Engineering Geology 170 (2014) 1–10

Contents lists available at ScienceDirect

Engineering Geology

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

An improvement to MLR model for predicting liquefaction-induced lateral spread using multivariate adaptive regression splines

Anthony.T.C. Goh *, W.G. Zhang

School of Civil & Environmental Engineering, Nanyang Technological University, Nanyang Avenue, 639798, Singapore

ARTICLE INFO

Article history:
Received 15 July 2013
Received in revised form 9 December 2013
Accepted 14 December 2013
Available online 22 December 2013

Keywords:
Multiple linear regression
Multivariate adaptive regression splines
Lateral spread
Liquefaction
Nonlinear model
Multi-variable problem

ABSTRACT

Soil liquefaction during earthquakes can result in ground movements that cause damage to buildings and lifelines. Lateral spreading is one form of earthquake-induced ground movements that have caused extensive damage in previous earthquakes. The lateral displacement is dependent on many factors including the earthquake magnitude, thickness and particle size of the liquefiable subsoils and the depth of the groundwater. A number of analytical and empirical methods have been proposed to predict the magnitude of the lateral displacement. One common empirical method is the MLR model which is based on multiple linear regression (MLR) analysis of a database of observed case histories. It is proposed in this paper to use a nonparametric regression procedure known as multivariate adaptive regression splines (MARS), as an improvement to the current MLR model to predict the liquefaction induced lateral displacement. First the basis of the MARS method and its associated procedures are explained in detail. Results are then presented to show the accuracy of the proposed approach, in comparison to the commonly used multiple regression approach. Analysis of observed case histories indicated that the MARS outperforms MLR in terms of predictive accuracy. MARS automatically models non-linearities and interactions between variables without making any specific assumptions. Furthermore, it is able to provide the relative importance of the input variables and give insights of where significant changes in the data may occur.

1. Introduction

During an earthquake, liquefaction occurs in saturated sand deposits, due to excess pore water pressure increase. It can cause serious to destructive damage to structures. The liquefaction mechanism includes ground subsidence, flow failure and lateral spreading, among other effects. Lateral spreading involves the movement of relatively intact soil blocks on a layer of liquefied soil towards a free face or down a gentle slope. It can also induce different forms of ground deformation, the magnitudes of which range from a few centimeters to several meters. Susceptibility to liquefaction-induced lateral spreading is dependent on a number of factors such as the depth of the groundwater table, the physical and mechanical properties of the subsoils, and the intensity and duration of the ground shaking. The large number of factors involved presents challenges in developing simplified analytical solutions to estimate the magnitude of the lateral displacement. A rigorous numerical model must consider dynamic and three dimensional effects as well as the anisotropic and heterogeneous nature of liquefiable soil deposits. Moreover, accurate constitutive modeling of a liquefiable soil is a difficult problem, even when considerable laboratory testing is undertaken. Such efforts are hampered by the difficulty in obtaining representative, "undisturbed" testing samples from the in situ deposit. In practice, empirical models based on case histories have remained the more popular assessment method in the past decades.

Various empirical approaches have also been proposed (e.g., Hamada et al., 1986; Youd and Perkins, 1987; Bartlett and Youd, 1992a, 1992b, 1995; Shamoto et al., 1998; Bardet et al., 1999, 2002; Rauch and Martin, 2000; Youd et al., 2002; Zhang et al., 2004; Al Bawwab, 2005; Zhang and Zhao, 2005; Javadi et al., 2006; Kanibir et al., 2006; Aydan et al., 2008). Table 1 shows some of the more common empirical models. The corresponding parameter descriptions are listed in Table 2.

The most widely used method is the multiple linear regression (MLR) approach originally proposed by Bartlett and Youd (1995) in which two different site conditions are considered: (1) lateral spread towards a free face (e.g., river) and (2) lateral spread down gentle ground slopes where a free face is absent or far away. Based on database records from case histories, empirical models were developed for estimating horizontal ground displacement from liquefaction-induced lateral spread. The original procedure was later revised (Youd et al., 1999). The most updated version of the equations is as follows (Youd et al., 2002).

For free-face (ff) conditions:

$$\log D_h = -16.713 + 1.532M - 0.012R - 1.406\log(R^*) + 0.592\log(W) + 0.540\log(T_{15}) + 3.413\log(100 - F_{15}) - 0.795\log(D50_{15} + 0.1) \quad (1)$$

* Corresponding author. Tel.: +65 6790 5271.
E-mail address: ctcgoh@ntu.edu.sg (A.T.C. Goh).

0013-7952/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.enggeo.2013.12.003

Table 1
Some empirical models for prediction of the lateral spread.

Method	Subset	Model
Hamada et al. (1986)		$D_h = 0.75H^{1/2}\theta^{1/3}$ (*H* is the thickness of the liquefied soil, in meters; θ is the slope of either the ground surface or the base of the liquefied soil, in percent)
Youd and Perkins (1987)		$\log LSI = -3.49 - 1.86\log R + 0.98M$ (Liquefaction Severity Index *LSI* is defined as the general maximum magnitude of ground failure displacement, measured in millimeters divided by 25)
Bardet et al. (1999)	ff	$\log(D_h+0.01) = -17.372 + 1.248M - 0.923\log R - 0.014R + 0.685\log W + 0.3\log T_{15} + 4.826\log(100-F_{15}) - 1.091D50_{15}$
	gs	$\log(D_h+0.01) = -14.152 + 0.988M - 1.049\log R - 0.011R + 0.318\log S + 0.619\log T_{15} + 4.287\log(100-F_{15}) - 0.705D50_{15}$
Youd et al. (2002)	ff	$\log D_h = -16.713 + 1.532M - 1.406\log R^* - 0.012R + 0.592\log W + 0.540\log T_{15} + 3.413\log(100-F_{15}) - 0.795\log(D50_{15} + 0.1\,\text{mm})$
	gs	$\log D_h = -16.213 + 1.532M - 1.406\log R^* - 0.012R + 0.338\log S + 0.540\log T_{15} + 3.413\log(100-F_{15}) - 0.795\log(D50_{15} + 0.1\,\text{mm})$
Javadi et al. (2006)	ff	$D_h = -234.1\frac{1}{M^2RW} - 156\frac{1}{M^2} - 0.008\frac{F_{15}}{R^2T_{15}} + 0.01\frac{WT_{15}}{R} - 2.9\frac{1}{F_{15}} - 0.036\frac{MT_{15}^2D50_{15}^2}{R^2W} + 9.4\frac{M}{RF_{15}} - 4\times10^{-6}\frac{MR^2}{D50_{15}} + 3.84$
	gs	$D_h = -0.027\frac{T_{15}^2F_{15}}{M^2} + 0.05\frac{RT_{15}}{M^2D50_{15}} + 0.44\frac{1}{MR^2ST_{15}} - 0.03R - 0.02\frac{M}{ST_{15}} - 5\times10^{-5}\frac{MR}{D50_{15}^2} + 0.075M^2 - 2.4$

Note: ff is for free-face condition; gs is for gently sloping ground.

and for gently sloping (gs) ground:

$$\log D_h = -16.213 + 1.532M - 0.012R - 1.406\log(R^*) + 0.338\log(S) + 0.540\log(T_{15}) + 3.413\log(100-F_{15}) - 0.795\log(D50_{15}+0.1). \quad (2)$$

The MLR model of Youd et al. (2002) is the most commonly used approach to estimate the liquefaction-induced lateral spread, because of the simplicity of the mathematical model and the easy interpretability of the input variables. However, its predictive capacity can only be considered as satisfactory with a coefficient of determination value R^2 of less than 0.84 (as discussed in Section 3.2), and it has been found to be less accurate for displacements smaller than 1.5 m. In addition, the MLR model assumes independence of the input variables and does not reflect the interaction/correlation effects between the seismic, geometric and soil parameters.

A soft computing technique known as artificial neural networks (ANN) has also been successfully applied to estimate the lateral displacement based on case records (Wang and Rahman, 1999; Chiru-Danzer et al., 2001). The main advantage of ANN over other regression techniques is the ability to capture and represent the nonlinear interaction among the multitude of variables of the system without having to assume the form of the relationship between the variables. Some limitations of neural networks include model interpretability, computational intensity, slow convergence speed and over-fitting problems.

In this paper, the liquefaction-induced lateral spread database used by Youd et al. (2002) has been reanalyzed using a procedure developed by Friedman (1991) known as multivariate adaptive regression splines (MARS). No prior knowledge of the form of the function is required in MARS. Besides the good predictive accuracy of MARS, its other advantages include its capacity to find the complex data mapping in high-dimensional data and produce simple, much easier to interpret models, its processing speed and its ability to estimate the relative importance of the input variables.

2. Details of MARS

Friedman (1991) introduced MARS as a statistical method for fitting the relationship between a set of input variables and dependent variables. MARS is a nonlinear nonparametric method based on a divide and conquer strategy in which the training data sets are partitioned into separate piecewise linear segments (splines) of differing gradients (slope). No specific assumption about the underlying functional relationship between the input variables and the output is required. The end points of the segments are called knots. A knot marks the end of one region of data and the beginning of another. The resulting piecewise curves (known as basis functions), give greater flexibility to the model, allowing for bends, thresholds, and other departures from linear functions.

MARS generates basis functions by searching in a stepwise manner. It searches over all possible univariate knot locations and across interactions among all variables. An adaptive regression algorithm is used for selecting the knot locations. MARS models are constructed in a two-phase procedure. The forward phase adds functions and finds potential knots to improve the performance, resulting in an overfitting model. The backward phase involves pruning the least effective terms. An open MARS source code from Jekabsons (2010) is used in carrying out the analyses presented in this paper.

Table 2
Parameters and parameter descriptions used in Youd et al. (2002).

Parameters		Parameter description
Output	$\log(D_h)$	D_h, the estimated lateral ground displacement, in meters
Input	M	The moment magnitude of the earthquake
	R	The nearest horizontal distance from the site to the seismic energy source, in kilometers
	$\log(R^*)$	$R^* = 10^{(0.89M - 5.64)} + R$, the modified source distance, in kilometers
	$\log(W)$ or $\log(S)$	$\log(W)$ for free face condition, W, the free-face ratio defined as the height (H) of the free face divided by the horizontal distance (L) from the base of the free face to the point in question, in percent; $\log(S)$ for gentle slope ground, S, the ground slope, in percent
	$\log(T_{15})$	T_{15} is the cumulative thickness of saturated granular layers with corrected blow counts, $(N_1)_{60}$, less than 15, in meters
	$\log(100 - F_{15})$	F_{15} is the average fines content for granular materials included within T_{15}, in percent
	$\log(D50_{15} + 0.1)$	$D50_{15}$ is the average mean grain size for granular materials within T_{15}, in millimeters