Expert Systems with Applications 38 (2011) 10114–10123

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Predictive model-based for the critical submergence of horizontal intakes in open channel flows with different clearance bottoms using CART, ANN and linear regression approaches

Mohammad Karim Ayoubloo [a,1], H. Md. Azamathulla [b,*], Ebrahim Jabbari [c,1], Morteza Zanganeh [c,1]

[a] Tehran Firefighting and Safety Services Organization, Tehran, Iran
[b] River Engineering and Urban Drainage Research Centre (REDAC), Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Penang, Malaysia
[c] College of Civil Engineering, Iran University of Science and Technology, Narmak, P.O. Box 16765-163, Tehran, Iran

ARTICLE INFO

Keywords:
CART
ANN
Critical submergence

ABSTRACT

This study presents the development of classification and regression tree (CART), artificial neural network (ANN) and linear regression approaches to predict the critical submergence in an open channel flow for different clearance bottoms. To use the models for application purposes and cover the wide range of inputs, the nondimensional parameters are employed to train and test. The testing results show that all three approaches satisfactorily estimate the critical submergence with margin differences. Also, committee models arithmetic mean-based for the testing results of the tree mentioned approaches are presented as the best models. A comparison between the present study and empirical approaches is carried out which indicates the proposed approaches outperform the empirical formulas expressed in the literature. In addition, committee models are presented as the more generalized approaches by *AIC* criterion. The results also indicate that the variations of the best approach (committee)-predicted and observed the normalized critical submergence with the intake pipe diameter versus the number of the testing data follow favorably a similar trend. Finally, a sensitivity analysis shows that the ratio of the velocity in an intake pipe to the velocity in a channel is the significant parameter in the estimation of the critical submergence.

1. Introduction

Vertical and horizontal intakes are one of the most important parts of hydraulic sets such as a river for irrigation or a reservoir for power generation and industrial purposes. Insufficient water height above a pipe intake (submergence) can lead to vortex formation and air entry. These phenomena may cause the problems such as loss of hydraulic machinery performance or erosion and vibration in pipelines (Denny, 1956). Therefore, an accurate estimation of submergence depth above intakes in related hydraulic systems design has to be considered.

The critical submergence of intakes has been extensively investigated in laboratory or experimentally. Denny (1956) carried out one of the first studies on the critical submergence and air-entraining vortices at pump sumps. He found that entry of 1% air volume to a vortex caused up to 15% reduction of pump performance. Several studies that yielded empirical expressions to predict the critical submergence of intakes have also been done (Kocabas & Yildrim, 2002; Odgaard, 1986; Reddy & Pickford, 1972; Swaroop, 1973). Generally, presented relations to estimate the critical submergence were based on Froud number, Reynolds number, the critical height of intake, Weber number and circulation. Yildirim and Kocabas (1995, 1998, 2002), and Yildrim, Kocabas, and Gulcan (2000) determined the critical submergence for intakes from an open channel flow and still water reservoirs using potential theory and dimensional analysis. Furthermore, the critical submergence for a rectangular intake has been studied by Yildrim (2004). Recently, Ahmad, Rao, and Mittal (2008) proposed predictors for the critical submergence of horizontal intakes for the different bottom clearance (vertical distance of intake to bottom of tank) in open channel flows.

It is very hard to find an exact relation between the effective parameters on the estimation of the critical submergence for a hydraulic system. Empirical formulas based on laboratory or experimentally data are used to predict the critical submergence. Also, due to complexity of the process, these conventional formulae based on regression approach may not be able to present complete relations in the process estimation. Hence, investigators have tried to estimate the critical submergence more accurate than empirical formulas. In the past few years, soft computing and data mining techniques such classification and regression tree (CART) and artificial neural networks (because of flexibility, ability to

* Corresponding author.
E-mail address: mdazmath@gmail.com (H. Md. Azamathulla).
[1] Tel.: +60 45995867.

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.02.073

Notations

AIC	Akaike information criterion	R	correlation coefficient
ANN	artificial neural network	Re	intake Reynolds number
BP	back propagation	MSE	mean square error
C	clearance bottom	$R(t)$	simply the weighted within node variance
d_i	pile diameter	Sc	critical submergence
E	error function	SI	scatter index
FF	fed forward	t	node for CART
f_i	the value of the frequency field	U_i	flow velocity in the intake pipe
Fr	intake Froude number	U_∞	flow velocity in the flume
G	gravitational acceleration	We	Weber number
LSD	least squared deviation	X	independent variable
m	measured value	Y	dependent variable
N	number of observations	y_i	is the value of the target field
N^H	number of hidden layer neurons	w	the weight for ANN
N^I	number of inputs	ν	kinematic viscosity
N^{TR}	umber of training samples	ω_i	the value of the weighting field for record i
$N_W(t)$	the weighted number of records	μ	dynamic viscosity
MLP	multi layer perceptron	σ	surface tension
o_i	output value of model	ρ	mass density
Q_i	intake discharge		

generalize and power to approximate nonlinear and complex system such as estimation of the critical submergence) have been widely used in many engineering problems.

About CART applications in the sciences as a new model can mention to soil properties prediction in environmental science (Henderson, Bui, Moran, & Simon, 2005), risk management analysis in petroleum pipeline construction (Dey, 2002), prediction of significant wave height (Mahjoobi & Etemad-Shahidi, 2008) and estimation of wave-induced scour around a pile (Ayoubloo & Etemad-Shahidi A., 2008b; Ayoubloo, Etemad-Shahidi, & Mahjoobi, 2010). ANN models have been used to estimate scour around piles (Kambekar & Deo, 2003), below spillways (Azamathulla, Deo, and Deolalikar, 2005, 2008), downstream of grade-control structures (Guven & Gunel, 2008) and the critical submergence of a vertical intake (Kocabas, Unal, & Unal, 2008). Also, combinations of fuzzy inference system with ANN (ANFIS) have been employed to predict wave characteristics (Kazeminezkad, Etemad-Shahidi, & Mousavi, 2005; Mahjoobi, Etemad-Shahidi, & Kazeminezhad 2008), water level in reservoir (Chang & Chang, 2005) and pile group scour (Bateni & Jeng, 2007). All aforementioned studies presented CART and ANN as the applicable and powerful tools compared with traditional regression schemes.

This study aims to investigate the skills of CART and ANN methods in prediction of the critical submergence of horizontal intakes in open channel flows and to compare with those obtained from well known empirical formulas. As a conventional method, linear regression-based equations, predicting the critical submergence, are also developed.

The two recently introduced soft computing techniques i.e. CART and ANN and linear regression are first briefly described. This is followed by a short discussion on the used experimental data set involving a total of 324 experimental data for the different bottom clearance. Afterwards, governing nondimensional parameters are discussed. Finally, the two developed data mining approaches and linear regression are applied to predict the scour depth to illustrate their efficiency and performance. The performance metrics of the employed methods for solving this problem are presented in forms of quantitative statistical measures and scatter plots and compared with those obtained from well known empirical formulae. The results show that the proposed methods are capable of predicting the critical submergence with an acceptable degree of accuracy. Furthermore, the excellent association between observed and committee model-yielded the critical submergence has proven the superior performance of this technique over all presented soft computing methods as well as commonplace empirical techniques.

1.1. Regression trees and CART algorithm

The classification and regression trees (CART) method of Breiman, Friedman, Olshen, and Stone (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:

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

$$\bar{y}(t) = \frac{1}{N_W(t)} \sum_{i \in t} \omega_i f_i y_i, \tag{2}$$

$$N_W(t) = \sum_{i \in t} \omega_i f_i, \tag{3}$$

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:

$$Q(s,t) = R(t) - R(t_L) - R(t_R), \tag{4}$$

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 $Q(s, t)$.