https://doi.org/10.1016/j.jclepro.2021.128205

Unconfined compressive strength of bio-cemented sand: state-of-the-art review and MEP-MC-based model development

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, ORCID: https://orcid.org/0000-0003-4154-7304

Abstract: This study presents a state-of-the-art review on the unconfined compressive strength (UCS) of bio-cemented sand treated by microbially induced calcite precipitation (MICP), followed by the high-performance prediction using a typical machine learning algorithm. First, various influencing parameters affecting the UCS of bio-cemented sand are identified, such as initial relative density, angularity of particle shape, bacterial concentration, precipitated calcium carbonate content, temperature and degree of saturation. Besides, the particle size distribution, urea and calcium concentration, and initial pH level also influence the UCS of the bio-cemented sand, but the effects remain contradictory or unclear. Following the state-of-the-art review, a big database covering 351 bio-cemented sand samples is developed, with the UCS as the output and seven influencing parameters (median grain size, coefficient of uniformity, initial void ratio, optical density of bacterial suspension, urea concentration, calcium concentration and precipitated calcium carbonate content) as inputs for the correlation. The multi expression programming (MEP) method combined with the Monte-Carlo (MC) method is proposed to develop the prediction models. All data groups randomly generated from the database are with 80% of the samples as the training sets and 20% as the testing sets. Finally, the optimal prediction model is selected with the lowest mean absolute error and based on analyses of monotonicity, sensitivity and robustness regarding more general applications.

Keywords: Unconfined compressive strength; bio-cemented sand; microbially induced calcite precipitation (MICP); multi expression programming (MEP); Monte-Carlo (MC) method

1. Introduction

Due to the growing population and urbanization all over the world, the expansion of civil infrastructure is imperative to meet the basic societal needs. During the infrastructure development, the construction is often limited by foundations with incompetent soil conditions. In this way, the soil with poor engineering properties must be improved accordingly. Generally, there are two categories of conventional soil improvement approaches: physical technology and chemical grouting (Mujah et al. 2017). The physical technology includes surcharge loading, pile installation, deep compaction (Wang and Chen 2019; Wang et al. 2019a, 2019b; Zhao et al. 2019; Chen et al. 2020), etc., normally consuming substantial energy resources, and causing soil disturbance and noise in the vicinity. The chemical grouting comprises of the treatment with lime, cement, fly ash (Yin and Zhou 2009; Yuan et al. 2016; Shen et al. 2017; Chen et al. 2019b), etc., which is chemically toxic and environmentally contaminating. Considering the limitations of the current technologies used for soil improvement, a new sustainable, efficient, economic and environmentally friendly approach is needed.

In the past two decades, a novel sustainable and economic bio-treatment method, microbially induced calcite precipitation (MICP), has been proposed and widely used on sand, based on the application of urease-active bacteria and cementitious solution (Mitchell and Santamarina 2005; DeJong et al. 2006; Whiffin et al. 2007; Ivanov and Chu 2008; Chu et al. 2012; DeJong et al. 2013; Montoya et al. 2013; Gomez et al. 2017; Mujah et al. 2017; Wang et al. 2017; Jiang et al. 2020; Liu et al. 2020; Tang et al. 2020). After the MICP treatment, the

unconfined compressive strength (UCS) is the most commonly used indicator to evaluate the strength of the bio-cemented soil. To date, the UCS of the bio-cemented sand by MICP has been studied mainly by the laboratory test (for example, van Paassen et al. 2010; Cheng et al. 2013; Chu et al. 2014). Although numerous tests have been performed, the prediction on the UCS of the bio-cemented sand is still limited to some linear or curvilinear fittings, simply considering the influencing parameter of the calcium carbonate content (Cheng et al. 2013, 2014; Chu et al. 2014; Soon et al. 2014; Zhao et al. 2014; Cheng et al. 2017; Mujah et al. 2019; Wen et al. 2019; Wang et al. 2020b). However, the UCS of the MICP-treated sand was affected by other influencing parameters such as soil property, bacterial concentration, cementitious solution concentration (Ivanov and Chu 2008; DeJong et al. 2013; Mujah et al. 2017; Wang et al. 2017; Jiang et al. 2020; Tang et al. 2020), etc. Hence, to identify influencing factors with their importance to the UCS based on all previous studies should be helpful for the engineering practice of such method, and worth carrying out.

Moreover, to better promote the application of the bio-treatment method, a more comprehensive model with explicit expressions considering a big database and various influencing parameters should be developed. To fulfil the purpose of solving nonlinear and complex problems with a big database, a genetic programming method namely multi expression programming (MEP) has been verified to be an appropriate alternative. This machine learning algorithm was firstly proposed by Oltean and Dumitrescu (2002), encoding multi programs into a single linear chromosome and aiming at generating the best encoded solution through the calculation process. Due to the advantages of high predicting precision,

high efficiency and easy implementation, the MEP method has been widely used in the field of geotechnical engineering for the last two decades, such as the prediction of compressive and tensile strengths (Baykasoğlu et al. 2008), cyclic soil response (Shahnazari et al. 2010), soil liquefaction (Alavi and Gandomi 2012), soil compaction characteristics (Wang and Yin 2020), air-entry value of compacted soils (Wang et al. 2020a), etc. Therefore, it is feasible to use the MEP approach to predict the UCS of the MICP-treated sand, as supported by the previous successful applications of the MEP approach in geotechnical engineering. In addition, to avoid local minimum problem, the Monte-Carlo (MC) method is worth being applied to the MEP approach to obtain a more reliable optimal prediction model.

In this study, the state-of-the-art review on the UCS of the bio-cemented sand is first conducted with the discussion of various influencing parameters. Then, a big database of samples is developed by collecting the UCS of the bio-cemented sand and the relevant influencing parameters. The MEP approach combined with the MC method is proposed to develop the prediction models using data groups randomly generated with each group encompassing 80% of the samples as the training data and 20% as the testing data. Finally, the optimal prediction model is selected with the lowest mean absolute error and based on analyses of monotonicity, sensitivity and robustness regarding more general applications.

2. UCS of bio-cemented soil: State-of-the-art review

In this study, the bio-treatment specifically indicates the approach of microbially induced calcite precipitation (MICP). The bio-chemical reactions of the MICP process can be

expressed as:

$$CO(NH_2)_2 + 2H_2O \xrightarrow{\text{urease}} 2NH_4^+ + CO_3^{-2} \quad (1)$$

$$Ca^{2+} + CO_3^{-2} \rightarrow CaCO_3 \quad (2)$$

During the MICP process, the urea hydrolysis occurs under the catalysing effect of the urease produced by the urease-active bacteria [Eq. (1)]. Then, the crystal bonding compound of calcium carbonate develops in combination of the Ca^{2+} in the solution and the produced CO_3^{-2} [Eq. (2)], bonding loose soil particles and increasing soil strength and stiffness. For the MICP treatment, there are mainly two ways of introducing the bacteria to soil: injection and premixing method (Mujah et al. 2017). During the injection method, the bacterial suspension is flushed through the compacted soil column, followed by a specific retention period to ensure the bacteria is attached to the sand particles before injecting the cementitious solution. In the premixing method, the bacterial suspension is mixed with the soil prior to the introduction of the cementitious solution. After both treatment methods, a relatively stiff sand column can be obtained for the UCS test. In the literature, several influencing factors were reported to affect the UCS value of the bio-cemented, as follows.

2.1. Soil properties

During the MICP treatment process, the precipitated calcium carbonate crystals locate in the pores between the soil particles, further serving as the bonding bridge to connect the adjacent particles and thus increasing the strength and stiffness. As the matrix material, the effect of various soil properties on the MICP treatment efficiency was investigated, including mineral composition, particle size, gradation, particle shape, initial void ratio, etc.

Among the aforementioned influencing parameters, the effects of particle size, gradation, particle shape and relative density on the UCS of bio-cemented sand were reported, while the relevant investigation on the effect of mineral composition remained scarce. Because the pores between soil particles serve as the flow channel for the bacteria (typical unicellular size around 1-3 μm, Ivanov and Chu 2008) and cementitious solutions, the soil particle size is important for the MICP treatment. As stated by Mitchell and Santamarina (2005), the preferable soil particle size ranges from 50 μm to 10 mm for the MICP process. Zhao et al. (2014) used the MICP approach to treat two kinds of sand with different median grain sizes d_{50} as 460 μm and 330 μm. The UCS testing results on the two bio-cemented sands indicated the UCS was higher for the sample with a bigger d_{50}. This observation was also validated by Amarakoon and Kawasaki (2016) for the measurement of the UCS of two bio-cemented sands (d_{50} as 600 μm and 200 μm). Nevertheless, Cui et al. (2016) stated a contradictory correlation between the UCS and the soil particle size: the bigger the particle size, the lower the UCS value. Regarding the gradation, Mahawish et al. (2018) demonstrated that the gap-graded particle size distribution led to a higher UCS of the bio-cemented soil. However, in this study, the d_{50} of each soil was also different, which might contribute to the difference of the testing UCS. Using the mixture of round and angular glass beads as the matrix, Xiao et al. (2019b) measured the UCS of bio-cemented mixture at different proportions of the matrix materials. For a given cementitious solution, the UCS was higher for the material with more angular particles.

The initial state (relative density or void ratio) is another significant influencing factor on

the UCS of the bio-cemented sand. According to the testing result on the UCS of bio-cemented silica sand at various relative densities by Al Qabany and Soga (2013), for a given calcium carbonate content, the UCS showed a higher value when the relative density was higher. Similar observation was identified by Rowshanbakht et al. (2016), by performing UCS test on the bio-cemented silica sand with various relative densities. They attributed the enhancement of UCS by increasing the relative density to more grain-to-grain contacts and thus better cementation.

2.2. Bacterial concentration

As the catalyst, the urease-active bacteria are important for the MICP treatment. On the one hand, the bacterial cells with highly active urease enzymes hydrolyse urea within the microbe, producing a big amount of CO_3^{2-} for the formation of calcium carbonate. On the other hand, because of the higher concentration of the nutrient close to the surface, the negatively charged bacterial cell is attached to the sand particle, acting as the nucleation site for the bonding crystals (DeJong et al. 2010). As a result, the bacterial concentration significantly affects the cementation effect of sand columns.

Zhao et al. (2014) treated one silica sand with the *Sporosarcina pasteurii* strain solution with optical density of the bacterial suspension OD_{600} (reflecting the bacterial concentration) equalling to 0.3, 0.6, 0.9, 1.2 and 1.5. As the OD_{600} was raised from 0.3 to 1.5, the UCS of the bio-cemented sand increased from 0.34-0.53 MPa to 2.01-2.22 MPa. Mujah et al. (2019) injected different concentrations of *Sporosarcina pasteurii* strain suspension for the MICP treatment of a silica sand, with the optical density OD_{600} as 1.24, 2.36 and 4.46. At a given

calcium carbonate content, the bio-cemented sample exhibited a higher UCS when treated by a higher optical density of the bacteria. Except for the bio-cemented sand, Soon et al. (2014) used the *Bacillus megaterium* strain suspension for the MICP treatment on a residual soil at various cell concentrations. The UCS testing results indicated that for the cases of 0.25 mol/L and 0.5 mol/L cementation reagent, when the cell concentration rose from 1×10^6 cfu/ml to 1×10^8 cfu/ml, the UCS of the bio-cemented soil increased at the amplitude from 25%-26% to 57-69%. Sharma and Ramkrishnan (2016) and Al-Salloum et al. (2017) used the *Sporosarcina pasteurii* strain suspension to treat fine-grained soil and cement mortar with various bacterial cell concentrations, respectively. They both drew consistent conclusion that the higher cell concentration led to a higher UCS of the bio-cemented material.

2.3. Cementitious solution concentration

For the MICP treatment, the urea and the calcium chloride mixture normally serve as the cementitious solution. The production of CO_3^{2-} is traced from the hydrolysis of the urea [Eq. (1)]. The precipitation of the bonding crystal needs the reaction between both Ca^{2+} and CO_3^{2-} [Eq. (2)]. Thus, the concentration and the proportion of each reagent significantly influences the treatment efficiency and the UCS of the bio-cemented soil.

Al Qabany and Soga (2013) treated a silica sand with the equimolar amount of urea and calcium chloride at various concentrations. The UCS testing results showed that the 1 mol/L cementitious solution produced most of the null strength samples. At a given precipitated calcium carbonate content, the sample treated by lower cementation concentration led to a higher UCS. Soon et al. (2014) used equimolar concentration of urea and calcium chloride to

treat a residual soil. It was indicated that with 1.0 mol/L cementation reagent, no measurable improvement of UCS was identified, while the cases of 0.25 mol/L and 0.5 mol/L cementation reagent contributed similar improvement of the UCS. Mujah et al. (2019) also used equimolar amount of urea and calcium carbonate for the cementitious solution to the MICP treatment of a silica sand. At a given calcium carbonate content, the sample with the lower cementation concentration possessed a higher UCS. However, opposite observation about the effect of cementitious solution concentration was identified by Zhao et al. (2014) and Wen et al. (2019), using the treatment of stirring the silica sand sample in the batch reactor filled by the cementation media. They stated that the sample with a higher cementation concentration induced a higher UCS.

Except for the MICP treatment with equimolar amount of urea and calcium chloride, different proportions of urea and calcium chloride were also used in some studies with a fixed ratio (Nafisi et al. 2020; Wang et al. 2020b). Nevertheless, the effect of the proportion of urea and calcium chloride on the UCS of the bio-cemented soil was not investigated.

2.4. Calcium carbonate content

Calcium carbonate is the production of the MICP treatment process for the porous materials. Based on the scanning electron microscope (SEM) observation, DeJong et al. (2010) revealed that the calcium carbonate precipitated mainly between the states of “uniform (calcium carbonate precipitated at an equal thickness around each soil particle, inducing lower smaller bonding)” and “preferential (calcium carbonate precipitated at particle-particle contact, leading to higher strength)”. Owing to the precipitated calcium carbonate at the

particle-particle contact, the strength and the stiffness of the bio-cemented sand are efficiently raised.

In the literature, different soils were treated by the MICP approach through various treatment methods. However, regardless of different investigations, all the previous studies drew a consistent conclusion that for a given condition, the UCS of the bio-cemented soil shows a higher value at a higher calcium carbonate content (van Paassen et al. 2010; Al Qabany and Soga 2013; Cheng et al. 2013, 2014; Chu et al. 2014; Soon et al. 2014; Zhao et al. 2014; Choi et al. 2016; Cheng et al. 2017; Li et al. 2018; Mahawish et al. 2018, 2019; Mujah et al. 2019; Wen et al. 2019; Xiao et al. 2019a; Nafisi et al. 2020; Wang et al. 2020b; Zhao et al. 2020).

2.5. Temperature

During the MICP treatment, the temperature influences the urease activity of the bacteria, thus affecting the formation of the bonding crystals and the UCS. Cheng et al. (2014b) performed the UCS test for the bio-cemented silica sand samples at two temperatures of 25 ℃ and 50 ℃. Although three times of calcium carbonate content were identified for the sample treated under 50 ℃, the crystal size was relatively small (2-5 μm), simply covering the particle surface. By contrast, the crystal produced under 25 ℃ was bigger (15-20 μm), effectively bonding the sand particles, thus inducing a higher UCS. Peng and Liu (2019) conducted the MICP treatment on a silica sand at various temperatures: 10 ℃, 15 ℃, 20 ℃, 25 ℃ and 30 ℃. Consistent results were obtained: as the temperature increased, the UCS decreased accordingly.

2.6. pH

The bacteria used in the MICP process normally favours the alkaline environment (Ferris et al. 2004). As summarised by Tang et al. (2020), the pH level affects the precipitation rate and production, morphological characteristics of the bonding crystal and bacteria adsorption. Hence, the initial pH level influences the final UCS of the bio-cemented soil. Cheng et al. (2014b) applied three pH levels (neutral, acid and alkaline) on the MICP treatment of sand columns. The highest UCS appeared for the case with neutral environment, while the calcium carbonate content increased as the pH level increased. Keykha et al. (2017) conducted the MICP treatment on a silty soil (passing 425-μm sieve) at various initial pH levels of 5, 6, 8, and 9. After the specific curing period, the UCS presented a higher value when the pH level was higher.

2.7. Degree of saturation

The degree of saturation during the MICP treatment is another influencing factor, reported to affect the distribution of calcium carbonate and thus the UCS (Cheng et al. 2013). From this study, it was indicated that for a given calcium carbonate content, the sample treated at lower degree of saturation led to a higher UCS.

2.8. Summary and database development

As mentioned previously, the UCS of the bio-cemented sand is influenced by various factors. However, due to the lack of data from the previous publications, the present study only considers the available influencing parameters: median grain size d_{50}, coefficient of uniformity C_u, initial void ratio e_0, optical density of bacterial suspension OD_{600}, urea

concentration M_u, calcium concentration M_{Ca} and calcium carbonate content F_{Ca}. Based on this criterion, totally 351 bio-cemented sand samples were collected from the literature (van Paassen et al. 2010; Al Qabany and Soga 2013; Cheng et al. 2013, 2014a; Zhao et al. 2014; Cheng et al. 2017; Mahawish et al. 2018, 2019; Mujah et al. 2019; Wen et al. 2019; Xiao et al. 2019a; Nafisi et al. 2020; Wang et al. 2020b), with the details of the influencing parameters and the UCS shown in the supplementary material. Note that to stay consistent with the other studies, only the data of the treatment under 100% degree of saturation was collected from Cheng et al. (2013). In the supplementary database, the samples are ordered alphabetically by the name of the first author.

Table 1 lists the descriptive statistics of the variables in the database, with the minimum, maximum, mean and standard deviation values for each variable. The frequency histogram of each variable is plotted in Fig. 1. The median grain size locates in the range from 0.12 mm to 1.6 mm (Table 1 and Fig. 1a), which is common for sand. The majority of the samples have the median grain size smaller than 0.3 mm. The distribution of the coefficient of uniformity is scattered, with more than 310 samples showing the value between 1 and 2 and a small number of samples possessing the value between 6 and 7 (Fig. 1b). The distribution of the initial void ratio is also not uniform, with closely 280 samples locating in the range from 0.6 to 0.7 (Fig. 1c). However, the distributions of the bacterial optical density, urea concentration, calcium concentration and calcium carbonate content are relatively uniform, with the majority showing the values between 1.6-2.4, 0.9 mol/L-1.2 mol/L, 0.9 mol/L-1.2 mol/L and 5%-10%, respectively (Figs. 1d-1g). From Table 1 and Fig. 1h, it is observed that most

samples (around 280 samples) provide the UCS from 0.05 MPa to 3 MPa, while the highest UCS shows the value of 14.23 MPa.

To understand the overall monotonic variation trend between the UCS and each input variable, Fig. 2 depicts the basic linear fittings. In this figure, the coefficient of determination R^2 is applied to assess the fitting precision as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (h_i - t_i)^2}{\sum_{i=1}^{n} (h_i - \overline{h_i})^2} \tag{3}$$

where h_i and t_i are the actual and the predicted UCS values for the ith output; $\overline{h_i}$ is the average of the actual UCS values. The R^2 ranges between 0 and 1, showing higher prediction precision with the R^2 value closer to 1.

It is observed from Fig. 2 that the best linear fitting appears for the relationship between the UCS and the calcium carbonate content, with the R^2 equal to 0.551 (Fig. 2g). When the calcium carbonate content increases, the UCS increases accordingly. The fittings for the input variables of median grain size (Fig. 2a) and coefficient of uniformity (Fig. 2b) present a relatively higher prediction accuracy than the remaining ones. With the increase of median grain size or the decrease of coefficient of uniformity, an increasing trend is identified. In terms of the initial void ratio (Fig. 2c), the optical density of bacterial suspension (Fig. 2d), the urea concentration (Fig. 2e) and the calcium concentration (Fig. 2f), the fittings are poor, with the R^2 lower than 0.01. This indicates that the monotonicity between the UCS and each of these four input variables is also poor, although the fittings all show a slightly increasing trend as each input variable increases. On the contrary, for the variation trends related to these

four influencing parameters, a peak value is identified in the middle, with the values of initial void ratio, the optical density of bacterial suspension, the urea concentration and the calcium concentration at around 0.66, 3.25, 1.00 mol/L and 1.00 mol/L, respectively. In general, the basic linear fittings in Fig. 2 are not capable of predicting the UCS separately as indicated by the low R^2 value no more than 0.551. Hence, a more comprehensive model is needed to connect the UCS and all the influencing parameters to improve the prediction accuracy.

3. Model development

3.1. Multi expression programming

To address the complicated nonlinear problem of this case, the MEP method is adopted. A MEP chromosome can encode multiple solutions (programs), starting by the creation of a random population of computer programs. Each gene is constituted by the terminal (element in the terminal set) and/or the function symbol (element in the function set). To ensure the appropriate generation of the chromosome, the first gene of a chromosome must be a terminal randomly chosen from the terminal set. The following genes with a function have a pointer towards the function arguments, presenting the expression indices with lower values than the position of this gene in the chromosome.

Before the MEP calculation, various code parameters need to be set. The population size defines the number of programs in the population. The number of generation is the number of calculations before a program runs to the end. The code length is defined as the number of genes, which is constant during the calculation. The crossover and the mutation

probability are the probability of an offspring subjected to the operators of crossover and mutation, respectively. If the crossover type is set as uniform, the offspring genes are randomly taken from one parent to another. The replication number is the number of developed chromosomes after the calculation. During the calculation, the best expression is evolved by repeating the following steps until the termination condition is obtained (Oltean and Grosan 2003): (i) selecting two parents by a procedure of binary tournament, then recombining them with a fixed crossover probability; (ii) obtaining two offspring from the previous step; (iii) mutating the offspring, if the offspring is better than the worst one, replacing the worst individual in the current population with the best of them.

An example of an MEP chromosome is demonstrated below:

$P_0 : x_1$

$P_1 : x_2$

$P_2 : + P_0, P_1$

$P_3 : x_3$

$P_4 : x_4$

$P_5 : pow\ P_3, P_4$

$P_6 : \times P_2, P_5$

The terminal and the function sets in this example are $\{x_1, x_2, x_3, x_4\}$ and $\{+, pow, \times\}$, respectively. By reading from top to bottom, the MEP genes can be converted into computer programs. Fig. 3 shows the corresponding gene trees for this example. The genes 0, 1, 3 and 4 represent a single terminal as: $P_0 = x_1$, $P_1 = x_2$, $P_3 = x_3$, $P_4 = x_4$, respectively. Gene 2

indicates the operation $+$ on the operands at the locations of the genes 0 and 1, defining the expression as $P_2 = x_1 + x_2$. Similarly, gene 5 indicates the operation pow on the operands at the genes 3 and 4, determining the expression as $P_5 = x_3^{x_4}$. By indicating the operation $\times$ on the operands at the genes 2 and 5, gene 6 is expressed as $P_6 = (x_1 + x_2)x_3^{x_4}$. Each of these expressions can be considered as a possible solution for the problem. The best expression is chosen to represent the chromosome, after controlling the fitness value f of all expressions as (Oltean and Grosan 2003):

$$f = \min_{i=1,k} \left\{ \sum_{j=1}^{L} \left| F_j - O_j^i \right| \right\} \tag{4}$$

where L is the number of fitness cases; k is the number of chromosome genes; F_j is the expected value for the fitness case j; O_j^i is the value returned (for the fitness case j) by the i th expression in the current chromosome.

3.2. Development of MEP-MC-based models

To strengthen the robustness of the prediction model for the UCS of the bio-cemented sand, the Monte-Carlo (MC) method is adopted as follows. To form the database for training and testing, the 80% of whole samples (280 samples) are randomly selected as the training data, while the remaining 20% of the samples (71 samples) are the testing data. To complete the Monte-Carlo analysis, hundreds of the processes of data generation are repeated. For each data group, a corresponding optimal model with evaluated prediction performance is obtained via the machine learning algorithm. According to the results of prediction performance, a final model ranking for hundreds-trained models is achieved.

For each data group, the training and testing data are submitted to the MEP. Various

hyper-parameters are set, following the optimal values used by Wang and Yin (2020) and Wang et al. (2020a). Table 2 lists the details of the hyper-parameter setting. For each case, four function sets are considered, thus inducing totally 1,200 times of calculations in this study. Through the MEP calculation, an optimal solution is obtained for the training data of each group, with the lowest MAE value (the optimal function set is determined accordingly), defined as:

$$MAE = \frac{\sum_{i=1}^{n} |h_i - t_i|}{n} \tag{5}$$

where n is the number of the output variables. The lower MAE indicates the higher precision of the prediction model.

From the calculation process in this study, it has been shown that 300 data groups are enough. Fig. 4 (a) plots the variation of the optimal MAE value for the training data of each data group. It can be observed that for the randomly generated data groups, the optimal MAE of the training data locates in a narrow range from 0.346 to 0.524. Using the optimal function set for each data group, the corresponding MAEs of the testing data are also obtained using the source code, as shown in Fig. 4 (b). Compared to the training data, the variation of the testing MAE is more scattered, ranging from 0.325 to 0.675.

To evaluate the performance of the optimal model developed by each data group on the whole database, the overall MAE_{All} is determined as:

$$MAE_{All} = \frac{\sum_{i=1}^{n_{Train}+n_{Test}} |h_i - t_i|}{n_{Train} + n_{Test}} = \frac{MAE_{Train} n_{Train} + MAE_{Test} n_{Test}}{n_{Train} + n_{Test}} \tag{6}$$

where n_{Train} and n_{Test} are the number of outputs in the training (280 samples) and testing (71 samples) data, respectively; MAE_{Train} and MAE_{Test} are the MAE of the training and testing data, respectively. Fig. 5 plots the variation of the overall MAE_{All} of the random data groups, ranging in a narrow range from 0.380 to 0.546. In particular, the overall MAE_{All} of numerous data groups locates close to the lowest value of 0.380. This convergence trend indicates that 300 data groups are sufficient to develop an optimal model for predicting the UCS of the bio-cemented sand.

3.3. Characteristics of prediction models

To demonstrate the model performance, five data groups with the lowest overall MAE_{All} (0.380, 0.382, 0.384, 0.388 and 0.389) are selected. These best five data groups are denominated as Group A, B, C, D and E following the order of the group number, as shown in Fig. 5. The optimal solution codes for the five data groups can be referred to the supplementary material.

By reading the optimal solution code, the explicit expressions for UCS of the bio-cemented sand in Group A are derived as:

$$UCS = \left[A_4 + \frac{(A_1 - e_0)(A_1 + A_4)}{A_2^{e_0}(F_{Ca} - 1)} \right] \frac{A_6}{A_7} \tag{7}$$

where A_1, A_2, A_3, A_4, A_5, A_6 and A_7 are parameters as:

$$A_1 = d_{50}^{\ d_{50}} \tag{8}$$

$$A_2 = (A_1 - d_{50})d \tag{9}$$

$$A_3 = M_{Ca}^{\ (OD_{600} - d_{50} F_{Ca} A_1 A_2)} \tag{10}$$

$$A_4 = F_{Ca} - 2e_0 + d_{50} A_1 A_2 \tag{11}$$

$$A_5 = A_2^{e_0}\left(\frac{A_3}{d_{50}A_1A_2} + OD_{600}^{\,2}\right) \tag{12}$$

$$A_6 = \frac{A_2^{e_0}(F_{Ca}-1)}{(A_1+A_4)(2e_0)^{A_1A_2}A_3^{A_5(A_1-e_0)}} \tag{13}$$

$$A_7 = \frac{2e_0 + A_4 + \dfrac{C_u}{A_2^{e_0}(F_{Ca}-1)}}{A_1+A_4} \tag{14}$$

The best explicit expressions for the UCS of bio-cemented sand of Group B training data are determined as:

$$UCS = B_2 + 2(M_u - M_{Ca})^{B_5} - B_6 \tag{15}$$

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

$$B_1 = OD_{600} - \frac{d_{50}}{M_u} \tag{16}$$

$$B_2 = \frac{F_{Ca}}{4C_u - d_{50}^{\,2C_u}} - e_0^{\left(\frac{d_{50}}{M_u}+2C_u\right)} \tag{17}$$

$$B_3 = M_u^{\frac{d_{50}}{M_u}} \tag{18}$$

$$B_4 = \frac{F_{Ca}B_1}{4C_u - d_{50}^{\,2C_u}}\left(\frac{d_{50}}{M_u}\right)^{F_{Ca}B_1} \tag{19}$$

$$B_5 = \frac{1}{C_u^{\,B_3}}\left(\frac{d_{50}}{M_u}\right)^{(4C_uF_{Ca}B_1-B_3)} \tag{20}$$

$$B_6 = \left(2B_4 - M_{Ca}B_1 - \frac{B_4}{B_1}\right)\left(B_3^{\,\frac{d_{50}C_u}{M_u}} - M_u\right) \tag{21}$$

The optimal expressions for the UCS of bio-cemented sand of Group C training data are listed as:

$$UCS = \frac{(M_u - M_{Ca})^{C_2}}{d_{50}^{\,d_{50}}} + \frac{C_3}{(M_u - d_{50}^{\,d_{50}})C_6} + C_7 \tag{22}$$

where C_1, C_2, C_3, C_4, C_5, C_6 and C_7 are parameters as:

$$C_1 = \lg F_{Ca} \tag{23}$$

$$C_2 = d_{50}^{F_{Ca}^2} \tag{24}$$

$$C_3 = \frac{d_{50} C_1^4}{d_{50} + F_{Ca}} \tag{25}$$

$$C_4 = d_{50} C_1^4 + d_{50}^{d_{50}} C_1^2 - \frac{d_{50} C_1^2 + M_{Ca} C_3}{C_1 + C_2 + d_{50}} \tag{26}$$

$$C_5 = \frac{C_3 \lg(d_{50} C_1^4 + d_{50}^{d_{50}} C_1^2)}{M_u - d_{50}^{d_{50}}} \tag{27}$$

$$C_6 = \frac{OD_{600}}{C_1^2 - C_3 - d_{50}} + C_5 \tag{28}$$

$$C_7 = C_4 e_0^{\lg(M_{Ca} C_3)} \tag{29}$$

The best explicit expressions for the UCS of bio-cemented sand of Group D training data are listed as:

$$UCS = D_7 D_8 + \frac{(D_8 - D_5)^2}{D_6} - \frac{[(M_{Ca} - 1) OD_{600} D_3]^2}{d_{50} - M_{Ca} - D_2} \tag{30}$$

where D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8 and D_9 are parameters as:

$$D_1 = d_{50}^{d_{50}} \tag{31}$$

$$D_2 = 4(D_1 - d_{50}) d_{50} + 2 e_0^{F_{Ca}} \tag{32}$$

$$D_3 = \left[\frac{(D_1 - d_{50}) d_{50} D_1}{D_2} \right]^{D_2} \tag{33}$$

$$D_4 = [(D_1 - d_{50})(D_2 + M_{Ca})]^{D_9} \tag{34}$$

$$D_5 = \frac{(D_1 - d_{50}) d_{50} D_1 D_9}{D_2} \tag{35}$$

$$D_6 = (D_1 - d_{50})^2 + M_u - M_{Ca} \tag{36}$$

$$D_7 = [(D_1 - d_{50})(D_2 + M_{Ca})]^{D_6^{D_8}} \tag{37}$$

$$D_8 = (M_u - M_{Ca})^{D_4} + D_5 \tag{38}$$

$$D_9 = D_3 + OD_{600} + F_{Ca} - OD_{600} M_{Ca} \tag{39}$$

The optimal explicit expressions for the UCS of bio-cemented sand of Group E training data are determined as:

$$UCS = \frac{E_7}{\lg(d_{50}^{F_{Ca}})} + \frac{(OD_{600} E_1^4)^2}{E_6} - E_3 + E_4 - E_7 \tag{40}$$

where E_1, E_2, E_3, E_4, E_5, E_6 and E_7 are parameters as:

$$E_1 = \lg F_{Ca} \tag{41}$$

$$E_2 = \left(\frac{F_{Ca}^2 d_{50}^{F_{Ca}^2}}{F_{Ca}^{F_{Ca}}} \right)^{d_{50} E_1} \tag{42}$$

$$E_3 = E_1 (M_u - M_{Ca})(E_1^4 + d_{50}) \tag{43}$$

$$E_4 = \frac{2E_3^{E_2}}{E_5} + E_1^2 + d_{50} E_1 (E_1^4 + d_{50}) \tag{44}$$

$$E_5 = (d_{50} E_1 + M_{Ca})^{M_{Ca}} \tag{45}$$

$$E_6 = \left(\frac{M_u F_{Ca}}{E_4} \right)^{2E_1(E_1^4 + d_{50})} \tag{46}$$

$$E_7 = d_{50}(E_3 + \lg E_5) \tag{47}$$

On the whole, all the five models consider the input variables of median grain size, optical density of bacterial suspension, calcium concentration and calcium carbonate content. In the models of Group C, D and E, the coefficient of uniformity is not used. In addition, the initial void ratio and the urea concentration are ignored by the model of Group E and Group A, respectively.

4. Results and discussions

4.1. Model validity

To assess the prediction accuracy of the developed models, three indicators of mean absolute error MAE [see Eq. (5)], root mean squared error $RMSE$ and coefficient of determination R^2 [see Eq. (1)] are used as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(h_i - t_i)^2}{n}} \tag{48}$$

For a given group of data, the higher value of R^2, or the lower values of MAE or $RMSE$ indicate the higher precision of the model. Fig. 6 is depicted for the comparison between the predicted and reference UCS values of the training data for the five selected groups. To quantify the prediction precision, the indicators of MAE, $RMSE$ and R^2 are presented in this figure. Due to the low MAE value for the five models during the selection process, the predicted UCS values all have a good agreement with the reference ones. The prediction accuracy all shows a relatively high level, with the R^2 higher than 0.89.

In terms of the corresponding testing data, Fig. 7 shows the comparison between the predicted and reference UCS values of the testing data using the developed models of the five groups. For Group C, the prediction accuracy of the testing data is higher than the that of the training data, showing the R^2 as 0.951. The prediction precisions of the testing data for the other four groups are lower than their respective training data. However, the R^2 values for these four models are still higher than 0.80, indicating that these four models are acceptable

regarding the prediction accuracy. In summary, the selected five models are all appropriate to predict the UCS of the bio-cemented sand by considering the prediction accuracy of the supplementary database.

4.2. Monotonicity analysis

To have a more general application of the developed models, further evaluation is needed. The monotonicity analysis is an efficient way to evaluate the feasibility of the developed models, by comparison with the original database using the variation trend between the UCS and each input variable (Wang and Yin 2020; Wang et al. 2020a). In this analysis, the examined input variable changes, while the other input variables stay constant. Note that the monotonicity sometimes depends on settings. Table 3 lists the basic setting of the input variables that is reasonable for the monotonicity analysis, corresponding to the mean value of each parameter except for the urea and calcium concentrations (see Table 1). Because the urea and calcium concentrations are manually controlled during the MICP treatment, the most common values (1 mol/L for both parameters) in the database are used for the monotonicity analysis. Using the prediction models and the basic setting of the input variable in Table 3, the relationship between the predicted UCS and the examined input variable can be obtained. Note that if a specific input variable is not considered by a model, the corresponding monotonicity analysis is not performed on the input variable for that model.

Fig. 8 presents the variation of the UCS with the median grain size, coefficient of uniformity, initial void ratio, optical density of bacterial suspension, urea concentration, calcium concentration and precipitated calcium carbonate content in the monotonicity

analysis. From Fig. 8a, it can be seen that except for the model of Group A, the UCS generally increases with the increase of median grain size, as the original database (Fig. 2a). In addition, the variation between the UCS and the median grain size is not smooth for the models of Group A, C and D. When the coefficient of uniformity increases, the predicted UCS by the models of Group A and B decreases monotonically (Fig. 8b), which is consistent with the original database (Fig. 2b). As illustrated previously for Figs. 2c-2f, the variation of the UCS with the initial void ratio, optical density of bacterial suspension, urea concentration or calcium concentration is scattered, with low fitting accuracy ($R^2 < 0.01$) for the linear monotonic relationship. Therefore, the monotonicity analysis for these four influencing parameters cannot be considered to evaluate the feasibility of the developed models, although some models fit part of variation trend for the original database (Figs. 2 and 8). The linear monotonic fitting shows a relatively high accuracy ($R^2 = 0.551$) for the variation between the UCS and the precipitated calcium carbonate content in the original database (Fig. 2g). Similar to this trend, the predicted UCS increases smoothly when the precipitated calcium carbonate increases for all the five models (Fig. 8g).

From the monotonicity analysis, the models of Group A, C and D show a poor monotonicity in terms of median grain size, while the other two models are appropriate. Regarding the coefficient of uniformity and particularly the precipitated calcium carbonate content, all the five models are reliable.

4.3. Sensitivity analysis

To assess the significance of the input variable on the predicted UCS, the sensitivity analysis

on the whole database is performed. For a specific input variable x_i, the sensitivity R_{sen} can be determined as:

$$R_{sen} = \frac{\sum_{i=1}^{N}(x_i t_i)}{\sqrt{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} t_i^2}} \tag{49}$$

where t_i is the predicted UCS using the prediction models; N is the number of samples in the supplementary database ($N = 351$). The sensitivity value R_{sen} ranges from 0 to 1, showing the related strength between predicted output variable and each input variable. The relevance between the output variable and the specific input variable is stronger, when the R_{sen} shows a value closer to 1.

Fig. 9 shows the result of the sensitivity analysis about the relevance of the input variable on the predicted UCS for the five models. If a specific input variable is ignored in a model, the corresponding sensitivity is not applicable for that case. It can be observed from Fig. 9 that for a specific input variable, the sensitivity value is close for the models of different groups. On the whole, the relevance of the precipitated calcium carbonate content on the predicted UCS is the strongest for all the five models, with the sensitivity value higher than 0.85. The relevance of the median grain size on the predicted UCS ranks the secondly significant. These observations are in consistency with the original database (Fig. 2): the fittings between the UCS and the precipitated calcium carbonate content, the median grain size have the highest precisions, suggesting that these two input variables present the strongest relevance to predict the UCS. This in turn verifies the feasibility of the developed five models in terms of the sensitivity analysis. By contrast, the relevance of the coefficient

of uniformity on the predicted UCS is the poorest in the models of Group A and B. Furthermore, this input variable is not considered in the other three models, indicating that this variable is the least important for the prediction of UCS of the bio-cemented sand. For the initial void ratio, optical density of bacterial suspension, urea and calcium concentrations, the sensitivity stays at a relatively high value of 0.6.

4.4. Robustness analysis

Robustness analysis is another key aspect to assess the ratio of the number of samples with reasonable UCS values to the total analysing samples, with an appropriate group of input variables (Jin et al. 2019; Jin and Yin 2020; Wang et al. 2020a). In this way, the robustness analysis further validates the feasibility of the prediction models. To quantify the robustness, a robustness ratio is defined as (Jin et al. 2019; Jin and Yin 2020; Wang et al. 2020a):

$$r = \frac{\text{Samples in the reasonable range}}{\text{Total testing samples}} \tag{50}$$

The samples are considered reasonable, with the predicted UCS values falling in the range from 0.05 MPa to 14.23 MPa according to the original database (Table 1). To generate the input variables, the seven variables (median grain size, coefficient of uniformity, initial void ratio, optical density of bacterial suspension, urea concentration, calcium concentration and precipitated calcium carbonate content) are assumed to be independent to each other, while obeying the lognormal distribution. Following the mean value and the standard deviation of each input variable (as shown in Table 1), 15,000 samples are randomly generated following the lognormal distribution. The generated samples with each input variable exceeding its minimum or maximum threshold are deleted. Because the urea

concentration is controlled equal to or higher than the calcium concentration in the original database, the input samples with the urea concentration smaller than the calcium concentration are also deleted. Finally, totally 4,295 samples are chosen to perform the robustness analysis.

Fig. 10 presents the number of samples locating in the specific range of predicted UCS values for the five models, along with the robustness ratio. It is observed from Fig. 10 that using the models of Group A, B and E, the majority of the samples show the predicted UCS ranging from 0.05 MPa to 3 MPa. Using the other two models, the samples mainly present the predicted UCS values from 3 MPa to 6 MPa. For all the models except for that of Group D, a small number of samples (less than 10%) show the predicted UCS from 6 MPa to 14.23 MPa. Regarding the samples with the predicted UCS smaller than 0.05 MPa beyond the reasonable range, the models of Group A, D and E have more samples than the other two models. With the predicted UCS exceeding 14.23 MPa of the reasonable range, about 9% of the samples are generated by the model of Group D. As a result, the robustness ratio for the model of Group D is the lowest as 86.71%. By contrast, the robustness ratios of the other four models are relatively high, showing the value all above 90%, indicating that these four models are reliable from the robustness analysis.

4.5. Discussions

On the whole, all five models predict the UCS of the bio-cemented sand well for both the training and the testing data (Figs. 6 and 7). From the monotonicity analysis, the models of Group A, C and D show a poor monotonicity dealing with the median grain size (Fig. 8).

Through the sensitivity analysis, the relevance of the input variable on the predicted UCS of the bio-cemented sand corresponds well with the original database for all the five models (Fig. 9). By the robustness analysis, the model of Group D is not reliable, with a low robustness ratio, while the other four models are robust for the prediction of UCS of bio-cemented sand (Fig. 10). In consideration of all the validation analysis, it can be concluded that the models of Group A, C and D have deficiencies in either validation process. Hence, the model of Group B is recommended for predicting the UCS of bio-cemented sand, with high prediction accuracy, feasible monotonicity, sensitivity and robustness, followed by the model Group E (more complicated expressions and slightly weaker robustness) and then others. It should be pointed out that the optimal result has also been revealed by selecting 10 best models with the same procedure.

There are also limitations in this study. Firstly, due to the lack of sufficient data, the influencing parameters, such as temperature, pH and degree of saturation, are not considered during the model development. Secondly, the input variable of the precipitated calcium carbonate content is an overall estimation of the tested samples. But as the previous studies stated, only the effective bonding crystals would increase the soil strength. This issue is not considered in the models, which needs to be further investigated (DeJong et al. 2010). Thirdly, the supplementary database encompasses testing samples treated by different methods. The treatment method might bring some prediction errors, while this issue is also not considered in the present models. More studies are needed to address this issue. Despite these limitations, the optimal prediction model can still be obtained to predict the UCS of the bio-cemented

sand in a feasible manner.

Besides, MEP is just one good choice in this study. Using the developed database, other excellent machine learning based methods (Jin et al. 2019; Jin and Yin 2020; Chen et al. 2019a; Zhang et al. 2020, 2021) can also be used to propose prediction models according to logic of this study.

5. Conclusions

In this study, the unconfined compressive strength (UCS) of the bio-cemented sand by microbially induced calcite precipitation (MICP) was firstly reviewed. Various influencing parameters affecting the UCS of the bio-cemented sand to different extents and with different effects were summarised. In general, higher initial relative density, higher angularity of particle shape, higher bacterial concentration, higher precipitated calcium carbonate content, lower temperature or lower degree of saturation led to a higher UCS of the bio-cemented sand. However, the effects of particle size, urea and calcium concentration, and initial pH level on the resulted UCS of the bio-cemented sand were contradictory or not clear.

Following the reported influencing parameters as the inputs and the UCS as the output, a big database covering 351 samples was developed by collecting data from the literature. The MEP approach was combined with the MC method to develop prediction models for the UCS of the bio-cemented sand. Each randomly generated data group has 80% of the samples as the training data and 20% as the testing data. After training and testing by the MEP-MC-based method, five models with the lowest overall mean absolute error were selected for further

validation by the prediction precision on the database, the analysis of monotonicity, sensitivity and robustness.

The same procedure was applied to more best models to ensure obtaining the optimal model. After various validations, the prediction model of Group B is suggested to predict the UCS of the bio-cemented sand with high prediction accuracy and reliable behaviours of monotonicity, sensitivity and robustness.

This study provides database of UCS of the bio-cemented sand and the logic of model development, which can be used for developing models with other excellent methods.

Notations

A_1, A_2, A_3, A_4, A_5, A_6, A_7 Parameters for Group A model

B_1, B_2, B_3, B_4, B_5, B_6 Parameters for Group B model

C_1, C_2, C_3, C_4, C_5, C_6, C_7 Parameters for Group C model

C_u Coefficient of uniformity

d_{50} Median grain size

D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9 Parameters for Group D model

e_0 Initial void ratio

E_1, E_2, E_3, E_4, E_5, E_6, E_7 Parameters for Group E model

f Fitness value of MEP expressions

F_{Ca} Calcium carbonate content

F_j Expected value for fitness case j

h_i Actual output variable

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

k Number of chromosome genes

L Number of fitness cases

M_{Ca} Calcium concentration

M_u Urea concentration

MAE Mean absolute error

MAE_{All} Mean absolute error of whole database

MAE_{Test} Mean absolute error of testing data

MAE_{Train} Mean absolute error of training data

n Number of outputs

n_{Test} Number of outputs in testing data

n_{Train} Number of outputs in training data

N Number of samples

OD_{600} Optical density of the bacterial suspension

O_j^i Value returned (for fitness case j) by the i th expression in the current chromosome

r Robustness ratio

R^2 Coefficient of determination

R_{sen} Sensitivity

$RMSE$ Root mean squared error

t_i Predicted output variable

x_i Input variable

UCS Unconfined compressive strength

Data Availability Statement

All data used during the study are available from the corresponding author by request.

Acknowledgements

This research was financially supported by the Research Grants Council (RGC) of Hong Kong Special Administrative Region Government (HKSARG) of China (Grant No.: R5037-18F).

References

Alavi AH, Gandomi AH (2012) Energy-based numerical models for assessment of soil liquefaction. Geosci Frontiers 3(4): 541-555

Al Qabany A, Soga, K (2013) Effect of chemical treatment used in MICP on engineering properties of cemented soils. Géotechnique 63(4): 331–339

Al-Salloum Y, Abbas H, Sheikh Q I, Hadi S, Alsayed S, Almusallam T (2017) Effect of some biotic factors on microbially-induced calcite precipitation in cement mortar. Saudi J Bio Sci 24(2): 286-294

Amarakoon GGNN, Kawasaki S (2016) Factors affecting the improvement of sand properties treated with microbially-induced calcite precipitation. In Geo-Chicago 2016 (pp. 72-83)

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

Chen RP, Liu QW, Wang HL, Liu Y, Ma QL (2020) Performance of geosynthetic- reinforced pile-supported embankment on soft marine deposit. P I Civil Eng- Geotech Eng in press https://doi.org/10.1680/jgeen.19.00136

Chen RP, Zhang P, Kang X, Zhong ZQ, Liu Y, Wu HN (2019a) Prediction of maximum surface settlement caused by EPB shield tunneling with ANN methods. Soils Found 59: 284–295

Chen YZ, Zhou WH, Liu F, Yi S (2019b) Exploring the effects of nanoscale zero-valent iron (nZVI) on the mechanical properties of lead-contaminated clay. Can Geotech J 56(10): 1395-1405

Cheng L, Cord-Ruwisch R, Shahin MA (2013) Cementation of sand soil by microbially induced calcite precipitation at various degrees of saturation. Can Geotech J 50(1): 81-90

Cheng, L., Shahin, M.A., and Cord-Ruwisch, R (2014a) Bio-cementation of sandy soil using microbially induced carbonate precipitation for marine environments. Géotechnique 64(12): 1010-1013

Cheng L, Shahin MA, Cord-Ruwisch R, Addis M, Hartanto T, Elms C (2014b) Soil stabilisation by microbial-induced calcite precipitation (MICP): investigation into some physical and environmental aspects. In 7th International Congress on Environmental Geotechnics, Melbourne, Australia, p1105–1112

Cheng L, Shahin MA, Mujah D (2017) Influence of key environmental conditions on microbially induced cementation for soil stabilization. J Geotech Geoenviron Eng 143(1): 04016083

Choi SG, Wang K, Chu J (2016) Properties of biocemented, fiber reinforced sand. Constr Build Mat 120: 623-629

Chu J, Stabnikov V, Ivanov V (2012) Microbially induced calcium carbonate precipitation on surface or in the bulk of soil. Geomicrobiol J 29(6): 544-549

Chu J, Ivanov V, Naeimi M, Stabnikov V, Liu HL (2014) Optimization of calcium-based bioclogging and biocementation of sand. Acta Geotech 9(2): 277-285

Cui MJ, Zheng JJ, Lai HJ (2016) Experimental study of effect of particle size on strength of bio-cemented sand. Rock Soil Mech 37(s2): 397-402 (in Chinese)

DeJong JT, Fritzges MB, Nüsslein K (2006) Microbially induced cementation to control sand response to undrained shear. J Geotech Geoenviron Eng 132(11): 1381-1392

DeJong JT, Mortensen BM, Martinez BC, Nelson DC (2010) Bio-mediated soil improvement. Ecol Eng 36(2): 197-210

DeJong JT, Soga K, Kavazanjian E, Burns S, van Paassen LA, Al Qabany A, Aydilek A, Bang SS, Burbank M, Caslake LF, Chen CY, et al (2013) Biogeochemical processes and geotechnical applications: progress, opportunities and challenges. Géotechnique 63(4): 287-301

Ferris FG, Phoenix V, Fujita Y, Smith RW (2004) Kinetics of calcite precipitation induced by ureolytic bacteria at 10 to 20 ℃ in artificial groundwater. Geochimica et Cosmochimica Acta, 68(8): 1701-1710

Gomez MG, Anderson CM, Graddy CM, DeJong JT, Nelson DC, Ginn TR (2017) Large-scale comparison of bioaugmentation and biostimulation approaches for biocementation of sands. J Geotech Geoenviron Eng 143(5): 04016124

Ivanov V, Chu J (2008) Applications of microorganisms to geotechnical engineering for bioclogging and biocementation of soil in situ. Rev Environ Sci Bio 7(2): 139-153

Jiang NJ, Tang CS, Hata T, Courcelles B, Dawoud O, Singh DN (2020) Bio-mediated soil improvement: The way forward. Soil Use Manage 36(2): 185-188

Jin YF, Yin ZY (2020) An intelligent multi-objective EPR technique with multi-step model selection for correlations of soil properties. Acta Geotech 15: 2053–2073

Jin YF, Yin ZY, Zhou WH, Yin JH, Shao JF (2019) A single-objective EPR based model for creep index of soft clays considering L2 regularization. Eng Geol 248: 242–255

Keykha HA, Asadi A, Zareian M (2017) Environmental factors affecting the compressive strength of microbiologically induced calcite precipitation-treated soil. Geomicrobiol J 34(10): 889-894

Li H, Li C, Zhou T, Liu S, Li L (2018) An improved rotating soak method for MICP-treated fine sand in specimen preparation. Geotech Test J 41(4): 805-814

Liu B, Zhu C, Tang CS, Xie YH, Yin LY, Cheng Q, Shi B (2020) Bio-remediation of desiccation cracking in clayey soils through microbially induced calcite precipitation (MICP). Eng Geol 264: 105389

Mahawish A, Bouazza A, Gates WP (2018) Effect of particle size distribution on the bio-cementation of coarse aggregates. Acta Geotech 13(4): 1019-1025

Mahawish A, Bouazza A, Gates WP (2019) Unconfined compressive strength and visualization of the microstructure of coarse sand subjected to different biocementation levels. J Geotech Geoenviron Eng 145(8): 04019033

Mitchell JK, Santamarina JC (2005) Biological considerations in geotechnical engineering. J Geotech Geoenviron Eng 131(10): 1222-1233

Montoya BM, DeJong JT, Boulanger RW (2013) Dynamic response of liquefiable sand improved by microbial-induced calcite precipitation. Géotechnique 63(4): 302-312

Mujah D, Shahin MA, Cheng L (2017) State-of-the-art review of biocementation by microbially induced calcite precipitation (MICP) for soil stabilization. Geomicrobiol J 34(6): 524-537

Mujah D, Cheng L, Shahin MA (2019) Microstructural and geomechanical study on biocemented sand for optimization of MICP process. J Mat Civil Eng 31(4): 04019025

Nafisi A, Mocelin D, Montoya BM, Underwood S (2020) Tensile strength of sands treated with microbially induced carbonate precipitation. Can Geotech J 57(10): 1611-1616

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. In: Technical Report, UBB-01–2002. Babes-Bolyai University, Cluj-Napoca

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

Peng J, Liu Z (2019) Influence of temperature on microbially induced calcium carbonate precipitation for soil treatment. PloS One 14(6): e0218396

Rowshanbakht K, Khamehchiyan M, Sajedi RH, Nikudel MR (2016) Effect of injected bacterial suspension volume and relative density on carbonate precipitation resulting from microbial treatment. Ecol Eng 89: 49-55

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

Sharma A, Ramkrishnan R (2016) Study on effect of microbial induced calcite precipitates on strength of fine grained soils. Perspectives Sci 8: 198-202

Shen SL, Wang ZF, Cheng WC (2017) Estimation of lateral displacement induced by jet grouting in clayey soils. Géotechnique 67(7): 621-630

Soon NW, Lee LM, Khun TC, Ling HS (2014) Factors affecting improvement in engineering properties of residual soil through microbial-induced calcite precipitation. J Geotech Geoenviron Eng 140(5): 04014006

Tang CS, Yin LY, Jiang NJ, Zhu C, Zeng H, Li H, Shi B (2020) Factors affecting the performance of microbial-induced carbonate precipitation (MICP) treated soil: a review. Environ Earth Sci 79(5): 1-23

van Paassen LA, Ghose R, van der Linden TJ, van der Star WR, van Loosdrecht MC (2010) Quantifying biomediated ground improvement by ureolysis: large-scale biogrout experiment. J Geotech Geoenviron Eng 136(12): 1721-1728.

Wang HL, Chen RP (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 HL, Chen RP, Cheng W, Qi S, Cui YJ (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 HL, Chen RP, Liu QW, Kang X (2019b) Investigation on geogrid reinforcement and pile efficacy in geosynthetic-reinforced pile-supported track-bed. Geotext Geomembranes 47(6):755-766

Wang HL, Yin ZY (2020) High performance prediction of soil compaction parameters using multi expression programming. Eng Geol 276: 105758

Wang HL, Yin ZY, Zhang P, Jin YF (2020a) Straightforward prediction for air-entry value of compacted soils using machine learning algorithms. Eng Geol 279: 105911

Wang Y, Konstantinou C, Soga K, DeJong JT, Biscontin G, Kabla AJ (2020b) Enhancing strength of MICP-treated sandy soils: from micro to macro scale. arXiv preprint arXiv:2006.15760

Wang Z, Zhang N, Cai G, Jin Y, Ding N, Shen D (2017) Review of ground improvement using microbial induced carbonate precipitation (MICP). Marine Geores Geotech 35(8): 1135-1146

Wen K, Li Y, Liu S, Bu C, Li L (2019) Development of an improved immersing method to enhance microbial induced calcite precipitation treated sandy soil through multiple treatments in low cementation media concentration. Geotech Geol Eng 37(2): 1015-1027

Whiffin VS, van Paassen LA, Harkes MP (2007) Microbial carbonate precipitation as a soil improvement technique. Geomicrobiol J 24(5): 417-423

Xiao Y, He X, Evans TM, Stuedlein AW, Liu H (2019a) Unconfined compressive and splitting tensile strength of basalt fiber–reinforced biocemented sand. J Geotech Geoenviron Eng 145(9): 04019048

Xiao Y, Stuedlein AW, Ran J, Evans TM, Cheng L, Liu H, van Paassen LA, Chu J (2019b) Effect of particle shape on strength and stiffness of biocemented glass beads. J Geotech Geoenviron Eng 145(11): 06019016

Yin JH, Zhou WH (2009) Influence of grouting pressure and overburden stress on the interface resistance of a soil nail. J Geotech Geoenviron Eng 135(9): 1198-1208

Yuan Y, Shen SL, Wang ZF, Wu HN (2016) Automatic pressure-control equipment for horizontal jet-grouting. Auto Constr 69: 11-20

Zhang P, Yin ZY, Jin YF, Chan THT (2020) A novel hybrid surrogate intelligent model for creep index prediction based on particle swarm optimization and random forest. Eng Geol 265: 105328

Zhang P, Yin ZY, Jin YF (2021) State-of-the-art review of machine learning applications in constitutive modeling of soils. Arch Comput Meth Eng https://doi.org/10.1007/s11831-020-09524-z

Zhao LS, Zhou WH, Geng X, Yuen KV, Fatahi B (2019) A closed-form solution for column-supported embankments with geosynthetic reinforcement. Geotex Geomembr 47(3): 389-401

Zhao Q, Li L, Li C, Li M, Amini F, Zhang H (2014) Factors affecting improvement of engineering properties of MICP-treated soil catalyzed by bacteria and urease. J Mat Civil Eng 26(12): 04014094

Zhao Y, Fan C, Ge F, Cheng X, Liu P (2020) Enhancing strength of MICP-treated sand with scrap of activated carbon-fiber felt. J Mat Civil Eng 32(4): 04020061

Table 1. Descriptive statistics of variables

Variable	Minimum	Maximum	Mean	Standard deviation
d_{50} (mm)	0.120	1.600	0.385	0.310
C_u	1.17	6.23	1.79	1.138
e_0	0.43	1.04	0.64	0.074
OD_{600}	0.30	4.46	1.95	1.194
M_u (mol/L)	0.10	1.50	0.73	0.339
M_{Ca} (mol/L)	0.10	1.50	0.71	0.334
F_{Ca} (%)	1.49	29.47	8.78	6.491
UCS (MPa)	0.05	14.23	1.79	1.973

Table 2. Parameter setting of the MEP code for each data group

Parameter	Setting
Population size	3000
Number of generation	3000
Code length	50
Crossover probability	0.9
Crossover type	Uniform
Mutation probability	0.01
Function set	+, −, ×, /; +, −, ×, /, pow; +, −, ×, /, lg; +, −, ×, /, pow, lg
Terminal set	Problem input
Replication number	10

Table 3. Basic setting of input variables for monotonicity analysis

Soil parameter	Value
d_{50} (mm)	0.385
C_u	1.79
e_0	0.64
OD_{600}	1.95
M_u (mol/L)	1
M_{Ca} (mol/L)	1
F_{Ca} (%)	8.78

Figure captions

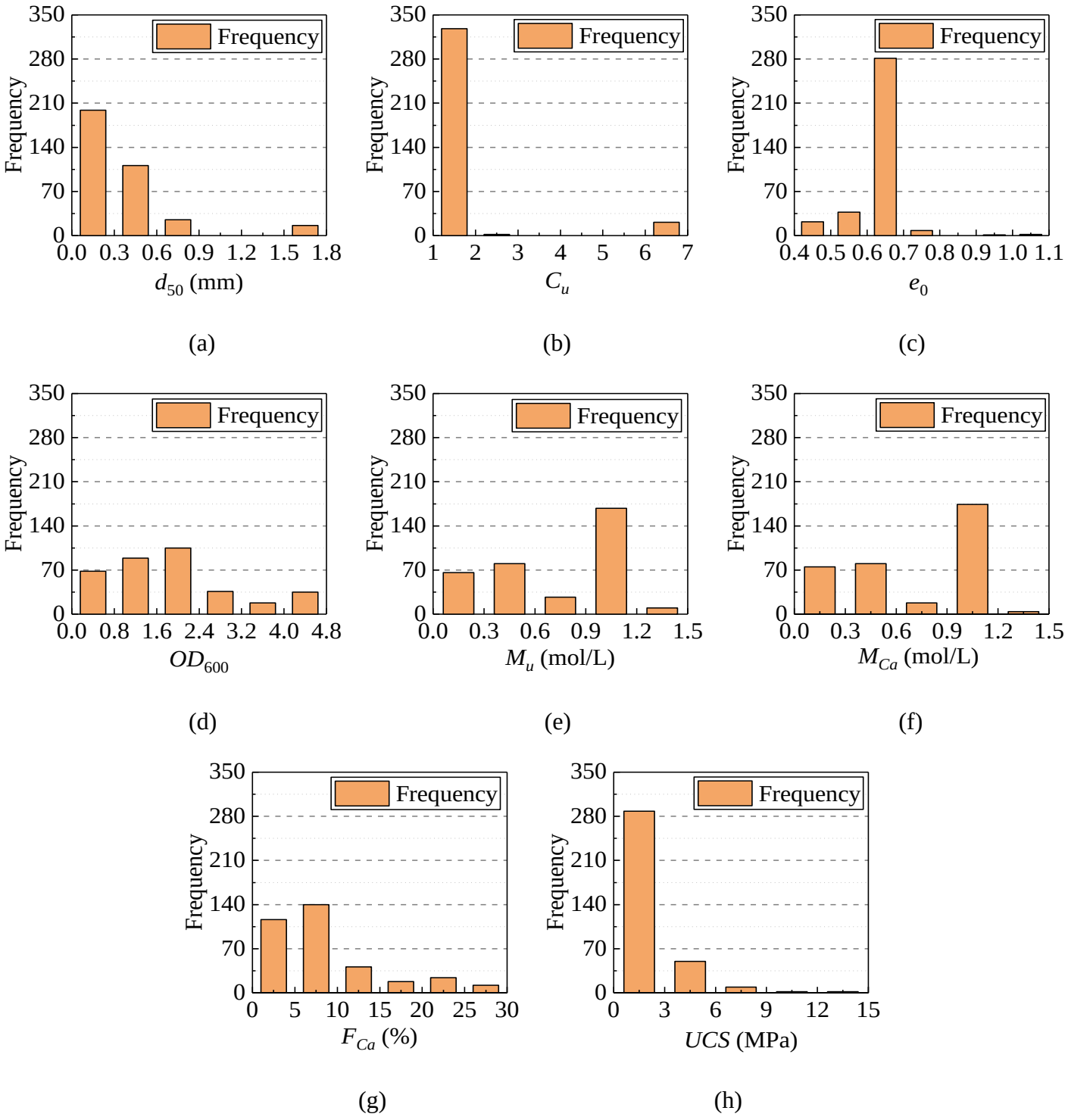


Fig. 1. Frequency histograms of the variables: (a) median grain size; (b) coefficient of uniformity; (c) initial void ratio; (d) optical density of bacterial suspension; (e) urea concentration; (f) calcium concentration; (g) precipitated calcium carbonate content; (h) unconfined compressive strength

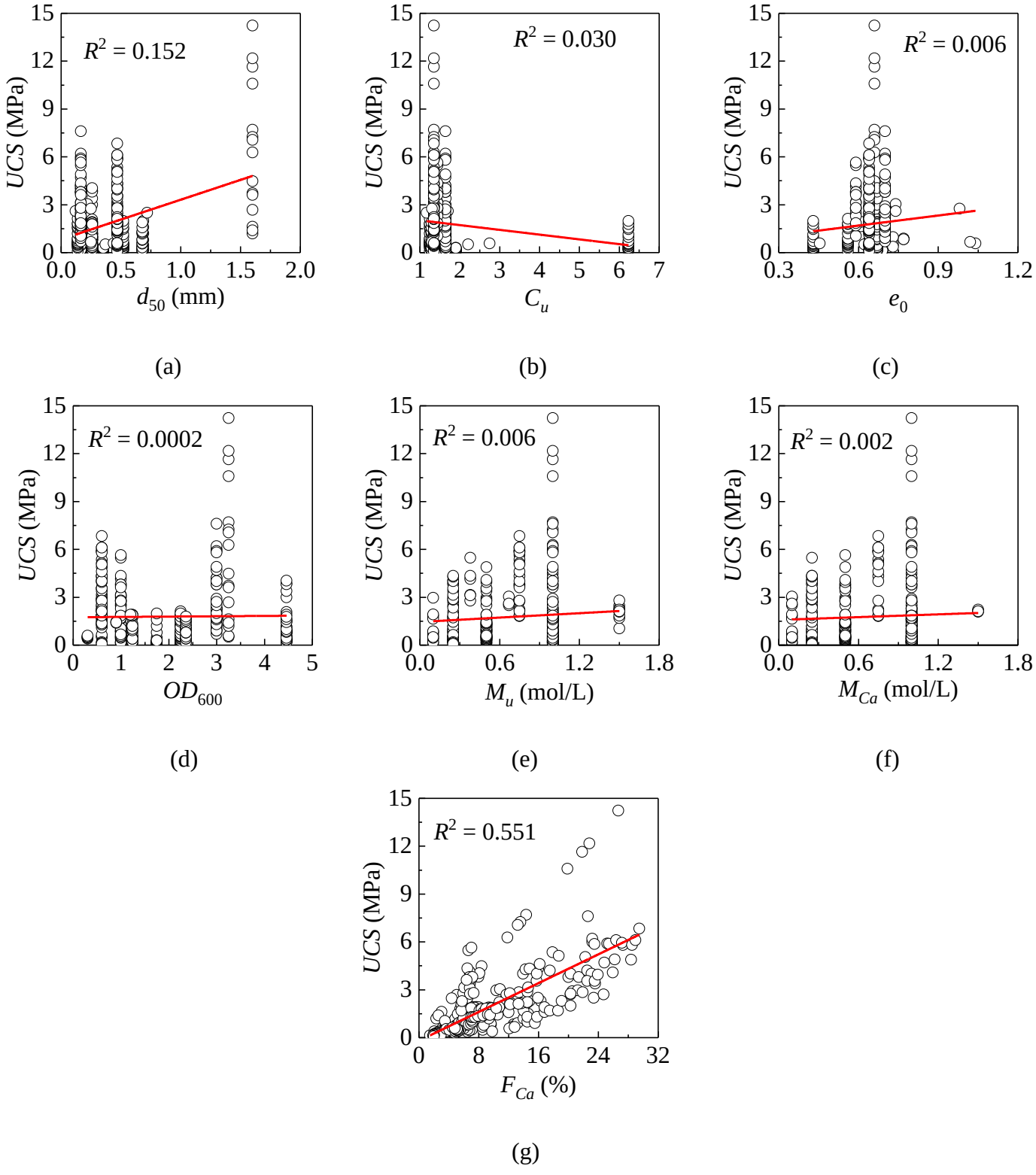


Fig. 2. Basic linear fittings between unconfined compressive strength and each input variable: (a) median grain size; (b) coefficient of uniformity; (c) initial void ratio; (d) optical density of bacterial suspension; (e) urea concentration; (f) calcium concentration; (g) precipitated calcium carbonate content

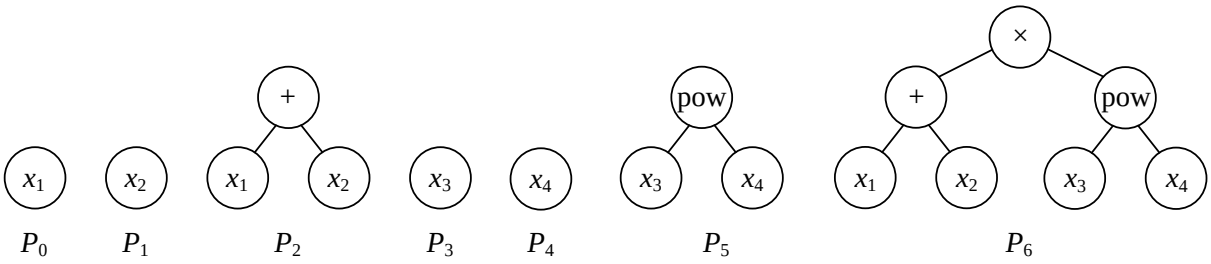
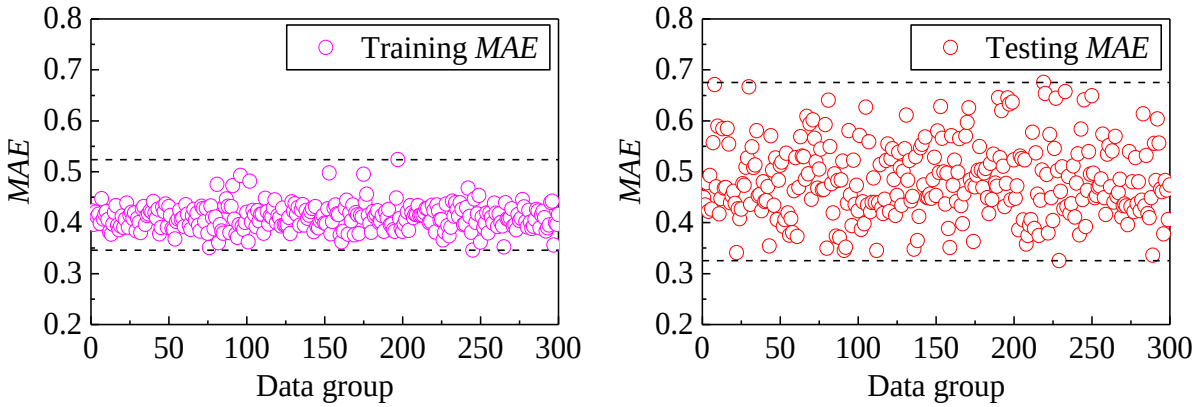


Fig. 3. Expression of a MEP chromosome represented by gene trees



(a) (b)

Fig. 4. Mean absolute error of (a) training data and (b) testing data for each data group

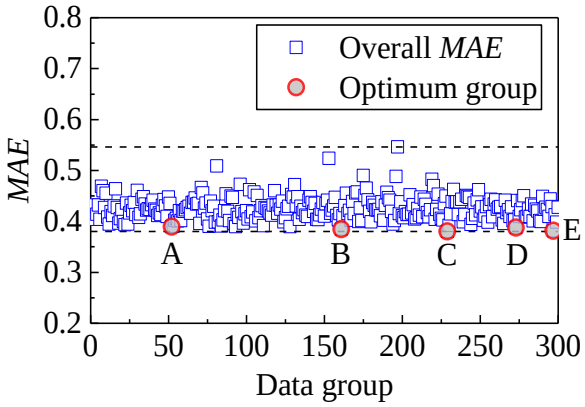


Fig. 5. Mean absolute error of the whole database for each data group

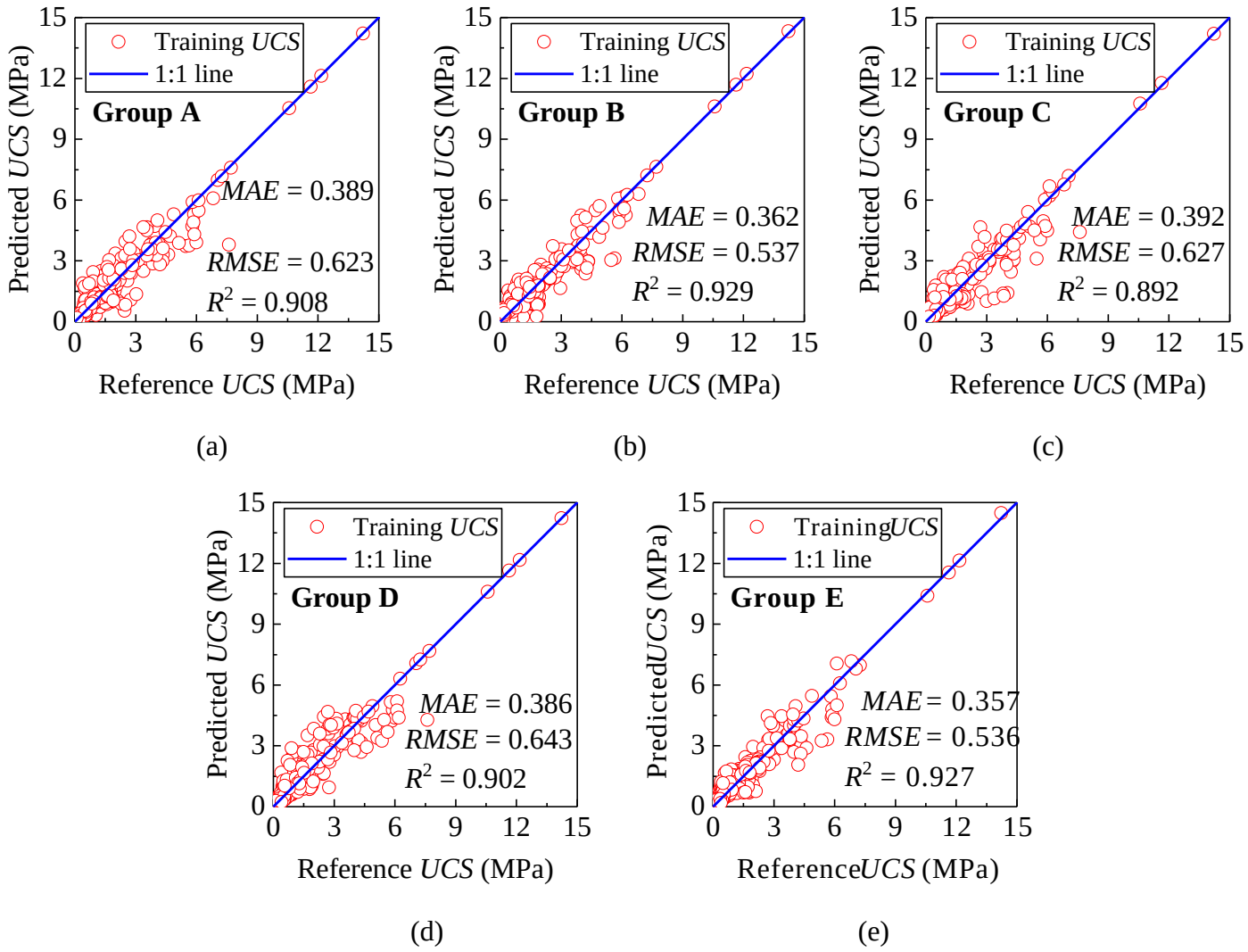


Fig. 6. Comparison of predicted and reference unconfined compressive strength for the training data of (a) Group A; (b) Group B; (c) Group C; (d) Group D and (e) Group E

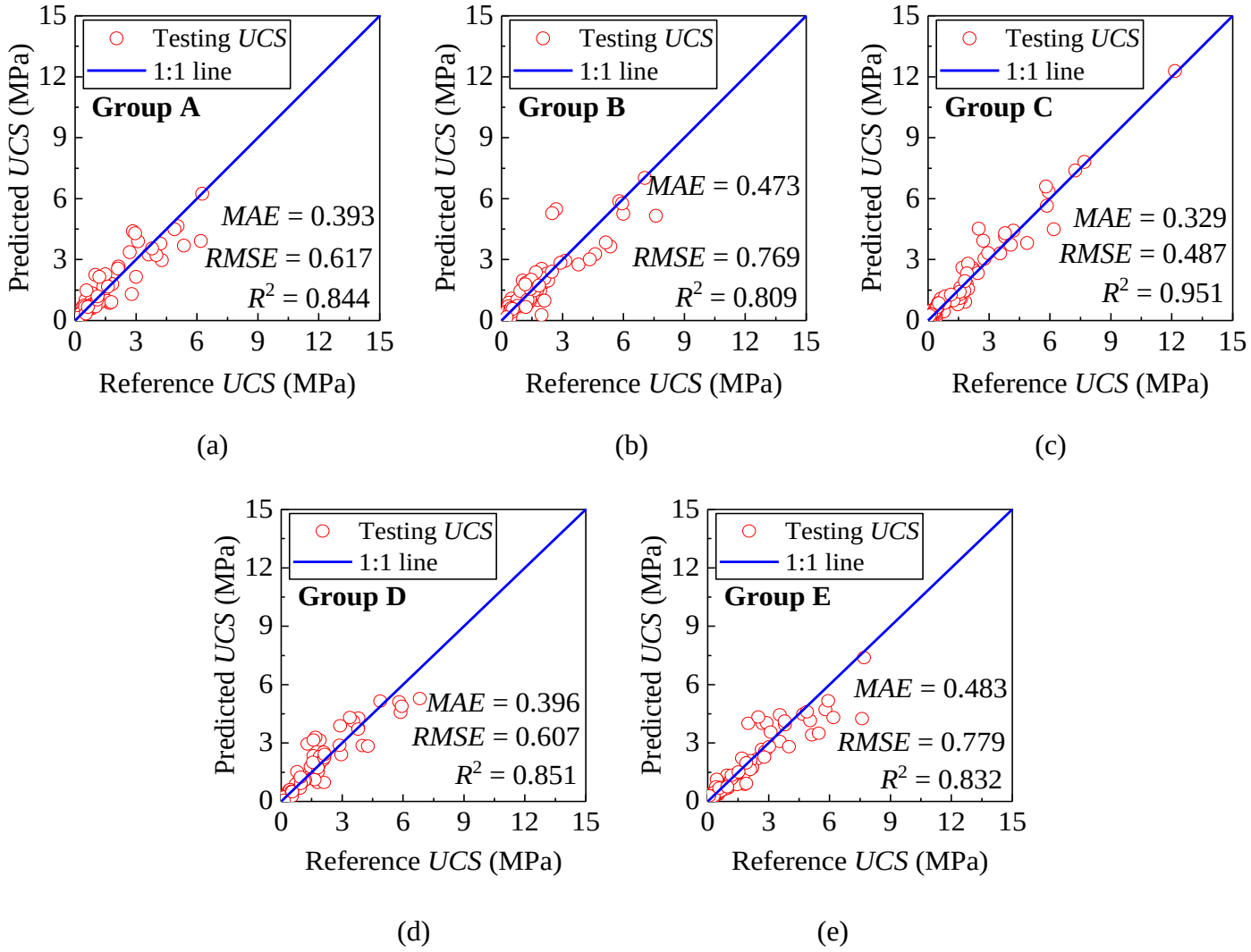


Fig. 7. Comparison of predicted and reference unconfined compressive strength for the testing data of (a) Group A; (b) Group B; (c) Group C; (d) Group D and (e) Group E

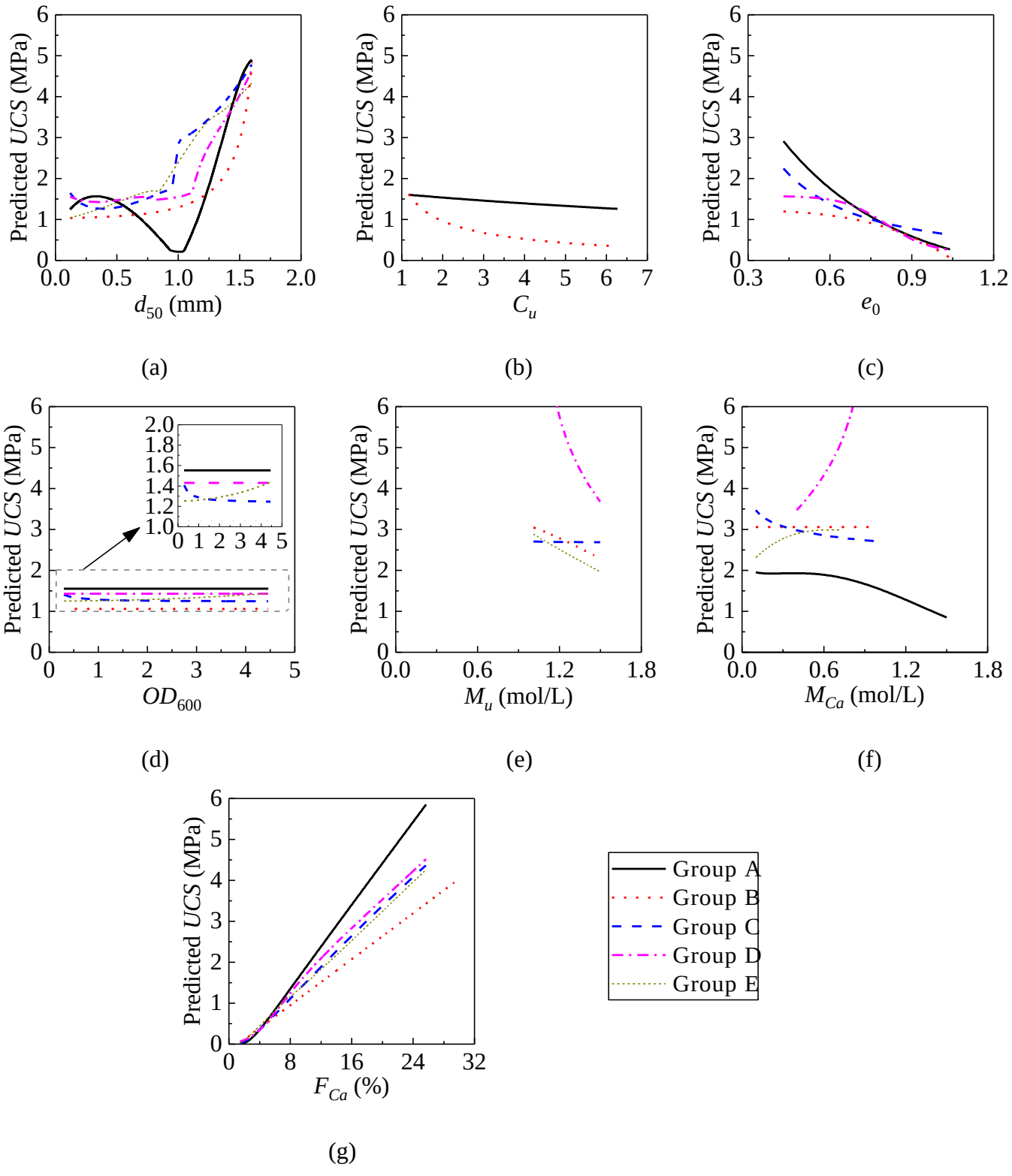


Fig. 8. Monotonicity analysis of the predicted unconfined compressive strength versus (a) median grain size; (b) coefficient of uniformity; (c) initial void ratio; (d) optical density of bacterial suspension; (e) urea concentration; (f) calcium concentration; (g) precipitated calcium carbonate content

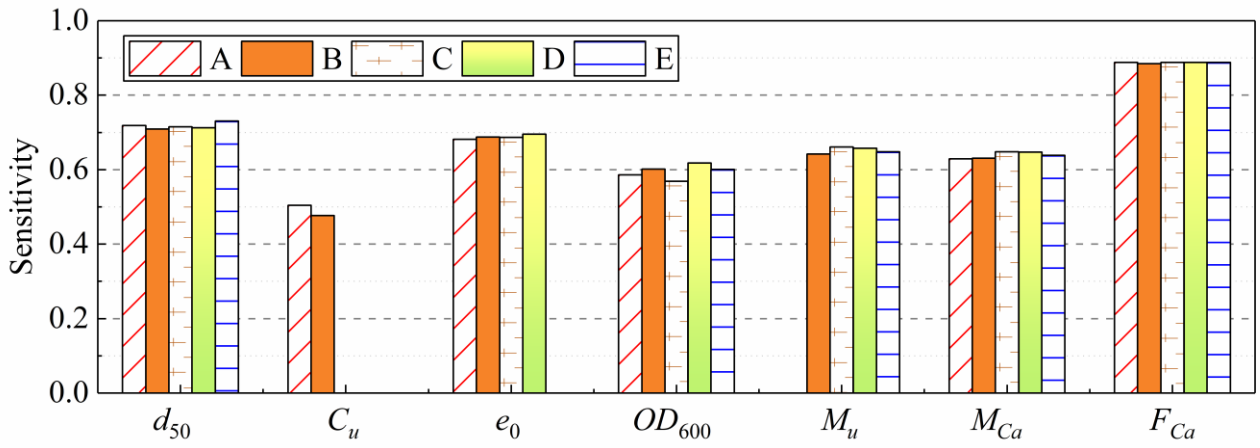


Fig. 9. Sensitivity analysis about the relevance of the input variable on the predicted unconfined compressive strength

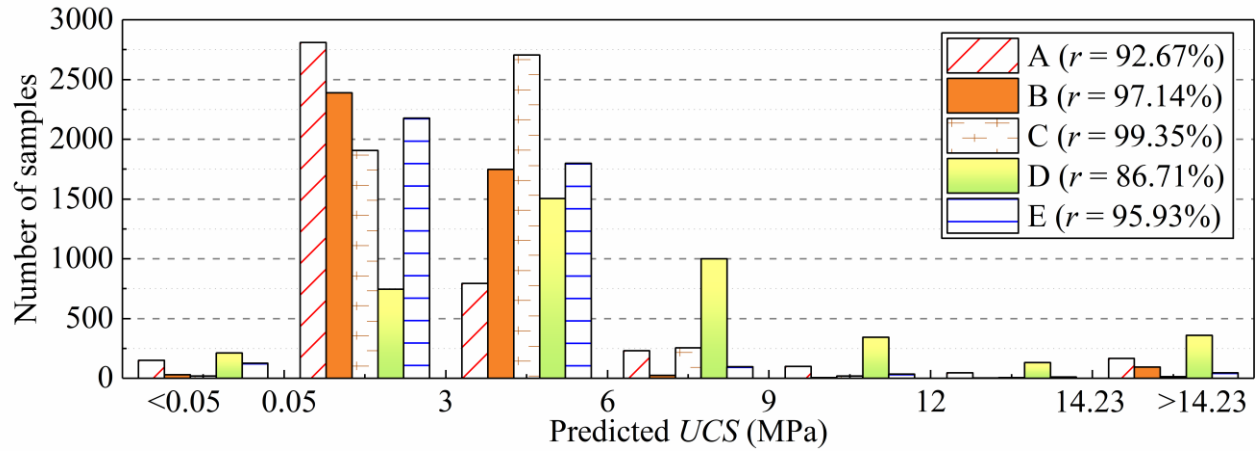


Fig. 10. Robustness analysis of the input variable on the predicted unconfined compressive strength