Gene-expression programming to predict the local scour depth at downstream of stilling basins

M. Mesbahi[a,*], N. Talebbeydokhti[b], S.-A. Hosseini[a] and S.-H. Afzali[b]

a. *Department of Civil Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.*
b. *Department of Civil and Environmental Engineering, Shiraz University, Shiraz, Iran.*

Abstract. This research introduces an approach to predict maximum scour depth, scour hole length, and the length of the deepest point of the scour hole at equilibrium conditions at downstream of stilling basins by using the GEP. Five non-dimensional parameters in terms of physical properties of bed sediments, stilling basin length, tail water depth, and discharge of spillway were considered as input variables to evaluate the scour hole geometry. The GEP model was developed using experimental datasets collected from literature. Results of the GEP models were compared with those obtained using ANFIS and non-linear regression analysis. Performances indicated that the proposed GEP models to characterize scour hole geometry produced more accurate results than the other methods. In addition, results of sensitivity analysis to define the most effective independent parameters on scour hole geometry were reported. Finally, proper application of the proposed model has been confirmed as the GEP-based best formulation, which is a useful soft computing tool for prediction of scour hole geometry at downstream of a stilling basin.

1. Introduction

Scouring process at downstream of hydraulic structures is issued as one of the most significant features in river engineering which jeopardize the stability and safety of structures. Occasionally, hydraulic jump at the downstream of stilling basin with high velocity causes high local shear stress which basically exceeds the threshold shear stresses of bed materials. Within the scour phenomenon, deposition of sediments downstream from reservoir can lead to increase in the tail water depth. Hence, decreasing the total available head for power production is met. To obtain the optimum design of stilling basins as an energy dissipater for decreasing the

destructive influences of scour process, hydraulic engineers should take into account an accurate prediction of maximum scour depth due to hydraulic jump with regard to the economic problems. Evaluating the local scour at downstream of hydraulic structures is one of the most attractive issues in water sciences [1,2].

Experimental studies on the scouring phenomenon at downstream of a ski-jump bucket, grade-control structures, sluice gates, and stilling basins have been conducted comprehensively. From previous investigations, it was established that the exiting traditional equations based on laboratory as well as prototype observations in hydraulic structures were met and also it was the most remarkable to obtain an accurate prediction of the maximum scour depth at downstream of hydraulic structures. Laboratory and field investigations indicated which effects of different variables on the scour process could not be characterized comprehensively. In case of experimental and field studies, lack of facilities for conducting the experiments

---

*. *Corresponding author. Tel.: +98 714 3523911*
   E-mail addresses: mesbahi@iaua.ac.ir (M. Mesbahi);
   taleb@shirazu.ac.ir (N. Talebbeydokhti);
   abbas_hoseyni@srbiau.ac.ir (S.-A. Hosseini);
   afzali@shirazu.ac.ir (S.-H. Afzali)

and the range of observed variables were two chief shortcomings causing over- (or under-) prediction of the scour depth [1,3,4-8].

Within a non-linear system with high degree of complexity, generalization of fundamentally functional relationships between input-output variables with the mathematical approaches is in lack of validation to obtain accurately efficient performance. Empirical equations based linear and non-linear regression methods are limited to the range of experimental data and field input-output data collections. In this way, it should be said that deriving equations have not high generalization capacity to be applied for presenting an accurate prediction of the scour depth at hydraulic structures [9,10].

In the recent decade, in case of scouring at downstream of hydraulic structures, different Artificial Intelligence (AI) models such as Artificial Neural Networks (ANNs), Artificial Neuro-Fuzzy Inference System (ANFIS), Genetic Programming (GP), Gene-Expression Programming (GEP), and Group Method of Data Handling (GMDH) have been applied to predict the scour depth. From these applications, predictive methods based on iterative and evolutionary algorithms were established, and promisingly good validations were yielded for the measured dataset in comparison with empirical equations based regression [11-22].

The main object of this study is GEP approach that has been developed to predict the geometry of scour hole at downstream of a stilling basin due to hydraulic jump. In fact, gene-expression programming can be utilized to find optimum estimation models with the least error and the best fit. Also, GEP model characterizes problems with high degree of complexity in form of simply explicit equations. Results of training and testing stages are investigated to evaluate scour hole for three different flow conditions. Also, the performance of GEP models is compared with those of the ANFIS and traditional equations based non-linear regression analysis [23,24].

The advantages of a system like GEP are clear by nature, but the most important ones should be emphasized. First, the chromosomes are simple entities: linear, compact, relatively small, and easy to manipulate genetically (replicate, mutate, recombine, transpose, etc.). Second, the ETs are exclusively the expression of their respective chromosomes; they are the entities upon which selection acts and, according to fitness, they are selected to be reproduced with modification. During reproduction, there are the chromosomes of the individuals, not the ETs, which are reproduced with modification and transmitted to the next generation [25].

On account of these characteristics, GEP is extremely versatile and it greatly surpasses the existing evolutionary techniques. Indeed, in the most complex problem presented in previous work, GEP surpasses GP by more than four orders of magnitude [25].

2. AI approaches for the scour modeling at downstream of hydraulic structures

Prior to applications of the AI models for prediction of the scour at downstream of a hydraulic structure, Azamathulla et al. (2005) investigated geometry of scour hole downstream of ski-jump bucket using different artificial neural networks [12]. They found that Feed Forward Back Propagation Neural Network (FFBP-NN) had performed relatively well, compared to the Radial Basis Function Neural Network (RBF-NN) and equations based non-linear regression. Neural networks models are "black boxes" and have limited ability to explicitly identify possible causal relationships, so, it may be more difficult to use them in the field; they require greater computational resources; they are prone to over-fitting; their development is empirical; and there remain many methodological issues to be resolved [11,12,15].

Azamathulla et al. (2007) predicted local scour at downstream of flip bucket using ANFIS model [26]. They developed the ANFIS models using experimental input-output datasets. Results of performances indicated that the ANFIS model characterized the best model in comparison with the empirical equations based non-linear regression models. In addition, Several researchers represented ANNs, ANFIS, and GP approaches based field datasets for predicting the maximum scour depth at downstream of ski-jump bucket [11,27,28]. Their proposed models were in compromisingly good agreement with the observed scour depth values. Other researchers predicted scour downstream of grade-control structures using the GP and ANNs methods. GP algorithm operates with populations of individuals that can encode simple computer programs, functions, and solutions to our problem. Similarly, as in the case of genetic algorithms, the efficiency of the evaluation function greatly impacts the efficiency of the whole algorithm and therefore, as well, the application of genetic programming. For that reason, it is important to implement fast evaluation of individuals. From their studies, it was found that AI models produced more efficient results than those obtained using empirical methods [16,18,21].

Goel and Pal (2009) represented ANNs and Support Vector Machines (SVM) for evaluation of scour depth downstream of grade-control structures. They found that equations, based on SVM, work better for large-scale. Also, as other research results from the SVM based modeling show a better performance in comparison to the back propagation ANN and empirical Equation. Performances for training and testing stages indicated that the SVM model gave

better prediction of the scour depth than both exiting equations and FFBP-NN models [13].

Farhoudi et al. (2010) applied adaptive ANFIS to predict the maximum scour depth and its location in downstream of stilling basins due to hydraulic jump. They found that ANFIS predicts the characteristics of scour hole with more efficient results [29]. The overall performance of ANFIS model is superior to that of the ANN model when compared to error based criteria. Further, it is required to collect field data of scour to train the GP approach and validate its usefulness [17].

Also, Guven (2011) used a multi-output Descriptive Neural Network (DNN) model to predict scour hole geometry downstream of hydraulic structures [15]. The suggested method shows that the explicit formulation, extracted from DNN, can replace the conventional regression equations with high accuracy. Luacelli and Goistolisti (2011) applied Evolutionary Polynomial Regression (EPR) based field and experimental datasets for prediction of maximum scour depth downstream of grade-control structures [19].

Samadi et al. (2012) have predicted the scour process below free overfall spillways by assessment of M5' model tree and Classification And Regression Trees (CART). Statistical error parameters used for qualitative and quantitative representation of results indicated that model trees were more accurate than the CART model for evaluating the scour depth. The research results of soft computing approaches were compared with those of empirical equations and showed that the soft computing methods are more accurate and have better performance than the empirical equations [22].

Multi-Objective Evolutionary Polynomial Regression (MO-EPR) is a new modeling technique that combines numerical regression and evolutionary computing to local scouring downstream of GCS (Growing Cell Structures) modeling. MO-EPR performs an evolutionary-based multi-objective optimization in the space of solutions, using three conflicting objective functions describing accuracy of the candidate models. MO-EPR performs a set of optimal data models of different accuracy and complexity [19].

Sreeja (2012) evaluated empirical equations for predicting scour at downstream of ski-jump spillway using laboratory and field datasets. Finally, he proposed a generalized equation with more efficiency using experimental and field data reported in the literature [30].

Meantime, Azamathulla (2013 & 2014) studied local scour depth at downstream of the ski-jump bucket spillways. He gave some comments on the previous investigations and suggested constructive corrections [31,32].

Najafzadeh et al. (2014b) have applied the GMDH networks based laboratory datasets for prediction of scour hole geometry at downstream of a ski-jump bucket [33]. Their studies have shown that GMDH networks give relatively good results in comparison with the ANNs, ANFIS, GP, and empirical models. Meantime, Najafzadeh and Lim (2014) improved Neuro-Fuzzy GMDH network using Particle Swarm Optimization (PSO) algorithm to predict the local scour depth downstream of sluice gates. Performances of the Neuro-Fuzzy Group Method of Data Handling Particle Swarm Optimization (NF-GMDH-PSO) network produced more accurate results than those of the other empirical equations [20].

2.1. Analysis of effective parameters on the scouring downstream of stilling basins

In the previous investigations about scouring prediction at hydraulic structures, effective parameters on the scour hole geometry have been considered, including those of tail water condition, physical properties of bed material, hydraulic conditions of hydraulic jump, and geometry of hydraulic structures [1,12,16,19,29,33,34]. In this way, for scour modeling at downstream of stilling basins, the functional relationship between input-output parameters can be expressed as follows:

$$f\left(q, g, d_{50}, \sigma_g, d_2, h_d, L, \rho_s, \rho_w, d_s, l_s, l_h\right) = 0, \qquad (1)$$

in which q, g, d_{50}, σ_g, d_w, h_d, L, ρ_s, ρ_w, d_s, l_s, l_h and l_s are the unit discharge of spillway, acceleration due to gravity, median sediment size, geometric standard deviation of sediment size distribution, tail water depth, spillway height, length of stilling basin, density of sediment, density of water, maximum scour depth, length of the scour hole, and length of the deepest point of the scour hole, respectively [1,33,34].

Previous investigations for prediction of the scour hole geometry downstream of hydraulic structures indicated that grouped non-dimensional variables produced good results in comparison with those of dimensional variables [12,15,23,24,33,35]. Therefore, using the Buckingham theorem, non-dimensional equations in functional forms were used to develop the model as follows:

$$\frac{d_s}{d_w} = f\left(\frac{q}{\sqrt{gd_w^3}}, \frac{h_d}{d_w}, \sigma_g, \frac{d_{50}}{d_w}, \frac{L}{d_w}, \frac{\rho_s}{\rho_w}\right), \qquad (2)$$

$$\frac{d_s}{d_w} = f\left(\frac{q}{\sqrt{gd_w^3}}, \frac{h_d}{d_w}, \sigma_g, \frac{d_{50}}{d_w}, \frac{L}{d_w}, \frac{\rho_s}{\rho_w}\right), \qquad (3)$$

$$\frac{l_h}{d_w} = f\left(\frac{q}{\sqrt{gd_w^3}}, \frac{h_d}{d_w}, \sigma_g, \frac{d_{50}}{d_w}, \frac{L}{d_w}, \frac{\rho_s}{\rho_w}\right), \qquad (4)$$

in which $q/\sqrt{gd_w^3}$ is the Froude number (Fr) due to tail water. The ratio of sediment density to water density, ρ_s/ρ_w, would be constant and can be eliminated from the analysis. In this way, Eqs. (2)-(4) are applied to
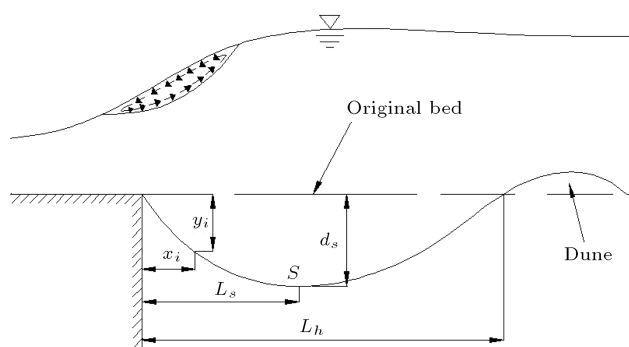
Figure 1. Definition sketch for characterizations of scour hole geometry downstream of a stilling basin.

develop the gene-expression programming for predicting the scour hole geometry at downstream of stilling basin [24,33,35]. To perceive the effective parameters on the scour depth, the process of scour phenomenon downstream of a spillway due to hydraulic jump is illustrated in Figure 1 [36].

Datasets were reported from different conditions of flow, bed sediments, and scales of experimental models. Farhoudi (1979) conducted laboratory works with three types of bed sediments including sand, coarse bakelite, and fine Bakelite [36]. Also, median sediment sizes were between 0.15 and 0.85 mm. In his tests program, three sizes of physical models categorized as small, medium, and large were built to which the lengths of stilling basins were fixed at 41.5, 83, and 166 cm. Heights of spillways were 0.1, 0.2, and 0.4 m. Local scour experiments due to the hydraulic jump have been carried out in three various conditions downstream of stilling basin, i.e. free (balanced tail water), repelled (low-tail water), and drowned (high-tail water) hydraulic jump. In the free jump status, the tail water depth is efficient to produce hydraulic jump on the apron [36].

In this status, hydraulic jump occurs over the movable bed sediments in downstream of stilling basin, leading in an intense sediment transport process due to development of the large eddies in the scour hole and washing the sediments back towards the apron and pile adjacent to its end. Whenever the piled sediment reaches enough height to interrupt the passage of flow filaments leavings the apron, re-entertainment of the piled sediment particles is observed. This mechanism of the scour phenomenon is dominant at initiation of each experiment that has been performed by an oscillation action in the scour hole location. In low tail water conditions, the scour hole depth increases and water level raises tended to move the front of the jump towards of the apron. In addition, it is seen that flow separation mechanism is developed further along the downstream face of the scour hole [36].

Through the drowned hydraulic jump with high tail water conditions, the tail water depth is such that

it forces the jump back toward the spillway and forms the jump adjacent to the spillway toe. As tail water depth increases, the hydraulic jump inclines to be more submerged and occurs closer to the drop toe [36].

In this study, effects of different scales on performing the predictive data-exiting approaches and regression based techniques are neglected. There may exist the expectation that the capacity generalization of models reduces. Meantime, this feature was issued in previous studies [2,12,16].

Datasets used for scour modelling are represented in three groups based on tail water conditions. The first group of datasets is the maximum of scour hole depth related to the high tail water condition (ratio of tail water depth, d_w, to the conjugate depth of hydraulic jump, d_2, is 1.25 ($d_w/d_2 = 1.25$)). In this condition, 442, 156, and 221 series datasets are devoted to prediction of d_s/d_w, l_s/d_w, and l_h/d_w, respectively. The second group of datasets is the maximum of scour hole length pertained to the balanced tail water conditions ($d_w/d_2 = 1$). For the balanced condition, 418, 156, and 262 series datasets are devoted to prediction of d_s/d_w, l_s/d_w, and l_h/d_w, respectively. Finally, the third one is the distance between the end of stilling basin and the maximum of scour hole depth in low tail water status ($d_w/d_2 = 0.78$). For this condition, 356, 150, and 206 series datasets are related to prediction of d_s/d_w, l_s/d_w, and l_h/d_w, respectively. Generally, 1216, 462, and 689 datasets are collected for prediction of the maximum scour depth, length of the deepest point of the scour hole, and scour hole length downstream of stilling basins, respectively. Each of datasets is divided into two parts for training and testing stages. It is clear that the selected models are inherently very sensitive to the number of datasets used for training. In this way, for each condition, 75% of datasets are selected by trial and error for the training stage, and the remaining of them (25%) are devoted to test the model. Also, Table 1 presents the ranges of input-output parameters to evaluate scour hole geometry downstream of stilling basins [31,32,36].

2.2. Development of the gene-expression programming

Most recently, a new technique, called GEP, was developed which was an extension of the GP approach. The GEP is a search model that evolves computer programs in forms of mathematical expressions, decision trees, and logical expressions. In addition, the GEP model has attracted the attention of investigators in prediction of characteristics in hydraulic problems. This research represents GEP models for evaluation of scour hole geometry at downstream of stilling basins. The GEP approach is coded in form of linear chromosomes which are then expressed into Expression Trees (ETs) [37-40].

Table 1. Range of input-output parameters for scour hole modeling.

Tail water conditions	Fr	$\dfrac{d_{50}}{d_w}$	$\dfrac{L}{d_w}$	$\dfrac{h_d}{d_w}$	σ_g	$\dfrac{d_s}{d_w}$	$\dfrac{l_s}{d_w}$	$\dfrac{l_h}{d_w}$
High $(d_w/d_2 = 1.25)$	0.1507-3.46	0.000706-0.0119	3.73-19.88	0.898-4.79	1.28-1.35	0.0398-0.9132	0.205-8.963	0.404-7.831
Balance $(d_w/d_2 = 1)$	0.3407-0.3413	0.00154-0.0132	5.073-10.633	1.222 -2.562	1.28-1.35	0.0403-2.063	—	—
	0.225-0.329	0.00117-0.0129	3.905-10.395	0.941-2.504	1.28-1. 31	—	0.216-6.143	—
Low $(d_w/d_2 = 0.78)$	0.329-0.341	0.0015-0.0132	4.96-10.63	1.195-2 .56	1.28-1.35	—	—	0.2304-12.78
	0.4781-0.4788	0.00192-0.0166	6.357-13.32	1.532-3.211	1.28-1.35	0.238-4.179	0.792-12.34	2.366-16.04

In fact, the ETs are sophisticated computer programming which are usually evolved to solve a practical problem, and are selected according to their fitness at solving that problem. The corresponding empirical expressions can be obtained from these trees structures. A population of the ETs will discover traits, and therefore will adapt to the particular problem it is employed to solve [37-40].

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 item is evaluated as follows:

$$f_i = \sum_{j=1}^{C_t} \left(M - |C_{(i,j)-T_j}| \right), \qquad (5)$$

in which M, $C_{(i,j)}$, and T_j are the selection range, the value returned by the individual chromosome i for fitness case j, and the largest value for fitness case j, respectively [37-40].

In the second stage, the set of terminals, T, and the set of functions, F, are selected to generate the chromosomes. In this study, the terminal includes five independent parameters in form of $T \left(\frac{d_s}{d_w}, \frac{l_s}{d_w}, \text{ and } \frac{l_h}{d_w} \right) = Fr, \frac{h_d}{d_w}, \sigma_g, \frac{d_{50}}{d_w}, \text{ and } \frac{L}{d_w}$ [37-40].

To find the appropriate function set, it is necessary to peer review previous investigations of scour problems in this area. In this way, four basic operators $(+, -, *, /)$ and basic mathematical functions ($\sqrt{}$, Power, Sin, Cos, Exp) are applied to scour hole geometry modeling. The third step is to configure the chromosomal architecture. The fourth step is selection of liking function. Finally, for the fifth stage, the sets of genetic operators that cause variation and their rates was chosen. The other details related to the architecture of GEP modeling are expressed in the literature. In this study, characterizations of scour hole geometry in forms of d_s/d_w, l_s/d_w, and l_h/d_w are predicted using the GEP model for different tail water conditions (High, Balance, and Low) [37].

The best explicit form of the GEP approach for scour hole geometry in high tail water condition is given as follows:

$$d_s/d_w = 1.502 + 0.03479(L/d_w)^2 + 0.03479 \sin(L/d_w)$$

$$- 0.03479 \cos \left(\frac{\sigma_g}{d_{50}/d_w} \right) - 0.3557 \frac{L/d_w}{(\sigma_g)^2}$$

$$- 0.3557(\sigma_g)^2 - 0.3557(h_d/d_w)^2$$

$$+ 2.927 \sin(\text{Fr}^2), \qquad (6)$$

$$l_s/d_w = \left(1017 \sin(\text{Fr})\text{Fr}^2 - 738.3\sigma_g \text{Fr}^2 \sin(\text{Fr}) \right) \text{Fr}$$

$$+ 1.263 + 0.002509(h_d/d_w)^2(\text{Fr}^2)$$

$$\left(\frac{h_d/d_w}{d_{50}/d_w} - 8.539 \right), \qquad (7)$$

$$l_h/d_w = 0.386 + 0.6101 \frac{(\text{Fr} - L/d_w) \sin(\text{Fr})}{(\text{Fr} - \sigma_g) + 0.4584}$$

$$- 22150 \frac{(h_d/d_w - 3.3288)d_{50}/d_w(\text{Fr}+0.25245)}{L/d_w(L/d_w)^2}$$

$$- 0.01972 \cos \left((L/d_w)^2 \right) . (L/d_w)^2 . \sin(\text{Fr}). \quad (8)$$

Furthermore, the functional set and the operational parameters applied in the proposed GEP models are presented in Table 2.

2.3. Adaptive neuro-fuzzy inference system

The ANFIS, first introduced by Jang (1993), is an approximator which is capable of approximating any real continuous function on a compact set to any degree of accuracy [17]. The basic structure of fuzzy modeling, commonly known as Fuzzy Inference System (FIS), is based on the knowledge that can be inferred from

Table 2. Functional set and the operational parameters used in the proposed GEP models.

Parameter	Value
Population size	100
Number of generations	500
Tournament size	5
Max. gene	4
Max. tree depth	4
Operators and functions	($\sqrt{}$, $+$, $-$, $*$, $/$, Power, Sin, Cos, Exp)

Table 3. Fuzzy rules used to predict the scour hole geometry by the ANFIS model.

Fuzzy rule ($y_{\text{output}} = d_s/d_w$, l_s/d_w, and l_h/d_w)
1. If (Fr is Fr mf1*) and (d_{50}/d_w is d_{50}/d_w mf1) and (L/d_w is L/d_w mf1) and (σ_g is σ_g mf1) and (h_d/d_w is h_d/d_w mf1) then (y_{output} is y_{output} mf1)
2. If (Fr is Fr mf2*) and (d_{50}/d_w is d_{50}/d_w mf2) and (L/d_w is L/d_w mf2) and (σ_g is σ_g mf2) and (h_d/d_w is h_d/d_w mf2) then (y_{output} is y_{output} mf2)
3. If (Fr is Fr mf3*) and (d_{50}/d_w is d_{50}/d_w mf3) and (L/d_w is L/d_w mf3) and (σ_g is σ_g mf3) and (h_d/d_w is h_d/d_w mf3) then (y_{output} is y_{output} mf3)

*: mf1, mf2, and mf3 are membership functions.

available data or verbal information. In this study, three FISs were applied to evaluate characterizations of scour hole geometry. Each of the ANFIS models has five inputs and one output. The ANFIS models were trained using a hybrid algorithm which was a combination of the gradient descent and least square method. In this way, the best value of the learning rate in back propagation, μ, was adjusted 0.01 for the ANFIS models. Also, each of these proposed models were generated using 3 rules for the scour hole prediction (Table 3) [41,42].

2.4. Multiple non-linear regressions

Non-linear relationships based regression models between the dependent variables of d_s/d_w, l_s/d_w, and l_h/d_w and independent parameters were utilized to find the least squares fit to those observations. In triple conditions of tail water, non-linear regressions were conducted to predict scour hole geometry. In this way, equations given by non-linear regression were presented to evaluate d_s/d_w, l_s/d_w, and l_h/d_w in different conditions of tail water as follows:

$$d_s/d_w = \mathrm{Fr}^{1.081} \cdot \left(\frac{d_{50}}{d_w}\right)^{-0.234} \cdot \left(\frac{L}{d_w}\right)^{-0.387} \cdot \left(\frac{h_d}{d_w}\right)^{0.869}$$
$$\cdot \sigma_g^{-1.954}, \tag{9}$$

$$l_s/d_w = \mathrm{Fr}^{0.9497} \cdot \left(\frac{d_{50}}{d_w}\right)^{-0.146} \cdot \left(\frac{L}{d_w}\right)^{0.128} \cdot \left(\frac{h_d}{d_w}\right)^{0.476}$$
$$\cdot \sigma_g^{2.2056}, \tag{10}$$

$$l_h/d_w = \mathrm{Fr}^{1.0419} \cdot \left(\frac{d_{50}}{d_w}\right)^{-0.0605} \cdot \left(\frac{L}{d_w}\right)^{1.0192} \cdot \left(\frac{h_d}{d_w}\right)^{-0.311}$$
$$\cdot \sigma_g^{0.6501}. \tag{11}$$

3. Results and discussion

The results of the proposed GEP, ANFIS models, and traditional equations based regression technique are expressed in this section. The statistical parameters

are correlation coefficient (R), Root Mean Square Error (RMSE), Scatter Index (SI), BIAS, and Mean Absolute Percentage of Error (MAPE) and can be defined to evaluate error indicators in the training and testing stages:

$$R = \frac{\sum_{i=1}^{M} \left(Y_{i(\text{Actual})} - \bar{Y}_{(\text{Actual})}\right)\left(Y_{i(\text{Model})} - \bar{Y}_{(\text{Model})}\right)}{\sqrt{\sum_{i=1}^{M} \left(Y_{i(\text{Actual})} - \bar{Y}_{(\text{Actual})}\right)^2}}$$
$$\cdot \frac{1}{\sqrt{\sum_{i=1}^{M} \left(Y_{i(\text{Model})} - \bar{Y}_{(\text{Model})}\right)^2}}, \tag{12}$$

$$\mathrm{RMSE} = \left[\frac{\sum_{i=1}^{M} \left(Y_{i(\text{Model})} - Y_{i(\text{Actual})}\right)^2}{M}\right]^{1/2}, \tag{13}$$

$$\mathrm{MAPE} = \frac{1}{M}\left[\frac{\sum_{i=1}^{M} \left|Y_{i(\text{Model})} - Y_{i(\text{Actual})}\right|}{\sum_{i=1}^{M} Y_{i(\text{Actual})}} \times 100\right], \tag{14}$$

$$\mathrm{BIAS} = \frac{\sum_{i=1}^{M} \left(Y_{i(\text{Model})} - Y_{i(\text{Actual})}\right)}{M}, \tag{15}$$

$$\mathrm{SI} = \frac{\mathrm{RMSE}}{(1/M)\sum_{i=1}^{M} y_{i(\text{Actual})}}, \tag{16}$$

where $Y_{i(\text{Model})}$ is the predicted values (network output), $Y_{i(\text{Actual})}$ is the observed values (target), $\bar{Y}$ is the mean values, and M is the total of events.

Results of the performances of GEP, ANFIS models, and traditional equations based non-linear regression to predict the scour hole geometry are presented in Table 4.

From Table 4, statistical error parameters for training stage indicate that the GEP approach produces the maximum scour depth with relatively good performances in forms of RMSE = 0.35, MAPE = 0.0426, and SI = 0.551. For prediction of l_s/d_w, the best formulation included four effective parameters of

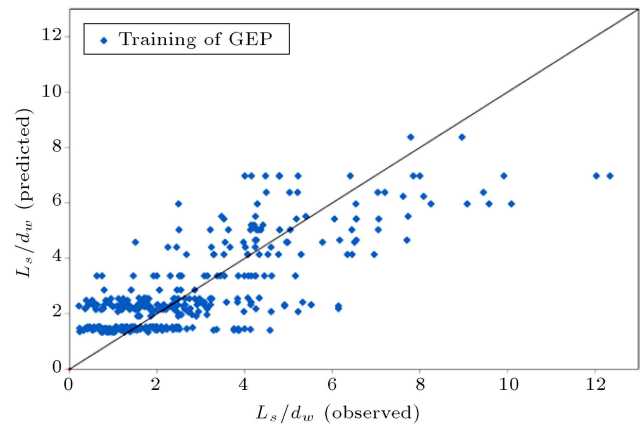**Table 4.** Performances of the proposed models for prediction of scour hole geometry.

Parameters	GEP (training stage)				
	R	RMSE	MAPE	BIAS	SI
d_s/d_w	0.657	0.349	0.0426	0	0.551
l_s/d_w	0.764	1.372	0.103	0	0.473
l_h/d_w	0.831	1.733	0.0612	0	0.418
Parameters	GEP (testing stage)				
	R	RMSE	MAPE	BIAS	SI
d_s/d_w	0.61	0.34	0.134	0.001	0.547
l_s/d_w	0.77	1.64	0.337	-0.069	0.587
l_h/d_w	0.816	1.802	0.19	-0.207	0.441
Parameters	ANFIS				
	R	RMSE	MAPE	BIAS	SI
d_s/d_w	0.61	0.344	0.138	0.011	0.552
l_s/d_w	0.79	1.6	0.322	-0.121	0.495
l_h/d_w	0.794	1.9	0.206	-0.26	0.463
Parameters	Non-linear regression				
	R	RMSE	MAPE	BIAS	SI
d_s/d_w	0.286	0.548	0.144	-0.062	0.878
l_s/d_w	0.747	1.96	0.365	-0.645	0.607
l_h/d_w	0.222	5.82	0.287	-0.284	1.438
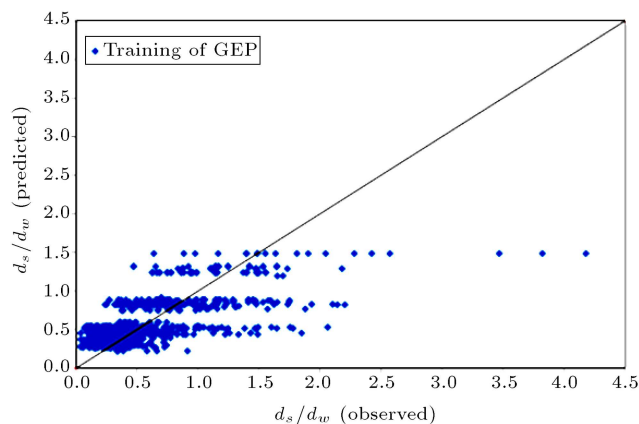
Figure 2. Performance of the training stage of GEP model for prediction of d_s/d_w.

Figure 3. Performance of the training stage of GEP model for prediction of l_s/d_w.
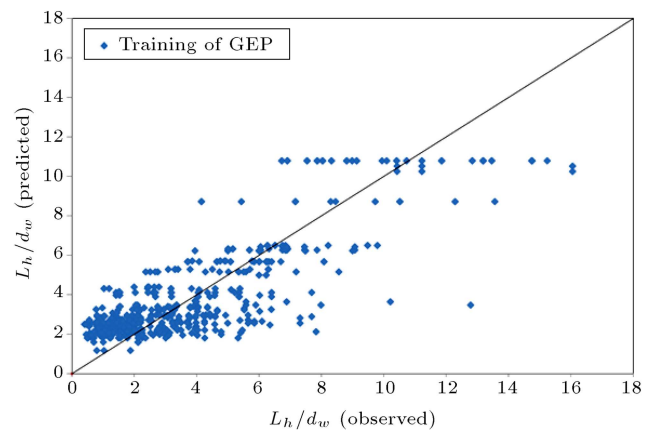
**Figure 4.** Performance of the training stage of GEP model for prediction of l_h/d_w

d_{50}/d_w, h_d/d_w, σ_g, and Fr. Also, high accuracy of the GEP model ($R = 0.76$) was met in comparison with that obtained for d_s/d_w ($R = 0.65$). The proposed GEP model for evaluating l_h/d_w has given $R = 0.831$, RMSE = 1.733, and MAPE = 0.0612. Qualitative training results for the proposed GEP model in predicting the d_s/d_w, l_s/d_w, and l_h/d_w parameters are illustrated in Figures 2 to 4, respectively.

For prediction of d_s/d_w, performing the testing results indicated that the GEP model provided relatively acceptable accuracy in terms of BIAS = 0.001, compared to the ANFIS model (BIAS = 0.011).

Also, according to the statistical error parameters of R, RMSE, MAPE, and SI, the proposed GEP ($R = 0.61$, RMSE = 0.35, MAPE = 0.134, and SI = 0.547) produced approximately the same performance in comparison with that obtained using the ANFIS model ($R = 0.61$, RMSE = 0.344, MAPE = 0.138, and SI = 0.552). Furthermore, the ANFIS model presented prediction of l_s/d_w with relatively higher accuracy ($R = 0.79$, RMSE = 1.6, and SI = 0.495) than that yielded using the GEP model ($R = 0.77$, RMSE = 1.64, and SI = 0.587), whereas the BIAS given by the GEP model (BIAS = -0.069) is lower than that obtained by the ANFIS model (BIAS = -0.121). For predicting the l_h/d_w parameter, the proposed GEP approach indicated remarkably higher accuracy in terms of $R = 0.816$, RMSE = 1.802, and BIAS = -0.207, compared to that given by the ANFIS model ($R = 0.794$, RMSE = 1.9, and BIAS = -0.26). Performances of testing results for the proposed GEP models in predicting the d_s/d_w, l_s/d_w, and l_h/d_w parameters are given in Figures 5 to 7, respectively.
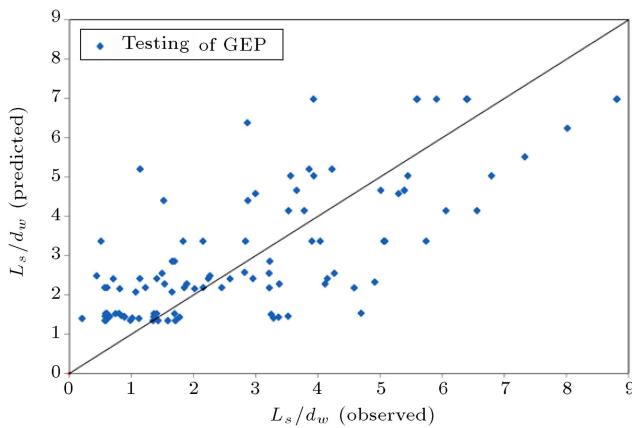
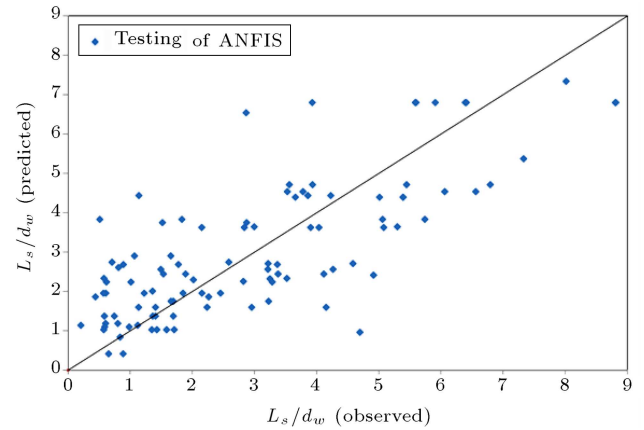In addition, performances of the ANFIS models in predicting the non-dimensional parameters of d_s/d_w,
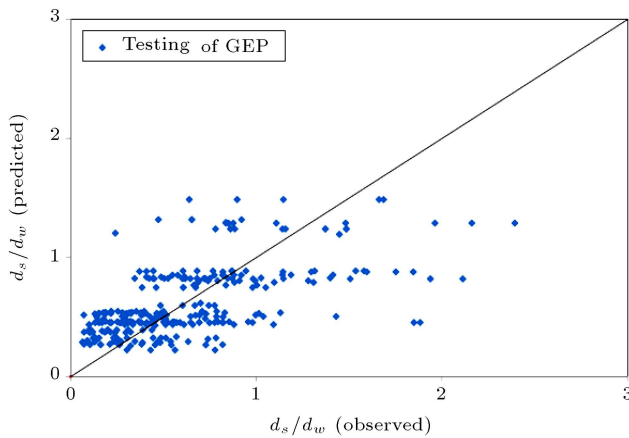
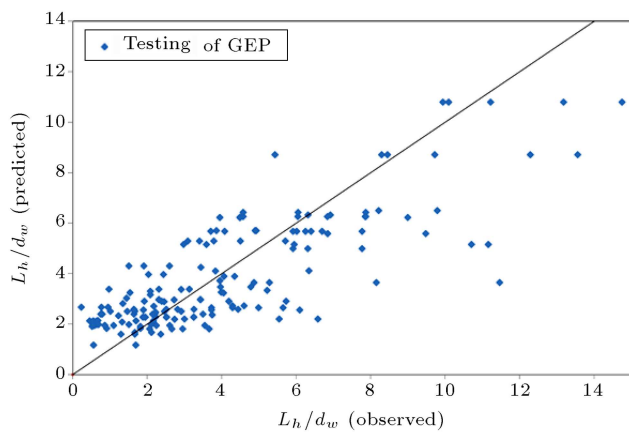**Figure 5.** Performance of the GEP model for prediction of d_s/d_w.
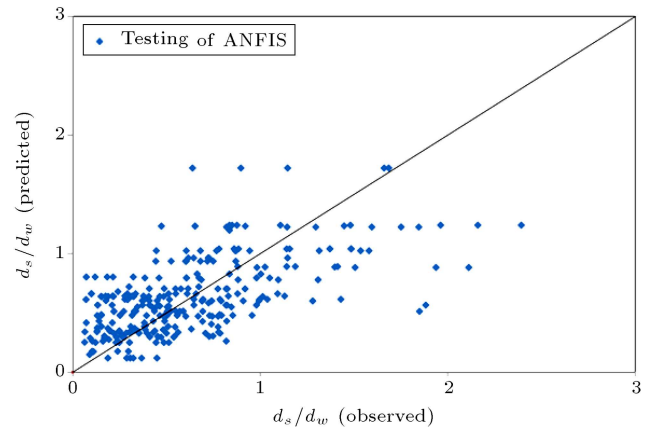
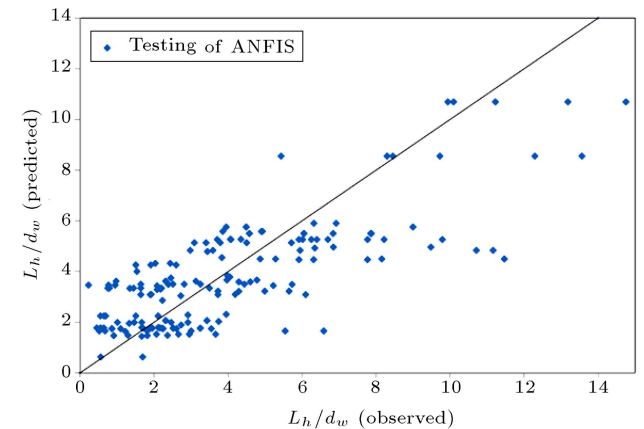**Figure 8.** Performance of the ANFIS model for prediction of d_s/d_w.

Figure 6. Performance of the GEP model for prediction of l_s/d_w.

Figure 9. Performance of the ANFIS model for prediction of l_s/d_w.

Figure 7. Performance of the GEP model for prediction of l_h/d_w.

Figure 10. Performance of the ANFIS model for prediction of l_h/d_w.

l_s/d_w, and l_h/d_w are illustrated in Figures 8 to 10, respectively. Eqs. (9) and (11) based non-linear regression were applied to predict the scour hole geometry downstream of stilling basin.

Statistical results of these equations are given in Table 4. Eq. (9) provided the maximum scour depth with significantly lower accuracy in terms of R = 0.286, RMSE = 0.548 and scatter index = 0.878 than those of the ANFIS and GEP models. However, Eq. (9) based approach failed to yield more accurate prediction of the scour depth downstream of stilling basin. For the length of scour hole, Eq. (10) provided

the higher difference between the predicted and the observed values by parameter BIAS = -0.645 than those given by the ANFIS (BIAS = -0.121) and GEP models (BIAS = -0.069).

From Table 4, it should be said that Eq. (11) provided remarkably low accuracy (R = 0.222) and high error of prediction (RMSE = 5.82 and SI = 1.432) compared to the GEP (RMSE = 1.802 and SI = 0.441) and ANFIS (RMSE = 1.96 and SI = 0.463) models. The lack of validation for Eqs. (9) to (11) corresponds to their limitation in effective parameters, which commonly causes them to fail to present the accurate feature for physical meaning of scour hole geometry. Results of performances for the non-linear regression equations are given in Figures 11 to 13.

4. Sensitivity analysis

To determine the relative significance of each independent parameter (input variables) on scour hole geometry downstream of stilling basin, sensitivity analyses were conducted. The proposed GEP models were selected for this purpose. In order to perform the sensi-
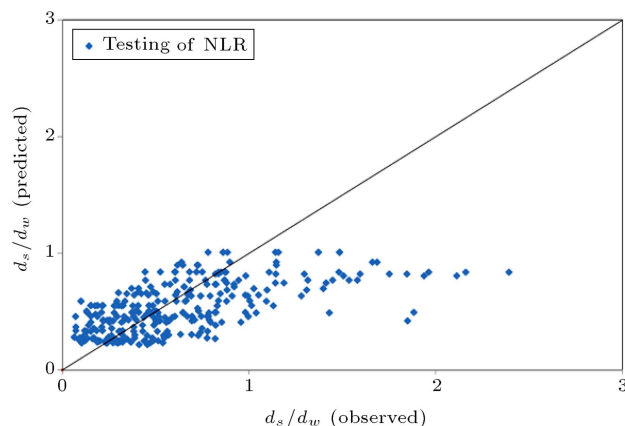
Figure 11. Performance of the non-linear regression model for prediction of d_s/d_w.
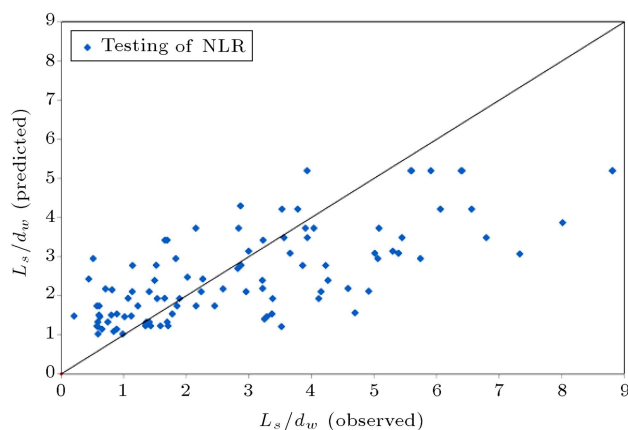
Figure 12. Performance of the non-linear regression model for prediction of l_s/d_w.
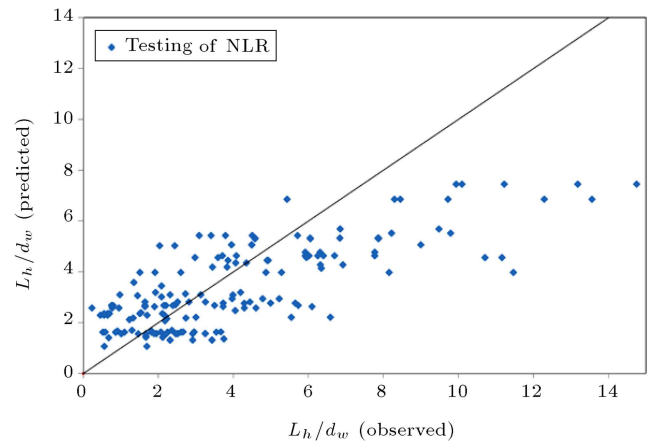
Figure 13. Performance of the non-linear regression model for prediction of l_h/d_w.

tivity analyses, each time, one parameter of Eqs. (2) to (4) was eliminated to evaluate the effect of that input parameter. In this way, statistical error parameters of R, RMSE, MAPE, BIAS, and SI were considered to evaluate performance of sensitivity analysis. Results of performances are given in Table 5.

For prediction of d_s/d_w, elimination of Fr from the list of input variables, compared to the other parameters, leads to relatively high error in evaluation (R = 0.587, RMSE = 0.342, and SI = 0.57). From Table 5, it is found that the parameter related to the bed sediment size, d_{50}/d_w, is the second important input in the form of BIAS = 0.0042 and also, other statistical errors have provided approximately the same values for sensitivity analysis of the l/d_w and h_d/d_w parameters. In addition, the parameter σ_g is defined as the least important parameter in the maximum scour depth with R = 0.64, RMSE = 0.33, and SI = 0.531.

For prediction of l_s/d_w downstream of stilling basin, L/d_w is defined as the most important input variable (RMSE = 1.67, BIAS = -0.102, and SI = 0.518), compared to the other parameters. Meantime, performance of sensitivity analysis indicates that the d_{50}/d_w parameter has the least influence on the l_s/d_w with R = 0.779, MAPE = 0.334, and SI = 0.5083. From Table 5, it is found that the parameter related to the bed sediment size, d_{50}/d_w, is the second important input in form of BIAS = 0.0042 and also, other statistical errors provide approximately the same values for sensitivity analysis of the L/d_w and h_d/d_w parameters. The h_d/d_w parameter related to the geometry of spillway in terms of R = 0.767, RMSE = 1.66, and MAPE = 0.348 is defined as the second important parameter in the scour hole length.

Sensitivity analysis was carried out for l_h/d_w, which demonstrated that the parameter was assigned as the most significant input variable (R = 0.774, RMSE = 1.94, MAPE = 0.208, and SI = 0.476) in comparison with the other non-dimensional parame-

Table 5. Results of sensitivity analysis.

Function	R	RMSE	MAPE	BIAS	SI
$d_s/d_w = f(d_{50}/d_w, L/d_w, h_d/d_w, \sigma_g)$	0.587	0.342	0.134	0.0264	0.57
$d_s/d_w = f(Fr, L/d_w, h_d/d_w, \sigma_g)$	0.612	0.342	0.135	0.0042	0.547
$d_s/d_w = f(Fr, d_{50}/d_w, h_d/d_w, \sigma_g)$	0.616	0.34	0.134	0.0018	0.545
$d_s/d_w = f(Fr, d_{50}/d_w, L/d_w, \sigma_g)$	0.612	0.341	0.134	0.0008	0.547
$d_s/d_w = f(Fr, d_{50}/d_w, L/d_w, h_d/d_w)$	0.64	0.33	0.131	-0.0017	0.531
$l_s/d_w = f(d_{50}/d_w, L/d_w, h_d/d_w, \sigma_g)$	0.778	1.64	0.347	-0.1301	0.508
$l_s/d_w = f(Fr, L/d_w, h_d/d_w, \sigma_g)$	0.779	1.64	0.334	-0.0967	0.508
$l_s/d_w = f(Fr, d_{50}/d_w, h_d/d_w, \sigma_g)$	0.77	1.67	0.336	-0.102	0.518
$l_s/d_w = f(Fr, d_{50}/d_w, L/d_w, \sigma_g)$	0.767	1.66	0.348	-0.0731	0.516
$l_s/d_w = f(Fr, d_{50}/d_w, L/d_w, h_d/d_w)$	0.776	1.64	0.341	-0.0892	0.508
$l_h/d_w = f(d_{50}/d_w, L/d_w, h_d/d_w, \sigma_g)$	0.774	1.94	0.208	0.146	0.476
$l_h/d_w = f(Fr, L/d_w, h_d/d_w, \sigma_g)$	0.816	1.803	0.194	-0.158	0.441
$l_h/d_w = f(Fr, d_{50}/d_w, h_d/d_w, \sigma_g)$	0.814	1.809	0.194	-0.136	0.443
$l_h/d_w = f(Fr, d_{50}/d_w, L/d_w, \sigma_g)$	0.775	1.94	0.2033	-0.196	0.474
$l_h/d_w = f(Fr, d_{50}/d_w, L/d_w, h_d/d_w)$	0.81	1.87	0.197	-0.184	0.446

ters. Furthermore, the h_d/d_w parameter is the second important variable on the l_h/d_w with $R = 0.779$, RMSE = 1.94, MAPE = 0.203, and SI = 0.474. From Table 5, statistical errors parameters in terms of $R =$ 0.816, RMSE = 1.803, MAPE = 0.19, and SI = 0.441 indicate that d_{50}/d_w has the least effect on the d_{50}/d_w parameters.

5. Conclusion

In this research, efficiency of the GEP model based relationship to predict the scour hole geometry in terms of the maximum scour depth, scour hole length, and the length of the deepest point of the scour hole at downstream of stilling basin has been studied. The GEP approach was designed using five non-dimensional parameters and three output variables.

Results of performances for the proposed GEP models indicated more accurate prediction of scour hole characterizations than those for the ANFIS and non-linear regression models. According to the quantitative comparisons, statistical error parameters indicated that the GEP approach predicted the maximum scour depth with lower errors of BIAS = 0.001, RMSE = 0.35, MAPE = 0.134 and higher correlation coefficient of $R = 0.97$, compared to the ANNs, ANFIS (MAPE = 0.138, BIAS = 0.011, and SI = 0.552) and non-linear regression techniques ($R = 0.286$, RMSE = 0.548, MAPE = 0.134). From the sensitivity analysis, carried out by the GEP model, it can be concluded that Fr, L/d_w, and h_d/d_w were the most important parameters in prediction of d_s/d_w, l_s/d_w, and l_h/d_w, respectively.

Finally, it is proven that the GEP model based on

the best relationship is a powerful technique to solve scour problems at downstream of stilling basins as well as the other AI models.

Nomenclature

AI	Artificial Intelligence
ANNs	Artificial Neural Networks
ANFIS	Adaptive Neuro-Fuzzy Inference System
GP	Genetic Programming
GEP	Gene-Expression Programming
GMDH	Group Method of Data Handling
FFBP-NN	Feed Forward Back Propagation Neural Network
RBF-NN	Radial Basis Function Neural Network
SVM	Support Vector Machines
DNN	Descriptive Neural Network
EPR	Evolutionary Polynomial Regression
CARTs	Classification And Regression Trees
PSO	Particle Swarm Optimization
NF-GMDH-PSO	Neuro-Fuzzy Group Method of Data Handling Particle Swarm Optimization
ETs	Expression Trees

References

1. Farhoudi, J. and Khalili Shayan, H. "Investigation on local scour downstream of adverse stilling basins", *Ain Shams Engineering Journal*, **5**, pp. 361-375 (2014).

2. Najafzadeh, M. and Azamathulla, H.Md. "Neuro-fuzzy GMDH systems to predict the scour pile groups due to waves", *Journal of Computing in Civil Engineering (ASCE)*, pp. 1943-5487 (2014).

3. Bormann, N.E. and Julien, P.Y. "Scour downstream of grade-control structures", *Journal of Hydraulic Engineering*, **117**(5), pp. 579-594 (1991).

4. Chee, S.P. and Padiyar, P.V. "Erosion at the base of flip buckets", *Canadian Engineering Journal*, **52**(11), pp. 22-24 (1969).

5. D'Agostino, V. and Ferro, V. "Scour on alluvial bed downstream of grade-control structures", *Journal of Hydraulic Engineering*, **130**(1), pp. 24-36 (2004).

6. Incyth, L.H.A. "Estudio sobre modelo del aliviadero de la Presa Casa de Piedra", *Informe Final, Argentina*, DOH-044-03-82, Ezeiza (1982).

7. Lenzi, M.A. and Comiti, F. "Local scouring and morphological adjustments in steep channels with check-dam sequences", *Geomorphology*, **55**(1-4), pp. 97-109 (2003).

8. Martin, R.B.F. "Scouring of rocky river beds by free jet spillways", *International Water Power Dam Construction*, **27**(4), pp. 152-153 (1975).

9. Termini, D. "Bed scouring downstream of hydraulic structures under steady flow conditions", *Experimental Analysis of Space and Time Scales and Implications for Mathematical Modeling*, Catena, **84**, pp. 125-135 (2011).

10. Veronese, A. "Erosion de Fondoa Valle diuno Scarico", Annali dei Lavori Publicci, **759**, pp. 717-726 (1937).

11. Azamathulla, H.Md., Deo, M.C. and Deolalikar, P.B. "Alternative neural networks to estimate the scour below spillways", *Advances in Engineering Software*, **38**(8), pp. 689-698 (2008a).

12. Azamathulla, H.Md., Deo, M.C. and Deolalikar, P.B. "Neural networks for estimation of scour downstream of a ski-jump bucket", *Journal of Hydraulic Engineering (ASCE)*, **131**(10), pp. 898-908 (2005).

13. Goel, A. and Pal, M. "Application of support vector machines in scour prediction on grade-control structures", *Engineering Applications of Artificial Intelligence*, **22**, pp. 216-223 (2009).

14. Guven, A. and Azamathulla, H.Md. "Gene-expression programming for flip-bucket spillway scour", *Water Science & Technology*, **65**(11), pp. 1982-1987 (2012b).

15. Guven, A. "A multi-output descriptive neural network for estimation of scour geometry downstream from hydraulic structures", *Advances in Engineering Software*, **42**, pp. 85-93 (2011).

16. Guven, A. and Gunal, M. "Genetic programming approach for prediction of local scour downstream hydraulic structures", *Journal of Irrigation and Drainage Engineering*, **134**(2), pp. 241-249 (2008a).

17. Jang, J.S.R. "ANFIS: adaptive-network-based fuzzy inference system", *IEEE. Trans. Syst., Man Cyber*, **23**(3), pp. 665-685 (1993).

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

19. Laucelli, D. and Giustolisi, O. "Scour depth modelling by a multi-objective evolutionary paradigm", *Environmental Modelling & Software*, **26**, pp. 498-509 (2011).

20. Najafzadeh, M. and Lim, S.Y. "Application of improved neuro-fuzzy GMDH to predict scour downstream of sluice gates", *Earth Science Informatics*, DOI: 10.1007/s12145-014-0144-8 (2014).

21. Najafzadeh, M. and Barani, Gh.A. "Discussion of genetic programming to predict river pipeline scour by H. Md. Azamathulla and A. Ab. Ghani", *Journal of Pipeline Systems Engineering and Practice (ASCE)*, **4**(4), pp. 1949-1204 (2013).

22. Samadi, M., Jabbari, E. and Azamathulla, H.Md. "Assessment of $M5'$ model tree and classification and regression trees for prediction of scour depth below free overfall spillways", *Neural Computing and Applications*, DOI 10.1007/s00521-012-1230-9 (2012).

23. Sattar, A.M.A. "Gene expression models for the prediction of longitudinal dispersion coefficients in transitional and turbulent pipe flow", *J. Pipeline Syst. Eng. Pract.*, **5**(1), 04013011 (2014a).

24. Sattar, A.M.A. "Gene expression models for prediction of dam breach parameters", *Journal of Hydroinformatics*, **16**(3), pp. 550-571 (2014b).

25. Candida, F. "Gene expression programming: A new adaptive algorithm for solving problems", *Complex Systems*, **13**(2), pp. 87-129 (2001).

26. Azamathulla, H.Md., Ghani, A.A. and Zakaria, N.A. "An ANFIS based approach for predicting the scour below flip-bucket spillway", *Riverside Kuching*, Sarawak, Malaysia, June 6-8 (2007).

27. Azamathulla, H.Md., Guven, A. and Demir, Y.K. "Linear genetic programming to scour below submerged pipeline", *Ocean Engineering*, **38**(8-9), pp. 995-1000 (2011).

28. Azamathulla, H.Md., Ghani, A.A., Zakaria, N.A. and Guven, A. "Genetic programming to predict bridge pier scour", *Journal of Hydraulic Engineering (ASCE)*, **136**(3), pp. 165-169 (2010).

29. Farhoudi, J., Hosseini, S.M. and Sedghi-Asl, M. "Application of neuro-fuzzy model to estimate the characteristics of local scour downstream of stilling basins", *Journal of Hydroinformatics*, **12**(2), pp. 201-211 (2010).

30. Sreeja, C.K.P. "Evaluation of selected equations for predicting scour at downstream of ski-jump spillway using laboratory and field data", *Engineering Geology*, **129-130**, pp. 98-103 (2012).

31. Azamathulla, H.Md. "Comment on 'Evaluation of selected equations for predicting scour at downstream of ski-jump spillway using laboratory and field data' by C. Kumar and P. Sreeja.", *Engineering Geology*, **152**(1), pp. 210-211 (2014).

32. Azamathulla, H.Md. "Comment on 'Evaluation of selected equations for predicting scour at downstream of ski-jump spillway using laboratory and field data' by C. Kumar and P. Sreeja", *Engineering Geology*, **152**(1), pp. 210-211 (2013).

33. Najafzadeh, M., Barani, Gh.A. and Hessami-Kermani, M.R. "Group method of data handling to predict scour at downstream of a ski-jump bucket spillway", *Earth Science Informatics*, **7**, pp. 231-248 (2014).

34. Oliveto, G. and Coumniello, V. "Local scour downstream of positive-step stilling basins", *Journal of Hydraulic Engineering (ASCE)*, **135**(10), pp. 846-851 (2009).

35. Najafzadeh, M., Barani, Gh.A. and Hessami Kermani, M. "Estimation of pipeline scour due to waves by GMDH", *Journal of Pipeline System Engineering and Practice (ASCE)*, **5**(3), pp. 1949-1204 (2014c).

36. Farhoudi, J. "Scaling relationship for local scour downstream of stilling basins", Thesis presented to Southampton University at Southampton, England, in partial fulfillment of the requirements for the degree of Doctor of Philosophy (1979).

37. Azamathulla, H.Md. and Haque, A.A.M. "Prediction of scour depth at culvert outlets using gene-expression programming", *International Journal of Innovative Computing, Information and Control*, **8**(7B), pp. 5045-5054 (2012).

38. Ferreria, C., *Gene-Expression Programming: Mathematical Modeling by an Artificial Intelligence*, GEP Book, 2nd Ed., Spriner, Heidelberg (2006).

39. Ferreria, C. "Gene-expression programming in problem solving", *The 6th Online Word Conference on Soft Computing in Industrial Applications (Invited tutorials)* (2001a).

40. Ferreria, C. "Gene-expression programming: A new adaptive algorithm for solving problems", *Complex Systems*, **13**(2), pp. 87-129 (2001b).

41. Jang, J.S.R., Sun, C.T. and Mizutani, E., *Neuro-Fuzzy and Soft Computing, Englewood Cliffs*, NJ: Prentice-Hall, 640 pages (1997).

42. Jang, J.S.R. and Sun, C.T. "Neuro-fuzzy modelling and control", *Proc. IEEE*, **83**(3), pp. 378-406 (1995).

Biographies

Masoud Mesbahi received his BS degree in Civil Engineering from Islamic Azad University in 1996, and his MS degree in Civil Engineering from Islamic Azad University of South-Tehran in 2002. He is currently a PhD degree student at Science and Research Branch of Tehran-Islamic Azad University. His research interests include scouring at downstream of stilling basin and artificial intelligence approaches. He has presented and published several papers in various national and international conferences.

Nasser Talebbeydokhti is Professor of Civil and Environmental Engineering in the School of Engineering at Shiraz University. He is an active member of the Academy of Sciences and the Editor-in-Chief of Iranian Journal of Science and Technology, Transaction of Civil Engineering, which is an ISI journal. He is also Editor-in-Chief of Journal of Water Engineering and Associate Editor of Hydraulic Engineering Journal. Professor Talebbeydokhti has published more than 100 journal papers and more than 200 conference papers. He has been recognized as prominent hydraulic engineer. His book entitled Analysis and Design of Gabion Structure was awarded as the best engineering book. Professor Talebbeydokhti is member of promotion committee in the School of Engineering and member of research committee in several water, waste water, and environmental engineering companies in Iran.

Seyed Abbas Hosseini obtained his BS in Civil Engineering, MS in Hydraulic Structure and PhD degree in Water Engineering from Sharif University of Technology. He is currently Assistant Professor of Technical & Engineering Department at Islamic Azad University, Science and Research Branch, Tehran, Iran. His research interests are experimental and numerical modeling of incompressible flow, sediment transport, and scouring in hydraulic structures.

Seied Hosein Afzali received his BS degree in Civil Engineering from Shiraz University, Shiraz, Iran, and MS degree from Amir Kabir University of Technology, Tehran, Iran, in Hydraulic Structures and PhD degree in the same field from Shiraz University, Shiraz, Iran. He is currently Assistant Professor of Civil Engineering at Shiraz University, Iran. His main research interests are computational fluid dynamics, hydraulic, optimization, river mechanics, and groundwater.