Applied Ocean Research 32 (2010) 34–39

Contents lists available at ScienceDirect

Applied Ocean Research

journal homepage: www.elsevier.com/locate/apor

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

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

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

ARTICLE INFO

Article history:
Received 13 October 2009
Received in revised form
12 May 2010
Accepted 12 May 2010
Available online 3 June 2010

Keywords:
Artificial Neural Network
CART
Pile
Regression tree
Scour depth
Wave

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 the 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.

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 waves 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 difficult 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

* Corresponding author. Tel.: +98 21 77240399; fax: +98 21 77240398.
E-mail address: etemad@iust.ac.ir (A. Etemad-Shahidi).

0141-1187/$ – see front matter
doi:10.1016/j.apor.2010.05.003

investigated by Khosronejad et al. [12]. Bateni and Jeng [13] also combined ANNs with a 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 a sigmoidal function (F), which is defined below:

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

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:

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

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 the 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 an 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:

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

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

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

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 $Q(s, t)$.

Table 1
Ranges of data set used to train and test the network.

Parameter	Range
Grain size ($d \times 10^{-5}$)	18–58 m
Pile diameter (D)	0.01–0.20 m
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 ($Re \times 10^5$)	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

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]:

$$Re = \frac{U_m D}{\nu} \quad (7)$$

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

$$KC = \frac{U_m T}{D} \quad (9)$$

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

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

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.