GEOSCIENCE FRONTIERS 3(4) (2012) 541–555

available at www.sciencedirect.com

China University of Geosciences (Beijing)

GEOSCIENCE FRONTIERS

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

RESEARCH PAPER

Energy-based numerical models for assessment of soil liquefaction

Amir Hossein Alavi [a,*], **Amir Hossein Gandomi** [b]

[a] *School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran*
[b] *College of Civil Engineering, Tafresh University, Tafresh, Iran*

Received 20 September 2011; received in revised form 5 December 2011; accepted 9 December 2011
Available online 28 December 2011

KEYWORDS
Soil liquefaction;
Capacity energy;
Linear genetic programming;
Multi expression programming;
Sand;
Formulation

Abstract This study presents promising variants of genetic programming (GP), namely linear genetic programming (LGP) and multi expression programming (MEP) to evaluate the liquefaction resistance of sandy soils. Generalized LGP and MEP-based relationships were developed between the strain energy density required to trigger liquefaction (capacity energy) and the factors affecting the liquefaction characteristics of sands. The correlations were established based on well established and widely dispersed experimental results obtained from the literature. To verify the applicability of the derived models, they were employed to estimate the capacity energy values of parts of the test results that were not included in the analysis. The external validation of the models was verified using statistical criteria recommended by researchers. Sensitivity and parametric analyses were performed for further verification of the correlations. The results indicate that the proposed correlations are effectively capable of capturing the liquefaction resistance of a number of sandy soils. The developed correlations provide a significantly better prediction performance than the models found in the literature. Furthermore, the best LGP and MEP models perform superior than the optimal traditional GP model. The verification phases confirm the efficiency of the derived correlations for their general application to the assessment of the strain energy at the onset of liquefaction.

* Corresponding author.
E-mail addresses: ah_alavi@hotmail.com, am_alavi@civileng.iust.ac.ir (A.H. Alavi), a.h.gandomi@gmail.com, ag72@uakron.edu (A.H. Gandomi).

Peer-review under responsibility of China University of Geosciences (Beijing).
doi:10.1016/j.gsf.2011.12.008

Production and hosting by Elsevier

1. Introduction

Soil liquefaction is one of the most complex phenomena studied in geotechnical earthquake engineering. Liquefaction is commonly considered as a specific feature of loose and saturated sandy soils. Liquefaction usually occurs when the pore water pressure increases to carry the overburden stress. Therefore, soil immediately loses most of its strength leading to extreme deformations, flow of water and suspension of sediment (Darve, 1996). Numerous studies have focused on analyzing the liquefaction phenomenon since it is one of the major sources for failures of critical structures. Several procedures are developed to evaluate

the liquefaction potential in the field. The available liquefaction evaluation procedures are categorized into three main groups (Green, 2001): (1) stress-based procedures, (2) strain-based procedures, and (3) energy-based procedures. The stress-based procedure (Seed and Idriss, 1971) is the most widely-used liquefaction assessment method. This approach is mainly empirical and based on laboratory and field observations. The stress method has continually been refined as a result of newer studies and increase in the number of liquefaction case histories (e.g., Youd et al., 2001). The main criteria in the stress-based procedure are the shear stress level and number of cycles. Despite the continuous revisions and extensions of the stress-based method, the uncertainty on the subject of random loading still exists (Green, 2001; Baziar and Jafarian, 2007). Dobry et al. (1982) proposed the strain-based procedure as an alternative to the empirical stress-based procedure. This method was derived from the mechanics of two interacting idealized sand grains and then generalized for natural soil deposits (Green, 2001; Baziar and Jafarian, 2007).

The energy concept has widely been used in the theories of elasticity and plasticity, potential energy surface for constitutive law and energy principles (Desai and Siriwardane, 1984). The basic elements of both the stress and strain methods are incorporated in the formulation of the energy-based method. In this method, the amount of total strain energy at the onset of liquefaction is obtained from laboratory testing or field recorded data. In a typical cyclic (triaxial or simple shear) laboratory test, the stress, strain and pore pressure time histories are recorded. Hysteresis loops can be generated from these stress and strain time histories. Fig. 1 illustrates a typical hysteresis loop from a typical stress-controlled cyclic triaxial test. The strain energy for each cycle of loading is equivalent to the area inside the hysteresis loop (Ostadan et al., 1996). In other words, this area represents the dissipated energy per unit volume of the soil mass (Green, 2001). This is based on the idea that during deformation of cohesionless soils under dynamic loads part of the energy is dissipated into the soil (Nemat-Nasser and Shokooh, 1979). The instantaneous energy and its summation over time intervals are computed until the onset of liquefaction. The summation of the energy at this time is used as the measures of the capacity of the soil sample against initial liquefaction occurrence in terms of the strain energy (capacity energy).

To predict liquefaction, this strain energy is compared with the strain energy imparted by earthquake to the sand layer during the seismic design event. The experiments revealed that the build-up of the excess pore pressure is proportional to the total strain energy in all loading cycles up of the initial liquefaction. This observation has prompted the formulation of the energy-based approach. Since the late 1970s, numerous energy-based procedures have been proposed for evaluating the liquefaction potential of soil deposits (Liang, 1995; Green, 2001). The use of strain energy concept is a logical step in the evolution of liquefaction evaluation of soils for two reasons (Baziar and Jafarian, 2007). The first reason is that seismologists have long been quantifying the energy released during earthquakes and have determined simple correlations with common seismological parameters. The second reason is that some pioneer researchers developed functional relationships correlating the energy density dissipated into the cohesionless soils to the pore pressure build-up (Nemat-Nasser and Shokooh, 1979).

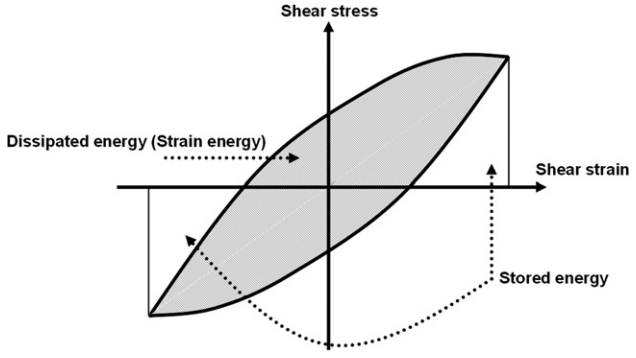

Figure 1 A typical hysteresis shear stress–strain loop (Green, 2001).

The energy-based approach has several advantages in comparison with the other existing methods to evaluate the liquefaction potential of soils. Some of the most important advantages of this approach are well summarized by Voznesenskya and Nordal (1999) and Dief and Figueroa (2001). However, the complexity of the liquefaction behavior suggests the necessity of developing more comprehensive models to assess it.

Genetic programming (GP) (Koza, 1992; Banzhaf et al., 1998) is a developing subarea of evolutionary algorithms inspired from the Darwin's evolution theory. GP may generally be defined as a specialization of genetic algorithms (GA) where the solutions are computer programs rather than binary strings. Linear genetic programming (LGP) (Brameier and Banzhaf, 2007) is a new branch of GP. LGP operates on programs represented as linear sequences of instructions of an imperative programming language (Brameier and Banzhaf, 2007). Multi expression programming (MEP) (Oltean and Dumitrescu, 2002) is another recent variant of GP that uses a linear representation of chromosomes. The modeling capabilities of LGP and MEP have been shown by researchers (Oltean and Grosşan, 2003; Baykasoglu et al., 2008). In contrast with traditional GP and other soft computing tools, applications of LGP and MEP in the field of civil engineering are new and restricted to a few areas (Alavi et al., 2010a; Gandomi et al., 2010a; Alavi and Gandomi, 2011).

In this research, the LGP and MEP techniques were utilized to obtain generalized relationships between the energy per unit volume dissipated during liquefaction and the soil initial parameters. A traditional GP analysis was performed to benchmark the LGP and MEP-based correlations. Further, the prediction performance of the derived correlations was compared with that of different models found in the literature.

2. Review of energy-based liquefaction evaluation models

Contrary to the stress-based and strain-based approaches, the energy-based procedures use various measures of energy as the base parameters to quantify demand (the load imparted to the soil by the earthquake) and capacity (the demand required to induce liquefaction). The energy-based liquefaction evaluation procedures are mainly grouped into approaches developed using earthquake case histories, and those developed from laboratory data (Green, 2001).

2.1. Analytical and empirical models

Numerous researches are conducted to develop energy-based models for the evaluation of the liquefaction potential (Towhata and Ishihara, 1985; Liang et al., 1995). The necessity to obtain

calibration parameters for many of the existing pore pressure models limits their usefulness (Baziar and Jafarian, 2007). In recent years, Green et al. (2000) developed an energy-based model on the basis of the stress-controlled cyclic triaxial test data on sand samples. Several models were developed relating the soil capacity energy, shear strain amplitude, and some of the sandy soil initial parameters on the basis of a series of laboratory cyclic shear and centrifuge tests (Figueroa et al., 1994; Liang, 1995; Rokoff, 1999; Dief and Figueroa, 2001). Most of these relationships were derived by performing a multiple linear regression (MLR) analysis. Some of the most well-known models in this field and their corresponding correlation coefficient (R) values are shown in Table 1. Only two groups of researchers, (Wang et al., 1997; Baziar and Jafarian, 2007), have taken into account the important role of the fines content in the evaluation of the liquefaction behavior.

2.2. Soft computing-based models

Several computer-aided pattern recognition, data classification and soft computing approaches have been recently employed for solving problems in civil engineering. Artificial neural networks (ANNs), support vector machine (SVM), relevance vector machine (RVM), and Bayesian updating are well-known branches of such systems. These methods are widely employed for the behavior modeling of different civil engineering tasks (Yilmaz et al., 2002; Al-Anazi and Babadagli, 2010; Ghorbani et al., 2010). Also, they have been used for the evaluation of the liquefaction potential (Goh, 1994; Goh, 2002; Cetin et al., 2004; Pal, 2006; Goh and Goh, 2007; Samui, 2007; Oommen et al., 2008; Oommen and Baise, 2010; Oommen et al., 2010). However, applications of these techniques to the energy-based assessment of the liquefaction resistance are very limited. In this context, Baziar and Jafarian (2007) developed an ANN model for evaluation of the liquefaction potential based on the energy concepts. Chen et al. (2005) presented a seismic wave energy-based method with back-propagation neural networks to assess the liquefaction probability. In that work, back-propagation neural networks were used to simulate Fourier spectrum of seismic wave acceleration. Then, seismic wave energy was obtained by integration of the pseudo-spectrum. Despite the acceptable performance of ANNs, they have some fundamental disadvantages. A notable limitation of ANNs is that they are not usually capable of generating practical prediction equations. Moreover, they require the structure of the network to be identified in advance (Alavi et al., 2011). Recently, Baziar et al. (2011) utilized an evolutionary approach based on GP for estimation of capacity energy of liquefiable soils.

3. Genetic programming

GP is a symbolic optimization technique with a great ability to evolve computer programs based on the Darwin's evolution theory. Koza (1992) introduced GP as an extension of genetic algorithms (GAs). The main difference between GP and GA is related to the representation of the solution. A string of numbers is created by GA to represent the solution, while the GP solutions are computer programs commonly represented as tree structures. GAs are generally used in parameter optimization to evolve the best values for a given set of model parameters. GP, on the other hand, gives the basic structure of the approximation model together with the values of its parameters (Torres et al., 2009; Gandomi and Alavi, 2011).

In addition to traditional tree-based GP, there are other types of GP where programs are represented in different ways. These are linear and graph-based GP (Banzhaf et al., 1998; Poli et al., 2007). The emphasis of the present study is placed on the linear GP techniques. Several linear variants of GP have recently been proposed such as LGP and MEP. The linear variants of GP make a clear distinction between the genotype and phenotype of an individual. In these variants, individuals are represented as linear strings that are decoded and expressed like nonlinear entities (trees) (Oltean and Grosşan, 2003). There are some main reasons for using linear GP. Basic computer architectures are fundamentally the same now as they were twenty years ago, when GP began. Nearly all computer architectures represent programs in a linear fashion. Also, computers do not naturally run tree-shaped programs. Hence, slow interpreters have to be used as part of tree-based GP. Conversely, the use of an expensive interpreter is avoided by evolving the binary bit patterns and the algorithm can run several orders of magnitude faster (Poli et al., 2007; Alavi and Gandomi, 2011).

Table 1 Different energy-based models for liquefaction assessment.

Equation	Authors	Expression	R	Test description
Eq. (1)	Figueroa et al. (1994)	$\text{Log}(W) = 2.002 + 0.00477\sigma'_{\text{mean}} + 0.0116D_{\text{r}}$	0.968	27 torsional shear-controlled strain liquefaction tests on Reid Bedford sand
Eq. (2)	Liang (1995)	$\text{Log}(W) = 2.062 + 0.0039\sigma'_{\text{mean}} + 0.0124D_{\text{r}}$	0.962	9 strain-controlled torsional triaxial experiments on Reid Bedford sand
Eq. (3)	Liang (1995)	$\text{Log}(W) = 2.484 + 0.00471\sigma'_{\text{mean}} + 0.00052D_{\text{r}}$	0.997	13 strain-controlled torsional triaxial experiments on LSFD sand
Eq. (4)	Dief and Figueroa (2001)	$\text{Log}(W) = 1.164 + 0.0124\sigma'_{\text{mean}} + 0.0209D_{\text{r}}$	0.971	30 centrifuge liquefaction tests on Nevada sand
Eq. (5)	Dief and Figueroa (2001)	$\text{Log}(W) = 2.4597 + 0.00448\sigma'_{\text{mean}} + 0.00115D_{\text{r}}$	0.986	30 centrifuge liquefaction tests conducted at a scale of 60g's on Reid Bedford sand
Eq. (6)	Baziar and Jafarian (2007)	$\text{Log}(W) = 2.1028 + 0.004566\sigma'_{\text{mean}} + 0.005685D_{\text{r}} + 0.001821\text{FC} - 0.02868C_{\text{u}} + 2.0214D_{50}$	0.806	284 cyclic triaxial, torsional and simple shear element tests on sands

σ'_{mean} (kPa): soil initial effective mean confining pressure; D_{r} (%): initial relative density after consolidation; FC (%): percentage of fines content; C_{u}: coefficient of uniformity; D_{50} (mm): mean grain size; W (J/m^3): measured strain energy density required for triggering liquefaction.

3.1. Linear genetic programming

LGP has a linear structure similar to the DNA molecule in biological genomes. LGP uses sequences of imperative instructions as genetic material. Typical structures of programs generated by LGP and traditional tree-based GP are shown in Fig. 2. An LGP program can be considered as a data flow graph. This program is represented as a linear sequence of instructions of an imperative programming language (like C/C++) (see Fig. 2a). On the other hand, the structure of the program evolved by tree-based GP is like a tree expressed in the functional programming language (like LISP) (see Fig. 2b) (Brameier and Banzhaf, 2001, 2007).

In the LGP system described here, an individual program is interpreted as a variable-length sequence of simple C instructions. The instruction set or function set of LGP consists of arithmetic operations, conditional branches, and function calls. The terminal set of the system is composed of variables and constants. The instructions are restricted to operations that accept a minimum number of constants or memory variables, called registers (r), and assign the result to a destination register, e.g., $r_0 = r_1 + 1$. A part of a linear genetic program in C code is represented in Fig. 3. In this figure, register $r[0]$ holds the final program output (Gandomi et al., 2010a).

Automatic Induction of Machine code by Genetic Programming (AIMGP) is a particular variant of LGP. AIMGP stores the programs as linear strings of native binary machine code. The evolved programs are directly executed by the processor during the fitness calculation. The AIMGP execution speed is much higher than GP since no interpreter or complex memory handling is involved (Nordin, 1994; Gandomi et al., 2010a). Here are the steps which the modified steady-state machine code LGP algorithm follows for a single run (Brameier and Banzhaf, 2007; Gandomi et al., 2010a):

I. Initializing a population of randomly generated programs and calculating their fitness values.
II. Running a tournament. In this step four programs are selected from the population randomly. They are compared based on their fitness. Two programs are then picked as the winners and two as the losers.
III. Transforming the winner programs. After that, two winner programs are copied and transformed probabilistically into two new programs via crossover and mutation operators.
IV. Replacing the loser programs in the tournament with the transformed winner programs. The winners of the tournament remain unchanged.
V. Repeating steps two through four until termination or convergence conditions are satisfied.

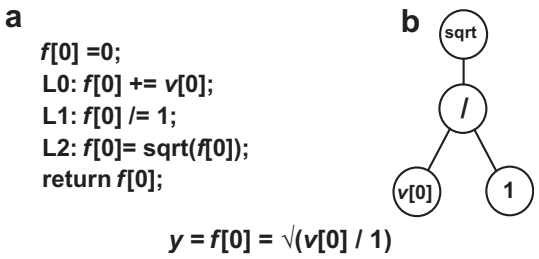

Figure 2 A comparison of the GP program structures. (a): LGP; (b): tree-based GP.

```
L0:  f[0]-=v[0];
L1:  f[0]*=3;
L2:  f[0]/=v[1];
L3:  f[0]+=v[0];
L4:  f[0]/=2;
L5:  f[0]=cos(f[0]);
L6:  f[0]+=f[0];
```

Figure 3 An excerpt of a linear genetic program.

Comprehensive descriptions of the basic parameters used to direct a search for a linear genetic program are provided by Brameier and Banzhaf (2007).

3.2. Multi expression programming

MEP is another subarea of GP. It was first introduced by Oltean and Dumitrescu (2002). Linear chromosomes are used by MEP for solution encoding. This technique encodes multiple computer programs in a single chromosome. A program with the best fitness represents the chromosome. The MEP decoding process is not more complicated than other GP variants storing a single program in a chromosome (Alavi et al., 2010a). The steady-state algorithm of MEP starts by the creation of a random population of computer programs. MEP uses the following steps to evolve the best program until a termination condition is reached (Oltean and Grosşan, 2003; Alavi et al., 2010a):

I. Selection of two parents using a binary tournament procedure (Koza, 1992) and recombination of them with a fixed crossover probability.
II. Obtaining two offspring by the recombination of two parents.
III. Mutation of the offspring and replacement of the worst individual in the current population with the best of them (if the offspring is better than the worst individual in the current population).

The representation of the MEP solutions is similar to the procedure followed by C and Pascal to convert expressions into machine code (Aho et al., 1986). Functions and terminals are a part of a population member created by MEP. The terminal and function symbols are elements in the terminal and function sets, respectively. A function set can contain the basic arithmetic operations or any other mathematical functions. The terminal set can contain numerical constants, logical constants and variables. Each gene encodes a terminal or a function symbol. The first symbol in a chromosome is a terminal symbol. An example of a MEP chromosome is as given below:

1: a
2: b
3: −1, 2
4: /2, 3

The function set for the above example includes "–" and "/". a and b are the elements of the terminal set. The MEP individuals are converted into programs by reading the chromosome top–down starting with the first position. In this example, genes 1 and 2 encode simple expressions which are $E_1 = a$ and $E_2 = b$. Gene 3 indicates the operation "–" on the operands located at positions 1 and 2. Therefore, gene 3 encodes the expression: $E_3 = a - b$. Gene 4 indicates the operation "/" on the operands located at positions 2 and 3. Therefore, gene 4 encodes the expression: $E_4 = b/(a - b)$. Each of the above expressions can be considered as a possible solution. The MEP chromosomes can be illustrated as a forest of trees rather than a single tree because of their multi expression representation (see Fig. 4). The best expression is selected after controlling the fitness of all expression in an MEP chromosome using the following equation (Oltean and Grosşan, 2003):

$$f = \min_{i=1,\,m}\left\{\sum_{j=1}^{n}\left|E_j - O_j^i\right|\right\} \tag{1}$$

in which n is the number of fitness cases; E_j is the expected value for the fitness case j; O_j^i is the value returned for the j^{th} fitness case by the i^{th} expression encoded in the current chromosome, and m is the number of chromosome genes (Alavi et al., 2010a).

4. Developing numerical correlations for the energy-based liquefaction assessment

The mechanical analysis of the liquefaction phenomenon shows that the volume variation rate imposed by the material flow rule (i.e., dilation angle) has to be lower than the volume variation rate imposed by the loading path. In this case, the effective stresses are decreasing possibly to zero. Thus, the dilation angle plays a significant role in liquefaction process (Darve, 1996). According to the experimental and theoretical studies, the dilation angle is mainly influenced by the granular material, relative density and initial confining pressure (Li and Dafalias, 2002). Therefore, in a rational manner the main parameters which affect the liquefaction potential are the grain size distribution of material, fine contents, initial relative density, and initial effective mean confining pressure. This paper considers the feasibility of using the LGP and MEP approaches to obtain meaningful relationships between the level of energy required for the liquefaction of sands and the above mentioned parameters. The LGP and MEP-based relationships were developed using two different combinations of the predictor variables. The first combination consisted of most of the soils initial parameters as follows:

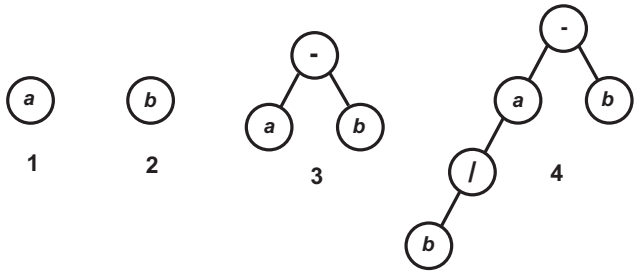

Figure 4 Expressions encoded by an MEP chromosome and represented as trees.

$$\mathrm{Log}(W) = f\left(\sigma'_{\text{mean}},\, D_{\text{r}},\, \text{FC},\, C_{\text{u}},\, D_{50}\right) \tag{2}$$

where W is the measured strain energy density required for triggering liquefaction (capacity energy). This capacity energy is the accumulative area of stress–strain loops up to the liquefaction triggering (see Fig. 1). The input variables used to develop the prediction correlations are listed below:

- Soil initial effective mean confining pressure (σ'_{mean}).
- Initial relative density after consolidation (D_{r}).
- Percentage of fines content (FC).
- Coefficient of uniformity (C_{u}).
- Mean grain size (D_{50}).

σ'_{mean} is related to the initial shearing resistance of the soil and D_{r} represents relative density. FC, C_{u} and D_{50} are the grain size characteristics of soils. The significant influence of σ'_{mean} and D_{r} in determining W is well understood (Nemat-Nasser and Shokooh, 1979; Figueroa et al., 1994; Liang, 1995). The grain size distribution notably affects the liquefaction characteristics of sands (Seed and Idriss, 1971; Figueroa et al., 1998). The strong effect of FC, C_{u} and D_{50} to determine W was previously demonstrated by a few researchers (Figueroa et al., 1998; Baziar and Jafarian, 2007). As expected, coarser soils require higher unit energy for liquefaction than finer soils.

In order to conduct a fair comparison between the results obtained herein and those of the previous studies, the number of the predictor variables was reduced to two parameters, i.e., σ'_{mean} and D_{r}. These parameters are the most widely-used parameters in the available energy-based pore pressure build-up models for the liquefaction assessment. Hence, the formulation of the liquefaction capacity energy was considered to be as follows:

$$\mathrm{Log}(W) = f\left(\sigma'_{\text{mean}},\, D_{\text{r}}\right) \tag{3}$$

The best LGP and MEP-based formulas were chosen on the basis of a multi-objective strategy as given below:

i The simplicity of the model, although this was not a predominant factor.
ii. Providing the best fitness value on the training set of data.

Correlation coefficient (R), root mean squared error (RMSE) and mean absolute error (MAE) were used to evaluate the performance of the proposed correlations. R, RMSE and MAE are given in the form of formulas as follows:

$$R = \frac{\sum_{i=1}^{n}(h_i - \overline{h}_i)(t_i - \overline{t}_i)}{\sqrt{\sum_{i=1}^{n}(h_i - \overline{h}_i)^2 \sum_{i=1}^{n}(t_i - \overline{t}_i)^2}} \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(h_i - t_i)^2}{n}} \tag{5}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|h_i - t_i| \tag{6}$$

where h_i and t_i are, respectively, the actual and predicted output values for the i^{th} output; $\overline{h}_i$ and $\overline{t}_i$ are, respectively, the average of the actual and predicted outputs, and n is the number of samples.

4.1. Experimental database and data preprocessing

A comprehensive database of previously published cyclic tests was used for the development of the proposed correlations (Baziar and Jafarian, 2007). The database consists of 216 cyclic triaxial (Green, 2001), 61 cyclic torsional shear (Towhata and Ishihara, 1985; Liang, 1995), 6 cyclic simple shear (VELACS project) (Arulmoli et al., 1992), and 18 liquefaction triggering centrifuge (Dief, 2000) tests data. The database includes the measurements of several variables such as σ'_{mean} (kPa), D_r (%), FC (%), C_u, D_{50} (mm), and W (J/m^3). To visualize the distribution of the samples, the data are presented by frequency histograms (Fig. 5). Furthermore, the database contains results of some element tests under random loading. Two criteria that indicate the liquefaction triggering are: (1) initial liquefaction ($r_u = 1$) and (2) double amplitude of strain of 5% ($\varepsilon_{DA} = 5\%$), whichever occurs first (Baziar and Jafarian, 2007).

Cross validation is a widely-used method for model evaluation. In the present study, one of the most well-known types of cross validation, called hold-out method was used. This method is based on randomly division of data sets into training and testing subsets. The training data are used for the learning process. The testing data are employed to measure the performance of the obtained model on data that play no role in building it. The hold-out validation avoids the overlap between training data and test data leading to a more accurate estimate for the generalization performance of the algorithm. The advantage of this method is that it takes less time to compute compared with the other cross validation procedures such as K-fold cross validation approach (Refaeilzadeh et al., 2009). However, the evaluation may depend heavily on which data points end up in the training set and which end up in the test set. Thus, the evaluation may be significantly different depending on how the division is made. To deal with this problem, for the LGP and MEP analyses, a trial study was conducted to find a consistent data division. The selection was such that the statistical properties (e.g., mean and standard deviation) of the training and testing subsets were similar to each other. Out of the 301 data, 226 data were used as the training data and 75 sets were taken for the testing purpose. Although normalization is not strictly necessary in the GP-based analysis, better results are often reached after normalizing the variables. Further, normalization speeds up the learning process. These are mainly due to influence of unification of the variables, no matter their range of variation (Alavi et al., 2010b). Thus, the input and output variables were normalized between 0 and 1. Selection of the optimal method for normalizing the data was on the basis of both controlling several normalization methods (Swingler, 1996) and the simplicity of the method. The ranges, normalized forms, and

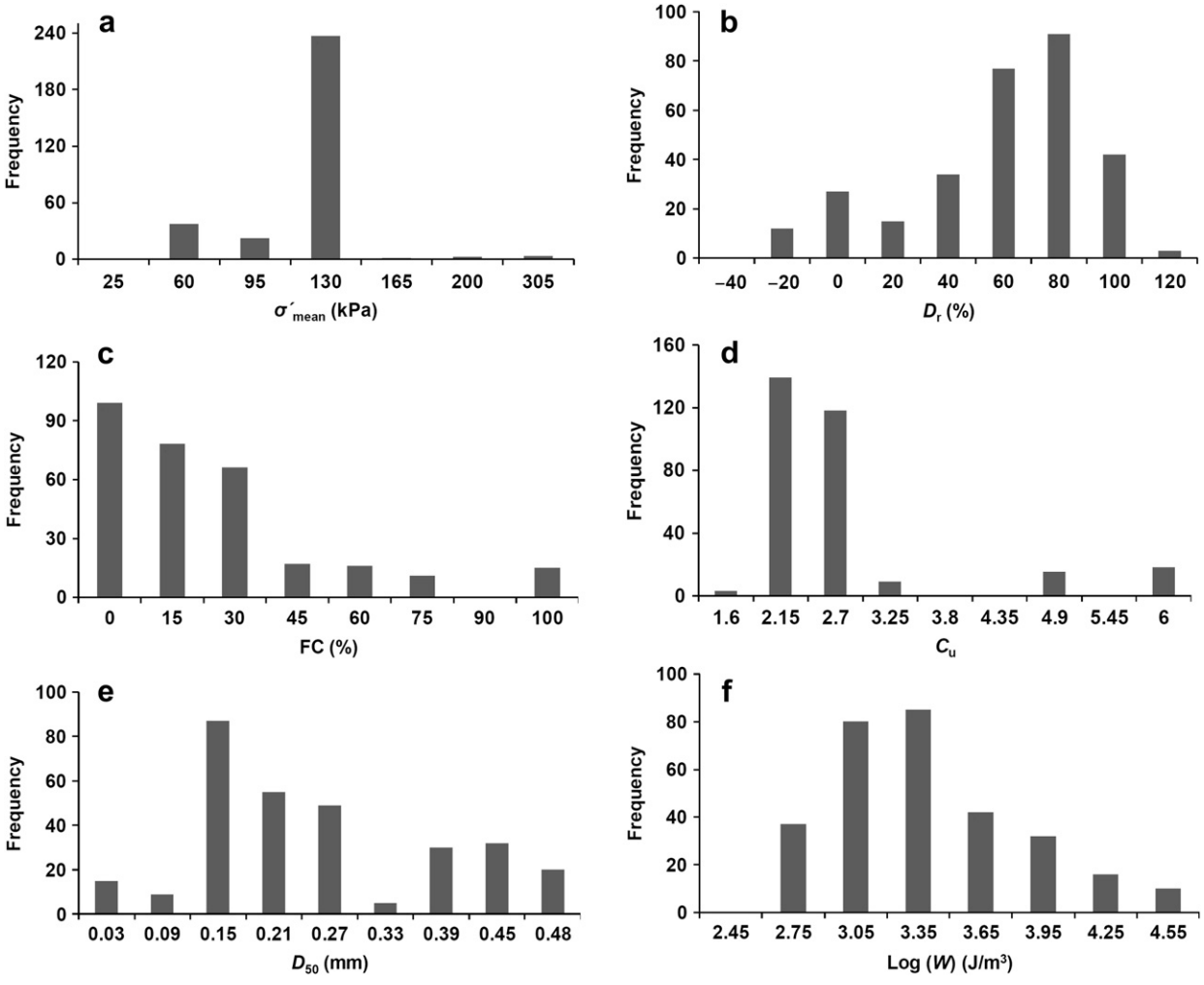

Figure 5 Histograms of the input and output variables.

statistics of different variables involved in the model development are given in Table 2.

4.2. Model development and analysis using LGP

The available database was used for the training and testing of the LGP prediction correlations. Using two different sets of the input parameters, two LGP-based formulas were obtained. Various parameters are involved in the LGP algorithm. The parameter selection affects the model generalization capability of LGP. The number of programs in the population that LGP will evolve is set by the population size. A run will take longer with a larger population size. The maximum number of tournaments sets the outer limit of the tournaments that will occur before the program terminates the run. The proper number of population and tournaments depends on the number of possible solutions and complexity of the problem. Mutation and crossover rates are the probabilities that an offspring will be subjected to the mutation and crossover operations, respectively (Koza, 1992; Gandomi et al., 2010b). The lengths of the evolved programs in runs can be controlled by initial and maximum program size parameters. The initial program size parameter sets the size of the programs in the first population at the start of each run. The maximum program size parameter sets the maximum length of the other programs evolved during each run (Brameier and Banzhaf, 2007). Several runs were conducted to come up with a parameterization of LGP that provided enough robustness and generalization to solve the problem. The LGP parameters were changed for different runs. The parameters were selected on the basis of both previously suggested values (Francone, 2001; Mukkamala et al., 2004; Baykasoglu et al., 2008; Gandomi et al., 2010a,b) and making several preliminary runs and observing the performance behavior. Three optimal levels were set for the population size (10,000, 15,000, 25,000) and two levels were considered for the crossover rate (50%, 90%). The mutation rate was set to 90%. Although most GP systems use a low mutation rate, numerical experiments showed that considering high mutation rates improves the generalization capability of LGP (Banzhaf et al., 1996; Brameier and Banzhaf, 2001; Francone, 2001; Gandomi et al., 2010a,b). This might be due to the significant effect of exchanging a variable on the program flow during the mutation operation (Brameier and Banzhaf, 2001). The success of the LGP algorithm usually increases with increasing the initial and maximum program size parameters. In this case, the complexity of the evolved functions increases and the speed of the algorithm decreases. These parameters are measured in bytes. The initial program size was set to 80 bytes. Two optimal values (128, 256) were considered for the maximum program size as tradeoffs between the running time and the complexity of the evolved solutions. The number of demes is related to the way that the population of programs is divided. Note that demes are semi-isolated subpopulations that evolution proceeds faster in them in comparison to a single population of equal size (Brameier and Banzhaf, 2007). Herein, the number of demes was set to 20. In this study, four basic arithmetic operators ($+$, $-$, $\times$, $/$) and basic mathematical functions ($\sqrt{}$, sin, cos) were utilized to get the optimum LGP models. There are $3 \times 2 \times 2 = 12$ different combinations of the parameters. All of these combinations were tested and 10 replications for each combination were carried out. This makes 120 runs for each of the combinations of the predictor variables. Therefore, the overall number of runs was equal to 120 $\times$ 2 (number of the input combinations) $=$ 240. A fairly large number of tournaments (900,000) were tested on each run to find models with minimum error. To evaluate the fitness of the evolved programs, the average of the squared raw errors was used. For each case, the program was run until there was no longer significant improvement in the performance of the models or the runs terminated automatically. For the analysis, a computer software called Discipulus (Conrads et al., 2004) was used which works on the basis of the AIMGP platform. Discipulus is a fast LGP system written for the Wintel platform. It operates directly on machine code (Foster, 2001; Deschaine and Francone, 2002). Discipulus can be regarded as an efficient modeling tool for complex problems because its speed permits conducting many runs in realistic timeframes. This leads to deriving consistent, high-precision models with little customization. Furthermore, it is well-designed to prevent overfitting and to evolve robust solutions (Francone and Deschaine, 2004).

4.2.1. LGP-based capacity energy correlations

The LGP-based formulations of the strain energy density required for triggering liquefaction, W (J/m^3), are as given below:

$$\mathrm{Log}(W)_{\mathrm{LGP,\,I}} = \frac{5}{4}\Big(2\sigma'_{\mathrm{mean},\,n}D_{\mathrm{r},\,n} + D_{\mathrm{r},\,n}D_{50,\,n} + D_{\mathrm{r},\,n}D^2_{50,n} \times \Big(\sigma'_{\mathrm{mean},\,n} + D_{50,\,n} - \Big(3\sigma'_{\mathrm{mean},\,n} - 6\mathrm{FC}_n + 4C_{\mathrm{u},\,n}\Big)^2 - 1\Big) + 2\Big) \quad (7)$$

$$\mathrm{Log}(W)_{\mathrm{LGP,\,II}} = \frac{5}{2} + 5\sigma'_{\mathrm{mean},\,n}D_{\mathrm{r},\,n} - 5\Big(\sigma'_{\mathrm{mean},\,n}D_{\mathrm{r},\,n}\Big)^2 \quad (8)$$

where $\sigma'_{\mathrm{mean},\,n}$, $D_{\mathrm{r},n}$, FC_n, $C_{\mathrm{u},n}$ and $D_{50,n}$, respectively, denote the soil initial effective mean confining pressure, initial relative density after consolidation, percentage of fines content, coefficient of uniformity, and mean grain size in their normalized forms (see Table 2). Fig. 6 shows a comparison of the experimental versus predicted liquefaction capacity energy using the LGP correlations.

Table 2 The variables used in model development.

Parameters	Minimum	Maximum	Standard deviation	Skewness	Kurtosis	Mean	Normalized form
Inputs							
σ'_{mean} (kPa)	27.8	294	31.28	2.12	16.17	94.91	σ'_{mean}/300
D_{r} (%)	−44.5	105.1	32.56	−0.82	−0.03	49.29	(D_{r} + 40)/150
FC (%)	0	100	25.88	1.79	2.79	19.68	(FC + 40)/150
C_{u}	1.57	5.88	1.09	2.08	3.83	2.42	C_{u}/6
D_{50} (mm)	0.03	0.46	0.13	0.47	−1.01	0.23	D_{50}/0.5
Output							
Log (W) (J/m^3)	2.48	4.54	0.45	0.71	−0.07	3.25	Log (W)/5

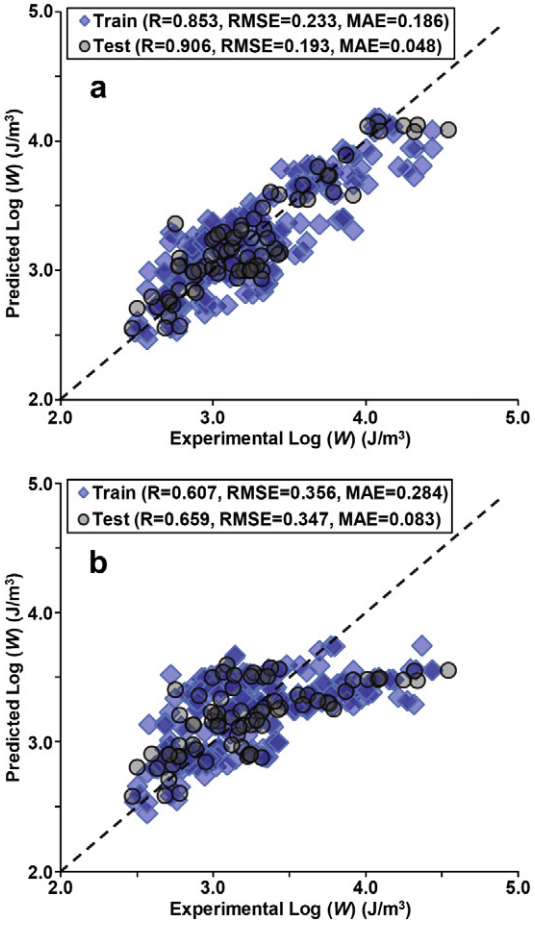

Figure 6 Experimental versus predicted liquefaction capacity energy using the LGP correlations. (a): Eq. (7); (b): Eq. (8).

4.3. Model development and analysis using MEP

Two separate MEP prediction equations were obtained for the liquefaction capacity energy. The parameter selection will affect the model generalization capability of MEP. The number of generation sets the number of levels the MEP algorithm uses before the run terminates. The number of expressions encoded by each MEP chromosome is equal to the chromosome length. This parameter directly influences the size of the search space and the number of solutions explored within the search space. Similar to LGP, several runs were conducted to find efficient parameters. The MEP parameters were changed for different runs. The parameters were chosen based on both some suggested values (Oltean, 2004; Grosan and Abraham, 2006; Baykasoglu et al., 2008; Alavi et al., 2010a) and after a trial and error approach. Three optimal levels were set for the population size (250, 500, 1000) and two levels were taken for the crossover rate (50%, 90%). The mutation rate was set to 10%. The success of the MEP algorithm usually increases with increasing the chromosome length (Oltean and Dumitrescu, 2002; Oltean and Grosşan, 2003). Two optimal values equal to 20 and 50 genes were selected for the chromosome length. Basic arithmetic operators (+, −, ×, /) and mathematical functions (exp, sin, cos) were utilized to get the optimum MEP models. Napierian logarithm function was further considered in this case. Since the best obtained formula considering this function form was not precise, it was not presented herein. All of these combinations were tested and 10 replications for each combination were carried out. The overall number of runs was equal to $3 \times 2 \times 2 \times 10 \times 2$ (number of the input combinations) $= 240$. A fairly large number of generations were tested on each run to find models with minimum error. The program was run until the runs automatically terminated. The fitness the programs evolved by MEP were calculated using Eq. (1). For the analysis, source code of MEP (Oltean, 2004) in C++ was utilized.

4.3.1. MEP-based capacity energy correlations

The MEP-based prediction equations for the capacity energy, W (J/m^3), are as given below:

$$\mathrm{Log}(W)_{\mathrm{MEP,\,I}} = \frac{5}{2} + 5D_{\mathrm{r},\,n}\left(\frac{\sigma'_{\mathrm{mean},\,n}}{2} - \frac{D_{50,n}}{2}\left(\frac{\mathrm{FC}_n}{2} + \frac{D_{50,n}}{2} \times \left(2\sigma'_{\mathrm{mean},\,n} - D_{50,n}\left(\frac{\frac{\sigma'^2_{\mathrm{mean},\,n}}{2} + 2\mathrm{FC}_n - 2C_{\mathrm{u},\,n}}{\mathrm{FC}_n} + 1\right) + \frac{\sigma'^2_{\mathrm{mean},\,n}}{4}\right) - \frac{1}{2}\right)\right) \quad (9)$$

$$\mathrm{Log}(W)_{\mathrm{MEP,\,II}} = \frac{5}{2} + 5\sigma'_{\mathrm{mean},\,n}D_{\mathrm{r},n}\left(1 - \frac{\sigma'_{\mathrm{mean},\,n}}{2}\right) \quad (10)$$

and $\sigma'_{\mathrm{mean},\,n}$, $D_{\mathrm{r},n}$, FC_n, $C_{\mathrm{u},n}$ and $D_{50,n}$ are the predictor variables in their normalized forms as shown in Table 2. Fig. 7 presents a comparison of the experimental versus predicted liquefaction capacity energy using the MEP-based equations.

4.4. Model development using traditional GP

A traditional tree-based GP analysis was performed to compare the linear variant of GP, i.e., LGP and MEP, with a classical GP approach. After developing and controlling several models with different combinations of the input parameters, the best GP model was selected and presented as the optimal model. Similar to the LGP and MEP-based analyses, the input and output variables were normalized between 0 and 1. Several runs were conducted considering different values for the GP parameters. A large number of generations were tested to find a model with minimum error. Different levels were selected for the population size within the range of 200–800. From experimental trials, the rates of crossover and mutation were set to optimal values equal to 90% and 10%, respectively. Linear error function was adopted as the fitness function. The maximum tree depth directly influences the size of the search space and the number of solutions explored within the search space. An optimal value equal to 8 was considered for this parameter. GPLAB (Silva 2007), in conjunction with subroutines coded in MATLAB, was used to implement the tree-based GP algorithm. The traditional GP-based formulation of W in terms of σ'_{mean}, D_{r}, FC, C_{u} and D_{50} is as given below:

$$\mathrm{Log}(W) = \frac{20}{\left(\left(\left(7 - \left(\sigma'_{\mathrm{mean},\,n} + D_{\mathrm{r},\,n} + D^2_{50,\,n}\right)\right) - \left(\sigma'_{\mathrm{mean},\,n}(D_{\mathrm{r},\,n} - \mathrm{FC}_n)\right)\right) + C_{\mathrm{u},\,n}\right)} \quad (11)$$

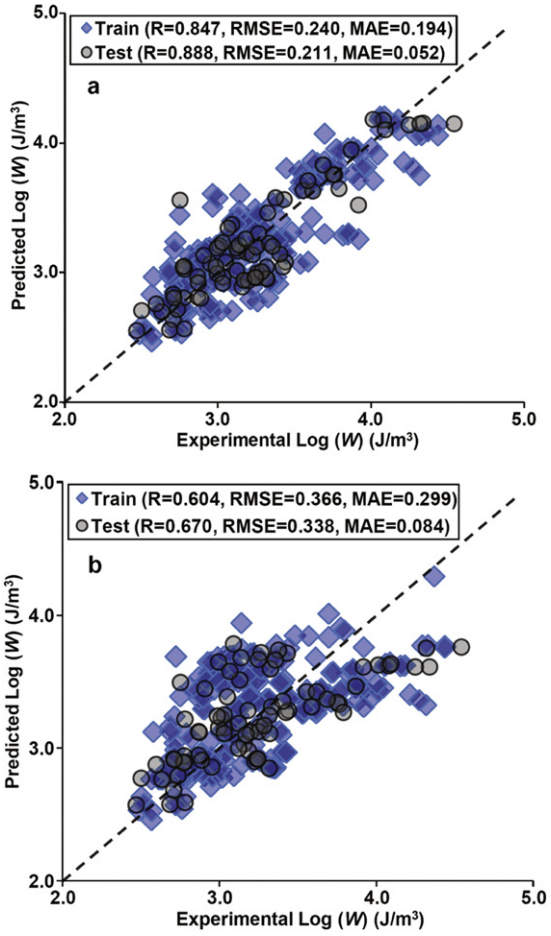

Figure 7 Experimental versus predicted liquefaction capacity energy using the MEP correlations. (a): Eq. (9); (b): Eq. (10).

Fig. 8 illustrates the experimental against predicted capacity energy using the GP model.

5. Comparison of the energy-based numerical correlations

Different equations were obtained for the assessment of the liquefaction resistance of sand–silt mixtures. Performance statistics of the models obtained by LGP, MEP, standard GP, and the conventional MLR-based equations for the entire data (301 data sets) are summarized in Table 3. Fig. 9 visualizes a comparison of the predictions made by different models. Since the other existing energy-based pore pressure build-up models need calibration parameters, it was not possible to evaluate their performance on the available database. Comparing the performance of the proposed relationships, it can be seen from Figs. 6 and 7, and Table 3 that Eq. (7) created by LGP has produced better results than Eq. (9) evolved by MEP on the training, testing, and entire data. With the exception of the testing data, the same results are obtained on the training and whole of data by Eq. (8) of LGP compared with Eq. (10) generated by MEP. The results demonstrate that the LGP and MEP-based formulas with five inputs significantly outperform those using two inputs. Also, the best LGP and MEP models (Eqs. (7) and (9)) have produced better results than the best GP model. As shown in Table 3, the LGP and

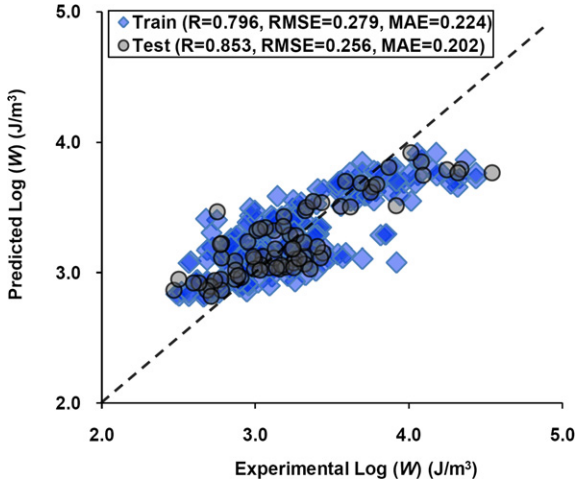

Figure 8 Experimental versus predicted liquefaction capacity energy using the GP correlation.

MEP-based formulas provide considerably better results than the regression models proposed by Liang (1995), Dief and Figueroa (2001) and Figueroa et al. (1994). Although most of the existing linear regression-based models yield accurate results for their relevant database, they cannot successfully work for the current database. This is due to nonlinearity in the liquefaction development. Furthermore, as shown in Table 1, the test results used for the calibration of the existing regression models are less than those considered for the development of the proposed models. Major differences among the correlations using two and five of soil initial parameters imply the necessity of using five predictor variables (σ'_{mean}, D_r, FC, C_u, and D_{50}) for the performed analyses.

In order to control the external validation of the best LGP and MEP models, a new criterion was checked on the testing data sets. Smith (1986) stated that if correlation coefficient (R) value provided by a model is higher than 0.8 and the error values (e.g., RMSE and MAE) are low, the predicted and measured values are strongly correlated with each other. Golbraikh and Tropsha (2002) suggested that at least one slope of regression lines (k or k')

Table 3 Overall performance of the energy-based correlations for the liquefaction assessment.

Model	Performance		
	R	RMSE	MAE
Model inputs: σ'_{mean}, D_r			
Figueroa et al. (1994)	0.57	0.453	0.373
Liang (1995)	0.55	0.458	0.379
Liang (1995)	0.36	0.509	0.375
Dief and Figueroa (2001)	0.58	0.62	0.475
Dief and Figueroa (2001)	0.43	0.51	0.374
LGP, Eq. (8)	0.62	0.354	0.283
MEP, Eq. (10)	0.62	0.359	0.294
Model inputs: σ'_{mean}, D_r, FC, C_u, D_{50}			
Baziar and Jafarian (2007)	0.81	0.262	0.208
LGP, Eq. (7)	0.87	0.224	0.178
MEP, Eq. (9)	0.86	0.233	0.187
GP, Eq. (11)	0.81	0.274	0.219

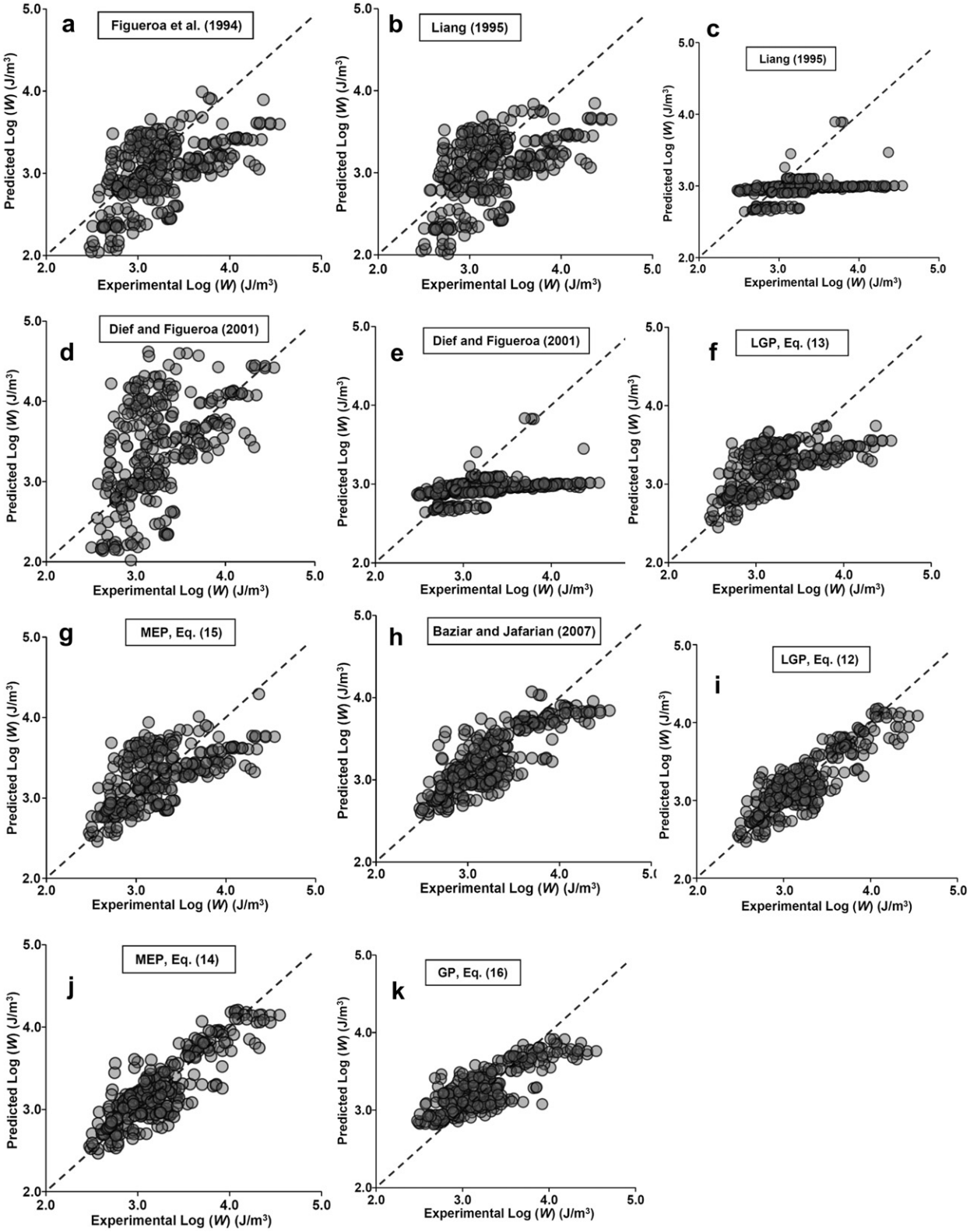

Figure 9 A comparison between the experimental and predicted capacity energy values using different models.

through the origin should be close to 1. Furthermore, the squared correlation coefficient between the predicted and measured values (Ro^2), and the correlation coefficient between the measured and predicted values (Ro'^2) should be close to 1 (Roy and Roy, 2008; Alavi et al., 2011). The considered validation criteria and the relevant results obtained by the models are presented in Table 4. As it is seen, the derived models fully satisfy the required conditions.

Table 4 Statistical parameters of the best LGP and MEP correlations for the external validation.

Item	Formula	Condition	LGP, Eq. (7)	MEP, Eq. (9)
1	R	$0.8 < R$	0.906	0.888
2	$k=\frac{\sum_{i=1}^{n}(h_i \times t_i)}{h_i^2}$	$0.85 < k < 1.15$	1.004	1.006
3	$k'=\frac{\sum_{i=1}^{n}(h_i \times t_i)}{t_i^2}$	$0.85 < k' < 1.15$	0.993	0.99
4	$R_0^2=1-\frac{\sum_{i=1}^{n}(t_i-h_i^0)^2}{\sum_{i=1}^{n}(t_i-\bar{t}_i)^2}$ $h_i^0=k \times t_i$	Should be close to 1	0.999	0.998
5	$R_0'^2=1-\frac{\sum_{i=1}^{n}(h_i-t_i^0)^2}{\sum_{i=1}^{n}(h_i-\bar{h}_i)^2}$ $t_i^0=k' \times h_i$	Should be close to 1	0.996	0.994

h_i: actual output value for the i^{th} output; t_i: predicted output value for the i^{th} output; n: number of samples.

One of the significant advantages of LGP and MEP is that they directly learn from experimental data. Thus, these methods are suitable for extracting the functional relationships for the cases where the underlying relationships are unknown or the physical meaning is difficult to be explained. Contrary to these methods, conventional methods (e.g., regression and finite element method) assume the structure of the model in advance, which may be suboptimal (Alavi et al., 2011). The best solutions obtained by means of LGP and MEP are determined after controlling millions of linear and nonlinear models. That is why the derived models can proficiently take into account the interactions between the dependent and independent variables. However, for more reliability, the results of the LGP and MEP-based analyses are suggested to be treated as a complement to conventional computing techniques. In any case, the important role of engineering judgment in interpretation of the results obtained should seriously be taken into consideration (Cabalar and Cevik, 2009; Alavi et al., 2011). It is worth mentioning that the LGP and MEP algorithms are parameter sensitive. Their performance could be improved by using any form of optimally controlling the parameters of the run (e.g., GAs) (Dimopoulos and Zalzala, 2001).

6. Sensitivity analysis

A sensitivity analysis was conducted to determine the contributions of the variables to the prediction of the strain energy. To perform the sensitivity analysis, frequency values (Francone, 2001) of the input parameters were obtained. A frequency value equal to 100% for an input indicates that this variable has been appeared in 100% of the best thirty programs evolved by LGP and MEP. This is a common approach in the GP-based analyses (Francone, 2001; Alavi et al., 2010a; Gandomi et al., 2010a). Baziar and Jafarian (2007) categorized σ'_{mean} and D_r into one group referred to as *Intergranular Contact Density*. They considered FC as a single category controlling the potential of the pore pressure build-up. C_u and D_{50} were classified as *Grain Size Characteristics* or *Textural Properties*. A similar categorization to that defined by Baziar and Jafarian (2007) was considered in this study.

The frequency values of the input parameters are presented in Fig. 10. According to Fig. 10a, the capacity energy is more sensitive to D_r and σ'_{mean} than the other inputs. It can also be observed from Fig. 10b that the capacity energy is more dependent on D_r in comparison with σ'_{mean}. As it is seen, the results obtained by the LGP and MEP formulations are in agreement with each other. It is notable that σ'_{mean} and D_r are the most widely-used parameters directly incorporated in the majority of the previous published models.

7. Parametric analysis

For further verification of the LGP and MEP-based correlations, a parametric analysis was performed in this study. The main goal was to find the effect of each parameter on the capacity energy (W). The parametric analysis investigates the response of the predicted W from the models to a set of hypothetical input data

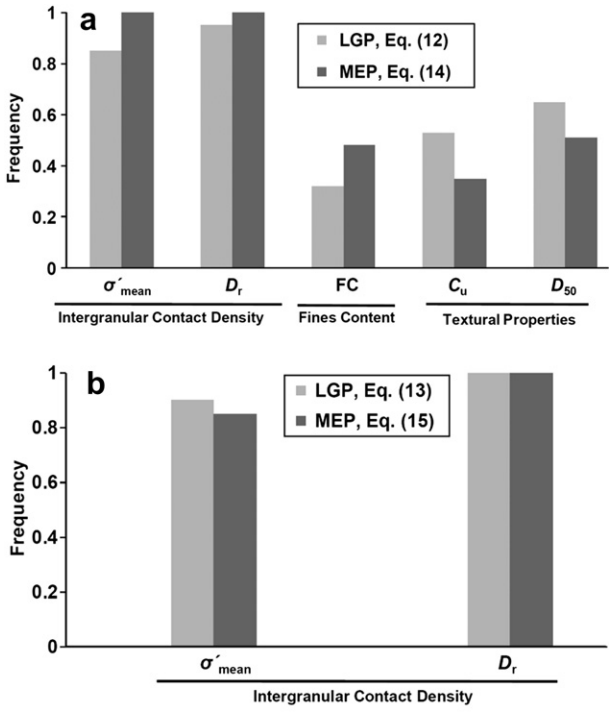

Figure 10 Contributions of the predictor variables in the LGP and MEP models.

generated over the training ranges of the minimum and maximum data. For this aim, one predictor variable was changed at a time while the other seismic variables were kept constant at the average values of their entire data sets. A set of synthetic data for the single varied parameter was generated by increasing the value of this in increments (Alavi et al., 2011). These variables were presented to the prediction models and W was calculated. This procedure was repeated using another variable until the responses of the models were tested for all of the predictor variables (Alavi et al., 2011). Fig. 11 presents the tendency of the W predictions to the variations of σ'_{mean}, D_r, FC, C_u, and D_{50}.

The results of the parametric analysis for Eqs. (7)–(10) indicate that the capacity energy of sands continuously increases due to increasing σ'_{mean}, D_r and D_{50}, and decreases with increasing C_u. These results are in close agreement with the results of the laboratory studies carried out by other researchers (Lee and Seed, 1976; Liang, 1995; Polito and Martin, 2001).

The susceptibility of sands deposits with silt content to liquefaction is higher than clean sands (Baziar and Jafarian, 2007; Polito and Martin, 2001). However, there is not a general agreement about the effect of silt content on the liquefaction resistance of sands (Baziar and Jafarian, 2007). Naeini and Baziar (2004) and Xenaki and Athanasopoulos (2003) showed that the liquefaction resistance of sand–silt mixtures decreases when non-plastic FC increases up to 35% and 44%, and afterward the resistance starts increasing. Polito and Martin (2001) performed a laboratory parametric study utilizing cyclic triaxial tests to clarify the effects of non-plastic fines on the liquefaction potential of sands. Fig. 12 shows a plot of cyclic resistance versus silt content for specimens of Yatesville sand and silt presented by Polito and Martin (2001). The marked drop in the cyclic resistance occurs as the silt content exceeds the limiting silt content (about 35%). The largest amount of silt that can be accommodated in the voids created by the sand skeleton is called the limiting silt content and occurs between 25% and 45% for most sands (Polito and Martin, 2001). In the present study, the results of the parametric analysis for FC indicate that the energy-based liquefaction resistance of sand–silt deposits increases when FC increases up to about 30% and thereafter it

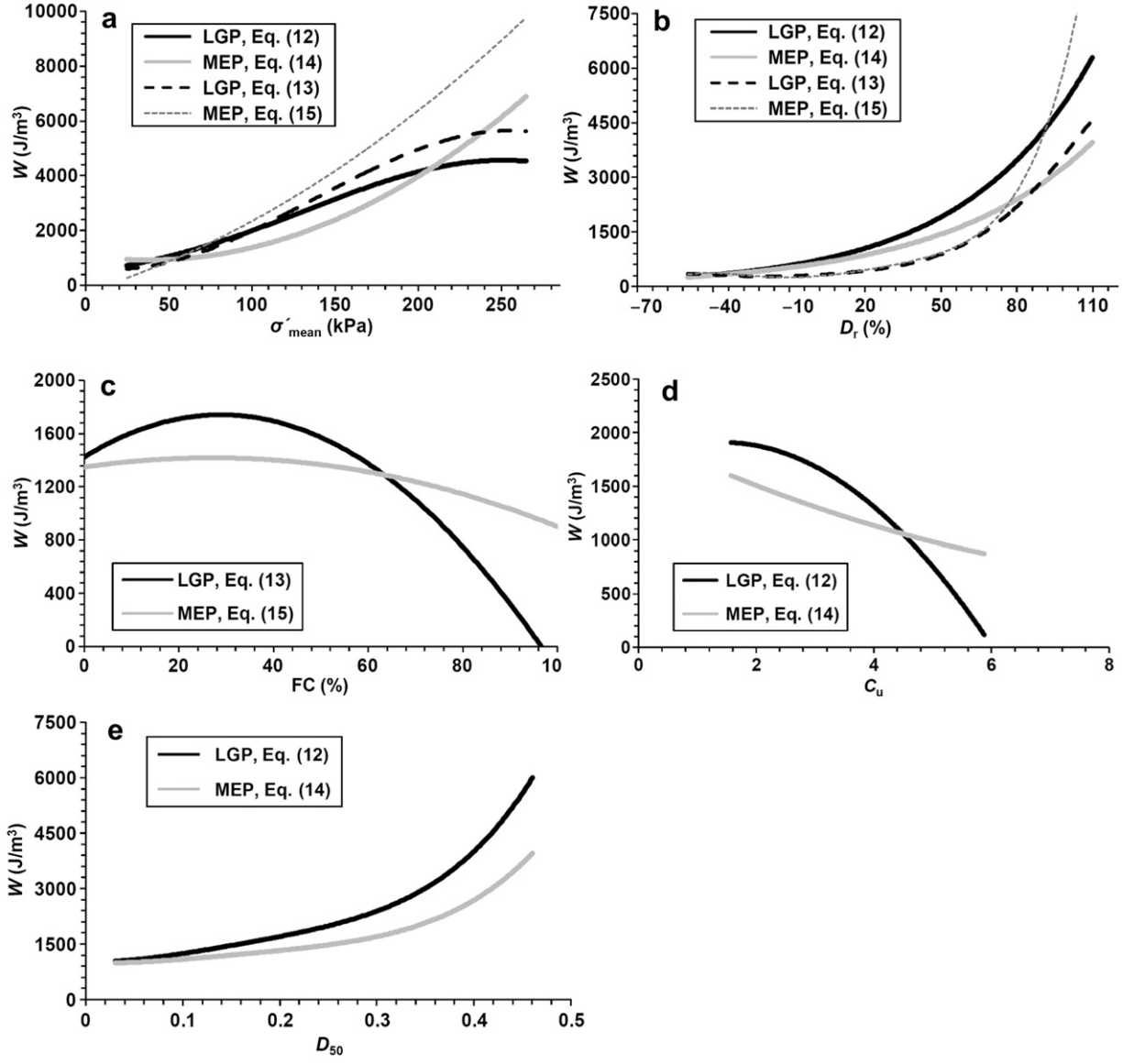

Figure 11 Parametric analysis of the capacity energy in the LGP and MEP-based correlations.

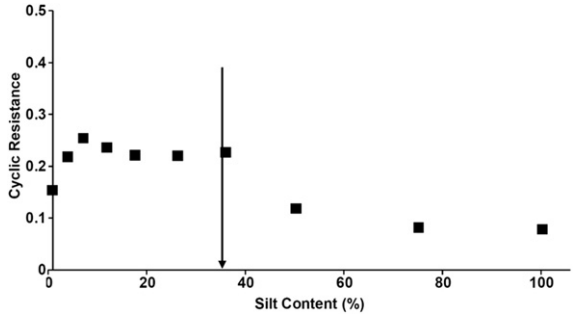

Figure 12 Variations of the cyclic resistance with silt content for Yatesville silty sand specimens (Polito and Martin, 2001).

starts decreasing (see Fig. 11c). A comparison between Figs. 11c and 12 reveals that the trends obtained by the proposed correlations, specifically the LGP model, are soundly similar to those reported by Polito and Martin (2001).

8. Conclusions

In the present study, new empirical correlations were derived to estimate the amount of the strain energy required up to the liquefaction triggering using LGP and MEP. Two different combinations of the influencing variables were considered for the development of the LGP and MEP-based correlations. The first combination of the input parameters consisted of σ'_{mean}, D_r, FC, C_u, and D_{50}. The second combination was comprised the most widely-used parameters in the energy-based pore pressure build-up models for the liquefaction assessment, i.e., σ'_{mean} and D_r. A traditional GP analysis was performed to benchmark the LGP and MEP correlations. Major findings obtained in this research are as follows.

i. The LGP and MEP-based correlations give good estimations of the capacity energy of sandy soils. On average, the LGP and MEP formulas developed upon the same sets of the predictor variables reach a similar prediction performance. The validity of the models was checked for a part of the database beyond the training data domain. The validation phases confirm the efficiency of the models for their general application to the capacity energy estimation.
ii. The best LGP and MEP models perform superior than the optimal traditional GP model. Due to the high nonlinearity in the liquefaction development, the proposed nonlinear correlations produce considerably better outcomes over the existing linear regression-based models.
iii. The correlations that were developed using σ'_{mean}, D_r, FC, C_u, and D_{50} remarkably outperform those using σ'_{mean} and D_r. As the other researchers have mentioned, the sensitivity analysis results indicate that D_r and σ'_{mean} are much more effective to explain the variations of the capacity energy than other soil initial parameters.
iv. The results of the parametric analysis were confirmed with the results of the experimental studies presented by other researchers. The results indicate that the developed correlations are robust and efficaciously incorporate the underlying physical relations governing the liquefaction behavior.
v .The LGP and MEP approaches have a great ability to specify the structure of the model using only the experimental data. The models derived using these techniques are suggested to be used for pre-design purposes. Furthermore, they may be used as a quick check on solutions developed by more time consuming and in-depth deterministic analyses.
vi. It has been shown that the prediction accuracy of the machine learning-based models using a single training and testing set can vary significantly (Oommen and Baise, 2010). Applying a K-fold cross validation method to the performance evaluation can be an efficient approach to cope with this issue.

References

Aho, A., Sethi, R., Ullman, J., 1986. Compilers: Principles, Techniques, and Tools. Addison-Wesley, Reading, MA.

Al-Anazi, A., Babadagli, T., 2010. Automatic fracture density update using smart well data and artificial neural network. Computers & Geosciences 36, 335–347.

Alavi, A.H., Ameri, M., Gandomi, A.H., Mirzahosseini, M.R., 2011. Formulation of flow number of asphalt mixes using a hybrid computational method. Construction and Building Materials 25, 1338–1355.

Alavi, A.H., Gandomi, A.H., 2011. A robust data mining approach for formulation of geotechnical engineering systems. Engineering Computations 28 (3), 242–274.

Alavi, A.H., Gandomi, A.H., Mousavi, M., Mollahasani, A., 2010b. High-precision modeling of uplift capacity of suction caissons using a hybrid computational method. Geomechanics and Engineering 2 (4), 253–280.

Alavi, A.H., Gandomi, A.H., Sahab, M.G., Gandomi, M., 2010a. Multi expression programming: a new approach to formulation of soil classification. Engineering with Computers 26 (2), 111–118.

Arulmoli, K., Muraleetharan, K.K., Hosain, M.M., Fruth, L.S., 1992. VELACS Laboratory Testing Program – Soil Data Report. The Earth Technology Corporation, Irvine, Calif. Report to the National Science Foundation, Washington, DC. <http://gees.usc.edu/velacs>

Banzhaf, M., Francone, F.D., Nordin, P., 1996. The effect of extensive use of the mutation operator on generalization in genetic programming using sparse data sets. In: Proceedings of the International Conference on Evolutionary Computation, volume 1141 of LNCS. Berlin, Germany, pp. 300–309.

Banzhaf, W., Nordin, P., Keller, R., Francone, F., 1998. Genetic Programming – An Introduction. On the Automatic Evolution of Computer Programs and its Application. dpunkt/Morgan Kaufmann, Heidelberg/San Francisco.

Baykasoglu, A., Gullub, H., Canakcı, H., Ozbakır, L., 2008. Prediction of compressive and tensile strength of limestone via genetic programming. Expert Systems with Applications 35 (1–2), 111–123.

Baziar, M.H., Jafarian, Y., 2007. Assessment of liquefaction triggering using strain energy concept and ANN model: capacity energy. Soil Dynamics and Earthquake Engineering 27 (12), 1056–1072.

Baziar, M.H., Jafarian, Y., Shahnazari, H., Movahed, V., Tutunchian, M.A., 2011. Prediction of strain energy-based liquefaction resistance of sand–silt mixtures: an evolutionary approach. Computers & Geosciences 37 (11), 1883–1893.

Brameier, M., Banzhaf, W., 2001. A comparison of linear genetic programming and neural networks in medical data mining. IEEE Transactions on Evolutionary Computation 5 (1), 17–26.

Brameier, M., Banzhaf, W., 2007. Linear Genetic Programming. Springer Science + Business Media, LLC, New York.

Cabalar, A.F., Cevik, A., 2009. Genetic programming-based attenuation relationship: an application of recent earthquakes in turkey. Computers & Geosciences 35, 1884–1896.

Cetin, K.O., Seed, R.B., Der Kiureghian, A., Tokimatsu, K., Harder, L.F., Kayen, R.E., Moss, R.E.S., 2004. Standard penetration test-based probabilistic and deterministic assessment of seismic soil liquefaction potential. Journal of Geotechnical and Geoenvironmental Engineering 130 (12), 1314–1340.

Chen, Y.R., Hsieh, S.C., Chen, J.W., Shih, C.C., 2005. Energy-based probabilistic evaluation of soil liquefaction. Soil Dynamics and Earthquake Engineering 25 (1), 55–68.

Conrads, M., Dolezal, O., Francone, F.D., Nordin, P., 2004. Discipulus LiteTM-fast Genetic Programming Based on AIM Learning Technology. Register Machine Learning Technologies Inc., Littleton, CO.

Darve, F., 1996. Liquefaction phenomenon of granular materials and constitutive instability. International Journal of Computer Aided Methods in Engineering – Engineering Computations 13 (7), 5–28.

Desai, C.S., Siriwardane, H.J., 1984. Constitutive Laws for Engineering Materials: With Emphasis on Geologic Materials. Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Deschaine, L.M., Francone, F.D., 2002. Comparison of Discipulus™ Linear Genetic Programming Software with Support Vector Machines, Classification Trees, Neural Networks and Human Experts. Register Machine Learning Technologies Inc.. White Paper. Available online at: www.rmltech.com/Comparison.White.Paper.pdf (accessed 25.01.09).

Dief, H.M., 2000. Evaluating the liquefaction potential of soils by the energy method in the centrifuge. PhD Dissertation, Department of Civil Engineering, Case Western Reserve University, Cleveland, OH.

Dief, H.M., Figueroa, J.L., 2001. Liquefaction assessment by the energy method through centrifuge modeling. In: Zeng, X.W. (Ed.), Proceedings of the NSF International Workshop on Earthquake Simulation in Geotechnical Engineering. CWRU, Cleveland, OH.

Dimopoulos, C., Zalzala, A.M.S., 2001. Investigating the use of genetic programming for a classic one-machine scheduling problem. Advances in Engineering Software 32 (6), 489–498.

Dobry, R., Ladd, R.S., Yokel, F.Y., Chung, R.M., Powell, D., 1982. Prediction of Pore Water Pressure Build-up and Liquefaction of Sands during Earthquakes by the Cyclic Strain Method. National Bureau of Standards, US Department of Commerce, US Governmental Printing Office, Building Science Series, Washington, DC, p. 138.

Figueroa, J.L., Saada, A.S., Liang, L., Dahisaria, M.N., 1994. Evaluation of soil liquefaction by energy principles. Journal of Geotechnical and Geoenvironmental Engineering 20 (9), 1554–1569.

Figueroa, J.L., Saada, A.S., Rokoff, M.D., Liang, L., 1998. Influence of grain-size characteristics in determining the liquefaction potential of a soil deposit by the energy method. In: Proceedings of the International Workshop on the Physics and Mechanics of Soil Liquefaction. Baltimore, Maryland, USA, pp. 237–245.

Foster, A., 2001. Review: Discipulus: a commercial genetic programming system. Genetic Programming and Evolvable Machines 2 (2), 201–203.

Francone, F., 2001. Discipulus Pro™ Owner's Manual. Register Machine Learning Technologies. Littleton, Colorado, USA.

Francone, F.D., Deschaine, L.M., 2004. Extending the boundaries of design optimization by integrating fast optimization techniques with machine-code-based, linear genetic programming. Information Sciences 161, 99–120.

Gandomi, A.H., Alavi, A.H., 2011. Multi-stage genetic programming: a new strategy to nonlinear system modeling. Information Sciences 181 (23), 5227–5239.

Gandomi, A.H., Alavi, A.H., Sahab, M.G., Arjmandi, P., 2010a. Formulation of elastic modulus of concrete using linear genetic programming. Journal of Mechanical Science and Technology 24 (6), 1011–1017.

Gandomi, A.H., Alavi, A.H., Sahab, M.G., 2010b. New formulation for compressive strength of CFRP confined concrete cylinders using linear genetic programming. Materials and Structures 43 (7), 963–983.

Ghorbani, M.A., Khatibi, R., Aytek, A., Makarynskyy, O., Shiri, J., 2010. Sea water level forecasting using genetic programming and comparing the performance with Artificial Neural Networks. Computers & Geosciences 36 (5), 620–627.

Goh, A.T.C., 1994. Seismic liquefaction potential assessed by neural networks. Journal of Geotechnical Engineering 120 (9), 1467–1480.

Goh, A.T.C., 2002. Probabilistic neural network for evaluating seismic liquefaction potential. Canadian Geotechnical Journal 39 (1), 219–232.

Goh, A.T.C., Goh, S.H., 2007. Support vector machines: their use in geotechnical engineering as illustrated using seismic liquefaction data. Computers and Geotechnics 34, 410–421.

Golbraikh, A., Tropsha, A., 2002. Beware of q^2. Journal of Molecular Graphics and Modelling 20 (4), 269–276.

Green, R.A., 2001. Energy-based evaluation and remediation of liquefiable soils. PhD dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Green, R.A., Mitchell, J.K., Polito, C.P., 2000. Energy-based excess pore pressure generation model for cohesionless soils. In: Proc. John Booker Memorial Symposium. Sydney, NSW, Australia, pp. 16–17.

Grosan, C., Abraham, A., 2006. Stock Market Modeling Using Genetic Programming Ensembles, Genetic Systems Programming. Studies in Fuzziness and Soft Computing. Springer Verlag, Germany, pp. 131–146.

Koza, J., 1992. Genetic Programming, on the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge (MA).

Lee, K.L., Seed, H.B., 1976. Cyclic stress conditions causing liquefaction of sand. Journal of the Soil Mechanics and Foundations Division 93, 47–70 (SMl).

Li, X.S., Dafalias, Y.F., 2002. Constitutive modeling of inherently anisotropic sand behavior. Journal of Geotechnical and Geoenvironmental Engineering 128 (10), 868–880.

Liang, L., Figueroa, J.L., Saada, A.S., 1995. Liquefaction under random loading: a unit energy approach. Journal of Geotechnical and Geoenvironmental Engineering 121, 776–781 (GT11).

Liang, L., 1995. Development of an energy method for evaluating the liquefaction potential of a soil deposit. PhD dissertation, Department of Civil Engineering, Case Western Reserve University, Cleveland, OH.

Mukkamala, S., Sung, A.H., Abraham, A., 2004. Modelling intrusion detection systems using linear genetic programming approach. IEA/AIE 2004, LNAI 3029, 633–642.

Naeini, S.A., Baziar, M.H., 2004. Effect of fines content on steady-state strength of mixed and layered samples of a sand. Soil Dynamics and Earthquake Engineering 24, 181–187.

Nemat-Nasser, S., Shokooh, A., 1979. A unified approach to densification and liquefaction of cohesionless sand in cyclic shearing. Canadian Geotechnical Journal 16 (4), 659–678.

Nordin, P.J., 1994. A compiling genetic programming system that directly manipulates the machine code (Chapter 14). In: Kenneth, E., Kinnear, J.R. (Eds.), Proceedings of the International Conference on Advances in Genetic Programming. MIT Press, USA, pp. 311–331.

Oltean, M., Dumitrescu, D., 2002. Multi Expression Programming. Technical Report, UBB-01-2002. Babeş-Bolyai University, Cluj-Napoca, Romania.

Oltean, M., Grosşan, C., 2003. A comparison of several linear genetic programming techniques. Advances in Complex Systems 14 (4), 1–29.

Oltean, M., 2004. Multi expression programming source code. Available at: http://www.mep.cs.ubbcluj.ro/.

Oommen, T., Baise, L.G., 2010. Model development and validation for intelligent data collection for lateral spread displacements. Journal of Computing in Civil Engineering 24 (6), 467–477.

Oommen, T., Baise, L.G., Vogel, R.M., 2010. Validation and application of empirical liquefaction models. Journal of Geotechnical and Geoenvironmental Engineering 136 (12), 1618–1633.

Oommen, T., Misra, D., Twarakavi, N.K.C., Prakash, A., Sahoo, B., Bandopadhyay, S., 2008. An objective analysis of support vector machine based classification for remote sensing. Mathematical Geoscience 40 (4), 409–424.

Ostadan, F., Deng, N., Arango, I., 1996. Energy-based Method for Liquefaction Potential Evaluation, Phase I. Feasibility Study. U.S. Department of Commerce, Technology Administration, National Institute of Standards and Technology, Building and Fire Research Laboratory.

Pal, M., 2006. Support vector machines-based modelling of seismic liquefaction potential. International Journal for Numerical and Analytical Methods in Geomechanics 30 (10), 983–996.

Poli, R., Langdon, W.B., McPhee, N.F., Koza, J.R., 2007. Genetic programming: an introductory tutorial and a survey of techniques and applications. Technical Report CES-475, ISSN: 1744–8050.

Polito, C.P., Martin, J.R., 2001. Effects of non-plastic fines on the liquefaction resistance of sands. Journal of Geotechnical and Geoenvironmental Engineering 127 (5), 408–415.

Refaeilzadeh, P., Tang, L., Liu, H., 2009. Cross validation. In: Proceedings of Encyclopedia of Database Systems, pp. 532–538.

Rokoff, M.D., 1999. The influence of grain-size characteristics in determining the liquefaction potential of a soil deposit by the energy method. M.S. thesis, Department of Civil Engineering, Case Western Reserve University, Cleveland, OH.

Roy, P.P., Roy, K., 2008. On some aspects of variable selection for partial least squares regression models. QSAR & Combinatorial Science 27, 302–313.

Samui, P., 2007. Seismic liquefaction potential assessment by using Relevance Vector Machine. Earthquake Engineering and Engineering Vibration 6 (4), 331–336.

Seed, H.B., Idriss, I.M., 1971. Simplified procedure for evaluating soil liquefaction potential. Journal of the Soil Mechanics and Foundations Division 97, 1249–1274 (SM8).

Silva, S., 2007. GPLAB: a genetic programming toolbox for MATLAB. Available online at: <http://gplab.sourceforge.net>.

Smith, G.N., 1986. Probability and Statistics in Civil Engineering. Collins, London.

Swingler, K., 1996. Applying Neural Networks a Practical Guide. Academic Press, New York.

Torres, R.S., Falcão, A.X., Gonçalves, M.A., Papa, J.P., Zhang, B., Fan, W., 2009. A genetic programming framework for content-based image retrieval. Pattern Recognition 42 (2), 283–292.

Towhata, I., Ishihara, K., 1985. Shear work and pore water pressure in undrained shear. Soils and Foundations 25 (3), 73–84.

Voznesenskya, E.A., Nordal, S., 1999. Dynamic instability of clays: an energy approach. Soil Dynamics and Earthquake Engineering 18 (2), 125–133.

Wang, G., Takemura, J., Kuwano, J., 1997. Evaluation of excess pore water pressures of intermediate soils due to cyclic loading by energy method. In: Proceedings of International Conference on Computer Methods and Advances in Geomechanics. Rotterdam, Netherlands, pp. 2215–2220.

Xenaki, V.C., Athanasopoulos, G.A., 2003. Liquefaction resistance of sand–silt mixtures: an experimental investigation of the effect of fines. Journal of Soil Dynamics Earthquake Engineering 23, 183–194.

Yilmaz, S., Demircioglu, C., Akin, S., 2002. Application of artificial neural networks to optimum bit selection. Computers & Geosciences 28 (2), 261–269.

Youd, T.L., Idriss, I.M., Andrus, R.D., Arango, I., Castro, G., Christian, J.T., Dobry, R., Finn, W.D.L., Harder, L.F., Hynes, M.E., Ishihara, K., Koester, J.P., Liao, S.S.C., Marcuson, W.F., Martin, G.R., Mitchell, J.K., Moriwaki, Y., Power, M.S., Robertson, P.K., Seed, R.B., Stokoe, K.H., 2001. Liquefaction resistance of soils: summary report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of liquefaction resistance of soils. Journal of Geotechnical and Geoenvironmental Engineering 127 (4), 297–313.