EVALUATION OF LIQUEFACTION POTENTIAL OF SOIL BASED ON STANDARD PENETRATION TEST USING MULTIVARIATE ADAPTIVE REGRESSION SPLINES & MULTI-GENE GENETIC PROGRAMMING

Rupashree Sahoo
Assistant Deputy Director, Department of Water Resources, Government of Odisha, Bhubaneshwar – 751 001.
rupashree.lucy@gmail.com

Sarat Kumar Das
Associate Professor, Department of Civil Engineering, NIT Rourkela, Sundergarh – 769 008.
saratdas@rediffmail.com

ABSTRACT: The present study discusses the evaluation of liquefaction potential of soil based on post liquefaction standard penetration test (SPT) database available in literature using two artificial intelligence techniques; multivariate adaptive regression splines (MARS) and multi-gene genetic programming (MGGP). Two models are formed by taking different input variables for predicting liquefaction susceptibility of soils. For the first model the input variables are critical depth (D), critical N-value (Nm), fine content (FC), clay size content (CC), D50, peak ground acceleration (a_max), cyclic stress ratio (CSR), critical (N1)_{so} value. Similarly, for second model the input variables are Cyclic Stress Ratio (CSR) and Critical (N1)_{so} value. The overall success rate for first model is obtained as 96.53% and 96.88% from MARS and MGGP respectively, which are better than the available results obtained by ANN (88.37%) and SVM (94.19%) models. Similarly, for second model the overall success rate is 92.36% for MARS model which is better than the available ANN (88.37%) model. A comparative study is also made between the results obtained by MARS and MGGP model and it is found that the MGGP model is more efficient than MARS model.

Keywords: Liquefaction Susceptibility, Standard Penetration Test, MARS, MGGP.

1 Introduction

Liquefaction of saturated sandy soils during earthquakes causes building settlement or tipping, sand blows, lateral spreading, ground cracks, landslides, dam instability, high embankment failures and other hazards. Prediction of liquefaction susceptibility of saturated sandy soils due to an earthquake is an important task in earthquake geotechnical engineering. Since it is very difficult to get high-quality undisturbed samples of sandy soils, in situ tests have been used to determine the liquefaction resistance of saturated sandy soils. The standard penetration test (SPT) data is the most common method to predict the liquefaction potential of soil (Seed and De Alba 1986). Because of its reliability, speed, economy and continuity of profiling, the cone penetration test (CPT) is considered a superior technique than SPT. Liquefaction analysis based on probabilistic and statistical methods have been done by many researchers. Recently, artificial neural network (ANN) (Goh 1994; Juang et al. 2000; Hanna et al. 2007; Samui and Sitharam 2011) model, support vector machine (SVM) (Pal 2006; Goh and Goh 2007; Samui and Sitharam 2011) model are found to more efficient in predicting the liquefaction susceptibility compared to empirical methods. Another technique, called the genetic programming (GP), developed by Koza (1992), mimics biological evolution of living organisms and makes use of principle of genetic algorithm (GA). It is also called as ‘grey box’ model. GP helps in achieving greatly simplified model formula compared to ANN model, but a trade-off is made between the complexity of the formula and accuracy of the model. Another class of model may be termed as ‘white box’ model is the multivariate adaptive regression spline (MARS) developed based on statistical model developed by Friedman (1991). MARS can adjust any functional form, hence suitable for exploratory data analysis. Samui et al. (2011) observed that the MARS model for uplift capacity of suction caisson has better statistical performance comparable to ANN and FEM model. In the present study an attempt has been made to evaluate the liquefaction of soil based on post liquefaction SPT database present in literature (Hwang and Yang 2001) using MARS and GP. A comparative study among the developed MARS model, GP model and the available ANN model are made in terms of rate of successful liquefaction and non-liquefaction cases.

2 Methodology

In the present study liquefaction index (LI) is predicted by MARS and GP techniques to separate liquefaction
and non-liquefaction cases. MARS, and variant of

gene programming multi-gene genetic programming
(MGGP) have been used in limited geotechnical
engineering problems, and are not very common to
gene programming professionals, hence are

discussed in brief as follows.

2.1 Multivariate Adaptive Regression Splines
(MARS)

MARS is a nonparametric regression procedure
proposed by Friedman (1991) that does not assume any
functional relationship between independent and
dependent variables. MARS uses the regression data to
construct this relation and forms some sets of
coefficients and basis functions. In other words, it can
be said that this method is based on “divide and
conquer” strategy, which divides the input parameters
into groups or say regions, each having its own
regression equation. A special advantage of MARS lies
in its ability to estimate the contributions of the basis
functions so that both the additive and the interactive
effects of the predictors are allowed to determine the
response variable. The general form of a MARS
predictor is as follows:

\[ F(x) = \beta + \sum_{i=1}^{n} B_i a_i \]  

(1)

Where

\[ \beta = \text{Intercept}, \ B_i = \text{Basis functions}, \ a_i = \text{Coefficient of} \]

\[ \text{Basis function} I, \ n = \text{no of basis functions.} \]

The MARS algorithm for estimating the model function
F(x) consists of two algorithms (Friedman 1991)

(i) The forward stepwise algorithm: Here, forward
stepwise search for the basis function takes place with
the constant basis function, the only one present
initially, and the process stops when a user-specified
maximum basis function value is reached. This model
typically over fits the data and so a backward deletion
procedure is applied.

(ii) The backward stepwise algorithm: The purpose of
this algorithm is to prevent from over-fitting by
decreasing the complexity of the model without
degrading the fit to the data.

In the present study the subroutine ‘EARTH’ written in
“R” language is used to implement the MARS method.
The R is a system for statistical computation and
graphics, used in various statistical problems related to
engineering, medical, economics etc. It is very easy to
work on R and consists of small syntax and can take data
from standard spread sheet.

2.2 Genetic Programming

Genetic programming model is composed of nodes,
which resembles a tree structure and, thus, it is also
known as a GP tree. Nodes are the elements either from a
functional set or terminal set. A functional set may
include arithmetic operators (+, *, Ï„, or –), mathematical
functions \( \sin(\cdot), \cos(\cdot), \tanh(\cdot) \) or \( \ln(\cdot) \), Boolean
operators (AND, OR, NOT, etc.), logical expressions
(IF, or THEN) or any other suitable functions defined by
the user. The terminal set includes variables (like \( x_1, \ x_2, \ x_3, \ etc. \)) or constants (like 3, 5, 6, 9, etc.) or both. The
functions and terminals are randomly chosen to form a
GP tree with a root node and the branches extending from
each function nodes to end in terminal nodes.

Initially, a set of GP trees, as per user defined population
size, are randomly generated using various functions
and terminals assigned by the user. The fitness criterion
is calculated by the objective function and it determines
the quality of each individual in the population
competing with rest. At each generation a new
population is created by selecting individuals as per the
merit of their fitness from the initial population and then
implementing various evolutionary mechanisms like
reproduction, crossover and mutation to the functions
and terminals of the selected GP trees. The new
population then replaces the existing population. This
process is iterated until the termination criterion, which
can be either a threshold fitness value or maximum
number of generations, is satisfied.

2.3 Multi-gene genetic programming

Multi-gene genetic programming (MGGP) is a variant
of GP and is designed to develop input output
relationship of a system in terms of empirical
mathematical model which is weighted linear
combination of outputs from a number of GP trees. It is
also referred as symbolic regression. Each tree
represents lower order nonlinear transformations of
input variables termed as “gene” and the linear
combination of these genes are termed as “multi gene”.

In the MGGP model development it is important to
make a tradeoff between accuracy and complexity in
terms of maximum allowable number of genes \( (G_{max}) \)
and maximum depth of GP tree \( (d_{max}) \) (Searson et al.
2010). The linear coefficients termed as weights of
Gene-1 and Gene-2 \( (c_1 \ and \ c_2) \) and the bias \( (c_0) \) of the
model are obtained from the training data using
statistical regression analysis (ordinary least square
method). In addition to the standard GP evolution
mechanisms as discussed earlier, there are some special
MGGP crossover mechanisms (Searson et al. 2010) which
allow the exchange of genes between individuals.
Similarly, MGGP also provides six methods of mutation.
of genes (Gandomi and Alavi 2012b). The general form of MGGP based model of the present study can be presented as:

\[ L_{IP} = \sum_{i=1}^{n} f[X, f(X), b_i] + b_0 \]  

(2)

where \( L_{IP} \) is the predicted value of liquefaction index (LI), \( F \) is the function created by the MGGP process referred herein as liquefaction index function, \( X \) is the vector of input variables = \{(\( N_1 \) ), CSR7.5\}, where \((\( N_1 \) )\) is the corrected blow count and CSR7.5 is the cyclic stress ratio adjusted to the benchmark earthquake (moment magnitude, \( M_w \), of 7.5) as presented by Juang et al. (2000), \( b_i \) is constant, \( f \) is the MGGP function defined by the user, and \( n \) is the number of terms of model equation. The MGGP as per Searson et al. (2010) is used and the present model is developed and implemented using Matlab (MathWorks Inc 2010).

3 Results and Discussion

In the present study, database comprising of 288 numbers of Standard Penetration Test (SPT), of liquefaction case history of Chi-Chi, Tawaiin, 1999 earthquake is used (Hwang and Yang 2001). Out of these 288 cases, 164 cases are identified as liquefied cases and 124 are non-liquefied cases, based on the field test values. Here 202 cases are selected randomly for training and the rest 88 are selected for testing. Samui and Sitharam (2011) and Muduli and Das (2013) also have used the above stated database with the same number of training and testing data while developing ANN, SVM and MGGP based liquefaction models. In case of MARS and MGGP approach, normalization or scaling of data sets are not required as in the case of ANN and SVM approach. For MARS modelling of liquefaction index, following assumptions or predictions were made:

For all liquefied cases, the liquefaction index (LI) is assumed as 1 and for non-liquefied cases, the LI = 0. The Training and Testing performance (%) are calculated by using the following formula:

\[ \text{Training Performance (\%) or Testing Performance (\%)} = \left( \frac{\text{No of data predicted accurately by MARS}}{\text{Total data}} \right) \]  

(3)

The various statistical parameters coefficient of efficiency (E), average absolute error (AAE) (Das and Basudhar 2008) are also used to compare the results.

3.1 MARS Modelling for Liquefaction Index (LI).

For predicting the liquefaction index using MARS modelling, two models were considered, which were chosen on the basis of number of input parameters considered for modelling.

The successful prediction rates for liquefied cases are shown in Table 1. It can be seen from the above table that Model 1 gives better prediction accuracy for the occurrence of liquefaction than the Model 2. The various statistical performances of the two models are depicted in Table 2. The final equation for prediction of Liquefaction Index based on MARS model no 1 is given as:

\[ LI = 1.143 + \sum_{i=1}^{7} B_i \times a_i \]  

(4)

Where \( B_i \) are the basis functions and \( a_i \) are the constants.

Table 1 MARS modelling showing input parameters, overall, training and testing performances.

<table>
<thead>
<tr>
<th>Model No</th>
<th>Input parameters</th>
<th>Overall Performance (%)</th>
<th>Training Performance (%)</th>
<th>Testing Performance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Depth, Nm, FC, CC, a max, CSR, N1(60), LIm</td>
<td>96.53</td>
<td>92.1</td>
<td>90.37</td>
</tr>
<tr>
<td>2</td>
<td>CSR, N1(60)</td>
<td>92.36</td>
<td>91.58</td>
<td>95.35</td>
</tr>
</tbody>
</table>

Table 2 statistical performances obtained by MARS modelling.

<table>
<thead>
<tr>
<th>Statistical Performances</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Performance (%)</td>
<td>96.53</td>
<td>92.36</td>
</tr>
<tr>
<td>R</td>
<td>0.93</td>
<td>0.928</td>
</tr>
<tr>
<td>E</td>
<td>0.859</td>
<td>0.855</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.186</td>
<td>0.187</td>
</tr>
<tr>
<td>AAE</td>
<td>0.035</td>
<td>0.035</td>
</tr>
</tbody>
</table>

From the above table it is concluded that, based on sensitivity analysis, CSR is the most important variables among all.

3.2 MGGP Modelling for Liquefaction Index (LI).

For predicting the liquefaction index using MGGP modelling, only one model was formulated by considering all the input variables. The successful prediction rates for liquefied cases are given in Table 7. The equation for prediction of Liquefaction Index based on MGGP model 1 is given below:
Evaluation of liquefaction potential of soil based on standard penetration test using Multivariate Adaptive Regression Splines & Multi-Gene Genetic Programming

\[ LI = 0.03 \times D_{50} - 0.03 \times D - \frac{2.96}{6_{10.53}} + 0.02 \times (D_{50} + (N_i)_{so}) \times (CSR - 4.732) + 0.03 \times a_{max}(N_i)_{so} - \frac{8.769 \times CSR}{(CSR + (N_i)_{so})} + \]

\[ \frac{0.0002 \times (N_i)_{so}(N_m - CC)}{a_{max}} + 2.425 \]  

(5)

Generally, the efficiency of different models are compared firstly on testing data then on training data (Das and Basudhar 2008). It is evident from the above table that, for the same database as presented by (Samui and Sitharam 2011) the accuracy of prediction of liquefaction susceptibility of soil based on ANN model were 94.55 and 83.77 % for training and testing respectively.

Table 7 MGGP modelling showing input parameters, overall, training and testing performances.

<table>
<thead>
<tr>
<th>Input parameters</th>
<th>Overall Performance</th>
<th>Training Performance</th>
<th>Testing Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth, ( N_m ), FC, CC, ( a_{max} ), CSR, ( N_{(60)} )</td>
<td>96.88</td>
<td>97.03</td>
<td>96.51</td>
</tr>
</tbody>
</table>

In case of MARS the accuracy of prediction by model no 1 were found out to be 92.1 and 90.37 for training and testing respectively, whereas in the case of model no 2 the accuracy of prediction obtained were 91.58 and 95.35 % respectively. Similarly, in case of MGGP the accuracy of prediction by model no 1 were found out to be 97.03 and 96.51 respectively, whereas in case of model no 2, the accuracy of prediction were 94.55 and 94.19. Thus, it can be stated that MGGP model in both the cases outperforms the MARS and ANN Models.

4 Conclusions

The present paper successfully adopted MARS and GP for prediction of Liquefaction Index of a soil. The developed models using MARS and MGGP have shown good predictive abilities than ANN, but in comparison between the four above stated methods, the developed MGGP models outperforms the MARS and ANN models. The equation developed (Equation 5) by MGGP model can be helpful to the geotechnical engineers for predicting liquefaction susceptibility of soils.

5 References


