

Predicting scour depth at seawalls using GP and ANNs

Ali Pourzangbar, Aniseh Saber, Abbas Yeganeh-Bakhtiary and
Lida Rasoul Ahari

ABSTRACT

Accurate prediction of maximum scour depth is important for the optimum design of seawall structure. Owing to the complex interaction of the incident waves, sediment bed, and seawalls, the prediction of the scour depth is not an easy task to accomplish. Undermining the recent experimental and numerical advancement, the available empirical equations have limited accuracy and applicability. The aim of this study is to investigate the application of robust data-mining methods including genetic programming (GP) and artificial neural networks (ANNs) for predicting the maximum scour depth at seawalls under the broken and breaking waves action. The performance of GP and ANNs models has been compared with the existing empirical formulas employing statistical measures. The results indicated that both the GP and ANNs models functioned significantly better than the existing empirical formulas. Furthermore, the capability of GP was used to produce meaningful mathematical rules, and an analytical formula for predicting the maximum scour depth at seawalls under breaking and broken waves' attacks was developed by utilizing GP.

Key words | artificial neural networks (ANNs), breaking wave, broken wave, genetic programming (GP), scour depth, seawalls

Ali Pourzangbar (corresponding author)
School of Civil Engineering,
Iran University of Science & Technology,
Tehran,
Iran
E-mail: a_pourzangbar@sut.ac.ir

Aniseh Saber
Architecture Department, School of Engineering,
University College of Nabi Akram,
Tabriz,
Iran

Abbas Yeganeh-Bakhtiary
School of Civil Engineering,
Iran University of Science & Technology,
Tehran,
Iran
and
Department of Mechanical Engineering, Institute
Teknologi Brunei (ITB),
Brunei, Darussalam

Lida Rasoul Ahari
School of Electrical, IT and Computer Sciences,
Islamic Azad University of Qazvin,
Iran

INTRODUCTION

Coastal protection structures such as seawalls and breakwaters are constructed to protect harbors and coasts against attack from waves and to provide a self-sheltered area. The malfunction of coastal structures can result in major socio-economic and environmental problems such as coastal inundation and flooding. The design and construction of coastal protection structures are very costly and time-consuming, and they require skilled labor, therefore, the optimum design of such structures is essential. Scour at seawalls is one of the most important aspects of the stability of the structure. Oumeraci (1994) and Lillycrop & Hughes (1993) claim that scour can cause significant structural instability which can lead to structural failure. Hence, predicting scour depth at seawalls is of great importance in the coastal engineering discipline. The incident wave

climate, geomorphological properties of sediment bed, and structural configurations are key parameters in the prediction of the scour depth at seawalls.

Several studies have been conducted on the non-breaking wave-induced scour at coastal structures (e.g., De Best *et al.* 1971; Xie 1981, 1985; Sumer & Fredsøe 2000; Sumer *et al.* 2005; Lee & Mizutani 2008); conversely, there are few studies available on the broken wave-induced scour. For the case of the non-breaking wave, it is shown that the scour and sediment deposition patterns in front of the coastal structures are governed by the action of standing waves. However, broken wave-induced scour patterns are different from those of non-breaking waves (Tsai *et al.* 2009).

Fowler (1992) conducted laboratory studies to investigate scour at the toe of a vertical seawall under the action of

broken waves. He reported that the scour depth of monochromatic wave terrain exceeded the scour depth associated with the irregular wave terrain, and developed an empirical formula for predicting the maximum scour depth at seawalls by using the irregular wave data set as follows:

$$\frac{S_{\max}}{H_0} = \sqrt{22.72 \frac{h_{\text{toe}}}{L_0} + 0.25} \quad (1)$$

where S_{\max} is the maximum scour depth, and H_0 and L_0 are, respectively, the wave height and length at the deep water condition, and h_{toe} is water depth at the toe of the seawall. The above formula was developed based on laboratory tests with constant structure configuration and bed slope, single type of bed sediment, and limited relative water depth at the toe of the seawall ($h_{\text{toe}}/L_0 < 0.15$); therefore, the formula is limited to a small range of data.

Later, Sutherland *et al.* (2006) performed laboratory measurements to study scour at seawalls with sloping and vertical structure and suggested an empirical formula for predicting the scour depth at seawalls based on Iribarren number, I_r , as:

$$\frac{S_{\max}}{H_{si}} = 1.3I_r + 0.169 \quad (2)$$

in which,

$$I_r = \frac{\tan \beta}{\sqrt{H_{si}/L_p}} \quad (3)$$

where H_{si} is the incident significant wave height, L_p the spectral peak wavelength, and β is the bed slope. This study showed that with the increase of Iribarren number and bed slope, the relative scour depth increases. It was also indicated that for the broken wave-induced scour, unlike the non-breaking case, the wall slope affects the relative scouring depth adversely. The above empirical formula considers the effects of bed slope, as well as the wave steepness and the type of wave breaking; however, the formula does not account for the relative water depth at the toe of the seawall structure, the bed sediment properties, wall slope, and structure configuration.

Comparing the results of Fowler (1992) with Sutherland *et al.* (2006) shows that the relative scour depth increases with the increase in the relative water depth at the toe for $h_{\text{toe}}/L_p \leq 0.12$, and decreases with an increase in the relative water depth for $h_{\text{toe}}/L_p \geq 0.12$. It is also evident that the scour depth due to plunging breaker is larger than that of spilling breaker.

At the same time, Tsai *et al.* (2009) carried out experiments to study the scour depth at the toe of seawalls on a steep beach slope under the action of broken waves. Tsai *et al.*'s (2009) laboratory measurements indicated that the scour depth increases with the increase in the steepness of incident wave, however, the increase in the relative water depth at the toe results in the reduction of scour depth. The latter finding fits well with Sutherland *et al.*'s (2006) results. The relation between relative scour depth at seawalls (S_{\max}/H_0) and the breaker type has been investigated by Tsai *et al.* (2009), and the findings indicate that plunging breaker scour depth is higher than that of spilling breaker or that of non-breaking waves. Equations (4) and (6) are the empirical relations developed by Tsai *et al.* (2009) for prediction of relative scour depth in terms of Iribarren number and the relative water depth at the toe of seawall, respectively:

$$\frac{S_{\max}}{H_0} = 0.92I_r - 0.18 \quad (4)$$

in which,

$$I_r = \frac{\tan \beta}{\sqrt{H_0/L_0}} \quad (5)$$

$$\frac{S_{\max}}{H_0} = \frac{0.231}{[\sinh(2\pi h_{\text{toe}}/L_0)]^{1.35}} \quad (6)$$

Tsai *et al.*'s (2009) formulas, similar to Sutherland *et al.*'s (2006), are not capable of predicting the maximum scour depth by considering all of the effective parameters in the scour processes. The existing regression-based equations available for predicting scour depth at seawalls are generally limited in applicability, due to experimental conditions and parameters tested. Therefore, large uncertainties are associated with the existing relations which inevitably increase the safety factor and construction cost.

Hence, a comprehensible model for scour depth prediction is essential.

In recent years, numerical models have been employed to predict the scour depth around coastal structures. Gislason *et al.* (2000), Chen (2006), and Hajivalie *et al.* (2012) developed numerical models to study the scour in front of breakwaters under non-breaking wave action. The application of numerical models is mainly limited, due to mathematical complexity of modeling scour parameters and time intensity of simulation runtime. Soft computing methods such as artificial neural networks (ANNs), ANFIS and genetic programming (GP) do not have the complexity of numerical models, therefore they have been employed for the prediction of scour depth (e.g., Kambekar & Deo 2003; Azamathullah *et al.* 2005; Azamathulla *et al.* 2008, 2010; Guven *et al.* 2009; Kazeminezhad *et al.* 2010; Etemad-Shahidi & Ghaemi 2011; Zanganeh *et al.* 2011; Azamathulla 2012a, 2012b; Guven & Azamathulla 2012; Yeganeh-Bakhtiary *et al.* 2012; Pourzangbar *et al.* 2013).

Although in several studies soft computing methods were employed to predict the scour depth at hydraulic structures, there is no study available yet which employed soft computing methods to predict the scour depth at seawalls under the broken waves' attack. The main purpose of this study is to develop a robust model for the prediction of scour depth at seawalls under the breaking and broken waves' attack. Hence, the capabilities of GP and ANN were tested for developing an accurate and comprehensive predictive model.

GP developed by Koza (1992) is a generalized form of genetic algorithm (GA) (Goldberg 1989). The main advantage of the GP models is the capability of providing meaningful mathematical expressions. Recently, GP has been employed successfully in suspended sediment modeling (Aytek & Kisi 2008), determination of the most effective parameters (Pourzangbar 2012), seawater level forecasting (Ghorbani *et al.* 2010), short-term water table depth fluctuations' prediction (Shiri & Kisi 2011), horizontal intakes in open channel flow (Azamathulla & Ahmad 2012), and reservoir operation (Fallah-Mehdipour *et al.* 2013). In the current study, the models are trained and tested by applying the data from Fowler (1992), Sutherland *et al.* (2006), and Tsai *et al.*'s (2009) experimental data sets. To verify the developed models, the predicted results were

compared with those of the measurements and empirical relations.

GP AND ANN DEVELOPMENT

Genetic programming

GP, first developed by Koza (1992), is a robust method employed for prediction, classification, and function finding. GP is a generalized form of genetic algorithms (Goldberg 1989); however, there are differences between GP and GA, due to the nature of data representation and final solutions. In GP, the individuals are non-linear chromosomes with different sizes and shapes (called parse trees); however, the individuals have linear structures and fixed size in GA. Moreover, unlike GA and the soft computing methods like ANNs, there is no presumptive structure about the relationship between the independent and dependent variables in GP, but the appropriate objective function and its coefficients and parameters can be determined for any given data set. During the training step, the outcome of GP is called solution, and it is in the form of a parse tree or a mathematical expression, is continually evolving and never fixed. The GP model has two components: (1) the functional set of operators such as arithmetic operations ($-$, $+$, \times , \div), logical functions, mathematical functions (\sqrt{x} , $\tan x$, $\sin h x$, x^2 , ...) and domain specific functions; and (2) the independent and dependent variables and the random coefficients and the constant values referred to as the terminal set.

GP must accomplish the process of evolution consisting of a step-by-step procedure as below:

1. Creating an initially selected random population of the models (solutions) by randomly picking up the defined variables and operations.

In this step, GP produces a certain number of models (referred to as the initial population) by randomly combining the independent variables, constants, and defined operations.

2. Evaluating the fitness of each model (solution or individual) by using a fitness function like root mean square error (RMSE) and selecting out the parents (the individuals who deserve to yield offspring (new solutions)).

The process of parent selection includes various selection methods, two of which are: *ranking* in which the models are selected based on their fitness values and better performances, and *tournament* which is selecting the fittest models as the parents by randomly picking up a certain number of models for special times.

3. Producing new individuals by applying the GP operators to the parents.

The most famous operations used in GP are as shown below:

- **Crossover:** This operation produces two offspring by replacing the two parts (the crossover fragments) of two parents. In other words, the former offspring is produced by replacing the crossover fragment of the first parent with the crossover fragment of the second one (Koza 1992).
- **Mutation:** This operation causes a random change in the structure of a parent. In other words, an offspring is

produced by removing a random part of a parent (the function or the terminal referred to as the mutation point) and inserting another randomly generated subtree at that point.

- **Reproduction:** It has an effect on one parent and produces a child. Reproduction is responsible for keeping a parent in the new population without alteration.
4. Repeating the production of offspring by following steps 2 and 3 up to a certain number (the generation number) and replacing the new offspring with the previous ones.
 5. Iterating the mentioned process (steps 2 to 4) until the termination condition, e.g., the maximum number of generation or fitness function performance, is satisfied (Ferreira 2006).

The implementation procedure of GP is represented in Figure 1. For more information, a good explanation of various concepts of GP can be found in Koza (1992).

In order to solve a problem using GP, the user must accomplish the following preparatory steps.

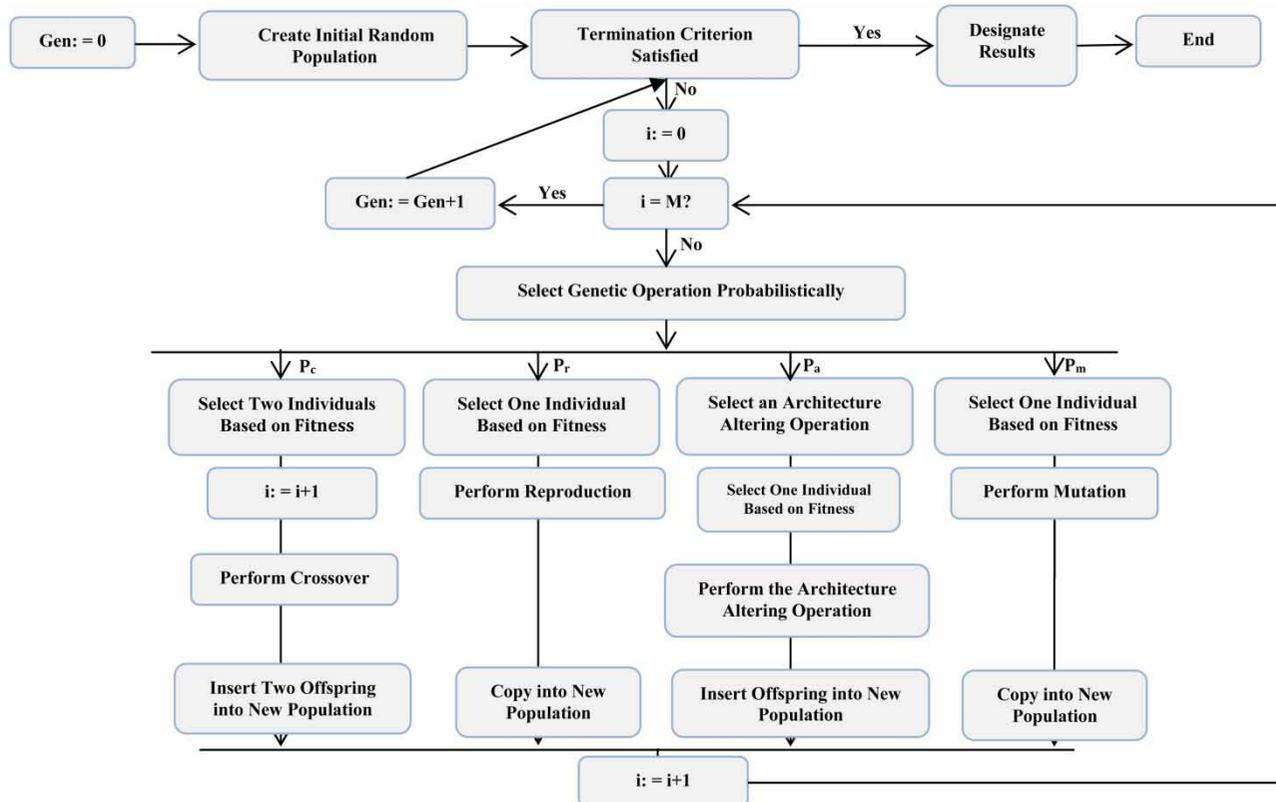


Figure 1 | Flowchart of GP (Koza 1992).

- Determining the set of terminals that correspond to the independent and dependent variables.
- Determining the set of functions, which is rather challenging since the inappropriate function set may change the problem entity, therefore the function set is based on the previous investigations and existing equations in this study.
- Defining the fitness measure, which evaluates how good a particular evolved model can solve the problem. (In this study, RMSE has been chosen as the fitness function.)
- Determining the controlling parameters, such as the chromosomal architecture, the genetic operators' rates, and the genes linking function.

These parameters can be used to control the run. One of the main problems related to the GP application is called bloat phenomenon. During bloat phenomenon, the program size (depth of parse trees) starts growing without any corresponding improvement in the model fitness. The bloat phenomenon results in the nested models that are hard to interpret and are computationally expensive. The nested models may give no sense about the physical basis of the studied phenomenon (Poli & McPhee 2008). Applying parsimony pressure coefficient to the GP models may be regarded as a proper method to the limitation of the parse tree depth as described by Poli & McPhee (2008).

- Choosing the termination condition for terminating a run and accepting the result. (According to Koza (1992), a

specified maximum number of generations or specified perfect level of performance can be the most proper criterion to stop the current run.)

The functional set and the operational parameters used in the GP modeling during this study are presented in Table 1. Other parameters are the default values of version 4.0.954 (Enterprise Edition) of GeneXpro Tools (2006) software application, which is used in this study to evolve the GP models.

Artificial neural networks

ANNs are the new versions of the parallel information processing systems that simulate the human brain behavior to provide a random mapping between an input vector and an output one. Similar to the human brain, which is composed of more than 10 billion interconnected cells (called neurons), ANNs are composed of a certain number of computational elements called neurons (the detailed information of ANN structure and modeling process is available in the Appendix, available with the online version of this paper).

Rogers & Dowla (1994) proposed the following criteria for the determination of hidden layer neurons' number:

$$N^H \leq 2N^L + 1 \quad (7a)$$

$$N^H \leq N^{TR} / [N^L + 1] \quad (7b)$$

where N^H stands for the number of hidden layer neurons, N^L is the number of input parameters (here $N^L = 5$), and N^{TR}

Table 1 | Characteristics of employed GP

| Parameter | Description of parameter | Setting of parameter |
|-----------------|---------------------------------------|---|
| P ₁ | Function set | +, −, ×, ÷, √x, e ^x , x ² , ∛x, tanx, sinhx |
| P ₂ | Number of chromosomes | 30 |
| P ₃ | Head size | 8 |
| P ₄ | Number of genes | 3 |
| P ₅ | Linking function | Addition |
| P ₆ | Fitness function | RMSE |
| P ₇ | Mutation rate | 0.044 |
| P ₈ | One-point and two-point recombination | 0.3 |
| P ₉ | Gene transposition | 0.1 |
| P ₁₀ | Constants per gene | 2 |
| P ₁₁ | Range of constants | −10 to 10 |

stands for the number of training data sets (here $N^{TR} = 31$). According to Equations (7a) and (7b), the number of hidden layer neurons must be less than five neurons ($N^H < 5$) for the current study.

Feed-forward network with standard back propagation algorithm is the most commonly used neural network in many studies (Jain & Deo 2007). In this study, a three-layer feed-forward network with Levenberg–Marquardt back propagation training algorithm is employed for the prediction of scour depth at seawalls under the broken waves' action. The optimized network was achieved by decreasing gradient weight and bias learning function. The learning rate and the iteration, resulting from the trial-and-error process, were 0.01 and 1,000, respectively. The log-sigmoid function was also employed in the optimum network as a transfer function.

EFFECTIVE PARAMETERS

By reviewing previous studies of the scour at seawalls due to the broken waves' action, it can be concluded that the broken wave-induced scour at seawalls mainly depends on three distinctive groups of parameters: the incident wave characteristics, the bed sediment properties, and the seawall (structure) configuration. The functional relationship between the maximum scour depth and influential parameters can be given as Equation (8):

$$S_{max} = f(H_0, L_0, h_{toe}, d_{50}, G_s, U_{fm}, P, \alpha, \beta, \nu) \quad (8)$$

where d_{50} , G_s are the sediment mean diameter and the specific gravity of sediment, respectively, ν stands for the fluid kinematic viscosity, P is the structure permeability index, α is the seawall slope in degree, and U_{fm} is the shear velocity at the undisturbed bed calculated by:

$$U_{fm} = U_m \sqrt{0.15 f_w} \quad (9)$$

where U_m is the maximum wave orbital velocity at the bed just above the wave boundary and f_w is the wave friction coefficient.

Regarding the dimensional analysis, the dimensionless form of the effective parameters (Equation (8)) can be

expressed as:

$$\frac{S_{max}}{H_0} = f \left\{ Cr, \frac{H_b}{L_0}, \frac{d_{50}}{H_0}, Ir_b, \frac{h_b}{L_0}, \frac{h_{toe}}{L_0} \right\} \quad (10)$$

where Cr is the reflection coefficient, and H_b/L_0 , h_b/L_0 , and h_{toe}/L_0 are the normalized breaking wave height, the normalized water depth at the breaking point, and the relative water depth at the toe of the structure, respectively, d_{50}/H_0 is the normalized mean diameter of bed sediment, and Ir_b is the breaking surf similarity parameter given by $Ir_b = \tan\beta / \sqrt{H_b/L_0}$. The dimensionless parameters reviewed in Equation (10) account for the effects of interactions among the bed sediment, the broken wave, and the seawall during the scour process, and they are used for developing the GP and ANN models. The reflection coefficient in Equation (10) considers the effects of seawall configurations (i.e., permeability and front wall slope) on the scour. For seawalls with a small reflection coefficient, unlike the non-breaking wave-induced scour, the scour depth is larger than that of structures with large wave reflection (Sutherland *et al.* 2006; Tsai *et al.* 2009). The surf similarity parameter describes the impacts of the wave breaking type on the scour depth. The scour depth due to plunging breaker is larger than that of spilling breaker or non-breaking waves (Sutherland *et al.* 2006; Tsai *et al.* 2009).

It is indicated that the pattern of sediment erosion and deposition is significantly affected by the mode of sediment transport. Also, in the case of the broken wave-induced scour, the wave breaking and turbulence result in suspended sediment transport. Therefore, the characteristics of the broken wave and turbulence are more dominant than the wave shear stress or Shields parameter (Sutherland *et al.* 2006). The normalized water depths at the toe (h_{toe}/L_0), the breaking point (h_b/L_0), and the normalized breaking wave height (H_b/L_0) indicate the interaction between the broken wave and bed sediment.

RESULTS AND DISCUSSION

Available experimental data

The key mechanism of breaking wave-induced scour is completely different from that of non-breaking wave-induced

scour. In other words, in order to have a physically sound prediction, only the breaking waves' data set must be utilized for scour due to breaking waves. Therefore, in this paper, only breaking waves' data set is used for developing models with ANN and GP.

The waves' regularity only affects the amount of maximum scour depth, and it does not have any influence on the scour physics. For instance, the maximum scour depths associated with regular waves are larger than those of irregular waves (Sumer & Fredsøe 2000). Therefore, both regular and irregular waves are used to develop predictive models in this paper. In brief, in this paper, only the breaking/broken waves' data set is used for developing models. This data set includes the regular and irregular waves.

According to previously mentioned statements, in this study, the broken or breaking waves data sets of Fowler (1992), Sutherland *et al.* (2006), and Tsai *et al.* (2009) were used for developing both the ANN and GP models. Fowler's data set includes 18 tests with irregular waves and 4 tests with regular waves for the vertical seawall (Fowler 1992). Sutherland *et al.*'s data set contains 35 tests under irregular waves' attack for the vertical and inclined (1:2 slope) front-wall slopes (Sutherland *et al.* 2006). Tsai *et al.* (2009) performed 25 tests for scour under regular wave conditions, mainly for non-breaking waves on seawalls. In total, 41 data points related to the breaking/broken waves scour are used to develop the models.

The data range of modeling parameters for training and testing are presented in Table 2. The training data sets contain 70% of the whole data (28 data points) and the remaining 30% (13 data points) are employed for testing.

Model assessment

This paper presents both the GP and ANN approaches for predicting the relative maximum scour depth (S_{max}/H_0) at seawalls. The models are developed by utilizing different combinations of the governing parameters (Equation (10)). The developed models indicated that the relative scour depth was not sensitive to the normalized mean diameter of the bed sediment (d_{50}/H_0); moreover, applying d_{50}/H_0 as the input parameter led to more complex models with no considerable increase in accuracy. Thus, the effect of sediment size on the relative scour depth is negligible because the range of the sediment size is very low (0.0004–0.0016) in the experimental data set utilized for developing the ANN and GP models. All of the parameters in Equation (10), except the normalized sediment size (d_{50}/H_0), were used for developing the GP and ANN models.

To achieve the optimum ANN results, the number of hidden layer neurons was determined by a trial-and-error method (Table 3). The trial-and-error process revealed that a hidden layer with four neurons results in the best performance of the ANN model. The performance of developed ANN and GP models in predicting the relative scour depth (S_{max}/H_0) at seawalls was compared with the empirical relations proposed by Fowler (1992) and Sutherland *et al.* (2006). Statistical error parameters, such as correlation coefficient (CC), $RMSE$, $BIAS$, and scatter index (SI) were determined to evaluate the accuracy of the developed model and compare it with the existing empirical approaches. Equations (11)–(14) present the statistical

Table 2 | Ranges of the parameters employed to train and test the GP and ANN models

| Parameters | Train range | Test range | Minimum | Average | Maximum |
|---------------|---------------|---------------|---------|---------|---------|
| Cr | 0.260–0.576 | 0.270–0.576 | 0.260 | 0.397 | 0.576 |
| Ir_b | 0.057–0.870 | 0.061–0.780 | 0.057 | 0.272 | 0.870 |
| H_b/L_0 | 0.016–0.087 | 0.009–0.087 | 0.016 | 0.044 | 0.087 |
| h_b/L_0 | 0.019–0.094 | 0.009–0.094 | 0.009 | 0.048 | 0.094 |
| h_{toe}/L_0 | –0.009–0.073 | –0.009–0.055 | –0.009 | 0.023 | 0.073 |
| D_{50}/H_0 | 0.0004–0.0016 | 0.0004–0.0016 | 0.0004 | 0.0006 | 0.0016 |
| S_{max}/H_0 | 0.125–0.782 | 0.194–0.803 | 0.125 | 0.450 | 0.803 |

Table 3 | Trial and error for determining the best number of hidden layer neurons of the employed ANN model

| Hidden layer neurons' number | Data set | CC | RMSE | SI | BIAS |
|------------------------------|-------------------|-------|-------|--------|--------|
| 3 | Training data set | 0.855 | 0.096 | 21.57% | -0.012 |
| | Testing data set | 0.862 | 0.115 | 24.74% | -0.070 |
| | All data sets | 0.846 | 0.102 | 22.61% | -0.029 |
| 4 | Training data set | 0.913 | 0.075 | 16.84% | -0.010 |
| | Testing data set | 0.885 | 0.102 | 21.93% | -0.064 |
| | All data sets | 0.895 | 0.084 | 18.58% | -0.026 |
| 5 | Training data set | 0.726 | 0.135 | 30.28% | 0.039 |
| | Testing data set | 0.961 | 0.050 | 10.75% | -0.009 |
| | All data sets | 0.779 | 0.116 | 25.84% | 0.025 |

formulas used in this paper:

$$CC = \frac{\sum_{i=1}^N (O_i - \overline{O_m})(P_i - \overline{P_m})}{\sqrt{\sum_{i=1}^N (O_i - \overline{O_m})^2 \times \sum_{i=1}^N (P_i - \overline{P_m})^2}} \quad (11)$$

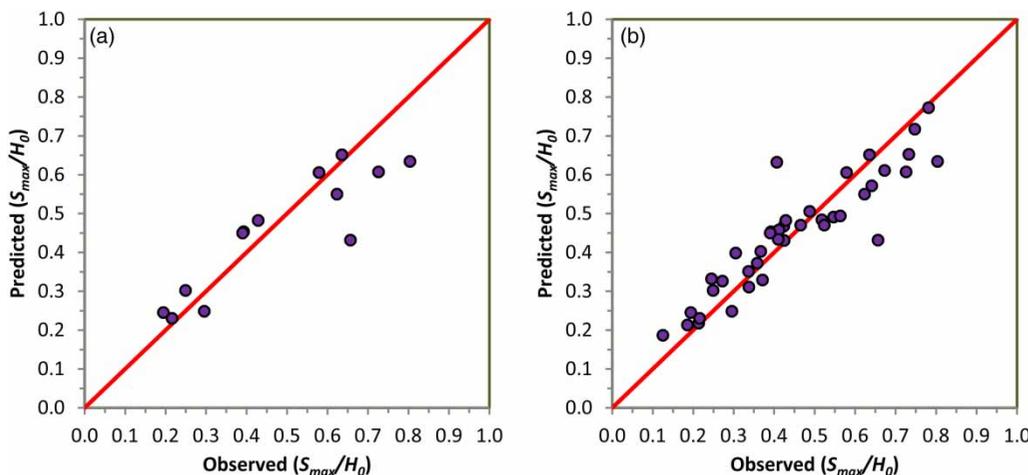
$$RMSE = \sqrt{\frac{\sum_{i=1}^N (P_i - O_i)^2}{N}} \quad (12)$$

$$SI = \frac{\sqrt{\sum_{i=1}^N (P_i - O_i)^2 / N}}{\overline{O_m}} \times 100\% \quad (13)$$

$$BIAS = \frac{\sum_{i=1}^N (P_i - O_i)}{N} \quad (14)$$

where O_i and P_i represent the observed and predicted values, respectively, N is the number of observed data, and $\overline{P_m}$ and $\overline{O_m}$ are the corresponding mean values of the predicted and observed parameters, respectively.

Figures 2 and 3 present the comparison between the observed relative scour depth against the predicted values of the developed GP and ANNs models, respectively. The comparison between the outcomes of GP and ANNs models (Figures 2 and 3) shows that the overall performance of the GP model is better than that of the ANN model in terms of statistical parameters (Table 4). Considering all data sets, the CC of the GP model (0.912) is higher than ANNs (0.895). For the relative scour depth higher than 0.7, both ANNs and GP models underestimate the scour depth. This would be as a result of the lack of sufficient

**Figure 2** | Comparison between measured and predicted relative maximum scour depth (S_{max}/H_0) by GP for (a) test data set and (b) all data sets.

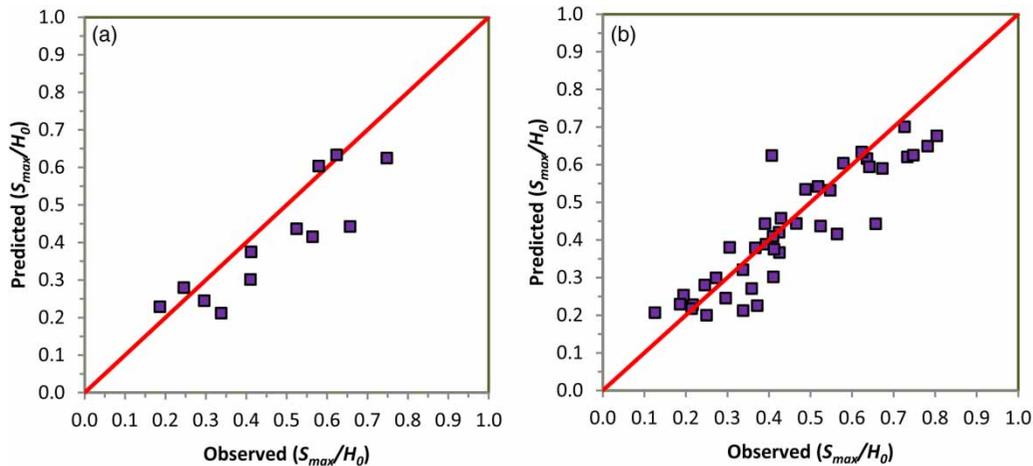


Figure 3 | Comparison between measured and predicted relative maximum scour depth (S_{max}/H_0) by ANN for (a) test data set and (b) all data sets.

Table 4 | Performance indices of various approaches in prediction of relative scour depth

| Model (Equation) | Used data set | CC | RMSE | SI | BIAS |
|--------------------------|-----------------------------------|-------|-------|--------|---------|
| ANN | Testing data set | 0.885 | 0.102 | 21.93% | -0.064 |
| | All data sets | 0.895 | 0.084 | 18.58% | -0.026 |
| GP | Testing data set | 0.896 | 0.095 | 19.95% | -0.023 |
| | All data sets | 0.912 | 0.075 | 16.71% | -0.0003 |
| | Fowler (1992) data set | 0.863 | 0.109 | 21.92% | 0.023 |
| | Sutherland et al. (2006) data set | 0.924 | 0.078 | 18.50% | -0.023 |
| Fowler | Fowler (1992) data set | 0.850 | 0.121 | 23.35% | 0.031 |
| | All data sets | 0.164 | 0.536 | 119.1% | 0.369 |
| Sutherland et al. (2006) | Sutherland et al. (2006) data set | 0.792 | 0.140 | 31.47% | -0.071 |
| | All data sets | 0.043 | 0.355 | 75.1% | 0.049 |

data points for this range. Empirical equations and soft computing approaches use utterly different mechanisms for developing their final results. Nevertheless, the same trend was also observed in the prediction of the empirical formulas (Figure 3). Thus, a lack of data points and their low discrepancy for a larger amount of scour depths result in various models' underestimation.

Figure 4 indicates the performance of empirical formulas (Fowler (1992) and Sutherland et al.'s (2006) proposed equations) in prediction of all data sets utilized in this paper (collected from published literature). The figure reveals that the Fowler (1992) formula overestimates the relative scour depth and the main drawback of Sutherland et al.'s (2006) formula is its approximately constant prediction (0.37) for a

large range of the relative scour depth (0.20–0.80). In other words, this figure shows that these empirical equations are not applicable for data sets which are beyond their experimental data. This is because of the use of limited input parameters and limited range of data sets.

Figure 5 illustrates the performance of each of the models of Fowler (1992), Sutherland et al. (2006), and the GP model for the prediction of data sets of Fowler and Sutherland et al. Considering Fowler's (1992) data set, Figure 5 demonstrates that the GP model outperforms Fowler's formula in prediction of maximum scour depth. Similarly, the GP model performs significantly better than Sutherland et al.'s (2006) formula in prediction of Sutherland et al.'s experimental data set. The results reveal that

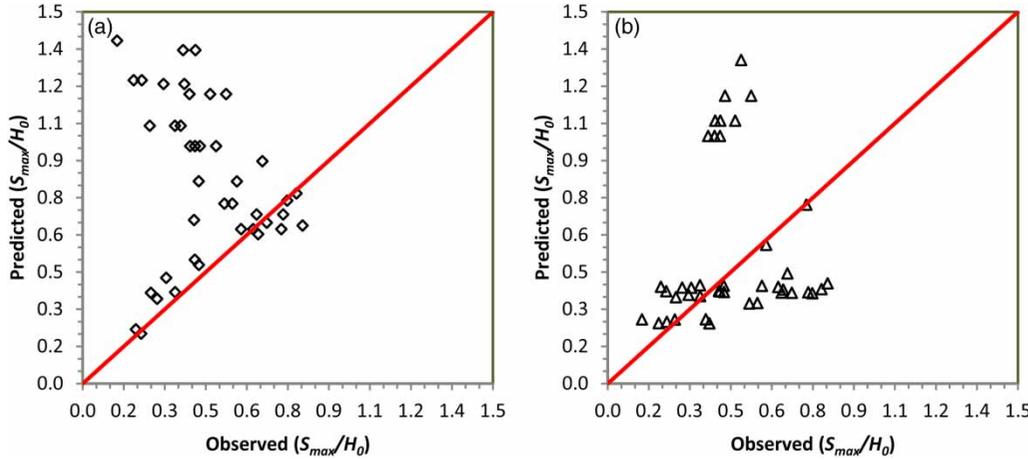


Figure 4 | Comparison between measured and predicted relative maximum scour depth (S_{max}/H_0) for all data sets by (a) Fowler (1992) and (b) Sutherland et al. (2006).

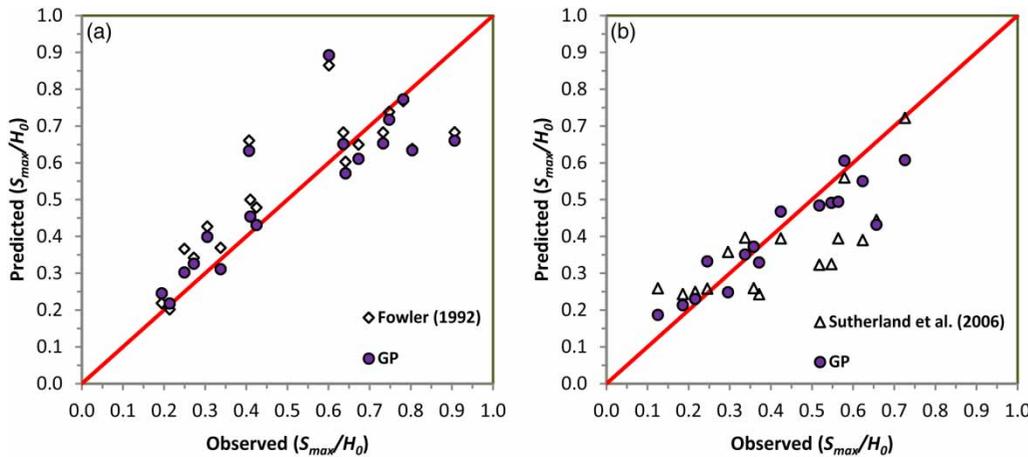


Figure 5 | Comparison between measured and predicted relative maximum scour depth (S_{max}/H_0) by GP and empirical equations for (a) Fowler (1992) data set and (b) Sutherland et al. (2006) data set.

the empirical formulas are capable of predicting the relative scour depth for a limited data range with a fair accuracy; nevertheless, the empirical formulas do not make an accurate prediction for scour depth beyond the range of their experimental data sets. The statistical error parameters (Table 4) confirm the higher accuracy of developed ANNs and GP predictions in comparison with the empirical formulas. In summary, the empirical formulas failed to predict the relative scour depth for the conditions beyond the range of their experimental data sets; nevertheless, the presented approaches (the GP and ANNs models) performed well for a large range of data. Besides higher accuracy, the key advantage of GP is the capability of

producing accurate and meaningful mathematical expression which can be easily used for predicting the relative scour depth at seawalls for different ranges of parameters. The developed GP model (Equation (15)) for predicting relative scour depth is as follows:

$$\frac{S_{max}}{H_0} = \tan \left\{ \frac{h_{toe}/L_0}{Cr - \sqrt{\exp(-0.31 - (h_{toe}/L_0))}} + \frac{(h_b/L_0)(H_b/L_0)}{[\exp(20.34cr^2)]^2} + 0.79 \left\{ \sqrt[3]{0.68 + Ir_b} - \tan \left(\sqrt[3]{\frac{h_b}{L_0}} \right) \right\} \right\} \quad (15)$$

To understand the physical trend of the GP evolved equation (Equation (15)), a parametric evaluation with varying input parameters is very necessary (Kazeminezhad et al. 2010). To perform this, variation of the relative maximum scour depth against the reflection coefficient (Cr) and the relative water depth at the toe (h_{toe}/L_0) was investigated. Figures 6 and 7 show the variation trend of S_{max}/H_0 against Cr , h_b/L_0 , and h_{toe}/L_0 , Ir_b , respectively, when the other input parameters contributing to Equation (15) have fixed amounts. Figure 6(a) indicates the S_{max}/H_0 variation when Cr varies from 0.270 to 0.561 when the h_b/L_0 , h_{toe}/L_0 ,

H_b/L_0 , and Ir_b is equal to 0.017, 0.006, 0.016, and 0.258, respectively. As expected, Figure 6(a) indicates a reduction in the relative scour depth with an increase in the reflection coefficient, which demonstrates that the proposed equation (Equation (15)) is in line with the physical facts and the results of the laboratory studies of Sutherland et al. (2006) and Tsai et al. (2009). Similarly, the effect of the relative water depth at the toe (h_{toe}/L_0) on the relative maximum scour depth (S_{max}/H_0) has been investigated. To perform this, the Cr , h_b/L_0 , H_b/L_0 , and Ir_b were considered to be 0.349, 0.061, 0.066, and 0.78, respectively, and h_{toe}/L_0

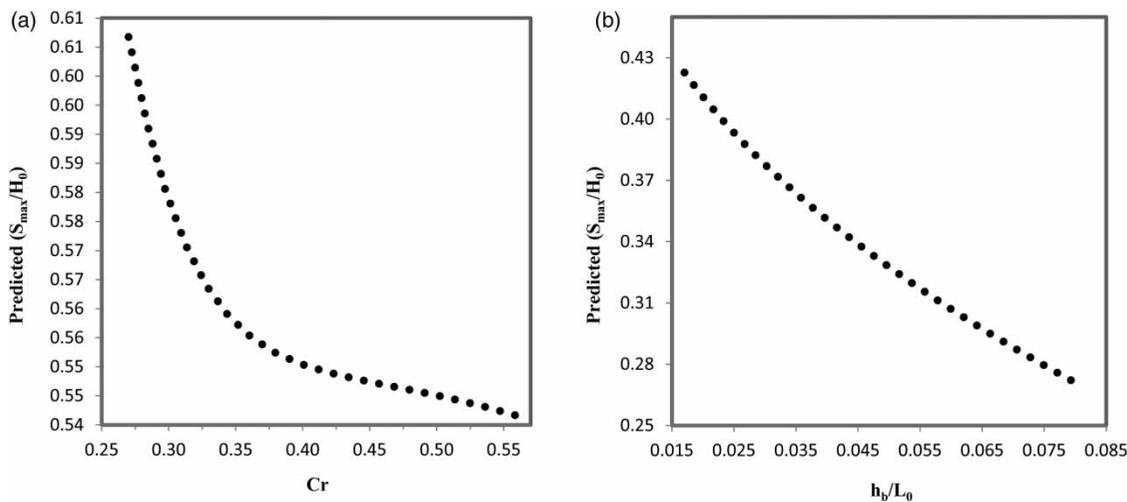


Figure 6 | Variation of the predicted scour depth, using the GP evolved model (Equation (15)): (a) against the reflection coefficient for h_b/L_0 , h_{toe}/L_0 , H_b/L_0 , and Ir_b equal to 0.017, 0.006, 0.016, and 0.258, respectively, and (b) against h_b/L_0 for Cr , h_{toe}/L_0 , H_b/L_0 , and Ir_b equal to 0.4, 0.037, 0.039, and 0.066, respectively.

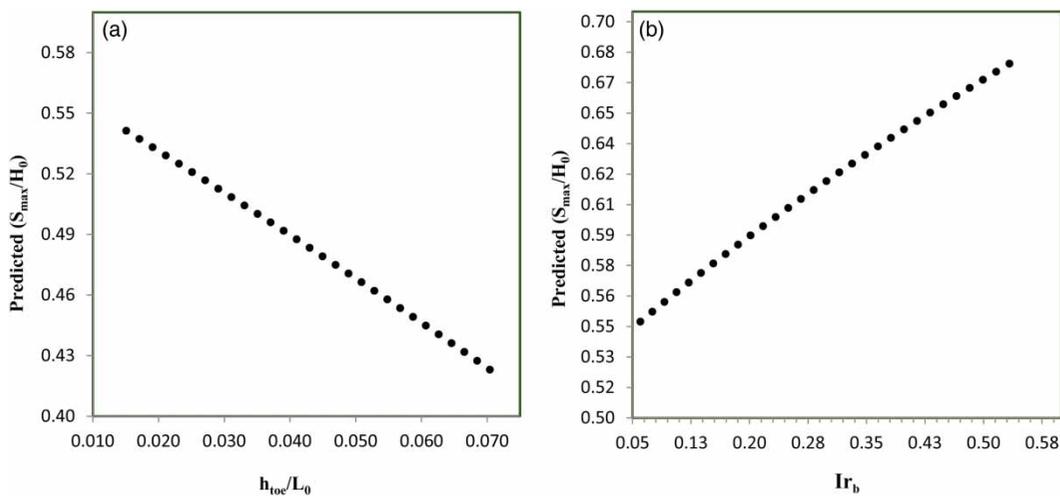


Figure 7 | The variation of the predicted scour depth, using the GP evolved model (Equation (15)): (a) against the relative water depth at the toe (h_{toe}/L_0) for Cr , h_b/L_0 , H_b/L_0 , and Ir_b equal to 0.349, 0.061, 0.066, and 0.78, respectively, and (b) against Ir_b for Cr , h_{toe}/L_0 , H_b/L_0 , and h_b/L_0 equal to 0.27, 0.006, 0.016, and 0.017, respectively.

varies in the range of 0.015–0.071. As seen in Figure 7(a), similar to the results of Fowler (1992) and Sutherland *et al.* (2006), the predicted relative scour depths resulting from Equation (15) were reduced by increasing the relative water depth at the toe. Thus, it is obvious that the proposed equation (Equation (15)) is in line with the existing formulas and the developed concepts.

In order to achieve the models' reliability, a box plot of various models' discrepancy is very useful (Etemad-Shahidi & Ghaemi 2011). As indicated in Figure 8, it was found that the empirical equations of Fowler (1992) and Sutherland *et al.* (2006) are more conservative when compared with the soft computing evolved models. Moreover, Figure 8 shows that the lower and upper quartiles of data average in empirical equations have significantly different numbers which indicate their uncertainty. This gap is not as large as it is in soft computing models. Having larger box height, the empirical formulas need larger safety factors to cover all the range of predicted scour depths. Nevertheless, the soft computing evolved models are more accurate and more reliable when compared to empirical formulas.

As mentioned earlier, the GP evolved equations are very accurate and physically sound. Nevertheless, as indicated in Equation (15), the GP models are mathematically complicated, so their interpretation is not as easy as the empirical formulas. The main reason for producing complicated equations by GP is due to the GP concept. For evolving models, GP uses a set of input parameters, operations, and functions. The best combination of the mentioned variables results in a GP final model. Moreover, ANN is like a black box model. However, the GP and ANN models' accuracy and capability in predicting various phenomena have been proven in many studies. The GP evolved model (Equation (15)) is completely meaningful, in such a way that the variation trend of maximum scour depth against input parameters are closely aligned with the experimental results (Figures 6 and 7). Moreover, it is comparable with the existing empirical formulas from the viewpoint of accuracy, applicability, and reliability. Regarding the mentioned factors, the GP evolved model (Equation (15)) can be seen as a better option to the existing empirical equations and can be complementary to them in predicting maximum scour depth under the action of breaking waves.

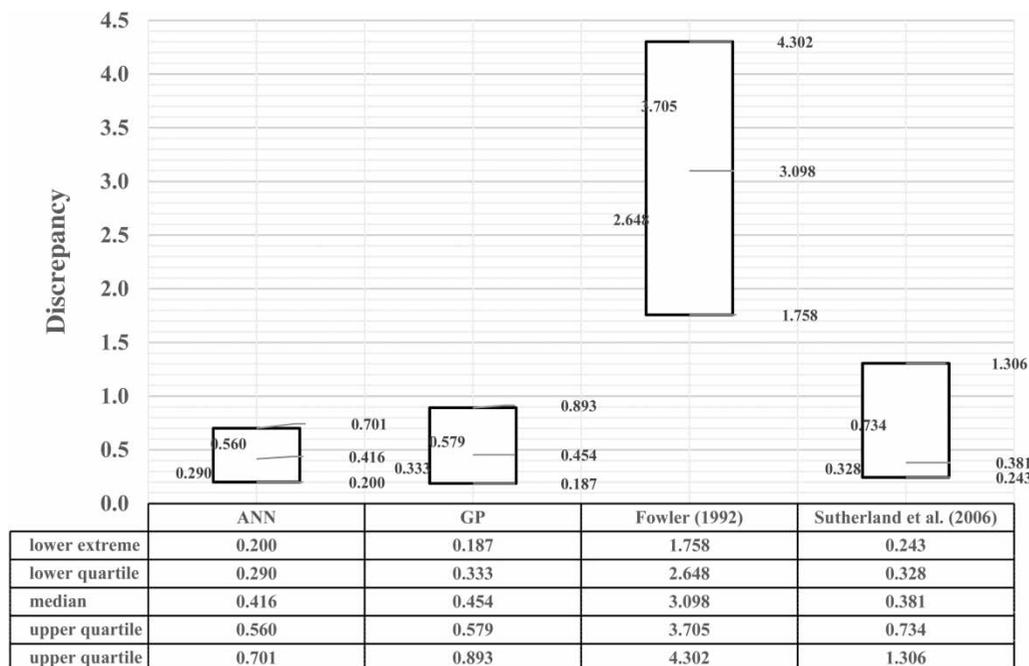


Figure 8 | Box plot of evolved models and empirical formulas.

Sensitivity analysis

Sensitivity analysis and the determination of the relative significance of the input parameters are important issues in developing an accurate prediction model. Employing irrelevant or insignificant input parameters can lead to such complex models that are very difficult to evaluate and interpret. For sensitivity analysis, many studies have employed the exclusion of parameters' approach, where the significance of each parameter was separately evaluated through statistical measures (e.g., Etemad-Shahidi & Ghaemi 2011). One of the major capabilities of GP is its inherent power in determining the significant variables and gradually omitting the input variables that do not contribute beneficially to the models' accuracy (Jayawardena et al. 2005). Several models have been developed with different combinations of input parameters to achieve the best GP output. The sensitivity analysis of developed GP models indicated that S_{max}/H_0 is mostly affected by Cr and followed by h_b/L_0 , h_{toe}/L_0 , Ir_b , and H_b/L_0 , respectively. To study the sensitivity of the developed GP models, Liong et al.'s (2002) approach is implemented, where only one input parameter varies while the others are constant and a variation of ± 15 , ± 10 , ± 5 for each input parameter is considered at each stage. The influence of the modification procedures on the proposed formula for prediction of S_{max}/H_0 is measured in terms of average percentage change (APC) as:

$$APC = \frac{100}{N} \sum_{i=1}^N \left\{ \frac{(S_{max}/H_0)_{org} - (S_{max}/H_0)_{mod}}{(S_{max}/H_0)_{org}} \right\}_i \quad (16)$$

where $(S_{max}/H_0)_{org}$ is the predicted relative scour depth proposed by GP using the original values of the input variables,

and $(S_{max}/H_0)_{mod}$ is the modified GP predicted relative scour depth due to the variation of a particular variable and N is the number of data points. The procedure is repeated for all of the input variables. The significance of input parameters resulting from sensitivity analysis is presented in Table 5.

SUMMARY AND CONCLUSION

This study explored the capabilities of GP and ANN methods for predicting the broken wave-induced scour depth at seawalls. The laboratory data sets of Fowler (1992), Sutherland et al. (2006), and Tsai et al. (2009) were used for developing the models. Statistical error measures were utilized for determining the performance of the GP and ANN models and comparing it with the empirical formulas. The result obtained clearly shows that the GP and ANN models are more accurate when compared to the empirical relations, while the relative scour depth proposed by the GP model achieved higher accuracy than those of the ANN predictions. In addition to higher accuracy, the main advantage of GP, unlike ANNs, is to make predictions by generating simple and meaningful mathematical expression, which can be utilized in predicting the scour depth for a wide range of data.

Further analyses of the results reveal that the effect of the normalized mean diameter of sediment (d_{50}/H_0) on the relative scour depth (S_{max}/H_0) was negligible. Therefore, the prediction models were developed utilizing the effective parameters on the relative scour depth (S_{max}/H_0) including the relative water depth at the toe (h_{toe}/L_0), the reflection coefficient (Cr), the relative water depth at the breaking

Table 5 | Average percentage change (APC) in relative scour depth due to variation of variable values

| Considered variable | Percentage change in variable (Equation (16)) | | | | | | Significance order |
|---------------------|---|--------|--------|--------|--------|--------|--------------------|
| | - 15 | - 10 | - 5 | + 5 | + 10 | + 15 | |
| Cr | -10.399 | -6.485 | -3.076 | 2.897 | 5.765 | 8.763 | 1 |
| Ir_b | 2.312 | 1.532 | 0.762 | -0.753 | -1.498 | -2.234 | 4 |
| H_b/L_0 | -0.457 | -0.288 | -0.136 | 0.123 | 0.236 | 0.338 | 5 |
| h_b/L_0 | -3.889 | -2.548 | -1.253 | 1.214 | 2.392 | 3.537 | 2 |
| h_{toe}/L_0 | -3.144 | -2.109 | -1.061 | 1.075 | 2.165 | 3.269 | 3 |

point (h_b/L_0), the normalized broken wave height (H_b/L_0), and the breaking surf similarity parameter (I_{rb}). The proposed formula by GP fits well with the developed concepts in laboratory studies. The significance of the input parameters on the relative scour depth was evaluated with a sensitivity analysis. The results of sensitivity analysis show that the relative scour depth (S_{max}/H_0) is mainly influenced by the reflection coefficient, and the relative water depth at the breaking point. Despite better performance of the GP model in comparison to that of the ANN predictions of relative scour depth, both the GP and ANN models are promising techniques for predicting the broken wave-induced scour at seawalls. However, one of the obvious limitations of this work is that quite a small data set was used. It can be recommended to carry out studies with larger data sets when they become available.

ACKNOWLEDGEMENTS

The authors would like to extend their special thanks to Professor Solomatine as the Editor of Journal of Hydroinformatics for handling this paper. In addition, the authors would like to sincerely thank Professors J. Fowler, J. Sutherland, C. Obhrai, R. J. S. Whitehouse, A. M. C. Pearce, C. P. Tsai, and H. B. Chen for providing their valuable experimental data for this paper.

REFERENCES

- Aytek, A. & Kişi, Ö. 2008 A genetic programming approach to suspended sediment modelling. *Journal of Hydrology* **351**, 288–298.
- Azamathulla, H. M. 2012a Gene expression programming for prediction of scour depth downstream of sills. *Journal of Hydrology* **460**, 156–159.
- Azamathulla, H. M. 2012b Gene-expression programming to predict scour at a bridge abutment. *Journal of Hydroinformatics* **14** (2), 324–331.
- Azamathulla, H. M. & Ahmad, Z. 2012 GP approach for critical submergence of intakes in open channel flows. *Journal of Hydroinformatics* **14** (4), 937–943.
- Azamathulla, H. M., Ghani, A. A., Zakaria, N. A., Lai, S. H., Chang, C. K., Leow, C. S. & Abohasan, Z. 2008 Genetic programming to predict ski-jump bucket spillway scour. *Journal of Hydroinformatics* **20** (4), 477–484.
- Azamathulla, H. M., Ghani, A. A., Zakaria, N. A. & Guven, A. 2010 Genetic programming to predict bridge pier scour. *Journal of Hydraulic Engineering* **136**, 165–169.
- Azamathullah, H. M., Deo, M. C. & Deolalikar, P. B. 2005 Neural networks for estimations of scour downstream of ski-jump bucket. *Journal of Hydraulic Engineering* **131** (10), 898–908.
- Chen, B. 2006 The numerical simulation of local scour in front of a vertical-wall breakwater. *Journal of Hydrodynamics Ser. B* **18** (3, Supplement), 134–138.
- De Best, A., Bijker, E. W. & Wichers, J. E. W. 1971 Scouring of sand in front of a vertical breakwater. In: *Conference on Port and Ocean Engineering under Arctic Conditions*, Norwegian Institute of Technology, Trondheim, Norway, pp. 1077–1086.
- Etemad-Shahidi, A. & Ghaemi, N. 2011 Model tree approach for prediction of pile groups scour due to waves. *Ocean Engineering* **38** (13), 1522–1527.
- Fallah-Mehdipour, E., Bozorg Haddad, O. & Mariño, M. A. 2013 Developing reservoir operational decision rule by genetic programming. *Journal of Hydrodynamics* **15** (1), 103–119.
- Ferreira, C. 2006 *Gene Expression Programming, Mathematical Modeling by an Artificial Intelligence*, 2nd edn. Springer-Verlag, Heidelberg, Germany.
- Fowler, J. E. 1992 *Scour Problems and Methods for Prediction of Maximum Scour at Vertical Seawalls*. Technical Report CERC 92-16. Coastal Engineering Research Center, Department of the Army Waterways Experiment Station, Vicksburg, MS, USA.
- GeneXproTools 4.0.954 Enterprise Edition 2006 Available at: <http://www.gepsoft.com>.
- Ghorbani, M. A., Khatibi, R., Aytek, A., Makarynsky, O. & Shiri, J. 2010 Sea water level forecasting using genetic programming and comparing the performance with Artificial Neural Networks. *Computers and Geosciences* **36** (5), 620–627.
- Gislason, K., Fredsoe, J., Mayer, S. & Sumer, B. M. 2000 The mathematical modeling of the scour in front of the toe of a rubble-mound breakwater. In: *Proceedings of the 27th International Coastal Engineering Conference, Sydney, Australia*.
- Goldberg, D. E. 1989 *Genetic Algorithms for Search, Optimization and Machine Learning*. Addison-Wesley Publishing Co., Boston, MA, USA.
- Guven, A. & Azamathulla, H. M. 2012 Gene expression programming for flip bucket spillway scour. *Water Science and Technology* **65** (11), 1982–1987.
- Guven, A., Azamathulla, H. M. & Zakaria, N. A. 2009 Linear genetic programming for prediction of circular pile scour. *Ocean Engineering* **36**, 985–991.
- Hajivalie, F., Yeganeh-Bakhtiary, A., Houshanghi, H. & Gotoh, H. 2012 Euler-Lagrange model for scour in front of vertical breakwater. *Applied Ocean Research* **34**, 96–106.
- Jain, P. & Deo, M. C. 2007 Real-time wave forecasts off the western Indian coast. *Applied Ocean Research* **29**, 72–79.
- Jayawardena, A. W., Muttill, N. & Fernando, T. M. K. G. 2005 Rainfall-runoff modeling using genetic programming. In: *International Congress on Modeling and Simulation, Advances and Applications for Management and Decision*

- Making. Modeling and Simulation Society of Australia and New Zealand, Melbourne, Australia, pp. 1841–1847.
- Kambekar, A. R. & Deo, M. C. 2003 Estimation of pile group scour using neural networks. *Applied Ocean Research* **25** (4), 225–234.
- Kazeminezhad, M. H., Etemad-Shahidi, A. & Yeganeh-Bakhtiary, A. 2010 An alternative approach for investigation of the wave-induced scour around pipelines. *Journal of Hydroinformatics* **12** (1), 51–65.
- Koza, J. 1992 *Genetic Programming, On the Programming of Computers by Natural Selection*. MIT Press, Cambridge, MA, USA.
- Lee, K.-H. & Mizutani, N. 2008 Experimental study on scour occurring at a vertical impermeable submerged breakwater. *Applied Ocean Research* **30** (2), 92–99.
- Lillycrop, W. J. & Hughes, S. A. 1993 *Scour Hole Problems Experienced by the Corps of Engineers, Data Presentation and Summary*. Miscellaneous papers, CERC-93-2. US Army Engineers Waterways Experiment Station, Coastal Engineering Research Center, Vicksburg, MS, USA.
- Liong, S. Y., Gautam, T. R., Khu, S. T., Babaovic, V. & Muttill, N. 2002 Genetic programming, a new paradigm in rainfall-runoff modeling. *Journal of American Water Resources Association* **38** (3), 219–231.
- Oumeraci, H. 1994 Review and analysis of vertical breakwater failure-lessons learned. *Coastal Engineering* **22**, 3–29.
- Poli, T. & McPhee, N. F. 2008 *Covariant Parsimony Pressure for Genetic Programming*. Technical Report CES-480. Department of Computing and Electronic Systems, University of Essex, UK.
- Pourzangbar, A. 2012 Determination of the most effective parameters on scour depth at seawalls using genetic programming (GP). In: *The 10th International Conference on Coasts, Ports and Marine Structures (ICOPMASS 2012)*, Tehran, Iran.
- Pourzangbar, A., Yeganeh-Bakhtiary, M. A. & Hajivalie, F. 2013 Predicting scour depth at the trunk section of the seawalls using artificial neural networks. In: *The 7th National Conference on Civil Engineering*, Zahedan, Iran.
- Rogers, L. L. & Dowla, F. U. 1994 Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. *Water Resources Research* **30** (2), 457–481.
- Shiri, J. & Kişi, Ö. 2011 Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. *Computers and Geosciences* **37** (10), 1692–1701.
- Sumer, B. M. & Fredsøe, J. 2000 Experimental study of 2D scour and its protection at a rubble-mound breakwater. *Coastal Engineering* **40** (1), 59–87.
- Sumer, B. M., Fredsøe, J., Lamberti, A., Zanuttigh, B., Dixen, M., Gislason, K. & Di Penta, A. F. 2005 Local scour at roundhead and along the trunk of low crested structures. *Coastal Engineering* **52**, 995–1025.
- Sutherland, J., Obhrai, C., Whitehouse, R. J. S. & Pearce, A. M. C. 2006 Laboratory tests of scour at a seawall. In: *Proceedings of 3rd International Conference on Scour and Erosion, Gouda, The Netherlands*, (CD-ROM).
- Sydenham, P. H. & Thorn, R. 2005 *Handbook of Measuring System Design*. Vol. 3. John Wiley and Sons, pp. 901–908.
- Tsai, C. P., Chen, H. B. & You, S. S. 2009 Toe scour of seawall on a steep seabed by breaking waves. *Journal of Waterway, Port, Coastal, and Ocean Engineering* **135** (2), 61–68.
- Xie, S. L. 1981 *Scouring Patterns in Front of Vertical Breakwaters and Their Influence on the Stability of the Foundations of the Breakwaters*. Department of Civil Engineering, Delft University of Technology, The Netherlands, p. 61.
- Xie, S. L. 1985 Scouring patterns in front of vertical breakwaters. *Acta Oceanologica Sinica* **4** (1), 153–164.
- Yeganeh-Bakhtiary, A., Ghorbani, M. A. & Pourzangbar, A. 2012 Determination of the most important parameters on scour at coastal structures. *Journal of Civil Engineering and Urbanism* **2** (2), 68–71.
- Zanganeh, M., Yeganeh-Bakhtiary, A. & Bakhtyar, R. 2011 Combined Particle Swarm Optimization and Fuzzy Inference System model for estimation of current-induced scour beneath marine pipelines. *Journal of Hydroinformatics* **13** (3), 558–573.

First received 17 June 2015; accepted in revised form 9 December 2016. Available online 2 February 2017