Gene-Expression Programming, Evolutionary Polynomial Regression, and Model Tree to Evaluate Local Scour Depth at Culvert Outlets

Mohammad Najafzadeh[1] and Ali Reza Kargar[2]

Abstract: Protection of the downstream of culvert outlets against scour process, as a water conveyance structure, is a highly significant issue in design of culverts. Frequent field and experimental investigations were carried out to produce a relationship between the scour depth due to the governing variables. However, existing empirical equations do not always provide a precise estimation of the scour depth due to the complexity of the scour phenomena. In this investigation, gene-expression programming (GEP), model tree (MT), and evolutionary polynomial regression (EPR) are utilized to predict the scour depth downstream of culvert outlets. Input variables—considering effective parameters on the scour depth—were defined as sediment size at downstream, geometry of culvert outlets, and flow characteristics in upstream and downstream. Experimental datasets to develop the models were collected from different literature. Performances of the proposed models for the training and testing phases were assessed using several statistical measures. Results of performances indicated that EPR provided the lowest level of precision including index of agreement (IOA = 0.958) and root mean squared error (RMSE = 0.419) for prediction of local scour depth at culvert outlets than those obtained using MT (IOA = 0.947 and RMSE = 0.471) and GEP (IOA = 0.943 and RMSE = 0.487). In terms of accuracy, all proposed equations extracted from artificial intelligence approaches had remarkable superiority to the traditional equations. Ultimately, it has been proven that mathematical expressions given by evolutionary computing tools had sufficient generalization to present an accurate prediction of the local scour depth with respect to preserving physical meaning of results. **DOI: 10.1061/(ASCE)PS.1949-1204.0000376.** *© 2019 American Society of Civil Engineers.*

Author keywords: Gene-expression programming; Evolutionary polynomial regression; Model tree; Culvert outlet; Scour depth; Statistical measures.

Introduction

Culvert outlets are hydraulic structures that control flow. Occasionally, culvert outlets are utilized for management of excessive run off. One of the most remarkable considerations for designing of culvert outlets is estimation of scour depth at outlets. Scouring process at outlets can occur due to several shortcomings in its designing. Scouring at outlets for both free and submerged flow is the most significant factor leading to the irreversible damages (e.g., Abt et al. 1985, 1987; Lim 1995). In this way, extensive laboratory investigations were carried out to identify effective parameters on the scour depth at culvert outlets. A large number of experimental research works were reported from several laboratory works. In fact, several empirical equations were extracted from experimental datasets on the basis of restricted ranges of observed data (Laushey et al. 1967; Opie 1967; Bohan 1970; Abt et al. 1984, 1985, 1987; Abida and Townsend 1991; Lim 1995; Liriano et al. 2002). In the case of traditional equations, performances of experiments indicated that there is no a general equation to provide a precise estimation of scour depth at culvert outlets which can be validated for a board range of experimental datasets.

In the recent decade, some artificial intelligence (AI) approaches such as artificial neural networks (ANNs), adaptive neuro-fuzzy inference system (ANFIS), genetic programming (GP), gene-expression programming (GEP), group method of data handling (GMDH), and support vector machine (SVM) have been applied to predict the local scour depth at various hydraulic structures such as downstream of grade-control structures, ski-jump bucket spillways, bridge abutments, bridge piers, culvert outlets, downstream of sill structure, and below offshore structures (e.g., Azmathullah et al. 2005; Guven and Gunal 2008a, b; Azamathulla and Ghani 2010a; Ghazanfari-Hashemi et al. 2011; Azamathulla 2012a, b; Guven and Azamathulla 2012; Najafzadeh and Sarkamaryan 2018; Ebtehaj et al. 2018). In the case of scour depth prediction at culvert outlets, it should be noted that a large number of studies conducted by AI approaches were a suitable platform in order to reach the scour depth prediction with permissible level of accuracy rather than empirical equations (Liriano and Day 2001; Azamathulla and Ghani 2010b; Azamathulla and Haque 2012, 2013; Najafzadeh 2015). Among mentioned AI models, GP, GEP, and GMDH approaches have the capability of describing a relationship among input and output variables for different realms of scouring problems. On the contrast, the most primary concern related to the ANNs and ANFIS approaches is that these models have not only a black box nature but also relatively voluminous computation (e.g., Azamathulla and Ghani 2010a, b; Azamathulla and Haque 2012, 2013; Najafzadeh 2015).

Recently, evolutionary polynomial regression (EPR) and model tree (MT) techniques extracted an input-output model based best

[1]Assistant Professor, Dept. of Water Engineering, Faculty of Civil and Surveying Engineering, Graduate Univ. of Advanced Technology, P.O. Box 76315116, Kerman 7631133191, Iran (corresponding author). Email: moha.najafzadeh@gmail.com; m.najafzadeh@kgut.ac.ir

[2]Assistant Professor, Dept. of Rock Mechanics, School of Mining Engineering, College of Engineering, Tehran Univ., P.O. Box 1439957131, Tehran 1417466191, Iran. Email: ar.kargar@ut.ac.ir

Note. This manuscript was submitted on January 2, 2018; approved on October 9, 2018; published online on March 26, 2019. Discussion period open until August 26, 2019; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Pipeline Systems Engineering and Practice***, © ASCE, ISSN 1949-1190.

formulations to obtain physical meaning of governing parameters in the scour depth prediction problems. In fact, these two AI models were used to understand scouring process at hydraulic structures, as evaluation of group piers scour depth under waves and currents conditions (Etemad-Shahidi and Ghaemi 2011; Ghaemi et al. 2013), geometry of scour hole downstream of spillway (Pal et al. 2012; Samadi et al. 2014), prediction of scour depth downstream of grade-control structures and sluice gates (Laucelli and Giustolisi 2011; Najafzadeh et al. 2018b), bed of rectangular channels (Najafzadeh et al. 2018a), and below pipelines induced currents (Najafzadeh and Sarkamaryan 2018). Furthermore, EPR and MT technique based explicit formulations have not been applied yet in the prediction of the local scour depth at culvert outlets.

In this study, EPR, MT, and GEP based formulations were employed to predict the scour depth at culvert outlets. These AI techniques were developed using experimental datasets. Performance of the proposed models were investigated in both training and testing stages by means of several statistical measures. Results of this study were compared with those obtained using empirical equations-based regression approaches. Additionally, parametric study of the AI techniques was carried out to perceive physical meaning of results. Ultimately, Fisher test was conducted to select AI model with the best performance.

Review of Scouring Investigations at Culvert Outlets

Investigations to determine the scour depth at culvert outlets have been carried out widely in different experimental conditions. Laushey et al. (1967) and Opie (1967) initiated experiments set up for culvert outlet with pipe cross-section and uniform bed sediments. In addition, a large number of experiments were performed to find physical meaning different governing variables on local scour depth at culvert outlets (e.g., Laushey et al. 1967; Opie 1967; Bohan 1970; Ruff et al. 1982; Abt et al. 1984, 1985, 1987; Abida and Townsend 1991; Lim 1995; Aderibigbe and Rajaratnam 1998; Day et al. 2001; Liriano et al. 2002).

Ruff et al. (1982) conducted a large number of experiments to investigate effects of flow conditions, bed material, geometry of pipe culvert on the scour hole geometry.

They concluded that the maximum scour depth downstream of culvert outlets was located approximately between 0.3 and 0.4 of the maximum scour length. Abt et al. (1984) performed several experiments to evaluate effects five non-cohesive bed materials on the geometry of hole scour at culvert outlets. From his experiments, he proposed a design curves for estimation of scour hole dimensions for practical uses. Additionally, scour hole dimensions was correlated to the discharge intensity. A simple expression-based formulation was given in which dependency of the scour depth hole on a variety of non-cohesive materials based upon various discharge values, culvert diameter, mean grain diameter and material standard deviation was as follows:

$$\frac{d_s}{d_0} = 3.67 Fr_d^{0.57} \left(\frac{d_{50}}{d_0}\right)^{0.4} \sigma_g^{-0.4} \tag{1}$$

where d_s = local scour depth at culvert outlet; d_0 = pipe diameter for circular outlets and outlet height for non-circular outlets; d_{50} = median sediment size; σ_g = geometric standard deviation of bed sediments; and Fr_d = densimetric Froude number. Fr_d is computed as

$$Fr_d = \frac{U_0}{\sqrt{g(\frac{\rho_s}{\rho} - 1)d_{50}}} \tag{2}$$

where ρ = mass density of water; U_0 = mean velocity at outlet; ρ_s = mass density of bed material; and g = acceleration due to gravity.

Abt et al. (1985) carried out experiments to investigate slope effects on the dimensions of the scour hole at downstream of culvert outlets. Culvert slope was varied between 0 and 10%. Results indicated that increasing in culvert slope increase the scour depth. Also, the maximum scour hole width decreases with increasing of slope higher than 5%. Scour length decreased with increase in slope from 0 to 5%. Abt et al. (1987) performed experiments with various culvert shape including arch, square, and rectangular cross-sections. They concluded that scour length observed from square and rectangular cross section is 40% more than that obtained using circular culvert outlets. Blaisdell and Anderson (1988) carried out a comprehensive investigation of scour at downstream of cantilevered culvert outlets. From their studies, they used the computed ultimate maximum scour hole depth to compute separate contour maps for discharges covering a wide range of anticipated discharges.

Abida and Townsend (1991) have set up experiments for local scour at downstream of box-culvert outlets. From their observations, for condition of very shallow tail water, maximum scour depth decrease with decreasing of tail-water depth. Also, Lim (1995) performed experiments for scour at un-submerged and full flowing culvert outlets. Analysis of the observed data indicated that local scour depth was evaluated as a function of densimetric Froude number (Fr_d) as follow:

$$\frac{d_s}{d_0} = 0.45 Fr_d \tag{3}$$

Furthermore, Aderibigbe and Rajaratnam (1998) used dune armored bed for reduction of maximum scour depth at downstream of culvert outlets. They observed that the average reduction of the scour depth was approximately 60%.

Day et al. (2001) investigated tail water depth and model scale on scouring process downstream of culvert outlets. They used pipe culverts with diameters (d_0) 0.013, 0.020, 0.052,0.146, and 0.311 m. Also, sediment gradations with d_{50} of 0.38, 0.59, 1.4, 4.4, and 7.9 mm were applied for experiments setup. For ratio of tail-water depth (H) to the culvert diameter (d_0) between 0.5 and 2, following equation-based regression was obtained:

$$\frac{d_s}{d_0} = 0.877 \left(\frac{H}{d_0}\right)^{-0.37} Ln(Fr_d) + 0.2 Ln\left(\frac{H}{d_0}\right) - 0.24 \tag{4}$$

Liriano et al. (2002) studied effects of turbulent flow structures on the scour at culvert outlets. They analyzed the main velocities, turbulence intensities, Reynolds stresses, and near-bed bursting structure. Also, experimental studies indicated that initial formation of the scour hole is related to the mean velocity exceeding the critical velocity.

Furthermore, a large number of experimental investigations have been reported by Federal Highway Administration (FHWA). They proposed several empirical equations for validation of observed data sets. Result of their works indicated that laboratory datasets have substantial restrictions which are related to the conditions of flow and bed materials (e.g., Kerenyi et al. 2003, 2007).

Azamathulla and Haque (2012) employed GEP model on the basis of evolutionary computing for prediction of the local scour depth at culvert outlets. They proposed following equation for this purpose as

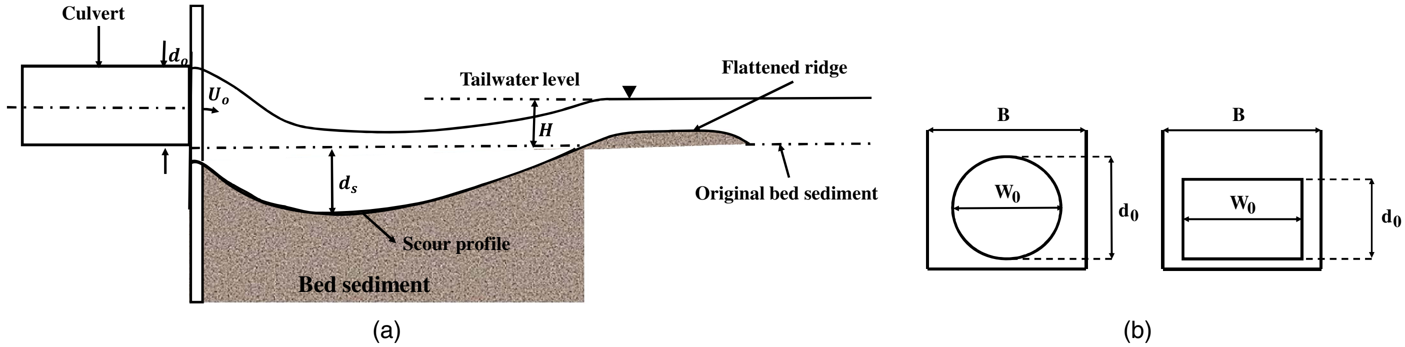

Fig. 1. Scour process at culvert outlet: (a) definition of hydraulic parametrers physical sediment properties; and (b) configuration of geometric parameters of and B, d_0, and W_0.

$$\frac{d_s}{d_0} = \left[\frac{-6.62 + Fr_d}{\left(\sqrt{Fr_d}/\frac{H}{d_0}\right) + \frac{9.65}{\sigma_g}}\right] + \left[\frac{d_{50}}{d_0} - \frac{H/d_0}{e^{\left[2.34+\left(\frac{d_{50}}{d_0}\right)\right]} + e^{\sigma_g}}\right] + \left(Fr_d\sqrt{\sigma_g \cdot \frac{H}{d_0}}\right)^{0.5} \quad (5)$$

Data Presentation for Scour Depth Modeling

Prior to laboratory works for investigating the effective parameters on the scour depth at culvert outlets, the scour depth depends on geometry of outlets cross-section, various flow condition at upstream and downstream, and physical characteristics of bed materials at outlet of culvert (e.g., Opie 1967; Abt et al. 1987; Abida and Townsend 1991; Lim 1995; Liriano et al. 2002). Therefore, the following function is proposed for scour depth prediction:

$$d_s = f(\rho, \mu, U_0, d_0, B, H, W_0, \rho_s, g, d_{50}, \sigma_g) \quad (6)$$

where μ, d_0, H, B, and W_0 are the dynamic viscosity of water, the pipe diameter for circular outlets and outlet height for non-circular outlets, water depth at downstream of receiving channel, channel width, and outlet width, respectively. To visualize the effective parameters on the scour depth at outlet of culvert for unsubmerged flow conditions, general feature of scouring process was sketched in Fig. 1(a). Moreover, in Fig. 1(b), d_0, B, and W_0 were specified for both circular and box culverts.

On the basis of Buckingham π-theorem and choosing the ρ, U_0, and d_0 as repeating variables, a set of nine non-dimensional parameters was resulted by performance of dimensional analysis:

$$\pi_9 = f_1(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6, \pi_7, \pi_8) \quad (7)$$

in which $\pi_1 = B/d_0$, $\pi_2 = H/d_0$, $\pi_3 = W_0/d_0$, $\pi_4 = d_{50}/d_0$, $\pi_5 = \rho_s/\rho$, $\pi_6 = \sigma_g$, $\pi_7 = g \cdot d_0/U_0^2$, $\pi_8 = \rho \cdot U_0 \cdot d_0/\mu$, and $\pi_9 = d_s/d_0$. In addition, the fifth dimensionless parameter is introduced as specific gravity G_s. Axiomatically, performance of algebraic operations on dimensionless parameters always produces other non-dimensional parameters. With getting inspiration of previous investigations, it was found that densimetric Froude number (Fr_{d}) was applied in estimation of the local scour depth at culvert outlet instead of normal Froude number. Hence, with the aid of fourth, fifth, and seventh dimensionless parameters, densimetric Froude number can be calculated as

$$\pi' = \left(\frac{1}{\pi_7} \cdot \frac{1}{\pi_4} \cdot \frac{1}{\pi_5 - 1}\right)^{0.5} = \frac{U_0}{\sqrt{g \cdot d_{50} \cdot (G_s - 1)}} \quad (8)$$

With respect to Eqs. (7) and (8), Eq. (6) was rewritten as

$$d_s/d_0 = f_2(B/d_0, Fr_d, H/d_0, W_0/d_0, d_{50}/d_0, \sigma_g, \rho \cdot U_0 \cdot d_0/\mu) \quad (9)$$

where $\rho \cdot U_0 \cdot d_0/\mu$ = Reynolds number for flow passing through the culvert.

Due to previous experimental results, the G_s ratio can was neglected in the Eq. (9) because it varies in a very restricted range from a practical point of view. Furthermore, effects of Reynolds number of pipe culvert ($\mathrm{Re_p}$) on the local scour depth at culvert outlets has remained meaningless (e.g., Abt et al. 1984; Abida and Townsend 1991; Lim 1995; Liriano et al. 2002). Therefore, $\mathrm{Re_p}$ is removed from Eq. (9). In this way, Eq. (9) can be re-written as

$$d_s/d_0 = f_2(B/d_0, Fr_d, H/d_0, W_0/d_0, d_{50}/d_0, \sigma_g) \quad (10)$$

Table 1. Range of datasets used for development of the proposed models

References	Parameters						
	Fr_d	H/d_0	W_0/d_0	B/d_0	d_{50}/d_0	σ_g	d_s/d_0
Laushey et al. (1967)	1.04–3.37	0.5	1	5.6–9.4	0.16–0.4	1.3	0.42–1.16
Bohan (1970)	12.97–88.61	1	0.3–0.75	16–71	0.0007–0.0032	1.37	2.12–5.22
Rajaratnam and Diebel (1981)	4.45–37.1	0.196–3.385	0.75–1	1–86	0.0022–0.0827	1.27–1.37	0.66–6.53
Ruff et al. (1982)	5.53–13.47	0–0.45	1	11.93–61	0.0054–0.03	1.3–4.78	0.579–3.46
Abt et al. (1987)	7.13–23.13	0.45	1–2	61	0.02	1.33	2.03–4.06
Abida and Townsend (1991)	0.59–3.81	0.05–1.55	0.67–3.34	6.578	0.006–0.017	1.22–2.24	0.19–2.4
Lim (1995)	1.91–24.6	31.7–55	15–26	38.46–66.67	0.06–0.11	1.25	0.81–4.87
Day et al. (2001)	0.05–0.94	0.5–2	1	7.4	0–0.03	1.4	0.18–1.89

J. Pipeline Syst. Eng. Pract., 2019, 10(3): 04019013

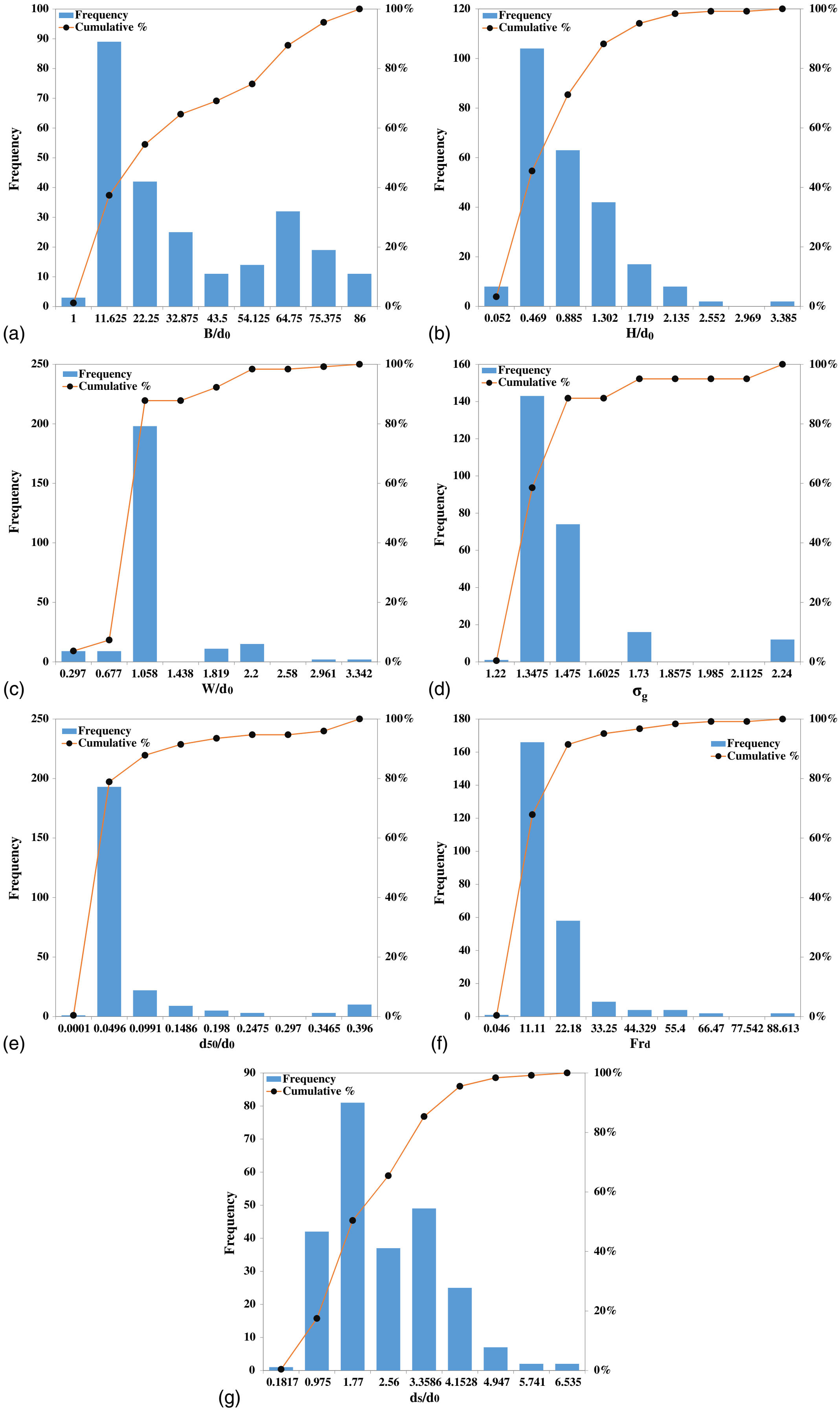

Fig. 2. Histograms of variables used in the models' implementation: (a) B/d_0; (b) H/d_0; (c) W_0/d_0; (d) σ_g; (e) d_{50}/d_0; (f) Fr_d; and (g) d_s/d_0.

In the case of scour depth estimation at culvert outlets, previous investigations have established that use of dimensionless parameters had the capability to provide better performance of AI models in comparison with that of dimensional parameters (e.g., Azamathulla and Ghani 2010b; Azamathulla and Haque 2012, 2013; Najafzadeh 2015). In this study, Eq. (10) was applied to develop the proposed models for prediction of scour depth at culvert outlets. The datasets used consisted of 246 samples collected from literature (Laushey et al. 1967; Bohan 1970; Rajaratnam and Diebel 1981; Ruff et al. 1982; Abt et al. 1987; Abida and Townsend 1991; Lim 1995; Day et al. 2001).

Table 1 presents ranges of input-output parameters for the scour depth prediction. Out of the dataset, about 80% (197 datasets) and 20% (69 datasets) were dedicated randomly to perform training and testing stages, respectively. Additionally, to conceptualize distribution pertained to the non-dimensional parameters, experimental datasets have been demonstrated by means of frequency histograms in Fig. 2.

Frameworks of Models Implementation

In this section, descriptions of the GEP, MT, and EPR modeling approaches were succinctly presented. And the way of developing the proposed AI methods are carried out to extract the best relationships for prediction of scour depth at culvert outlets.

Development of GEP Model

GEP is a relatively contemporary extension of the GP approach. Basically, general structure of GEP is programmed in terms of mathematical expressions. The GEP approach is coded in forms of linear chromosomes, being then expressed into Expression Trees (ETs) (e.g., Azamathulla and Haque 2012; Ferreira 2006, 2001).

In fact, the ETs are computer programming with high level of knowledge extraction which are occasionally employed to present an efficient solution for a practical problem. Additionally, ETs are opted on the basis of their fitness values obtained through solution of practical problem. From every ET, a mathematical expression is released in a way that summation of all equations is equal to general solution of GEP model. Through ETs, there is a population which will discover traits, and thus will adapt to the particular problems. This is indicative of having sufficient time and acceptable values for setting parameters of GEP, and consequently accurate predictions extracted by GEP model would be remained unmasked.

Development of the GEP approach includes five steps. The first step is to select the fitness function, f_i, of an individual program (i). This function is evaluated as follows:

$$f_i = \sum_{j=1}^{C_t} (M - |C_{(i,j)} - T_j|) \tag{11}$$

in which M, $C_{(i,j)}$, and T_j are the selection range, value returned by the individual chromosome i for fitness case j, the largest value for fitness case j.

In the second stage, the set of terminals T and the set of function F were selected to generate the chromosomes. In the present study, the terminals includes a set of six independent variables, as noted in Eq. (11), in form of $T(d_s/d_0) = \{B/d_0, H/d_0, W_0/d_0, d_{50}/d_0, Fr_d, \sigma_g\}$.

In the case of finding the best function set, it is essential to consider general mathematical shape of empirical equations given in literature. As mentioned in introduction section, four basic operators (+,–,*,/) and basic mathematical functions ($\sqrt{}$, power, exp) were applied in empirical equations given in literature (Abt et al. 1984; Lim 1995; Day et al. 2001) so as to predict the local scour at culvert outlets. In this way, development of GEP was carried out on the basis of these considerations due to obtaining physical meaning of GEP results. The third step is to configure the chromosomal architecture. The fourth step is selection of liking function. Ultimately, for the fifth stage, a set of genetic operators such as mutation, inversion, one (or two)-point recombination, gene recombination, and gene transposition are inevitably obtained which each operator has a specific rate. The other in-depth information about structure of GEP technique were detailed in the literature (Ferreira 2001, 2006).

Table 2. Characterization of the proposed GEP model

Parameter	Description	Setting
P_1	Function set	+, −, ×, /, exp, power, $\sqrt{}$
P_2	Mutation rate	0.00138
P_3	Inversion rate	0.00546
P_4	One point and two-point recombination rate respectively (%)	0.277
P_5	Gene recombination rate	0.00277
P_6	Gene transportation rate	0.00277
P_7	Maximum tree depth	5
P_8	Number of gene	3
P_9	Number of chromosomes	30
P10	Number of generation	2,000

Through development of GEP for the local scour depth at culvert outlets, the functional set and the operational parameters applied in this study were presented in Table 2. The best formulation of GEP model, as a function of input-output variables, was obtained as

$$\frac{d_s}{d_0} = \sqrt{\sqrt{2(Fr_d)} + \sqrt{\left(\frac{d_{50}}{d_0} + Fr_d\right) \cdot \frac{H}{d_0} \cdot \frac{d_{50}}{d_0}}} + Fr_d - \left(Fr_d + \frac{d_{50}}{d_0}\right) \cdot (6.283 - Fr_d) \tag{12}$$

In addition, the expression trees related to the above formulation was illustrated in Figs. 3(a–c). In Fig. 3(c), constant values demonstrated in the third sub-tree is 6.283 and the actual variables are the H/d_0, d_{50}/d_0, and Fr_d. From Eq. (12), it can be found that many input variables including σ_g, B/d_0, and W_0/d_0 have no influence on the scour depth at culvert outlets. This finding was in good agreement with results of experimental datasets reported by Day et al.(2001). On the other hand, as seen in Eq. (4) and H/d_0, d_{50}/d_0, and Fr_d variables could play a crucial role in estimation of scour depth at culvert outlets.

Development of MT Model

Among the data mining techniques, MT, as a machine learning classifier, has the capability in order to present an efficient solution to the practical problem by dividing it into several sub-problems (sub-domains) which are individually distinctive from each other and consequently the ultimate results of MT for a primary domain is a combination of these sub-domains. In fact, MT technique is capable of performing a multivariable regression model for input-output variables trapped in each sub-domain (Quinlan 1992; Wang and Witten 1997; Etemad-shahidi and Ghaemi 2011; Pal et al. 2012). As a considerable advantage, MT approach has the ability to present solution to the class of continuous problems by means of piecewise linear techniques. In this way, nonlinear complicated

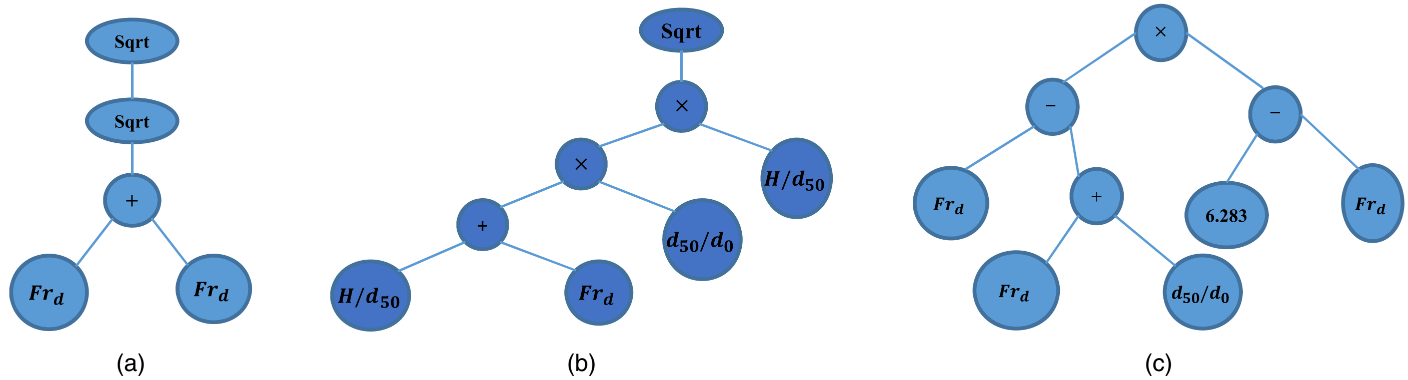

Fig. 3. Optimal expression tree (ET) structures extracted from GEP performance: (a) sub-ET1; (b) sub-ET2; and (c) sub-ET3.

models can be converted to several linear simple and easy-to-use models with an acceptable precision level of approximation. Development of MT has two main steps: (1) creating the tree structures; and (2) extracting knowledge from it. To obtain more perception of MT performance, a schematic diagram for the tree structures-building approach within nine linear regression models in Fig. 4. In fact, knowledge extraction from the MT structure related to each sub-domains was illustrated in Fig. 4(a) and additionally a general tree structure of MT approach was sketched in Fig. 4(b). According to the Figs. 4(a and b), the data samples illustrated on X_1 axis was divided into two main segments, including $X_1 \leq 5$ and $X_1 > 5$. If X_1 is smaller than (or is equal to) 5, four linear models are created. In this way, If $X_2 \leq 3$, then Model 1 is generated. Otherwise three linear models are appeared in a way that If $X_2 \leq 7$ then Model 4 is obtained. If $X_1 \leq 2$ then Model 9 is produced. Otherwise, Model 7 is created. Moreover, there is similar trend in order to create the rest of linear models for range of $X_1 > 5$ in Fig. 4.

Based on the domain-splitting criterion, various efficient approaches such as M5 model was frequently employed to develop MT technique. M5 approach firstly generates a regression tree by means of splitting the instance search space within a particular process characterized by recurrence. This process is performed to minimize variations of the intra-subset in values (or quantities) from the root to the node and through the branch (Wang and Witten 1997). This variation is evaluated by means of the standard, deviation of the values which stretch out through the branch from the root to the node, being carried out by computing the desired reduction in error values from testing every variable (or attribute) at the node. The input variable that maximum level of desired error reduction is opted. According to the Quinlan (1992) investigations, this process will be continued until the standard deviation reduction becomes less than a certain percent of the standard deviation of the original dataset or when only a small quantity of data samples (just below 5%) remain. Through MT approach, standard deviation reduction (SDR) factor:

$$SDR = sd(E) - \sum_i \frac{|E_i|}{|E|} sd(E_i) \tag{13}$$

in which E, E_i, and sd are the set of samples that reach the node, the set that results of splitting the node on the basis of the chosen variable, and standard deviation, respectively. The M5 utilizes

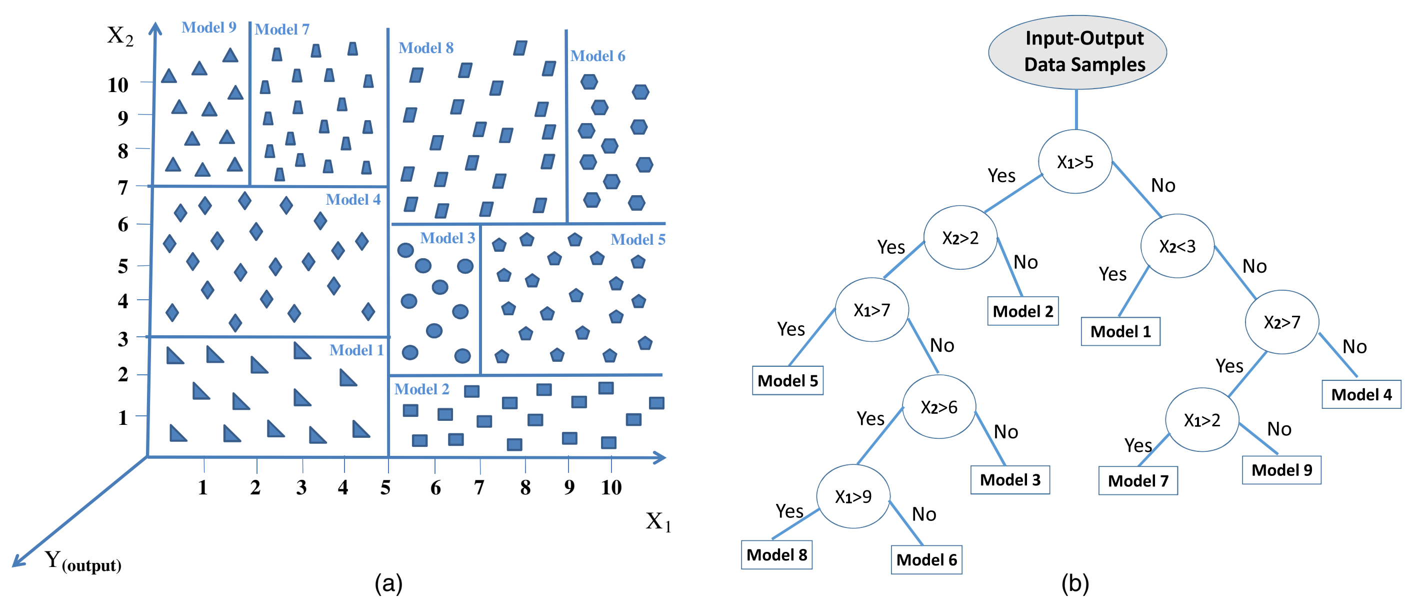

Fig. 4. Splitting the input space and prediction by MT for a input-output dataset: (a) splitting of the input space ($X_1 \times X_2$) into nine subspaces using M5 model tree; and (b) schematic diagram of creating nine linear models using a set of IF-THEN rules.

Rule:1
IF

$$Fr_d > 6.99$$

THEN

$$\frac{d_s}{d_0} = 0.7406 + 0.0083\frac{B}{d_0} + 0.485\frac{H}{d_0} + 0.7168\frac{W_0}{d_0} + 9.628\frac{d_{50}}{d_0} + 0.0348Fr_d$$

Rule: 2
IF

$$Fr_d \leq 2.437$$

$$\sigma_g > 1.35$$

THEN

$$\frac{d_s}{d_0} = -5.1079 + 0.824\frac{B}{d_0} - 0.098\frac{H}{d_0} - 0.0299\sigma_g - 0.748\frac{d_{50}}{d_0} + 0.522Fr_d$$

Otherwise,
Rule: 3

$$\frac{d_s}{d_0} = 0.726 - 0.4079\frac{H}{d_0} - 0.223\sigma_g + 0.2648Fr_d$$

Fig. 5. Linear equations extracted from MT.

the *sd* parameter as an error criterion of the class values that reach a node. Generally, as tree structure grew, linear model is created for every inner node. This model is generated on the basis of values corresponded to that node and consequently all the applied test variables in the sub-tree emanated from that node. Other processes of M5 including pruning and smoothing the trees were detailed in literature (e.g., Etemad-Shahidi and Ghaemi 2011; Pal et al. 2012).

In the current research, the proposed MT approach has six non-dimensional inputs and one output parameter. MT technique was developed using three rules in form of linear multivariable equations for the local scour depth evaluation. These three linear equations were given Fig. 5. According to Fig. 5, Fr_d and σ_g were assigned as splitting parameters to create linear models. In the first rule given by MT, corresponding linear model had no dependency on σ_g variable. The second linear model indicated that variations of d_s/d_0 versus W_0/d_0 has remained meaningless and additionally W_0/d_0, B/d_0, and d_{50}/d_0 have not been included in the third linear model. Furthermore, the second linear equation should be simultaneously valid for two ranges of $Fr_d \leq 5.492$ and $Fr_d \leq 2.437$. For instance, $Fr_d = 4$ is not valid $Fr_d \leq 2.437$ and consequently it cannot be satisfied in Rule 2. Overall, from Fig. 5, it can be said that all linear equations are easy-to-use with simple mathematical expressions in order to predict the local scour depth at culvert outlets.

Development of EPR Model

EPR is one of the data-driven models which has the capability to present symbolic relationships in order to characterize complicated systems. EPR works on the basis of a global search approach. In fact, global search applied in the EPR technique is genetic algorithm (GA) (Savic et al. 2006; Giustolisi and Savic 2006). The general mathematical expression given by EPR are composed of several additive terms multiplied by as many coefficients as

$$\hat{Y} = a_0 + \sum_{j=1}^{m} a_j \cdot (X_1)^{\mathbf{ES}(j,1)} \cdot \ldots \cdot (X_k)^{\mathbf{ES}(j,k)} \cdot f\left((X_1)^{\mathbf{ES}(j,k+1)} \cdot \ldots \cdot (X_k)^{\mathbf{ES}(j,2k)}\right) \quad (14)$$

where m is the maximum number of mathematical terms, a_0 is the bias term, a_j is a set of coefficients for mathematical relationship, $\mathbf{X}_i$ are the input variables (or independent candidate) for EPR model, $\hat{\mathbf{Y}}$ is the output variable predicted by EPR model, k is the number of input variables, function f whose general mathematical structure is selected by user, and **ES** is a set of user-specified exponents (Giustolisi and Savic 2006, 2009; Savic et al. 2009; Laucelli and Giustolisi 2011; Laucelli et al. 2012).

In the case of EPR development, three basic objective functions are required to be considered. In this way, EPR searches m-dimensional formulations by means of two or three objectives as (1) the maximization of model accuracy; (2) the minimization of number of model coefficients; and (3) the minimization of the number of input variables applied in the EPR modeling. In fact, multi-objective genetic algorithm (MOGA) has been employed in the EPR structure to find optimal mathematical relationship (e.g., Savic et al. 2009; Altomare et al. 2013). EPR-MOGA is performed in two various media. In the first place, setting parameters of EPR such as type of objective functions, number of mathematical terms, range of exponents, and type of mathematical formulation used in the EPR modeling are defined in Excel software. Secondly, all the coefficients of mathematical relationship related to the EPR technique are optimized by means of GA coded in MATLAB. Several mathematical relationships were obtained within training stage and ultimately the best equation with the highest level of accuracy when observed values (output variable) are compared with values predicted by EPR model.

In the current study, to develop EPR mode for prediction of the local scour depth at culvert outlets, W_0/d_0, B/d_0, d_{50}/d_0, H/d_0, Fr_d and σ_g were considered as input variables. Furthermore, the range of exponents **EX** is [−2; −1.5; −1; −0.5; 0; 0.5; 1; 1.5; 2]; the maximum number of mathematical terms (m) is equal to 5. All the non-negative coefficients a_j were allowed to be considered in equations given by EPR model with existence of bias a_0. To obtain the optimum equation by EPR, values of setting parameters were given in Table 3.

With respect to defaults of proposed EPR model, the most precise equation was obtained as

$$\begin{aligned}\frac{d_s}{d_0} = &+1.0793\frac{(Fr_d)^{0.5}}{\sigma_g}\exp\left(-0.5\sigma_g - 2\frac{d_{50}}{d_0}\right) + 0.54262\sigma_g^{0.5}\exp\left(-2\frac{d_{50}}{d_0}\right)\\ &+ 5.4148\times10^{-5}\left(\frac{B}{d_0}\right)^2\left(\frac{W_0}{d_0}\right)^{0.5}\sigma_g^2(Fr_d)^{0.5}\exp\left(-2\frac{H}{d_0} + 0.5\frac{W_0}{d_0} - 0.5\sigma_g - 2\frac{d_{50}}{d_0}\right)\\ &+ 5.1069\times10^{-6}\left(\frac{B}{d_0}\right)^2\left(\frac{H}{d_0}\right)^{1.5}\left(\frac{W_0}{d_0}\right)^2\sigma_g^2\left(\frac{d_{50}}{d_0}\right)^{1.5}(Fr_d)^2\exp\left(+0.5\frac{W_0}{d_0} - 1.5\frac{d_{50}}{d_0}\right)\\ &+ 0.0028595\left(\frac{B}{d_0}\right)^2\left(\frac{H}{d_0}\right)^2\left(\frac{W_0}{d_0}\right)^{1.5}\sigma_g^2\left(\frac{d_{50}}{d_0}\right)(Fr_d)\exp\left(-2\frac{H}{d_0} - 0.5\sigma_g - 2\frac{d_{50}}{d_0}\right) + 0.15872\end{aligned} \quad (15)$$

Results and Discussion

The results of proposed models approach and empirical equations have been given in this section. For evaluation of statistical analysis performances, index of agreement (IOA), root mean square error (RMSE), Akaike information criterion (AIC), and scatter index (SI) can be defined as follows:

$$IOA = 1 - \frac{\sum_{i=1}^{NT} (d_s/d_{oi(Predicted)} - d_s/d_{oi(Observed)})^2}{\sum_{i=1}^{NT} [|d_s/d_{oi(Predicted)} - \overline{d_s}/d_{o(Observed)}| + |d_s/d_{oi(Observed)} - \overline{d_s}/d_{o(Observed)}|]^2} \tag{16}$$

$$RMSE = \left[\frac{\sum_{i=1}^{NT} (d_s/d_{0i(Predicted)} - d_s/d_{0i(Observed)})^2}{NT}\right]^{1/2} \tag{17}$$

$$AIC = NT \cdot \ln(NT \cdot RMSE^2) + 2NOV \tag{18}$$

$$SI = \frac{\sqrt{(1/NT) \sum_{i=1}^{NT} ((d_s/d_{0i(Predicted)} - \overline{d_s}/d_{oPredicted}) - (d_s/d_{0i(Observed)} - \overline{d_s}/d_{o(Observed)}))^2}}{(1/NT) \sum_{i=1}^{NT} d_s/d_{0i(Observed)}} \tag{19}$$

where NT = population size of observations; and NOV = number of independent variables. The IOA as a standardized criterion for evaluation of the proposed model prediction error ranging from 0 to 1. A value of 1 shows the most permissible performance and additionally 0 shows that the proposed model stands at lowest level of accuracy without an agreement between observed values and predicted ones. Furthermore, AIC has the capability of evaluating relative quality of statistical performances for a given datasets. Negative values of AIC indicates better performance of the proposed model in comparison with positive ones.

Quantitative results of the proposed models were given in Table 4. IOA (0.962) and RMSE (0.465) given by MT indicated better performance for the training stage than those obtained using EPR (IOA = 0.957 and RMSE = 0.495) and GEP (IOA = 0.941 and RMSE = 0.582). Moreover, MT provided the lowest value of AIC (750.95) rather than GEP (AIC = 839.81) and EPR (AIC = 776.71). In the case of SI measure, MT model with SI of 0.231 has superiority to the GEP (SI = 0.289) and EPR (SI = 0.246). Generally, it can be said that MT, introduced in form of linear formulations, predicted the local scour depth with permissible level of precision in comparison with other proposed models. Qualitative performance of the proposed models for training stage were illustrated in Fig. 6. From Fig. 6, for observed d_s/d_0 between 0 and 2, all the proposed models indicated relatively significant over-prediction and additionally, for $d_s/d_0 = 2$–7, almost dimensionless scour depth have placed in an allowable error bound.

Table 3. Setting parameters for the proposed EPR

Parameter description	Parameter setting
Function set	Exponential
Type of model	Statical
Type of mathematical expression	$\hat{y} = a_0 + \sum_{j=1}^{m} a_j (X_1)^{ES(j,1)} \ldots (X_K)^{ES(j,K)} \times f((X_1)^{ES(j,K+1)}) \ldots f((X_K)^{ES(j,2K)})$
Exponents range	[−2, −1.5, −1, −0.5, 0, 0.5, 1, 1.5, 2]

Table 4. Results of performances for the proposed models and empirical equations

Model	IOA	RMSE	AIC	SI
Training stage				
EPR	0.957	0.495	776.71	0.246
MT	0.962	0.465	750.95	0.231
GEP	0.941	0.582	839.81	0.289
Testing stage				
EPR	0.958	0.419	117.588	0.181
MT	0.947	0.471	129.013	0.207
GEP	0.943	0.487	132.264	0.214
Eq. (1) [proposed by Abt et al. (1984)]	0.761	1	202.97	0.428
Eq. (3) [proposed by Lim (1995)]	−4.127	4.648	353.27	1.737
Eq. (4) [proposed by Day et al. (2001)]	0.655	1.205	221.01	0.406
Eq. (5) [proposed by Azamathulla and Haque (2012)]	0.077	1.972	269.25	0.823

In the testing stage, with respect to IOA and RMSE values, Eq. (14) extracted from EPR produced the local scour depth at culvert outlets with relatively lower computational error (IOA = 0.958 and RMSE = 0.419) in comparison with GEP (IOA = 0.943 and RMSE = 0.487) and MT (IOA = 0.947 and RMSE = 0.471). Also, SI value was indicative of being better performance of EPR (0.181) related to the MT (SI = 0.207) and GEP (SI = 0.217). Overall, it should be noted that ERP model with representation of easy-to-use mathematical expression had the most accurate prediction of local scour depth compared with the other proposed methods. On the other hand, MT and GEP have stood the second and third ranks in terms of accuracy, respectively. In terms of qualitative comparisons, performances of the proposed AI approaches for testing stage were demonstrated in Fig. 7. From Fig. 7, for the observed d_s/d_0 between 0.25 and 1, all the proposed models had insignificant over-prediction, whereas for the observed $d_s/d_0 = 1$ and 5, MT had relatively higher level of under-prediction in comparison with GEP and EPR techniques. In fact, for this range of dimensionless scour depth, almost points on Fig. 7 have placed in the permissible error bound for prediction of the scour depth at culvert outlets.

In the current study, four empirical equations were used to evaluate the local scour depth at culvert outlets. Statistical results of traditional equations were given in Table 4. Eq. (1), proposed by Abt et al. (1984), could present relatively accurate estimation of local scour depth at culvert outlets with IOA of 0.761 and RMSE of 1 in comparison with Eq. (3) given by Lim (1995) (IOA = −4.127 and RMSE = 4.648). From Table 4, it can be found that

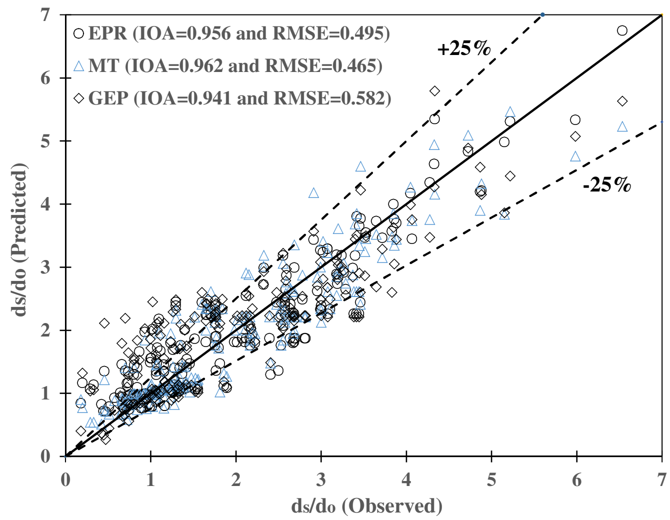

Fig. 6. Scatter plot of observed and predicted scour depth at culvert outlets for training of the proposed models.

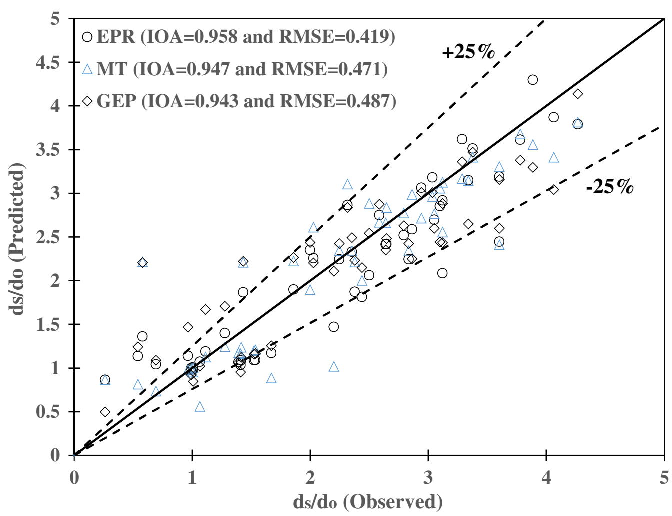

Fig. 7. Scatter plot of observed and predicted scour depth at culvert outlets for testing of the proposed models.

Day et al. (2001) equation [Eq. (4)] has stood at the second rank in terms of prediction precision with IOA of 0.655 and RMSE of 1.25 among empirical equations. Moreover, AIC and SI values produced by Eq. (4) indicated this superiority to the Eq. (3) (AIC $= 353.27$ and SI $= 1.737$) and Eq. (4) (AIC $= 221.01$ and SI $= 0.406$). Ultimately, statistical measures presented in Table 4 indicated that Eq. (5) given by Azamathulla and Haque (2012), provided the local scour depth with the largest computational errors (IOA $= 0.077$ and RMSE $= 1.972$) compared with the other traditional equations. In the case of comparisons, performances of the traditional equations for testing stage have been presented in Fig. 8. From Fig. 8, for observed values of d_s/d_0 between 0.6 and 1.6, Eq. (1) provided the local scour depth with low level of under-prediction (or over-prediction). This means that a fair amount of points have placed out of $\pm 35\%$ bound and additionally the rest of points have placed in an allowable range of computational error for observed $d_s/d_0 = 0.2$–6. As seen in Fig. 8, Eq. (3), given by Lim (1995), has produced the scour depth with tangible level of under-prediction for

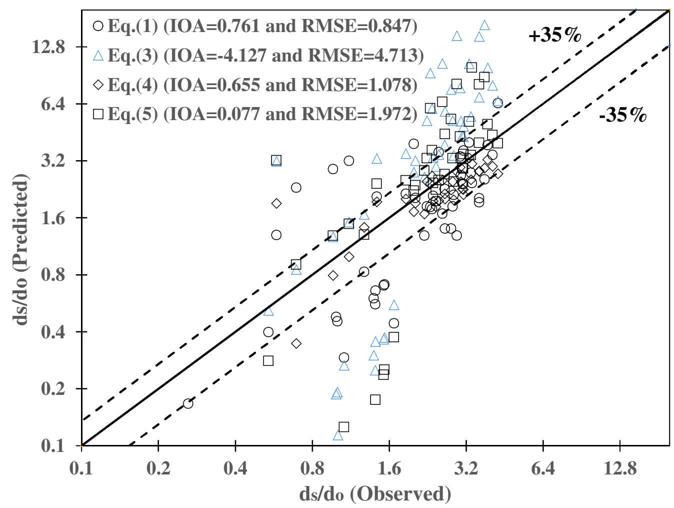

Fig. 8. Scatter plot of observed and predicted scour depth at culvert outlets for testing of the empirical equations.

the observed $d_s/d_0 = 0.8$–1.6 and, for the observed d_s/d_0 between 1.6 and 6.4, remarkable over-prediction was illustrated. Qualitative performance of Eq. (4) showed that almost d_s/d_0 values have placed in the proposed error bound. Moreover, for the observed d_s/d_0 between 0.8 and 1.6, a few points were indicative of under prediction and additionally this trend can be seen for the observed $d_s/d_0 = 1.6$ and 6.4.

Through this study, in the terms of a comparison between proposed models and empirical equations, results of Table 4 indicated that application of AI models based on evolutionary computing for a wide range of experimental datasets and various experimental conditions had better performance than the traditional models. As a major drawback, all empirical equations mentioned in this investigation were proposed merely for a limited range of datasets. This means that these equations did not have the capability to generalize the local scour depth at culvert outlets.

Parametric Study

In the first place, remarkable results of previous experimental investigations indicated that densimetric Froude number (Fr_d) was considered as the most effective variable on the local scour depth at culvert outlets (e.g., Lim 1995; Day et al. 2001). In this way, effects of Fr_d on the local scour depth was investigated with respect to performance of the proposed AI approaches. Fig. 9 illustrated variations of dimensionless local scour depth versus densimetric Froude number for all proposed AI techniques. From Fig. 9, it can be inferred that d_s/d_0 had a dramatic upward trend for Fr_d between 0.05 and 20. For $Fr_d = 20$–88.61, even though slope of variations has decreased, general pattern given in Fig. 8 had an upward trend. Overall, variations of d_s/d_0 versus Fr_d, demonstrated in Fig. 9, were in good agreement with those experimental results reported in literature (Laushey et al. 1967; Bohan 1970; Rajaratnam and Diebel 1981; Abt et al. 1984; Lim 1995).

Fisher Test for the Proposed AI Techniques

In this section, the analysis of variance (ANOVA) approach was carried out to assess statistical reliability pertained to the AI techniques used in this study. In this way, Fisher test was applied to evaluate performance of AI models and empirical equations.

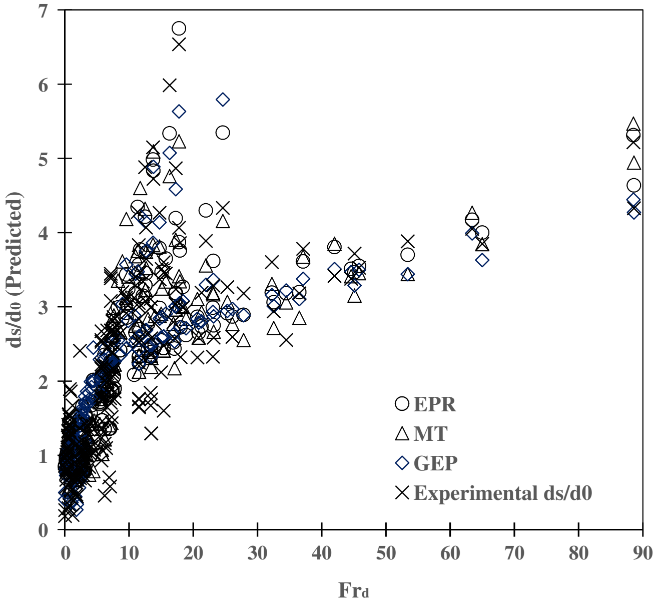

Fig. 9. Variations of d_s/d_0 versus Fr_d.

Through the Fisher test, to assess the hypothesis asserting that value of variation expressed by the regression model is higher than the variation expressed by the averages, the F ratio was employed. In the Fisher test, it is claimed that null hypothesis is accepted if $F_0 > F_{\propto,k,n-p}$, where $\propto$ = significant level; k = number of independent variables; $p = k + 1$; and n = size of the data sample. For all the proposed AI techniques and traditional equations, values of $\propto$, k, and $n - p$ are fixed 0.05, 6, and 49, respectively. From the Fisher test, $F_{0.05,6,49}$ is 2.295 with respect to the F distribution table. Additionally, F_0 is computed as (Hair et al. 1995)

$$F_0 = \frac{MS_R}{MS_E} \tag{20}$$

where MS_R = regression mean square; and MS_E = error mean square. MS_R and MS_E are computed as

$$MS_R = \frac{SSR}{k} \tag{21}$$

$$MS_E = \frac{SSE}{n-p} \tag{22}$$

Moreover, SSR and SSE are the sum of squares due to regression and the sum of squares of error, respectively, being computed as

$$SSR = \sum_{i=1}^{NT} (d_s/d_{0i(Predicted)} - d_s/d_{0i(Observed)})^2 \tag{23}$$

$$SSE = \sum_{i=1}^{NT} (d_s/d_{0i(Predicted)} - \overline{d_s}/d_{0i(Observed)})^2 \tag{24}$$

Statistical results of Fisher test have been given in Table 5. For the d_s/d_0 prediction, with respect to $F_{0.05,6,49}$, F_0 values obtained by EPR ($F_0 = 1.386$) and MT ($F_0 = 1.737$) models have accepted hypothesis and consequently indicated accurate estimation of d_s/d_0 compared with GEP approach ($F_0 = 2.352$). As seen in Table 5, all the empirical equations have rejected hypothesis of Fisher test. For instance, Eq. (2) produced higher level of F_0 (5.505) in comparison with $F_{0.05,6,49}$ value and consequently Eq. (2), proposed by Abt et al. (1984), is indicative of being lower capability in the d_s/d_0 estimation.

Table 5. Analysis of variance for the scour depth prediction

Model	MS_R	MS_E	F_0	State of hypothesis
EPR	1.438	1.037	1.386	Accept
MT	1.815	1.045	1.737	Accept
GEP	1.94	0.825	2.352	Reject
Eq. (1)	8.211	1.922	4.272	Reject
Eq. (2)	176.44	32.05	5.505	Reject
Eq. (4)	11.867	3.417	3.473	Reject
Eq. (5)	31.76	7.727	4.11	Reject

Conclusions

In the current investigation, EPR, MT, and GEP techniques in terms of mathematical expressions were developed to evaluate the local scour depth at culvert outlets. To obtain the optimum relationships on the basis of the proposed AI models, a dimensional analysis was used to extract dimensionless parameters affecting on the scour process at culvert outlets. Performance of the proposed techniques for training and testing stages were assessed using some statistical indices. Beside, empirical equations proposed by Abt et al. (1984), Lim (1995), Liriano and Day (2001), and Azamathulla and Haque (2012) were employed to compare with the EPR, GEP, and MT. From this study, fundamental conclusions were understood as follows:

- Quantitative results of statistical measures in the training stage indicated that linear equations given by MT provided the local scour depth with relatively accurate estimation (RMSE = 0.465 and AIC = 750.95) in comparison with GEP (RMSE = 0.582 and AIC = 839.81) and GEP (RMSE = 0.495 and AIC = 776.71).
- In the testing stage, performance of AI models demonstrated that EPR method predicted the scour depth with higher level of precision (RMSE = 0.419 and AIC = 117.588) than those obtained using GEP (RMSE = 0.487 and AIC = 132.264) and MT (RMSE = 0.471 and AIC = 129.013) techniques. Overall, it can be concluded that Eq. (15) extracted from EPR performance was selected as the best model in terms of precision rather than linear and non-linear equations given by GEP and MT.
- Performance of empirical equations indicated that Eq. (3), proposed by Lim (1995), provided the local scour depth with higher amount of computational error (RMSE = 4.648 and AIC = 353.27) than other traditional equations. From qualitative comparisons among empirical equations, Eq. (3) had significant over-prediction of the local scour depth for the observed d_s/d_0 between 1.6 and 6.4.
- Graphical variations of d_s/d_0 versus Fr_d indicated that results of the proposed AI models had a remarkable coincidence with those reported in literature. To put it another way, an upward trend demonstrated in Fig. 8 confirmed that the scour depth at culvert outlets was a function fundamentally of densimetric Froude number. This findings was in well agreement with Lim (1995) investigations.
- In order to assign the best model with permissible level of accuracy, performance of Fisher test indicated that EPR and MT were capable to present relatively lower level of computational error rather than GEP model.

References

Abida, H., and R. D. Townsend. 1991. "Local scour downstream of box-culvert outlets." *J. Irrig. Drain. Eng.* 117 (3): 425–440. https://doi.org/10.1061/(ASCE)0733-9437(1991)117:3(425).

Abt, S. R., J. F. Ruff, and F. K. Doehring. 1985. "Culvert slope and shape effects on outlet scour." *J. Hydraul. Eng.* 111 (10): 1363–1367. http://doi.org/10.1061/(ASCE)0733-9429(1985)111:10(1363).

Abt, S. R., R. L. Kloberdanz, and C. Mendoza. 1984. "Unified culvert scour determination." *J. Hydraul. Eng.* 110 (10): 1475–1479. https://doi.org/10.1061/(ASCE)0733-9429(1984)110:10(1475).

Abt, S. R., J. F. Ruff, F. K. Doehring, and C. A. Donnell. 1987. "Influence of culvert shape on outlet scour." *J. Hydraul. Eng.* 113 (3): 393–400. https://doi.org/10.1061/(ASCE)0733-9429(1987)113:3(393).

Aderibigbe, O., and N. Rajaratnam. 1998. "Effect of sediment gradation on erosion by plane turbulent wall jets." *J. Hydraul. Eng.* 124 (10): 1034–1042. https://doi.org/10.1061/(ASCE)0733-9429(1998)124:10(1034).

Altomare, C., X. X. Gironella, and D. Laucelli. 2013. "Evolutionary data-modelling of an innovative low reflective vertical quay." *J. Hydroinf.* 15 (3): 763–779. https://doi.org/10.2166/hydro.2012.219.

Azamathulla, H. M. 2012a. "Gene expression programming for prediction of scour depth downstream of sills." *J Hydrol.* 460–461 (Aug): 156–159. https://doi.org/10.1016/j.jhydrol.2012.06.034.

Azamathulla, H. M. 2012b. "Gene-expression programming to predict scour at a bridge abutment." *J. Hydroinf.* 14 (2): 324–331. https://doi.org/10.2166/hydro.2011.135.

Azmathullah, H. M., M. C. Deo, and P. B. Deolalikar. 2005. "Neural networks for estimation of scour downstream of a ski-jump bucket" *J. Hydraul. Eng.* 131 (10): 898–908. https://doi.org/10.1061/(ASCE)0733-9429(2005)131:10(898).

Azamathulla, H. M., and A. A. Ghani. 2010a. "Genetic programming to predict river pipeline scour." *J. Pipeline Syst. Eng. Pract.* 1 (3): 127–132. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000060.

Azamathulla, H. M., and A. A. Ghani. 2010b. "ANFIS-based approach for predicting the scour depth at culvert outlets." *J. Pipeline Syst. Eng. Pract.* 2 (1): 35–40. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000066.

Azamathulla, H. M., and A. A. M. Haque. 2012. "Prediction of scour depth at culvert outlets using gene-expression programming." *INSPEC (IET)* 8 (7B): 5045–5054.

Azamathulla, H. M., and A. A. M. Haque. 2013. "Knowledge extraction from trained neural network scour model at culvert outlets." *Neural Comput. Appl.* 23 (7–8): 2137–2141. https://doi.org/10.1007/s00521-012-1164-2.

Blaisdell, F. W., and C. L. Anderson. 1988. "A comprehensive generalized study of scour at cantilevered pipe outlets." *J. Hydraul. Res.* 26 (4): 357–376. https://doi.org/10.1080/00221688809499198.

Bohan, J. P. 1970. *Erosion and riprap requirements at culvert and storm-drain outlets.* Research Rep. No. H-70-2. Vicksburg, MS: US Army Engineer Waterways Experiment Station.

Day, R. A., S. L. Liriano, and W. R. White. 2001. "Effect of tailwater depth and model scale on scour at culvert outlets." *Proc. Inst. Civ. Eng. Water Marit. Eng.* 148 (3): 189–198. https://doi.org/10.1680/wame.2001.148.3.189.

Ebtehaj, I., H. Bonakdari, F. Moradi, B. Gharabaghi, and Z. S. Khozani. 2018. "An integrated framework of extreme learning machines for predicting scour at pile groups in clear water condition" *Coastal Eng* 135: 1–15. https://doi.org/10.1016/j.coastaleng.2017.12.012.

Etemad-Shahidi, A., and N. Ghaemi. 2011. "Model tree approach for prediction of pile groups scour due to waves." *Ocean Eng.* 38 (13): 1522–1527. https://doi.org/10.1016/j.oceaneng.2011.07.012.

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

Ferreira, C. 2006. *Gene-expression programming, mathematical modeling by an artificial intelligence.* Berlin: Spriner.

Ghaemi, N., A. Etemad-Shahidi, and B. Ataie-Ashtiani. 2013. "Estimation of current-induced pile groups scour using a rule based method." *J. Hydroinf.* 15: 516–528. https://doi.org/10.2166/hydro.2012.175.

Ghazanfari-Hashemi, S., A. Etemad-Shahidi, M. H. Kazeminezhad, and A. R. Mansoori. 2011. "Prediction of pile group scour in waves using support vector machines and ANN." *J. Hydroinf.* 13 (4): 609–620. https://doi.org/10.2166/hydro.2010.107.

Giustolisi, O., and D. A. Savic. 2006. "A symbolic data-driven technique based on evolutionary polynomial regression." *J Hydroinf.* 8 (3): 207–222. https://doi.org/10.2166/hydro.2006.020b.

Giustolisi, O., and D. A. Savic. 2009. "Advances in data-driven analyses and modelling using EPR-MOGA." *J Hydroinf.* 11 (3–4): 225–236. https://doi.org/10.2166/hydro.2009.017.

Guven, A., and H. M. Azamathulla. 2012. "Gene-expression programming for flip-bucket spillway scour." *Water Sci. Technol.* 65 (11): 1982–1987. https://doi.org/10.2166/wst.2012.100.

Guven, A., and M. Gunal. 2008a. "Genetic programming approach for prediction of local scour downstream hydraulic structures." *J. Irrig. Drain. Eng.* 134 (2): 241–249. https://doi.org/10.1061/(ASCE)0733-9437(2008)134:2(241).

Guven, A., and M. Gunal. 2008b. "Prediction of scour downstream of grade-control structures using neural networks." *J. Hydraul. Eng.* 134 (11): 1656–1660. https://doi.org/10.1061/(ASCE)0733-9429(2008)134:11(1656).

Hair, J. F., R. E. Anderson, R. L. Tathan, and W. Black. 1995. *Multivariate data analysis with readings*. 4th ed. Englewood Cliffs, NJ: Prentice-Hall.

Kerenyi, K., J. S. Jones, and S. Stein. 2003. *Bottomless culvert scour study: Phase 1 laboratory report.* Rep. No. FHWA-RD-O2-078. Washington, DC: Federal Highway Administration.

Kerenyi, K., J. S. Jones, and S. Stein. 2007. *Bottomless culvert scour study: Phase 2 laboratory report.* Rep. No. FHWA-HRT-07-026. Washington, DC: Federal Highway Administration.

Laucelli, D., L. Berardi, A. Doglioni, and O. Giustolisi. 2012. "EPR-MOGA-XL: An excel based paradigm to enhance transfer of research achievements on data-driven modeling." In *Proc., 10th Int. Conf. on Hydroinformatics HIC 2012*, edited by R. Hinkelmann, M. H. Nasermoaddeli, S. Y. Li-ong, D. Savic, and P. Fröhle. Hamburg, Germany: Politecnico di Bari.

Laucelli, D., and O. Giustolisi. 2011. "Scour depth modelling by a multi-objective evolutionary paradigm." *Environ. Model. Software* 26 (4): 498–509. https://doi.org/10.1016/j.envsoft.2010.10.013.

Laushey, L. M., U. Kappus, and P. Ofwonam. 1967. "Magnitude and rate of scour at culvert outlets." In *Proc., 12th IAHR Congress*, 338–345. Fort Collins, CO: Colorado State Univ.

Lim, S. Y. 1995. "Scour below unsubmerged full flowing culvert outlets." *Proc. Inst. Civ. Eng. Marit. Eng.* 112 (2): 136–149. https://doi.org/10.1680/iwtme.1995.27659.

Liriano, S. L., and R. A. Day. 2001. "Prediction of scour depth at culvert outlets using neural networks." *J. Hydroinf.* 3 (4): 231–238. https://doi.org/10.2166/hydro.2001.0021.

Liriano, S. L., R. A. Day, and W. R. White. 2002. "Scour at culvert outlets as influenced by the turbulent flow structure." *J. Hydraul. Res.* 40 (3): 367–376. https://doi.org/10.1080/00221680209499951.

Najafzadeh, M. 2015. "Neurofuzzy-based GMDH-PSO to predict maximum scour depth at equilibrium at culvert outlets." *J. Pipeline Syst. Eng. Pract.* 7 (1): 06015001. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000204.

Najafzadeh, M., and S. Sarkamaryan. 2018. "Extraction of optimal equations for evaluation of pipeline scour depth due to currents." *Proc. Inst. Civ. Eng. Marit. Eng.* 171 (1): 1–10. https://doi.org/10.1680/jmaen.2017.10.

Najafzadeh, M., J. Shiri, and M. Rezaie-Balf. 2018a. "New expression-based models to estimate scour depth at clear water conditions in rectangular channels" *Mar. Georesour. Geotechnol.* 36 (2): 227–235. https://doi.org/10.1080/1064119X.2017.1303009.

Najafzadeh, M., A. Tafarojnoruz, and S. Y. Lim. 2018b. "Prediction of local scour depth downstream of sluice gates using data-driven models" *J. Hydraul. Eng.* 23 (2): 195–202. https://doi.org/10.1080/09715010.2017.1286614.

Opie, T. R. 1967. "Scour at culvert outlets." M.Sc. thesis, Dept. of Civil Engineering, Colorado State Univ.

Pal, M., N. K. Singh, and N. K. Tiwari. 2012. "M5 model tree for pier scour prediction using field dataset." *KSCE J. Civ. Eng.* 16 (6): 1079–1084. https://doi.org/10.1007/s12205-012-1472-1.

Quinlan, J. R. 1992. "Learning with continuous classes." In *Proc., Australian Joint Conf. on Artificial Intelligence*, 343–348. Wuhan, China: Scientific Research Publishing.

Rajaratnam, N., and M. Diebel. 1981. "Erosion below culvert-like-structures." In *Proc., 6th Canadian Hydrotechnical Conf.*, 469–486. Ottawa: Canadian Society for Civil Engineering.

Ruff, J. F., S. R. Abt, C. Mendoza, A. Shaikh, and R. Kloberdanz. 1982. *Scour at culvert outlets in mixed bed materials.* Rep. No. FHWA-RD-82-11. Washington, DC: Federal Highway Administration.

Samadi, M., E. Jabbari, and H. Md. Azamathulla. 2014. "Assessment of M5' model tree and classification and regression trees for prediction of scour depth below free overfall spillways." *Neural Comput. Appl.* 24 (2): 357–366. https://doi.org/10.1007/s00521-012-1230-9.

Savic, D. A., O. Giustolisi, L. Berardi, W. Shepherd, S. Djordjevic, and A. Saul. 2006. "Modelling sewer failure by evolutionary computing." *Proc. ICE Water Manage.* 159 (2): 111–118. https://doi.org/10.1680/wama.2006.159.2.111.

Savic, D. A., O. Giustolisi, and D. Laucelli. 2009. "Asset deterioration analysis using multi-utility data and multi-objective data mining" *J. Hydroinf.* 11 (3–4): 211–224. https://doi.org/10.2166/hydro.2009.019.

Wang, Y., and I. H. Witten. 1997. *Induction of model trees for predicting continuous lasses*. Working Paper 96/23. Hamilton, New Zealand: Dept. of Computer Science, Univ. of Waikato.