USING ANNS AND REGRESSION TREES APPROACHES TO ESTIMATE SCOUR DEPTH AROUND A CIRCULLAR PILE DUE TO WAVE IN MEDIUM DENSE SILT AND SAND BED

Mohammad Karim Ayoubloo
Iran University of Science and Technology
Mo_ayoubloo@civileng.iust.ac.ir
Tel: +989192162883

Amir Etemad-Shahidi
Iran University of Science and Technology
Etemad@iust.ac.ir
Tel: +982173913170

Javad Mahjoobi
Ministry of Energy, Water Research Institute
Jmahjoobi@gmail.com
Tel: +989151254211

ABSTRACT: Prediction of scour around a pile due to oscillatory wave action is very important in many offshore and marine engineering problems. Because of complexity of scour process, most of the empirical formulas are unable to estimate scour hole depth accurately. Artificial Neural Networks (ANNs) and regression trees are efficient procedures to understand and model complex systems with ambiguous relations. A Multi Layer Perceptron (MLP) is one of the most common kinds of ANNs and has been used to map input-output systems while CART algorithm was employed for building and evaluating regression trees. In the present study, two input sets were employed to estimate scour depth: (a) dimensional parameters such as bed grain size, pile diameter, wave period, wave height, maximum flow velocity and maximum shear velocity (b) nondimensional parameters such as pile Reynolds number, Shields parameter, Keulegan-Carpenter number, grain Reynolds number, sediment number and relative density. Output parameter was nondimensional equilibrium scour depth. The tests results reveal that a MLP with back propagation learning rule and CART model based on nondimensional parameters can predict scour hole depth better than the existing empirical formula. Also, a sensitivity analysis was carried out and it showed that Keulegan-Carpenter number and wave height are the most important parameters in scour process.

Keywords: Artificial Neural Network; Pile; Regression tree; Scour; Wave

INTRODUCTION

Piles are extensively used as a principle element in marine and coastal structures. When a pile is installed on the sea floor, because of interaction between wave and current and a pile, scour occurs in the vicinity of the pile. Commonly scour leads to changes in the capacity of the sediment migration on an erodible bed. Scour process can finally cause to complete failure and collapse of a marine structure. Therefore, accurate estimation of scour depth around a pile is very important and has to be considered.

Many studies about scour around a pile in wave have been performed during the past decades. Sumer et al. (1992) and Kobayashi and Oda (1994) displayed Keulegan-Carpenter number as the main parameter governing the scour process. Also Sumer et al. (1992) presented an experimental expression to predict scour depth. The influence of pile geometry, using square and circular pile, has been investigated by Sumer et al. (1993). Further Sumer and Fredsoe (1998) carried out a laboratory study around different arrangements of pile groups. An investigation of the scour around group of piles in the field due to oscillatory wave has been performed by Bayram and Larson (2000). They found a distinct correlation between scour depth and Keulegan-Carpenter number. Recently Sumer et al. (2007) implemented an experimental investigation on wave scour around a circular pile in three kind of soil so called dense silt (with relative density of Dr = 0.74), medium dense silt (with Dr = 0.38) and sand (with Dr = 0.23). They showed when the bed was dense silt the scour depth was increased by a factor of 1.6-2 relative to two others.

It is very difficult to find a mathematical model that exactly shows the scour process developing under effect of wave and current. There several governing parameters interacting with each other complicatedly in the scour process. In spite of existing investigations, there is no reliable and

general formula to predict the scour depth around a pile. Artificial Neural Networks (ANNs) are efficient and simple procedures to model and map an input-output complex system. Recently due to flexibility of ANNs , these models have been used in hydraulic (Azmathullau et al., 2005), coastal and ocean engineering (Khosronejad et al., 2003; Kambekar and Deo, 2003) and to predict wave and pile group scour (Kazeminzhad et al., 2005; Bateni and jeng, 2007).
As mentioned before, estimation of scour depth is basically an uncertain process and therefore is difficult to perform by means of empirical equations. Hence, Regression Trees can also be suited to this problem since this approach is 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 and their application does not require knowledge of the underlying physical process as a precondition.
However, using of the ANN and regression trees to estimate scour depth around a pile due to wave in medium dense silt and sand bed has not been experimented before. In this paper, CART algorithm (Breiman et al., 1984) is employed for building and evaluating regression trees. CART builds classification and regression trees for predicting continuous (regression) and categorical predictor variables (classification).

ANN

ANNs have emerged from studies of how human brain performs operation. Human brain is composed of billions of individual processing units namely neurons that are interconnected. Each neuron input, x_1-x_n, is weighted by the values w_1-w_n. a threshold in the node is described by an additional constant input of w_0. The output, o, is obtained by summing the weighted inputs to the neuron and passing the result through a nonlinear function (Fig. 1), F, as defined below:

$$o = F(\sum_{i=1}^{n} w_i x_i + w_0) \tag{1}$$

$$F(z) = \frac{1}{1+e^{-z}} \tag{2}$$

where F is sigmoidal function.

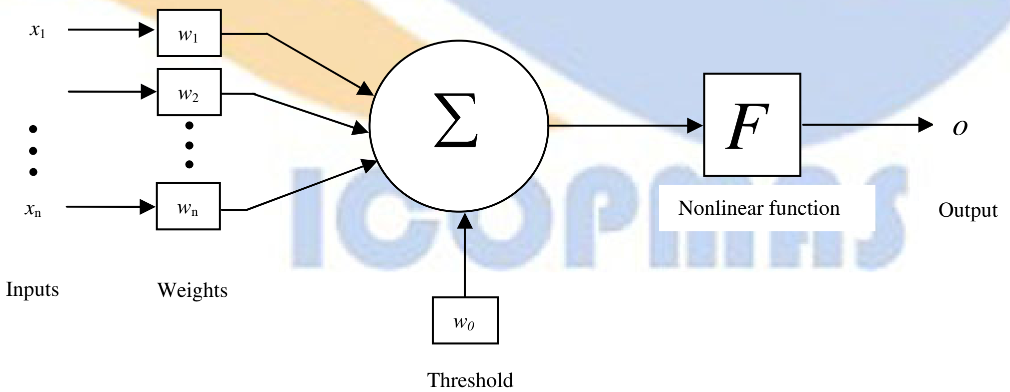

Fig1: A neuron

The most popular neural network is a Multi Layer Perceptron (MLP). A simple MLP is made up of an input, a hidden and an output layer as shown in Fig. 2. Each layer

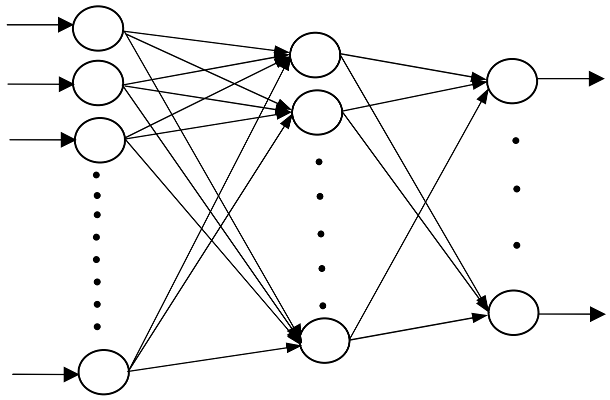

Fig. 2: A schematic MLP

consists of number of neurons. Data flows through in one direction only, from input to output layer. Hence, this type of network is called a Feed Forward Network. Once the topology of the network is assigned and the network is trained. The training procedure searches for the minimum of the error function (average sum squared) as clarified below:

$$E = \frac{1}{2}\sum (o_i - t_i)^2 \quad (3)$$

where o_i and t_i are the output and target values of ith output node, respectively.
A number of algorithms have been proposed for training a MLP, determining and adopting the weights, and the most popular is a Back Propagation algorithm. At the first weights are chosen randomly and then fixed by the following process:

$$\Delta w\,(l+1) = \alpha \Delta w\,(l) - \eta \,\partial E / \partial w \quad (4)$$

Where w is the weight between any two nodes; Δw_l and Δw_{l+1} are changes in the weights at lth and $(l+1)$th iteration respectively; α is momentum factor and η is learning rate.

REGRESSION TREES AND CART ALGORITHM

The Classification and Regression Trees (CART) method of Breiman et al. (1984) is another data mining tool that generates binary decision trees. A decision tree is an arrangement of tests that prescribes an appropriate test at every step in an analysis. A Decision tree is a tree in which each branch node represents a choice between a number of alternatives and each leaf node represents a classification or decision.
Regression tree building centers on three major components: (1) a set of questions of the form: is $X \le d$? where X is a variable and d is a constant. The response to such questions is yes or no; (2) goodness of split criteria for choosing the best split on a variable and (3) the generation of summary statistics for terminal nodes.
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 resubstitution estimate of risk for the node (Breiman et al., 1984). It is defined as:

$$R(t) = \frac{1}{N_W(t)} \sum_{i \in t} \omega_i f_i \left(y_i - \bar{y}(t) \right)^2 \quad (5)$$

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:

$$Q(s,t) = R(t) - R(t_L) - R(t_R) \quad (6)$$

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 $Q(s, t)$.

Stopping rules control how the algorithm decides when to stop splitting nodes in the tree. Tree growth proceeds until every leaf node in the tree triggers at least one stopping rule. Any of the following conditions will prevent a node from being split:

a. All records in the node have the same value for all predictor fields used by the model.
b. The number of records in the node is less than the minimum parent node size (user defined).
c. If the number of records in any of the child nodes resulting from the node's best split is less than the minimum child node size (user defined).
d. The best split for the node yields a decrease in impurity that is less than the minimum change in impurity (user defined).

In regression trees, each terminal node's predicted category is the weighted mean of the target values for records in the node (Breiman et al., 1984). This weighted mean is calculated as:

$$\bar{y}(t) = \frac{1}{N_W(t)} \sum_{i \in t} \omega_i f_i y_i \quad (7)$$

where $N_W(t)$ is defined as:

$$N_W(t) = \sum_{i \in t} \omega_i f_i \quad (8)$$

GOVERNING PARAMETERS

Dimensional analysis and the earlier studies (Sumer et al., 1992; Bayram an Larson, 2000; Sumer et al., 2007) presented some of dimensional and nondimensional parameters that may affect on scour process more than the others such as: bed grain size (d), pile diameter (D), wave period (T), wave Height (H), maximum flow velocity (U_m), maximum shear velocity (U_{fm}), pile Reynolds number (Re), Shields parameter (θ), grain Reynolds number (Re_d), sediment number (Ns) and relative density (Dr). The nondimensional numbers are defined as:

$$\mathrm{Re} = \frac{U_m D}{\nu}, \theta = \frac{U_{fm}^2}{g(Gs-1)d}, KC = \frac{U_m T}{D}, \mathrm{Re}_d = \frac{U_{fm} d}{\nu}, Ns = \frac{U_m}{\sqrt{g(Gs-1)d}}, Dr = \frac{e_{\max} - e}{e_{\max} - e_{\min}} \quad (9)$$

where g is the acceleration due to gravity, Gs the specific gravity, ν the kinematic viscosity, e void ratio and e_{max} and e_{min} maximum and minimum void ratios respectively.

For statistical comparison of predicted and observed scour depth the Mean Absolute Error (MAE), the Root Mean Square Error ($RMSE$) and correlation coefficient (R^2) were used as defined:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |t_i - o_i| \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (t_i - o_i)^2}{N}} \quad (11)$$

$$R^2 = \frac{\left(\sum_{i=1}^{N} (o_i - \bar{o})(t_i - \bar{t}) \right)^2}{\sum_{i=1}^{N} (o_i - \bar{o})^2 \sum_{i=1}^{N} (t_i - \bar{t})^2} \quad (12)$$

where o_i and t_i are network output and target values for ith output, $\bar{o}$ and $\bar{t}$ are the average of network output and target values and N is the total number of events considered.

DATA SET FOR ANN

The laboratory data reported by Sumer et al. (2007) were used for training and likewise testing the network. Three kinds of soil were used in their work but the data of two kinds (medium dense silt and sand) are used in the present study, because the scour depth in dense silt was not agreed with the Sumer et al. (1992) experimental equation (Sumer et al., 2007). Input parameters was divided into dimensional (d, D, T H, U_m and U_{fm}) and nondimensional (Re, θ, KC, Re_d, Ns and Dr). Normalized scour depth (S/D) was the output. Other parameters such as Gs, g and ν were constant. The number of observations was 38 which 28 were including in training and 10 in testing the ANN. The training data were not involved in the testing. The ranges of the inputs and output are presented in table 1.

Table 1: Ranges of data employed to train and test the ANNs and CART

Parameter	Range
Grain size ($d\times10^{-5}$)	6-14.7 m
Pile diameter (D)	0.015-0.08 m
Wave period (T)	1.2-2.5 s
Wave height ($H\times10^{-2}$)	5.5-17.5 m
Maximum flow velocity (U_m)	0.128-0.37 m/s
Maximum shear velocity (U_{fm})	0.014-0.024 m/s
Pile Reynolds number ($Re\times10^{4}$)	0.24-3
Sheilds parameter (θ)	0.094-0.36
Keulegan-Carpenter number (KC)	7-22
Sediment number (Ns)	2.62-7.58
Relative density (Dr)	0.23-0.38
Nondimensinal equilibrium scour depth (S/D)	0-0.5

SCOUR DEPTH PREDICTION-DIMENSIONAL

An ANN model was designed to estimate normalized scour depth (S/D) in a dimensional form. As mentioned, in this case d, D, T, U_m and U_{fm} were employed as dimensional inputs. Training of the network has been done by training data and verifying the accuracy of the network has been performed by testing data. Several topologies by trail and error were examined so that to arrive the best result accuracy. To prevent overfitting during the training of the ANN, the number of nodes of the hidden layer was chosen by use of the expression presented by Huang and Foo (2002):

$M \leq 2Z+1$

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

Dimensional network had one hidden layer with 13 neurons. Input and output layer had 6 and 1 neurons, respectively. After training, the trained network was performed by testing data. As seen in Fig. 3 and Table 2, a high value of 0.958 of the R^2 and low values of MAE and $RMSE$ of 0.021 and 0.027 respectively, show ANN clearly able to estimate S/D by the effects of dimensional parameters.

Table 2: Statistical and error parameters of two approaches to estimate scour depth with dimensional parameters

Approach	R^2	MAE	$RMSE$
ANN	0.978	0.021	0.027
CART	0.169	0.101	0.126

Also Fig. 3 shows a comparison between observed and predicted *S/D* by CART algorithm for dimensional testing data. The error statistics of the models generated by dimensional parameters for testing data are given in Table 2. As seen, the low value of R^2 and the relative high values of *MAE* and *RMSE* shows that CART model can not understand the relation between dimensional parameters and scour depth. From the comparison between ANN and CART models, it can be deduced that ANN was more skillful and accurate than CART in prediction of *S/D* with dimensional parameters.

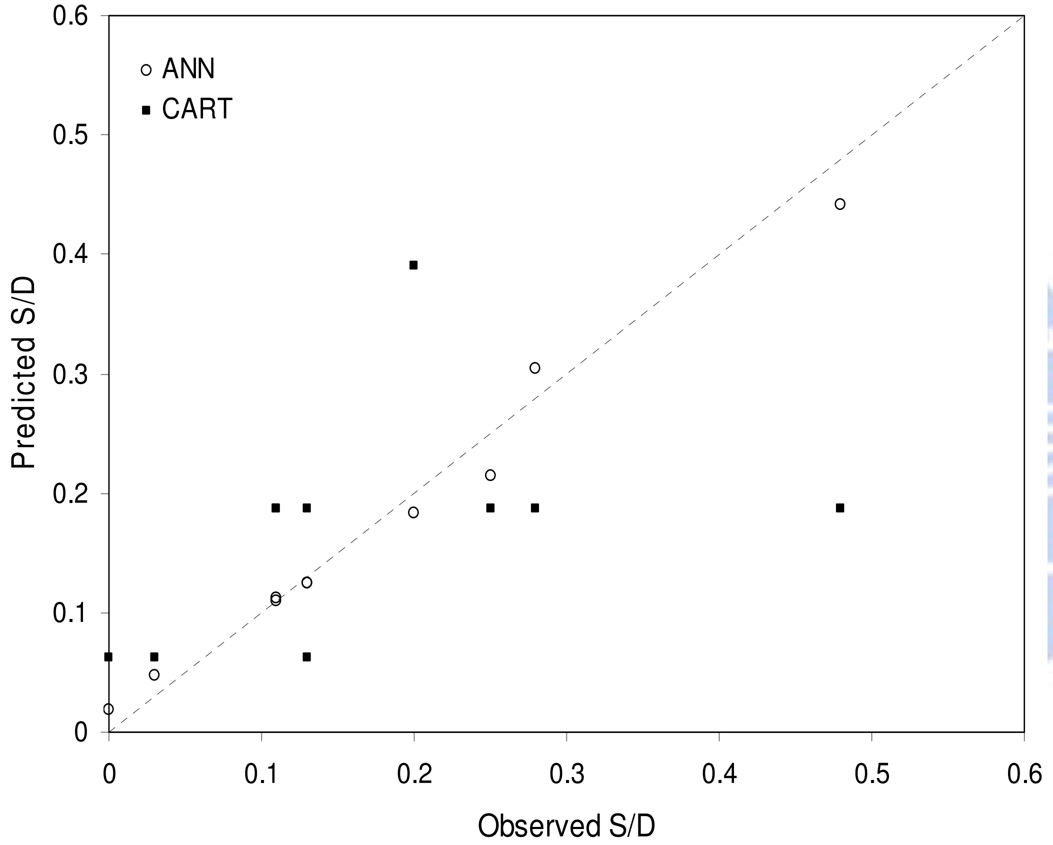

Fig. 3: Comparison between observed and predicted *S/D* by dimensional ANN and CART for testing data set

SCOUR DEPTH PREDICTION-NONDIMENSIONAL

Another ANN model by nondimensional parameters (Re, θ, KC, Re_d, Ns and Dr) was also created. The ANN topology in nondimensional form is the same as dimensional form. The nondimensional ANN performed as well as dimensional model to predict *S/D* (Fig. 4). This is confirmed by the values of 0.953, 0.022 and 0.030 of the R^2, *MAE* and *RMSE* that showed almost these parameters remained constant (Table 3).

By contrast with ANN results, CART algorithm 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 nondimensional data. The testing results of CART model developed by nondimensional data (Table 3) have the

Table 3: Statistical and error parameters of two approaches to estimate scour depth with nondimensional parameters

Approach	R^2	*MAE*	*RMSE*
ANN	0.953	0.022	0.030
CART	0.852	0.040	0.056

R^2, *MAE* and *RMSE* of 0.852, 0.038 and 0.050, respectively. These values indicate that an increase in the R^2 (5.04 times) and a decrease in errors parameters (2.67 times in *MAE* and 2.52 times in *RMSE*) relative to those of CART model trained with dimensional data. Therefore, the testing results indicate that regression tree and ANN based on nondimensional data performed good for predicting scour depth but ANN result was approximately better than CART results**.**

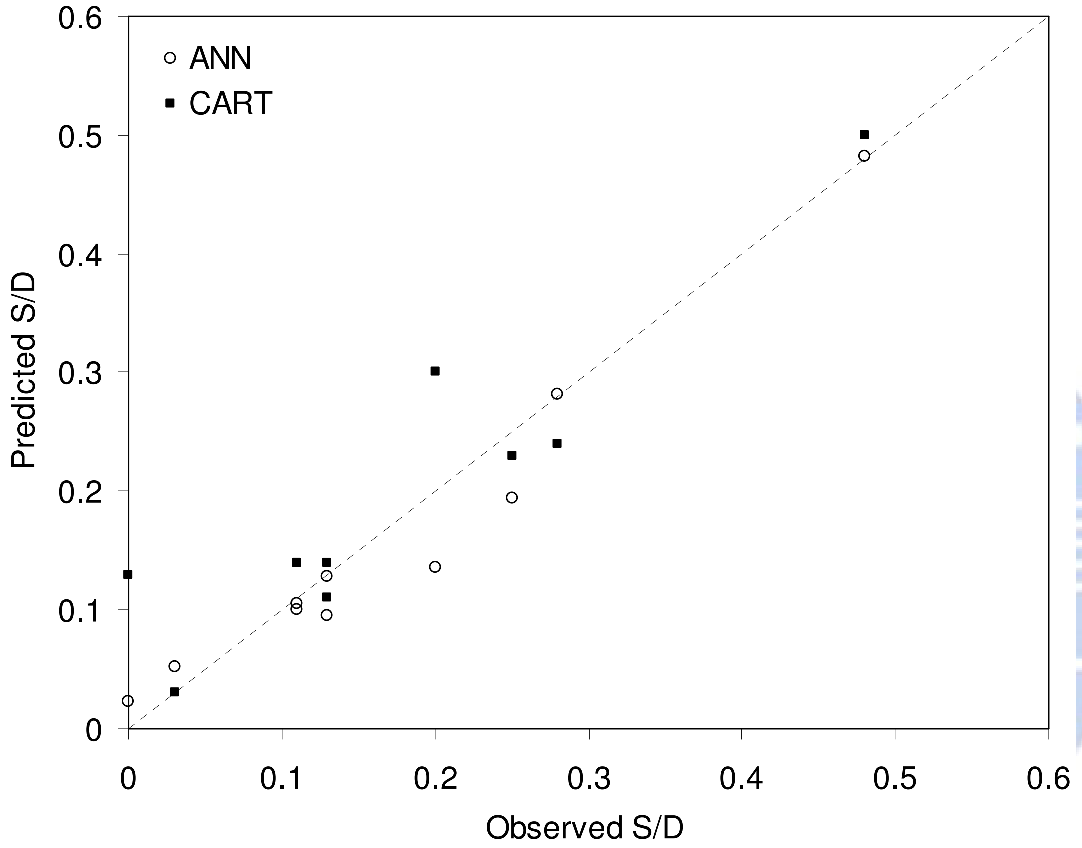

Fig. 4: Comparison between observed and predicted *S/D* by dimensionless ANN for testing data set

COMPARISON

To present ANN and CART models as efficient approaches for predicting *S/D*, a comparison between ANN and CART and the existing empirical formula was accomplished. To evaluate scour depth around a pile due to wave action, an empirical expression has been reported by Sumer et al. (1992):

$$S/D = 1.3\{1 - \exp[-0.03(KC - 6)]\}: for\ KC \geq 6 \qquad (13)$$

Observed versus predicted scour depth of ANN and CART models and conventional expression for the testing data is plotted in Fig. 5.

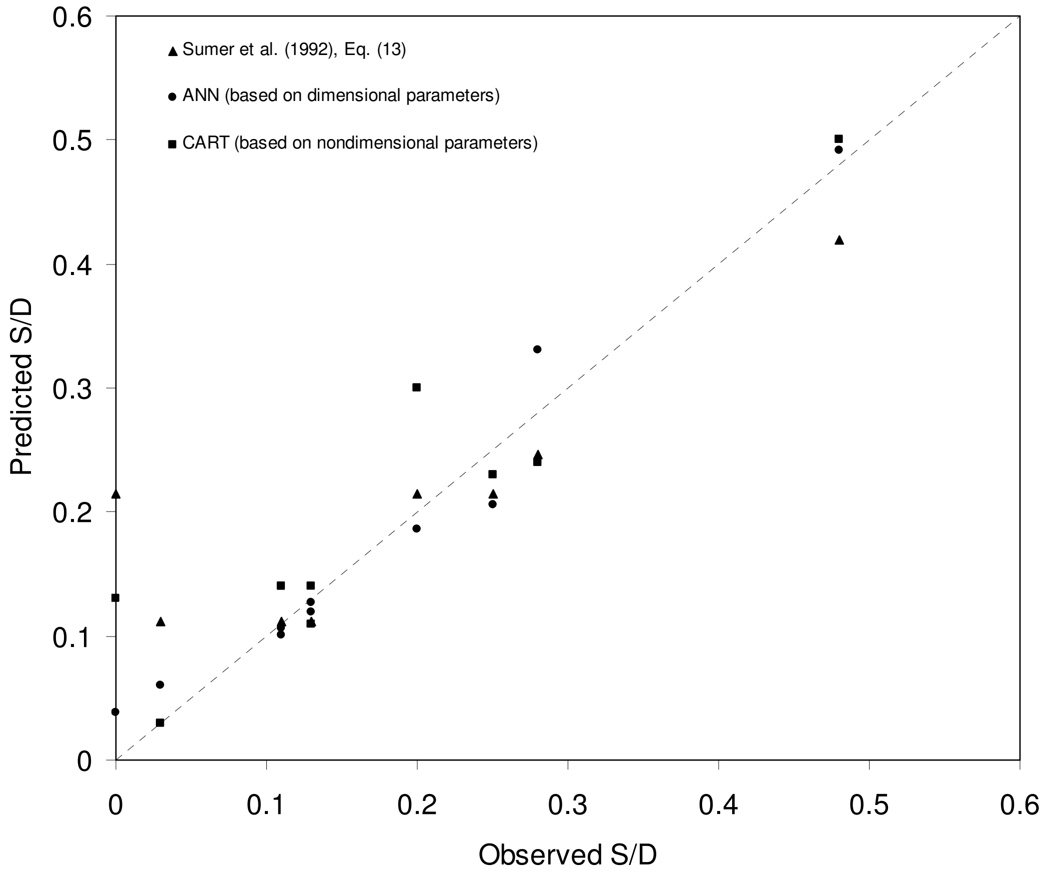

Fig. 5: Comparison between the present study and the empirical approach

As shown in Table 4, ANN model corresponding values of the R^2, *MAE* and *RMSE* are 0.978, 0.021 and 0.027 that present 42.71% increasing in the R^2 and decreasing about 55.31% and 64.93% in *MAE* and *RMSE* relative to the Eq. 13 proposed by Sumer et al. (1992). Also, CART model shows increasing in the R^2 about 24.37% and decreasing in the *MAE* and *RMSE* values about 19.14% and 35.06% respectively relative to Eq. 13. These results show that both ANN and CART models can predict the scour depth more accurately than the empirical method.

Table 4: Statistical and error parameters of the approaches to estimate scour depth

Approach	R^2	*MAE*	*RMSE*
Sumer et al.(1992), Eq. 13	0.685	0.047	0.077
ANN based on dimensional parameters	0.978	0.021	0.027
CART based on nondimensional parameters	0.852	0.038	0.050

SENSITIVITY ANALYSIS

The ANNs involving the effect of each parameter (dimensional and dimensionless) separately were retrained and tested to predict normalized scour depth. These tests revealed more impressive dimensional and nondimensional variables on scour process. Table 5 compares the ANN models that regard individual dimensionals as input.

Table 5: Statistical and error parameters of ANNs based on dimensional parameters individually

ANN based on	R^2	*RMSE*	*MAE*
Bed grain size (d)	0.005	0.1340	0.1006
Pile diameter (D)	0.028	0.1639	0.1209
Wave height (H)	**0.440**	**0.1077**	**0.0744**
Wave period (T)	0.016	0.1360	0.1045
Maximum flow velocity (U_m)	0.156	0.1221	0.0870
Maximum shear velocity (U_{fm})	0.312	0.1119	0.0831

As shown in Table 5, predicted normalized scour depth by wave height is more admissible than the other dimensionals. Also Table 6 shows the ANNs generated by dimensionless parameters that presents *S/D* has strong dependency of *KC* number. The result of Sumer et al. (1992), scour process is mainly governed by *KC* number, confirms and emphasizes the last result of ANN about *KC* dependency.

Table 6: Statistical and error parameters of ANNs based on dimensionless parameters individually

ANN based on	R^2	*RMSE*	*MAE*
Pile Reynolds number (Re)	0.042	0.1345	0.0940
Shields parameter (θ)	0.343	0.1090	0.0773
Keulegan-Carpenter number (KC)	**0.650**	**0.0781**	**0.0568**
Grain Reynolds number (Re_d)	0.380	0.1075	0.0787
Sediment number (Ns)	0.215	0.1218	0.0826
Relative density (Dr)	0.005	0.1340	0.1006

SUMMARY AND COCLUSIONS

The ANN regression trees were applied to forecast scour hole depth around a pile in medium dense silt and sand bed by Sumer et al. (2007) data. Two ANNs and two CART models were generated by dimensional and dimensionless parameters. The results show that the neural networks trained by dimensional and nondimensional parameters can predict normalized scour depth acceptable and their difference in the error parameters (*MAE* and *RMSE*) and the R^2 is very negligible but dimensional ANN predicts slightly accurately. Also, the ANN model established using dimensional parameters estimates scour depth more accurately than the empirical technique based on statistical curve fitting. Also, the CART models shows that these models can predict scour hole depth with nondimensional parameters more accurate than empirical formula. The results of the sensitivity analysis present *H* and *KC* number as more effective dimensional and nondimensional parameters to predict normalized scour depth, respectively.

ACKNOWLEDGMENTS

The authors would like to thank Eng. M. Rastegar Mojarad for his assistant. Also, the first author wishes to special thank to Mrs. N. Hashemian Natanzi and Mrs. M. Irani.

REFFRENCES

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

Kobayashi, T., Oda, K., 1994, Experimental study on developing process of local scour around a vertical cylinder, Proc. 24th Coast. Engrg. Conf. ASCE, Newyork, 1284-1297.

Sumer, B.M., Christiansen, N., Fredsoe, J., 1993. Influence of cross section on wave scour around piles. Journal of Waterway, Port, Coastal and Ocean Engineering, ASCE 119 (5), 477-495.

Sumer, B.M., Fedsoe, J., 1998, Wave scour around group of vertical piles, J. Wtrwy., Port, Coast. And Oc. Engrg., ASCE 124 (5), 348-256.

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

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

Azmathullau, H.Md., Deo, M.C., Deolalikar, P.B, 2005, Neural networks for estimation of scour downstream of a ski-jump bucket, Journal of Hydraulic Engineering, ACSE 131 (10), 898-908.

Khosronejad, A., Montazer, G.A., Ghodsian, M., 2003. Estimation of scour hole properties around vertical pile using ANNs. International Journal of Sediment Reaserch 18(4), 290-300.

Kambekar, A.R., Deo, M.C., 2003. Estimation of pile group using neural networks. Applied Ocean Reaserch 25, 225-234.

Kazeminezkad, M.H., Etemad-Shahidi, A., Mousavi, S.J., 2005. Application of fuzzy inference system in the prediction of wave parameters. Ocean Engineering 32, 1709-1725.

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

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

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