An intelligent multi-objective EPR technique with multi-step model selection for correlations of soil properties

Yin-Fu JIN[1] and Zhen-Yu YIN[1*]

Affiliation:

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, Fax +852 23346389, E-mail: zhenyu.yin@polyu.edu.hk; zhenyu.yin@gmail.com

Abstract: Current multi-objective Evolutionary Polynomial Regression (EPR) methodology has difficulties on decision ~~make~~ making of optimal EPR model. This paper proposes an intelligent multi-objective optimization based EPR technique with multi-step automatic model selection procedure. A newly developed multi-objective differential evolution algorithm (MODE) is adopted to improve the optimization performance. The proposed EPR process is composed of two stages: (1) intelligent roughing model selection and (2) model delicacy identification. In the first stage, besides of two objectives (model accuracy and model complexity), the model robustness measured by robustness ratio is considered as an additional objective in the multi-objective optimization. In the second stage, a new indicator named selection index is proposed and incorporated to find the optimal model. After intelligent roughing selection and delicacy identification, the optimal EPR model is obtained considering the combined effects of correlation coefficient, size of polynomial terms, number of involved variables, robustness ratio and monotonicity. To show the practicality of the proposed EPR technique, three illustrative cases helpful for geotechnical design are presented: (a) modelling of compressibility, (b) modelling of undrained shear strength, and (c) modelling of hydraulic conductivity. For each case, a practical formula with better performance in comparison to various existing empirical equations is finally provided. All results demonstrate that the proposed intelligent MODE-based EPR technique is efficient and effective.

Key words: multi-objective optimization; artificial intelligence; compressibility; undrained shear strength; permeability; robustness

1 Introduction

In geotechnical engineering, the soil properties are important for design and the service evaluation of post-construction [1-10] and also for the constitutive modelling [11-17]. To quickly get soil properties, the empirical equation correlating the key property to some basically physical properties (e.g., soil density, void ratio, and Atterberg limits) is more convenient. However, when the relationship between the objective property and basically physical properties is highly nonlinear, the empirical approach becomes helpless. In contrast, the data mining techniques incorporating the artificial intelligence ~~is~~ are suitable for such problems.

Numerical regression through artificial intelligence is the most forceful and usually applied to solve the problem of finding the optimal model via fitting the observed data, such as back-propagation neural network (BPNN) [18-25], evolutionary neural network (ENN) [26], random forest [27,28], support vector machines (SVMs) [29,30], genetic programming (GP) [31] and Bayesian-related methods [32,5,6,33,34]. Among numerous artificial intelligence methods used in data-mining, the Evolutionary Polynomial Regression (EPR) is a data-driven modelling hybrid technique[35], based on evolutionary computing, that combines the best features of conventional numerical regression techniques with the genetic programming/symbolic regression technique. EPR is suitable for modelling physical phenomena [36] because of two features: (i) the introduction of prior knowledge about the physical system/process; and (ii) the production of symbolic formulae, enabling data mining to discover patterns which describe the desired parameters. Comparison showed that the EPR is superior to other data-mining techniques (such as extreme learning machine) [37,38]. Recently, the EPR technique has developed rapidly, as it provides advantages in modelling nonlinear complex problems. Such successful applications include modelling of clay compressibility [39,40], evaluation of liquefaction potential of sand [41,42], prediction of water quality parameters[38], prediction of soil saturated water content [43], settlement prediction of foundations [44-46], evaluation of pile bearing capacity [47-49], modelling of soil mechanical behaviour [50-53] and modelling of clay creep index [54]. Furthermore, the development of optimization algorithms can also effectively support EPR [55,39,56].

In general, EPR employs the single-objective genetic algorithm (SOGA) to find the optimal exponents [57,49,58,42,41,51]. Other single-objective optimization algorithms guaranteeing the global optimal solution (e.g., Differential Evolution algorithm[56], Particle Swarm Optimization[28], Ant Colony Optimization[59], Artificial Bee Colony algorithm[60]) can also be adopted in the EPR process. According to [61,62], however, the SOGA-based EPR has the following drawbacks: (1) the performance decreases with increasing the number of polynomial terms, and (2) the results are often difficult to interpret. Actually, the obtained models can be ranked according to their fitness or model complexity. However, the models ranked according to the model complexity requires some subjective judgment, and consequently this process is usually biased by the user's experience rather than being purely based on some mathematical criteria [63]. Furthermore, generalization ability and estimation of robustness aren't possible using the SOGA-based EPR methodology.

To overcome these drawbacks, it is possible to use a multi-objective optimization (MOOP) algorithm in EPR [62,61], such as the MOGA [64], NSGA [65] and other multi-objective algorithms, taking into account the many factors that influence formulae selection. The models obtained in common MOOP-based EPR methodology are ranked according to: (1) model complexity (i.e., the number of polynomial terms) and (2) model accuracy (i.e., model fitness). The generalization ability and robustness for identified formulae are usually overpassed in MOOP-based EPR methodology, which is rarely reported [62,61,66]. However, the advantage of MOOP-based EPR is that the number of objectives can be determined by the user to solve problems of interest. Thus, a further improvement of EPR can be achieved by implementing the MOOP strategy to optimise for robustness. The enlarged objectives for the MOOP-based EPR are as follows: (1) maximisation of model accuracy, (2) minimisation of the number of polynomial terms and (3) maximisation of model robustness (newly proposed and introduced to original MOOP-EPR procedure in this study). Note that the MOOP-based EPR can determine the Pareto front consisting of optimal formulae considering parsimony (number of constants and variables), accuracy and robustness. Moreover, the obtained Pareto front, composed of the set of Pareto optimal solutions which are not dominated by any other feasible solutions, will guide the model selection. In the end, a simple, reliable and robust EPR model can be achieved.

Therefore, this study is the first to propose an intelligent multi-objective EPR procedure with multi-step model selection ~~is proposed~~. The general EPR procedure is first introduced. Then, the flowchart of the proposed intelligent multi-objective EPR procedure is followed. The proposed technique has two stages: intelligent roughing model identification using multi-objective EPR and model delicacy selection. In the first stage, the multi-objective error function is composed of three objectives: term size, correlation coefficient and robustness ratio. A newly developed multi-objective differential evolution algorithm is adopted to improve the performance. In the second stage, a new selection index for selecting the optimal EPR model is defined and used. Finally, three typical cases with comparison to existing empirical equations are presented to show the practicality of the proposed EPR technique: (a) modelling of compressibility; (b) modelling of undrained shear strength; and (c) modelling of hydraulic conductivity.

2 General EPR procedure

EPR was first introduced by Giustolisi and Savic [35], which is a data-driven method based on evolutionary computing. A general EPR expression can be mathematically formulated as:

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

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. More details about EPR and the code can be found in [67] ~~http://www.hydroinformatics.it/~~.

Fig. 1 shows a typical flow chart for the EPR procedure. The general functional structure represented by $f\left(\mathbf{X}, a_j\right)$ in Eq.(1) is constructed from elementary functions using an optimization algorithm strategy. The building blocks (elements) of the structure are defined by the user based on the understanding of the physical process of interest. Selecting the feasible structures is conducted through an evolutionary process, while the parameters a_j are estimated by the least squares method.

3 Intelligent procedure of MODE-based EPR modelling

3.1 Flowchart of proposed intelligent multi-step selection EPR procedure

As stated by Wood [68], simple yet adequate models are favoured on the basis of practicality. The purpose of the proposed procedure is to ensure an optimal EPR model that has a reasonable balance between predictive capability and generalization ability. In this study, the proposed procedure involves two steps: (1) first detect all possible EPR models and (2) then identify the optimal one.

Fig. 2 presents the procedure of the proposed MOOP-based EPR technique, focusing on the estimation of model robustness, where $\boldsymbol{\theta}$ is the decision variables corresponding the exponents of the EPR model; *Comb* represents the number of variable combinations; and m is the number of polynomial terms of the EPR model. The first step is the intelligent roughing selection of all possible EPR models using multi-objective optimization. Note that any multi-objective optimization algorithm guaranteeing global Pareto front solutions can be employed in the proposed EPR procedure. In the proposed procedure, two additional variables *Comb* (an integer number) and m (an integer number) are included to the input, which is similar to the new SOGA-based EPR proposed by Jin et al.[54]. All variables are first generated randomly within their domains in the initial generation. Next, the possible variable combinations are selected according to the value of *Comb*, and then a possible term size for constructing the EPR model is chosen according to the value of m. Subsequently, the EPR model with unknown coefficients is obtained. Then, the vector of coefficient $\boldsymbol{a}$ is determined by least squares method between the observed and predicted data. So far, an entire EPR model is achieved. Next, three objectives are successively computed: (a) the term size normalized by the maximum term size is the first objective to assess model complexity and generalization ability; (b) the coefficient of determination R^2 is the second objective to assess model accuracy; and (c) the robustness ratio is the third objective to evaluate model robustness. All the objectives are transferred into the multi-objective differential evolution (MODE) algorithm, based on which all EPR models in the same generation are ranked and selected. Note that in order to keep the minimum multi-objective optimization, the second objective R^2 is replaced by 1-R^2, and the third objective, robustness ratio, is replaced by 1-robustness ratio in the MODE algorithm. The whole process exits when the stop criterion is reached; otherwise, the process continues to the next generation. As the number of generations increases, the eventual result of the first step is the finding of all possible EPR models with different numbers of term size, R^2 and robustness ratio on the Pareto front.

The second step, delicacy identification, is launched after the first step is completed. In this step, the optimal EPR model is finally identified. First, the obtained EPR models with R^2 lower than a value (e.g., 0.7) are discarded because the predictive ability of an EPR model must be guaranteed for the purpose of practice. Next, to deeply understand the monotonicity of model candidates, a monotonicity study is conducted on the involved physical properties for each candidate. The characteristics of monotonicity for a formula can basically hint whether it is physically correct or not. A model with monotonous variables is preferred for engineers. If a variable in the model candidate is monotonous, "1" is scored. Otherwise, "0" is scored. Then, the proportion of monotonous variables to total involved variables is calculated. The priority of each model is ranked according to the value of this proportion.

Then, all model candidates are ranked in terms of R^2, number of term size, number of involved variables, robustness ratio and monotonicity. For each indicator, the best one scores "1", the second scores "2", the third scores "3" and so on. Based on these ranking values, the selection index "s_index"$\in[0, 1]$, representing the possibility of a model's selection, is computed. It is defined as:

$$\text{s\_index} = 1 - \sum_{i=1}^{m} \left(w_i \frac{(\text{ind})_j^i}{\sum_{l=1}^{n} (\text{ind})_l^i} \right), \; j = 1,\; 2,\; \ldots,\; n \tag{2}$$

where n is the number of model candidates and $(\text{ind})_j^i$ is the ranking value for ith indicator of the jth model candidate; m is the number of indicators (m=5 in this study); w_i is the weight for each indicator (w_i=1/m in this study indicating an equal weight for each indicator). A high value for "s_index" indicates that the model has a high possibility of being selected.

Finally, the model candidate with the maximum value for "s_index" is selected as the optimum. In contrast to conventional EPR procedures (e.g., SOGA-based or MOOP-based), (1) the additional influence factor—robustness—on selecting EPR formula is incorporated into the proposed

MODE-based EPR; (2) the delicacy identification with defining a new selection index “s_index” to consider the model complexity, accuracy, robustness and monotonicity is proposed and implemented. After intelligent roughing selection and delicacy identification, the optimal EPR model is obtained considering model complexity, accuracy, robustness and monotonicity.

3.2 Error function

3.2.1 *Objective 1: Term size*

The number of term size can be an indicator in the estimation of model complexity. Low model complexity results in high generalization ability. Thus, the number of term size is one objective in the process of intelligent selection.

3.2.2 *Objective 2: Coefficient of determination R^2*

The performance of an EPR model is determined by fitness function. The coefficient of determination (R^2) is adopted as the fitness function, which is defined as:

$$R^2 = \frac{\sum_{i=1}^{N}\left(\mathbf{Y}_{\mathrm{m}} - \mathbf{Y}_{\mathrm{p}}\right)^2}{\sum_{i=1}^{N}\left(\mathbf{Y}_{\mathrm{m}} - \left(\frac{1}{N}\right)\sum_{i=1}^{N}\mathbf{Y}_{\mathrm{m}}\right)^2} \tag{3}$$

where N is the number of data points; $\mathbf{Y}_{\mathrm{m}}$ is the vector of observed values; and $\mathbf{Y}_{\mathrm{p}}$ is the vector of predicted values.

3.2.3 *Objective 3: Robustness ratio*

According to Jin et al.[54], an appropriate model has not only a good predictive ability and less complexity but also good robustness. A criterion representing the robustness proposed by Jin et al. [54] is adopted in the proposed MOOP-based EPR procedure:

$$\text{Robustness ratio}=\frac{\text{Samples falling in reasonable range}}{\text{Total samples}} \tag{4}$$

First, 10,000 samples randomly are generated from a reasonable joint distribution (e.g., multivariable lognormal distribution for most soil properties). Note that it supposes that variables (e.g., liquid limit (w_{L}), plastic limit (w_{P}), and plasticity index (I_{P})) are independent of each other. Then, the dependent variable (such as creep index in Jin et al.[54]) is predicted using each obtained EPR model. Finally, the robustness ratio is calculated according to the number of samples falling in the reasonable range. Note that the reasonable range for a concerned variable can be determined over a large amount of representative data, such as the collected data for training EPR model. Furthermore, sometimes, the engineer's experience can help determining the range.

3.3 Adopted MODE

To improve the performance of the proposed EPR process, the newly developed MODE by Jin et al. [55] was adopted. Fig. 3 shows the MODE flowchart, where μ is the population of individuals, λ is offspring and CR is crossover probability. In this MODE, a novel DE inspired recombination mutation operator proposed by Qi et al. [69] was adopted as follows:

$$\begin{cases} \boldsymbol{v}_i = \boldsymbol{x}_{r1} + F^i\left(\boldsymbol{x}_{r2} - \boldsymbol{x}_{r3}\right), \boldsymbol{x}_{r1} \text{ is the best individual among } [\boldsymbol{x}_{r1}\ \boldsymbol{x}_{r2}\ \boldsymbol{x}_{r3}], \text{ if } \left(rand < 0.6\right) \\ \boldsymbol{v}_i = \boldsymbol{x}_{best,i} + F^i\left(\boldsymbol{x}_{r1} - \boldsymbol{x}_{r2}\right) + F^i\left(\boldsymbol{x}_{r3} - \boldsymbol{x}_{r4}\right), \text{ otherwise} \end{cases} \tag{5}$$

where the indices r_1, r_2, r_3 and r_4 are distinct integers uniformly chosen from the set $\{1, 2, ..., N_p\}$; N_p is the number of individuals in one generation; $\left(\boldsymbol{x}_{r1} - \boldsymbol{x}_{r2}\right)$ and $\left(\boldsymbol{x}_{r3} - \boldsymbol{x}_{r4}\right)$ are difference vectors to mutate the corresponding parent $\boldsymbol{x}_i$; $\boldsymbol{x}_{best,i}$ is the best vector in the current generation i, which is randomly chosen as one of the top 100p% individuals in the current population with $p \in (0, 1]$, and in this case p was set to 0.1; and F^i is the mutation factor that is regenerated within [0.5, 1.0] at each generation.

After mutation, a binomial crossover is applied to offspring generated by crossover: ~~s~~

$$\mathbf{u}_{i,j} = \begin{cases} \mathrm{V}_{i,j}, & \text{if rand}\left(0,1\right) \le CR \text{ or } j = j_{rand} \\ \mathrm{x}_{i,j}, & \text{otherwise} \end{cases} \tag{6}$$

where rand(a, b) is a uniform random number in the interval [a, b] and is independently generated for each j and each i; j_{rand}=randint (1, D) is an integer randomly chosen from 1 to D and is newly generated for each i, D being the dimension of the problem; and the crossover probability $CR \in [0, 1]$, with CR=0.3 used in this study.

4 Applications to geotechnical properties of soils

To show the practicality of the proposed EPR procedure, it was applied to three cases, which covered typical geotechnical design activities: (a) modelling of compressibility, which is important for predicting the settlement of geotechnical structures; (b) modelling of undrained shear strength, which is important for predicting soil strength and analysing its failure probability; and (c) predicting hydraulic conductivity, which is extremely important for solving various hydrogeology as well as geotechnical and environmental problems.

4.1 Modelling of compressibility for remoulded clays

4.1.1 *Database*

The compressibility of a soil is usually measured by compression index C_c, defined as $C_c=\Delta e/\Delta\log(\sigma'_v)$, where e is void ratio and σ'_v is effective vertical stress. In traditional way, the C_c can be obtained from one-dimensional compression test or isotropic compression test prior to triaxial shear test. However, both kinds of tests would take a long time for obtaining the C_c. More than 50 clays with 200 measured points were collected from several references [70-84,39,85] and used in the proposed MODE-based EPR procedure. In the database, the initial void ratio (e_0), liquid limit (w_L) and plasticity index (I_P) were selected as the correlating variables of interest. It should be noted that the intrinsic compression index is directly related to the mineralogical composition of clays [86-90]. However, the datasets including the mineral fraction are limited, which hinders to develop an EPR based model involving the mineral fraction with excellent generalization ability. To assess the adequacy of the database, some indicators were determined, as noted in the statistics of variables summarized in Table 1.

4.1.2 *Discrepancy of current correlation formula*

According to previous studies [39], the compression index of remoulded clays can be correlated to various soil physical properties, such as water content w, initial void ratio e_0, liquid limit w_L, plastic limit w_P, and plastic index I_P. Table 2 summarizes the current empirical correlations of C_c using physical properties for remoulded clays. To assess their performance, all the collected data were predicted using some empirical correlations. Due to the unavailability of data, the correlations that involve G_s, e_L and A were not compared. The comparison between measurements and predictions is shown in Fig. 4. It was found that the correlation coefficients R^2 of all selected correlations are smaller than 0.8. From a practical view, the performance of all selected empirical correlations is not satisfactory and needs to be further improved.

4.1.3 *New EPR formulation of soil compressibility*

Based on the equations shown in Table 2, the e_0, w_L and I_P are the most typical properties and thus are selected as the correlating variables. A general structure of EPR expression for C_c is expressed:

$$C_c = f\left(e_0, w_L, I_P\right) + a_0 \quad (7)$$

where a_0 is a constant.

150 datasets randomly selected in the prepared database were used for training, and the remaining were used for testing. For simplicity, the value of exponent was constrained to [-2, 2] with a step size of 1. Also, the maximum number of terms was set to 8 according to Yin et al.[39]. For MODE, the number of the initial population was set to ten times that of decision variables, and the maximum generation was set to 200. Independent multiple runs were performed to avoid randomness. These settings of MODE will be used in the following cases. A total of seven combinations (= $C_3^1 + C_3^2 + C_3^3$), each containing different physical properties, were obtained, as summarized in Table 3. Thus, the maximum number of variable combinations *Comb* is 7, and the maximum number of polynomial terms *m* is 8.

4.1.4 *Results and discussion*

To follow the proposed MODE-based EPR procedure in details, the correlations of C_c for remoulded clays were obtained and presented here. To highlight the good performance of the proposed MODE-EPR over the other previous MOOP EPR techniques, a classic multi-objective genetic algorithm EPR "NSGA-II EPR" only considering two objectives (accuracy and model complexity) was selected for a fair comparison. Fig. 5(a) shows the Pareto fronts obtained by MODE and NSGA-II respectively. It can be seen that not only a better accuracy but also a better diversity (the term number of correlation varying from 1 to 8) of correlations were obtained by MODE compared to NSGA, which indicates the advantage of the proposed method over the others. Furthermore, the robustness and monotonicity are the add-values due to the new objective and the new workflow considered in the proposed MODE EPR, which however can't be guaranteed by previous MOOP EPR techniques. Fig. 5(b) shows all obtained models presented in the space of R^2, term size and robustness ratio. Since the robustness ratio for all models is almost the same and close to 1, the results are redisplayed in Fig. 5(c). It is found that the correlation coefficient increases as the number of polynomial terms increases. For practicality, models with R^2 greater than 0.8 were of concern and marked using a "blue star", as shown in Fig. 5(c), and the models with large number of term sizes but a slight increase in R^2 were discarded. The formulations of interest are summarized in Table 4. Each possible model was assigned a model number to make it easily identifiable. Table 5 presents the results of monotonicity analysis for variables involved in the EPR model of C_c. It can be seen that Model 1 has good mathematical characteristics. The results of ranking in terms of R^2, *Comb*, *m* and monotonicity for EPR models of C_c are summarized in Table 6. Based on these results, Model 1 was considered the optimal model for modelling C_c.

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

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

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

The higher the R^2 or lower the RMSE and MAE values are, the better the model's performance. A *u* value greater than 1.0 indicates over-estimation; a value less than 1.0 indicates underestimation. The best model is represented by a *u* value close to 1.0, and σ close to 0.

Table 7 summarizes the five indicators for the optimal model on training and testing data for C_c. The comparison of C_c between measurements and EPR predictions is shown in Fig. 6. All testing results demonstrate that the optimal EPR model can ~~accurately~~ reasonably predict the compression index using physical properties for given remoulded clays. Compared to the traditional way, the proposed EPR model can predict a value of C_c with enough accuracy and low experimental cost.

4.2 Modelling of undrained shear strength for clays

4.2.1 *Database*

To propose an EPR model of s_u, numerous experimental data given in published papers [91-95,71,96-100] were compiled to form the database. In the database, water content *w*, liquid limit

w_L, plastic limit w_P, plastic index I_P, sensitivity S_t, overconsolidation ratio OCR, effective in-situ vertical stress σ'_v and preconsolidation pressure σ'_p are treated as variables of interest. A total of 58 clays with 363 measured points were used in the training and testing. To assess the adequacy of the database, some indicators were determined, as noted in the statistics of variables summarized in Table 8. The undrained shear strength (s_u) can be evaluated in situ (such as the field vane (FV) test and the piezocone cone penetration (CPTU) test) as well as in laboratory tests (such as the undrained triaxial compression (TXC) and the direct simple shear (DSS) test). In the database, most of the s_u were measured from field vane shear test. According to [91], a correction is needed to convert s_u^{FV} into s_u(mob) due to the overestimation of s_u from standard FV tests with a high speed of rotation in the test. The s_u(mob) can be expressed as:

$$s_u(\mathrm{mob}) = \lambda \cdot s_u^{FV} \text{ with } \lambda = 1.5/(1 + w_L) \tag{10}$$

According to [101] and [102], s_u obtained from FV is somewhat comparable to s_u from DSS test results.

4.2.2 *Discrepancy of current correlation formula*

When s_u cannot be directly measured or the measurements are considered unreliable, s_u is commonly evaluated from transformation models using clay properties. Such models are usually empirical or semi-empirical, obtained by data fitting of measurements.

Table 9 summarizes some commonly used correlations of s_u. However, such models must be carefully applied and their limitations recognized, as soil properties, soil behaviour, and site geology may differ between the data source and where the transformation models are calibrated [91]. To evaluate their performances, a comparison between measurements and predictions for each model was conducted. All results are shown in Fig. 7. It can be seen that the predictions are far from the actual values for all selected models. The models that involve OCR are superior to other models. Of them, the best model is the one proposed by [103] as a function of OCR and S_t. However, none of models are satisfying for application to engineering practices.

4.2.3 *New EPR formulation of undrained shear strength*

Based on the performance of the equations shown in Fig. 6, w, w_L, OCR and S_t were selected as the correlating variables to construct an EPR model of $s_u(\text{mob})/\sigma'_v$. Then, the general structure of EPR expression is expressed as:

$$\frac{s_u(\text{mob})}{\sigma'_v} = f\left(w, w_L, \text{OCR}, S_t\right) + a_0 \tag{11}$$

where s_u(mob) is mobilized undrained shear strength; and a_0 is a constant.

300 data randomly selected in the prepared database were used for training and the remaining data were used for testing. The number of variable combinations ($Comb = C_4^1 + C_4^2 + C_4^3 + C_4^4$) is 15, and the maximum number of terms m is 8. All combinations are summarized in Table 10.

4.2.4 *Results and discussion*

Similar to the previous case, following the proposed MODE-based EPR procedure, the correlations of $s_u(\text{mob})/\sigma'_v$ for clays were obtained and presented here. Fig. 8 (a) shows all obtained models presented in terms of R^2, term size and robustness ratio. Since the robustness ratio for all models is almost the same and close to 1, the results are redisplayed in Fig. 8 (b). It is found that the R^2 increases slightly as the term size increases. Therefore, only three model candidates with fewer term sizes were of concern and marked using "blue stars". Table 11 gives the correlations of $s_u(\text{mob})/\sigma'_v$ with different numbers of term sizes. All equations are a function of OCR and water content, which are similar to the equations proposed by [101,104,105]. Table 12 presents the results of monotonicity analysis for variables involved in EPR models of $s_u(\text{mob})/\sigma'_v$. It can be seen that all selected models have good mathematical characteristics. Based on obtained preliminary results, all models were ranked in terms of R^2, *Comb*, *m* and monotonicity, and the selection index was then computed, as summarized in Table 13. Based on the results, "Model 1" was considered as the optimal model for modelling $s_u(\text{mob})/\sigma'_v$, which is a function of OCR.

Table 14 summarizes the five indicators for the optimal model on training and testing data for $s_u(\text{mob})/\sigma'_v$. The comparison of $s_u(\text{mob})/\sigma'_v$ between measurements and EPR predictions is shown in Fig. 6. It can be seen that most predicted points locate in the reasonable range [$\mathbf{Y_m}$=$\mathbf{Y_p}$$\pm$0.25]. All testing results demonstrate that the optimal EPR model can approximately predict undrained shear strength using physical properties for given clays.

4.3 Modelling of hydraulic conductivity for fine soils

4.3.1 *Database*

To propose an EPR correlation of hydraulic conductivity, a lot of experimental data [106,107,80,108-116] were collected to form a database. A total of 31 clays with 361 measured points were used in this case. To assess the adequacy of the database, some indicators were determined, as noted in the statistics of variables summarized in Table 15.

4.3.2 *Discrepancy of current correlation formula*

Up to now, numerous equations have been proposed to predict the saturated hydraulic conductivity of soils [117]. These equations are empirical and the hydraulic conductivity is commonly expressed as a function of the porosity and selected physical properties of the soils (e.g., I_p, w_L, w_P, I_L and percentage of clay minerals). All selected equations are summarized in Table 16. To assess the performance of all selected equations, the collected database was used to form a prediction. Equations that involve e_L imply that, at the liquid limit (e/e_L=1), the k value takes a constant value whatever the clay, which is obviously unreasonable. Due to the unavailability of data, equations that involve I_L [118] and clay minerals p [106] were not compared.

Fig. 10 shows the comparison of k between measurements and predictions for empirical correlations. It is found that none of the selected equations can well predict hydraulic conductivity using physical properties. It seems that only the correlation proposed by Sridharan and Nagaraj [119] is applicable for most measured points, but its performance is unsatisfying for the purpose of application . Therefore, a reliable and effective EPR correlation of hydraulic conductivity using soil physical properties will be presented in the next section.

4.3.3 *New EPR formulation of hydraulic conductivity*

Based on the equations shown in Table 16, the e, w_L and I_P are selected as the correlating variables. Besides these variables, the clay content also has an important influence on hydraulic conductivity [117]. To keep the relationship between k and e [120], these four properties were selected to build the following general structure of EPR expression:

$$\log(k) = f\left(CI, w_L, I_P\right)e + a_0 \tag{12}$$

where k is hydraulic conductivity; CI is clay content (the percentage of soil particle size <2 μm); w_L is liquid limit; I_P is plastic index; e is void ratio; and a_0 is a constant.

300 datasets randomly selected in the prepared database were used for training and the remaining data were used for testing. The number of variable combinations ($Comb = C_3^1 + C_3^2 + C_3^3$) is 7, and the maximum number of terms m is 8. All combinations of variables are summarized in Table 17.

4.3.4 *Results and discussion*

Following the proposed MODE-based EPR procedure, the correlations of hydraulic conductivity for fine soils were obtained and presented in details. Fig. 11(a) shows all obtained models presented in terms of R^2, term size and robustness ratio. Since the robustness ratio for all models is almost the same and close to 1, the results are redisplayed in Fig. 11(b). It is found that the correlation coefficient increases as the number of polynomial terms increases. For practicality, models with R^2 greater than 0.7 were selected, and the remaining models were discarded. For selected models with the same number of term size, only the model with the maximum R^2 was of concern, marked using a "blue star", as shown in Fig. 11(b). The formulations of model candidates are presented in Table 18. The results of monotonicity analysis for variables involved in the EPR model are summarized in Table 19. Apart from Model 1, the monotonicity of other model candidates is not good. To find the optimal model, all model candidates were ranked in terms of R^2, the number of term size, the number of variables, robustness ratio and monotonicity, and the selection index was also computed, as summarized in Table 20. The results indicate that Model 1, with three terms, is optimum.

Table 21 shows the summary of five indicators for the optimal model on training and testing data for k. The comparison of k between measurements and EPR predictions is shown in Fig. 12. It can be seen that all predicted points locate in the range (1/3~3) of actual values, which is acceptable for engineering practice [121]. All testing results demonstrate that the optimal EPR model can ~~accurately~~ reasonably predict hydraulic conductivity using physical properties for given fine clays.

Note that the predicted performance of S_u or k is not as good as C_c for remoulded clays. Compared to intrinsic C_c, the uncertainty of S_u or k is more significant. The value of S_u or k is obtained from field tests, which are affected by many factors, such as the soil spatial variability [122] and water chemical environment [123]. To quantify the uncertainty and predict a reasonable value of S_u or k, the optimization algorithm used in EPR process can be replaced by Bayesian parameter identification method [124,125]. Furthermore, increasing the number of polynomial terms would have slight improvement for such a problem but it also brings model complexity.

5 Conclusions

An intelligent MODE-based EPR modelling technique with multi-step model selection has been proposed in this study. The first stage ~~was~~ is roughing selection, in which an enhanced MOOP-based EPR procedure with three objectives ~~was~~ is proposed. The proposed EPR procedure ~~was~~ is set apart from common MOOP-based EPR procedures in that, besides considering model accuracy (fitting performance) and model complexity (term size), it ~~defined~~ defines and ~~adopted~~ adopts the robustness ratio to measure the robustness of an EPR model.

After roughing selection, the second stage, delicacy identification, ~~was~~ is launched, in which the optimal model ~~was~~ is finally selected. All model candidates ~~were~~ are respectively ranked in terms of R^2, number of term size, number of involved variables, robustness ratio and monotonicity. To find the optimal model, a selection index considering the combined effects of all indicators ~~was~~ is defined and used in the proposed procedure for a decision make on the optimal model.

To show the practicality of the proposed intelligent EPR technique, it ~~was~~ is applied to three cases, which covered typical geotechnical design activities: (a) modelling of compressibility; (b) modelling of undrained shear strength; and (c) modelling of hydraulic conductivity. Finally, three practical formulae ~~were~~ are obtained and evaluated with better performance comparing to existing ones. All comparisons demonstrate that the proposed MODE-EPR involving the indicator of model robustness with two-stage selection scheme is superior to the existing methods in terms of accuracy and robustness.

The performance of proposed correlations (i.e., C_c, S_u and k) would be better if more datasets or more variables are involved in the proposed selection procedure. However, the collection of data is a tedious and difficult work, which needs put in a lot of vigour and time. Furthermore, the representation of traditional machine learning methods or data mining methods is limited. In the future, the more advanced optimization algorithm or novel selection procedure may improve the performance of EPR.

Acknowledgments

This research was financially supported by a RIF project (Grant No.: 15209119, PolyU R5037-18F) from Research Grants Council (RGC) of Hong Kong Special Administrative Region Government (HKSARG) of China.

Reference

1. Jiang G, Chen W, Liu X, Yuan S, Wu L, Zhang C (2018) Field study on swelling-shrinkage response of an expansive soil foundation under high-speed railway embankment loads. Soils and Foundations 58 (6):1538-1552
2. Yuan S, Liu X, Buzzi O (2018) Effects of soil structure on the permeability of saturated Maryland clay. Géotechnique 69 (1):72-78
3. Kulhawy FH, Mayne PW (1990) Manual on estimating soil properties for foundation design. Electric Power Research Inst., Palo Alto, CA (USA); Cornell Univ., Ithaca …,
4. Jin Y-F, Yin Z-Y, Wu Z-X, Daouadji A (2018) Numerical modeling of pile penetration in silica sands considering the effect of grain breakage. Finite Elem Anal Des 144:15-29. doi:https://doi.org/10.1016/j.finel.2018.02.003
5. Qi X-H, Zhou W-H (2017) An efficient probabilistic back-analysis method for braced excavations using wall deflection data at multiple points. Computers and Geotechnics 85:186-198
6. Tan F, Zhou WH, Yuen KV (2018) Effect of loading duration on uncertainty in creep analysis of clay. Int J Numer Anal Methods Geomech 42 (11):1235-1254
7. Zhou W-H, Tan F, Yuen K-V (2018) Model updating and uncertainty analysis for creep behavior of soft soil. Computers and Geotechnics 100:135-143. doi:https://doi.org/10.1016/j.compgeo.2018.04.006
8. Chen WB, Yin JH, Feng WQ, Borana L, Chen RP (2018) Accumulated Permanent Axial Strain of a Subgrade Fill under Cyclic High-Speed Railway Loading. Int J Geomech 18 (5). doi:04018018, 10.1061/(asce)gm.1943-5622.0001119
9. Chen WB, Feng WQ, Yin JH, Borana L, Chen RP (2019) Characterization of permanent axial strain of granular materials subjected to cyclic loading based on shakedown theory. Construction and Building Materials 198:751-761. doi:10.1016/j.conbuildmat.2018.12.012
10. Chen WB, Liu K, Yin ZY, Yin JH (2020) Crushing and Flooding Effects on One-Dimensional Time-Dependent Behaviors of a Granular Soil. Int J Geomech 20 (2). doi:04019156 10.1061/(asce)gm.1943-5622.0001560
11. Yao Y, Sun D, Luo T (2004) A critical state model for sands dependent on stress and density. Int J Numer Anal Methods Geomech 28 (4):323-337
12. Yao Y, Sun D, Matsuoka H (2008) A unified constitutive model for both clay and sand with hardening parameter independent on stress path. Computers and Geotechnics 35 (2):210-222
13. Yao Y-P, Yamamoto H, Wang N-D (2008) Constitutive model considering sand crushing. Soils and Foundations 48 (4):603-608
14. Yao Y, Hou W, Zhou A (2009) UH model: three-dimensional unified hardening model for overconsolidated clays. Geotechnique 59 (5):451-469
15. Yao Y, Kong L, Hu J (2013) An elastic-viscous-plastic model for overconsolidated clays. Science China Technological Sciences 56 (2):441-457
16. Yao Y, Zhou A (2013) Non-isothermal unified hardening model: a thermo-elasto-plastic model for clays. Géotechnique 63 (15):1328

17. Yao Y-P, Kong L-M, Zhou A-N, Yin J-H (2014) Time-dependent unified hardening model: Three-dimensional elastoviscoplastic constitutive model for clays. Journal of engineering mechanics 141 (6):04014162
18. He S, Li J (2009) Modeling nonlinear elastic behavior of reinforced soil using artificial neural networks. Applied Soft Computing 9 (3):954-961. doi:https://doi.org/10.1016/j.asoc.2008.11.013
19. Rashidian V, Hassanlourad M (2014) Application of an artificial neural network for modeling the mechanical behavior of carbonate soils. Int J Geomech 14 (1):142-150. doi:10.1061/(asce)gm.1943-5622.0000299
20. Penumadu D, Zhao RD (1999) Triaxial compression behavior of sand and gravel using artificial neural networks (ANN). Comput Geotech 24:207-230
21. Basheer IA (2000) Selection of methodology for neural network modeling of constitutive hystereses behavior of soils. Comput-Aided Civ Inf 15:440–458
22. Habibagahi G, Bamdad A (2003) A neural network framework for mechanical behavior of unsaturated soils. Canadian Geotechnical Journal 40 (3):684-693
23. Turk G, Logar J, Majes B (2001) Modelling soil behaviour in uniaxial strain conditions by neural networks. Adv Eng Softw 32:805-812
24. Chen R, Zhang P, Wu H, Wang Z, Zhong Z (2019) Prediction of shield tunneling-induced ground settlement using machine learning techniques. Frontiers of Structural and Civil Engineering 13 (6):1363-1378
25. Chen RP, Zhang P, Kang X, Zhong ZQ, Liu Y, Wu HN (2019) Prediction of maximum surface settlement caused by earth pressure balance (EPB) shield tunneling with ANN methods. Soils and Foundations 59 (2):284-295. doi:10.1016/j.sandf.2018.11.005
26. Johari A, Javadi AA, Habibagahi G (2011) Modelling the mechanical behaviour of unsaturated soils using a genetic algorithm-based neural network. Comput Geotech 38 (1):2-13. doi:10.1016/j.compgeo.2010.08.011
27. Zhang P, Chen RP, Wu HN (2019) Real-time analysis and regulation of EPB shield steering using Random Forest. Automation in construction 106. doi:Unsp 10286010.1016/j.autcon.2019.102860
28. Zhang P, Yin Z-Y, Jin Y-F, Chan TH (2020) A novel hybrid surrogate intelligent model for creep index prediction based on particle swarm optimization and random forest. Eng Geol 265:105328
29. Kohestani VR, Hassanlourad M (2016) Modeling the mechanical behavior of carbonate sands using artificial neural networks and support vector machines. Int J Geomech 16 (1):04015038. doi:10.1061/(ASCE)
30. 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. Eng Geol 204:108-120. doi:https://doi.org/10.1016/j.enggeo.2016.02.009
31. Cabalar AF, Cevik A (2011) Triaxial behavior of sand–mica mixtures using genetic programming. Expert Systems with Applications 38 (8):10358-10367

32. Gamse S, Zhou W-H, Tan F, Yuen K-V, Oberguggenberger M (2018) Hydrostatic-season-time model updating using Bayesian model class selection. Reliability Engineering & System Safety 169:40-50
33. Tan F, Zhou W-H, Yuen K-V (2016) Modeling the soil water retention properties of same-textured soils with different initial void ratios. J Hydrol 542:731-743
34. Zhou W-H, Yuen K-V, Tan F (2012) Estimation of maximum pullout shear stress of grouted soil nails using Bayesian probabilistic approach. Int J Geomech 13 (5):659-664
35. Giustolisi O, Savic D (2006) A symbolic data-driven technique based on evolutionary polynomial regression. Journal of Hydroinformatics 8 (3):207-222
36. Abdul Ghani N, Shahin M, Nikraz H (2012) Use of evolutionary polynomial regression (EPR) for prediction of total sediment load of Malaysian rivers. Int J Eng 6 (5):262-277
37. Rezaie-Balf M, Kisi O (2017) New formulation for forecasting streamflow: evolutionary polynomial regression vs. extreme learning machine. Hydrology Research:nh2017283
38. Najafzadeh M, Ghaemi A, Emamgholizadeh S (2019) Prediction of water quality parameters using evolutionary computing-based formulations. Int J Environ Sci Technol 16 (10):6377-6396
39. 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. Eng Geol 210:158-167
40. 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 (3):211-224
41. Rezania M, Javadi AA, Giustolisi O (2010) Evaluation of liquefaction potential based on CPT results using evolutionary polynomial regression. Computers and Geotechnics 37 (1):82-92
42. Rezania M, Faramarzi A, Javadi AA (2011) An evolutionary based approach for assessment of earthquake-induced soil liquefaction and lateral displacement. Eng Appl Artif Intell 24 (1):142-153
43. Khoshkroudi SS, Sefidkouhi MAG, Ahmadi MZ, Ramezani M (2014) Prediction of soil saturated water content using evolutionary polynomial regression (EPR). Archives of Agronomy and Soil Science 60 (8):1155-1172
44. Shahnazari H, Shahin MA, Tutunchian MA (2014) Evolutionary-based approaches for settlement prediction of shallow foundations on cohesionless soils. Geotech Eng 12 (1):55-64
45. 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
46. Shahin MA (2014) State-of-the-art review of some artificial intelligence applications in pile foundations. Geoscience Frontiers 7 (1):33-44. doi:10.1016/j.gsf.2014.10.002
47. Ebrahimian B, Movahed V (2013) Evaluation of Axial Bearing Capacity of Piles in Sandy Soils by CPT Results. Evaluation 29:31
48. Ebrahimian B, Movahed V (2017) Application of an evolutionary-based approach in evaluating pile bearing capacity using CPT results. Ships and Offshore Structures 12 (7):937-953

49. Ahangar-Asr A, Javadi AA, 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
50. Javadi AA, Faramarzi A, Ahangar-Asr A (2012) Analysis of behaviour of soils under cyclic loading using EPR-based finite element method. Finite Elem Anal Des 58:53-65
51. Shahnazari H, Tutunchian MA, Rezvani R, Valizadeh F (2013) Evolutionary-based approaches for determining the deviatoric stress of calcareous sands. Comput Geosci 50:84-94
52. Faramarzi A, Alani AM, Javadi AA (2014) An EPR-based self-learning approach to material modelling. Computers & Structures 137:63-71
53. Nassr A, Javadi A, Faramarzi A (2018) Developing constitutive models from EPR‐based self‐learning finite element analysis. Int J Numer Anal Methods Geomech 42 (3):401-417
54. Jin Y-F, Yin Z-Y, Zhou W-H, Yin J-H, Shao J-F (2019) A single-objective EPR based model for creep index of soft clays considering L2 regularization. Eng Geol 248:242-255. doi:https://doi.org/10.1016/j.enggeo.2018.12.006
55. Jin Y-F, Yin Z-Y, Zhou W-H, Huang H-W (2019) Multi-objective optimization-based updating of predictions during excavation. Eng Appl Artif Intell 78:102-123. doi:https://doi.org/10.1016/j.engappai.2018.11.002
56. Yin Z-Y, Jin Y-F, Shen JS, Hicher P-Y (2018) Optimization techniques for identifying soil parameters in geotechnical engineering: Comparative study and enhancement. Int J Numer Anal Methods Geomech 42 (1):70-94. doi:10.1002/nag.2714
57. Ahangar-Asr A, Faramarzi A, Javadi AA (2010) A new approach for prediction of the stability of soil and rock slopes. Engineering computations 27 (7):878-893
58. Faramarzi A, Javadi AA, Alani AM (2012) EPR-based material modelling of soils considering volume changes. Comput Geosci 48:73-85
59. Dorigo M, Birattari M, Stützle T (2006) Ant colony optimization. Computational Intelligence Magazine, IEEE 1 (4):28-39
60. Karaboga D, Ozturk C (2011) A novel clustering approach: Artificial Bee Colony (ABC) algorithm. Applied Soft Computing 11 (1):652-657
61. Savic D, Giustolisi O, Laucelli D (2009) Asset deterioration analysis using multi-utility data and multi-objective data mining. Journal of Hydroinformatics 11 (3-4):211-224
62. Giustolisi O, Savic D (2009) Advances in data-driven analyses and modelling using EPR-MOGA. Journal of Hydroinformatics 11 (3-4):225-236
63. Young P, Parkinson S, Lees M (1996) Simplicity out of complexity in environmental modelling: Occam's razor revisited. Journal of applied statistics 23 (2-3):165-210
64. Goldberg DE, Corruble V, Ganascia J-G, Holland J (1994) Algorithmes génétiques: exploration, optimisation et apprentissage automatique. Addison-Wesley France,
65. Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. Evolutionary Computation, IEEE Transactions on 6 (2):182-197
66. Vassallo R, Doglioni A, Grimaldi G, Di Maio C, Simeone V (2016) Relationships between rain and displacements of an active earthflow: a data-driven approach by EPRMOGA. Natural hazards:DOI 10.1007/s11069-11015-12140-11069

67. Giustolisi O, Savic D (2006) Evolutionary Polynomial Regression (EPR) http://www.hydroinformatics.it/.
68. Wood DM (2003) Geotechnical modelling, vol 1. CRC Press,
69. Qi Y, Hou Z, Yin M, Sun H, Huang J (2015) An immune multi-objective optimization algorithm with differential evolution inspired recombination. Applied Soft Computing 29:395-410
70. Burland J (1990) On the compressibility and shear strength of natural clays. Geotechnique 40 (3):329-378
71. Horpibulsuk S, Shibuya S, Fuenkajorn K, Katkan W (2007) Assessment of engineering properties of Bangkok clay. Canadian Geotechnical Journal 44 (2):173-187
72. Hong S-J, Kim D-H, Lee M-J, Jie H-K, Lee W-J (2013) Evaluation of Compression Index for Natural Clay Using the Compression Characteristic of Reconstituted Clay. Journal of the Korean Geotechnical Society 29 (3):5-13
73. Hong Z-S, Zeng L-L, Cui Y-J, Cai Y-Q, Lin C (2012) Compression behaviour of natural and reconstituted clays. Geotechnique 62 (4):291-301
74. Hong Z, Yin J, Cui Y-J (2010) Compression behaviour of reconstituted soils at high initial water contents. Geotechnique 60 (9):691-700
75. Carrier III N (1985) Consolidation parameters derived from index tests. Geotechnique 35 (2):211-213
76. Nagaraj T, Murthy BS (1983) Rationalization of Skempton's compressibility equation. Geotechnique 33 (4):433-443
77. Nagaraj T, Murthy BS (1986) A critical reappraisal of compression index equations. Geotechnique 36 (1):27-32
78. Nath A, DeDalal S (2004) The role of plasticity index in predicting compression behaviour of clays. Electron J Geotech Eng 9:1-7
79. Park J, Koumoto T (2000) Compression characteristics of remolded clays. Transactions of the Japanese Society of Irrigation, Drainage and Reclamation Engineering (Japan)
80. Horpibulsuk S, Yangsukkaseam N, Chinkulkijniwat A, Du YJ (2011) Compressibility and permeability of Bangkok clay compared with kaolinite and bentonite. Appl Clay Sci 52 (1):150-159
81. Sridharan A, Nagaraj H (2000) Compressibility behaviour of remoulded, fine-grained soils and correlation with index properties. Canadian Geotechnical Journal 37 (3):712-722
82. Favre J-L, Hattab M (2008) Analysis of the 'Biarez–Favre'and 'Burland'models for the compressibility of remoulded clays. CR Geosci 340 (1):20-27
83. Tiwari B, Ajmera B (2011) New correlation equations for compression index of remolded clays. Journal of geotechnical and geoenvironmental engineering 138 (6):757-762
84. 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 (11):767-777
85. 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 (6):04017092
86. Tiwari B, Ajmera B (2012) New Correlation Equations for Compression Index of Remolded Clays. J Geotech Geoenviron 138 (6):757-762. doi:10.1061/(asce)gt.1943-5606.0000639

87. Habibbeygi F, Nikraz H, Koul BK, Iovine G (2018) Regression models for intrinsic constants of reconstituted clays. Cogent Geoscience 4 (1). doi:10.1080/23312041.2018.1546978
88. Giasi CI, Cherubini C, Paccapelo F (2003) Evaluation of compression index of remoulded clays by means of Atterberg limits. B Eng Geol Environ 62 (4):333-340. doi:10.1007/s10064-003-0196-3
89. Cerato AB, Lutenegger AJ (2004) Determining intrinsic compressibility of Fine-Grained soils. J Geotech Geoenviron 130 (8):872-877.
90. Cao S, Song W, Yilmaz E (2018) Influence of structural factors on uniaxial compressive strength of cemented tailings backfill. Construction and Building Materials 174:190-201
91. D'Ignazio M, Phoon K-K, Tan SA, Länsivaara TT (2016) Correlations for undrained shear strength of Finnish soft clays. Canadian Geotechnical Journal 53 (10):1628-1645. doi:10.1139/cgj-2016-0037
92. Ching J, Phoon K-K (2014) Transformations and correlations among some clay parameters — the global database. Canadian Geotechnical Journal 51 (6):663-685. doi:10.1139/cgj-2013-0262
93. Ching J, Phoon K-K, Chen C-H (2013) Modeling piezocone cone penetration (CPTU) parameters of clays as a multivariate normal distribution. Canadian Geotechnical Journal 51 (1):77-91. doi:10.1139/cgj-2012-0259
94. Leroueil S, Tavenas F, Samson L, Morin P (1983) Preconsolidation pressure of Champlain clays. Part II. Laboratory determination. Canadian Geotechnical Journal 20 (4):803-816. doi:10.1139/t83-084
95. Tavenas F, Leroueil S, La Rochelle P, Roy M (1978) Creep behaviour of an undisturbed lightly overconsolidated clay. Canadian Geotechnical Journal 15 (3):402-423
96. Tanaka H, Locat J, Shibuya S, Soon TT, Shiwakoti DR (2001) Characterization of Singapore, Bangkok, and Ariake clays. Canadian Geotechnical Journal 38 (2):378-400. doi:10.1139/t00-106
97. Hanzawa H, Fukaya T, Suzuki K (1990) Evaluation of engineering properties for an Ariake clay. Soils and Foundations 30 (4):11-24
98. Kim YT, Leroueil S (2001) Modeling the viscoplastic behaviour of clays during consolidation: application to Berthierville clay in both laboratory and field conditions. Canadian Geotechnical Journal 38 (3):484-497
99. Hawkins A, Larnach W, Lloyd I, Nash D (1989) Selecting the location, and the initial investigation of the SERC soft clay test bed site. Q J Eng GeolHydrogeol 22 (4):281-316
100. Ching J, Phoon K-K, Chen Y-C (2010) Reducing shear strength uncertainties in clays by multivariate correlations. Canadian Geotechnical Journal 47 (1):16-33
101. Jamiolkowski M New developments in field and laboratory testing or soils. In: Proc. 11th. Int. Conf. on SMFE., San Francisco, CA, 1985, 1985. pp 57-153
102. Chandler RJ (1988) The in-situ measurement of the undrained shear strength of clays using the field vane. In: Vane shear strength testing in soils: field and laboratory studies. ASTM International,
103. Ching J, Phoon K-K (2012) Modeling parameters of structured clays as a multivariate normal distribution. Canadian Geotechnical Journal 49 (5):522-545

104. Karlsrud K, Hernandez-Martinez FG (2013) Strength and deformation properties of Norwegian clays from laboratory tests on high-quality block samples. Canadian Geotechnical Journal 50 (12):1273-1293
105. Larsson R, Larsson R (2007) Skjuvhållfasthet: utvärdering i kohesionsjord. Statens geotekniska institut (SGI) Linköping,
106. Dolinar B (2009) Predicting the hydraulic conductivity of saturated clays using plasticity-value correlations. Appl Clay Sci 45 (1-2):90-94
107. Mishra AK, Ohtsubo M, Li L, Higashi T (2011) Controlling factors of the swelling of various bentonites and their correlations with the hydraulic conductivity of soil-bentonite mixtures. Appl Clay Sci 52 (1):78-84. doi:https://doi.org/10.1016/j.clay.2011.01.033
108. Sinha SK, Wang MC (2008) Artificial Neural Network Prediction Models for Soil Compaction and Permeability. Geotech Geol Eng 26 (1):47-64. doi:10.1007/s10706-007-9146-3
109. Al-Tabbaa A, Wood DM (1987) Some measurements of the permeability of kaolin. Géotechnique 37 (4):499-514. doi:10.1680/geot.1987.37.4.499
110. Sivapullaiah PV, Sridharan A, Stalin VK (2000) Hydraulic conductivity of bentonite-sand mixtures. Canadian Geotechnical Journal 37 (2):406-413. doi:10.1139/t99-120
111. Shen S-L, Wang J-P, Wu H-N, Xu Y-S, Ye G-L, Yin Z-Y (2015) Evaluation of hydraulic conductivity for both marine and deltaic deposits based on piezocone testing. Ocean Eng 110:174-182. doi:https://doi.org/10.1016/j.oceaneng.2015.10.011
112. Benson CH, Trast JM (1995) Hydraulic conductivity of thirteen compacted clays. Clays Clay Miner 43 (6):669-681
113. Zeng L-L, Hong Z-S, Cai Y-Q, Han J (2011) Change of hydraulic conductivity during compression of undisturbed and remolded clays. Appl Clay Sci 51 (1):86-93. doi:https://doi.org/10.1016/j.clay.2010.11.005
114. Shen S-L, Wu Y-X, Misra A (2017) Calculation of head difference at two sides of a cut-off barrier during excavation dewatering. Computers and Geotechnics 91:192-202
115. Shen SL, Xu YS (2011) Numerical evaluation of land subsidence induced by groundwater pumping in Shanghai. Canadian Geotechnical Journal 48 (9):1378-1392
116. Wu Y-X, Shen S-L, Yuan D-J (2016) Characteristics of dewatering induced drawdown curve under blocking effect of retaining wall in aquifer. J Hydrol 539:554-566
117. Chapuis RP (2012) Predicting the saturated hydraulic conductivity of soils: a review. Bull Eng Geol Environ 71 (3):401-434. doi:10.1007/s10064-012-0418-7
118. Berilgen SA, Berilgen MM, Ozaydin İK (2006) Compression and permeability relationships in high water content clays. Appl Clay Sci 31 (3):249-261
119. Sridharan A, Nagaraj HB (2005) Hydraulic conductivity of remolded fine-grained soils versus index properties. Geotechnical & Geological Engineering 23 (1):43
120. Taylor DW (1948) Fundamentals of soil mechanics. Soil Science 66 (2):161
121. Zeng L-l, Hong Z-s, Chen F-q (2012) A law of change in permeability coefficient during compression of remolded clays (in Chinese). Rock and Soil Mechanics 5:001
122. Dasaka SM, Zhang LM (2012) Spatial variability of in situ weathered soil. Géotechnique 62 (5):375-384. doi:10.1680/geot.8.P.151.3786

123. Bolt G (1956) Physico-chemical analysis of the compressibility of pure clays. Geotechnique 6 (2):86-93
124. Jin Y-F, Yin Z-Y, Zhou W-H, Horpibulsuk S (2019) Identifying parameters of advanced soil models using an enhanced transitional Markov chain Monte Carlo method. Acta Geotech 14 (6):1925-1947 doi:10.1007/s11440-019-00847-1
125. Jin Y-F, Yin Z-Y, Zhou W-H, Shao J-F (2019) Bayesian model selection for sand with generalization ability evaluation. Int J Numer Anal Methods Geomech 43 (14):2305-2327. doi:10.1002/nag.2979
126. Skempton AW, Jones O (1944) Notes on the compressibility of clays. Quarterly Journal of the Geological Society 100 (1-4):119-135
127. Wroth C, Wood D (1978) The correlation of index properties with some basic engineering properties of soils. Canadian Geotechnical Journal 15 (2):137-145
128. Biarez J, Hicher P-Y (1994) Elementary mechanics of soil behaviour: saturated remoulded soils. AA Balkema,
129. Skempton A (1954) Discussion of the structure of inorganic soil. Journal of American Society of Civil Engineers 80 (478):19-22
130. Larsson R (1980) Undrained shear strength in stability calculation of embankments and foundations on soft clays. Canadian Geotechnical Journal 17 (4):591-602
131. Mesri G (1989) A reevaluation of using laboratory shear tests. Canadian Geotechnical Journal 26 (1):162-164
132. Nishida Y, Nakagawa S Water permeability and plastic index of soils. In: Proceedings of IASH-UNESCO Symposium Tokyo, Pub, 1969. pp 573-578
133. Carrier WD, Beckman JF (1984) Correlations between index tests and the properties of remoulded clays. Géotechnique 34 (34):211-228
134. Nagaraj TS, Pandian NS, Raju PSRN (1993) Stress state-permeability relationships for fine-grained soils. Géotechnique 43 (2):333-336. doi:10.1680/geot.1993.43.2.333
135. Nagaraj TS, Pandian NS, Raju PSRN (1994) Stress-state—permeability relations for overconsolidated clays. Géotechnique 44 (2):349-352. doi:10.1680/geot.1994.44.2.349
136. Sivapullaiah PV, Sridharan A, Stalin VK (2000) Hydraulic conductivity of bentonite-sand mixtures. Canadian Geotechnical Journal 37 (37):406-413
137. Prakash K, Sridharan A (2002) Determination of liquid limit from equilibrium sediment volume. Géotechnique 52 (9):693-696. doi:10.1680/geot.2002.52.9.693

Tables

Table 1 Statistics of variables used in the database

Variable	Maximum	Minimum	Mean	Standard deviation
e_0	4.463	0.676	2.158	0.920
w_L (%)	166.20	25.00	68.89	26.61
I_P (%)	113.90	8.0	35.21	19.72
C_c	1.340	0.12	0.457	0.226

Table 2 Some formulations of correlation for the compression index C_c of remoulded soils

Formulations	References
$C_c = 0.007(w_L - 10)$	[126]
$C_c = 0.5 I_P G_s$	[127]
$C_c = 0.2237 e_L$	[76]
$C_c = 0.329\left[0.027(w - w_P) + 0.0133 I_P\left(1.192 + A^{-1}\right)\right]$	[75]
$C_c = 0.2343 e_L$	[77]
$C_c = 0.256 e_L - 0.04$	[70]
$C_c = 0.009(w_L\text{-}13)$	[128]
$C_c = 0.014(I_P + 3.6)$	[81]
$\frac{C_c}{n_0} = 0.0109 C_c + 0.0018$, with $n_0 = \frac{100 e_0}{1 + e_0}$	[79]
$C_c = 0.015 I_P - 0.0198$	[78]
$\frac{C_c}{n_0} = 1.0584 n_0 + 0.0885$, with $n_0 = \frac{e_0}{1 + e_0}$	[83]

Table 3 Variable combinations in EPR model of C_c

Comb	Number of variables	Combinations
1	1	[e_0]
2	1	[w_L]
3	1	[I_P]
4	2	[e_0, w_L]
5	2	[e_0, I_P]
6	2	[w_L, I_P]
7	3	[e_0, w_L, I_P]

Table 4 Optimal correlations of C_c with different number of term sizes

Model number	*Comb*	*m*	Proposed optimal correlation expression
1	5	1	$C_c = 0.1576 e_0 w_L + 0.193$
2	7	2	$C_c = 0.1022 e_0 w_L + 0.1337 \frac{e_0 I_p}{w_L} + 0.139$
3	7	3	$C_c = 0.1104 e_0 w_L + 0.2743 \frac{e_0 I_p}{w_L} - 0.0532 \left(\frac{e_0 I_p}{w_L} \right)^2 + 0.0561$
4	7	4	$C_c = 0.1653 e_0 w_L + 0.23 \frac{e_0 I_p}{w_L} - 0.0676 \left(\frac{e_0 I_p}{w_L} \right)^2 - 0.0699 \frac{(w_L)^2}{I_p} + 0.135$
5	7	5	$C_c = 0.3206 e_0 w_L - 0.0284 e_0 (w_L I_p)^2 + 0.0213 (e_0 w_L I_p)^2 - \frac{0.1412 (w_L)^2}{I_p} - 0.0540 (e_0)^2 w_L I_p + 0.2325$

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

Table 5 Results of monotonicity analysis for variables involved in EPR model of C_c

Model number	e_0	w_L	I_p	Monotonous variables /Total involved variables	Ranking
1	1	1	-	2/2	1
2	1	0	1	2/3	2
3	1	0	0	1/3	3

4	1	0	0	1/3	3
5	1	0	0	1/3	3

Table 6 Results of ranking in terms of R^2, *Comb*, *m* and monotonicity for EPR models of C_c

Model number	R^2	*Comb*	*m*	Robustness ratio	Monotonicity	Selection index
1	5	1	1	1	1	0.841
2	4	2	2	1	2	0.802
3	3	2	3	1	3	0.785
4	2	2	4	1	3	0.785
5	1	2	5	1	3	0.785

Table 7 Summary of five indicators for the optimal model on training and testing data for C_c

Comb	*m*	Training					Testing				
		R^2	RMSE	MAE	u	σ	R^2	RMSE	MAE	u	σ
5	2	0.875	0.0727	0.054	1.036	0.192	0.848	0.0695	0.052	1.039	0.212

Table 8 Statistics of variables used in the database for S_u

Variable	Maximum	Minimum	Mean	Standard deviation
w_L (%)	201.8	22.0	66.40	23.51
I_P (%)	73.9	2.7	27.59	8.14
w (%)	180.1	17.3	74.26	23.26
S_t	64.0	2.0	16.29	13.12
OCR	7.5	1.0	1.778	0.898
σ'_v (kPa)	163.0	7.5	48.71	24.37
σ'_p (kPa)	270.0	20.0	78.85	38.49
s_u^{FV} (kPa)	75.0	5.0	20.10	10.15

Table 9 Current formulations of correlation for s_u

Formulations	References
$\frac{s_u^{FV}}{\sigma'_p} \approx 0.11 + 0.0037 I_p$	[129]
$\frac{s_u^{FV}}{\sigma'_p} \approx 0.08 + 0.0055 I_p$	[130]
$\frac{s_u(\text{mob})}{\sigma'_p} \approx 0.22$	[131]
$\frac{s_u(\text{mob})}{\sigma'_v} \approx (0.23 \pm 0.04)\text{OCR}^{0.8}$	[101]
$\frac{s_u(\text{mob})}{\sigma'_v} \approx S \cdot \text{OCR}^{m}$	[101]
$\frac{s_u^{DSS}}{\sigma'_v} \approx \left(0.125 + \frac{0.205 w_L}{1.17}\right)\text{OCR}^{0.8}$	[105]
$\frac{s_u^{DSS}}{\sigma'_v} \approx (0.14 + 0.18 w)\text{OCR}^{(0.35+0.77w)}$	[104]
$\frac{s_u(\text{mob})}{\sigma'_v} \approx 0.229\text{OCR}^{0.823} S_t^{0.121}$	[103]

Table 10 Variable combinations in EPR model of $s_u(\text{mob})/\sigma'_v$

Comb	Number of variables	Combinations
1	1	[w]
2	1	[w_L]
3	1	[OCR]
4	1	[S_t]
5	2	[w, w_L]
6	2	[w, OCR]
7	2	[w, S_t]
8	2	[w_L, OCR]
9	2	[w_L, S_t]
10	2	[OCR, S_t]
11	3	[w, w_L, OCR]
12	3	[w, w_L, S_t]

13	3	[w, OCR, S_t]
14	3	[w_L, OCR, S_t]
15	4	[w, w_L, OCR, S_t]

Table 11 Optimal correlations of $s_u(\text{mob})/\sigma'_v$ with different numbers of term sizes

Model number	m	Proposed optimal correlation expression
1	1	$\frac{s_u(\text{mob})}{\sigma'_v}=0.2605+0.045\cdot\text{OCR}^2$
2	2	$\frac{s_u(\text{mob})}{\sigma'_v}=0.27+0.0445\cdot\text{OCR}^2-0.0066\cdot\left(\frac{1}{w\cdot\text{OCR}}\right)^2$
3	3	$\frac{s_u(\text{mob})}{\sigma'_v}=0.303+0.0481\cdot\text{OCR}^2-0.0344\frac{1}{w\cdot\text{OCR}}-0.0082\cdot w\cdot\text{OCR}^2$

Remark: w is in real number, not in percentage.

Table 12 Results of monotonicity analysis for variables involved in EPR model of $s_u(\text{mob})/\sigma'_v$

Model number	w	OCR	Monotonous variables /Total involved variables
1	-	1	1/1
2	1	1	2/2
3	1	1	2/2

Table 13 Results of ranking in terms of R^2, *Comb*, m and monotonicity for EPR models of $s_u(\text{mob})/\sigma'_v$

Model number	R^2	*Comb*	m	Robustness ratio	Monotonicity	Selection index
1	3	1	1	1	1	0.693
2	2	2	2	1	1	0.653
3	1	2	3	1	1	0.653

Table 14 Summary of five indicators for the optimal model on training and testing data for $s_u(\mathrm{mob})/\sigma'_v$

Comb	*m*	Training					Testing				
		R^2	RMSE	MAE	u	σ	R^2	RMSE	MAE	u	σ
3	1	0.78	0.129	0.090	1.086	0.347	0.87	0.127	0.097	1.236	0.553

Table 15 Statistics of variables used in the database for hydraulic conductivity

Variable	Maximum	Minimum	Mean	Standard deviation
e	3.339	0.578	1.434	0.505
w_L (%)	678	44	210.3	202.8
I_P (%)	622.7	19	171.5	191.7
CI (%)	85.7	11.5	59.9	21.4
k (m/s)	6.79E-09	7.07E-12	4.95E-10	8.98E-10

Table 16 Current correlation formula for predicting hydraulic conductivity

Formulations	References
$e=\left(0.01I_p+0.05\right)\cdot\left[10+\log k\left(\mathrm{cm/s}\right)\right]$	[132]
$k\left(m/s\right)=\dfrac{0.0174I_p^{-4.29}\left[e-0.027\left(w_p-0.242I_p\right)\right]}{1+e}$	[133]
$\dfrac{e}{e_L}=2.162+0.195\log k\left(cm/s\right)$	[134]
$\dfrac{e}{e_L}=2.28+0.233\log k\left(cm/s\right)$	[135]
$\log k\left(m/s\right)=\dfrac{\left(e-0.0535w_L-5.286\right)}{\left(0.0063w_L+0.2516\right)}$	[136]
$\dfrac{e}{e_L}=2.23+0.204\log k\left(cm/s\right)$	[137]
$k\left(m/s\right)=\left(0.00104I_p^{-5.2}\right)\dfrac{e^5}{1+e}$	[119]
$k\left(m/s\right)=\exp\left(-5.51-4\ln I_p\right)\left(e\right)^{7.52\exp\left(-0.25I_L\right)}$	[118]

$k\left(m/s\right)=\dfrac{6.31\times10^{-7}}{\left(I_{\mathrm{P}}-8.74p\right)^{3.03}}e^{2.66\left(I_{\mathrm{P}}-8.74p\right)^{0.234}}$	[106]

Remarks: e: void ratio; e_{L}: void ratio at liquid limit; w_{L} (%):liquid limit; w_{P} (%):plastic limit; I_{P} (%): plasticity index; I_{L}: liquidity index; p: the percentage of clay minerals in the soil divided by 100.

Table 17 Variable combinations in EPR model of k

Comb	Number of variables	Combinations
1	1	[CI]
2	1	[w_{L}]
3	1	[I_{P}]
4	2	[CI, w_{L}]
5	2	[CI, I_{P}]
6	2	[w_{L}, I_{P}]
7	3	[CI, w_{L}, I_{P}]

Table 18 Optimal correlations of k with different number of term sizes

Model number	m	Proposed optimal correlation expression
1	3	$\log k=\left(-1.0334I_{\mathrm{P}}+0.9435\dfrac{w_{\mathrm{L}}^2}{I_{\mathrm{P}}}+\dfrac{0.0762}{w_{\mathrm{L}}I_{\mathrm{P}}}\right)e-10.9919$
2	5	$\log k=\left(0.348\dfrac{w_{\mathrm{L}}^2}{I_{\mathrm{P}}CI}-0.063\left(\dfrac{CI}{w_{\mathrm{L}}I_{\mathrm{P}}}\right)^2-0.2427\dfrac{I_P^2CI}{w_{\mathrm{L}}}-0.1597\dfrac{I_P^2}{CI^2w_{\mathrm{L}}}+0.415\dfrac{CI^2}{w_{\mathrm{L}}^2I_{\mathrm{P}}}\right)e-11.092$
3	6	$\log k=\left(0.9925\dfrac{CIw_{\mathrm{L}}^2}{I_P^2}-1.2737\left(\dfrac{I_{\mathrm{P}}}{w_{\mathrm{L}}}\right)^2+0.2147\dfrac{w_{\mathrm{L}}^2}{I_{\mathrm{P}}CI}-0.7354\dfrac{CI}{I_P^2}+0.6546\dfrac{CI}{w_{\mathrm{L}}^2I_{\mathrm{P}}}-0.0277\dfrac{I_P^2}{CI}\right)e-11.0984$
4	7	$\log k=\left(\dfrac{-0.0119CI}{w_{\mathrm{L}}^2I_P^2}-0.8036\left(\dfrac{I_{\mathrm{P}}}{w_{\mathrm{L}}}\right)^2+\dfrac{0.3402w_{\mathrm{L}}^2}{I_{\mathrm{P}}CI}+0.3386CI-\dfrac{0.2798I_{\mathrm{P}}}{w_{\mathrm{L}}CI}-\dfrac{0.0395I_P^2}{CI}+\dfrac{0.6738CI}{w_{\mathrm{L}}^2}\right)e-11.0896$
5	8	$\log k=\left(-0.0612w_{\mathrm{L}}^2+0.204\left(\dfrac{CIw_{\mathrm{L}}}{I_{\mathrm{P}}}\right)^2-\dfrac{0.1061I_{\mathrm{P}}}{CI}+\dfrac{0.4191w_{\mathrm{L}}^2}{I_{\mathrm{P}}CI}+0.0006\left(CIw_{\mathrm{L}}I_{\mathrm{P}}\right)^2-\dfrac{0.4375I_P}{w_{\mathrm{L}}CI}+\dfrac{0.5711CI}{I_Pw_{\mathrm{L}}^2}-\dfrac{0.2243CI}{I_P^2w_{\mathrm{L}}}\right)e-11.1251$

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

Table 19 Results of monotonicity analysis for variables involved in EPR model of k

Model number	CI	w_L	I_p	e	Monotonous variables /Total involved variables
1	-	1	1	1	3/3
2	0	0	0	1	1/4
3	0	0	0	1	1/4
4	0	0	0	1	1/4
5	0	0	0	1	1/4

Table 20 Results of ranking in terms of R^2, *Comb*, m and monotonicity for EPR models of k

Model number	R^2	*Comb*	m	Robustness ratio	Monotonicity	Selection index
1	5	1	1	1	1	0.835
2	4	2	2	1	2	0.791
3	3	2	3	1	2	0.791
4	2	2	4	1	2	0.791
5	1	2	5	1	2	0.791

Table 21 Summary of five indicators for the optimal model on training and testing data for k

Comb	m	Training					Testing				
		R^2	RMSE	MAE	u	σ	R^2	RMSE	MAE	u	σ
6	3	0.740	1.71e-9	3.56e-10	1.25	0.99	0.761	1.67e-9	4.59e-10	1.36	0.98

Figure captions

Fig. 1 Typical flowchart of EPR procedure

Fig. 2 Procedure of intelligent multi-step selection MOOP-based EPR process

Fig. 3 Flow chart of proposed MODE

Fig. 4 Comparison of predictions and measurements for EPR model and empirical correlations

Fig. 5 Obtained Pareto front by proposed EPR procedure for C_c: (a) Pareto fronts obtained by MODE and NSGA-II; (b) Pareto front by MODE in space of R^2, robustness ratio and term size; (c) Pareto front by MODE in space of R^2 and term size

Fig. 6 Comparison of C_c between measurements and EPR predictions

Fig. 7 Comparison of $s_u(\text{mob})/\sigma'_v$ between predictions and measurements for empirical correlations

Fig. 8 Obtained Pareto front of proposed EPR procedure for $s_u(\text{mob})/\sigma'_v$

Fig. 9 Comparison of $s_u(\text{mob})/\sigma'_v$ between measurements and EPR predictions

Fig. 10 Comparison of k between predictions and measurements for empirical correlations

Fig. 11 Obtained Pareto front by proposed EPR procedure for k

Fig. 12 Comparison of k between measurements and EPR predictions

Figure 1

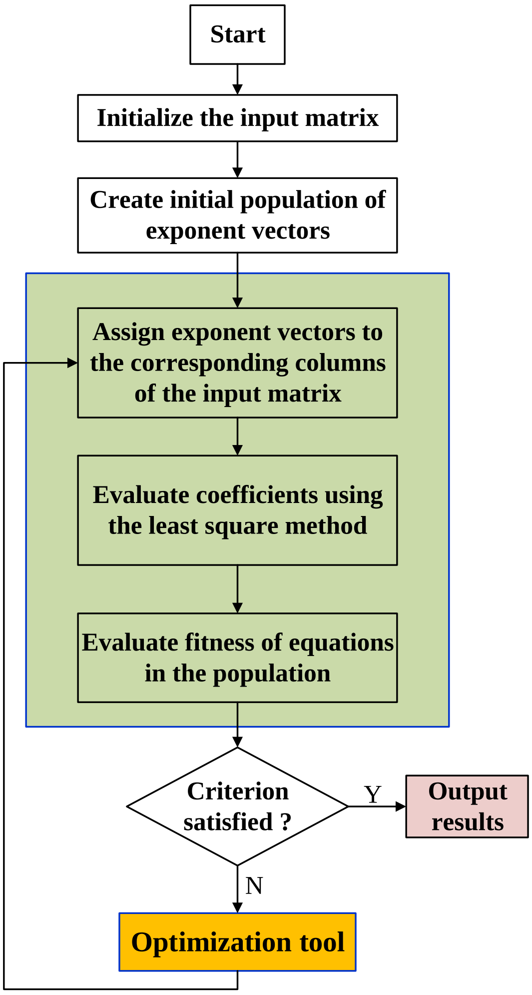
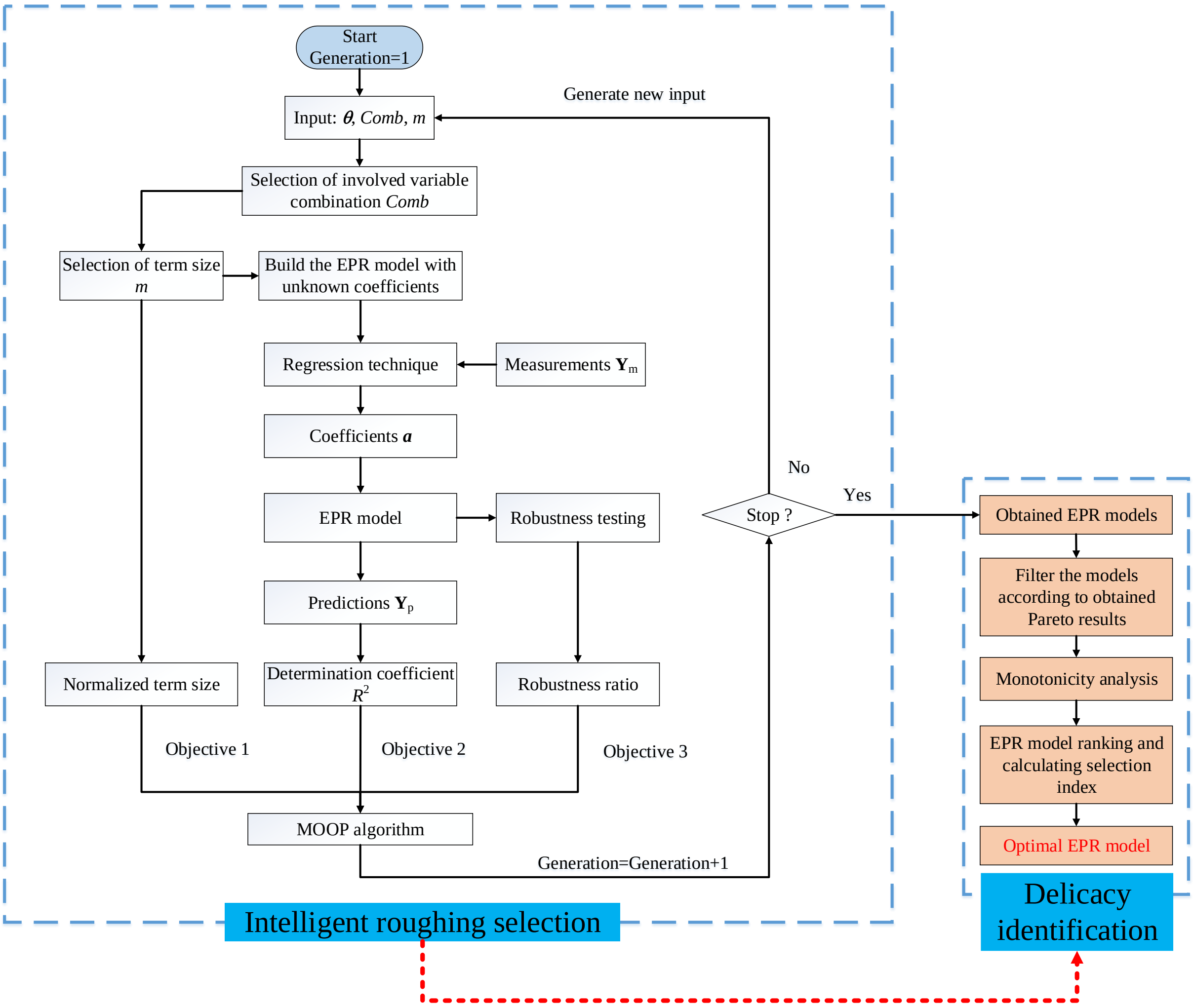
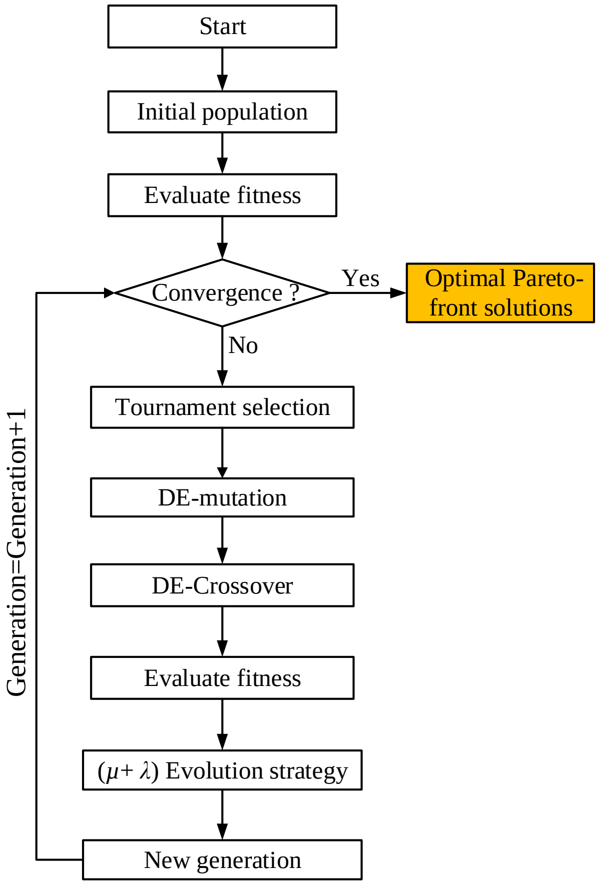

Start
Initialize the input matrix
Create initial population of exponent vectors
Assign exponent vectors to the corresponding columns of the input matrix
Evaluate coefficients using the least square method
Evaluate fitness of equations in the population
Criterion satisfied ?
Y
Output results
N
Optimization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

Figure 2



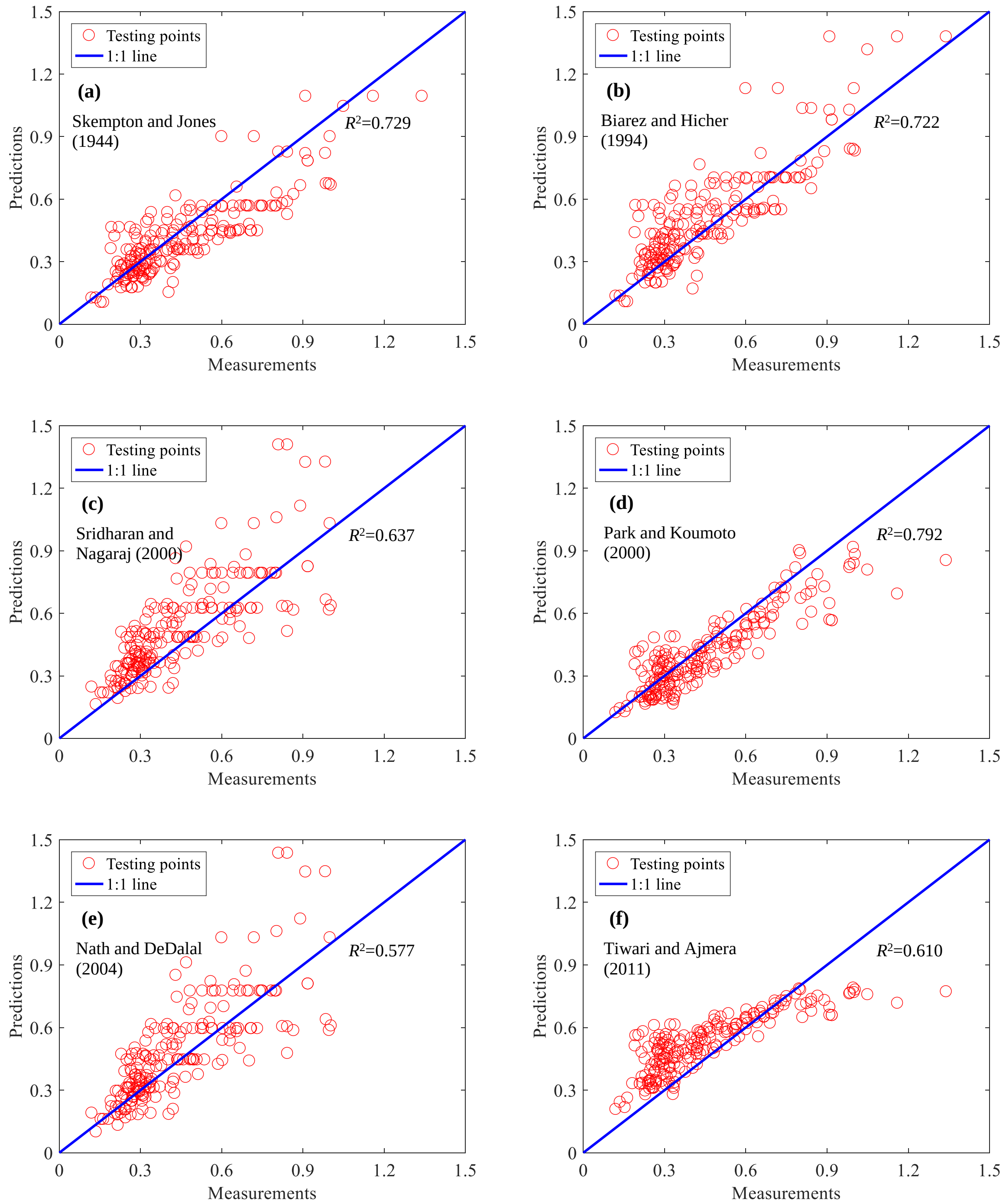
Start
Generation=1
Input: θ, *Comb*, *m*
Generate new input
Selection of involved variable combination *Comb*
Selection of term size *m*
Build the EPR model with unknown coefficients
Regression technique
Measurements $\mathbf{Y}_m$
Coefficients $\boldsymbol{a}$
EPR model
Robustness testing
Predictions $\mathbf{Y}_p$
Normalized term size
Determination coefficient R^2
Robustness ratio
Objective 1
Objective 2
Objective 3
MOOP algorithm
Generation=Generation+1
No
Yes
Stop ?
Obtained EPR models
Filter the models according to obtained Pareto results
Monotonicity analysis
EPR model ranking and calculating selection index
Optimal EPR model
Intelligent roughing selection
Delicacy identification

Figure 3



Start
Initial population
Evaluate fitness
Convergence ?
Yes
Optimal Pareto-front solutions
No
Tournament selection
DE-mutation
DE-Crossover
Evaluate fitness
(μ+ λ) Evolution strategy
New generation
Generation=Generation+1

Figure 4



Testing points
1:1 line
(a)
Skempton and Jones (1944)
R^2=0.729
(b)
Biarez and Hicher (1994)
R^2=0.722
(c)
Sridharan and Nagaraj (2000)
R^2=0.637
(d)
Park and Koumoto (2000)
R^2=0.792
(e)
Nath and DeDalal (2004)
R^2=0.577
(f)
Tiwari and Ajmera (2011)
R^2=0.610
Predictions
Measurements
0
0.3
0.6
0.9
1.2
1.5

Figure 5

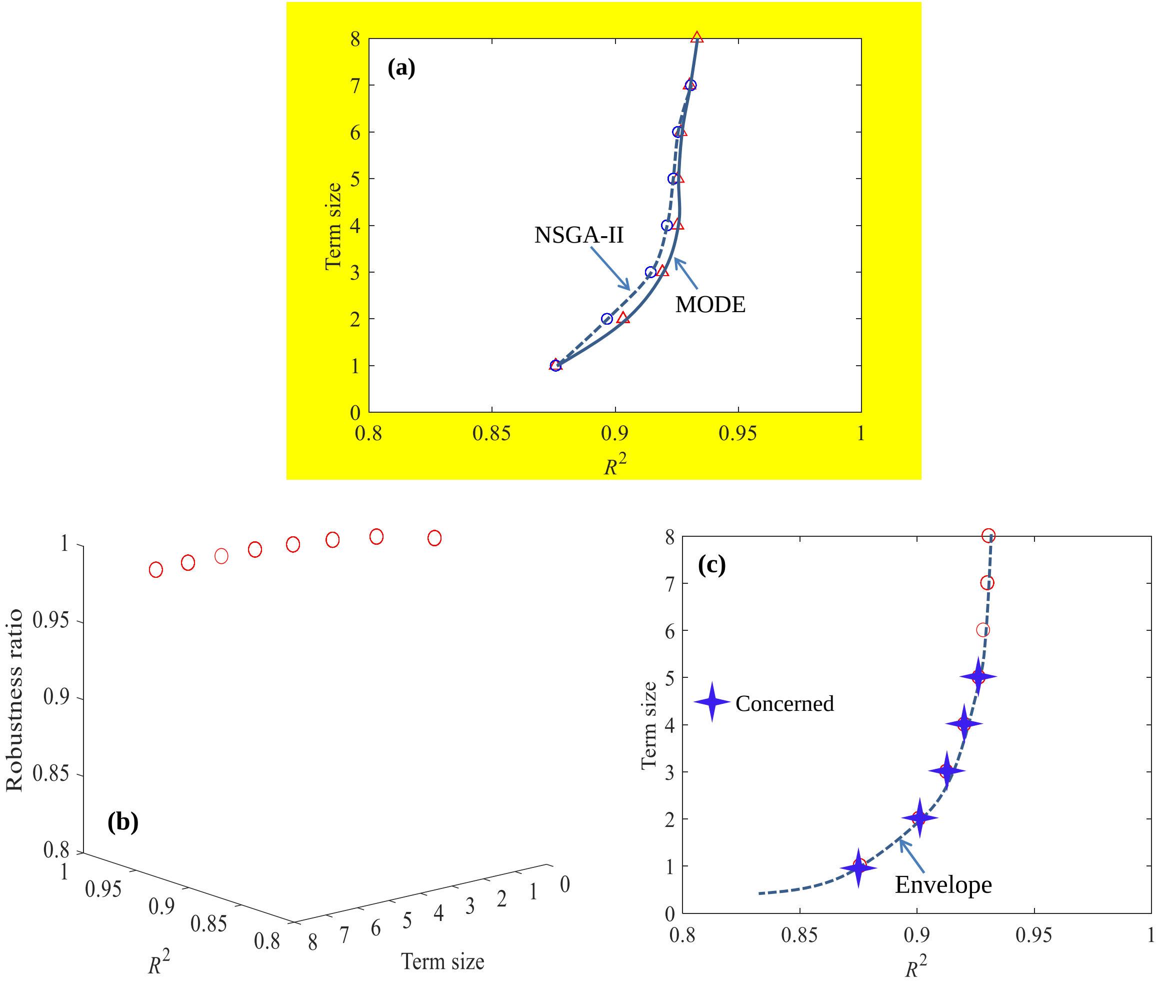
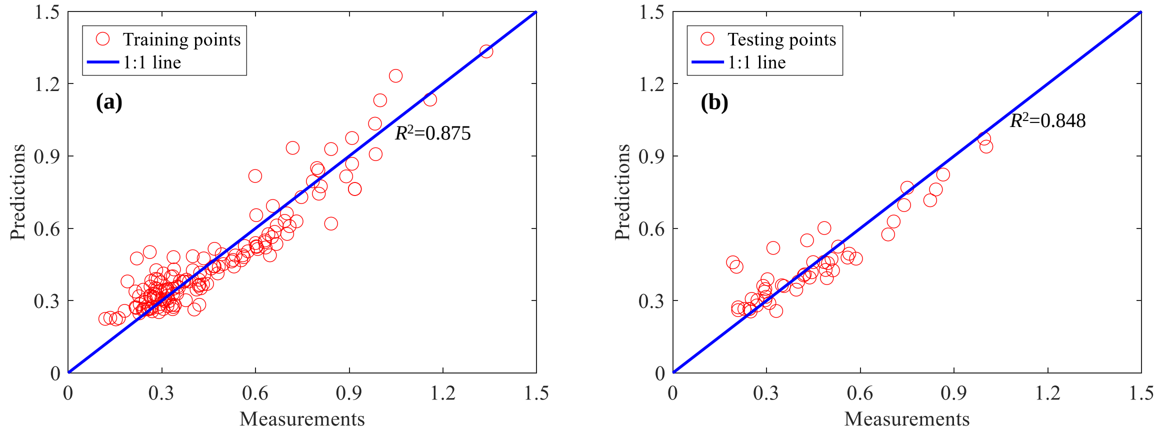
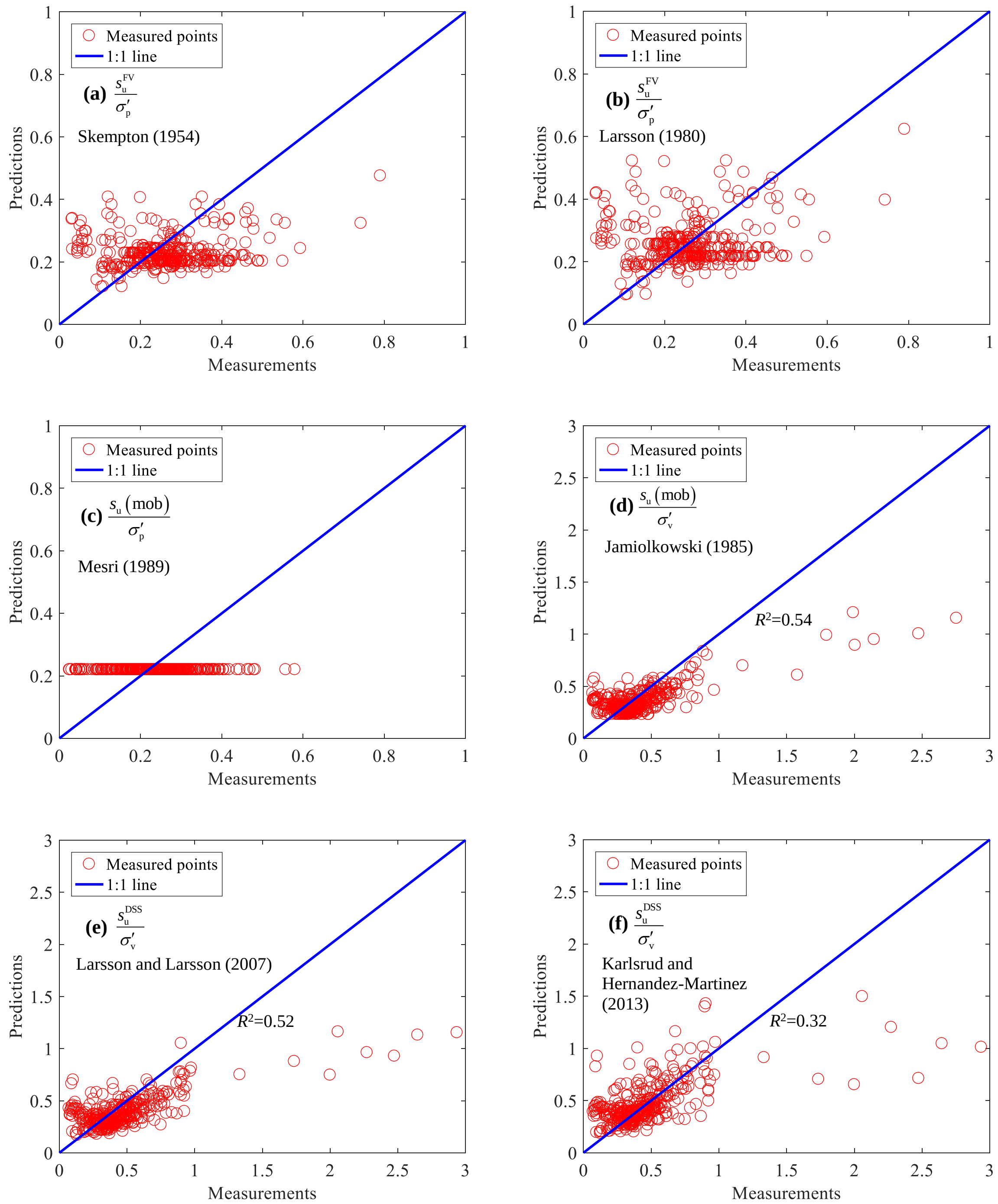

(a)
NSGA-II
MODE
Term size
R^2
(b)
Robustness ratio
Term size
R^2
(c)
Concerned
Envelope
Term size
R^2

Figure 6



Training points
1:1 line
(a)
R^2=0.875
Predictions
Measurements
Testing points
1:1 line
(b)
R^2=0.848
Predictions
Measurements

Figure 7



Measured points
1:1 line
(a) $\frac{s_u^{FV}}{\sigma'_p}$
Skempton (1954)
(b) $\frac{s_u^{FV}}{\sigma'_p}$
Larsson (1980)
(c) $\frac{s_u(\text{mob})}{\sigma'_p}$
Mesri (1989)
(d) $\frac{s_u(\text{mob})}{\sigma'_v}$
Jamiolkowski (1985)
R^2=0.54
(e) $\frac{s_u^{DSS}}{\sigma'_v}$
Larsson and Larsson (2007)
R^2=0.52
(f) $\frac{s_u^{DSS}}{\sigma'_v}$
Karlsrud and
Hernandez-Martinez
(2013)
R^2=0.32
Predictions
Measurements

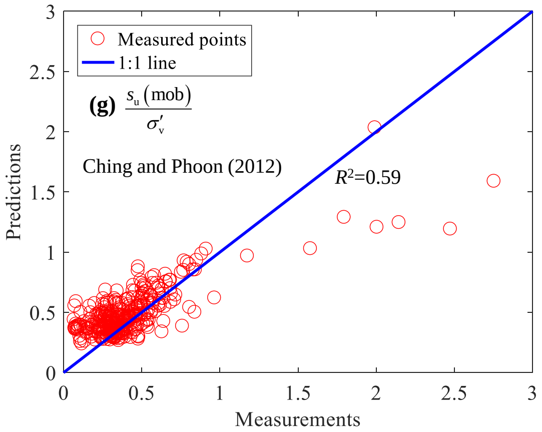
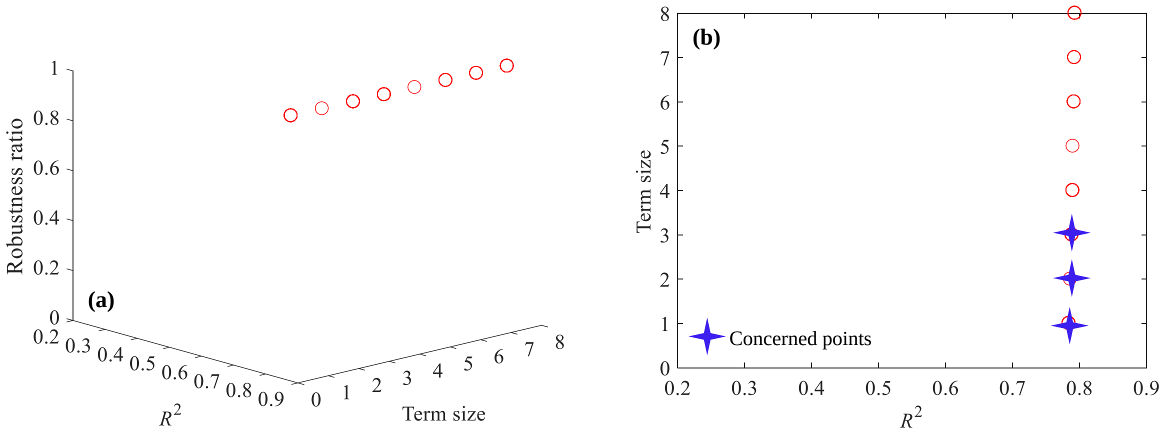
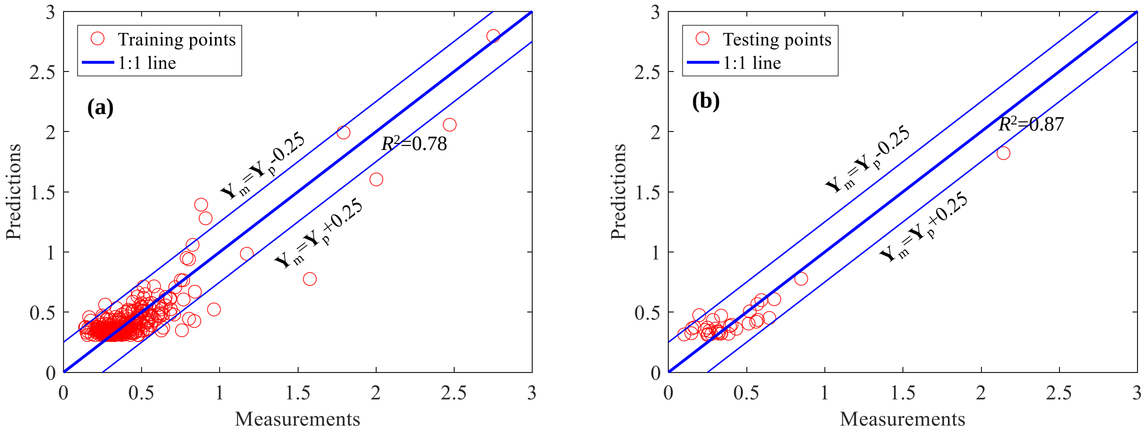

Measured points
1:1 line
(g) $\frac{s_u(\text{mob})}{\sigma'_v}$
Ching and Phoon (2012)
R^2=0.59
Predictions
Measurements
0
0.5
1
1.5
2
2.5
3

Figure 8



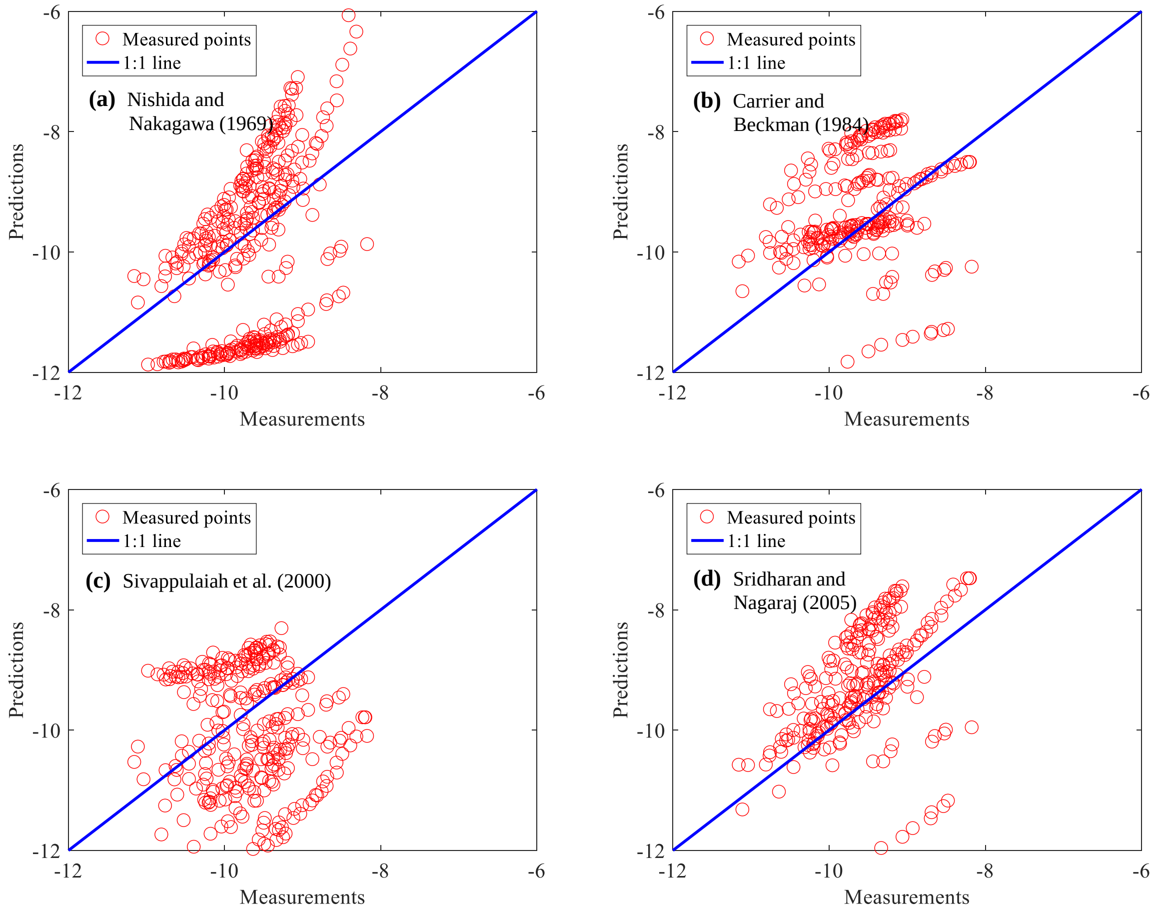
(a)
Robustness ratio
R^2
Term size
(b)
Term size
Concerned points
R^2

Figure 9



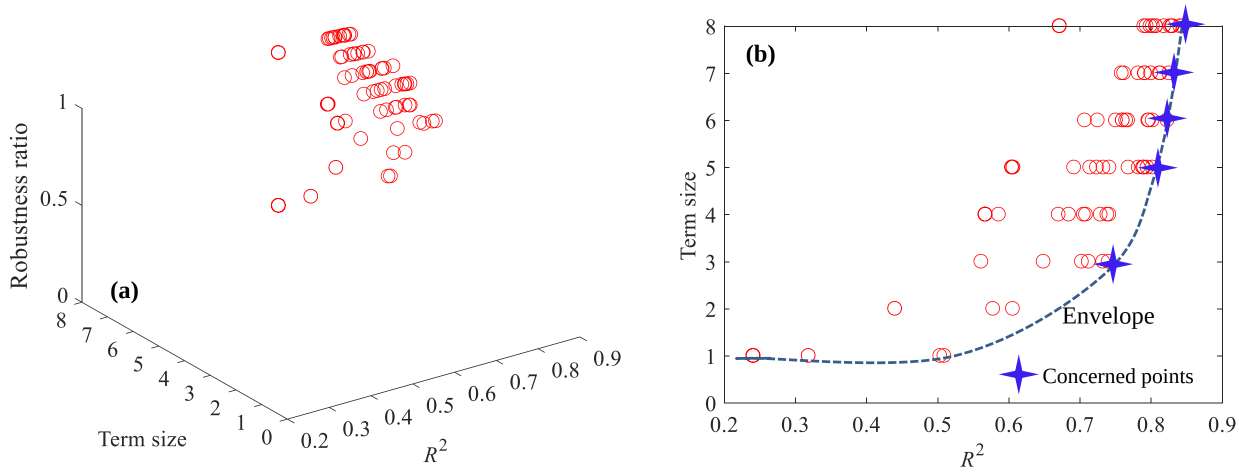
Training points
1:1 line
(a)
Y_m=Y_p-0.25
Y_m=Y_p+0.25
R²=0.78
Predictions
Measurements
Testing points
1:1 line
(b)
Y_m=Y_p-0.25
Y_m=Y_p+0.25
R²=0.87
Predictions
Measurements

Figure 10



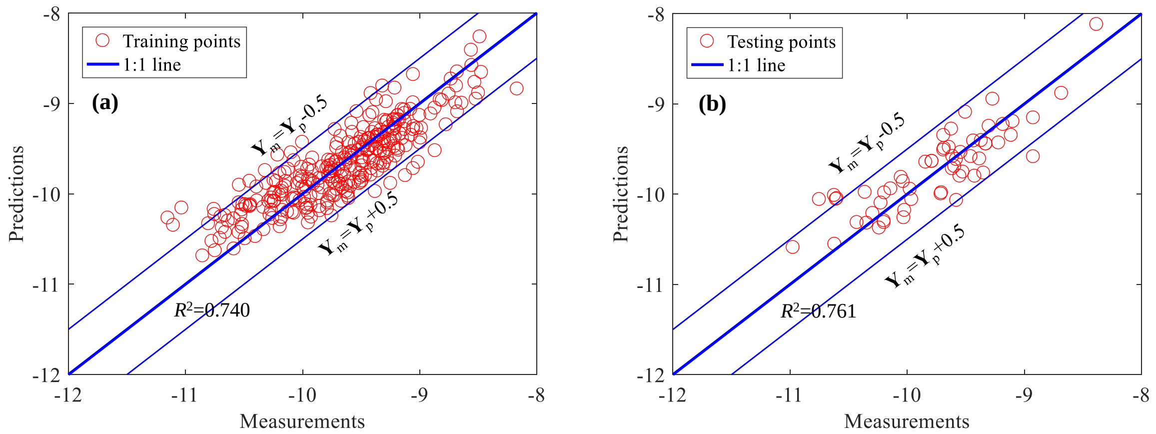
Measured points
1:1 line
(a) Nishida and Nakagawa (1969)
(b) Carrier and Beckman (1984)
(c) Sivappulaiah et al. (2000)
(d) Sridharan and Nagaraj (2005)
Predictions
Measurements
-6
-8
-10
-12

Figure 11



(a)
Robustness ratio
Term size
R^2
(b)
Term size
R^2
Envelope
Concerned points

Figure 12



Training points
1:1 line
(a)
$Y_m=Y_p-0.5$
$Y_m=Y_p+0.5$
R^2=0.740
Predictions
Measurements
Testing points
1:1 line
(b)
$Y_m=Y_p-0.5$
$Y_m=Y_p+0.5$
R^2=0.761
Predictions
Measurements