth downstream of a ski-jump bucket, as:

$$d_s = 6.914d_w \left(\frac{q}{\sqrt{gd_w^3}} \right)^{0.694} \left(\frac{H_1}{d_w} \right)^{0.0815} \left(\frac{R}{d_w} \right)^{-0.235} \left(\frac{d_{50}}{d_w} \right)^{0.196} (\emptyset)^{0.196} \quad (2)$$

where d_s , q , R , \emptyset , d_w , H_1 , d_{50} , g denote the equivalent scour depth, unit discharge, radius of the bucket, lip angle of bucket, tail water depth, height of water on the control structure, bed material size and gravitational acceleration, respectively.

3. Mason & Arumugam (1985):

$$d_s = 3.27 \left(\frac{q^{0.6}H_1^{0.05}d_w^{0.15}}{g^{0.2}d_{50}^{0.1}} \right) - d_w \quad (3)$$

where d_s , q , d_w , H_1 , d_{50} , g stand for equivalent scour depth, unit discharge, tail water depth, head between the upper (reservoir) water level and the tail water level, bed material size and gravitational acceleration, respectively.

4. Sofrelec (1980):

$$d_s = 2.3q^{0.6}H_1^{0.1} \quad (4)$$

where d_s , q , H_1 are the equivalent scour depth, unit discharge, and head between the upper (reservoir) water level and the tail water level, respectively.

Sharp-crested weir formulas

1. Yen (1987):

$$\frac{d_s}{\left(\frac{q^2}{g} \right)^{1/3}} = (6.42 - 3.1H^{0.1})g^{-H/600} \left(\frac{gH^5}{q^2} \right)^{20+H/600} \left(\frac{H}{d_s} \right)^{1/10} \left(\frac{h}{H} \right)^{3/20} \quad (5)$$

with the equilibrium scour depth (d_s), unit discharge (q), total head (H), head of water falling from the weir (h) and gravitational acceleration (g).

2. D'Agostino & Ferro (2004):

$$\frac{d_s}{z} = 0.540 \left(\frac{b}{z}\right)^{0.593} \left(\frac{h}{H}\right)^{-0.126} (d_{50})^{0.544} \left(\frac{d_{90}}{d_{50}}\right)^{-0.856} \left(\frac{b}{B}\right)^{-0.751} \quad (6)$$

where d_s , H , d_{50} , d_{90} , b , B , h and z denote the equilibrium scour depth, total head, mean sediment size, diameter for which 90% of grains' size is smaller than, width of the weir, width of channel, head of water falling from the weir, and weir height, respectively.

Inclined slope controlled structure formulas

1. Chee & Yuen (1985):

$$d_s = 0.6 \frac{q^{0.45} U_0^{0.55} \sin \lambda}{d_{50}^{0.1}} - z \quad (7)$$

2. Bormann & Julien (1991)

$$d_s = 0.7 \frac{q^{0.45} U_0 d_w^{0.12} \sin \lambda}{g^{0.73} d_{50}^{0.3}} \quad (8)$$

where in both Equations (7) and (8), d_s , U_0 , q , λ , d_{50} , d_w and z represent the equilibrium scour depth, jet velocity entering tail water, unit discharge, the face angle of the control structure, mean sediment size, tail water depth and weir height, respectively.

Input selection

Input selection is a key aspect in data driven model development and applications. Several approaches have been suggested and reviewed by investigators for determining the most relevant input variables to feed data driven models (ASCE 2000; Bowden et al. 2005). Among others, the input selection methods based on *a priori* knowledge of the modeled system (Bowden et al. 2005) might be a successful alternative for modeling scour depth, taking into consideration the physical parameters affecting this phenomenon. Since there are multitudes of parameters inflecting local scour depth, some very important parameters (obtained

through physical judgment) will be applied here as dimensionless input variables. A combination of this approach with sensitivity analysis was used here for better describing the most effective parameters. Once the optimum data driven approaches were selected, a sensitivity analysis was carried out to distinguish the irrelevant parameters.

Input selection for ski-jump bucket

The scour depth downstream of the ski-jump bucket might be taken as a function of the flow characteristics, the geometry of the ski-jump bucket, the bed material and water depth upstream and downstream of the spillway (Azamathulla et al. 2005). A functional relationship might be given as:

$$d_s = f(d_s, q, H_1, R, d_{50}, \varnothing, d_w, h_0, g, b) \quad (9)$$

where d_s , q , H_1 , R , d_{50} , \varnothing , d_w , h_0 , g and b denote the equilibrium scour depth, unit discharge, head between the upper (reservoir) water level and the tail water level, radius of the bucket, mean sediment size, lip angle of bucket, tail water depth, difference between upstream head and water jet surface on the jump bucket, gravitational acceleration, and width of spillway, respectively, (Figure 1(a)). Accordingly, the dimensionless parameters can be written as:

$$\frac{d_s}{d_w} = f\left(\frac{q^2}{gh_0^3}, \frac{h_0}{R}, \varnothing, \frac{R}{d_{50}}, \frac{q^2}{gH_1^3}, \frac{d_w}{d_{50}}, Fr_2, \frac{H_1}{d_w}\right) \quad (10)$$

where Fr_2 is the Froude number of flow downstream of the ski-jump bucket,

$$\left(Fr_2 = \frac{q}{\sqrt{gd_w^3}}\right). \quad (10a)$$

Consequently, nine input configurations were built here for studying the effect of the non-dimensional parameters on the scour depth. Table 1 sums up the introduced input configurations.

Input selection for sharp-crested weir

The scour depth downstream of a sharp-crested weir might be considered as a function of the flow characteristics, the

Table 1 | Input configurations of data driven models

Model	Input variables	Model	Input variables	Model	Input variables
<i>Ski-jump bucket data</i>					
MB1	$(q^2/g h_0^3)$	MB4	$(q^2/g h_0^3, h_0/R, \emptyset, R/d_{50})$	MB7	$(q^2/g H_1^3, d_w/d_{50}, Fr_2)$
MB2	$(q^2/g h_0^3, h_0/R)$	MB5	$(q^2/g H_1^3, d_w/d_{50})$	MB8	$(R/d_{50}, Fr_2, H_1/d_w)$
MB3	$(q^2/g h_0^3, h_0/R, \emptyset)$	MB6	$(q^2/g H_1^3, Fr_2)$	MB9	$(\emptyset, d_w/d_{50}, Fr_2, H_1/d_w)$
<i>Sharp edge overflow data</i>					
MS1	$(h/H, b/z)$	MS3	$(Fr_1, h/H, b/z, d_{50}/d_w)$	MS5	$(Fr_1, h/H, b/z, d_{50}/d_w, Fr_2)$
MS2	$(h/H, b/z, d_{50}/d_w)$	MS4	$(Fr_1, d_w/H)$	MS6	$(Fr_1, d_{50}/d_w, d_w/H)$
<i>Inclined slope controlled structure data</i>					
MI1	(Fr_2)	MI4	$(Fr_1, D_p/h, \lambda)$	MI7	(Fr_1, Fr_2)
MI2	$(Fr_2, D_p/h)$	MI5	$(Fr_2, D_p/h, \lambda, d_{50}/d_{90})$	MI8	$(Fr_1, d_w/d_{50})$
MI3	$(Fr_2, D_p/h, \lambda)$	MI6	$(Fr_1, D_p/h, \lambda, d_{50}/d_{90})$	MI9	$(Fr_1, Fr_2, \lambda, d_{50}/d_{90})$

geometry of the sharp-crested weir type of bed material and water depth upstream and downstream of the weir, as follows (D'Agostino & Ferro 2004):

$$d_s = f(q, H, b, d_{50}, d_w, h, z) \quad (11)$$

where d_s , q , H , d_{50} , d_w , h and z , stand for the equilibrium scour depth, unit discharge, upstream total head, mean sediment size, tail water depth, head of water falling from the weir, and weir height, respectively (Figure 1(b)).

Accordingly, the dimensionless expression would read:

$$\frac{d_s}{d_w} = f\left(\frac{h}{H}, \frac{b}{z}, \frac{d_{50}}{d_w}, \frac{d_w}{H}, Fr_1, Fr_2\right) \quad (12)$$

where Fr_1 is the Froude number of flow upstream ($Fr_1 = q/\sqrt{gh^3}$) and Fr_2 is the Froude number of flow downstream of a sharp-crested weir ($Fr_2 = q/\sqrt{gd_w^3}$). Using this non-dimensional form, six input configurations were constructed to feed data driven models (Table 1).

Input selection for inclined slope controlled structure

Considering the available effective parameters in literature, the scour depth downstream of an inclined slope controlled structure might be written as:

$$d_s = f(q, h, \lambda, d_{50}, d_w, d_{90}, D_p) \quad (13)$$

where d_s , q , λ , d_{50} , d_w , h , d_{90} , and D_p denote, respectively, the equilibrium scour depth, unit discharge, the angle of a control structure with horizontal surface, mean sediment size, tail water depth, head of water falling from the weir, diameter for which 90% of particles are finer, and the drop height. Accordingly, the dimensionless expression would read:

$$\frac{d_s}{d_w} = f\left(\frac{D_p}{h}, \lambda, \frac{d_{50}}{d_{90}}, \frac{d_w}{d_{50}}, Fr_1, Fr_2\right) \quad (14)$$

where $Fr_1 = q/\sqrt{gh^3}$ is the Froude number of flow upstream and $Fr_2 = q/\sqrt{gd_w^3}$ is the Froude number of flow downstream of a sharp-crested weir. Having selected these non-dimensional parameters, nine input configurations were defined to feed data driven models (Table 1).

Performance criteria

The models' performance was assessed using three statistical criteria, namely the Pearson's correlation coefficient (CC), the mean absolute error (MAE), Determination Coefficient (R^2) and the root mean square error (RMSE), expressions for which are as follows:

$$CC = \frac{\sum_{i=1}^N (O_i - \bar{O}_i)(P_i - \bar{P}_i)}{\sqrt{\sum_{i=1}^N (O_i - \bar{O}_i)^2 \sum_{i=1}^N (P_i - \bar{P}_i)^2}} \quad (15)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |O_i - P_i| \quad (16)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{N}} \quad (17)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O}_i)^2} \quad (18)$$

where O_i and P_i represent the measured and simulated scour depth values, respectively. \bar{O}_i and \bar{P}_i denote the mean of the measured and simulated values, respectively. N stands for the number of patterns.

RESULTS AND DISCUSSION

Model implementation

GEP models

In this study, four basic arithmetic operators (+, −, ×, /) and four basic mathematical functions ($\sqrt{\quad}$, exp, sin, cos) were utilized as the GEP function set. Further, each GEP model was evolved till the fitness function remained unchanged for 10,000 runs for each pre-defined number of genes (varying from 3 to 4), then the program was stopped. The model parameters and the size of the developed GEP models were then tuned (optimized) throughout refining (optimizing) the trained and fixed model as a starter. Table 2 lists the combination of all genetic operators used in this study. Further details about GEP applications in hydraulic structures studies might be found in e.g. Roushangar et al. (2014a, 2014b).

ANN models

Multi-layer FFNN can have more than one hidden layer, although some studies have shown that a single hidden layer might be applied for approximating complex nonlinear functions (Tsukamoto 1979). Therefore, one hidden layer-FFNN was used for modeling scour depth. A difficult task with

Table 2 | Parameters of GEP models used in this study

Description of parameter	Setting of parameter
Function set	+, −, *, /, $\sqrt{\quad}$, Exp, sin, cos
Chromosomes	30
Head size	7
Number of genes	3, 4
Linking function	Addition
Fitness function error type	RMSE
Mutation rate	0.044
Inversion rate	0.1
One-point recombination rate	0.3
Two-point recombination rate	0.3
Gene recombination rate	0.1
Gene transposition rate	0.1

FFNN application involves choosing its architecture, e.g. the number of hidden nodes and the learning rate. Here, the number of hidden layer nodes is determined by trial and error (through examining various network structures). Consequently, several architectures were examined with variable hidden layer nodes (from 2 to 7). The tangent sigmoid and pure linear functions were found to be appropriate when used as the hidden and output node activation functions, respectively. Feed-forward-BP, TRAINLM (Levenverg–Marquardt algorithm) and LEARNGDM (gradient descent momentum) were applied for network type, training function and adoption learning function, respectively. The training of the FFNN models was stopped when the acceptable level of error was achieved. FFNN was implemented using MATLAB.

Assessing different data pattern management scenarios

In applying GEP and FFNN techniques, a random pattern selection data management scenario was used to obtain the best train-test blocks size. Although this might not be comparable to the most capable data management techniques e.g. k-fold testing (Shao 1993), the present methodology would allow reduction of the over training risk and an increase in the model validity outside the applied train-test patterns. Accordingly, three scenarios including 70%–30%, 65%–35%, and 60%–40% were evaluated to select the best data-partitioning (management)

configuration. These values stand for the partitions of the data that have been used for training/testing phases. For instance, in the 70%–30% scenario, 70% of data patterns were used for training the models and 30% were reserved for testing. Figure 2 shows the R^2 values of the applied data management scenarios for the GEP and FFNN models. The figure clearly depicts that the 65%–35% mode gives the most accurate results, so this pattern was applied for establishing and testing the data driven models. Figure 2 also compares the testing R^2 values of the employed FFNN and GEP models for a ski-jump bucket, sharp-crested weir and inclined slope control structure. From the figure, it is clear that the R^2 values for all the employed ski-jump models are more than 0.5, with a maximum R^2 value corresponding to the MB4 model. Among the sharp-crested weir

models, the MS5 model represents the largest values of R^2 for both the GEP and FFNN techniques. Finally, for the inclined slope control structure, the maximum R^2 values correspond to those of the MI3 model.

Model validation

Ski-jump bucket

In this part, 62 data patterns (65% of all patterns) were used for training and 33 data patterns (35% of the patterns) were used for random testing. Table 3 represents the testing statistics for all nine ski-jump bucket scour depth models (fed with the aforementioned nine input configurations). The table clearly shows that the MB4 model has the highest accuracy. Figure 3(a) displays the observed and simulated scour depth values of GEP and FFNN models (fed with MB4 input configuration) for the testing stage.

Sharp-crested weir

From 203 patterns of the sharp-crested weir, 132 patterns were selected for training and 71 patterns were reserved

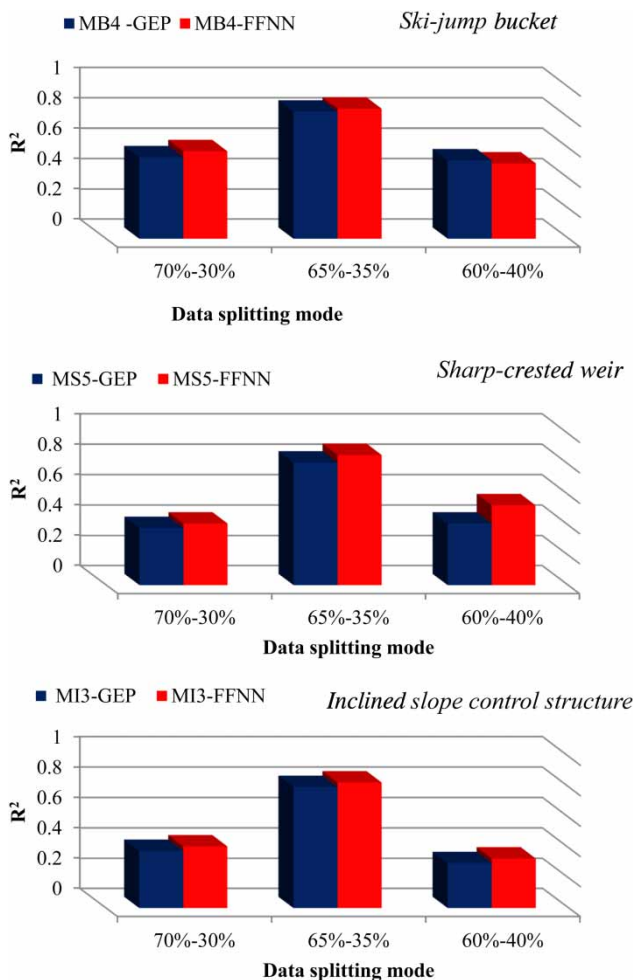


Figure 2 | R^2 values of employed data management scenarios.

Table 3 | Testing statistics of the GEP and FFNN models for the ski-jump bucket

Input configuration	Model	CC	R^2	RMSE	MAE	Hidden neuron no.
MB1	GEP	0.291	0.370	0.071	0.051	–
	FFNN	0.112	0.181	0.084	0.056	2
MB2	GEP	0.526	0.687	0.041	0.033	–
	FFNN	0.493	0.503	0.057	0.045	3
MB3	GEP	0.758	0.821	0.032	0.028	–
	FFNN	0.169	0.294	0.091	0.052	3
MB4	GEP	0.841	0.858	0.034	0.025	–
	FFNN	0.863	0.876	0.032	0.024	5
MB5	GEP	0.532	0.637	0.046	0.036	–
	FFNN	0.456	0.585	0.059	0.035	5
MB6	GEP	0.133	0.233	0.072	0.053	–
	FFNN	0.411	0.479	0.063	0.049	3
MB7	GEP	0.788	0.839	0.038	0.029	–
	FFNN	0.769	0.845	0.038	0.032	3
MB8	GEP	0.611	0.792	0.047	0.042	–
	FFNN	0.789	0.801	0.039	0.029	5
MB9	GEP	0.813	0.830	0.039	0.032	–
	FFNN	0.501	0.513	0.072	0.047	7

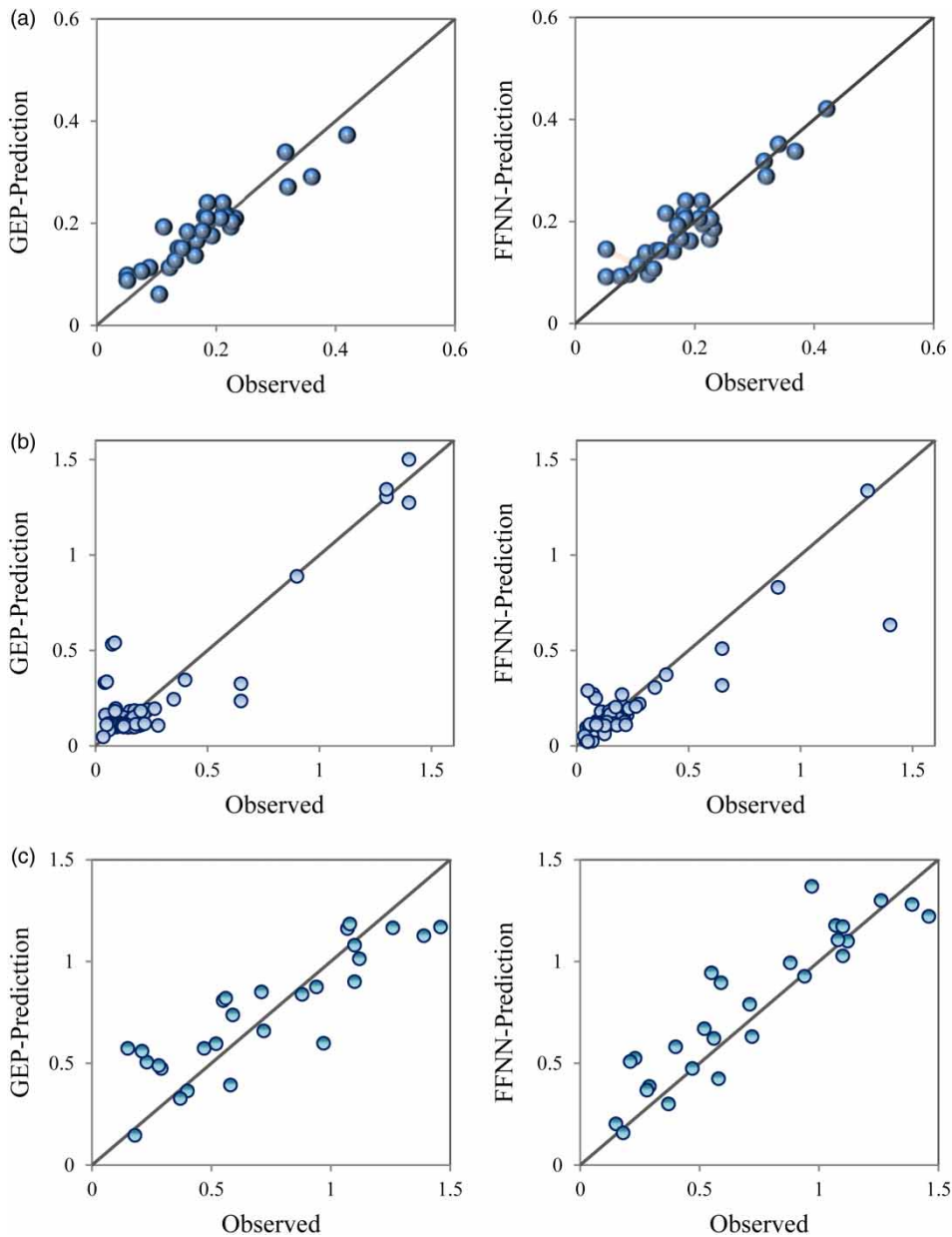


Figure 3 | Observed vs. predicted scour depths testing stage of GEP and FFNN models for (a) ski-jump bucket, (b) sharp-crested weir, and (c) inclined slope controlled structure.

for testing. Table 4 summarizes the statistical criteria of all models. From the statistics, the MS5-based GEP and FFNN models have the highest accuracy. Figure 3(b) presents the observed vs. simulated scour depth values.

Inclined slope controlled structure

From the existing 80 patterns of inclined slope controlled structure data, 53 patterns were used for training and 27

patterns were considered for testing. Table 5 sums up the corresponding error statistics for the GEP and FFNN models. From the table, it is observed that the models corresponding to the MI3 configuration present the highest accuracies. Figure 3(c) illustrates the observed vs. simulated scour depth values of GEP and FFNN models.

Comparing the results of all three categories, it is seen that the error values (in term of *RMSE* and *MAE*) are relatively lower for the ski-jump bucket. Although the *RMSE* and *MAE*

Table 4 | Testing statistics of the GEP and FFNN models for sharp edge overflow

Input configuration	Model	CC	R ²	RMSE	MAE	Hidden neuron no.
MS1	GEP	0.511	0.629	0.194	0.078	–
	FFNN	0.319	0.426	0.385	0.162	2
MS2	GEP	0.723	0.766	0.127	0.079	–
	FFNN	0.631	0.679	0.191	0.097	3
MS3	GEP	0.759	0.817	0.152	0.085	–
	FFNN	0.813	0.850	0.126	0.062	5
MS4	GEP	0.633	0.692	0.119	0.086	–
	FFNN	0.418	0.470	0.372	0.155	3
MS5	GEP	0.814	0.874	0.112	0.088	–
	FFNN	0.862	0.885	0.104	0.059	5
MS6	GEP	0.763	0.833	0.149	0.123	–
	FFNN	0.799	0.841	0.154	0.066	3

Table 5 | Testing statistics of the GEP and FFNN models for inclined slope controlled structure

Input configuration	Model	CC	R ²	RMSE	MAE	Hidden neuron no.
MI1	GEP	0.211	0.399	0.311	0.284	–
	FFNN	0.198	0.295	0.335	0.282	2
MI2	GEP	0.502	0.514	0.282	0.246	–
	FFNN	0.613	0.666	0.239	0.202	2
MI3	GEP	0.803	0.769	0.198	0.164	–
	FFNN	0.833	0.835	0.175	0.131	3
MI4	GEP	0.633	0.699	0.221	0.186	–
	FFNN	0.702	0.745	0.208	0.185	5
MI5	GEP	0.588	0.679	0.234	0.192	–
	FFNN	0.598	0.741	0.194	0.167	5
MI6	GEP	0.657	0.717	0.223	0.186	–
	FFNN	0.635	0.677	0.232	0.197	7
MI7	GEP	0.786	0.819	0.279	0.239	–
	FFNN	0.509	0.551	0.268	0.202	3
MI8	GEP	0.691	0.733	0.212	0.174	–
	FFNN	0.609	0.658	0.237	0.197	3
MI9	GEP	0.687	0.724	0.209	0.169	–
	FFNN	0.702	0.751	0.225	0.175	5

values are dimensional measures and might not be used for comparison between different cases, the difference between these measures provides information about the variances in the observed and simulated values. The analysis of these differences (not presented here) showed that the values are of small magnitude for ski-jump bucket models, emphasizing the lower variance values of the simulated values and the higher accuracy of the applied models. Nevertheless, the comparison of GEP and

FFNN models in all cases couldn't provide a general conclusion about the superiority of one technique over the other. While GEP surpasses the FFNN in the case of the ski-jump bucket, the results of the FFNN are more promising for the inclined slope control structure. However, the GEP models have the capability to give explicit mathematical expressions of the studied phenomena. To better describe and interpret the GEP formulation, it is necessary to use parsimony pressure to reduce the size of the parse trees. In the present study, parsimony pressure was also applied to condense the produced expressions. Figure 4 shows the Expression Tree (ET) of GEP-based models for all three categories of data. In the figure, d0, d1, d2 and d3 in part (a) denote q^2/gH_0^3 , ϕ , H_0/R and R/d_{50} , respectively, and d0, d1, d2, d3 and d4 in part (b) denote, q^2/gH_0^3 , b/z , d_m/d_w , H_0/H_1 and Fr_2 , respectively. Finally, in part (c), d0, d1 and d2 denote λ , Fr_0 and Z/H_0 , respectively.

The mathematical expression of these ETs reads as follows:

MB4 Model (ski-jump bucket data)

$$d_s = \sqrt{\frac{q^2}{gh_0^3} \left(0.99 + \frac{h_0}{R} \right) * \sqrt{\frac{h_0}{R}} * \sqrt{\frac{R}{d_{50}}} + \frac{q^2}{gh_0^3} (14.28\phi^2)} [9.28 \left(\frac{H_0}{R} \right)^{-1} + \frac{\phi H_0}{R}] + \cos \left\{ 3.78 \frac{H_0}{R} \left(\phi + \frac{q^2}{gH_0^3} \right) \right\} * \left(1 + 0.3 \frac{R}{d_{50}} \right) * 10^{-3}$$

MS5 Model (sharp-crested weir data)

$$d_s = A \tan (2Fr_2 * A \tan \left(\frac{h}{H} \right) * \sin \left(7.88 \left(\frac{d_{50}}{d_w} \right)^{1/3} \right) + \sin \left\{ A \tan \left[\left(\frac{d_{50}}{d_w} + 4.23 \right) * \frac{b}{z} * Fr_1^4 \right]^3 \right\}^2 + Fr_1^2 * Fr_2 * \cos \left[\left(0.088 + \frac{b}{z} \right) * \sin \left(\frac{b}{z} \right) \right]$$

MI3 (inclined slope controlled structure data)

$$d_s = \text{Exp} \left\{ (0.517 - \lambda)^3 - \left[\left(\frac{D_p}{h} \right)^2 + \left(\frac{0.45}{\lambda} \right) \right] \right\} + \text{Exp} \{ \cos (Fr_2)^3 + (Fr_2 + \lambda - 9.12) \} + A \tan \left\{ A \tan \left[\sin \left(\left(9.39 + \frac{D_p}{h} \right) - (1.24Fr_2) \right) \right]^2 \right\}$$

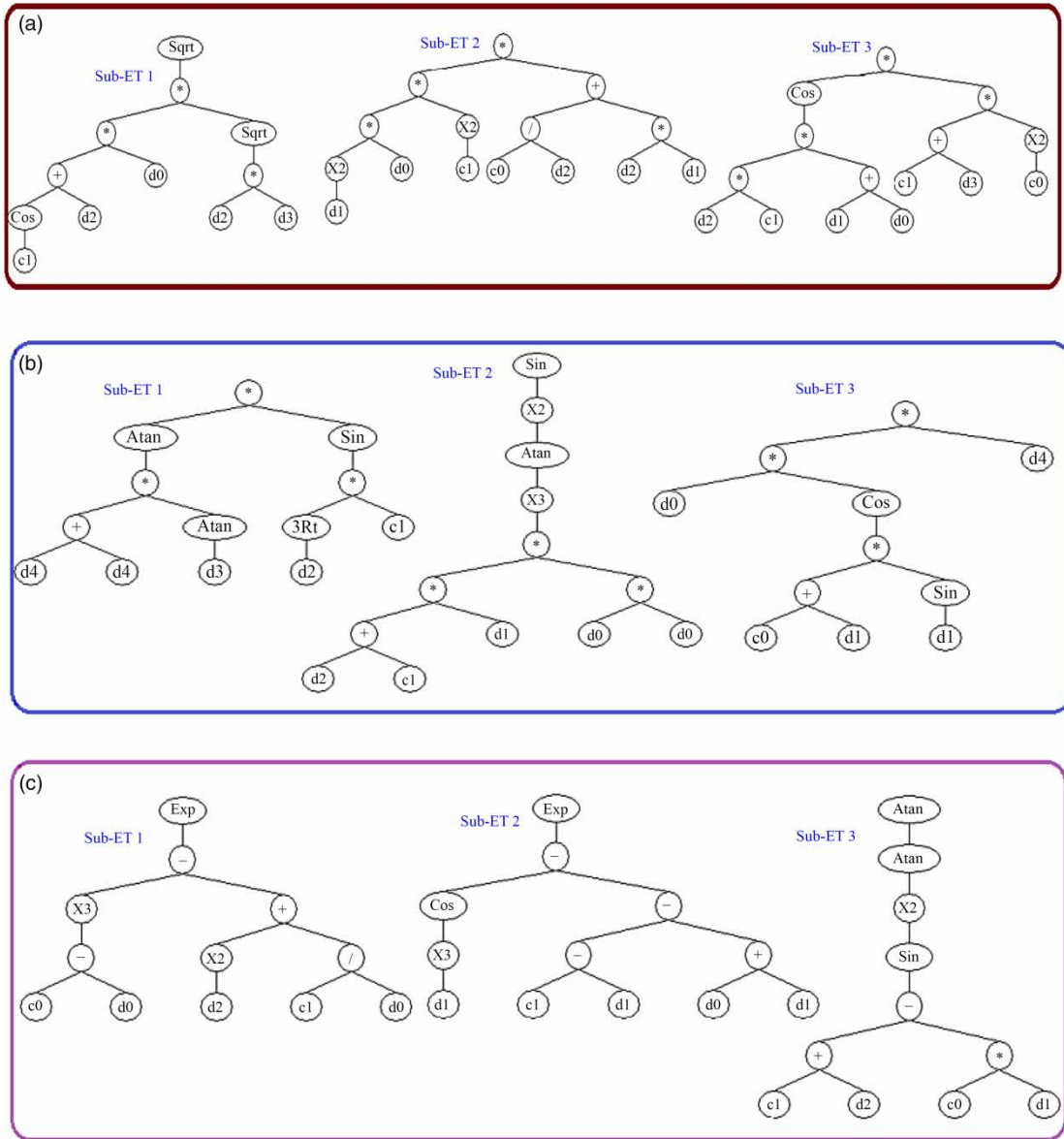


Figure 4 | ET for GEP formulation in (a) ski-jump bucket [MB4], (b) sharp-crested weir [MS5], and (c) inclined slope control structure [MI3].

As can be seen, GEP expressions have a high degree of complexity, which might be attributed to the non-linearity of the relations between the flow characteristics, the geometry of the grade control, the bed materials property and depth scour. The results confirm the importance of a proper input selection process in determining the model's accuracy and complexity.

Sensitivity analysis of the GEP and FFNN models

Based on the obtained results, the MB4, MS5 and MI3 models are selected as the best models for simulating scour depth for the ski-jump bucket, sharp-crested weir and inclined slope controlled structure cases, respectively. The MB4 model includes four dimensionless parameters,

i.e. (q^2/gh_0^3 , h_0/R , \emptyset , R/d_{50}). Table 6 shows the results of the sensitivity analysis of the GEP and FFNN models. According to the table, q^2/gh_0^3 is the most important parameter affecting the scour depth simulation downstream of the ski-jump bucket, while \emptyset shows the lowest impact.

In the case of the MS5 model, which includes four dimensionless parameters, i.e. (Fr_1 , d_{50}/d_w , h/H , b/z , Fr_2), the Fr_1 term seems to be the most important parameter affecting the scour depth downstream of the sharp-crested weir, whereas b/z shows the lowest impact (Table 6). Finally, for the slope controlled structure scour depth simulation, the MI3 model comprising 3 input parameters i.e. (λ , Fr_2 , D_p/h_0), according to Table 6, the omission of Fr_2 decreases the model accuracy to a great extent, so it seems that this parameter is the most influential factor in modeling

scour depth downstream of an inclined slope controlled structure.

Comparison of data driven and semi-empirical models

Performance criteria for semi-empirical formulas corresponding to the ski-jump bucket, sharp-crested weir and inclined slope, are listed in Table 7. The table also summarizes the results of the best data driven models. From the table, it is seen that the GEP and FFNN models are superior to semi-empirical models in estimating scour depth. Figures 5–7 compare the performance of the applied data driven and semi-empirical models in scour depth estimation for all three cases. Comparing the results, it may be stated that the GEP and FFNN methods produce more accurate results in the estimation of scour depth.

Comparisons of semi-empirical formulas and data driven model performances for the ski-jump bucket case (model: MB4), the sharp-crested wave case (model: MS5) and the inclined slope controlled structure case (model: M13) are shown in Figures 5, 6 and 7, respectively.

Table 6 | Sensitivity analyses of the best data driven models

Omitted variable	Model	CC	R ²	RMSE	MAE
Ski-jump bucket (MB4 model)					
q^2/gh_0^3	GEP	0.318	0.439	0.065	0.052
	FFNN	0.203	0.261	0.189	0.085
h_0/R	GEP	0.411	0.457	0.064	0.045
	FFNN	0.535	0.581	0.060	0.039
\emptyset	GEP	0.623	0.694	0.048	0.037
	FFNN	0.741	0.817	0.038	0.029
R/d_{50}	GEP	0.598	0.651	0.051	0.036
	FFNN	0.477	0.511	0.081	0.046
Sharp edge overflow (MS5 model)					
Fr_1	GEP	0.224	0.304	0.273	0.159
	FFNN	0.345	0.567	0.207	0.146
b/z	GEP	0.763	0.817	0.150	0.086
	FFNN	0.821	0.852	0.126	0.062
d_{50}/d_w	GEP	0.613	0.709	0.174	0.097
	FFNN	0.736	0.826	0.142	0.074
h/H	GEP	0.588	0.649	0.191	0.111
	FFNN	0.711	0.771	0.163	0.086
Fr_2	GEP	0.491	0.519	0.219	0.128
	FFNN	0.628	0.679	0.192	0.090
Inclined slope controlled structure (MI3 model)					
λ	GEP	0.411	0.498	0.282	0.246
	FFNN	0.398	0.439	0.187	0.095
Fr_2	GEP	0.122	0.232	0.338	0.290
	FFNN	0.233	0.394	0.302	0.243
D_p/h	GEP	0.321	0.413	0.298	0.232
	FFNN	0.285	0.382	0.306	0.237

Table 7 | Error statistics of the semi-empirical and data driven models of scour depth estimation

Model	CC	R ²	RMSE	MAE
<i>Ski-jump bucket</i>				
Veronese (1937)	0.393	0.425	0.221	0.155
Azamathulla (2012)	0.503	0.582	0.059	0.041
Sofrelec (1980)	0.399	0.457	0.233	0.172
Mason & Arumugam (1985)	0.413	0.506	0.142	0.096
GEP-MB4	0.841	0.858	0.034	0.027
FFNN-MB4	0.863	0.876	0.032	0.024
<i>Sharp edge over flow</i>				
Yen (1987)	0.612	0.661	0.576	0.337
D'Agostino & Ferro (2004)	0.598	0.658	0.356	0.184
GEP-MS5	0.814	0.874	0.112	0.079
FFNN-MS5	0.862	0.885	0.104	0.059
<i>Inclined slope controlled structure</i>				
Chee & Yuen (1985)	0.501	0.598	0.489	0.413
Bormann & Julien (1991)	0.557	0.632	0.717	0.570
GEP-MI3	0.803	0.769	0.198	0.164
FFNN-MI3	0.833	0.835	0.175	0.131

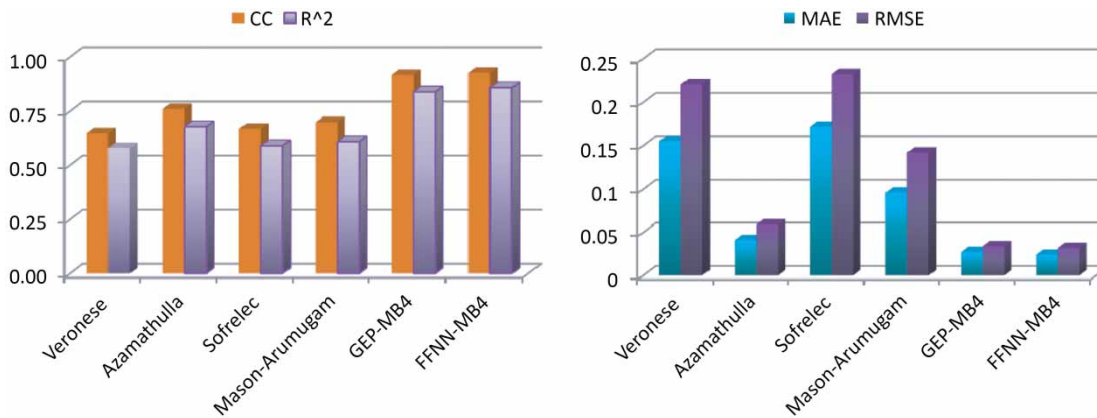


Figure 5 | Comparison of semi-empirical formulas and data driven model performances for the ski-jump bucket case (model: MB4).

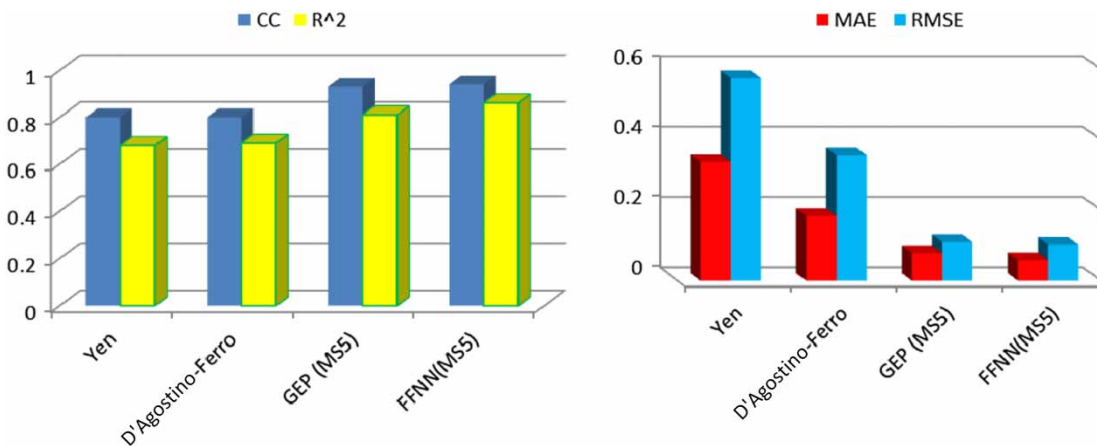


Figure 6 | Comparison of semi-empirical formulas and data driven model performances for the sharp-crested weir case (model: MS5).

CONCLUSIONS

The aim of this study was to assess the capabilities of data driven approaches in the estimation of scour depth downstream of grade control structures. Consequently, various GEP and FFNN models were developed and validated for three kinds of data, including the ski-jump bucket, sharp-crested weir and inclined slope controlled structures. These new techniques predicted scour depth downstream of grade control structures with reasonable accuracy from desirable data sets. The use of the GEP in the present study is another momentous part of scour depth estimation procedures for grade control structures. Since the GEP method can determine the most influential parameters on scour depth for all three structures, the use of a parsimonious model would result in higher accuracies than semi-empirical formulas.

The best semi-empirical formula R^2 values related to the ski-jump bucket (Azamathulla *et al.* 2005), sharp-crested weir (Yen 1987), and inclined slope controlled structures (Bormann & Julien 1991) are equal to 0.582, 0.661, and 0.632, respectively, which are less than the GEP and FFNN-related values for all of them. The obtained results revealed the superiority of FFNN and GEP techniques in predicting scour depth downstream of grade control structures.

The present study showed that GEP can be applied successfully to formulate scour depth downstream of grade control structures where interrelationships among the relevant variables are poorly understood.

Once the optimum model was identified for each data set, the most influential parameters affecting the scour depth were investigated through a sensitivity analysis. The results showed that for the ski-jump bucket, the

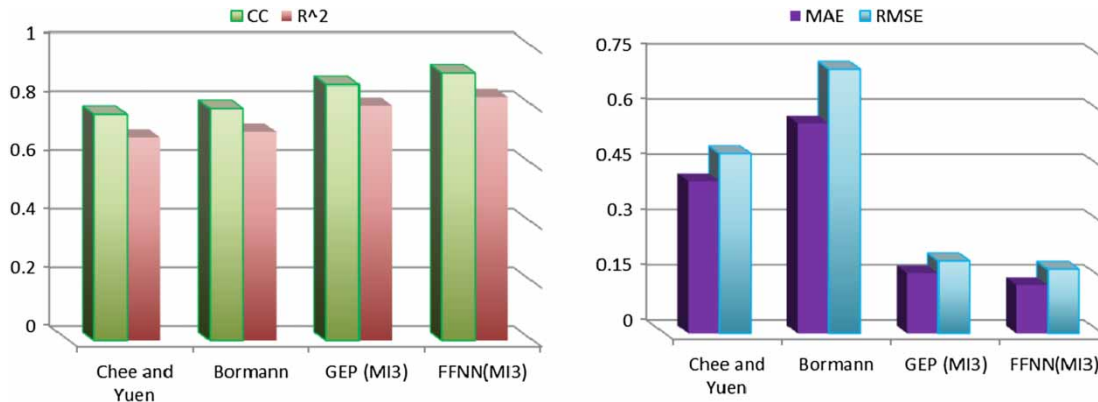


Figure 7 | Comparison of semi-empirical formulas and data driven model performances for the inclined slope controlled structure case (model: MI3).

optimum model includes four dimensionless parameters ($q^2/g h_0^3$, h_0/R , ϕ , R/d_{50}), from which $q^2/g h_0^3$ and ϕ showed the highest and least influence, respectively; for the sharp-crested weir, the $(Fr_1, d_{50}/d_w, h/H, b/z, Fr_2)$ parameters were found to be the most relevant factors on scour depth, where Fr_1 and b/z showed the highest and lowest effect on scouring; and for the inclined slope controlled structure, Fr_2 was the most influential parameter, while the λ showed the lowest effectiveness.

The results presented here are based on laboratory data, which have been captured from the existing literature. Further studies are needed to support the conclusions by using field data. Therefore, as the available data are limited for this issue, a comparison among different data management scenarios for feeding the data driven models would be valuable.

REFERENCES

- ASCE 2000 [Task Committee on application of artificial neural network in hydrology. Artificial neural network in hydrology. I: hydrology application. *Journal of Hydrologic Engineering* 5, 124–137.](#)
- Azamathulla, H. Md., Deo, M. C. & Deolalikar, P. B. 2005 [Neural networks for estimation of scour downstream of ski-jump bucket. *Journal of Hydraulic Engineering* 131, 898–908.](#)
- Azamathulla, H. Md. 2012 [Gene expression programming for prediction of scour depth downstream of sills. *Journal of Hydrology* 460–461, 156–159.](#)
- Azamathulla, H. Md. & Ghani, A. A. B. 2011 [ANFIS-based approach for predicting the scour depth at culvert outlets. *Journal of Pipelines System Engineering and Practice* 2, 35–40.](#)
- Azamathulla, H. Md. & Haque, A. A. M. 2012 [Prediction of scour depth at culvert outlets using gene expression programming. *International Journal of Innovative Computing, Information and Control \(ICIC International\)* 8, 5045–5054.](#)
- Azamathulla, H. Md. & Zakaria, N. A. 2007 [An ANFIS-based approach for predicting the scour below flip-bucket spillway. In: *2nd International Conference on Managing River in the 21st Century: Solution towards Sustainable River Basins, June 6–7, 2007, Riverside Kuching, Sarawak, Malaysia.*](#)
- Azamathulla, H. Md., Deo, M. C. & Deolalikar, P. B. 2005 [Neural networks for estimation of scour downstream of ski-jump bucket. *Journal of Hydraulic Engineering* 131, 898–908.](#)
- Azamathulla, H. Md., Deo, M. C. & Deolalikar, P. B. 2006 [Estimation of scour below spillways using neural networks. *Journal of Hydraulic Research* 44, 61–69.](#)
- Azamathulla, H. Md., Ghani, A. A. B., Zakaria, N. A., Lais, H., Chang, C. K., Leow, C. S. & Abuhasan, Z. A. 2008 [Genetic programming to predict ski-jump bucket spillway scour. *Journal of Hydrodynamics* 20, 447–484.](#)
- Azamathulla, H. Md., Ghani, A. A. B., Zakaria, N. A., Chang, C. K. & Abuhasan, Z. A. 2010 [Genetic programming approach to predict sediment concentration for Malaysian rivers. *International Journal of Ecological Economics and Statistics* 16, 53–64.](#)
- Babovic, V. 2000 [Data mining and knowledge discovery in sediment transport. *Computer-Aided Civil and Infrastructure Engineering* 15, 383–389.](#)
- Babovic, V. & Abbott, M. B. 1997 [The evolution of equations from hydraulic data Part II: applications. *Journal of Hydraulic Research* 35, 411–430.](#)
- Bertone, E., Stewart, R. A., Zhang, H. & Veal, C. 2015 [Data-driven recursive input-output multivariate statistical forecasting model: case of DO concentration prediction in Advancetown Lake, Australia. *Journal of Hydroinformatics* 17, 817–833.](#)
- Bonakdari, H. & Ebtehaj, I. 2015 [Comparison of data-driven approach in estimation of sediment transport in sewer pipe. In: *E-proceedings of the 36th IAHR World Congress, The Hague, The Netherlands.*](#)
- Bormann, N. & Julien, P. 1991 [Scour downstream of grade-control structures. *Journal of Hydraulic Engineering* 117, 579–594.](#)
- Bowden, G. J., Dandy, G. C. & Maier, H. R. 2005 [Input determination for neural network models in water resources](#)

- applications. Part 1. Background and methodology. *Journal of Hydrology* **301**, 75–92.
- Bureau of Indian Standards (BIS) 1985 *Criteria of Hydraulic Design of Bucket Type Energy Dissipaters*. BIS, 73651985, New Delhi, India.
- Chee, S. P. & Yuen, E. M. 1985 *Erosion of unconsolidated gravel beds*. *Canadian Journal of Civil Engineering* **12**, 559–566.
- D'Agostino, V. 1994 Indagine sullo scavo a valle di opera trasversale mediante modello di isico a fondomobile. *L'Energia Elettrica* **71**, 37–51 (in Italian).
- D'Agostino, V. & Ferro, V. 2004 *Scour on alluvial bed downstream of grade-control structures*. *Journal of Hydraulic Engineering* **130**, 24–37.
- Dargahi, B. 2003 *Scour development downstream of a spillway*. *Journal of Hydraulic Research* **41**, 417–426.
- Ferreira, C. 2001 *Gene expression programming: a new adaptive algorithm for solving problems*. *Complex Systems* **13**, 87–129.
- Guan, D., Melville, B. W. & Friedrich, H. 2014 *Flow patterns and turbulence structures in a scour hole downstream of a submerged weir*. *Journal of Hydraulic Engineering* **140**, 68–76.
- Güven, A. 2011 *A multi-output descriptive neural network for estimation of scour geometry downstream from hydraulic structure*. *Advances in Engineering Software* **42**, 85–93.
- Güven, A. & Günal, M. 2008 *Genetic programming approach for prediction of local scour downstream of hydraulic structures*. *Journal of Irrigation and Drainage Engineering* **134**, 241–249.
- Hamidi, N. & Kayaalp, N. 2008 *Estimation of the amount of suspended sediment in the Tigris River using artificial neural networks*. *Clean: Soil, Air, and Water* **36**, 380–391.
- Haykin, S. 1998 *Neural Networks: A Comprehensive Foundation*. Prentice Hall, New Jersey.
- Hornik, K. 1989 *Multilayer feedforward networks are universal approximators*. *Neural Networks* **2**, 359–366.
- Kisi, O. & Shiri, J. 2012 *River suspended sediment estimation by climatic variables implications: comparative study among soft computing techniques*. *Computers & Geosciences* **43**, 73–82.
- Kisi, O., Hosseinzadeh Dalir, A., Cimen, M. & Shiri, J. 2012 *Suspended sediment modeling using genetic programming and soft computing techniques*. *Journal of Hydrology* **450** (451), 48–58.
- Koza, J. R. 1992 *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press.
- Liao, K. W., Lu, H. J. & Wang, C. Y. 2012 *A probabilistic evaluation of pier-scour potential in the Gaoping River Basin of Taiwan*. *Journal of Civil Engineering and Management* **21**, 637–653.
- Lin, B. & Namin, M. M. 2005 *Modeling suspended sediment transport using an integrated numerical and ANNs model*. *Journal of Hydraulic Research* **43**, 302–310.
- Maier, H. R., Jain, A., Dandy, G. C. & Sudheer, K. P. 2010 *Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions*. *Environmental Modelling and Software* **25**, 891–909.
- Mason, P. J. & Arumugam, K. 1985 *Free jet scour below dams and flip buckets*. *Journal of Hydraulic Engineering* **111**, 220–235.
- Mossa, M. 1998 *Experimental study on the scour down-stream of grade-control structures*. In: *Proc., 26th Convegno di Idraulica e Costruzioni Idrauliche*, Catania, September, 3, 581–594.
- Nourani, V. & Andalib, G. 2015 *Daily and monthly suspended sediment load predictions using wavelet based artificial intelligence approaches*. *Journal of Mountain Science* **12**, 85–100.
- Nourani, V., Alami, M. T. & Aminfar, M. H. 2008 *A combined neural-wavelet model for prediction of watershed precipitation, Ligvanchai, Iran*. *Journal Environmental Hydrology* **16**, 1–12.
- Roushangar, K. & Alizadeh, F. 2015 *Studied suitability of different strategies in prediction of solid load discharge of an alluvial river*. In: *E-proceedings of the 36th IAHR World Congress*, The Hague, The Netherlands.
- Roushangar, K., Akhgar, S., Salmasi, F. & Shiri, J. 2014a *Modeling energy dissipation over stepped spillways using machine learning approaches*. *Journal of Hydrology* **508**, 254–265.
- Roushangar, K., Mehrabani, F. V. & Shiri, J. 2014b *Modeling river total bed material load discharge using artificial intelligence approaches (based on conceptual inputs)*. *Journal of Hydrology* **514**, 114–122.
- Roushangar, K., Mouaze, D. & Shiri, J. 2014c *Evaluation of genetic programming-based models for simulating friction factor in alluvial channels*. *Journal of Hydrology* **517**, 1154–1161.
- Shao, J. 1993 *Linear model selection by cross-validation*. *Journal of the American Statistical Association* **88**, 486–495.
- Sheikhalipour, Z. & Hassanpour, F. 2013 *Estimation of suspended sediment load using genetic expression programming*. *Journal of Civil Engineering and Urbanism* **3**, 292–299.
- Sofrelec, F. 1980 *Etude de factibilité du barrage de Kandadji*. p. 56.
- Tsukamoto, Y. 1979 *An approach to fuzzy reasoning method*. In: *Advances in Fuzzy Set Theory and Application* (M. M. Gupta, R. K. Ragade & R. R. Yager, eds). North, Amsterdam, pp. 137–149.
- Veronese, A. 1937 *Erosion di fondo a valle di un scarico*. *Annali Lavori Pubbl.* **75**, 717–726 (in Italian).
- Yang, C. T., Marsooli, R. & Aalami, M. T. 2009 *Evaluation of total load sediment transport using ANN*. *International Journal of Sediment Research* **24**, 274–286.
- Yen, C. L. 1987 *Discussion on free scours below dam, flip bucket, by Mason Peter J. & Arumugam Kanapathypilly*. *Journal of Hydraulic Engineering* **113**, 1200–1202.
- Yuhong, Z. & Wenxin, H. 2009 *Application of artificial neural network to predict the friction factor of open channel flow*. *Communications in Nonlinear Science and Numerical Simulation* **14**, 2373–2378.

First received 22 December 2015; accepted in revised form 12 May 2016. Available online 15 July 2016