https://dx.doi.org/10.1016/j.enggeo.2018.12.006

A single-objective EPR based model for creep index of soft clays considering L_2 regularization

Yin-Fu JIN[1], Zhen-Yu YIN[1*], Wan-Huan ZHOU[2], Jian-Hua YIN[1] and Jian-Fu SHAO[3]

Affiliation:

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

2 Department of Civil and Environmental Engineering, Faculty of Science and Technology, University of Macau, Macau, China

3 University of Lille, CNRS, Centrale Lille, FRE 2016 – LaMcube – Laboratoire de mécanique multiphysique multiéchelle, F-59000, Lille, France

* Corresponding author: Dr Zhen-Yu YIN, Tel. +852 34008470, Fax +852 23346389, E-mail: zhenyu.yin@polyu.edu.hk; zhenyu.yin@gmail.com

Abstract: A correlation of creep index (C_α) with high performance is highly recommended for soft clay engineering practice. Current empirical correlations are only acceptable for a few clays. This paper aims to propose a robust and effective evolutionary polynomial regression (EPR) model for C_α of clay. First, a database covering various clays is formed, in which 120 data are randomly selected for training and the remaining data are used for testing. To avoid overfitting, a novel EPR procedure using a newly enhanced differential evolution (DE) algorithm is proposed with two enhancements: (1) a new fitness function is proposed using the structural risk minimization (SRM) with L_2 regularization that penalizes polynomial complexity, and (2) an adaptive process for selecting the combination of involved variables and size of polynomial terms is incorporated. By comparing the predictive ability, model complexity, robustness and monotonicity, the EPR formulation for C_α involving clay content, plasticity index and void ratio with three terms is selected as the optimal model. A parametric study is then conducted to assess the importance of each input in the proposed model. All results demonstrate that the proposed model of C_α is simple, robust, and reliable for applications in engineering practice.

Key words: soft clays; Atterberg limits; creep; optimization; evolutionary polynomial regression; model selection

1 Introduction

Natural soft clays exhibit significant creep under both laboratory and in situ conditions after primary consolidation, which significantly influences the long-term safety of infrastructures in various fields, such as tunnelling (Ren et al. 2018; Shen et al. 2014; Wu et al. 2015; Wu et al. 2017), excavation (Feng et al. 2003; Jin et al. 2019; Wang et al. 2009; Zhang et al. 2013), embankment (Chai et al. 2018; Karstunen and Yin 2010; Rezania et al. 2017; Shen et al. 2005; Yin et al. 2011a; Zhu et al. 2014; Zhu et al. 2015), urban land subsidence (Shen et al. 2013; Shen and Xu 2011; Xu et al. 2016; Xu et al. 2012a; Xu et al. 2012b), etc. Usually, the creep property of soft clays is represented by the creep index $C_\alpha=\Delta e/\Delta\log(t)$, where e is void ratio and t is time during secondary compression. The creep index is a key parameter for most viscoplastic constitutive models applicable to the engineering practice (Yin et al. 2002; Yin and Cheng 2006; Yin et al. 2017a; Yin et al. 2015a; Yin et al. 2014; Yin et al. 2017b; Yin et al. 2010a; Yin and Karstunen 2011; Yin et al. 2011b; Yin et al. 2010b; Yin and Wang 2012; Zhu and Yin 2000), which is usually obtained by a conventional oedometer test. According to studies of (Yin et al. 2017b; Yin and Karstunen 2011; Yin et al. 2011b), the C_α corresponding to intact clays is not constant because of the effect of destructuration to the creep. In contrast, the C_α of reconstituted clay is an intrinsic property, which is the base for understanding the creep characteristic and thus more suitable to be adopted in practice (Mesri and Godlewski 1977). Because of this, the attention is paid to the C_α of reconstituted clay in this study.

The creep property should relate to the microstructure of clay (Yin et al. 2014; Yin et al. 2017b; Yin et al. 2008; Yin and Chang 2009). Unfortunately, the microstructure of clay is expensive to measure which may lead to a practical obstacle. Physical properties can somehow reflect the microstructure of clay. Thus practically, it is very convenient to get ideas of the intrinsic value of C_α only based on physical properties of clay. Some attempts have been made to correlate the C_α to some physical properties of soils (such as water content, void ratio, Atterberg limits)(Anagnostopoulos and Grammatikopoulos 2011; Nakase et al. 1988; Suneel et al. 2008; Yin 1999; Zeng et al. 2012a; Zeng and Liu 2010; Zhu et al. 2016). However, these correlations are only applicable for few clays and thus not enough reliable for soft clay engineering practice. Therefore, a robust and effective correlation between C_α and physical properties of clay is worth investigating.

Numerical regression is the most powerful and commonly applied form of regression used to solve the problem of finding the best model to fit the observed data (Alemdag et al. 2015; Gurocak et al. 2008; Gurocak et al. 2012; Zhang et al. 2015; Zhou et al. 2016). Evolutionary Polynomial Regression (EPR) is a recently developed hybrid regression method (Giustolisi and Savic 2006) that has advantages in modelling nonlinear complex problems. Applications in geotechnics include stability prediction of slopes (Ahangar-Asr et al. 2010; Doglioni et al. 2015; Gurocak et al. 2008), modelling of clay compressibility (Wu et al. 2018; Yin et al. 2016), modelling of permeability and compaction characteristics of soils (Ahangar-Asr et al. 2011), evaluation of liquefaction potential of sand (Rezania et al. 2011; Rezania et al. 2010), prediction of soil saturated water content (Khoshkroudi et al. 2014), settlement prediction of foundations (Ghorbani and Firouzi Niavol 2017; Shahin 2014; Shahnazari et al. 2014), evaluation of pile bearing capacity (Ahangar-Asr et al. 2014; Ebrahimian and Movahed 2013, 2017), pipeline failure prediction (Kakoudakis et al. 2017), modelling of soil behaviours (Faramarzi et al. 2014; Javadi et al. 2012; Nassr et al. 2018; Shahnazari et al. 2013), etc. These successful applications have demonstrated that the EPR technique is superior to other soft computing techniques, such as artificial neural networks (ANNs)(Kalinli et al. 2011), or genetic programming (GP) (Alemdag et al. 2015; Rezania et al. 2010). More recently, the development of optimization algorithms (Jin et al. 2017a; Jin et al. 2017b; Jin et al. 2016a, b; Jin et al. 2017c; Jin et al. 2018; Yin et al. 2016; Yin et al. 2018; Yin et al. 2017a) can improve the EPR technique in a more adaptive way. Thus, the optimization combined EPR technique is worth trying for the C_α of clay. However, in current EPR modelling most fitness functions are only based on training data error, such as the sum of squared errors (SSE) or coefficient of determination (COD). As a result, the proposed models are usually overfitting and weak in terms of the generalization ability. Therefore, the EPR technique needs to be improved to avoid overfitting and with good generalization ability.

In this study, a simple, robust and reliable correlation between the intrinsic C_α and physical properties of clay is proposed with improving and using the EPR technique. Firstly, a database including physical properties (clay content, Atterberg limits and void ratio) and C_α for various reconstituted clays is formed, in which 120 randomly selected data are used for training EPR model and the remaining data are used for testing. An efficient EPR procedure is proposed, in which the newly developed Nelder-Mead simplex differential evolution algorithm (NMDE) is employed as the optimization tool to search the optimal exponents; a fitness function based on structural risk minimization (SRM) with L_2 regularization is proposed and implemented; and an adaptive procedure for selecting the involved variables and the size of terms in EPR model is proposed and implemented. Then, six EPR models of C_α with different combinations of involved variables and different sizes of terms are obtained. Next, the optimal EPR model of C_α is selected among them based on the predictive ability, model complexity, robustness and monotonicity. Finally, a parametric study is conducted to assess the level of contribution of each physical property in the proposed model.

2 Database

2.1 Statistics and basic correlation analysis

Massive experimental data from various studies (Li et al. 2012; Yin 1999; Yin et al. 2015b; Zeng et al. 2012a; Zhu et al. 2016) were collected and used to develop the EPR model of intrinsic C_α. The clay content (*CI*), liquid limit (w_L), plastic limit (w_P), plasticity index (I_P) and void ratio (e) were treated as correlating variables of interest. Table 1 summarizes those physical properties and C_α for all selected reconstituted clays. Based on the data, the statistical analysis (maximum value, minimum value, mean value and standard deviation) for each property were conducted, summarized in Table 2.

The linear correlations between C_α and each main basic physical property (*CI*, w_L, w_P, I_P and e) are presented in Fig. 1. It is found that the C_α relatively highly correlates to e, followed by I_p and w_L, and very poorly correlates to w_P and *CI*. It is obvious that none of the basic correlations involving single physical property for predicting C_α is satisfactory for engineering purposes (R^2<0.8) and thus the performance of the correlation still needs to be improved. Note that the increasing of the ratio of clay to silt or sand can increase significantly the creep index (Yin, 1999), the *CI* should be considered in correlation.

2.2 Discrepancy of current correlation formula

Table 3 summarizes five existing empirical correlations of C_α, which were used to fit the database presented in Table 1. Fig. 2 shows the comparison between predictions and measurements for these empirical correlations. For the correlations only involving I_P (Nakase et al. 1988; Yin 1999), their performances are poor with a low value of correlation coefficient R^2. For the correlations proposed by Zeng et al. (2012a) and Zhu et al. (2016), their performances are almost the same, but the R^2 is still smaller than 0.8. Therefore, it is still recommended and should be practically useful to improve the correlation of C_α to physical properties.

Therefore, a novel correlation approach to find a reliable and reasonable correlation between C_α and physical properties was proposed, as presented below.

3 Differential evolution-based EPR modelling

3.1 General EPR procedure

The evolutionary polynomial regression (EPR) is a data-driven method based on evolutionary computing, aiming to search for polynomial structures representing a system, which was first introduced by Giustolisi and Savic (2006) with applications in the hydroinformatics and environment related problems. A general EPR expression can be mathematically formulated as:

$$y = \sum_{j=1}^{m} F\left(\mathbf{X}, f\left(\mathbf{X}\right), a_j\right) + a_0 \quad (1)$$

where y is the estimated vector of output of the process; a_0 is an optional bias; a_j is an adjustable parameter for the jth term; F is a function constructed by the process; $\mathbf{X}$ is the matrix of input variables; f is a function defined by the user; and m is the number of terms of the target expression.

According to Giustolisi and Savic (2006), the first step in identifying the model structure is to transfer Eq.(1) to the following vector form:

$$\mathbf{Y}_{N\times 1}\left(\boldsymbol{\theta}, \mathbf{Z}\right) = \left[\mathbf{I}_{N\times 1} \quad \mathbf{Z}^{j}_{N\times m}\right] \times \left[a_0 \quad a_1 \quad \dots \quad a_m\right]^{\mathrm{T}} = \mathbf{Z}_{N\times d} \times \boldsymbol{\theta}^{\mathrm{T}}_{d\times 1} \quad (2)$$

where $\mathbf{Y}_{N\times 1}\left(\boldsymbol{\theta}, \mathbf{Z}\right)$ is the least-squares (LS) estimator vector of N target values; $\boldsymbol{\theta}_{d\times 1}$ is the vector of d

(=m+1) parameters a_j and a_0 ($\boldsymbol{\theta}^{\mathrm{T}}$ is the transposed vector); and $\mathbf{Z}_{N\times d}$ is a matrix formed by **I** (unitary vector) for bias a_0, with m vectors of variables $\mathbf{Z}^j$. More details about the EPR can be found in Giustolisi and Savic (2006).

Fig. 3 shows a typical flow chart for the EPR procedure (Giustolisi and Savic 2006). The general functional structure represented by $f\left(\mathbf{X}, a_j\right)$ in Eq.(1) is constructed from elementary functions by EPR using an optimization algorithm strategy (such as genetic algorithm). Note that any optimization algorithm guaranteeing the global optimal solution can be employed in the EPR procedure. The building blocks (elements) of the structure are defined by the user based on understanding of the physical process. The selection of feasible structures to be combined is conducted through an evolutionary process, while the parameters a_j in Eq.(2) are estimated by the least squares method.

3.2 Implementation of NMDE in EPR modelling

In order to improve the efficiency of EPR modelling, the newly developed NMDE by Yin et al. (2018) was employed to select the useful input vectors from **X** to formulate the EPR. In NMDE the Nelder-Mead simplex (NMS) is used to accelerate the convergence speed (Fig. 4). Before performing the differential evolution (DE) mutation, all individuals are sorted based on their fitness value, and the best n+1 (n is the number of variables) is selected to perform the NMS. Based on the results of the NMS, the best individual is updated and then recombined with the N-(n+1) remaining individuals to perform the DE mutation. This process will be executed N times, resulting in a new population of N individuals. Then, the obtained population is applied to the crossover operation. To avoid a rapid loss of diversity, an elitism strategy is adopted when performing the selection, in which the 10% of individuals with the highest fitness are selected from the parents and children to survive to the next generation. The remainders are chosen by tournament selection from the mating pool composed of parents and children. The completion mechanism can help the NMDE to identify better solutions.

3.3 New fitness function considering L_2 regularization

The performance of an EPR procedure mainly depends on fitness function. A widely used fitness function is structural risk minimization (SRM) (Garg et al. 2017), which involves the addition of model complexity term (size of model) in the empirical error and punishes the model fitness based on its size. Another problem is that a relatively small amount of data will increase the risk to cause the model overfitting, making the training error small and the testing error particularly large, which would weaken the generalization ability of an EPR model. Then, the use of regularization/penalty functions (e.g., L_0, L_1 and L_2 regularizations) to avoid overfitting is suggested (Coelho and Neto 2017). Among various regularizations, the L_2 regularization is usually adopted (Ng 2004). Therefore, a modified mathematical formulation of SRM considering the L_2 regularization was adopted in this study, given as:

$$\mathrm{SRM} = \frac{\mathrm{SSE}}{N}\left(1-\sqrt{\left(\frac{n}{N}-\left(\frac{n}{N}\log\left(\frac{n}{N}\right)\right)+\left(\frac{\log\left(\frac{n}{N}\right)}{2N}\right)\right)}\right)^{-1}+\lambda\|\boldsymbol{\omega}\|_2^2 \quad (3)$$

with

$$\mathrm{SSE} = \sum_{i=1}^{N}\left(\mathbf{Y_m}-\mathbf{Y_p}\right)^2 \text{ and } \|\boldsymbol{\omega}\|_2^2 = \sum_{j=1}^{n}\boldsymbol{\omega}_j = \boldsymbol{\omega}^T\boldsymbol{\omega} \quad (4)$$

where N is the number of data points on which the SRM is computed; $\mathbf{Y}_\mathrm{m}$ is the vector of measured values; $\mathbf{Y}_\mathrm{p}$ is the vector of predicted values; $\boldsymbol{\omega}$ is the vector of model coefficients; λ is regularization parameter that requires manual adjustment to find an appropriate value.

3.4 Adaptive selection of correlating variables and term size

A reliable EPR model should have a reasonable trade-off between predictive ability and generalization ability. As stated by Wood (2003), simple yet adequate models are favoured on the basis of practicality. Therefore, an EPR procedure combining with model selection process should be proposed to ensure the model "simple" enough based on minimizing the training error. Then the model could also have a good generalization performance (e.g., the testing error is also small). In this case, the model selection involves two aspects: selecting the suitable combination of correlating variables and the appropriate size of terms.

Fig. 5 presents the proposed procedure, where $\boldsymbol{\theta}$ is the decision variables corresponding the exponents of EPR model; *Comb* represents the number of combination of correlating variables; *m* is the size of terms. Compared to the common EPR process, two additional variables *Comb* (an integer number) and *m* (an integer number) are added to the vector of optimization variable in the proposed procedure. Firstly, all variables in initial generation are generated randomly within their domains. Next, the possible combination of correlating variables is selected according to the value of *Comb* and then a possible term size is chosen according to the value of *m*. Subsequently, a generated EPR model with unknown coefficients according to Eq.(6) is attained. Then, the vector of coefficients $\boldsymbol{a}$ is determined by regression between the measurements and predictions. Finally, the fitness SRM with L_2 regularization is computed to evaluate the performance of EPR model, which determines whether the formula can survive to next generation in the DE-evolution. Once the stop criterion (e.g., the maximum number of generation) is reached, the whole process is exited; otherwise, the process will continue to the next generation.

With increasing the number of generations, the appropriate combination of correlating variables and term size will be automatically selected among numerous calculations. Moreover, through adjusting the regularization parameter, the most appropriate EPR model in terms of model complexity and generalization ability can be finally found.

3.5 Suggestion of regularization parameter

To find an appropriate value of regularization parameter λ to trade off the performance between prediction and generalization, several attempts to assign different values of λ are needed in the proposed EPR procedure. Firstly, the λ=0 is tested to check the general predictive performance of the EPR model without regularization. Then, a value of λ (e.g., 10^{-4}) that closes to zero is assigned to evaluate the effect of regularization on the selection of model. Note that the value of first tried λ is different for various concerned cases, because the value of λ is related to the coefficients of obtained equation. Next, based on investigated results, a series of calculation attempts using different values of λ are conducted based on which several formulas can be achieved. Among them, the formula that the predictive performance is similar to the one of the model with λ=0 is saved, otherwise discarded. The finally retained formulas have different number of term sizes. The most appropriate model will be eventually selected through the comparison of predictive ability and the number of term sizes, even other criteria depending on the problem on which for instance the robustness is more important or the accuracy.

4 EPR modelling of creep index

4.1 EPR modelling process for C_α

Since selecting (w_L, I_p) or (w_p, I_p) is physically the same for evaluating the C_α, based on the statistics results of database, four physical properties (CI, w_L, I_P and e) were selected as the correlating variables of interest to training the EPR model. To attain the nonlinear creep behaviour with a consecutively decreasing creep index C_α that fully relates to the soil density (Yin et al. 2015b; Zhu et al. 2016), a general structure of EPR expression for C_α was proposed as:

$$\ln\left(C_\alpha\right)=\sum_{j=1}^{m}\left[f\left(CI,\ w_L,\ I_P\right)e\right]+a_0 \tag{5}$$

which was further expressed as:

$$\ln\left(C_\alpha\right)=\left(\sum_{j=1}^{m}\left[a_j\left(CI\right)^{\theta_{j1}}\left(w_L\right)^{\theta_{j2}}\left(I_P\right)^{\theta_{j3}}\right]\right)e+a_0 \tag{6}$$

where a_0 is a constant in the EPR equation; a_j is the coefficient corresponding to j term and θ_j is the vector of exponent. Note that the use of logarithm in C_α can guarantee the positiveness of C_α.

To obtain an accurate and reasonable correlation, 120 data randomly selected in the prepared database were used for training and the remaining data were used for testing. For simplicity, the value of exponent was constrained to [-2, 2] with a step size to 1. Also, the maximum number of terms was set to 8 for restricting the model complexity. For NMDE, the number of initial population was set to ten times of decision variables and the maximum generation was set to 200. The probability of crossover CR is 0.3. For NMS, the tolerance for convergence was set to 10^{-4}. Independent multiple runs were performed to avoid randomness.

As shown in Fig. 5, for modelling C_α, the variable e is fixed to keep the unique relationship between C_α and e. A total of seven combinations ($=\mathrm{C}_3^1+\mathrm{C}_3^2+\mathrm{C}_3^3$) that each one contains different physical properties are obtained. Thus, the total number of combinations $Comb$ is 7 and the total number of term size m is 8. Following the proposed EPR procedure, the most appropriate EPR model for C_α in terms of model complexity and generalization ability can be finally found.

4.2 Analysis of results

To evaluate the performance of the obtained EPR model, five indicators are used. Besides the mean value u and standard deviation value σ, coefficient of determination (R^2), root mean square error (RMSE) index and mean absolute error (MAE) are expressed as:

$$R^2 = \frac{\sum_{i=1}^{N}\left(\mathbf{Y}_\mathrm{m}\right)^2 - \sum_{i=1}^{N}\left(\mathbf{Y}_\mathrm{m} - \mathbf{Y}_\mathrm{p}\right)^2}{\sum_{i=1}^{N}\left(\mathbf{Y}_\mathrm{m}\right)^2} \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{Y}_\mathrm{m} - \mathbf{Y}_\mathrm{p}\right)^2} \tag{8}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\mathrm{Y}_\mathrm{m} - \mathrm{Y}_\mathrm{p}\right| \tag{9}$$

Higher R^2 or lower RMSE and MAE values represent better model performance. Meanwhile, both the mean value "u" and the standard deviation value "σ" of $\mathbf{Y}_\mathrm{p}/\mathbf{Y}_\mathrm{m}$ were calculated. A "u" value greater than 1.0 indicates over-estimation and under-estimation otherwise.

Followed the suggestion of selecting λ, a series of calculation attempts using different values of λ (i.e., λ=0, 0.0001, 0.001, 0.01, 0.05 and 0.1) were carried out. Fig. 6 shows the evolution of model selection in terms of variable combination and size of terms for different values of λ. The results show that all models compete with each other to keep the diversity of population during the initial stage of EPR; then, with increasing the number of generations, the models having higher fitness (small training error) survive to next generation and the others with lower fitness are discarded; finally, the percentage of the models with good performance continues to rise to 100 %. Therefore, the most appropriate model represented by variable combination and size of terms are automatically selected using the proposed EPR procedure. The results demonstrate that the proposed EPR procedure combined with model selection is efficient.

Table 4 summarizes the expressions of all the proposed EPR models. It is found that all obtained EPR expressions contain the I_P, which demonstrates that the I_P has an important effect on the C_α. Fig. 7 shows the comparison of C_α between measurements and different EPR predictions for training and testing data. All obtained results are summarized in Table 5. With increasing the value of λ, a trend that the term size of obtained model decreases is found. Apart from the model with λ=0, the performance of other models is similar in terms of R^2, RMSE, MAE, u and σ. The performance of all EPR models is acceptable. Based on preliminary results, it seems that the performance of all EPR models is acceptable. However, considering the less complexity of an appropriate model, only EPR models with three terms (Eq.(11) and Eq.(12)) can be considered as the optimal models.

4.3 Robustness testing for proposed EPR models

An appropriate model has not only a good predictive ability and less complexity but also good robustness. The latter indicates the predicted values are always guaranteed reasonable for reasonable input values. To assess the robustness of each EPR model, the robustness tests were performed and a criterion representing the robustness was defined as:

$$\text{Robustness ratio}=\frac{\text{Samples located in reasonable range}}{\text{Total samples}} \tag{13}$$

The reasonable range for C_α in this case is [0.001~0.1] according to the statistical results, which is also applicable for most reconstituted soft clays (Yin et al. 2014; Yin et al. 2017b; Yin et al. 2015b; Zhu et al. 2016). To generate the testing samples, it supposes that variables (CI, w_L, I_p and e) are independent of each other and meet the multivariable lognormal distribution according to various studies (Cao and Wang 2014; Zhang et al. 2009; Zhang et al. 2017). According to the statistic results of used database shown in Table 2, the values of mean and standard deviation for each variable were employed to randomly generate 10000 samples from its lognormal distribution. Note that for the robustness testing, the samples should be adjusted according to the specific problem and the related variables, not 10000 samples for all cases. Then, the C_α was predicted by each proposed EPR model and the robustness ratio was then computed for each model. For each robustness test, the mean and standard deviation for samples locating in the reasonable range were also calculated.

Fig. 8 presents the results of robustness tests for four potential EPR models. It can be seen that the proposed EPR model (Eq.(11)) involving CI and I_P with 3 terms has the highest robustness ratio, and the mean and standard deviation values (0.0191$\pm$0.0125) predicted by this model are very close to the values (0.0182$\pm$0.0110) of the used database. Thus this EPR model can most probably give more reliable prediction of C_α on unseen data. Therefore, Eq. (11) is the optimal model in terms of robustness, followed by Eq.(12). These two formulas will be further examined.

4.4 Monotonicity and sensitivity analysis

The mathematical characteristics (e.g., monotonicity) of a formula can somehow imply whether it is physically correct or not. Thus, to deeply understand the mathematical characteristics of two proposed EPR models and select one of them as the optimum, a parametric study was conducted on the involving physical properties. Note that when one variable is being studied, the other two variables were fixed to their common values (CI=50 %, I_P=40 % and e=1.0) for Eq.(11) and (w_L=50 %, I_P=40 % and e=1.0) for Eq.(12). Fig. 9 and Fig. 10 show the results of parametric study for Eq.(11) and Eq.(12), respectively. For Eq.(11), it is found that with increasing the value of CI, the predicted values of C_α has a slight decrease; with increasing the value of I_P, the value of C_α increases quickly up to a point then increases slowly; However, for Eq.(12), the C_α decreases firstly and then increases with increasing w_L; with increasing the I_p, the C_α decreases, which is different from the investigation shown in Fig. 1 so that it is unreasonable. With increasing the void ratio, the C_α increases for both EPR models, which is in accordance with the findings by Yin et al. (2015b). Therefore, the EPR model involving CI and I_P with 3 terms for C_α is better in the monotonicity. Overall, this model was finally selected in terms of the predictive ability, model complexity, robustness and monotonicity.

To assess the importance of each input of the proposed EPR model on C_α, a sensitivity analysis was performed. The composite scaled sensitivity (CSS_j) analysis proposed by Hill (1998) was adopted, which indicates the amount of information provided by the i-th observations for the estimation of j-th parameter and is defined as:

$$CSS_j = \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}\left(\left(\frac{\partial y_i}{\partial x_j}\right)\cdot x_j\sqrt{\omega_i}\right)^2\right)} \qquad (14)$$

where y_i is the ith simulated value; x_j is the jth estimated parameter; $\partial y_i/\partial x_j$ is the sensitivity of the ith simulated value with respect to the jth parameter; N is the number of observations; ω_i is the weighting factor, which is related to the ith observation and can be evaluated based on the statistics (i.e. variance, or standard deviation, or coefficient of variation of the error of the observations). The composite scaled sensitivities indicate the total amount of information provided by the observations for the estimation of parameter j and measure the relative importance of the input parameters being simultaneously estimated.

To obtain a reliable sensitivity, the CSS_j was calculated on three different points for each involved variable and thus an average value was finally given. Fig. 11 shows the results of sensitivity analysis for C_α based on the proposed EPR model. The variable having the most significant influence on C_α is I_p, which has been highlighted by Yin (1999) and Nakase et al. (1988). The e and CI have a relatively minor important influence on predicting C_α. A slight higher sensitivity of CI well reveals the experimental results of Yin (1999).

4.5 Discussion

The CI is involved in the proposed EPR model, which implies the need of measurement of CI in laboratory. Comparing to Atterberg limits and void ratio, the measurement of CI is less conventional. Thus the need of CI will reduce the utility of the proposed model. Therefore, when the data of CI is not available, the EPR correlation (Eq.(12)) only involving Atterberg limits (w_L and I_p) can also be an alternative choice for predicting C_α although its monotonicity is worse.

In contrast to other techniques to obtain non-linear creep parameters of clays, such as trust-region reflective least squares algorithm (Le and Fatahi 2016; Le et al. 2015; Le et al. 2017), Simplex (Ye et al. 2016), genetic algorithm (Jin et al. 2017b; Yin et al. 2017a), the proposed EPR model only requires the basic physical information of soil samples and no additional laboratory tests (e.g., oedometer test, triaxial test) are needed. Moreover, the computational cost is less compared to numerous calculations on obtaining the fitness, sorting and selection for optimizations.

Since EPR is a data mining technique which is heavily dependent on the amount of data used, especially the range covered by the data, the formula obtained will be more applicable if more experimental data with wide range can be found and used. Currently, both EPR models are trained on limited data. Thus, their further performance needs more unseen data to verify. Moreover, since the EPR models are polynomial, it is inevitable to predict very unreasonable values on few special cases. As both models have high robustness ratios, this unreasonable probability should be slight. However, the robustness ratio is always smaller than 1, which means that the proposed ERP model fails in generating a reasonable value on some samples. Therefore, it is still necessary to pay attention when the predicted values are out of the proposed range of C_α.

For the applicability of proposed EPR model in real engineering practice, the basic physical properties (e.g., e, I_p and CI) of soil samples from the dominated soil layer can be easily measured in laboratory. Then, the intrinsic C_α (corresponding to reconstituted state) can be obtained using the proposed EPR model. This C_α is a key input parameter for many elasto-viscoplastic models (e.g. Kimoto and Oka, 2005; Yin et al., 2002; Yin et al., 2010, 2011), and then long-term performance of various real engineering structures (e.g. embankment, slope and tunnel) can be estimated by numerical simulations.

5 Conclusions

A simple, robust, and accurate EPR model for modelling C_α of reconstituted clays using physical properties has been proposed. Prior to EPR procedure, the database for training the EPR model was built, which contains clay content (CI), liquid limit (w_L), plastic limit (w_P), plasticity index (I_P), void ratio (e) and C_α. Based on the database, the statistical analysis and basic correlations between C_α and each physical property have then been conducted. The C_α is relatively highly correlated to e, followed by I_p and w_L, and very poorly correlated to w_P and CI.

To avoid overfitting and reduce the model complexity, a novel EPR procedure using a newly enhanced DE algorithm was proposed with two enhancements: (1) a new fitness function was proposed using the structural risk minimization (SRM) with L_2 regularization factor that penalizes polynomial complexity; (2) an adaptive process for selecting the combination of involved variables and size of polynomial terms was incorporated. The selection of regularization parameter was suggested.

To attain the nonlinear creep behaviour that the C_α consecutively decreases with increasing the soil density, a general structure of EPR expression for C_α was proposed, in which the CI, w_L and I_P were chosen as the dynamic correlating variables and the e was considered as a fixed variable. 120 data randomly selected in database were used for training and the remaining data were used for testing. The maximum size of terms was set to 8 and the total number of possible combinations of involved variables was 7.

Six EPR models with different variable combinations and size of model terms corresponding to different values of regularization parameter λ were firstly achieved. The performance of each model was compared using five indicators. Based on preliminary results, two EPR models with three terms were temporarily suggested as the optimal models. Then, a robustness testing was conducted on all obtained EPR models, in which the EPR model involving *CI* with three terms was selected as the optimal model. To deeply understand the mathematical characteristics of two proposed EPR models and select one of them as the optimum, a monotonicity analysis was conducted. Overall, the EPR model involving *CI* and I_P with 3 terms is finally recommended in terms of the predictive ability, model complexity, robustness and monotonicity. Hereafter, the sensitivity analysis of *CI*, I_p and e for the optimal model was carried out. The analysis results indicated that the I_p has the most significant influence on predicting C_α.

In the future, the proposed correlation will be applied to more engineering practices.

Acknowledgements

This research was financially supported by the National Natural Science Foundation of China (Grant No. 51579179).

Reference

Ahangar-Asr, A., Faramarzi, A. & Javadi, A.A. 2010. A new approach for prediction of the stability of soil and rock slopes. Engineering Computations, **27**, 878-893.

Ahangar-Asr, A., Faramarzi, A., Mottaghifard, N. & Javadi, A.A. 2011. Modeling of permeability and compaction characteristics of soils using evolutionary polynomial regression. Computers & Geosciences, **37**, 1860-1869.

Ahangar-Asr, A., Javadi, A.A., Johari, A. & Chen, Y. 2014. Lateral load bearing capacity modelling of piles in cohesive soils in undrained conditions: An intelligent evolutionary approach. Applied Soft Computing, **24**, 822-828.

Alemdag, S., Gurocak, Z., Cevik, A., Cabalar, A. & Gokceoglu, C. 2015. Modeling deformation modulus of a stratified sedimentary rock mass using neural network, fuzzy inference and genetic programming. Engineering Geology, **203**, 70-82, doi: 10.1016/j.enggeo.2015.12.002.

Anagnostopoulos, C. & Grammatikopoulos, I. 2011. A new model for the prediction of secondary compression index of soft compressible soils. Bulletin of Engineering Geology and the Environment, **70**, 423-427.

Cao, Z. & Wang, Y. 2014. Bayesian model comparison and selection of spatial correlation functions for soil parameters. Structural Safety, **49**, 10-17.

Chai, J.-C., Shen, J.S.-L., Liu, M.D. & Yuan, D.-J. 2018. Predicting the performance of embankments on PVD-improved subsoils. Computers and Geotechnics, **93**, 222-231.

Coelho, F. & Neto, J.P. 2017. A method for regularization of evolutionary polynomial regression. Applied Soft Computing, **59**, 223-228.

Doglioni, A., Crosta, G.B., Frattini, P., Melidoro, N.L. & Simeone, V. 2015. Predicting landslide displacements by multi-objective evolutionary polynomial regression. *Engineering Geology for Society and Territory-Volume 5*. Springer, 651-654.

Ebrahimian, B. & Movahed, V. 2013. Evaluation of Axial Bearing Capacity of Piles in Sandy Soils by CPT Results. Evaluation, **29**, 31.

Ebrahimian, B. & Movahed, V. 2017. Application of an evolutionary-based approach in evaluating pile bearing capacity using CPT results. Ships and Offshore Structures, **12**, 937-953.

Faramarzi, A., Alani, A.M. & Javadi, A.A. 2014. An EPR-based self-learning approach to material modelling. Computers & Structures, **137**, 63-71.

Feng, J., Chuhan, Z., Gang, W. & Guanglun, W. 2003. Creep modeling in excavation analysis of a high rock slope. Journal of Geotechnical and Geoenvironmental Engineering, **129**, 849-857.

Garg, A., Lam, J.S.L. & Panda, B. 2017. A hybrid computational intelligence framework in modelling of coal-oil agglomeration phenomenon. Applied Soft Computing, **55**, 402-412.

Ghorbani, A. & Firouzi Niavol, M. 2017. Evaluation of Induced Settlements of Piled Rafts in the Coupled Static-Dynamic Loads Using Neural Networks and Evolutionary Polynomial Regression. Applied Computational Intelligence and Soft Computing, **2017**.

Giustolisi, O. & Savic, D. 2006. A symbolic data-driven technique based on evolutionary polynomial regression. Journal of Hydroinformatics, **8**, 207-222.

Gurocak, Z., Alemdag, S. & Zaman, M.M. 2008. Rock slope stability and excavatability assessment of rocks at the Kapikaya dam site, Turkey. Engineering Geology, **96**, 17-27.

Gurocak, Z., Solanki, P., Alemdag, S. & Zaman, M.M. 2012. New considerations for empirical estimation of tensile strength of rocks. Engineering Geology, **145**, 1-8.

Han, J., Yao, Y. & Yin, Z. 2018. Influences of overconsolidation ratio on undrained creep behavior of overconsolidated saturated clay. Chinese Journal of Geotechnical Engineering, **40**, 426-430.

Javadi, A.A., Faramarzi, A. & Ahangar-Asr, A. 2012. Analysis of behaviour of soils under cyclic loading using EPR-based finite element method. Finite Elements in Analysis and Design, **58**, 53-65.

Jin, Y.-F., Wu, Z.-X., Yin, Z.-Y. & Shen, J.S. 2017a. Estimation of critical state-related formula in advanced constitutive modeling of granular material. Acta Geotechnica, **12**, 1329-1351, doi: 10.1007/s11440-017-0586-5.

Jin, Y.-F., Yin, Z.-Y., Riou, Y. & Hicher, P.-Y. 2017b. Identifying creep and destructuration related soil parameters by optimization methods. KSCE Journal of Civil Engineering, **21**, 1123-1134, doi: 10.1007/s12205-016-0378-8.

Jin, Y.-F., Yin, Z.-Y., Shen, S.-L. & Hicher, P.-Y. 2016a. Investigation into MOGA for identifying parameters of a critical-state-based sand model and parameters correlation by factor analysis. Acta Geotechnica, **11**, 1131-1145, doi: 10.1007/s11440-015-0425-5.

Jin, Y.-F., Yin, Z.-Y., Shen, S.-L. & Hicher, P.-Y. 2016b. Selection of sand models and identification of parameters using an enhanced genetic algorithm. International Journal for Numerical and Analytical Methods in Geomechanics, **40**, 1219-1240, doi: 10.1002/nag.2487.

Jin, Y.-F., Yin, Z.-Y., Shen, S.-L. & Zhang, D.-M. 2017c. A new hybrid real-coded genetic algorithm and its application to parameters identification of soils. Inverse Problems in Science and Engineering, **25**, 1343-1366, doi: 10.1080/17415977.2016.1259315.

Jin, Y.-F., Yin, Z.-Y., Wu, Z.-X. & Zhou, W.-H. 2018. Identifying parameters of easily crushable sand and application to offshore pile driving. Ocean Engineering, **154**, 416-429, doi: 10.1016/j.oceaneng.2018.01.023.

Jin, Y.-F., Yin, Z.-Y., Zhou, W.-H. & Huang, H.-W. 2019. Multi-objective optimization-based updating of predictions during excavation. Engineering Applications of Artificial Intelligence, 78, 102-123, doi: https://doi.org/10.1016/j.engappai.2018.11.002.

Kakoudakis, K., Behzadian, K., Farmani, R. & Butler, D. 2017. Pipeline failure prediction in water distribution networks using evolutionary polynomial regression combined with K-means clustering. Urban Water Journal, **14**, 737-742.

Kalinli, A., Acar, M.C. & Gündüz, Z. 2011. New approaches to determine the ultimate bearing capacity of shallow foundations based on artificial neural networks and ant colony optimization. Engineering Geology, **117**, 29-38, doi: https://doi.org/10.1016/j.enggeo.2010.10.002.

Karstunen, M. & Yin, Z.Y. 2010. Modelling time-dependent behaviour of Murro test embankment. Géotechnique, **60**, 735-749.

Khoshkroudi, S.S., Sefidkouhi, M.A.G., Ahmadi, M.Z. & Ramezani, M. 2014. Prediction of soil saturated water content using evolutionary polynomial regression (EPR). Archives of Agronomy and Soil Science, **60**, 1155-1172.

Kimoto, S. & Oka, F. 2005. An elasto-viscoplastic model for clay considering destructuralization and consolidation analysis of unstable behavior. Soils and Foundations, **45**, 29-42.

Le, T.M. & Fatahi, B. 2016. Trust-region reflective optimisation to obtain soil visco-plastic properties. Engineering Computations, **33**, 410-442.

Le, T.M., Fatahi, B. & Khabbaz, H. 2015. Numerical optimisation to obtain elastic viscoplastic model parameters for soft clay. International Journal of Plasticity, **65**, 1-21.

Le, T.M., Fatahi, B., Khabbaz, H. & Sun, W. 2017. Numerical optimization applying trust-region reflective least squares algorithm with constraints to optimize the non-linear creep parameters of soft soil. Applied Mathematical Modelling, **41**, 236-256.

Li, Q., Ng, C.W.W. & Liu, G.-b. 2012. Low secondary compressibility and shear strength of Shanghai Clay. Journal of Central South University, **19**, 2323-2332.

Mesri, G. & Godlewski, P.M. 1977. Time and stress-compressibility interrelationship. Journal of the Geotechnical Engineering Division, **103**, 417-430.

Nakase, A. & Kamei, T. 1986. Influence of strain rate on undrained shear characteristics of K0-consolidated cohesive soils. Soils and Foundations, **26**, 85-95.

Nakase, A., Kamei, T. & Kusakabe, O. 1988. Constitutive parameters estimated by plasticity index. Journal of geotechnical engineering, **114**, 844-858.

Nassr, A., Javadi, A. & Faramarzi, A. 2018. Developing constitutive models from EPR‐based self‐learning finite element analysis. International Journal for Numerical and Analytical Methods in Geomechanics, **42**, 401-417.

Ng, A.Y. 2004. Feature selection, L 1 vs. L 2 regularization, and rotational invariance. *Proceedings of the twenty-first international conference on Machine learning*. ACM, 78.

Qu, G., Hinchberger, S. & Lo, K. 2010. Evaluation of the viscous behaviour of clay using generalised overstress viscoplastic theory. Géotechnique, **60**, 777-789.

Ren, D.-J., Shen, S.-L., Arulrajah, A. & Wu, H.-N. 2017. Evaluation of ground loss ratio with moving trajectories induced in DOT tunnelling. Canadian Geotechnical Journal.

Rezania, M., Bagheri, M., Nezhad, M.M. & Sivasithamparam, N. 2017. Creep analysis of an earth embankment on soft soil deposit with and without PVD improvement. Geotextiles and Geomembranes.

Rezania, M., Faramarzi, A. & Javadi, A.A. 2011. An evolutionary based approach for assessment of earthquake-induced soil liquefaction and lateral displacement. Engineering Applications of Artificial Intelligence, **24**, 142-153.

Rezania, M., Javadi, A.A. & Giustolisi, O. 2010. Evaluation of liquefaction potential based on CPT results using evolutionary polynomial regression. Computers and Geotechnics, **37**, 82-92.

Shahin, M.A. 2014. State-of-the-art review of some artificial intelligence applications in pile foundations. Geoscience Frontiers, **7**, 33-44, doi: 10.1016/j.gsf.2014.10.002.

Shahnazari, H., Shahin, M.A. & Tutunchian, M.A. 2014. Evolutionary-based approaches for settlement prediction of shallow foundations on cohesionless soils. Geotech Eng, **12**, 55-64.

Shahnazari, H., Tutunchian, M.A., Rezvani, R. & Valizadeh, F. 2013. Evolutionary-based approaches for determining the deviatoric stress of calcareous sands. Computers & Geosciences, **50**, 84-94.

Shen, S.-L., Ma, L., Xu, Y.-S. & Yin, Z.-Y. 2013. Interpretation of increased deformation rate in aquifer IV due to groundwater pumping in Shanghai. Canadian Geotechnical Journal, **50**, 1129-1142.

Shen, S.-L., Wu, H.-N., Cui, Y.-J. & Yin, Z.-Y. 2014. Long-term settlement behaviour of metro tunnels in the soft deposits of Shanghai. Tunnelling and Underground Space Technology, **40**, 309-323.

Shen, S.L., Chai, J.C., Hong, Z.S. & Cai, F.X. 2005. Analysis of field performance of embankments on soft clay deposit with and without PVD-improvement. Geotextiles and Geomembranes, **23**, 463-485.

Shen, S.L. & Xu, Y.S. 2011. Numerical evaluation of land subsidence induced by groundwater pumping in Shanghai. Canadian Geotechnical Journal, **48**, 1378-1392.

Suneel, M., Park, L.K. & Im, J.C. 2008. Compressibility characteristics of Korean marine clay. Marine Georesources and Geotechnology, **26**, 111-127.

Wang, J., Xu, Z. & Wang, W. 2009. Wall and ground movements due to deep excavations in Shanghai soft soils. Journal of Geotechnical and Geoenvironmental Engineering, **136**, 985-994.

Wood, D.M. 2003. Geotechnical modelling. CRC Press.

Wu, H.-N., Shen, S.-L., Liao, S.-M. & Yin, Z.-Y. 2015. Longitudinal structural modelling of shield tunnels considering shearing dislocation between segmental rings. Tunnelling and Underground Space Technology, **50**, 317-323.

Wu, H.-N., Shen, S.-L. & Yang, J. 2017. Identification of tunnel settlement caused by land subsidence in soft deposit of Shanghai. Journal of Performance of Constructed Facilities, **31**, 04017092.

Wu, Z.-x., Ji, H., Yu, C. & Zhou, C. 2018. EPR-RCGA-based modelling of compression index and RMSE-AIC-BIC-based model selection for Chinese marine clays and their engineering application. Journal of Zhejiang University-SCIENCE A, **19**, 211-224.

Xu, Y.-S., Shen, S.-L., Ren, D.-J. & Wu, H.-N. 2016. Analysis of factors in land subsidence in Shanghai: a view based on a strategic environmental assessment. Sustainability, **8**, 573.

Xu, Y.S., Ma, L., Du, Y.J. & Shen, S.L. 2012a. Analysis of urbanisation-induced land subsidence in Shanghai. Natural hazards, 1-13.

Xu, Y.S., Ma, L., Shen, S.L. & Sun, W.J. 2012b. Evaluation of land subsidence by considering underground structures that penetrate the aquifers of Shanghai, China. Hydrogeology Journal, 1-12.

Ye, L., Jin, Y.-F., Shen, S.-L., Sun, P.-P. & Zhou, C. 2016. An efficient parameter identification procedure for soft sensitive clays. Journal of Zhejiang University SCIENCE A, **17**, 76-88, doi: 10.1631/jzus.A1500031.

Yin, J.-H. 1999. Properties and behaviour of Hong Kong marine deposits with different clay contents. Canadian Geotechnical Journal, **36**, 1085-1095.

Yin, J.-H. & Graham, J. 1989. Viscous–elastic–plastic modelling of one-dimensional time-dependent behaviour of clays. Canadian Geotechnical Journal, **26**, 199-209.

Yin, J.-H., Zhu, J.-G. & Graham, J. 2002. A new elastic viscoplastic model for time-dependent behaviour of normally and overconsolidated clays: theory and verification. Canadian Geotechnical Journal, **39**, 157-173.

Yin, J.H. & Cheng, C.M. 2006. Comparison of Strain-rate Dependent Stress-Strain Behavior from K o-consolidated Compression and Extension Tests on Natural Hong Kong Marine Deposits. Marine Georesources and Geotechnology, **24**, 119-147.

Yin, Z.-Y., Jin, Y.-F., Huang, H.-W. & Shen, S.-L. 2016. Evolutionary polynomial regression based modelling of clay compressibility using an enhanced hybrid real-coded genetic algorithm. Engineering Geology, **210**, 158-167.

Yin, Z.-Y., Jin, Y.-F., Shen, J.S. & Hicher, P.-Y. 2018. Optimization techniques for identifying soil parameters in geotechnical engineering: Comparative study and enhancement. International Journal for Numerical and Analytical Methods in Geomechanics, **42**, 70-94, doi: 10.1002/nag.2714.

Yin, Z.-Y., Jin, Y.-F., Shen, S.-L. & Huang, H.-W. 2017a. An efficient optimization method for identifying parameters of soft structured clay by an enhanced genetic algorithm and elastic–viscoplastic model. Acta Geotechnica, **12**, 849-867, doi: 10.1007/s11440-016-0486-0.

Yin, Z.-Y., Yin, J.-H. & Huang, H.-W. 2015a. Rate-Dependent and Long-Term Yield Stress and Strength of Soft Wenzhou Marine Clay: Experiments and Modeling. Marine Georesources & Geotechnology, **33**, 79-91.

Yin, Z.-Y., Zhu, Q.-Y., Yin, J.-H. & Ni, Q. 2014. Stress relaxation coefficient and formulation for soft soils. Géotechnique Letters, **4**, 45-51.

Yin, Z.-Y., Zhu, Q.-Y. & Zhang, D.-M. 2017b. Comparison of two creep degradation modeling approaches for soft structured soils. Acta Geotechnica, 1-19, doi: 10.1007/s11440-017-0556-y.

Yin, Z., Chang, C., Hicher, P. & Karstunen, M. 2008. Microstructural Modeling of Rate-dependent Behavior of Soft Soil. Proceeding of 12th IACMAG~ Goa, 862-868.

Yin, Z., Karstunen, M., Wang, J. & Yu, C. 2011a. Influence of features of natural soft clay on behaviour of embankment. Journal of Central South University of Technology, **18**, 1667-1676.

Yin, Z., Xu, Q. & Yu, C. 2015b. Elastic-Viscoplastic Modeling for Natural Soft Clays Considering Nonlinear Creep. International Journal of Geomechanics, **15**, A6014001, doi: doi:10.1061/(ASCE)GM.1943-5622.0000284.

Yin, Z.Y. & Chang, C.S. 2009. Microstructural modelling of stress-dependent behaviour of clay. International Journal of Solids and Structures, **46**, 1373-1388.

Yin, Z.Y., Chang, C.S., Karstunen, M. & Hicher, P.Y. 2010a. An anisotropic elastic-viscoplastic model for soft clays. International Journal of Solids and Structures, **47**, 665-677.

Yin, Z.Y. & Karstunen, M. 2011. Modelling strain-rate-dependency of natural soft clays combined with anisotropy and destructuration. Acta Mechanica Solida Sinica, **24**, 216-230.

Yin, Z.Y., Karstunen, M., Chang, C.S., Koskinen, M. & Lojander, M. 2011b. Modeling Time-Dependent Behavior of Soft Sensitive Clay. Journal of Geotechnical and Geoenvironmental Engineering, **137**, 1103-1113, doi: 10.1061/(asce)gt.1943-5606.0000527.

Yin, Z.Y., Karstunen, M. & Hicher, P.Y. 2010b. Evaluation of the influence of elasto-viscoplastic scaling functions on modelling time-dependent behaviour of natural clays. Soils and Foundations, **50**, 203-214, doi: 10.3208/sandf.50.203.

Yin, Z.Y. & Wang, J.H. 2012. A one-dimensional strain-rate based model for soft structured clays. Science China-Technological Sciences, **55**, 90-100, doi: 10.1007/s11431-011-4513-y.

Zeng, L., Hong, Z., Liu, S. & Chen, F. 2012a. Variation law and quantitative evaluation of secondary consolidation behavior for remolded clays. Chinese Journal of Geotechnical Engineering, **34**, 1496-1500.

Zeng, L., Hong, Z., Liu, S. & Chen, F. 2012b. Variation law and quantitative evaluation of secondary consolidation behavior for remolded clays. Chin J Geotech Eng, **34**, 1496-1500.

Zeng, L. & Liu, S. 2010. A Calculation Method of Secondary Compression Index for Natural Sedimentary Clays Using Void Index. *Geo-Shanghai International Conference. Shanghai*, 14-21.

Zhang, J.-F., Chen, J.-J., Wang, J.-H. & Zhu, Y.-F. 2013. Prediction of tunnel displacement induced by adjacent excavation in soft soil. Tunnelling and Underground Space Technology, **36**, 24-33.

Zhang, J., Zhang, L. & Tang, W.H. 2009. Bayesian framework for characterizing geotechnical model uncertainty. Journal of Geotechnical and Geoenvironmental Engineering, **135**, 932-940.

Zhang, L., Li, D.-Q., Tang, X.-S., Cao, Z.-J. & Phoon, K.-K. 2017. Bayesian model comparison and characterization of bivariate distribution for shear strength parameters of soil. Computers and Geotechnics, **95**, 110-118.

Zhang, W., Goh, A.T.C., Zhang, Y., Chen, Y. & Xiao, Y. 2015. Assessment of soil liquefaction based on capacity energy concept and multivariate adaptive regression splines. Engineering Geology, **188**, 29-37, doi: https://doi.org/10.1016/j.enggeo.2015.01.009.

Zhou, C., Yin, K., Cao, Y. & Ahmed, B. 2016. Application of time series analysis and PSO–SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China. Engineering Geology, **204**, 108-120, doi: https://doi.org/10.1016/j.enggeo.2016.02.009.

Zhu, G. & Yin, J.H. 2000. Elastic visco‐plastic consolidation modelling of clay foundation at Berthierville test embankment. International Journal for Numerical and Analytical Methods in Geomechanics, **24**, 491-508.

Zhu, Q.-Y., Jin, Y.-F., Yin, Z.-Y. & Hicher, P.-Y. 2013. Influence of natural deposition plane orientation on oedometric consolidation behavior of three typical clays from southeast coast of China. Journal of Zhejiang University SCIENCE A, **14**, 767-777.

Zhu, Q.-Y., Wu, Z.-X., Li, Y.-L., Xu, C.-J., Wang, J.-H. & Xia, X.-H. 2014. A modified creep index and its application to viscoplastic modelling of soft clays. Journal of Zhejiang University SCIENCE A, **15**, 272-281.

Zhu, Q.-Y., Yin, Z.-Y., Hicher, P.-Y. & Shen, S.-L. 2016. Nonlinearity of one-dimensional creep characteristics of soft clays. Acta Geotechnica, **11**, 887-900.

Zhu, Q.-Y., Yin, Z.-Y., Xu, C.-J., Yin, J.-H. & Xia, X.-H. 2015. Uniqueness of rate-dependency, creep and stress relaxation behaviors for soft clays. Journal of Central South University, **22**, 296-302.

Tables

Table 1 Summary of physical properties and creep index for all selected clays

Clay	CI /%	w_L /%	w_P /%	I_P /%	C_α	e	Reference
Haarajoki clay	65	88	26	62	0.0461	2.283	(Yin et al. 2015b; Zhu et al. 2016)
	65	88	26	62	0.0299	2.006	
	65	88	26	62	0.0226	1.733	
	65	88	26	62	0.0154	1.486	
	65	88	26	62	0.0140	1.273	
	65	88	26	62	0.0297	2.044	
	65	88	26	62	0.0329	1.779	
	65	88	26	62	0.0244	1.545	
	65	88	26	62	0.0173	1.338	
	65	88	26	62	0.0113	1.141	
	65	88	26	62	0.0332	2.058	
	65	88	26	62	0.0258	1.831	
	65	88	26	62	0.0187	1.602	
	65	88	26	62	0.0154	1.395	
	65	88	26	62	0.0108	1.196	
Suurpelto clay	80	80	23	57	0.0467	1.881	(Yin et al. 2015b; Zhu et al. 2016)
	80	80	23	57	0.0267	1.558	
	80	80	23	57	0.0166	1.092	
	80	80	23	57	0.0484	1.862	
	80	80	23	57	0.0203	1.635	
	80	80	23	57	0.0196	1.430	
	80	80	23	57	0.0143	1.272	
	80	80	23	57	0.0161	1.127	
	80	80	23	57	0.0548	2.080	
	80	80	23	57	0.0279	1.831	
	80	80	23	57	0.0196	1.579	
	80	80	23	57	0.0159	1.356	
	80	80	23	57	0.0131	1.147	
	80	80	23	57	0.0516	2.035	
	80	80	23	57	0.0288	1.782	
	80	80	23	57	0.0138	1.311	
	80	80	23	57	0.0117	1.104	
Mixed clay	78	45	26	19	0.0069	1.040	(Yin et al. 2015b; Zhu et al. 2016)
	78	45	26	19	0.0053	0.963	
	78	45	26	19	0.0062	0.883	

	78	45	26	19	0.0041	0.803	
	78	45	26	19	0.0037	0.729	
	78	45	26	19	0.0071	1.063	
	78	45	26	19	0.0053	0.809	
	78	45	26	19	0.0076	1.057	
	78	45	26	19	0.0062	0.982	
	78	45	26	19	0.0069	0.918	
	78	45	26	19	0.0051	0.843	
	78	45	26	19	0.0051	0.761	
	78	45	26	19	0.0058	0.997	
	78	45	26	19	0.0055	0.929	
	78	45	26	19	0.0051	0.858	
	78	45	26	19	0.0051	0.790	
Vanttila clay	53	98	30	68	0.0447	1.926	(Yin et al. 2015b; Zhu et al. 2016)
	53	98	30	68	0.0336	1.498	
	53	98	30	68	0.0315	1.287	
	53	98	30	68	0.0265	1.165	
	53	98	30	68	0.0272	1.075	
	53	98	30	68	0.0451	1.908	
	53	98	30	68	0.0230	1.677	
	53	98	30	68	0.0345	1.460	
	53	98	30	68	0.0249	1.252	
	53	98	30	68	0.0235	1.062	
	53	98	30	68	0.0357	1.942	
	53	98	30	68	0.0338	1.726	
	53	98	30	68	0.0239	1.518	
	53	98	30	68	0.0246	1.313	
	53	98	30	68	0.0196	1.121	
	53	98	30	68	0.0212	1.019	
	53	98	30	68	0.0302	1.999	
	53	98	30	68	0.0253	1.825	
Murro clay	26	88	34	54	0.0359	1.684	(Yin et al. 2015b; Zhu et al. 2016)
	26	88	34	54	0.0302	1.536	
	26	88	34	54	0.0269	1.385	
	26	88	34	54	0.0286	1.231	
	26	88	34	54	0.0233	1.099	
	26	88	34	54	0.0207	0.954	
	26	88	34	54	0.0311	1.691	
	26	88	34	54	0.0283	1.536	
	26	88	34	54	0.0281	1.4101	
	26	88	34	54	0.0237	1.260	

	26	88	34	54	0.0214	1.123	
	26	88	34	54	0.0233	1.394	
	26	88	34	54	0.0228	1.246	
	26	88	34	54	0.0184	1.107	
	26	88	34	54	0.0184	0.980	
	26	88	34	54	0.0269	1.446	
	26	88	34	54	0.0235	1.294	
	26	88	34	54	0.0228	1.145	
	26	88	34	54	0.0249	0.997	
	26	88	34	54	0.0258	1.411	
	26	88	34	54	0.0262	1.331	
	26	88	34	54	0.0214	1.183	
	26	88	34	54	0.0196	1.042	
Kaolin clay	83	40	20	20	0.0060	0.913	(Zhu et al. 2016)
	83	40	20	20	0.0065	0.841	
	83	40	20	20	0.0062	0.766	
	83	40	20	20	0.0060	0.689	
	83	40	20	20	0.0058	0.598	
Shanghai clay-1	33	51	26.4	24.6	0.0086	0.949	(Li et al. 2012)
	33	51	26.4	24.6	0.0083	0.857	
	33	51	26.4	24.6	0.0076	0.746	
	33	51	26.4	24.6	0.0072	0.680	
Shanghai clay-2	26	42.5	22.5	20	0.0076	0.861	(Zhu et al. 2016)
	26	42.5	22.5	20	0.0074	0.763	
	26	42.5	22.5	20	0.0069	0.671	
	26	42.5	22.5	20	0.0062	0.585	
Nanjing clay-9m	34	52	26	26	0.0154	1.015	(Zeng et al. 2012b)
	34	52	26	26	0.0142	0.915	
	34	52	26	26	0.0124	0.805	
	34	52	26	26	0.0116	0.699	
	34	52	26	26	0.0107	0.605	
	34	52	26	26	0.0093	0.518	
Wenzhou clay-10m	47.6	65	28	37	0.0228	1.334	(Zeng et al. 2012b)
	47.6	65	28	37	0.0212	1.207	
	47.6	65	28	37	0.0200	1.077	
	47.6	65	28	37	0.0184	0.938	
	47.6	65	28	37	0.0173	0.799	
	47.6	65	28	37	0.0152	0.654	
Lianyungang clay-12m	38	63	27	36	0.0179	1.101	(Zeng et al. 2012b)
	38	63	27	36	0.0147	0.834	
	38	63	27	36	0.0142	0.700	
Zhoushan clay	37	40.7	26.7	20	0.0071	0.882	(Zhu et al. 2013)

	37	40.7	26.7	20	0.0074	0.778	
	37	40.7	26.7	20	0.0069	0.669	
	37	40.7	26.7	20	0.0069	0.574	
	37	40.7	26.7	20	0.0058	0.482	
HKMC	27.5	60	28	32	0.0071	1.158	(Yin 1999)
	27.5	60	28	32	0.0046	0.894	
	27.5	60	28	32	0.0036	0.730	
	27.5	60	28	32	0.0034	0.610	
	27.5	60	28	32	0.0024	0.466	
Nanjing clay-7m	11.5	44	23	21	0.0111	0.901	(Zeng et al. 2012b)
	11.5	44	23	21	0.0100	0.846	
	11.5	44	23	21	0.0091	0.786	
	11.5	44	23	21	0.0089	0.704	
	11.5	44	23	21	0.0086	0.644	
	11.5	44	23	21	0.0074	0.562	
Wenzhou clay-4m	40.7	60	28	32	0.0208	1.404	(Zeng et al. 2012b)
	40.7	60	28	32	0.0192	1.289	
	40.7	60	28	32	0.0176	1.169	
	407	60	28	32	0.0164	1.032	
	40.7	60	28	32	0.0153	0.901	
	40.7	60	28	32	0.0144	0.786	
	40.7	60	28	32	0.0134	0.639	
Lianyungang clay-4m	40	86	31	55	0.0347	1.704	(Zeng et al. 2012b)
	40	86	31	55	0.0314	1.502	
	40	86	31	55	0.0284	1.289	
Kyuhoji clay	70	1.02	78.3	28.2	0.0320	1.02	(Kimoto and Oka 2005)
Illite clay	61	1.354	58	26	0.0092	1.354	(Yin and Graham 1989)
Merville clay	26	1.223	99	40	0.0158	1.223	(Han et al. 2018)
Kawasaki clay	22.3	1.07	55.3	25.9	0.0134	1.07	(Nakase and Kamei 1986)
Boston blue clay	57.6	1.181	45.4	21.7		1.1801	(Qu et al. 2010)

Table 2 Statistics of properties in the database

Properties	*CI* /%	w_L /%	w_P /%	I_p /%	*e*	C_α

Maximum value	83	98	34	68	2.284	0.0548
Minimum value	11.5	40	20	19	0.466	0.0037
Mean value	50.3	70.6	27.4	43.4	1.175	0.0182
Standard deviation	22.0	20.5	3.8	18.3	0.415	0.0110

Table 3 Summary of existing empirical correlations for C_α

Empirical correlations	Applicability	Reference
$C_\alpha = 0.00168 + 0.00033 I_p$	Remould clays	(Nakase et al. 1988)
$C_\alpha = 0.000369 I_p - 0.00055$	Remould clays	(Yin 1999)
$C_\alpha = \left(-0.0067 + 0.0115 e_L - 0.0016 (e_L)^2\right)(1+e)$	Remould clays	(Zeng et al. 2012a)
$C_\alpha = \left(-0.0274 + 0.0011 w_L - 0.00048 I_p\right)\left(\frac{w}{w_L}\right)^{0.7872-0.0369 w_L + 0.0619 I_p}$	Remould clays	(Zhu et al. 2016)
$C_\alpha = \left(0.0007 w_L - 0.0223\right)\left(\frac{w}{w_L}\right)^{0.014978 w_L - 0.23031}$	Remould clays	(Zhu et al. 2016)

where e_L is the void ratio corresponding to liquid limit; w is water content, equivalent to e for a given clay.

Table 4 Optimal correlations of C_α for different values of λ

λ	Proposed optimal correlation
0	$\ln(C_\alpha) = \left(0.9092 \frac{CI^2}{w_L I_P} + 0.6283 \frac{CI^2}{w_L^2 I_P} - 10.3212 (I_P CI)^2 - 0.1963 \left(\frac{CI}{w_L I_P}\right)^2 + 3.9741 (I_P)^2\right) e - 5.1618$ (10)
0.0001	$\ln(C_\alpha) = \left(0.3114 \frac{I_p^2}{CI} - 0.1229 \frac{1}{I_p^2} + 0.6455 \frac{1}{I_P}\right) e - 5.1308$ (11)

0.001 and 0.01	$\ln(C_\alpha)=\left(0.1265\frac{1}{w_L^2 I_P}-0.2463\frac{1}{I_p^2}+0.6264\left(\frac{w_L}{I_P}\right)^2\right)e-5.1098$	(12)
0.05 and 0.1	$\ln(C_\alpha)=\left(-0.0877\frac{1}{w_L I_p^2}+0.1998\left(\frac{w_L}{I_P}\right)^2+0.1497\frac{1}{w_L^2 I_p}+0.3846\frac{(w_L)^2}{I_P}\right)e-5.1660$	(13)

Remark: *CI*, w_L and I_P are in real number, not in percentage.

Table 5 Summary of indicators for all calculations of C_α with different values of λ

λ	*Comb*	*m*	Training					Testing				
			R^2	RMSE	MAE	*u*	σ	R^2	RMSE	MAE	*u*	σ
0	7	6	0.924	0.0042	0.0028	1.0464	0.0042	0.889	0.0053	0.0035	1.0284	0.1983
0.0001	4	3	0.896	0.0045	0.0032	1.0226	0.2204	0.863	0.0055	0.0036	1.0035	0.2218
0.001	6	3	0.892	0.0047	0.0032	1.0233	0.2219	0.853	0.0057	0.0038	1.0215	0.2219
0.01	6	3	0.892	0.0047	0.0032	1.0233	0.2219	0.853	0.0057	0.0038	1.0215	0.2219
0.05	6	4	0.895	0.0045	0.0031	1.0229	0.2221	0.857	0.0055	0.0037	1.013	0.2209
0.1	6	4	0.895	0.0045	0.0031	1.0229	0.2221	0.857	0.0055	0.001	1.013	0.2209

Figure captions

Fig. 1 Basic correlations between C_α and each physical property of soils

Fig. 2 Comparison between predictions and measurements for five empirical correlations

Fig. 3 Typical flowchart of EPR procedure

Fig. 4 Flowchart of NMDE

Fig. 5 Procedure of model selection combined with EPR process

Fig. 6 Evolution of model selection in terms of variable combination and size of terms

Fig. 7 Comparison of C_α between measurements and EPR predictions for different values of λ

Fig. 8 Distribution of C_α located in reasonable range in robustness testing

Fig. 9 Results of the C_α computed by Eq.(11) against (a) clay content, (b) plasticity index, and (c) void ratio

Fig. 10 Results of the C_α computed by Eq.(12) against (a) liquid limit, (b) plasticity index, and (c) void ratio

Fig. 11 Results of sensitivity analysis for EPR model of C_α

Figure 1

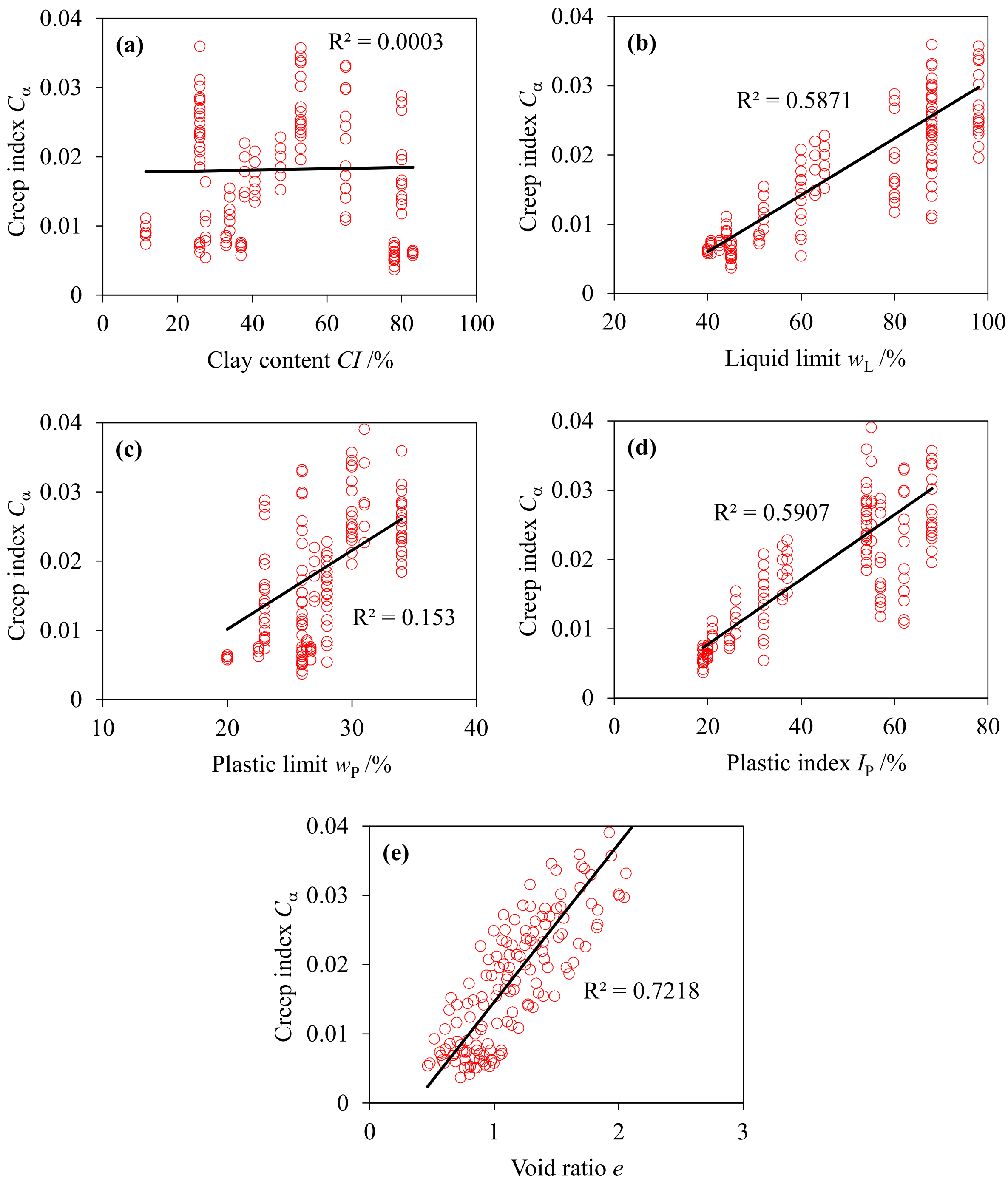

(a)
R² = 0.0003
Creep index Cα
Clay content CI /%
(b)
R² = 0.5871
Creep index Cα
Liquid limit wL /%
(c)
R² = 0.153
Creep index Cα
Plastic limit wP /%
(d)
R² = 0.5907
Creep index Cα
Plastic index IP /%
(e)
R² = 0.7218
Creep index Cα
Void ratio e

Figure 2

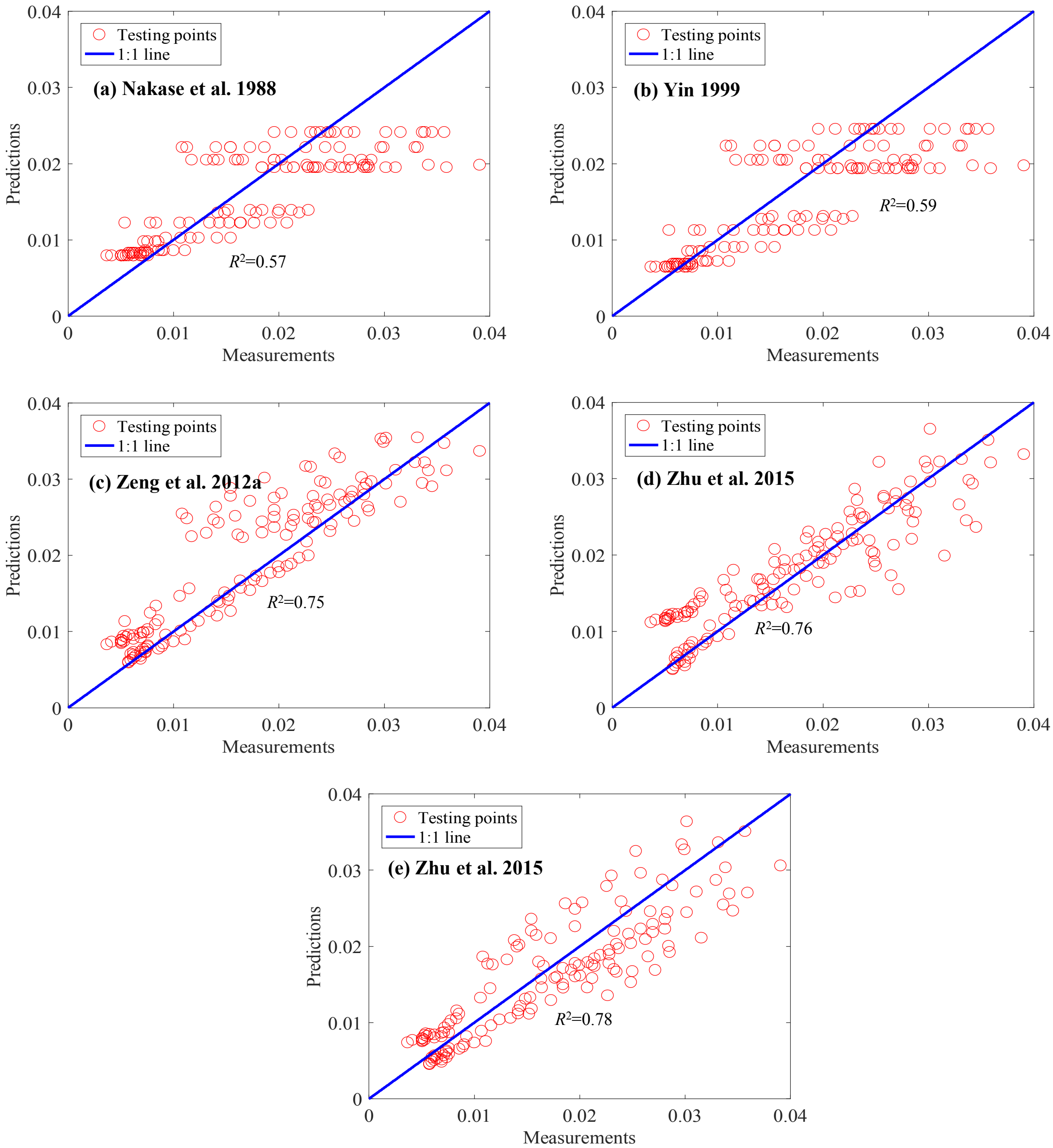

Testing points
1:1 line
(a) Nakase et al. 1988
R²=0.57
(b) Yin 1999
R²=0.59
(c) Zeng et al. 2012a
R²=0.75
(d) Zhu et al. 2015
R²=0.76
(e) Zhu et al. 2015
R²=0.78
Predictions
Measurements
0
0.01
0.02
0.03
0.04

Figure 3

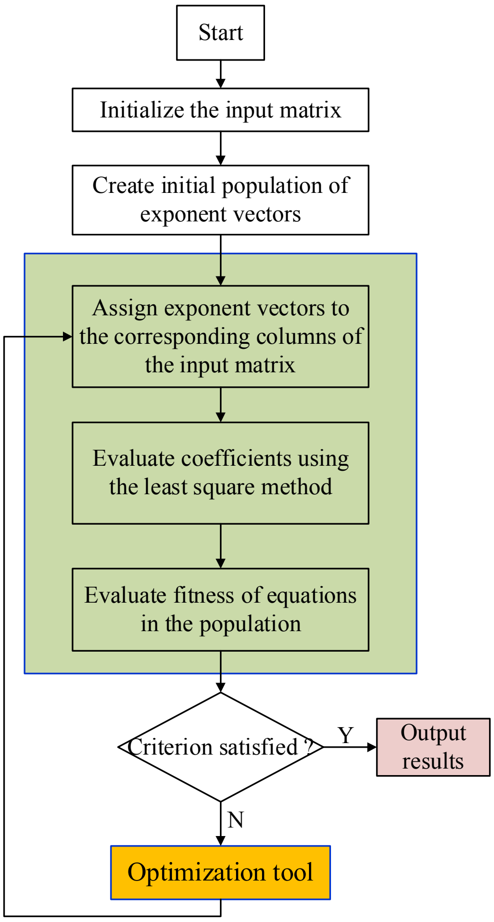
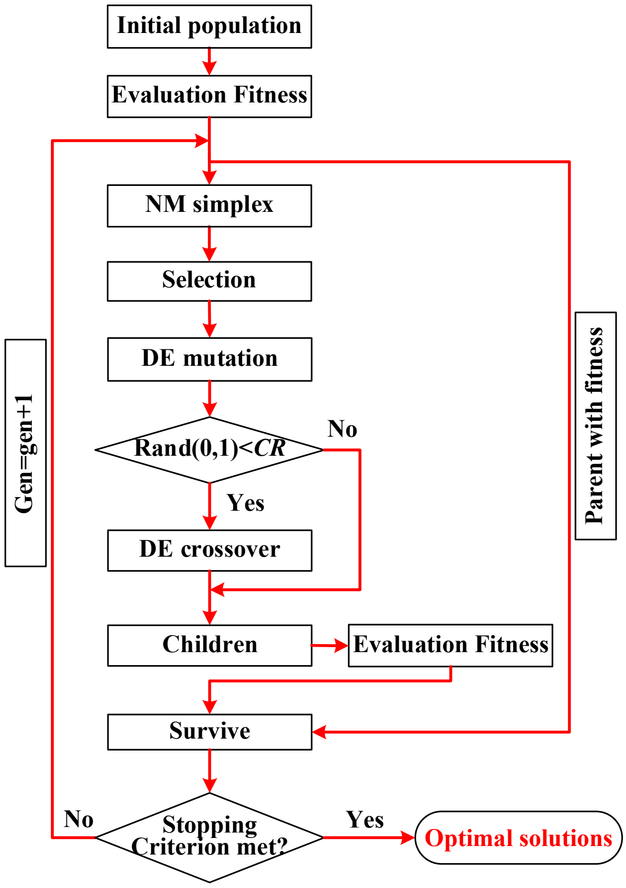
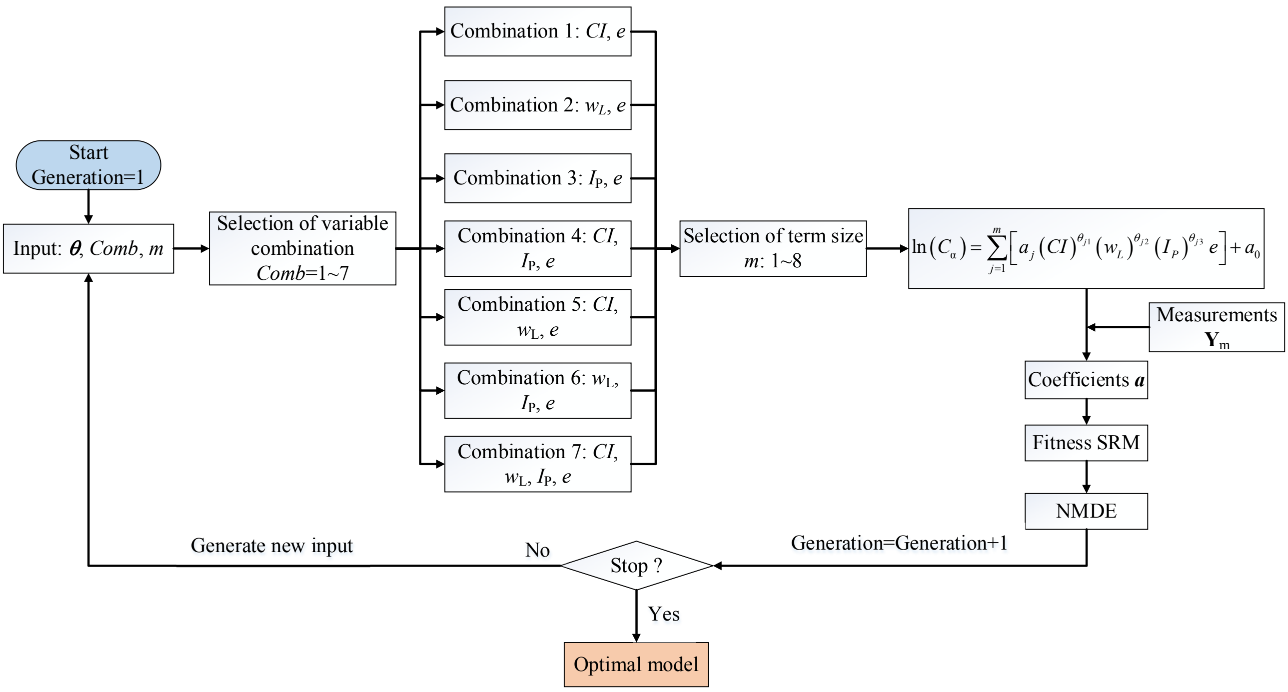

Start
Initialize the input matrix
Create initial population of exponent vectors
Assign exponent vectors to the corresponding columns of the input matrix
Evaluate coefficients using the least square method
Evaluate fitness of equations in the population
Criterion satisfied ?
Y
Output results
N
Optimization tool

Figure 4



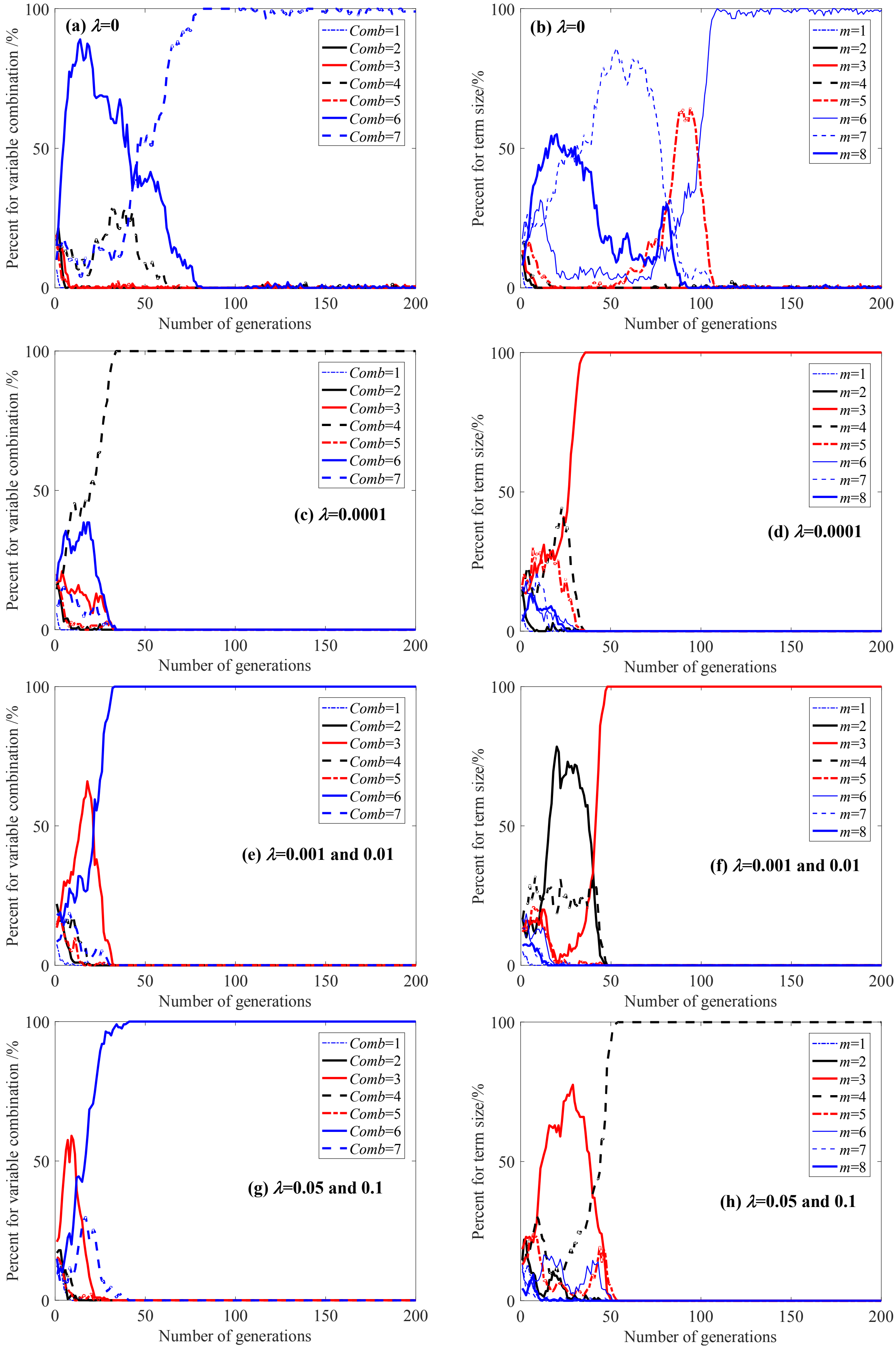
Initial population
Evaluation Fitness
NM simplex
Selection
DE mutation
Rand(0,1)<CR
No
Yes
DE crossover
Children
Evaluation Fitness
Survive
Stopping Criterion met?
No
Yes
Optimal solutions
Gen=gen+1
Parent with fitness

Figure 5



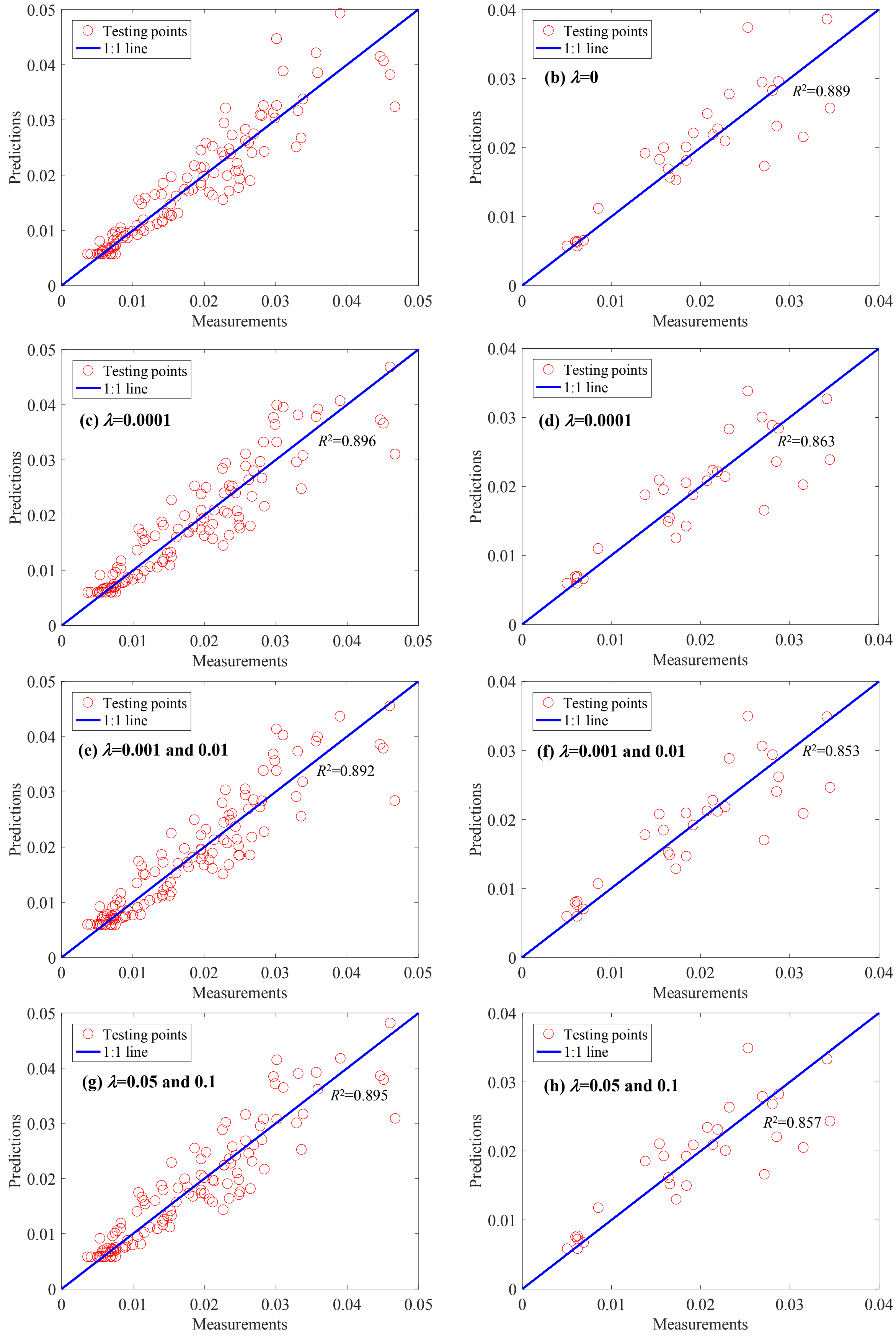
Start
Generation=1
Input: θ, Comb, m
Selection of variable combination Comb=1~7
Combination 1: CI, e
Combination 2: wL, e
Combination 3: IP, e
Combination 4: CI, IP, e
Combination 5: CI, wL, e
Combination 6: wL, IP, e
Combination 7: CI, wL, IP, e
Selection of term size m: 1~8
$\ln(C_\alpha) = \sum_{j=1}^{m}\left[a_j (CI)^{\theta_{j1}} (w_L)^{\theta_{j2}} (I_P)^{\theta_{j3}} e\right] + a_0$
Measurements Ym
Coefficients a
Fitness SRM
NMDE
Generation=Generation+1
Stop ?
No
Generate new input
Yes
Optimal model

Figure 6



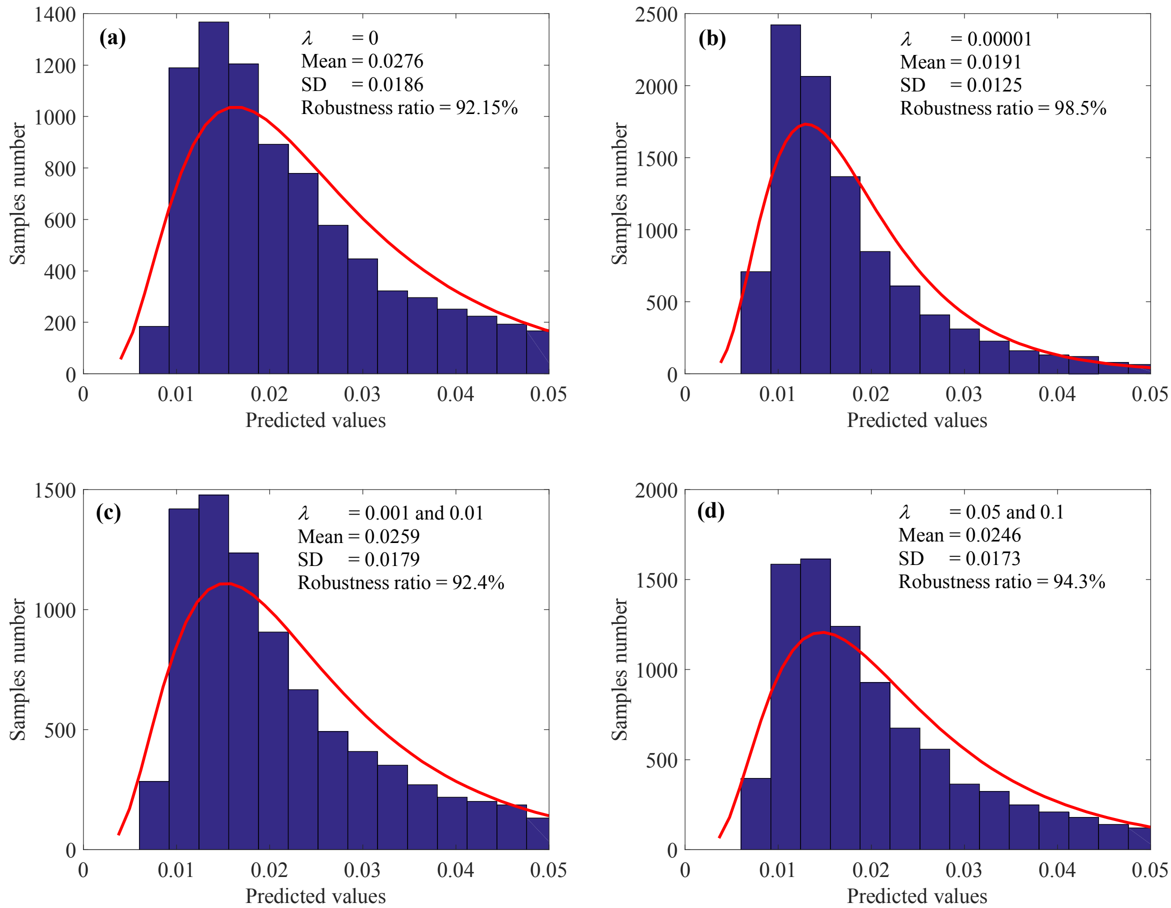
(a) λ=0
(b) λ=0
(c) λ=0.0001
(d) λ=0.0001
(e) λ=0.001 and 0.01
(f) λ=0.001 and 0.01
(g) λ=0.05 and 0.1
(h) λ=0.05 and 0.1
Comb=1
Comb=2
Comb=3
Comb=4
Comb=5
Comb=6
Comb=7
m=1
m=2
m=3
m=4
m=5
m=6
m=7
m=8
Percent for variable combination /%
Percent for term size/%
Number of generations
0
50
100
150
200

Figure 7



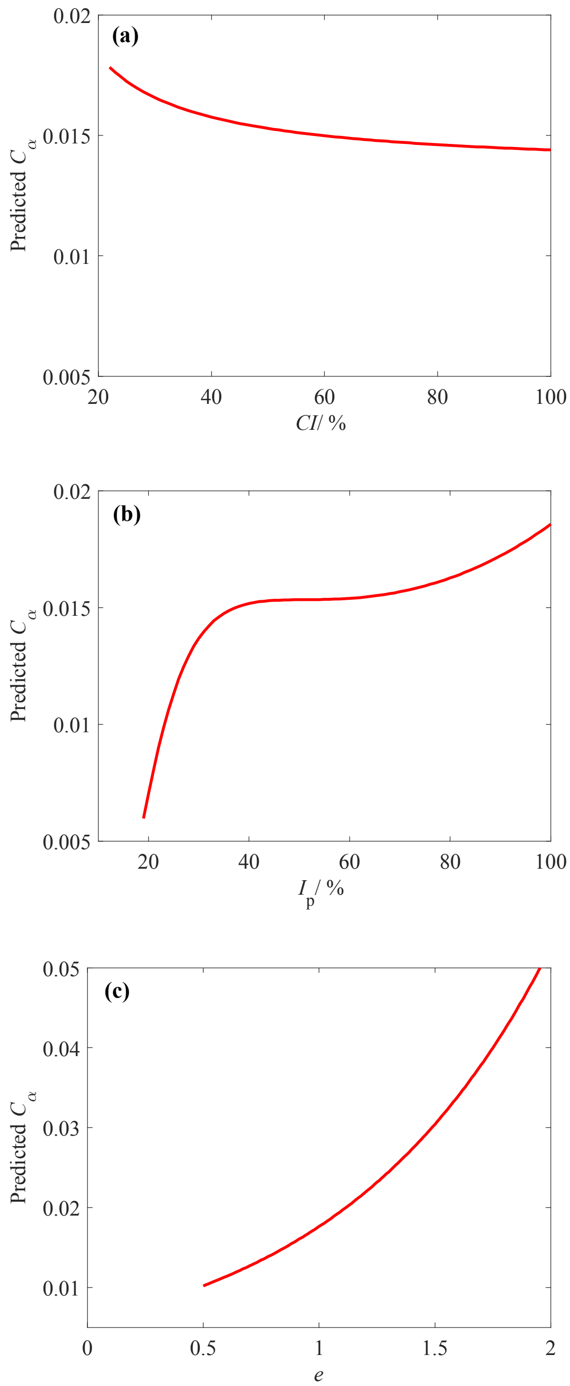
Testing points
1:1 line
Predictions
Measurements
(b) λ=0
R²=0.889
(c) λ=0.0001
R²=0.896
(d) λ=0.0001
R²=0.863
(e) λ=0.001 and 0.01
R²=0.892
(f) λ=0.001 and 0.01
R²=0.853
(g) λ=0.05 and 0.1
R²=0.895
(h) λ=0.05 and 0.1
R²=0.857

Figure 8



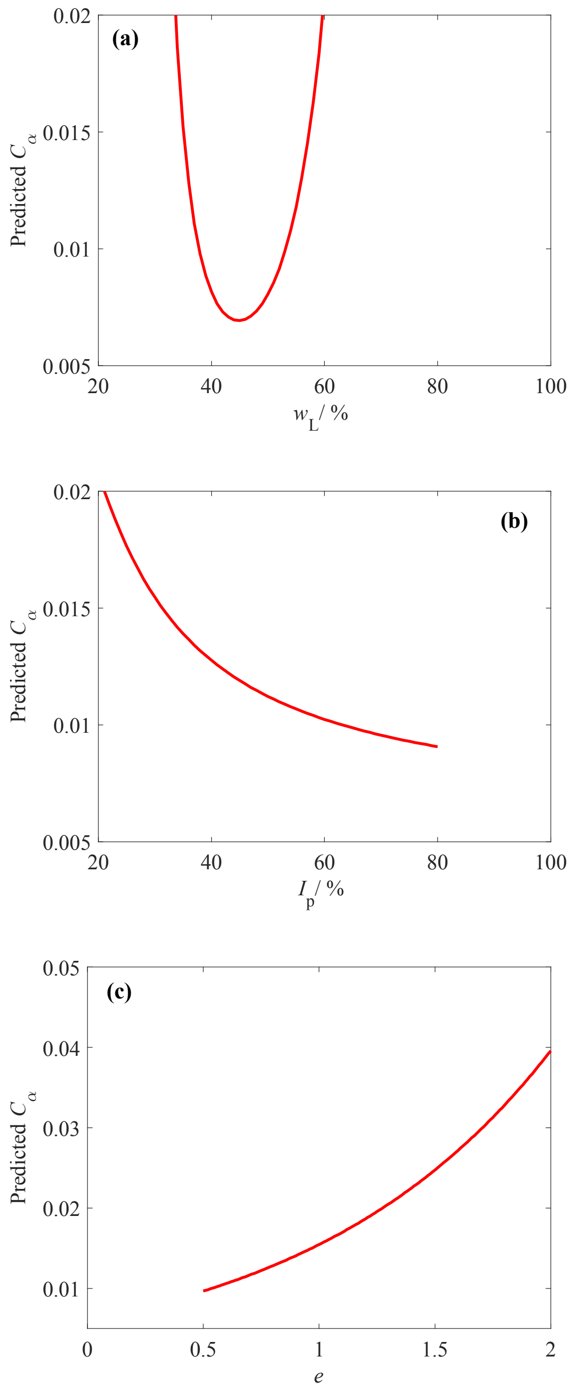
(a)
λ = 0
Mean = 0.0276
SD = 0.0186
Robustness ratio = 92.15%
Samples number
Predicted values
(b)
λ = 0.00001
Mean = 0.0191
SD = 0.0125
Robustness ratio = 98.5%
Samples number
Predicted values
(c)
λ = 0.001 and 0.01
Mean = 0.0259
SD = 0.0179
Robustness ratio = 92.4%
Samples number
Predicted values
(d)
λ = 0.05 and 0.1
Mean = 0.0246
SD = 0.0173
Robustness ratio = 94.3%
Samples number
Predicted values

Figure 9



(a)
Predicted C_α
0.02
0.015
0.01
0.005
20
40
60
80
100
CI/ %
(b)
Predicted C_α
0.02
0.015
0.01
0.005
20
40
60
80
100
I_p / %
(c)
Predicted C_α
0.05
0.04
0.03
0.02
0.01
0
0.5
1
1.5
2
e

Figure 10



(a)
Predicted C_α
0.02
0.015
0.01
0.005
20
40
60
80
100
w_L / %
(b)
Predicted C_α
0.02
0.015
0.01
0.005
20
40
60
80
100
I_p / %
(c)
Predicted C_α
0.05
0.04
0.03
0.02
0.01
0
0.5
1
1.5
2
e

Figure 11