Evaluation of regular wave scour around a circular pile using data mining approaches

M. K. Ayoubloo, A.Etemad-Shahidi*, J.Mahjoobi

School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran, P.O. Box 16765-163

Tel.: 9821 77240399, Fax: 98 21 77240398

*Corresponding Author, E-mail: etemad@iust.ac.ir

Abstract

An accurate estimation of scour depth around a pile is very difficult due to the complex behavior of flow around a pile structure on an erodible bed. In the current study, Regression Trees (RT) and Artificial Neural Networks (ANNs) as remedy data mining approaches are suggested to estimate the scour depth due to regular waves. These approaches were used to predict normalized scour depth as a function of two separate sets of parameters: (*i*) dimensional parameters and (*ii*) dimensionless parameters. The ANN trained by dimensional parameters provides more accurate results compared to that trained by dimensionless parameters. As opposed to the ANN model, the RT model based on dimensionless input parameters predicts normalized scour depth outperformed the one based on dimensional inputs. In addition, these models outperformed the existing empirical formulae. A committee model based on the geometric mean of the results of RT and ANN (developed by dimensionless parameters) is presented as the best model. To determine relative importance of input parameters in the prediction of the scour depth, a sensitivity analysis was then performed and it was found that the Keulegan-Carpenter number (*KC*) was found to be the most important one. The error statistics for two classes of KC ($KC < 10$ and $KC > 10$) indicated that the suggested approach performs better in the range of $KC < 10$ for the prediction of dimensionless scour depth.

Keywords: Artificial Neural Network; CART; Pile; Regression Tree; Scour depth; Wave

1. Introduction

Piles are one of the most important parts of a hydraulic structure used in pile-deck structures such as bridge piers and offshore platforms. A vertical pile is frequently employed as a foundation to support a hydraulic structure and transfer forces to the bed. The presence of a vertical pile located on an erodible bed changes the flow pattern around the pile. These changes can increase the local sediment transport and can lead to scouring around the pile.

The developed scour can be due to waves and currents or the combination of these two phenomena. One of the first studies on scour around obstructions due to random wave has been carried out by Palmer [1]. He found that the scour is independent of sediment characteristics for the range of studied median grain diameters (0.12 – 0.63 mm). Wang and Herbich [2] investigated scour around a pile due to the combination of wave and current. Sumer et al. [3] conducted an experimental study on the scour around a single circular pile exposed to waves. They conducted three sets of tests and normalized equilibrium scour depth (S) with pile diameter (D). They noted that the scour depth is mainly controlled by the Keulegan-Carpenter (KC) number and represented a formula for scour depth as a function of KC. A field study of the random wave induced scour around a group of piles has been reported by Bayram and Larson [4]. They developed an empirical relationship between scour depth and KC number that agreed with some earlier laboratory experiments. Myrhaug and Rue [5] using a stochastic procedure, suggested some equations for predicting normalized scour depth around piles in random waves. Recently, Sumer et al. [6] conducted experiments on wave scour around a circular pile in three types of soils with different relative densities. Using their data, Guven et al. [7] proposed a linear genetic programming for modeling the scour depth.

An accurate estimation of scour depth is hard to be accomplished by means of empirical equations. In the last decade, investigators have tried to improve the accuracy of scour depth estimation. Artificial Neural Networks (ANNs) have been widely used in hydraulic engineering problems because of their flexibility, ability to generalize and power to approximate nonlinear and complex phenomena. ANNs have been used to estimate scour below spillways [8], scour downstream of a ski-jump bucket [9] and scour downstream of grade-control structures [10]. Recently, Kambekar and Deo [11] used ANNs to estimate piles group scour. Their results indicated that ANNs could be a suitable procedure to predict scour geometries. Estimation of scour properties around a group of piles with feed-forward Multi Layer Perceptron (MLP) has been investigated by Khosronejad et al. [12]. Bateni and Jeng [13] also combined ANNs with Fuzzy Inference System (FIS) to predict the scour depth due to wave around pile groups. However, the application of the regression trees and ANNs to the prediction of scour depth around a single pile has not been tested yet.

Regression trees can be applied to this problem since they are primarily aimed at recognition of a complex pattern in a given set of input values. Regression trees are useful to model an input with the corresponding output. RT has been used for soil properties prediction in environmental science [14], risk management analysis in petroleum pipeline construction [15] and prediction of significant wave height [16]. In this paper, CART algorithm [17] is employed for building and evaluating regression trees. CART builds classification and regression trees for predicting continuous (regression) and categorical predictor variables (classification). This study aims to investigate the skills of the RT and ANN in the prediction of scour depth around a pile due to regular waves and to determine the relative importance of dimensional and dimensionless parameters on the scour process.

2. Data Mining Approaches

2.1. Artificial Neural Network

An Artificial Neural Network (ANN) is a simplified mathematical model to simulate Biological Neural Networks (BNNs) specifics. A typical neuron consists of n inputs. Each input is multiplied by the weight of input. Also, each neuron has a threshold value. A neuron uses nonlinear functions to determine outputs. The typical nonlinear function is sigmoidal function (F), which is defined below:

$$F(*) = \frac{1}{1+e^{-*}} \tag{1}$$

If $\sum_{j=1}^{n} w_{ij} x_{ij} \geq \phi_i$, then a neuron generates an activation signal R_i to determine output as shown below:

$$o_i = F_i\left(\sum_{j=1}^{n} w_{ij} x_{ij}\right) \tag{2}$$

where o_i is the output value, i is the number of neurons, j is number of inputs, x is the input value, w is the weight of input and ϕ is the threshold value.

2.2. CART Algorithm

The Classification and Regression Trees (CART) method of Breiman et al. [17] is another data mining tool that generates binary decision trees. CART is a nonparametric statistical methodology developed for analyzing classification issues either from categorical or continuous dependent variables. If the dependent variable is categorical, CART produces a classification tree. Otherwise, if the dependent variable is continuous, it produces a regression tree. The CART tree is constructed by splitting subsets of the data set using all predictor variables to create two child nodes repeatedly, beginning with the entire data set. The best predictor is chosen using a variety of impurity or diversity measures. The goal is to produce subsets of the data which are as homogeneous as possible with respect to the target variable [18]. In CART algorithm, for each split, each predictor is evaluated to find the best cut point (continuous predictors) or groupings of categories (nominal and ordinal predictors) based on improvement score or reduction in impurity [17].

In regression trees, the Least Squared Deviation (LSD) impurity measure is used for splitting rules and goodness of fit criteria**.** The LSD measure *R(t)* is simply the weighted within node variance for node *t*, and it is equal to the re substitution estimate of risk for the node [17]. It is defined as:

$$R(t)=\frac{1}{N_W(t)}\sum_{i\in t}\omega_i f_i\left(y_i-\bar{y}(t)\right)^2 \tag{3}$$

$$\bar{y}(t)=\frac{1}{N_W(t)}\sum_{i\in t}\omega_i f_i y_i \tag{4}$$

$$N_W(t)=\sum_{i\in t}\omega_i f_i \tag{5}$$

where $N_W(t)$ is the weighted number of records in node t, ω_i is the value of the weighting field for record i (if any), f_i is the value of the frequency field (if any), y_i is the value of the target field, and $\bar{y}(t)$ is the mean of the dependent variable (target field) at node t. The LSD criterion function for split s at node t is defined as:

$$Q(s,t)=R(t)-R(t_L)-R(t_R) \tag{6}$$

where $R(t_R)$ is the sum of squares of the right child node and $R(t_L)$ is the sum of squares of the left child node. The split s is chosen to maximize the value of $Q(s, t)$.

3. Governing parameters and data used

The most important dimensional and dimensionless parameters determining the scour depth around a pile due to regular waves may be recognized such as bed grain size (d), pile diameter (D), wave period (T), maximum flow velocity (U_m), maximum shear velocity (U_{fm}), pile Reynolds number (Re), Shields parameter (θ), Keulegan-Carpenter number (KC) and sediment number (Ns) defined below [3,11,13]:

$$\mathrm{Re}=\frac{U_m D}{\nu} \tag{7}$$

$$\theta=\frac{U_{fm}^2}{g(Gs-1)d} \tag{8}$$

$$KC=\frac{U_m T}{D} \tag{9}$$

$$Ns=\frac{U_m}{\sqrt{g(Gs-1)d}} \tag{10}$$

$$U_{fm} = (0.5f)^{1/2} U_m \tag{11}$$

where ν is the kinematic viscosity, *Gs* is the relative specific gravity, *g* is the gravitational acceleration and *f* is the wave friction factor.

For an ANN model the quality of database is very important [13]. Hence, training and testing data were obtained from the laboratory experiments of Sumer et al. [3] and Dey et al. [19]. The ranges of various parameters are summarized in Table 1 .It should be mentioned that the proposed models (ANNs and CARTs) are applicable within these ranges.

4. Scour depth prediction

Data set was separated into two input categories as inputs for prediction of the scour depth: dimensional (d, D, T, U_m and U_{fm}) and dimensionless (Re, θ, KC and Ns). Other influential parameters (Gs, g and ν) were constant during training and were not used as input directly. Scour depth normalized by pile diameter (S/D) was used as the output.

Dimensional and dimensionless data sets were divided into two parts for training and testing the models. The total numbers of data points were 88 which 75% of them were used for training and 25% for testing the models.

Statistical measures such as the Root Mean Square Error (*RMSE*), the correlation coefficient (*R*), the Bias and Scatter Index (*SI*) were employed for qualitative evaluation of the models. These measures are defined below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (m_i - o_i)^2}{N}} \tag{12}$$

$$R = \frac{\sum_{i=1}^{N}(o_i - \overline{o})(t_i - \overline{m})}{\sqrt{\sum_{i=1}^{N}(o_i - \overline{m})^2 \sum_{i=1}^{N}(t_i - \overline{m})^2}} \quad (13)$$

$$\text{Bias} = \overline{o} - \overline{m} \quad (14)$$

$$SI = \frac{RMSE}{\overline{m}} \quad (15)$$

where $\overline{o}$ and $\overline{m}$ are the average of the network output and the measured values and N is the total number of data points.

4.1 Dimensional inputs

As mentioned above, the data set was divided into two data subsets for training and testing the models. Here, grain bed size (d), pile diameter (D), wave period (T), maximum flow velocity (U_m) and maximum shear velocity (U_{fm}) were used as the inputs and normalized scour depth (S/D) was used as the output variable. A Multi Layer Perceptron (MLP) with Back Propagation (BP) learning rule was used to train the network. To prevent overfitting during the training of the ANN, the number of nodes of the hidden layer was chosen using expression given by Huang and Foo [20]:

$$M \le 2Z+1 \quad (16)$$

where M and Z are number of the nodes in hidden and input layers, respectively.

The number of the neurons of the input and output layers were 5 and 1, respectively. One hidden layer with 3 neurons was found to be the best topology. The comparison between predicted and observed values of the training and testing data using dimensional parameters for ANN, are shown in Fig. 1. Fig. 2 shows a comparison between observed and predicted S/D by CART algorithm for training and testing data. The error statistics of the models generated by dimensional parameters for testing data are given in Table 2. As seen, the correlation coefficient of the ANN model for testing data is about 13% more and *RMSE* and *SI* are about 64% less than those of CART model. From the comparison between Fig. 1 and Fig. 2, it can be deduced that ANN is more skillful and accurate than CART in prediction of *S/D*. Also, ANN model has a Bias of 0.018 that shows this model slightly overestimates *S/D*.

4.2 Dimensionless inputs

Here, the models are trained by dimensionless inputs such as pile Reynolds number (*Re*), Shields parameter (θ), Keulegan-Carpenter number (*KC*) and sediment number (*Ns*). The dimensionless parameters used in the present study are similar to those used in the prior studies [3, 4, 6, 11 and 13]. In the ANN model, the numbers of the neurons in input and output layers were 4 and 1, respectively. One hidden layer with 5 nodes was found as the optimum topology. The performance of this ANN model was not better than that of previous model (Table 2). This result is in line with the results of Kambekar and Deo [11] and Bateni and Jeng [13]. They also found that using dimensional parameters yields more accurate results for the estimation of scour depth. The *R* value of the testing was 0.924 that is about 4.4% less than that of ANN model developed by dimensional data (Table 2). Additionally, this model has a testing *RMSE* and *SI* of 0.075 and 0.203 respectively that are more than those of the ANN model trained by dimensional data. Bias of -0.005 showed that this model slightly underestimates normalized scour depth. Fig. 3 shows the comparison between predicted values for training and testing data of the trained ANN with dimensionless data set. As seen, the results yielded by the ANN model are satisfactory but not as accurate as those of the ANN model trained by the dimensional data set. However, it should be noted that the models based on dimensional parameters obtained from small scale measurements can not be used for design purposes.

In contrast with the ANN results, the CART model trained by dimensionless data was more accurate than that trained by dimensional data. Fig. 4 displays observed and predicted values of normalized scour depth for training and testing data of CART model trained by dimensionless data. The testing results of CART model developed by dimensionless data (Table 2) have *R*, *RMSE* and *SI* values of 0.955, 0.069 and 0.186, respectively. These values indicate an increase in the *R* (1.12 times) and a decrease in errors parameters (reduction by 57% in *RMSE* and *SI*) compared to those of CART model trained with dimensional data. Also, the testing results indicate that regression tree and ANN models based on dimensionless data performed approximately the same.

In brief, the regression tree can be preferred to the ANNs since it is a non parametric approach. In addition, in regression trees application there is no need to find the network parameters such as optimum numbers of hidden layers and neurons by the process of trial and error. Use of dimensional parameters may yield better results but these parameters are in a limited (laboratory) ranges and it is not possible to generalize the results to real (prototype) cases. By employing dimensionless parameters to train the data mining models, it is possible to generalize the results to real (prototype) cases.

In order to develop a more accurate model, a committee model generated by the combination of the results of ANN and CART models based on dimensionless parameters was also tested. The results were obtained by the simple geometric mean of the output values of ANN and RT models developed by non dimensional data, defined as:

$$o_C = \sqrt{o_A \times o_R} \quad (17)$$

where o_C, o_A and o_R are the outputs of the committee, ANN and regression tree models (based on dimensionless parameters), respectively. Fig. 5 compares the outputs of this committee model and the measured values. As shown in Table 2, the *R* value has increased about 0.4% and *RMSE* and *SI* have decreased about 17.2% relative to those of the best model (ANN based on dimensional parameters).

5. Comparison between present study and the empirical method

Results obtained by ANN and CART approaches for the prediction of normalized scour depth were also compared with that of existing empirical formulae. Sumer et al. [3] presented the following empirical expression for scour depth around a vertical circular pile due to regular oscillatory waves:

$$S/D = 1.3\{1 - \exp[-0.03(KC - 6)]\} : \text{for } KC \geq 6 \quad (18)$$

Several assumptions need to be considered in the development of regression models. The assumptions such as (a) randomness of errors with zero mean, normality of errors, homoscedasticity of errors and uncorrelatedness of errors may restrain the application of regression analysis (see also [21]).

In the traditional approach of statistical regression data had to follow a pre-defined distribution (normal distribution), while in data mining approaches the statistical distribution of the data does not need to be known [21].

Fig. 6 shows observed scour depth plotted against predicted scour depth using the empirical equation as well as data mining models.

Table 2 also, displays the statistical measures of the present study and the empirical formulae. As can be seen, the developed models predict the scour depth more accurately than the conventional approach. The ANN model based on dimensional parameters has *RMSE* and *SI* values of 0.058 and 0.157, respectively; i.e. 27% improvement compared to the error statistics of the best empirical formulae [3]. CART model generated by dimensionless parameters also, performed better than Sumer et al. [3] equation. In addition, the committee model (ANN and CART based on dimensionless parameters) estimates *S/D* better than equation 18. This is reflected in 39% decrease in *RMSE* and *SI*, 2.2% increase in the *R* and lower value of Bias (0.010) compared to those of Sumer et al. [3] formulae. Guven et al. [7] also proposed genetic programming (GP) and adaptive neuro-fuzzy system (ANFIS) models for prediction of scour due to regular waves. They showed that their models are superior to Sumer et al. [3] equation (Eq. 18).

6. Sensitivity analysis

To investigate the dependency of *S/D* to the input parameters, the most widely used measures, i.e. the linear correlation coefficient (*R*); which indicates the relationship between model output and inputs; *RMSE*, Bias and *SI* were used. To estimate the importance of different parameters, new models based on each parameter were trained and tested individually. It should be mentioned that parameters such as wave height and flow depth were not considered because they are not reported in the experiments and their effects are considered in maximum flow velocity. The results (Table 3) show that pile diameter is the most important dimensional parameter and the other important dimensional variable is *T*. Also, the conducted sensitivity tests show that *S/D* depends mainly on *KC* values. This result is in line with the results of the previous studies.

It is also interesting to examine the skills of the models as a function of *KC*. Two classes of *KC* values (Table 4) were defined for this purpose. As can be seen, for $KC < 10$ the error statistics are less than those of $KC > 10$. When *KC* is less than 10, scour process is directly related to the lee-wake vortex in the form of vortex shedding and when *KC* is more than 10, scour process is governed by combined horseshoe vortex and vortex shedding [3]. Hence, it could be conferred that scour process for $KC < 10$ is less complex (compared with $KC > 10$ regime) and the developed models and empirical methods show less errors in this regime.

7. Summary and conclusions

This study presents alternative soft computing tools for evaluation of regular waves induced scour around a circular pile. The performances of regression trees and ANNs approaches were demonstrated by showing their skills in prediction of scour depth. The secondary objective of this paper was to determine whether these models perform better than the conventional semi-empirical formulae. The selection of input variables to the network has a large impact on the model accuracy. Hence, two sets of parameters (dimensional and dimensionless) were utilized to analyze the models. The ANN model results showed that the scour depth could be best predicted by using dimensional data. The CART model based on dimensionless inputs predicted scour depth more accurately compared to the CART model based on dimensional inputs. In addition, this model outperformed the existing empirical formulae (Eq. 18). It was argued that the CART algorithm is preferred to the ANN since it is non parametric and does not require optimization of network parameters. Finally, it was shown that the combination (geometric mean) of the results of ANN and CART models performs the best (*RMSE*=0.048 and *SI* =0.13) and is more accurate than the existing approach (*RMSE*=0.079 and *SI* =0.21).

A sensitivity test was also carried out and it was shown that *KC* number is the most important parameter on scour process. Also, it was found that pile diameter has more influence on scour depth than the other dimensional parameters. In the prediction of *S/D* in *KC* $<$ 10 regime, the developed models and empirical methods showed less *RMSE* relative to that of *KC* $>$ 10 regime. The present study shows that data mining tools such as CART and ANN can model the systems with nonlinear and complex input-output relations smartly. The main limitation of these models is the range of applicability which is limited to the range of parameters used for the training. The used approaches could easily be extended to similar problem in hydraulic engineering such as analysis of scour around pile groups.

8. Acknowledgments

The authors appreciate comments of Dr. A. Yeganeh, M. Rastegar-Mojarad, M. H. Kazeminezhad and Mr. B. Salavati for their comments**.** Also, the first author wishes to thank Mrs. N. Hashemian Natanzi and Mrs. M. Irani for their support..

9. References

[1] Palmer H.D. Wave induced scour on the sea floor. Proceeding of the Civil Engineering, ASCE, New York, 1969; 703-716.

[2] Wang R.-K., Herbich J. B. Combined current and wave-produced scour around a single pile. CEO Rep. No. 269, Texas A&M University, 1983.

[3] Sumer B.M., Fredsoe J., Christiansen N. Scour around a vertical pile in waves. Journal of Waterway, Port, Coastal and Ocean Engineering 1992; 118 (1): 15-31.

[4] Bayram A., Larson M. Analysis of scour around a group of vertical piles in the field. Journal of Waterway, Port, Coastal and Ocean Engineering 2000; 126 (4): 215-220.

[5] Myrhaug D., Rue H. Scour around group of slender vertical piles in random waves. Applied Ocean Research 2005; 27 (1): 56-63.

[6] Sumer B.M., Hatipoglu F., Fredsoe J. Wave scour around a pile in sand, medium dense and dense silt. Journal of Waterway, Port, Coastal and Ocean Engineering 2007; 133 (1): 14-27.

[7] Guven A., Azmathulla H., Zakaria N. Linear genetic programming for prediction of circular pile scour. Ocean Engineering 2009; 36: 985-991.

[8] Azamathulla H. Md., Deo M. C., Deolalikar P. B. Estimation of scour below spillways using neural networks. Journal of Hydraulic Research 2006; 44(1): 61-69.

[9] Azmathulla H. Md., Deo M.C., Deolalikar P.B. Neural Networks for Estimation of Scour Downstream of a Ski-Jump Bucket. Journal of Hydraulic Engineering 2005; 131(10): 898-908.

[10] Guven A., Gunal M. Prediction of Scour Downstream of Grade-Control Structures Using Neural Networks. Journal of Hydraulic Engineering 2008; 134(11): 1656-1660.

[11] Kambekar A.R., Deo M.C. Estimation of pile group using neural networks. Applied Ocean Research 2003; 25: 225-234.

[12] Khosronejad A., Montazer G.A., Ghodsian M. Estimation of scour properties around vertical pile using ANNs. International Journal of Sediment Research 2003; 18(4): 290-300.

[13] Bateni S.M., Jeng D.-S. Estimation of pile group scour using adaptive neuro-fuzzy approach. Ocean Engineering 2007; 34: 1344-1354.

[14] Henderson B. L., Bui E. N., Moran C.J., Simon D. A. P. Australia-wide predictions of soil properties using decision trees. Geoderma 2005; 124 (3-4): 383-398.

[15] Dey P. K. Project risk management: A combined analytic hierarchy process and decision tree approach. Cost Engineering 2002; 44(3): 13–26.

[16] Mahjoobi J., Etemad-Shahidi A. An alternative approach for prediction of significant wave height based on classification and regression trees. Applied Ocean Research 2008; 30(3): 172-177.

[17] Breiman L., Friedman J.H., Olshen R.A., Stone C.J. Classification and Regression Trees. Belmont, Wadsworth Statistical Press, 1984.

[18] Kantardzic M. Data Mining: Concepts, Models, Methods, and Algorithms. John Wiley & Sons, 2003.

[19] Dey S., Sumer M., Fredsoe J. Control of scour at vertical circular piles under waves and current. Journal of Hydraulic Engineering 2006; 132(3): 270-279.

[20] Huang W., Foo S. Neural network modelling of salinity variation in Apalachicola river. Water Research 2002; 36:356–362.

[21] Maier, H. R., Dandy, G. C., The use of artificial neural networks for the prediction of water quality parameters. Water Resources Research 1996; 32(4): 1013-1022.

Figures Caption

Fig. 1. Comparison between observed and predicted normalized scour depth of ANN model for the dimensional data set

Fig. 2. Comparison between observed and predicted normalized scour depth of CART model for the dimensional data set

Fig. 3. Comparison between observed and predicted normalized scour depth of ANN model for the dimensionless data set

Fig. 4. Comparison between observed and predicted normalized scour depth of CART model for the dimensionless data set

Fig. 5. Comparison between observed and predicted normalized scour depth of the committee model results for dimensionless data set

Fig. 6. Comparison between the present study and the empirical approaches

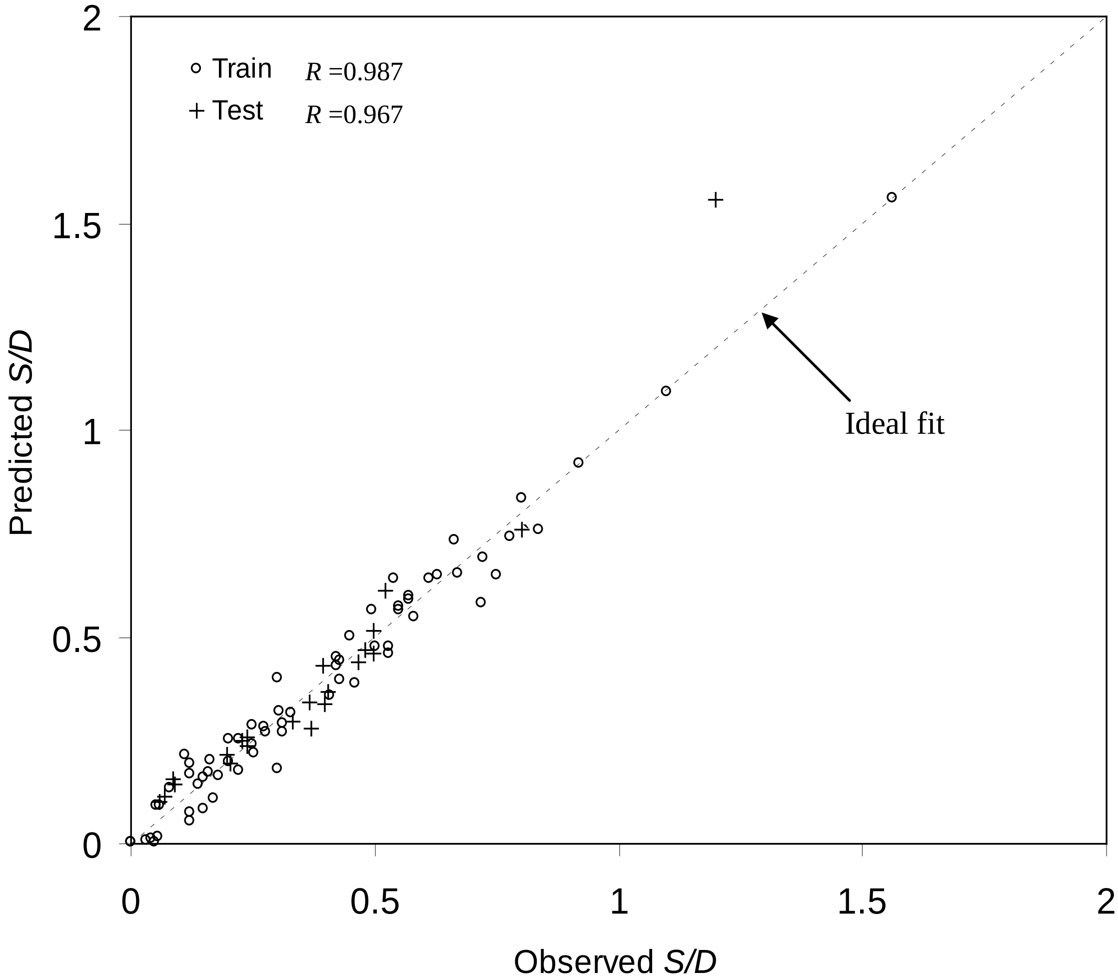

Fig. 1. Comparison between observed and predicted normalized scour depth of ANN model for the dimensional data set

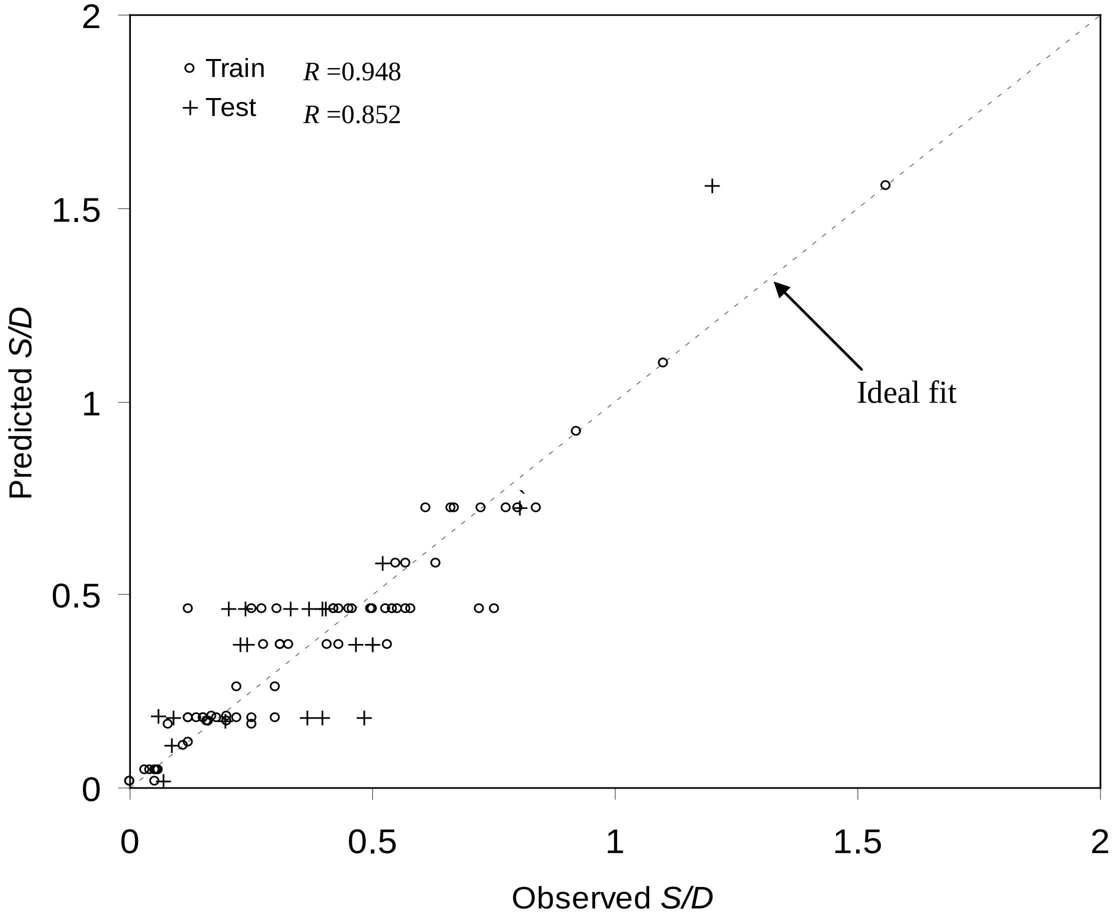

Fig. 2. Comparison between observed and predicted normalized scour depth of CART model for the dimensional data set

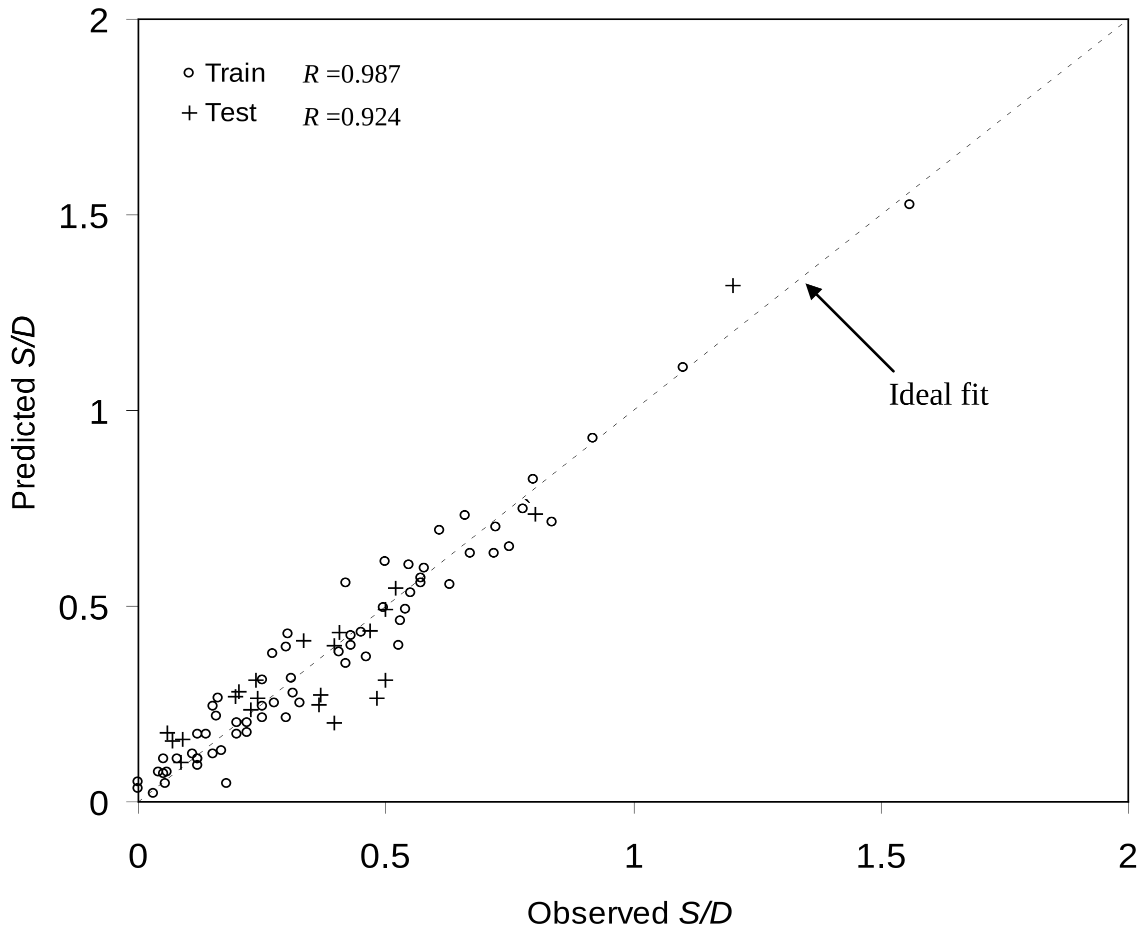

Fig. 3. Comparison between observed and predicted normalized scour depth of ANN model for the dimensionless data set

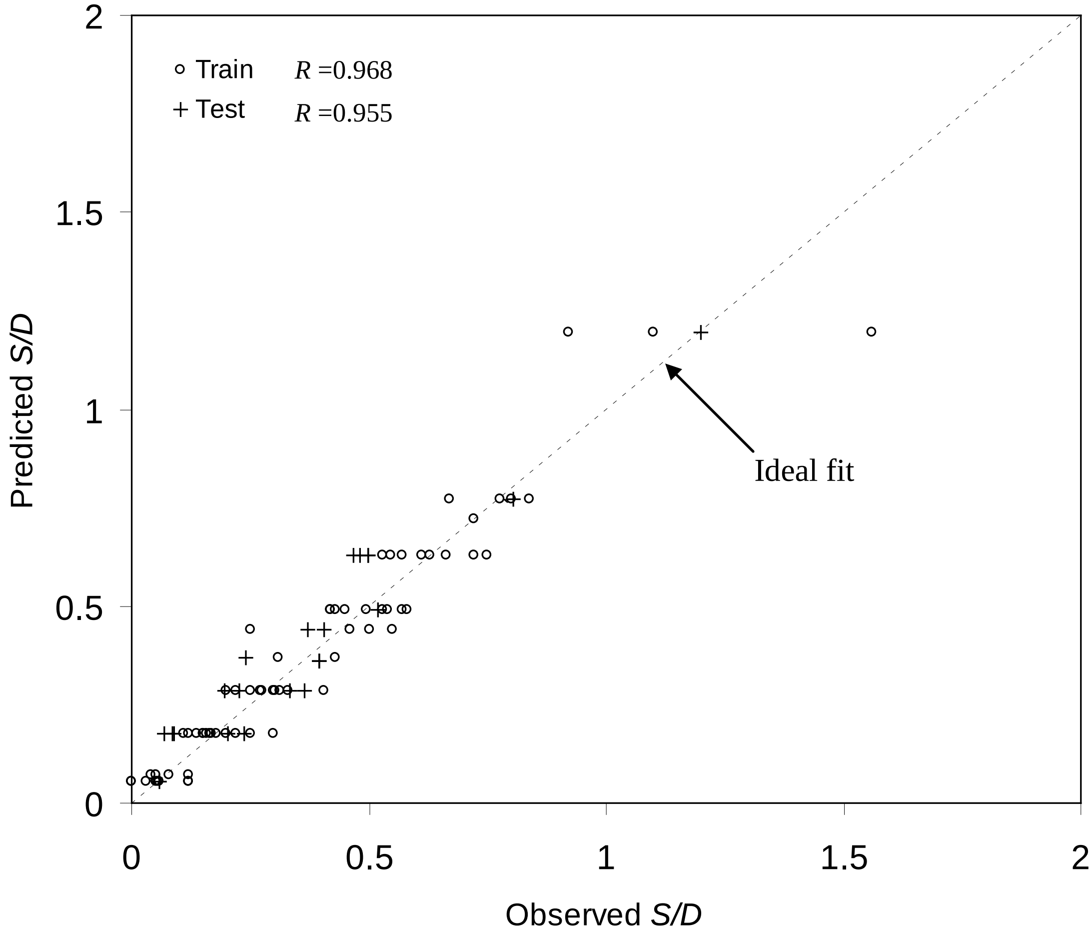

Fig. 4. Comparison between observed and predicted normalized scour depth of CART model for the dimensionless data set

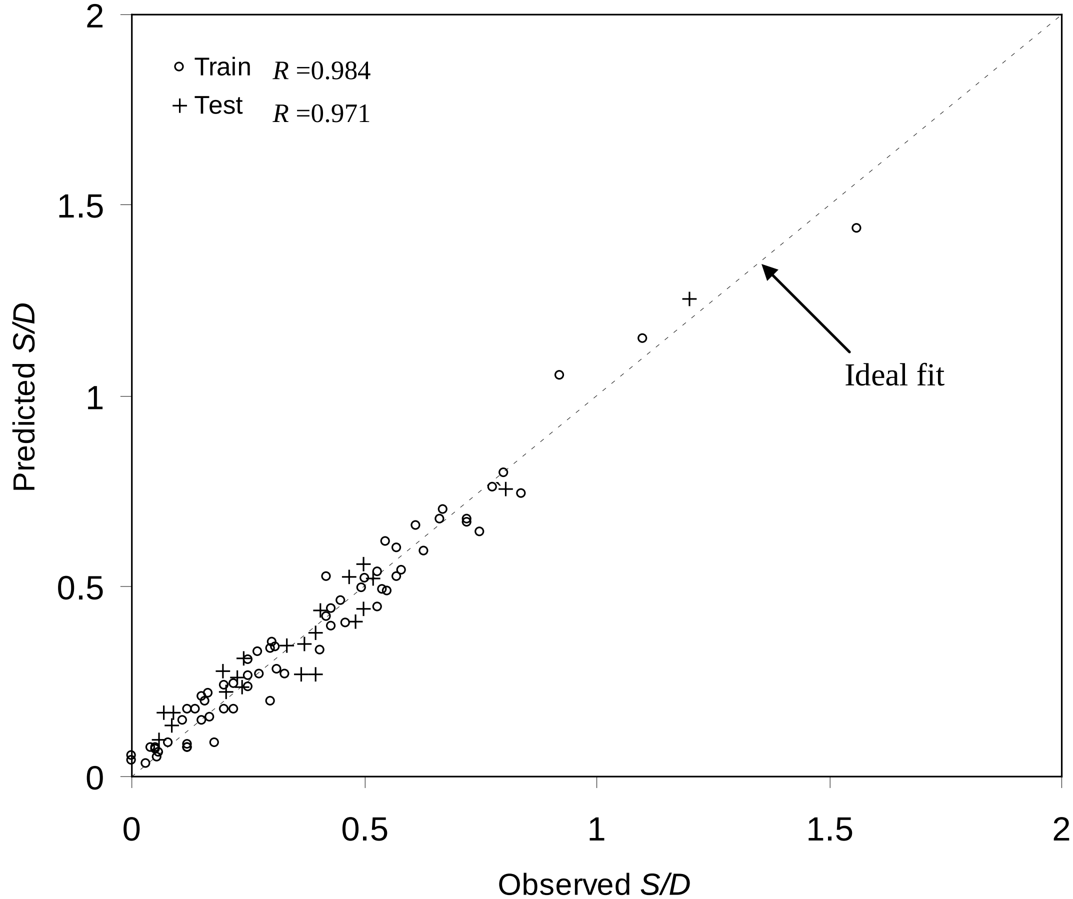

Fig. 5. Comparison between observed and predicted normalized scour depth of the committee model results for dimension data set

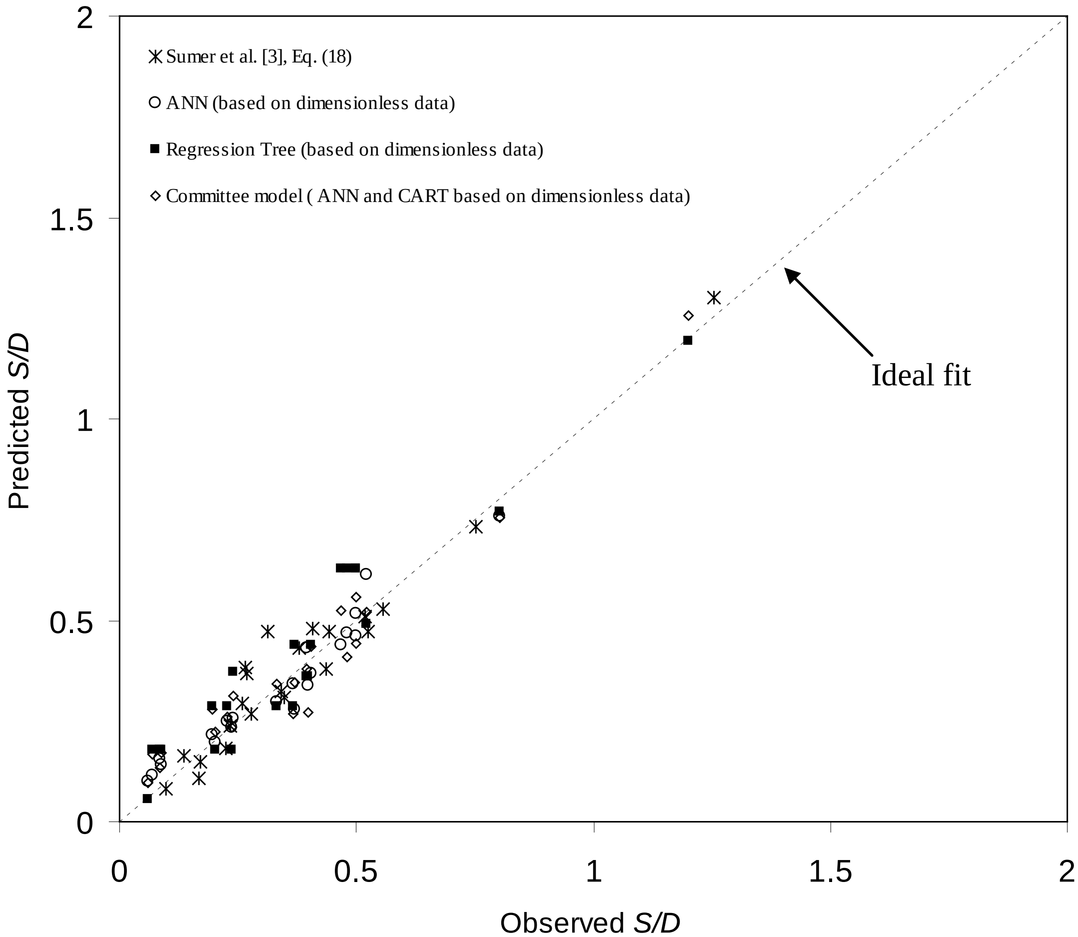

Fig. 6. Comparison between the present study and the empirical approach

Table1: Ranges of data set used to train and test the network

Parameter	Range
Grain size ($d\times10^{-5}$)	18-58 m
Pile diameter (D)	0.01-0.20m
Wave period (T)	1.19-588.24 s
Maximum flow velocity (U_m)	0.112-0.533 m/s
Maximum shear velocity (U_{fm})	0.013-0.025 m/s
Pile Reynolds number ($Re\times10^5$)	0.03-1.10
Shields parameter (θ)	0.04-0.22
Keulegan-Carpenter number (KC)	6.4-5626
Sediment number (Ns)	1.99-9.87
dimensionless equilibrium scour depth (S/D)	0-1.56

Table 2: Statistical error measures of various approaches to estimate *S/D*, testing data

Approach	*R*	*RMSE*	Bias	*SI*
Sumer et al.[3], Eq. (18)	**0.950**	**0.079**	**0.022**	**0.214**
ANN based on dimensional data	0.967	0.058	0.018	0.157
CART based on dimensionless data	0.955	0.069	0.040	0.186
Committee model results (ANN and CART based on dimensionless data)	**0.971**	**0.048**	**0.010**	**0.130**
CART based on dimensional data	0.852	0.161	0.026	0.436
ANN based on dimensionless data	0.924	0.075	-0.005	0.203

Table 3: Statistical error measures of ANNs based on each individual parameter

ANN based on	*R*	*RMSE*	Bias	*SI*
Bed grain size (d)	-0.477	0.282	0.205	0.764
Pile diameter (D)	**0.413**	**0.243**	**0.178**	**0.658**
Wave period (T)	0.390	0.272	0.190	0.737
Maximum flow velocity (U_m)	0.181	0.301	0.211	0.815
Maximum shear velocity (U_{fm})	0.159	0.288	0.195	0.780
Pile Reynolds number (Re)	0.559	0.191	0.115	0.517
Keulegan-Carpenter number (KC)	**0.941**	**0.110**	**0.101**	**0.298**
Shields parameter (θ)	0.521	0.282	0.211	0.764
Sediment number (Ns)	0.332	0.278	0.221	0.753

Table 4: The *RMSE* of two classes of *KC* using different approaches

Approach	*RMSE* (*KC*<10)	*RMSE* (*KC*>10)
Sumer et al.[3], Eq. (18)	0.065	0.128
ANN (based on dimensional data)	0.053	0.069
Regression tree (based on dimensionless data)	0.058	0.091
Committee model (ANN and CART based on dimensionless data)	0.040	0.059