https://dx.doi.org/10.1016/j.enggeo.2020.105758

High performance prediction of soil compaction parameters using multi expression programming

Han-Lin Wang[1] and Zhen-Yu Yin[1,*]

[1] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China.

[*] Corresponding author: Dr. Zhen-Yu YIN, Tel.: +852 34008470, Email: zhenyu.yin@polyu.edu.hk; zhenyu.yin@gmail.com

Abstract: Previous prediction models for soil compaction parameters were developed using limited data of specific soils and their accuracy also needs to be improved. This study presents the development of a new prediction model for the soil compaction parameters (i.e. optimum water content and maximum dry density) using the multi expression programming (MEP). Numerous soil compaction tests with a wide range of soil classifications and compaction energies are first collected to form a large database. Then, the optimal setting of the MEP code parameters is investigated and determined. The explicit formulations for the two key compaction parameters are finally proposed. The validity and the sensitivity analysis of the model are conducted. The results show that the proposed model enables to predict the soil compaction parameters for all kinds of soils in the database with high accuracy. The monotonicity analysis of the predicted compaction parameters with each input property (four physical properties of soil and one compaction energy) verifies the correctness and the validity of proposed model, showing consistency with the monotonicity concerning the actual data in the database. From the sensitivity analysis about the relevance of each input property on the predicted compaction parameters, it is indicated that the plastic limit and the fines content have more significant influences on the prediction results, while the effect of the liquid limit is the least pronounced.

Keywords: Genetic programming; soil compaction; optimum water content; maximum dry density; Atterberg limits; grain size distribution

Introduction

Soil compaction is the process of compacting the soil particles more closely together, by reducing the air void but maintaining the water content between the soil particles (Verma and Kumar 2019). Through soil compaction, the mechanical behaviors can be improved variously. Proctor (1933) recommended to compact the soil at a desired compaction energy with various water contents in the laboratory. In this way, the optimum water content and the maximum dry density can be obtained from the compaction curve. For the geotechnical practice, these two compaction parameters are widely used to maintain the long-term performance of numerous geotechnical structures, such as railway track-beds (Zhao et al. 2017; Chen et al. 2018; Wang et al. 2018a, 2018b, 2018c; Chen et al. 2019; Wang and Chen 2019; Wang et al. 2019a, 2019b; Zhao et al. 2019; Chen et al. 2020), highway embankments (Lim and Miller 2004; Kim et al. 2005; Rahman et al. 2008), earth dams (Di Matteo et al. 2009; Ardakani and Kordnaeij 2019), nuclear waste disposal (Delage et al. 2006; Sun et al. 2009), etc. Therefore, understanding and predicting the compaction parameters of different soils is of paramount importance for the construction and the maintenance of the geotechnical structures.

To date, the soil compaction parameters, i.e. optimum water content and maximum dry density, can be determined by both laboratory tests and analytical methods. As suggested by the widely accepted standards of ASTM D698 (ASTM 2012a, standard Proctor compaction) and ASTM D1557 (ASTM 2012b, modified Proctor compaction), the soil compaction curve can be obtained through the testing procedures of sample preparation, compaction with specified energy, measurement and calculation. In general, at least six tests need to be appropriately performed to identify the compaction curve with reasonable accuracy. Thus, the laboratory test for soil compaction is still time- and labor-consuming. To determine the soil compaction parameters more efficiently, several prediction models were proposed. Most of these models were built based on the regression analysis using limited data of specific soils (Blotz et al. 1998; Omar et al. 2003; Gurtug and Sridharan 2004; Sridharan and Nagaraj 2005; Di Matteo et al. 2009; Günaydın 2009; Mujtaba et al. 2013; Khuntia et al. 2015; Farooq et al. 2016). The prediction accuracies were scattered using these models, with the coefficients of determination ranging from 0.64 to 0.98. In addition, the prediction accuracy using these models appeared to decrease for a larger database (the number of samples for the regression analysis was smaller than 126 for these models, Verma and Kumar 2019).

To address the issue with a larger database and a higher accuracy, some machine learning techniques were applied to predict the compaction parameters. By the methods of artificial neural network (ANN, Sinha and Wang 2008) and evolutionary polynomial regression (EPR, Ahangar-Asr et al. 2011), the compaction parameters of 55 soil samples under standard Proctor compaction conditions were predicted. Ardakani and Kordnaeij (2019) used the group method of data handling (GMDH) type neural network to establish a more complex model to predict the compaction parameters for 212 samples. Based on 451 soil samples, Kurnaz and Kaya (2020) predicted the soil compaction parameters using GMDH–type neural network, support vector machine (SVM), Bayesian regularization neural network (BRNN), and extreme learning machine (ELM), while no explicit expressions were provided. On the whole, these prediction models show relatively higher coefficients of determination (from 0.90 to 0.98) than those using the regression analysis. However, in these studies, the soil type was limited. For example, in some studies, the soft clay with high plasticity was not considered, and in some others either the fine-grained soil or the coarse-grained soil could not be related. Furthermore, the influencing factors for the compaction parameters were not completely integrated in these models. Although the prediction accuracy can be guaranteed for a specific soil range in the previous studies, the issue about the limited soil type and the deficient consideration of the influencing factors might lead to predicting errors for the soils with different properties. Hence, for more convenient engineering application, a comprehensive model with explicit formulations to predict the soil compaction parameters needs to be developed, considering the overall range of soil types and influencing factors.

To fulfill this purpose, a genetic programming method namely multi expression programming (MEP) has proven to be an alternative and efficient approach to solve nonlinear and complex prediction problems. The MEP method was firstly proposed by Oltean and Dumitrescu (2002). Using this approach, the problem with multi computer programs can be encoded into a single chromosome. The best encoded prediction equations can be generated through the calculation process and be easily manipulated in practical cases. With the advantages of high efficiency, easy implementation and high predicting accuracy, the MEP approach has been developed rapidly in the field of geotechnical engineering for the last decade. Successful applications of this approach can be found in predicting compressive and tensile strengths (Baykasoğlu et al. 2008), peak ground acceleration (Cabalar and Cevik 2009), soil classification (Alavi et al. 2010), secant and reloading soil deformation moduli (Alavi et al. 2012), etc. Thereby, the prediction about the soil compaction parameters using

MEP is feasible, supported by previous relevant studies of this approach on the specific geotechnical problems.

In this study, the prediction model of soil compaction parameters, including the optimum water content and the maximum dry density, is developed using the approach of multi expression programming (MEP). A large database of soils encompassing various classifications (gravel, sand, silt, clay) are collected from the literature. The soils with the odd line number in the database are selected as the training data to generate the optimal formulations, which are then validated using the testing soils with the even line number in the database. Furthermore, the validity of the prediction model is evaluated by the analysis of monotonicity and sensitivity.

Multi expression programming (MEP)

Genetic algorithms (GA) is a stochastic method to search and to optimize the solution of a problem based on the principles of genetics and natural selection (Goldberg 1989). Through conventional optimization techniques, GA produces a series of binary strings to represent the solution. By evolving the string expression into the computer programs such as tree structures or functional programming language, the genetic programming (GP) was introduced as an extension of GA (Koza 1992). GP can be defined as a symbolic optimization technique, using Darwin's natural selection to solve a problem through computer programs. In GP, the main goal is to identify a program, connecting the known input and the known output variables based on the fitness function. Generally, GP can be categorized into three types: tree-based GP, graph-based GP and linear-based GP (Alavi et al. 2013; Cheng et al. 2020). Compared to the other two types of GP, the linear-based GP is more efficient as no expensive or slow interpreters are needed for this type. This leads to a more applicable value for the linear-based GP to improve the precision of the model in realistic timeframes.

In considerations of efficiency and accuracy, a representative linear-based GP approach so-called multi expression programming (MEP) was selected to predict the soil compaction parameters in this study. The MEP uses linear chromosomes to encode solutions. Multiple computer programs (solutions) can be encoded into a chromosome. By checking the fitness values of the programs, the best encoded solution is chosen to represent the chromosome. The MEP algorithm starts by developing a random population of computer programs. The following steps are repeated to develop the best program until reaching the termination condition (Oltean and Grosan 2003):

1. Two parents are selected using a procedure of binary tournament and recombined with a fixed crossover probability;
2. Two offspring are obtained through the recombination of two parents;
3. The offspring are mutated, and the worst individual in the current population is replaced by the best of them if the offspring is better than the worst individual in the current population.

The representation of MEP is similar to the way in which Pascal and C compliers convert mathematical expressions into machine code (Alavi et al. 2010, 2013). The MEP genes are represented by a string of expressions. The number of genes is defined as the chromosome length (code length), which is constant during the calculation. Each gene consists of one or two terminals (an element in the terminal set T) and a function symbol (an element in the function set F). Note that the first gene of a chromosome must be a terminal randomly chosen from the terminal set, to obtain the syntactically correct program. A gene with a function contains a pointer towards the function arguments. In a specific gene, the expressed terminal indices have lower values than the position of the gene in the chromosome.

An example of a MEP chromosome is illustrated as follows:

$P_0 : x_1$

$P_1 : x_2$

$P_2 : pow\ P_0,\ P_1$

$P_3 : x_3$

$P_4 :\ /\ P_2,\ P_3$

$P_5 : x_4$

$P_6 :\ +\ P_4,\ P_5$

In this example, the terminal set $T = \{x_1,\ x_2,\ x_3,\ x_4\}$ and the function set $F = \{pow,\ +,\ /\}$ are used. By reading the chromosome from top to bottom, the MEP genes can be translated into computer programs. The corresponding gene trees are plotted in Fig. 1. The genes 0, 1, 3 and 5 encode a single terminal as: $P_0 = x_1$, $P_1 = x_2$, $P_3 = x_3$, $P_5 = x_4$, respectively. Gene 2 indicates the operation pow on the operands at the locations 0 and 1 of the chromosome. Hence, gene 2 encodes the equation $P_2 = x_1^{x_2}$. Similarly, gene 4 indicates the operation / on the operands at the locations 2 and 3, developing the expression as $P_4 = x_1^{x_2} / x_3$. Finally, gene 6 can be expressed as $P_6 = x_1^{x_2} / x_3 + x_4$. Thus, the chromosome can be demonstrated as a forest of gene trees (Fig. 1), with multi expressions. The best expression is chosen after controlling the fitness of all expressions in a MEP chromosome (Oltean and Grosan 2003).

Model development

Geological database

To develop a prediction model of the soil compaction parameters for a wide range of soil classifications with high precision, 226 soil compaction tests were collected from the literature (Al-Khafaji 1993; Daniel and Wu 1993; Othman and Benson 1993; Shelley and Daniel 1993; Delage et al. 1996; Miller et al. 2002; Fleureau et al. 2003; Inci et al. 2003; Lim and Miller 2004; Clariá and Rinaldi 2007; Vassallo et al. 2007; Ito and Komine 2008; Shafiee 2008; Taïbi et al. 2008; Günaydın 2009; Horpibulsuk et al. 2009; Li 2009; Sawangsuriya et al. 2009; Agus et al. 2010; Akcanca and Aytekin 2012; Millogo et al. 2012; Heitor et al. 2013; Burton et al. 2014; Fox et al. 2014; Duong et al. 2014; Vega and McCartney 2015; Zhang et al. 2015; Jotisankasa and Taworn 2016; Kiliç et al. 2016; Al-Hussaini 2017; Wang et al. 2017; Chen et al. 2019; Di Sante 2019). The details of the soil properties can be downloaded and referred in the supplementary data. From the supplementary data, the soil properties including gravel content C_G, sand content C_S, fines content C_F, liquid limit LL, plastic limit PL, compaction energy E, optimum water content w_{opt} and maximum dry density $\rho_{d\max}$ are listed. Note that the gravel, sand and fines are differentiated by the grain size range of 75 mm to 4.75 mm, 4.75 mm to 0.075 mm and smaller than 0.075 mm, respectively (ASTM 2017). For the soil classification according to ASTM (2017), the investigated soils include lean clay (CL), silty clay (CL-ML), fat clay (CH), elastic silt (MH), silt (ML), clayey sand (SC), poorly graded sand with clay (SP-SC), well-graded sand with clay (SW-SC), silty sand (SM), clayey gravel (GC), poorly graded gravel with clay (GP-GC), well-graded gravel with clay (GW-GC) and silty gravel (GM). The samples in the database are ordered by the soil classification from clay to gravel. For the same classification, the samples are ordered alphabetically by the name of the author.

Table 1 lists the descriptive statistics of the soil properties in the present database, with the minimum, the maximum, the mean and the standard deviation of each variable. Fig. 2 plots the frequency histograms for each variable, also including the soil classification. From Fig. 2, it is indicated that for the present database, most soils have the gravel and the sand contents between 0 to 20 % (Figs. 2a and 2b), while the fines content in the soils lies mostly in the range of 80 % to 100 % (Fig. 2c). The majority of soils have the liquid limit from 0 to 50 (Fig. 2d). Around 32 soils have the liquid limit higher than 300, with high plasticity. For the plastic limit, it mainly locates in the range between 10 and 30 (Fig. 2e). The present database has a relatively higher amount of lean clay, fat clay, clayey sand and clayey gravel than the other soils (Fig. 2f). Most of the compaction tests were performed under the standard Proctor or the reduced compaction energy (no more than 600 kJ/m^3, Fig. 2g). About 30 modified Proctor compaction tests were also investigated. For the optimum water content and the maximum dry density, the highest frequency shows in the range of 10 % to 20 % and 1.6 Mg/m^3 to 1.8 Mg/m^3, respectively (Figs. 2h and 2i).

Figs. 3 and 4 depict the basic linear fittings of the optimum water content and the maximum dry density versus each input property, respectively. The coefficient of determination R^2 is used to evaluate the fitting accuracy, defined as:

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

where h_i and t_i are the actual and the predicted output values for the i th output; $\overline{h_i}$ is the average value of the actual outputs. The R^2 ranges from 0 to 1, showing the higher prediction accuracy with the R^2 value closer to 1. From Fig. 3, it can be observed that the optimum water content decreases with increase of gravel content, sand content or compaction energy, whereas it increases as the fines content, liquid limit or plastic limit increases. By contrast, the variation of the maximum dry density with each input property shows the opposite trend compared to the optimum water content (Fig. 4). For both compaction parameters, the fittings with the fines content or the plastic limit have relatively higher accuracy (with higher coefficient of determination R^2). The fittings with the compaction energy are the poorest. On the whole, all the R^2 values for the simple linear fittings are smaller than 0.55. A more comprehensive approach is needed to improve the prediction precision.

Model overview

From previous predictions on the soil compaction parameters, the properties of gravel content, sand content, fines content, liquid limit, plastic limit and compaction energy are the main influencing factors (Al-Khafaji 1993; Blotz et al. 1998; Omar et al. 2003; Gurtug and Sridharan 2004; Sridharan and Nagaraj 2005; Ito and Komine 2008; Sinha and Wang 2008; Di Matteo et al. 2009; Günaydın 2009; Ahangar-Asr et al. 2011; Mujtaba et al. 2013; Khuntia et al. 2015; Nagaraj et al. 2015; Farooq et al. 2016; Ardakani and Kordnaeij 2019; Kurnaz and Kaya 2020). However, as the soils are constituted by gravel, sand and fines, the gravel content can be calculated as the subtraction between 100 % and the contents of sand and fines. Thus, in the present prediction, the gravel content is not considered.

In order to develop an accurate model for predicting the soil compaction parameters and to check the validity of this model, the soils with the odd line number in the supplementary database are chosen as the training data (totally 113 samples). The soils with the even line number (totally 113 samples) in the supplementary database are set as the testing data to evaluate the developed model. In this way, both the training and the testing data cover all kinds of soil types in the present database, due to the order of the samples by the soil classification and then the name of the author. To assess the accuracy of the prediction, three indicators of mean absolute error MAE, root mean squared error $RMSE$ and coefficient of determination R^2 [see Eq. (1)] are introduced as:

$$MAE = \frac{\sum_{i=1}^{n} |h_i - t_i|}{n} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (h_i - t_i)^2}{n}} \quad (3)$$

where n is the number of outputs. For a specific group of data, the higher value of R^2 and the lower values of MAE and $RMSE$ indicate the higher precision of the model.

The source code of MEP (Oltean 2004) was used for the present analysis. Various code parameters need to be set for the program (Oltean and Dumitrescu 2002). The population size is defined as the number of programs in the population. The number of generation defines the number of calculations before the run of a program terminates. The increase of either parameter increases the running time of the program. The probabilities of crossover and mutation determine the probability of an offspring is subjected to the crossover and the mutation operators, respectively. The uniform crossover type indicates the offspring genes are randomly taken from one parent to another. The code length is the number of genes encoded in each chromosome. The replication number is the number of runs for each program (i.e. number of developed chromosomes).

New prediction model

To develop the model with optimal code parameters, three steps were followed. Firstly, typical initial code parameters were chosen from the suggested values of previous studies using the MEP approach to solve geotechnical problems. Trial and error method was applied on the training data to determine the initial optimal combination of the code parameters. Secondly, using the initial optimal combination as the basis, the investigation on the effect of each code parameter on the prediction accuracy was performed. That is to say, to study the influence of a specific code parameter, this parameter changes, while the other parameters are kept consistent with the initial optimal combination. Thirdly, the final optimal combination of the code parameters was adjusted according to the initial optimal setting and the parametric study. The explicit expressions were thus developed using the final optimal parametric combination.

Table 2 lists the initial parameter setting for the present MEP model of the training data, based on previous suggested values for the MEP method in geotechnical engineering (Oltean and Dumitrescu 2002; Baykasoğlu et al. 2008; Alavi et al. 2010, Shahnazari et al. 2010; Alavi et al. 2011, 2012, 2013; Mohammadzadeh et al. 2014). Note that the selection of the function set was also referred to previous predictions of the soil compaction parameters (Al-Khafaji 1993; Blotz et al. 1998; Omar et al. 2003; Gurtug and Sridharan 2004; Sridharan and Nagaraj 2005; Ito and Komine 2008; Sinha and Wang 2008; Di Matteo et al. 2009; Günaydın

2009; Ahangar-Asr et al. 2011; Mujtaba et al. 2013; Khuntia et al. 2015; Nagaraj et al. 2015; Farooq et al. 2016; Ardakani and Kordnaeij 2019; Kurnaz and Kaya 2020). From Table 2, it is indicated that for the initial calculation of either optimum water content or maximum dry density, the number of combination is $3\times3\times3\times3\times2=162$. From the trial and error calculations with the initial parameter setting (Table 2), the initial optimal parametric combination with the lowest MAE value was determined for both soil compaction parameters, as shown in Table 3.

Then, the code parametric investigation was conducted. For the parametric study, the investigated code parameter was changed in a reasonable range, with the other parameters were set as the values listed in Table 3. Figs. 5 and 6 present the effect of each code parameter on the prediction precision of the optimum water content and the maximum dry density, respectively. On the whole, similar trend about the prediction precision of the optimum water content and the maximum dry density can be identified with each code parameter. The prediction accuracy of the soil compaction parameters increases with the increase of population size, code length and number of generation in the beginning and tends to stabilize after the values of around 1000, 50 and 1000 are reached for each aforementioned code parameter. The crossover probability appears to have an insignificant influence on the prediction accuracy, while the crossover probability of 0.9 in the initial optimal set (Table 3) still generates a relatively lower MAE. As the mutation probability increases, the prediction accuracy decreases and tends to approach the minimum when the mutation probability reaches about 0.5.

From the parametric study in Figs. 5 and 6, the parameter setting in Table 3 is verified to induce the highest precision for the predicted model. However, the prediction accuracy with the code length of 50 enters the stable state, slightly lower than the case with the code length of 100 (Figs. 5b and 6b). Furthermore, the explicit expressions with the code length of 50 have only half number of the expressions compared to the case with the code length as 100. Therefore, in combined consideration of the complexity and the precision of the prediction expressions, the code length of 50 is thus chosen to develop the model. The other code parameter setting can be referred to Table 3 as the optimal combination. Using the aforementioned final code parameter setting, the optimal expressions for predicting the optimum water content of the training data are generated as follows:

$$w_{opt} = \frac{32}{PL} + \frac{A_1 + A_3}{A_2} + \frac{2A_1}{PL + E} \tag{4}$$

in which A_1, A_2 and A_3 are parameters as:

$$A_1 = \frac{PL^2}{1 + C_S} - \frac{2A_3}{LL} + C_S - C_F + 6 \tag{5}$$

$$A_2 = \frac{PL + E}{64} + C_F + PL + 9 \tag{6}$$

$$A_3 = (5 + C_F)(5 + PL) \tag{7}$$

The optimal expressions to predict the maximum dry density of the training data are developed as:

$$\rho_{d\max} = (\frac{2}{B_3})^{B_3} + \frac{2{B_3}^2 B_4}{C_F} - \frac{C_F}{B_6} + \frac{4C_S B_3}{C_F(C_F + 2LL - 2E)} \tag{8}$$

where B_1, B_2, B_3, B_4, B_5 and B_6 are parameters as:

$$B_1 = B_4 + E + \frac{C_S B_3}{C_F} \tag{9}$$

$$B_2 = \frac{B_4 + E}{B_3} + LL - E \tag{10}$$

$$B_3 = \lg PL \tag{11}$$

$$B_4 = \lg C_F \tag{12}$$

$$B_5 = \frac{C_F}{B_3} - \frac{2C_S {B_3}^2 B_4}{C_F} - \frac{B_1}{B_2} - PL - E - B_4 \tag{13}$$

$$B_6 = B_3(\frac{B_1}{B_2} - B_2 - B_5) \tag{14}$$

Fig. 7 plots the comparison of the predicted and reference soil compaction parameters using the developed model for the training data. It can be observed from Fig. 7 that both compaction parameters can be well predicted using the proposed model, showing slightly higher precision (higher R^2) for predicting the optimum water content than that of the maximum dry density. The values of MAE, $RMSE$ and R^2 for the prediction of the optimum water content are 1.206, 1.574 and 0.916, respectively. In terms of the prediction of the maximum dry density, the values of MAE, $RMSE$ and R^2 are estimated as 0.050, 0.069 and 0.872, respectively.

Results and discussions

Model validity

The soils with the even line number in the database are used as the testing data to evaluate the validity of the developed model. The comparisons between the predicted and the reference soil compaction parameters are presented in Fig. 8, also listing the values of MAE, $RMSE$ and R^2. The indicators showing high accuracy can be observed for both the optimum water content and the maximum dry density, with the R^2 values as 0.923 and 0.858, respectively. This suggests that the prediction of both compaction parameters using the proposed model is in good agreement with the testing data.

In addition, several previous prediction models of the soil compaction parameters are taken into comparison with the present model. Tables 4 and 5 list the expressions for predicting the optimum water content and the maximum dry density in the literature, respectively. In all the previous models, the liquid limit or the plastic limit was considered, while only some models included the compaction energy. In the models of Tables 4 and 5, the grain size distribution (the sand and the fines contents) was not taken into account.

To compare the prediction accuracy, the three indicators of MAE, $RMSE$ and R^2 are calculated for the testing data in the database using the present and the previous models. Fig. 9 shows the comparison result about each indicator for the predictions of the optimum water content. The R^2 for Al-Khafaji (1993), Blotz et al. (1998), Farooq et al. (2016) and Ito and Komine (2008) is not applicable, with negative values (Fig. 9c). It can be seen from Fig. 9 that the present model shows the highest prediction accuracy for the optimum water content, with the lowest MAE and $RMSE$, and the highest R^2. The predictions from the previous models present relatively poorer precision, with the R^2 values all smaller than 0.5. Note that by checking the predicted output optimum water content of the previous models, the predicting error mainly locates for the soils with high plasticity (when liquid limit is higher than 300 for the testing data). Nevertheless, using the present model, the predictions for the testing data show much higher precision, even for this kind of soil.

The comparison about the prediction accuracy of the maximum dry density for the testing data between the present and the previous studies is plotted in Fig. 10. The R^2 for Al-Khafaji (1993), Blotz et al. (1998), Farooq et al. (2016), Günaydın (2009) and Al-Khafaji

(1993)-2 is not applicable, with negative values (Fig. 10c). Similar to the optimum water content, the prediction of the maximum dry density using the present model shows the highest accuracy for the testing data. For the previous studies, the references of Sridharan and Nagaraj (2005), Nagaraj et al. (2015) and Günaydın (2009)-2 present relatively higher accuracies. By checking the predicting output values, these three models have lower accuracies for predicting the maximum dry density of the coarse-grained soils. Regarding the other previous models, the prediction error of the maximum dry density of the soils with high plasticity (when liquid limit is higher than 300) is the dominant issue to decrease the prediction precision.

From these comparisons, it is indicated that the present model enables the prediction of the compaction parameters for soils with a wide range of classifications, while the previous models have limitations for predicting the compaction parameters of some specific materials such as soils with high plasticity or coarse-grained soils.

Monotonicity analysis

To further validate the feasibility of the developed model for soil compaction parameters, the monotonicity analysis is performed. This analysis aims at implying whether the proposed expressions are correct or not (Jin et al. 2019; Jin and Yin 2020). Through the monotonicity analysis, one input soil property changes, while the other input soil properties stay constant. Using these input soil properties and the proposed model from the MEP, the variation of the predicted soil compaction parameter with the varying soil property can be identified. The basic setting of the soil properties for the monotonicity analysis is listed in Table 6. The sand content, the fines content, the liquid limit and the plastic limit are chosen as their respective mean value from the whole supplementary database (see Table 1). For the compaction energy, several fixed values are used generally (ASTM 2012a, 2012b). It is not reasonable to use the mean compaction energy as the basis for the monotonicity analysis. Thus, the commonly used compaction energy of 593 kJ/m^3 for the standard Proctor compaction test is selected. During the monotonicity calculation, the varying range of the input soil property cannot exceed the range defined by its minimum and maximum values as shown in Table 1. Note also that the sand content is not separately considered for the monotonicity analysis, because the cohesionless soil particles (including both the gravel and the sand) jointly influence the soil compaction parameters. To simulate the practical case, the summation of the contents of sand and fines cannot exceed 100 %.

Fig. 11 shows the variation of the predicted optimum water content with the fines content, the liquid limit, the plastic limit and the compaction energy, respectively. It is observed that the predicted w_{opt} increases monotonically with the increase of fines content, liquid limit or plastic limit, whereas the predicted value decreases monotonically as the compaction energy increases. These monotonic variations are consistent with the trend of the monotonicity concerning the actual optimum water content with each soil property in Fig. 3, suggesting the correctness of the proposed model.

The monotonicity analysis of the predicted maximum dry density with each input variable is presented in Fig. 12. Opposite to the optimum water content, the predicted $\rho_{d\max}$ decreases with the increase of fines content, liquid limit or plastic limit, but with the decrease of the compaction energy. This observation corresponds to that for the monotonicity of the actual database shown in Fig. 4. Thereby, the consistency between the monotonicity analysis and the database for the prediction of the maximum dry density also verifies the validity of the proposed model.

Sensitivity analysis

To have a better understanding about the importance of the input soil property on the predicted soil compaction parameters, the sensitivity analysis on the overall database is conducted. For a specific input variable m_i, the sensitivity R_{sen} can be calculated as (Momeni et al. 2014; Ardakani and Kordnaeij 2019):

$$R_{sen} = \frac{\sum_{i=1}^{N}(m_i t_i)}{\sqrt{\sum_{i=1}^{N} m_i^2 \sum_{i=1}^{N} t_i^2}} \tag{15}$$

where t_i is the predicted output variable using the proposed prediction model; N is the number of the soil compaction tests in the supplementary database (in this study, $N = 226$). The sensitivity value R_{sen} ranges from 0 to 1, showing the related strength between each input variable and the predicted output variable. With the R_{sen} value closer to 1, the specific input variable has a more significant influence on the predicted output variable.

Fig. 13 depicts the sensitivity distribution of each input property for the predicted optimum water content and maximum dry density regarding the whole database. All the five input variables affect the prediction of the soil compaction parameters using the proposed model (all sensitivity values are higher than 0.5), but with different impacts. For both soil compaction parameters, the sensitivities of the plastic limit and the fines content rank the highest, indicating that these two input parameters are the most significant properties affecting the prediction results. For the prediction of the optimum water content, the sensitivities of plastic limit and fines content are much higher than the others (Fig. 13a).

However, the distribution of the sensitivity is more uniform for the prediction of the maximum dry density: the sensitivities of sand content and compaction energy only show slightly smaller values than those of plastic limit and fines content (Fig. 13b). By contrast, the liquid limit has the lowest sensitivity value or the least influence for predicting both soil compaction parameters by the present model.

Conclusions

In this study, the prediction model of soil compaction parameters, i.e. the optimum water content and the maximum dry density, was developed using a method of genetic programming, namely multi expression programming (MEP). A large database of soils covering a wide range of classifications was formed by collecting data of published works.

The optimal parameter setting in the MEP source code was first investigated. The prediction accuracy of both soil compaction parameters increases with the increase of population size, code length or number of generation, but decreases as the mutation probability increases. The crossover probability has an insignificant influence on the prediction accuracy. Through this investigation, the final optimal setting of the code parameters was obtained (population size: 2000; code length: 50; crossover probability: 0.9; crossover type: uniform; mutation probability: 0.01; number of generation: 2000; function set: +, -, ×, /, pow, lg; terminal set: problem input; replication number: 10).

Then, using the final optimal combination of the code parameters, the explicit formulations for the prediction of the soil compaction parameters were obtained. The proposed model predicts the optimum water content and the maximum dry density with reasonable accuracy for the training data, with R^2 higher than 0.87. This model is also valid for predicting the soil compaction parameters from the testing data ($R^2 > 0.85$). Compared to the previous studies, the present model is more applicable and reliable for a wider range of soil classifications, even for the soils with high plasticity and coarse-grained soils from the database.

The validity of the proposed prediction model and the sensitivity of the input variables on the predicted results were further analyzed. The monotonicity analysis of the predicted compaction parameters shows similar variation trend as that for the actual data in the whole database, verifying that the proposed model is reasonable and valid. From the sensitivity analysis, the plastic limit and the fines content have more remarkable influences on the predictions of compaction parameters, while the effect of the liquid limit is the least significant.

The wide range of soils based prediction model requiring only four physical properties of soils and compaction energy should be practically useful and can be recommended for geotechnical application. For instance, for a given soil with the four basic physical properties known, the optimum water content and the maximum dry density can be estimated immediately by the proposed model assuming a big value of the compaction energy. Then the compaction energy can be easily determined according to the design required values of compaction parameters.

Notations

A_1, A_2, A_3 Parameters for predicting optimum water content

B_1, B_2, B_3, B_4, B_5, B_6 Parameters for predicting maximum dry density

C_F Fines content

C_G Gravel content

C_S Sand content

E Compaction energy

F Function set

h_i Actual output variable

$\overline{h_i}$ Average value of actual outputs

LL Liquid limit

m_i Input variable

MAE Mean absolute error

n Number of outputs

N Number of soil compaction tests in the database

PL Plastic limit

PI Plasticity index

R^2 Coefficient of determination

R_{sen} Sensitivity

$RMSE$ Root mean squared error

t_i Predicted output variable

T Terminal set

w_{opt} Optimum water content

$\rho_{d\max}$ Maximum dry density

$\gamma_{d\max}$ Maximum dry unit weight

Acknowledgement

The financial supports provided by the RIF project (Grant No. PolyU R5037-18F) from Research Grants Council (RGC) of Hong Kong are gratefully acknowledged.

References

Agus, S.S., Schanz, T., Fredlund, D.G., 2010. Measurements of suction versus water content for bentonite–sand mixtures. Can. Geotech. J. 47 (5), 583-594.

Ahangar-Asr, A., Faramarzi, A., Mottaghifard, N., Javadi, A.A., 2011. Modeling of permeability and compaction characteristics of soils using evolutionary polynomial regression. Comput. Geosci. 37 (11), 1860-1869.

Akcanca, F., Aytekin, M., 2012. Effect of wetting–drying cycles on swelling behavior of lime stabilized sand–bentonite mixtures. Environ. Earth Sci. 66 (1), 67-74.

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

Alavi, A.H., Gandomi, A.H., 2011. A robust data mining approach for formulation of geotechnical engineering systems. Eng. Computation. 28 (3), 242-274.

Alavi, A.H., Mollahasani, A., Gandomi, A.H., Bazaz, J.B., 2012. Formulation of secant and reloading soil deformation moduli using multi expression programming. Eng. Computation. 29 (2), 173-197.

Alavi, A.H., Gandomi, A.H., Nejad, H.C., Mollahasani, A., Rashed, A., 2013. Design equations for prediction of pressuremeter soil deformation moduli utilizing expression programming systems. Neural Comput. Appl. 23 (6), 1771-1786.

Al-Hussaini, O., 2017. Volume change behaviour of some geomaterials under combined influence of freeze-thaw and wet-dry cycles: An experimental investigation. PhD Thesis, Cardiff University, UK.

Al-Khafaji, A.N., 1993. Estimation of soil compaction parameters by means of Atterberg limits. Q. J. Eng. Geol. Hydroge. 26 (4), 359-368.

Ardakani, A., Kordnaeij, A., 2019. Soil compaction parameters prediction using GMDH-type neural network and genetic algorithm. Eur. J. Environ. Civil Eng. 23 (4), 449-462.

ASTM., 2012a. Standard test method for laboratory compaction characteristics of soil using standard effort. ASTM D698. West Conshohocken, PA: ASTM.

ASTM., 2012b. Standard test method for laboratory compaction characteristics of soil using modified effort. ASTM D1557. West Conshohocken, PA: ASTM.

ASTM., 2017. Standard practice for classification of soils for engineering purposes (unified soil classification system). ASTM D2487-17, West Conshohocken, PA.

Baykasoğlu, A., Güllü, H., Çanakçı, H., Özbakır, L., 2008. Prediction of compressive and tensile strength of limestone via genetic programming. Expert Syst. Appl. 35 (1-2), 111-123.

Blotz, L.R., Benson, C.H., Boutwell, G.P., 1998. Estimating optimum water content and maximum dry unit weight for compacted clays. J. Geotech. Geoenviron. Eng. 124 (9), 907–912.

Burton, G.J., Sheng, D., Campbell, C., 2014. Bimodal pore size distribution of a high-plasticity compacted clay. Géotech. Lett. 4 (2), 88-93.

Cabalar, A.F., Cevik, A., 2009. Genetic programming-based attenuation relationship: an application of recent earthquakes in turkey. Comput. Geosci. 35 (9), 1884-1896.

Chen, R.P., Liu, Q.W., Wang, H.L., Liu, Y., Ma, Q.L., 2020. Performance of geosynthetic-reinforced pile-supported embankment on soft marine deposit. P. I. Civil Eng.-Geotech. Eng. in press. doi: 10.1680/jgeen.19.00136.

Chen, R.P., Qi, S., Wang, H.L., Cui, Y. J., 2019. Microstructure and hydraulic properties of coarse-grained subgrade soil used in high-speed railway at various compaction degrees. J. Mater. Civil Eng. 31 (12), 04019301.

Chen, R.P., Wang, H.L., Hong, P.Y., Cui, Y.J., Qi, S., Cheng W., 2018. Effects of degree of compaction and fines content of the subgrade bottom layer on moisture migration in the substructure of high-speed railways. P. I. Mech. Eng. F-J. Rail Rapid Transit 234 (4), 1197–1210.

Cheng Z.L., Zhou W.H., Garg A., 2020. Genetic programming model for estimating soil suction in shallow soil layers in the vicinity of a tree, Eng. Geol. 268, 105506.

Clariá, J.J., Rinaldi, V.A., 2007. Shear wave velocity of a compacted clayey silt. Geotech. Test. J. 30 (5), 399-408.

Daniel, D.E., Wu, Y.K. 1993. Compacted clay liners and covers for arid sites. J. Geotech. Eng. 119 (2), 223-237.

Delage, P., Audiguier, M., Cui, Y.J., Howat, M.D., 1996. Microstructure of a compacted silt. Can. Geotech. J. 33 (1), 150-158.

Delage, P., Marcial, D., Cui, Y.J., Ruiz, X., 2006. Ageing effects in a compacted bentonite: a microstructure approach. Géotechnique 56 (5), 291-304.

Di Matteo, L., Bigotti, F., Ricco, R., 2009. Best-fit models to estimate modified proctor properties of compacted soil. J. Geotech. Geoenviron. Eng. 135 (7), 992-996.

Di Sante, M., 2019. On the compaction characteristics of soil-lime mixtures. Geotech. Geol. Eng. 38, 2335–2344.

Duong, T.V., Cui, Y.J., Tang, A.M., Dupla, J.C., Canou, J., Calon, N., Robinet, A., 2014. Investigating the mud pumping and interlayer creation phenomena in railway sub-structure. Eng. Geol. 171, 45-58.

Farooq, K., Khalid, U., Mujtaba, H., 2016. Prediction of compaction characteristics of fine-grained soils using consistency limits. Ara. J. Sci. Eng. 41 (4), 1319–1328.

Fleureau, J.M., Hadiwardoyo, S., Correia, A.G., 2003. Generalised effective stress analysis of strength and small strains behaviour of a silty sand, from dry to saturated state. Soils Found. 43 (4), 21-33.

Fox, P.J., Thielmann, S.S., Stern, A.N., Athanassopoulos, C., 2014. Interface shear damage to a HDPE geomembrane. I: Gravelly compacted clay liner. J. Geotech. Geoenviron. Eng. 140 (8), 04014039.

Goldberg, D., 1989. Genetic algorithms in search, optimization and machine learning. MA: Addison-Welsley.

Günaydın, O., 2009. Estimation of soil compaction parameters by using statistical analyses and artificial neural networks. Environ. Geol. 57 (1), 203.

Gurtug, Y., Sridharan, A., 2004. Compaction behaviour and prediction of its characteristics of fine grained soils with particular reference to compaction energy. Soils Found. 44 (5), 27–36.

Heitor, A., Indraratna, B., Rujikiatkamjorn, C., 2013. Laboratory study of small-strain behavior of a compacted silty sand. Can. Geotech. J. 50 (2), 179-188.

Horpibulsuk, S., Katkan, W., Naramitkornburee, A., 2009. Modified Ohio's curves: A rapid estimation of compaction curves for coarse-and fine-grained soils. Geotech. Test. J. 32 (1), 64-75.

Inci, G., Yesiller, N., Kagawa, T., 2003. Experimental investigation of dynamic response of compacted clayey soils. Geotech. Test. J. 26 (2), 125-141.

Ito, H., Komine, H., 2008. Dynamic compaction properties of bentonite-based materials. Eng. Geol. 98 (3-4), 133-143.

Jin, Y.F., Yin, Z.Y., 2020. An intelligent multi-objective EPR technique with multi-step model selection for correlations of soil properties. Acta Geotech. in press. doi: 10.1007/s11440-020-00929-5.

Jin, Y.F., Yin, Z.Y., Zhou, W.H., Yin, J.H., Shao, J.F., 2019. A single-objective EPR based model for creep index of soft clays considering L2 regularization. Eng. Geol. 248, 242-255.

Jotisankasa, A., Taworn, D., 2016. Direct shear testing of clayey sand reinforced with live stake. Geotech. Test. J. 39 (4), 608-623.

Khuntia, S., Mujtaba, H., Patra, C., Farooq, K., Sivakugan, N., Das, B.M., 2015. Prediction of compaction parameters of coarse grained soil using multivariate adaptive regression splines (MARS). Int. J. Geotech. Eng. 9 (1), 79-88.

Kiliç, R., Küçükali, Ö., Ulamiş, K., 2016. Stabilization of high plasticity clay with lime and gypsum (Ankara, Turkey). B. Eng. Geol. Environ. 75 (2), 735-744.

Kim, B., Prezzi, M., Salgado, R., 2005. Geotechnical properties of fly and bottom ash mixtures for use in highway embankments. J. Geotech. Geoenviron. Eng. 131 (7), 914-

924.

Koza, J.R., 1992. Genetic programming, on the programming of computers by means of natural selection. MIT Press, Cambridge, MA.

Kurnaz, T.F., Kaya, Y., 2020. The performance comparison of the soft computing methods on the prediction of soil compaction parameters. Arab. J. Geosci. 13 (4), 159.

Li, X., 2009. Dual-porosity structure and bimodal hydraulic property functions of coarse granular soils. Ph. D. Thesis, Hong Kong University of Science and Technology, Hong Kong.

Lim, Y.Y., Miller, G.A., 2004. Wetting-induced compression of compacted Oklahoma soils. J. Geotech. Geoenviron. Eng., 130 (10), 1014-1023.

Miller, C.J., Yesiller, N., Yaldo, K., Merayyan, S., 2002. Impact of soil type and compaction conditions on soil water characteristic. J. Geotech. Geoenviron. Eng., 128 (9), 733-742.

Millogo, Y., Morel, J.C., Traoré, K., Ouedraogo, R. 2012. Microstructure, geotechnical and mechanical characteristics of quicklime-lateritic gravels mixtures used in road construction. Constr. Build. Mat. 26 (1), 663-669.

Mohammadzadeh, D., Bazaz, J.B., Alavi, A.H., 2014. An evolutionary computational approach for formulation of compression index of fine-grained soils. Eng. Appl. Artif. Intel. 33, 58-68.

Momeni, E., Nazir, R., Armaghani, D.J., Maizir, H., 2014. Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN. Measurement, 57, 122-131.

Mujtaba, H., Farooq, K., Sivakugan, N., Das, B.M., 2013. Correlation between gradational parameters and compaction characteristics of sandy soils. Int. J. Geotech. Eng. 7 (4), 395-401.

Nagaraj, H.B., Reesha, B., Sravan, M.V., Suresh, M.R., 2015. Correlation of compaction characteristics of natural aoils with modified plastic limit. Transpor. Geotech. 2, 65–77.

Oltean, M., 2004, Multi expression programming source code. Available at: https://www.mepx.org/source_code.html

Oltean, M., Dumitrescu, D., 2002, Multi expression programming. Technical Report, UBB-01-2002, Babes-Bolyai University, Cluj-Napoca.

Oltean, M., Grosan, C., 2003. A comparison of several linear genetic programming techniques. Complex Syst. 14 (4), 285-314.

Omar, M., Shanableh, A., Basma, A., Barakat, S., 200). Compaction characteristics of granular soils in United Arab Emirates. Geotech. Geol. Eng. 21 (3), 283-295.

Othman, M. A., Benson, C.H., 1993. Effect of freeze–thaw on the hydraulic conductivity and morphology of compacted clay. Can. Geotech. J., 30 (2), 236-246.

Proctor, R.R., 1933. Fundamental principles of soil compaction. Eng. News-Record 111, 13.

Rahman, F., Hossain, M., Hunt, M.M., Romanoschi, S.A., 2008. Soil stiffness evaluation for compaction control of cohesionless embankments. Geotech. Test. J. 31 (5), 442-451.

Sawangsuriya, A., Edil, T.B., Bosscher, P.J., 2009. Modulus-suction-moisture relationship for compacted soils in postcompaction state. J. Geotech. Geoenviron. Eng., 135 (10), 1390-1403.

Shafiee, A., 2008. Permeability of compacted granule–clay mixtures. Eng. Geol. 97 (3-4), 199-208.

Shahnazari, H., Dehnavi, Y., Alavi, A.H., 2010. Numerical modeling of stress–strain behavior of sand under cyclic loading. Eng. Geol. 116 (1-2), 53-72.

Shelley, T.L., Daniel, D.E., 1993. Effect of gravel on hydraulic conductivity of compacted soil liners. J. Geotech. Eng. 119 (1), 54-68.

Sinha, S.K., Wang, M.C., 2008. Artificial neural network prediction models for soil compaction and permeability. Geotech. Geol. Eng. 26 (1), 47-64.

Sridharan, A., Nagaraj, H.B., 2005. Plastic limit and compaction characteristics of finegrained soils." P. I. Civil Eng.-Ground Improvement 9 (1), 17–22.

Sun, D.A., Cui, H., Sun, W., 2009. Swelling of compacted sand–bentonite mixtures. Appl. Clay Sci. 43 (3-4), 485-492.

Taïbi, S., Fleureau, J.M., Hadiwardoyo, S., Kheirbek-Saoud, S., 2008. Small and large strain behaviour of an unsaturated compacted silt. Eur. J. Environ. Civil Eng. 12 (3), 203-228.

Vassallo, R., Mancuso, C., Vinale, F., 2007. Effects of net stress and suction history on the small strain stiffness of a compacted clayey silt. Can. Geotech. J. 44 (4), 447-462.

Vega, A., McCartney, J.S., 2015. Cyclic heating effects on thermal volume change of silt. Environ. Geotech. 2 (5), 257-268.

Verma, G., Kumar, B., 2019. Prediction of compaction parameters for fine-grained and coarse-grained soils: a review. Int. J. Geotech. Eng. 1-8.

Wang, H.L., Cui, Y.J., Lamas-Lopez, F., Dupla, J.C., Canou, J., Calon, N., Saussine, G., Aimedieu, P., Chen, R.P., 2017. Effects of inclusion contents on resilient modulus and damping ratio of unsaturated track-bed materials. Can. Geotech. J., 54 (12), 1672-1681.

Wang, H.L., Chen, R.P., Qi, S., Cheng, W., Cui, Y.J., 2018a. Long-term performance of pile-supported ballastless track-bed at various water levels. J. Geotech. Geoenviron. Eng. 144 (6), 04018035.

Wang, H.L., Cui, Y.J., Lamas-Lopez, F., Calon, N., Saussine, G., Dupla, J.C., Canou, J., Aimedieu, P., Chen, R.P., 2018b. Investigation on the mechanical behavior of track-bed materials at various contents of coarse grains. Constr. Build. Mat. 164, 228-237.

Wang, H.L., Cui, Y.J., Lamas-Lopez, F., Dupla, J.C., Canou, J., Calon, N., Saussine, G., Aimedieu, P., Chen, R. P., 2018c. Permanent deformation of track-bed materials at various inclusion contents under large number of loading cycles. J. Geotech. Geoenviron. Eng. 144 (8), 04018044.

Wang, H.L., Chen, R.P., 2019. Estimating static and dynamic stresses in geosynthetic-reinforced pile-supported track-bed under train moving loads. J. Geotech. Geoenviron. Eng., 145 (7), 04019029.

Wang, H.L., Chen, R.P., Cheng, W., Qi, S., Cui, Y.J., 2019a. Full-scale model study on variations of soil stress in the geosynthetic-reinforced pile-supported track-bed with water level change and cyclic loading. Can. Geotech. J., 56 (1), 60–68.

Wang, H.L., Chen, R.P., Liu, Q.W., Kang, X., 2019b. Investigation on geogrid reinforcement and pile efficacy in geosynthetic-reinforced pile-supported track-bed. Geotext. Geomembranes 47 (6), 755-766.

Zhang, X., Mavroulidou, M., Gunn, M.J., 2015. Mechanical properties and behaviour of a partially saturated lime-treated, high plasticity clay. Eng. Geol. 193, 320-336.

Zhao, L.S., Zhou, W.H., Yuen, K.V., 2017. A simplified axisymmetric model for column supported embankment systems. Comput. Geotech. 92, 96-107.

Zhao, L.S., Zhou, W.H., Geng, X., Yuen, K.V., Fatahi, B., 2019. A closed-form solution for column-supported embankments with geosynthetic reinforcement. Geotext. Geomembranes 47 (3), 389-401.

Tables

Table 1. Descriptive statistics of variables

Variable	Minimum	Maximum	Mean	Standard deviation
C_G (%)	0.0	67.1	7.5	14.5
C_S (%)	0.0	89.0	29.5	23.3
C_F (%)	8.6	100.0	63.1	29.9
LL (%)	16.0	608.0	108.7	163.9
PL (%)	6.1	48.3	22.0	7.4
E (kJ/m^3)	154.5	2755	893.8	733.9
w_{opt} (%)	5.3	43.7	17.5	6.0
ρ_{dmax} (Mg/m^3)	1.09	2.33	1.75	0.20

Table 2. Initial parameter setting for the MEP calculation

Parameter	Setting
Population size	500, 1000, 2000
Code length	50, 100
Crossover probability	0.1, 0.5, 0.9
Crossover type	Uniform
Mutation probability	0.01, 0.1, 0.9
Number of generation	500, 1000, 2000
Function set	+, -, ×, /, pow, lg
Terminal set	Problem input
Replication number	10

Table 3. Initial optimal combination of MEP code parameters

Parameter	Setting
Population size	2000
Code length	100
Crossover probability	0.9
Crossover type	Uniform
Mutation probability	0.01
Number of generation	2000
Function set	+, -, ×, /, pow, lg
Terminal set	Problem input
Replication number	10

Table 4. Prediction models for optimum water content in the literature

No.	Reference	Model
(1)	Al-Khafaji (1993)	$w_{opt} = 0.14LL + 0.54PL$
(2)	Blotz et al. (1998)	$w_{opt} = (12.39 - 12.21\lg LL)\lg E + 0.67LL + 9.21$
(3)	Farooq et al. (2016)	$w_{opt} = 0.133LL + 0.02PI - 5.99\lg E + 28.60$
(4)	Günaydın (2009)	$w_{opt} = 0.8442PL + 0.1076$
(5)	Gurtug and Sridharan (2004)	$w_{opt} = (-0.344\lg E + 1.88)PL$
(6)	Ito and Komine (2008)	$w_{opt} = \dfrac{80.1PL}{15.6\ln E - 29.9}$
(7)	Nagaraj et al. (2015)	$w_{opt} = 0.76PL$
(8)	Sridharan and Nagaraj (2005)	$w_{opt} = 0.92PL$

Note: In the reference of (3), the plasticity index (*PI*) is used.

Table 5. Prediction models for maximum dry density in the literature

No.	Reference	Model
(1)	Al-Khafaji (1993)	$\rho_{d\max} = 2.27 - 0.019PL - 0.003LL$
(2)	Blotz et al. (1998)	$\gamma_{d\max} = (2.27\lg LL - 0.94)\lg E - 0.16LL + 17.02$
(3)	Farooq et al. (2016)	$\gamma_{d\max} = -0.055LL + 0.014PI + 2.21\lg E + 12.84$
(4)	Günaydın (2009)	$\gamma_{d\max} = -0.1008LL + 21.16$
(7)	Nagaraj et al. (2015)	$\gamma_{d\max} = 20.82 - 0.17PL$
(8)	Sridharan and Nagaraj (2005)	$\gamma_{d\max} = 0.23(93.3 - PL)$
(9)	Al-Khafaji (1993)-2	$\rho_{d\max} = 2.44 - 0.02PL - 0.008LL$
(10)	Günaydın (2009)-2	$\gamma_{d\max} = -0.2283PL + 21.88$

Note: In the references of (2), (3), (4), (7), (8) and (10), the maximum dry unit weight $\gamma_{d\max}$ is introduced. In the reference of (3), the plasticity index (*PI*) is used.

Table 6. Basic setting of soil properties for the monotonicity analysis

Soil parameter	Value
C_S (%)	29.5
C_F (%)	63.1
LL (%)	108.7
PL (%)	22.0
E (kJ/m^3)	593

Figure captions

Fig. 1. Expression of a MEP chromosome represented by gene trees

Fig. 2. Frequency histograms of the variables: (a) gravel content; (b) sand content; (c) fines content; (d) liquid limit; (e) plastic limit; (f) soil classification; (g) compaction energy; (h) optimum water content; (i) maximum dry density. Note: For brevity of Fig. 2f, the soil CL-ML is merged into CL; the soils SP-SC and SW-SC are merged into SC; the soils GP-GC and GW-GC are merged into GC.

Fig. 3. Basic linear fittings between optimum water content and each soil parameter: (a) gravel content; (b) sand content; (c) fines content; (d) liquid limit; (e) plastic limit; (f) compaction energy

Fig. 4. Basic linear fittings between maximum dry density and each soil parameter: (a) gravel content; (b) sand content; (c) fines content; (d) liquid limit; (e) plastic limit; (f) compaction energy

Fig. 5. Effects of code parameters on the prediction accuracy of optimum water content for the training data: (a) population size; (b) code length; (c) crossover probability; (d) mutation probability; (e) number of generation

Fig. 6. Effects of code parameters on the prediction accuracy of maximum dry density for the training data: (a) population size; (b) code length; (c) crossover probability; (d) mutation probability; (e) number of generation

Fig. 7. Comparison of MEP-predicted and reference output variables for the training data: (a) optimum water content; (b) maximum dry density

Fig. 8. Comparison of MEP-predicted and reference output variables for the testing data: (a) optimum water content; (b) maximum dry density

Fig. 9. Comparison of the prediction accuracy of optimum water content between the present and previous studies. Note: The coefficients of determination R^2 for the references of Al-Khafaji (1993), Blotz et al. (1998), Farooq et al. (2016), and Ito and Komine (2008) are

not applicable (being negative) for the testing data.

Fig. 10. Comparison of the prediction accuracy of the maximum dry density between the present and previous studies. Note: The coefficients of determination R^2 for the references of Al-Khafaji (1993), Blotz et al. (1998), Farooq et al. (2016), Günaydın (2009) and Al-Khafaji (1993)-2 are not applicable (being negative) for the testing data.

Fig. 11. Monotonicity analysis of the predicted optimum water content versus (a) fines content; (b) liquid limit; (c) plastic limit; (d) compaction energy

Fig. 12. Monotonicity analysis of the predicted maximum dry density versus (a) fines content; (b) liquid limit; (c) plastic limit; (d) compaction energy

Fig. 13. Sensitivity analysis about the effect of the input variables on the output variables: (a) optimum water content; (b) maximum dry density

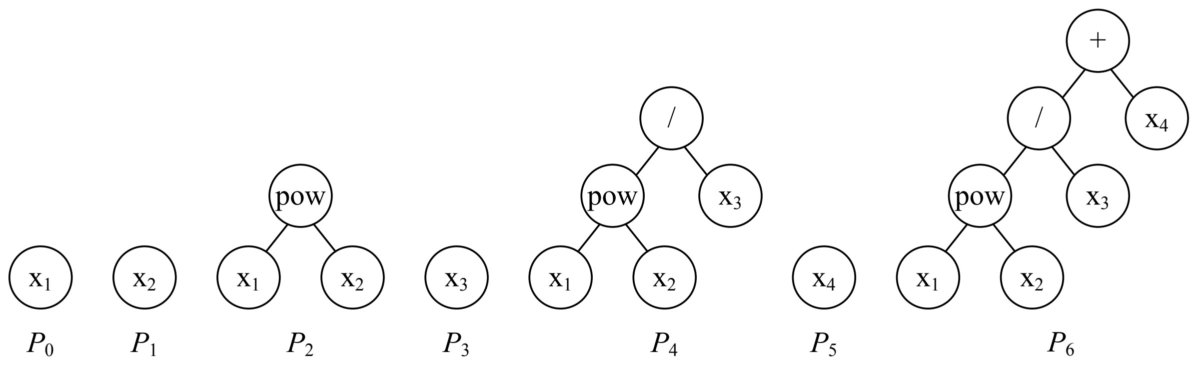


Fig. 1. Expression of a MEP chromosome represented by gene trees

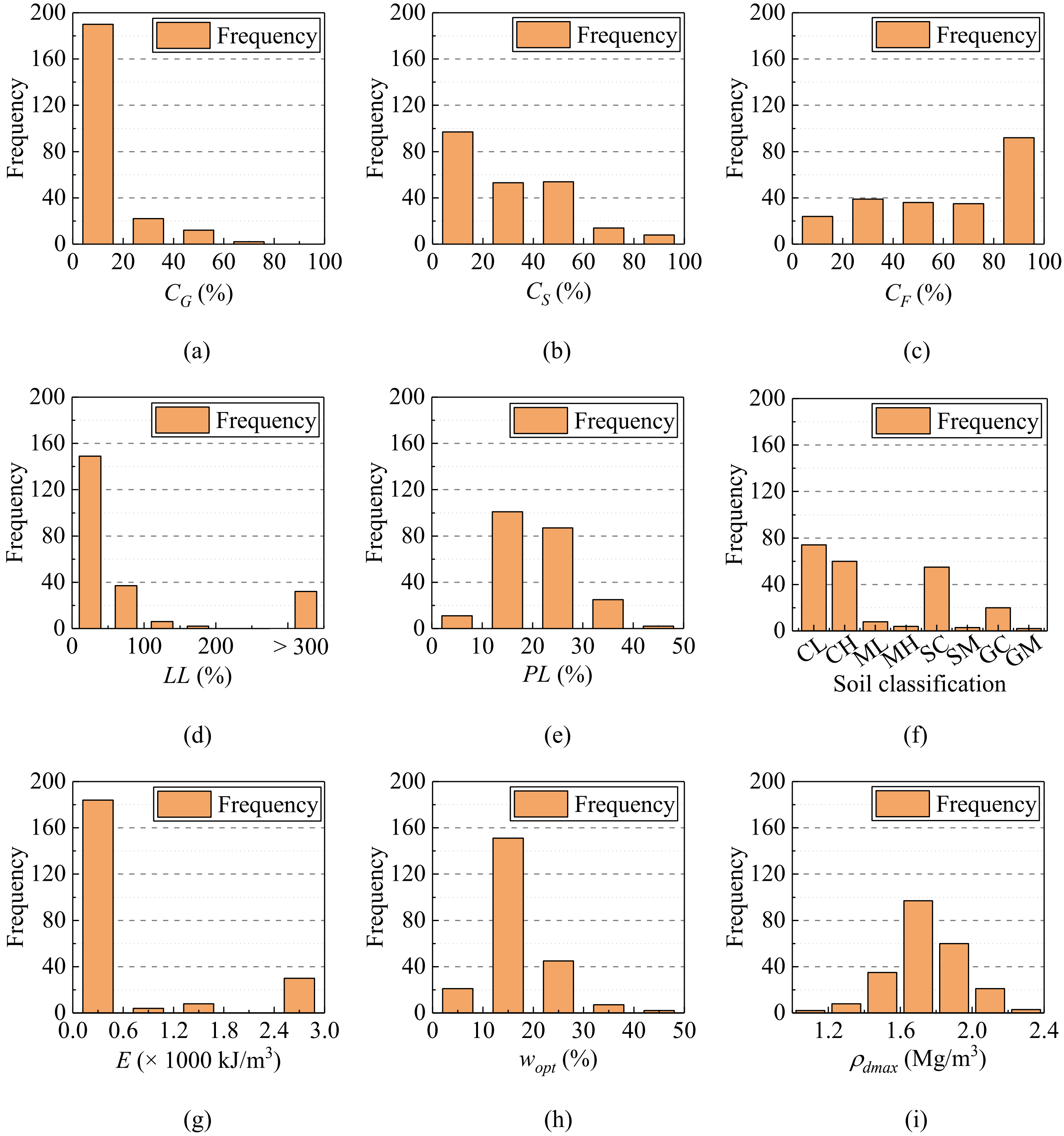


Fig. 2. Frequency histograms of the variables: (a) gravel content; (b) sand content; (c) fines content; (d) liquid limit; (e) plastic limit; (f) soil classification; (g) compaction energy; (h) optimum water content; (i) maximum dry density (Note: For brevity of Fig. 2f, the soil CL-ML is merged into CL; the soils SP-SC and SW-SC are merged into SC; the soils GP-GC and GW-GC are merged into GC.)

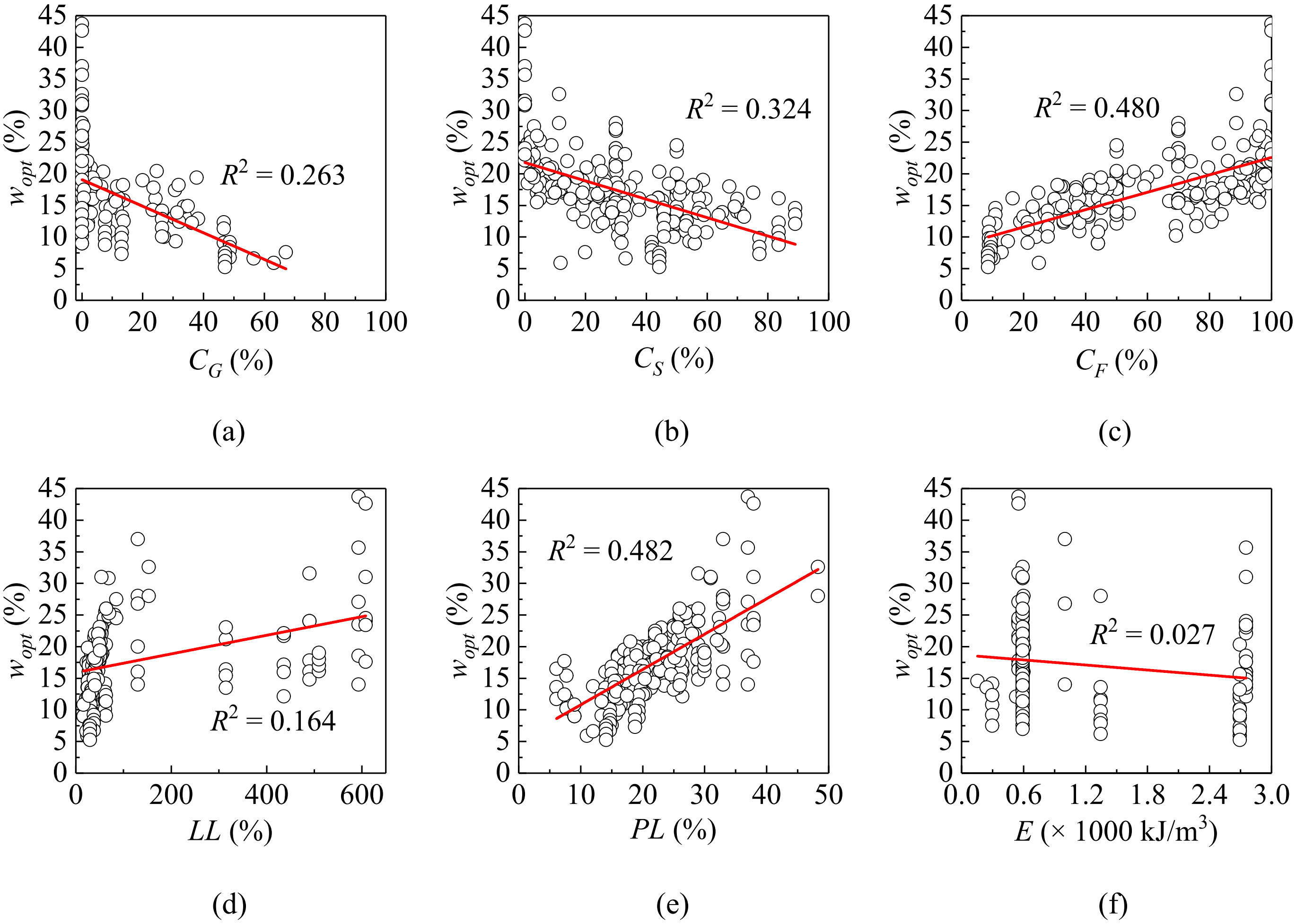


Fig. 3. Basic linear fittings between optimum water content and each soil parameter: (a) gravel content; (b) sand content; (c) fines content; (d) liquid limit; (e) plastic limit; (f) compaction energy

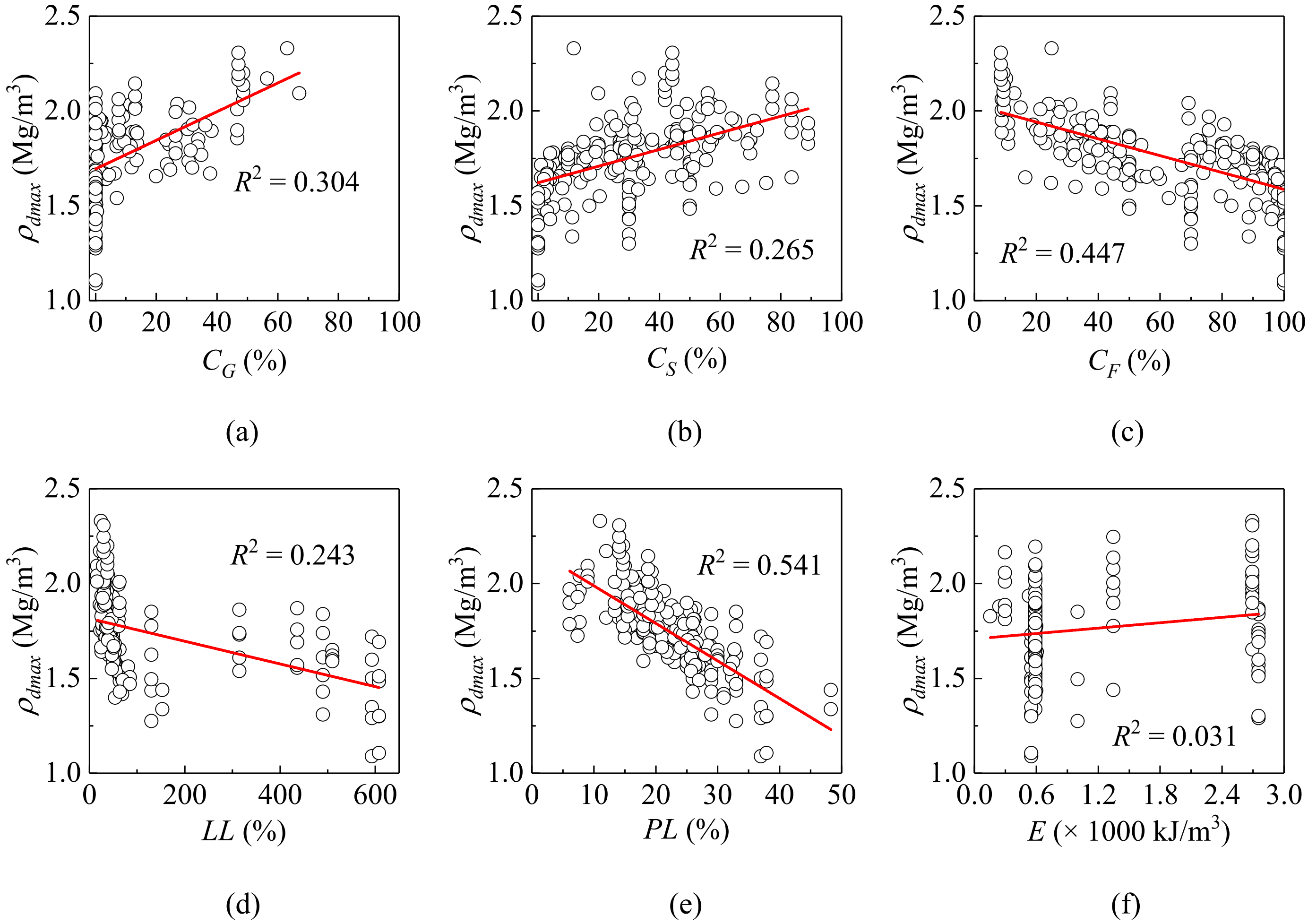


Fig. 4. Basic linear fittings between maximum dry density and each soil parameter: (a) gravel content; (b) sand content; (c) fines content; (d) liquid limit; (e) plastic limit; (f) compaction energy

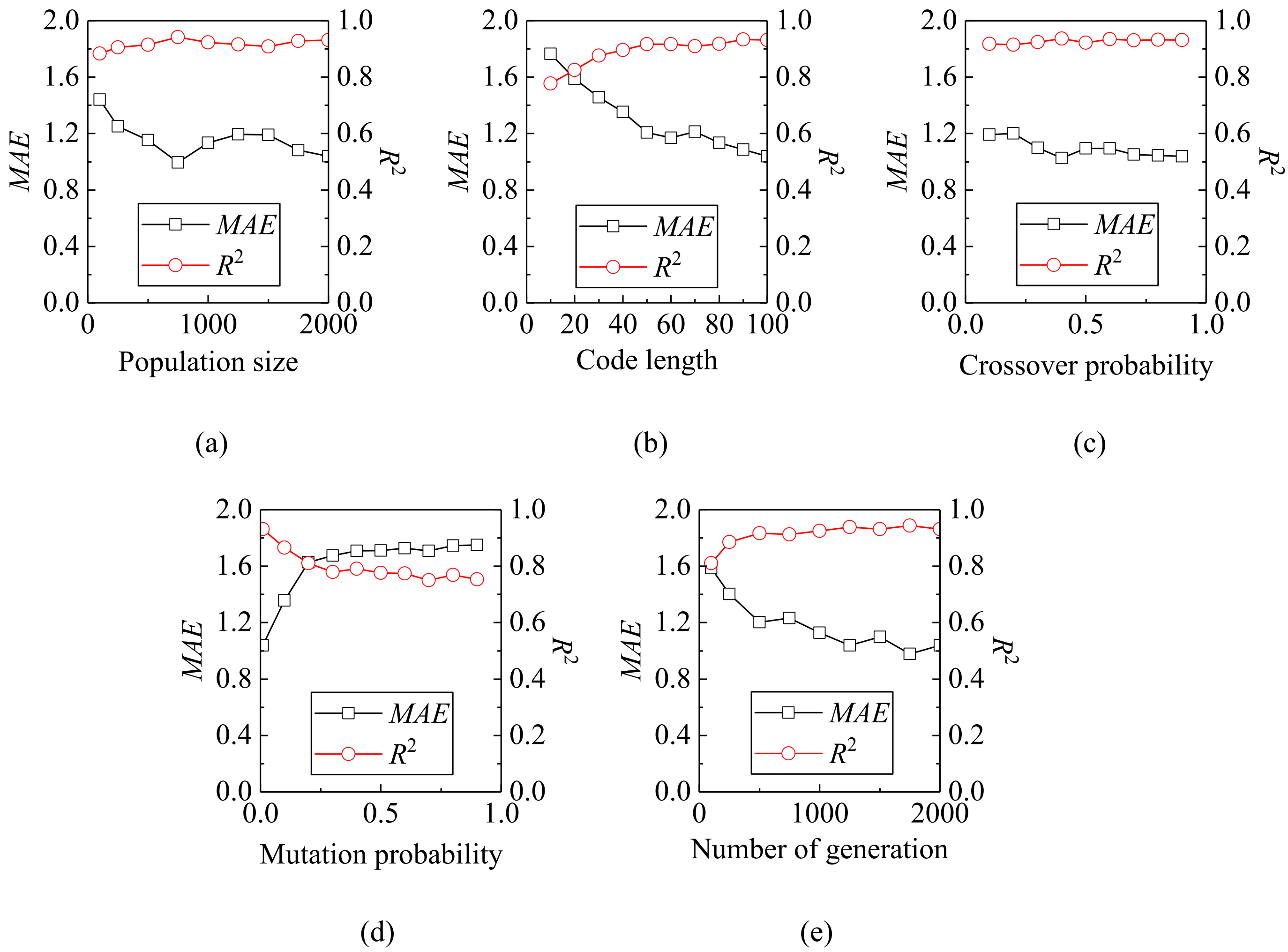


Fig. 5. Effects of code parameters on the prediction accuracy of optimum water content for the training data: (a) population size; (b) code length; (c) crossover probability; (d) mutation probability; (e) number of generation

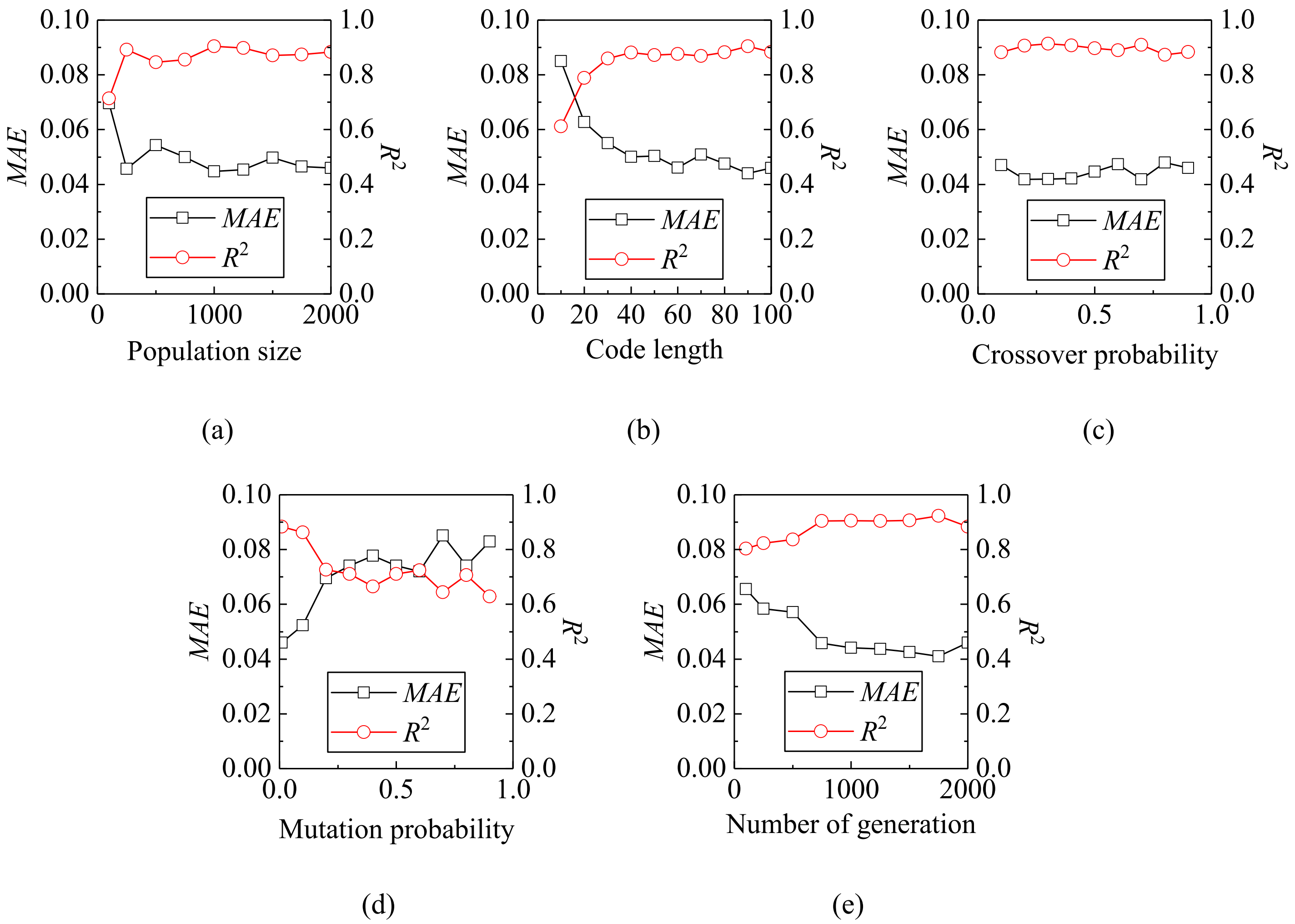
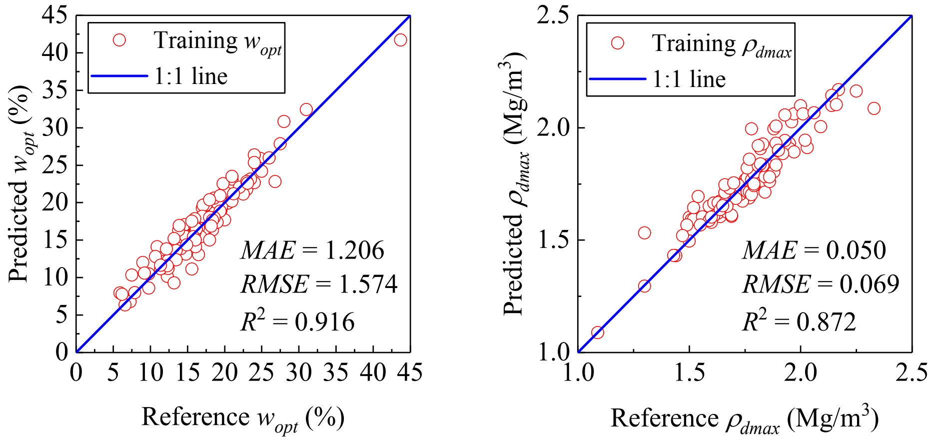
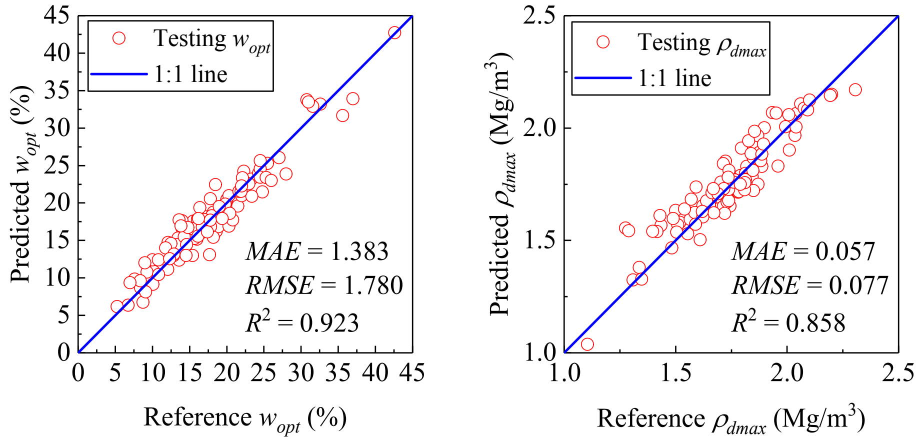


Fig. 6. Effects of code parameters on the prediction accuracy of maximum dry density for the training data: (a) population size; (b) code length; (c) crossover probability; (d) mutation probability; (e) number of generation



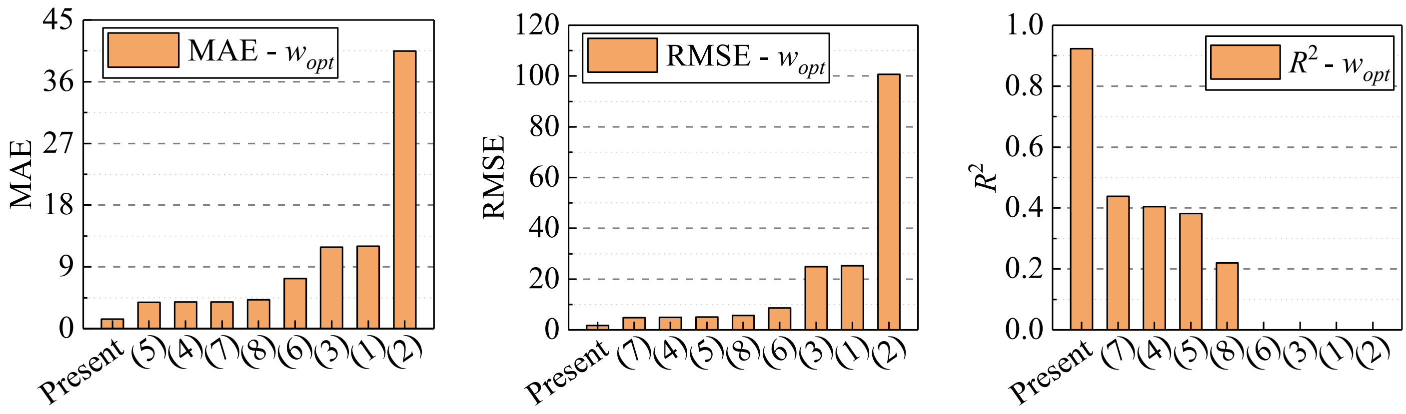
(a) (b)

Fig. 7. Comparison of MEP-predicted and reference output variables for the training data: (a) optimum water content; (b) maximum dry density



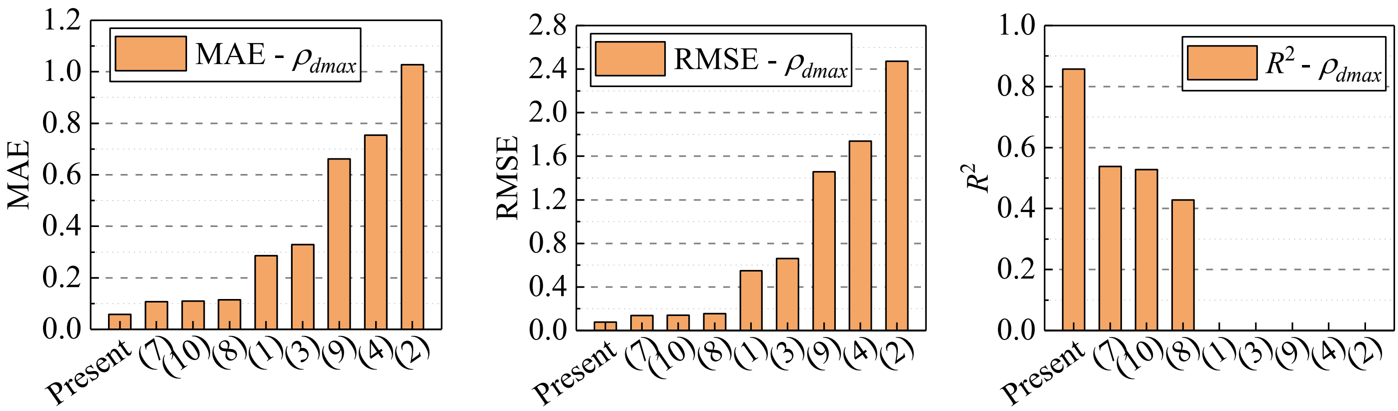
(a) (b)

Fig. 8. Comparison of MEP-predicted and reference output variables for the testing data: (a) optimum water content; (b) maximum dry density



(a) (b) (c)

Fig. 9. Comparison of the prediction accuracy of optimum water content between the present and previous studies (Note: The coefficients of determination R^2 for the references of Al-Khafaji (1993), Blotz et al. (1998), Farooq et al. (2016), and Ito and Komine (2008) are not applicable (being negative) for the testing data.)



(a) (b) (c)

Fig. 10. Comparison of the prediction accuracy of the maximum dry density between the present and previous studies (Note: The coefficients of determination R^2 for the references of Al-Khafaji (1993), Blotz et al. (1998), Farooq et al. (2016), Günaydın (2009) and Al-Khafaji (1993)-2 are not applicable (being negative) for the testing data.)

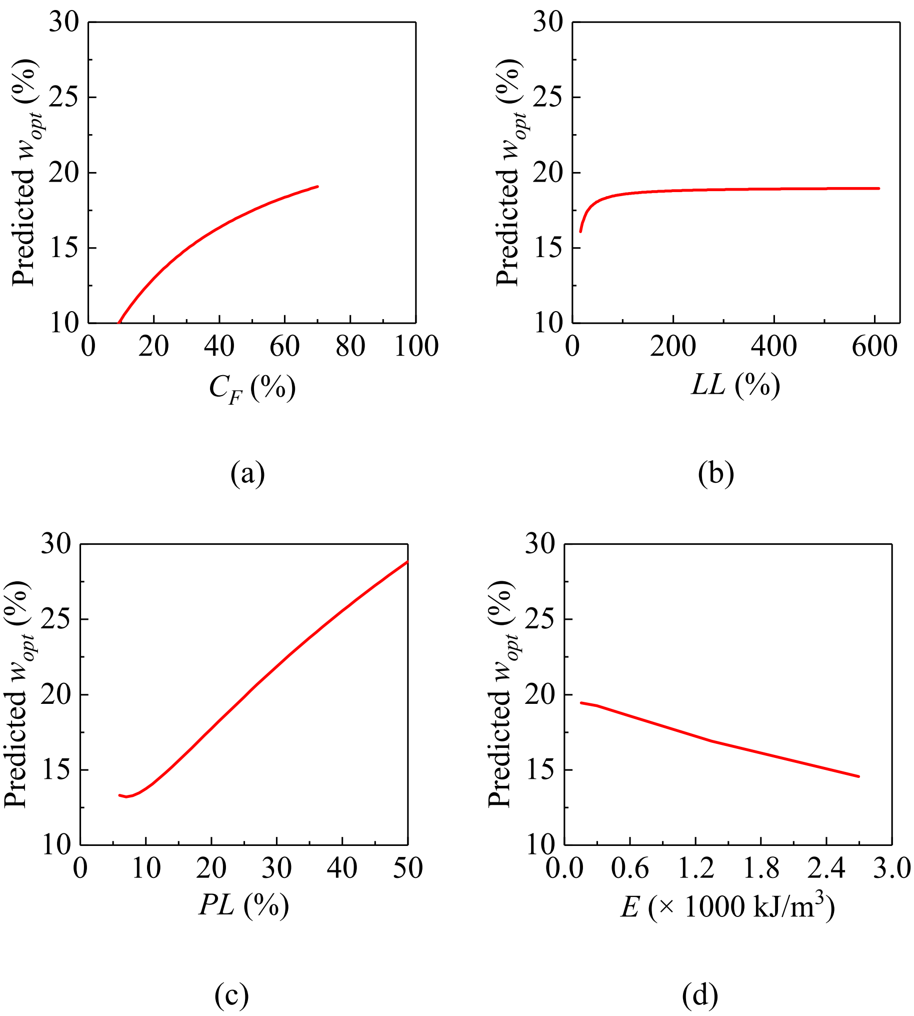


Fig. 11. Monotonicity analysis of the predicted optimum water content versus (a) fines content; (b) liquid limit; (c) plastic limit; (d) compaction energy

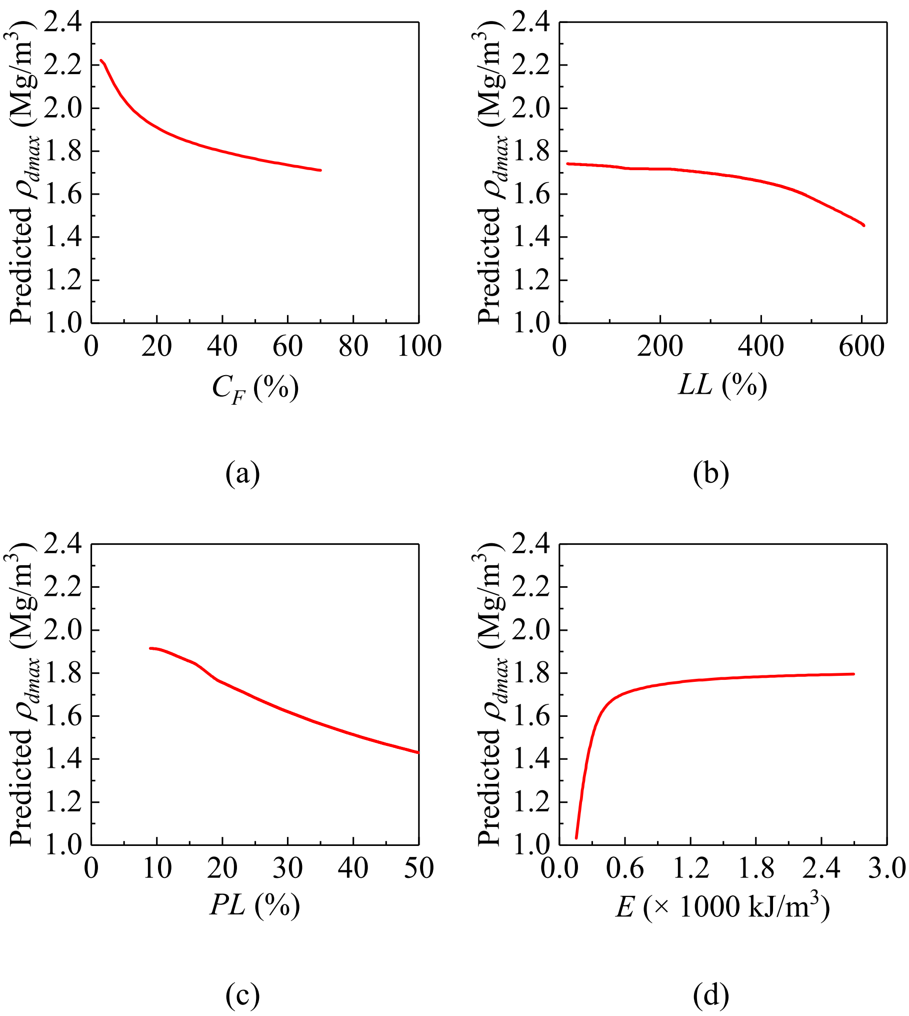
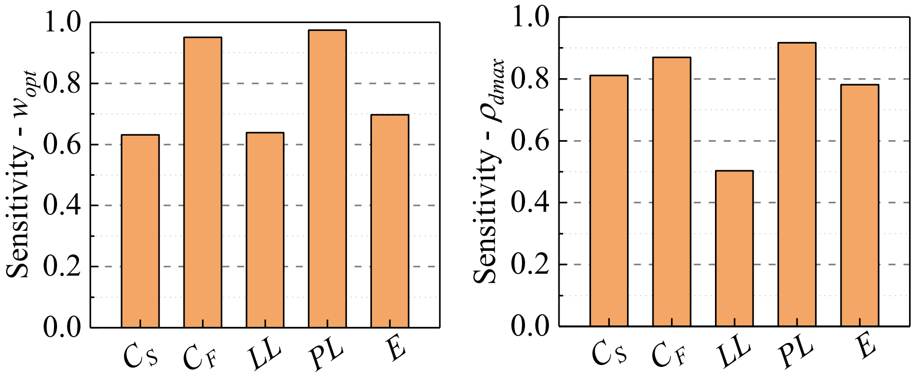


Fig. 12. Monotonicity analysis of the predicted maximum dry density versus (a) fines content; (b) liquid limit; (c) plastic limit; (d) compaction energy



(a) (b)

Fig. 13. Sensitivity analysis about the effect of the input variables on the output variables: (a) optimum water content; (b) maximum dry density