

# Gene-expression programming to predict scour at a bridge abutment

H. Md. Azamathulla

## ABSTRACT

The process involved in the local scour at an abutment is so complex that it makes it difficult to establish a general empirical model to provide accurate estimation for scour. This study presents the use of gene-expression programming (GEP), which is an extension of genetic programming (GP), as an alternative approach to estimate the scour depth. The datasets of laboratory measurements were collected from the published literature and used to train the network or evolve the program. The developed network and evolved programs were validated by using the observations that were not involved in training. The proposed GEP approach gives satisfactory results compared with existing predictors and artificial neural network (ANN) modeling in predicting the scour depth at an abutment.

**Key words** | artificial neural networks, bridge abutments, gene-expression programming, local scour, radial basis function

**H. Md. Azamathulla**  
River Engineering and Urban Drainage Research  
Centre (REDAC),  
Universiti Sains Malaysia,  
Engineering Campus,  
Seri Ampangan,  
14300 Nibong Tebal,  
Pulau Pinang,  
Malaysia  
E-mail: redacazamath@eng.usm.my;  
mdazmath@gmail.com

## NOTATION

$L$	length of abutment normal to the approaching flow
$F$	Froude number
$U$	mean velocity of approach flow
$U_c$	mean velocity of approach flow at the incipient motion of sediment
$b$	streamwise length of abutment (m)
$h$	depth of the approaching flow
$d_{50}$	particle mean diameter
$d_s$	equilibrium scour depth
$g$	gravitational acceleration
$R^2$	coefficient of determination
RMSE	root mean squared error
MAE	mean average error
$\rho$	fluid density
$\rho'_s$	buoyant sediment density
$\alpha$	spread
$\delta$	average absolute deviation.

doi: 10.2166/hydro.2011.135

## INTRODUCTION

Scour is a major cause of the failure of bridge abutments. Failure of bridges due to scour at their foundations consisting of abutments and piers is a common occurrence. Local scour at the foundations has long been a concern for engineers (Muzzammil 2010). In the safety evaluation of bridges, the local scour of bridge foundation material near a pier/abutment is therefore a significant issue (Azamathulla *et al.* 2010).

Interactions between the abutment and its erodible bed under strong current and/or wave conditions may cause scour at the bridge abutment. This process involves the complexities of both the three-dimensional flow pattern and sediment movement. Scour underneath the bridge abutment may expose a section of the bridge, causing it to become unsupported, leading to settlement and potentially structural failure. Generally the scour depth at abutments is predicted by three strategies such as the regime approach, the empirical approach, and the analytical or semi-empirical

approach. Even though many experimental and theoretical works were reported on scour prediction, there is a wide scope for further study in many applications. From the available literature, it is also revealed that the exact scour mechanism and effects of different parameters on scour depth are yet to be fully understood or explored (Dey & Barbhuya 2004a, b). The estimation of the scour characteristics at a bridge abutment continues to be a concern for hydraulic engineers.

A number of empirical formulae have been developed in the past to estimate the equilibrium scour depth at a bridge abutment, including Bateman *et al.* (2005) who developed a morphodynamic model to predict temporal evolution of local scour at bridge piers. There is therefore no single analytically derived equation which is valid for a wide range of flow conditions, bed material properties and abutment shape configurations, because of the difficulties of precise modeling of the phenomenon in a laboratory medium. Lack of understanding of complex flow conditions and simplified modeling of the phenomenon would lead to the pronounced modeling uncertainty. On the other hand, reliable field data are scarce, leading to calibration problems. Most commonly, regression relations are used to predict scour at a bridge abutment; however, regression analysis can have large uncertainties, which have major drawbacks pertaining to idealization of the complex scour process, and approximation and averaging widely varying prototype conditions. Thus, the computed scour depths can be far from the actual ones. Another important issue, apart from the complexity of the scour phenomenon involved, is due to the limitation of the regression analysis.

Predictive approaches such as artificial neural networks (ANN) (Azamathullah *et al.* 2005) and adaptive neuro-fuzzy inference systems (ANFIS) (Azamathulla *et al.* 2008; Muzzammil 2010) have been recently shown to yield effective estimates of scour around hydraulic structures. ANNs have been reported to provide reasonably good solutions for hydraulic-engineering problems, particularly for cases of the highly nonlinear and complex relationship among the input-output pairs in corresponding data (Azamathulla *et al.* 2010).

The objective of this study is to develop a predictive model for scour depth using GEP. The performance of the proposed GEP model is compared with a standard Radial

Basis Function (RBF) neural network and conventional regression-based equations. The explicit formulation of the GEP model is also presented.

## ANALYSIS OF LOCAL SCOUR AT BRIDGE ABUTMENT

The variables influencing the equilibrium scour depth ( $d_s$ ) at a bridge abutment perpendicular to the shoreline placed in uniform bed sediments are generally expressed in the following functional form, assuming a constant relative density of sediment and the absence of viscous effects (Dey & Barbhuya 2004a):

$$d_s = f(U, L, h, U_c, d_{50}, K_s) \quad (1)$$

where  $L$  is the length of the abutment perpendicular to the flow direction,  $h$  is the depth of the approach flow,  $U$  is the mean flow velocity,  $U_c$  is the critical velocity of bed sediment,  $d_{50}$  is the median size of the sediment and  $d_s$  is the equilibrium scour depth.  $K_s$  represents the abutment shape factor, being 1 for vertical-wall abutments, 0.82 for 45° wing-wall abutments and 0.75 for semicircular abutments (Melville 1992). As the scour depth at an abutment occurs when the excess approaching flow velocity ( $U_e$ ) is greater than zero, where  $U_e = U - 0.5U_c$ , Equation (1) may therefore be expressed as (Muzzammil 2010)

$$d_s = f(U_e, L, h, d_{50}, K_s). \quad (2)$$

Equation (2) may also be reduced in terms of a set of non-dimensional parameters of the form:

$$\frac{d_s}{L} = \Psi \left( F_e, \frac{h}{L}, \frac{d_{50}}{L}, K_s \right) \quad (3)$$

where  $F_e = U_e / (\Psi g L)^{0.5}$ ;  $\Psi = S - 1$  and  $S$  is the relative density of sediment particles.

The non-dimensional parametric representation in the present model has been justified by Dey & Barbhuya (2004a) for the different conditions of flows (Figure 1). The experimental data were collected from Dey & Barbhuya (2004a) and Ballio *et al.* (2009). The whole dataset consisting

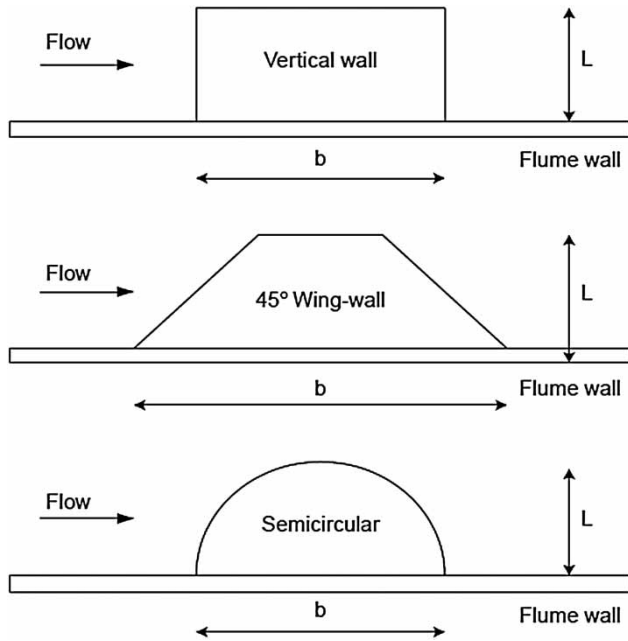


Figure 1 | Different types of abutments (Muzzammil 2010).

of 317 dataset. Figure 2 shows the definition figure of scour at a typical abutment. The time for running the experiment is generally considered to be an important variable of interest to avoid erroneous equilibrium scour depth. Lim & Chiew (2001) reported that the required time to reach the equilibrium scour at abutments in clear water scour is 3 to 8 d, depending on the flow and sediment conditions. Melville (1992) defined the time to reach the equilibrium scour condition such that the rate of increase of scour depth does not exceed 5% of the pier diameter in the succeeding 24 h period. Dey & Barbhuya (2004a) reported that, when a negligible (1 mm or less) difference of scour depth of a particular run was observed in an interval of 2 h after 48 h, it

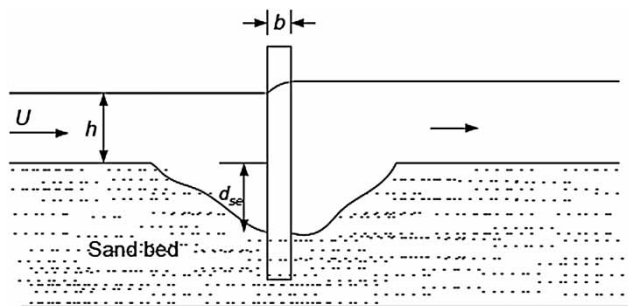


Figure 2 | Definition sketch of scour at a typical abutment (Muzzammil 2010).

was assumed that an equilibrium state has been achieved (Muzzammil 2010).

During the last two decades, researchers have noticed that the use of soft computing techniques as an alternative to conventional statistical methods based on controlled laboratory or field data yielded significantly better results. ANN and GP/GEP are the most widely used branches of soft computing in hydraulic engineering (Giustolisi 2004; Kizhisseri *et al.* 2005). Within the larger field of hydraulics, several researchers have dealt with the scour around and downstream of hydraulic structures using ANN (Azmathulla *et al.* 2005, 2006, 2008; Guven & Gunal 2008; Muzzammil 2010). Gene-expression programming (GEP), which is an extension of GP, has recently attracted the attention of researchers in the prediction of hydraulic characteristics. This study presents ANN and GEP as an alternative tool in the prediction of scour depth at a bridge abutment.

## OVERVIEW OF GEP

GEP, which is an extension of GP (Koza 1999), is a search technique that involves computer programs (e.g. mathematical expressions, decision trees, polynomial constructs and logical expressions). GEP computer programs are all encoded in linear chromosomes, which are then expressed or translated into expression trees (ETs). ETs are sophisticated computer programs that have usually evolved to solve a particular problem and are selected according to their fitness at solving that problem.

GEP is a full-fledged genotype/phenotype system, with the genotype totally separated from the phenotype, whereas in GP, genotype and phenotype are mixed together in a simple replicator system. As a result, the fully fledged genotype/phenotype system of GEP surpasses the old GP system by a factor of 100–60,000 (Ferreira 2001a, b).

Initially, the chromosomes of each individual in the population are generated randomly. Then, the chromosomes are expressed, and each individual is evaluated based on a fitness function and selected to reproduce with modification, leaving progeny with new traits. The individuals in the new generation are, in their turn, subjected to some developmental processes, such as expression of the genomes, confrontation of the selection environment and

reproduction with modification. These processes are repeated for a predefined number of generations or until a solution is achieved (Ferreira 2008a, b). The functionality of each genetic operator included in the GEP system has been explained by Guven & Aytek (2009).

## DEVELOPMENT OF NEURAL NETWORK MODEL

ANNs provide a random mapping between an input and an output vector, typically consisting of three layers of neurons, namely input, hidden and output, with each neuron acting as an independent computational element. Neural networks derive their strengths from the high degree of freedom associated with their architecture. Prior to application, the network is trained to observed datasets from the published data and Professor Dr Francesco Ballio of Politecnico di Milano, Milano, Italy also kindly provided additional scour data resulting from his previous work. This feeds the network with input and output pairs and determines the values of connection weights, bias or centers (see Figure 3 for an example). The training may require many epochs (presentation of complete datasets once to the network), being carried out until the training sum of square errors reaches a specified error goal. Concepts involved behind these training schemes are outlined in the ASCE Task Committee (2000). A neural network toolbox contained within the MATLAB package was used in this study. The usual feed-forward type of network was trained using radial basis function (RBF). Out of the total of 317 input-output pairs, about 75% (238 sets), selected randomly, and were used for training, whereas the remaining 25% (79 sets) were employed for testing. As dictated by the use of a Gaussian function, all patterns were normalized within the range of (0.0, 1.0) before their use. The RBF

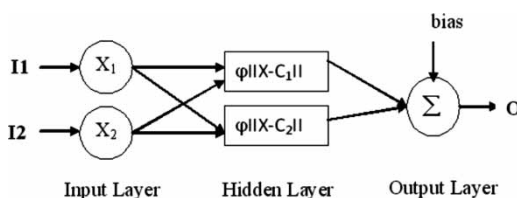


Figure 3 | RBF neural network architecture.

network (four inputs, 36 hidden neurons and one output) (as in Equation (3)) was trained by using various values of spread ( $\alpha$ ) between 0 and 1. A spread constant  $\alpha$  for the radial basis layer returns a network with weights and biases such that the outputs are exactly for given targets. The value of 0.01 was selected as it yielded the best performance for the training data.

## DEVELOPMENT OF GEP MODEL

In this section, the scour depth at an abutment is modeled using the GEP approach. Initially, the ‘training set’ is selected from the entire dataset and the rest is used as the ‘testing set’. Once the training set is selected, one could say that the learning environment of the system is defined. The modeling also includes five major steps to prepare to use GEP. The first is to choose the fitness function. For this problem, the fitness,  $f_i$ , of an individual program,  $i$ , is measured by

$$f_i = \sum_{j=1}^{C_i} (M - |C_{(i,j)} - T_j|) \quad (4)$$

where  $M$  is the range of selection,  $C_{(i,j)}$  is the value returned by the individual chromosome  $i$  for fitness case  $j$  (out of  $C_i$  fitness cases) and  $T_j$  is the target value for fitness case  $j$ . If  $|C_{(i,j)} - T_j|$  (the precision)  $\geq 0.01$ , then the precision is 0 and  $f_i = f_{\max} = C_i M$ . In this case,  $M = 100$  is used; therefore,  $f_{\max} = 1,000$ . The advantage of this kind of fitness function is that the system can find the optimal solution by itself.

Second, the set of terminals  $T$  and the set of functions  $F$  are chosen to create the chromosomes. In this problem, the terminal set consists of single independent variable, i.e.,  $T = \{h\}$ . The choice of the appropriate function set is not so clear; however, a good guess is helpful if it includes all the necessary functions. In this study, four basic arithmetic operators (+, -, \*, /) and some basic mathematical functions ( $v$ ) are utilized.

The third major step is to choose the chromosomal architecture, i.e., the length of the head and the number of genes. Initially we used a single gene and two head lengths and increased the number of genes and heads one at a time during each run while we monitored the training and testing

performances of each model. It was observed that more than two genes and a head length greater than 8 did not significantly improve the training and testing performance of GEP models. Thus, the head length,  $l_h = 8$ , and three genes per chromosome are employed for each GEP model in this study.

The fourth major step is to choose the linking function. In this study, addition and multiplication operators are used as linking functions, and it is observed that linking the sub-ETs by addition gives better fitness (Equation (4)) values. The fifth and final step is to choose the set of genetic operators that cause variation and their rates. A combination of all genetic operators (mutation, transposition and cross-over) is used for this purpose (Table 2).

The GEP model was developed using the same input variables as with an ANN-RBF model as parameters in

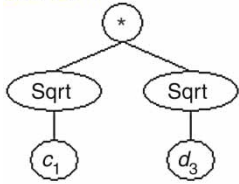
Equation (3), namely Froude number, the relative flow depth, relative sediment particle size,  $K_s$  is the shape factor and normalized equilibrium scour depth ( $d_s/L$ ) as the input and output patterns, respectively. Both of these combinations of inputs have been used for the GEP and ANN models.

The simplified analytic form of the proposed GEP model may be expressed as

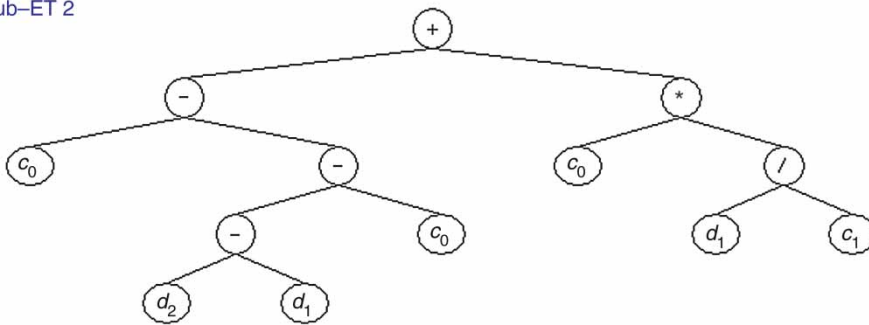
$$\frac{d_s}{L} = \sqrt{4.34K_s} + \left( -0.99 - \left\langle \left( \frac{h}{L} - F_e \right) + 0.99 \right\rangle + (-0.99) \frac{F_e}{1.97} \right) + \left[ \frac{h}{L} + \frac{(-3.78/5.45)}{K_s + (d_{50}/L)} \right] \frac{1}{(\sqrt{(h/L)})/(h/L)} \quad (5)$$

and the corresponding expression trees are shown in Figure 4.

Sub-ET 1



Sub-ET 2



Sub-ET 3

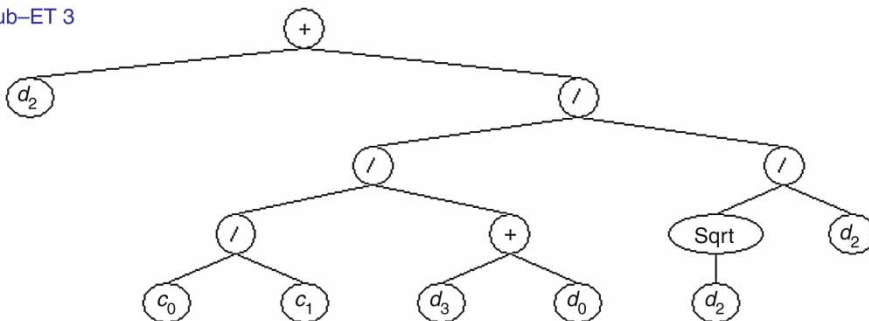


Figure 4 | Expression tree (ET) for GEP formulation ( $d_0 = d_{50}/L$ ,  $d_1 = F$ ,  $d_2 = h/L$  and  $d_3 = K_s$ ).

## RESULTS AND DISCUSSION

In this study, different combinations of input data (non-dimensional dataset) were explored to assess their influence on the scour depth modelling (Table 3). The GEP model was developed and tested for predicting abutment scour depth.

**Table 1** | Range for data of scour parameters used for estimation of equilibrium scour depth

Item	Parameters	Range
1	Abutment length $L$ (m)	0.04–0.4
2	Flow depth $h$ (m)	0.058–0.25
3	Mean velocity $U$ (m/s)	0.219–0.72
4	Sediment size $d_{50}$ (mm)	0.26–5
5	Scour depth $d_s$ (m)	0.053–0.47 0.053–0.29
6	Shape factor (Melville 1992)	1 for rectangular, 0.82 for 45° wing wall and 0.75 for semicircular abutment.

**Table 2** | Parameters of the optimized GEP model

Parameter	Description of parameter	Setting of parameter
$P_1$	Function set	+, -, *, /, v
$P_2$	Mutation rate (%)	44
$P_3$	Inversion rate (%)	10
$P_4$	One-point and two-point recombination rate, respectively (%)	30,30
$P_5$	Gene recombination rate	0.1
$P_6$	Gene transportation rate	0.1

**Table 3** | Sensitivity analysis for independent parameters for the testing set

Model	RMSE	MAE	$R^2$
$\frac{d_s}{L} = \Psi\left(F_e, \frac{h}{L}, \frac{d_{50}}{L}, K_s\right)$	0.845	0.656	0.89
$\frac{d_s}{L} = \Psi\left(F_e, \frac{h}{L}, \frac{d_{50}}{L}\right)$	0.935	0.845	0.85
$\frac{d_s}{L} = \Psi\left(F_e, \frac{h}{L}, K_s\right)$	0.986	0.943	0.83
$\frac{d_s}{L} = \Psi\left(F_e, \frac{d_{50}}{L}, K_s\right)$	0.918	0.726	0.741

## Training and testing results of GEP modeling

The performance of GEP in training and testing sets is validated in terms of the common statistical measures  $R^2$  (coefficient of determination), RMSE (root mean square error), MAE (mean average error) and  $\delta$  (average absolute deviation).

Table 1 shows the range of variation of collected data for this study and its parameters. The functional set and operational parameters used in GEP modeling during this study are listed in Table 2.

The performance of all models was compared using four error measures:

$$R^2 = 1 - \frac{\sum_{i=1}^N (o_i - t_i)^2}{\sum_{i=1}^N (o_i - \bar{o}_i)^2} \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (o_i - t_i)^2}{N}} \quad (7)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |o_i - t_i| \quad (8)$$

$$\delta = \frac{\sum (o_i - t_i)}{\sum o_i} * 100 \quad (9)$$

where  $t_i$  denotes the target values of equilibrium scour depth (cm), while  $o_i$  and  $\bar{o}_i$  denotes the observed and averaged observed values of equilibrium scour depth (cm), respectively, and  $N$  is the number of data points.

Table 3 compares the GEP model with one of the independent parameters removed in each case and any independent parameter from the input set that yielded larger RMSE, MAE and lower  $R^2$  values also removed. These four independent parameters affect  $d_s/L$ ; thus, the functional relationship given in Equation (3) is used for the GEP model in this study. The GEP approach resulted in a highly nonlinear relationship between  $d_s/L$  and the input parameters and the GEP model had the highest accuracy and the lowest error (Table 3). The testing performance of the proposed GEP model revealed a high generalization capacity with  $R^2 = 0.89$ , RMSE = 0.845, MAE = 0.656 and  $\delta = 7.8$ .

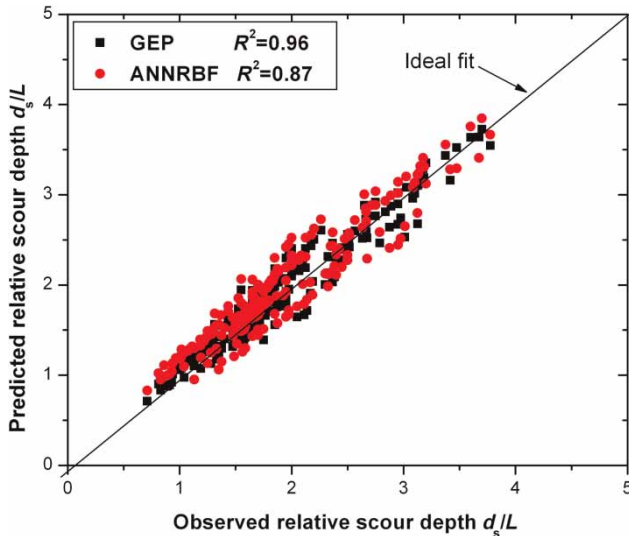


Figure 5 | Observed versus predicted relative scour depth (training).

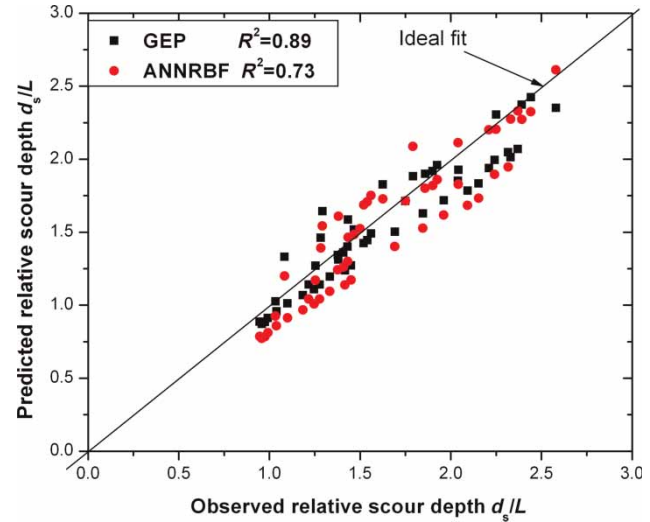


Figure 6 | Observed versus predicted scour depth: validation (testing).

Table 4 | Comparison of models for non-dimensional set performance of the GEP and ANN-RBF

Models	$R^2$		RMSE		MAE		$\delta$	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
GEP	0.96	0.89	0.29	0.845	0.279	0.656	3.7	7.8
ANN-RBF	0.87	0.73	0.48	1.073	0.083	0.071	11.45	15.67

A non-dimensional parameter in the Equation (3) sensitivity analysis shows that the dimensionless shape factor parameter ( $K_s$ ) and  $d_{50}/L$  have respectively the most and the least effect on normalized scour depth.

To assess the performance of the GEP model, observed equilibrium scour depth values were plotted against the predicted ones. Figures 5 and 6 illustrate the results with the performance indices between predicted and observed data for the training and validating (testing) datasets for dimensional parameters. The result of the grouped variables combination data shows a high coefficient of determination ( $R^2 = 0.96$ ); also RMSE (= 0.546) in the case of ANN-RBF has  $R^2 = 0.87$  and RMSE = 0.48 in both training and validation periods but this variation is low compare with  $R^2$  variation (Table 4). The results of an ANFIS-based approach for prediction of scour depth by Muzzammil (2010) are also interesting but not produced any expression like Equation (5) for general use in designs.

## CONCLUSION

The application of the relatively new soft computing approach of genetic programming to predict the local scour depth at an abutment of a bridge is described. A GEP and an ANN-RBF model were developed to predict the values of relative scour depth from the laboratory measurements. A new approach was presented to estimate equilibrium depth of scour at a bridge abutment from optimum datasets with the GEP and ANN modeling techniques. The application of the GEP in this study is another important contribution to scour-depth estimation methodologies for bridges. The present study indicates that employing the original dataset yielded a network that can predict measured scour depth at bridge abutments more accurately than standard regression analysis. The overall performance of the GEP model is superior to the ANN model.

## ACKNOWLEDGEMENTS

The author wishes to express his sincere gratitude to Universiti Sains Malaysia for funding a short-term grant to conduct this ongoing research (Grant No. 304.PREREDAC.6035262).

## REFERENCES

- American Society of Civil Engineers (ASCE) Task Committee 2000 The ASCE Task Committee on Application of artificial neural networks in hydrology. *J. Hydrol. Engng.* **5** (2), 115–137.
- Azamathulla, H. Md., Deo, M. C. & Deolalikar, P. B. 2005 Neural networks for estimation of scour downstream of ski-jump bucket. *J. Hydraul. Engng.* **131** (10), 898–908.
- Azamathulla, H. Md., Deo, M. C. & Deolalikar, P. B. 2006 Estimation of scour below spillways using neural networks. *IAHR, J. Hydraul. Res.* **44** (1), 61–69.
- Azamathulla, H. Md., Deo, M. C. & Deolalikar, P. B. 2008 Alternative neural networks to estimate the scour below spillways. *Adv. Engng. Software* **39** (8), 689–698.
- Azamathulla, H. Md., Ghani, A. A., Zakaria, N. A. & Aytac, G. 2010 Genetic programming to predict bridge pier scour. *J. Hydraul. Engng.* **136** (3), 165–169.
- Ballio, F., Teruzzi, A. & Radice, A. 2009 Constriction effects in clear-water scour at abutments. *J. Hydraul. Engng.* **135** (2), 140–145.
- Bateman, A., Fernandez, M. & Parker, G. 2005 Morphodynamic model to predict temporal evolution of local scour at bridge piers. In *Proc. 4th IAHR Symposium on River, Coastal and Estuarine Morphodynamics (RCEM 2005), Urbana, Illinois, 4–7 October*. Taylor & Francis, London, pp. 311–920.
- Dey, S. & Barbhuiya, A. K. 2004a Clear water scour at abutments. *Proc. Inst. Civil Engng. Wat. Mngmnt.* **157**, 77–97.
- Dey, S. & Barbhuiya, A. K. 2004b Clear water scour at abutments in thinly armored beds. *J. Hydraul. Engng.* **130** (7), 622–634
- Ferreira, C. 2001a Gene expression programming in problem solving. In *Proc. 6th Online World Conference on Soft Computing in Industrial Applications*. Invited tutorial, pp. 10–24.
- Ferreira, C. 2001b Gene expression programming: a new adaptive algorithm for solving problems. *Complex Syst.* **13** (2), 87–129.
- Güven, A. & Aytak, A. 2009 A new approach for stage-discharge relationship: gene-expression programming. *J. Hydrol. Engng.* **14** (8), 812–820.
- Güven, A. & Günel, M. 2008 Genetic programming approach for prediction of local scour downstream hydraulic structures. *J. Irrigat. Drainage Engng.* **134** (2), 241–249
- Giustolisi, O. 2004 Using genetic programming to determine Chèzy resistance coefficient in corrugated channels. *J. Hydroinf.* **6** (3), 157–173
- Koza, J. R. 1999 *Genetic Programming: On the Programming of Computers by means of Natural Selection*. MIT Press, Cambridge, MA.
- Kizhisseri, A. S., Simmonds, D., Rafiq, Y. & Borthwick, M. 2005 An evolutionary computation approach to sediment transport modeling. In *Proc. 5th International Conference on Coastal Dynamics, Barcelona, Spain, April 4–8*. ASCE, New York.
- Lim, F. H. & Chiew, Y. M. 2001 Parametric study of riprap failure around bridge piers. *J. Hydraul. Res.* **39** (1), 61–72.
- Melville, B. W. 1992 Local scour at bridge abutments. *J. Hydraul. Engng.* **118** (4), 615–631.
- Muzzammil, M. 2010 ANFIS approach to the scour depth prediction at a bridge abutment. *J. Hydroinf.* **12** (4), 474–485.

First received 22 October 2010; accepted in revised form 11 March 2011. Available online 18 June 2011