Soft Computing (2023) 27:16265–16278
https://doi.org/10.1007/s00500-023-08053-8

APPLICATION OF SOFT COMPUTING

Prediction of creep index of soft clays using gene expression programming

Xinhua Xue[1] · Chubing Deng[1]

Accepted: 13 March 2023 / Published online: 22 March 2023

Abstract

The creep index plays an important role in calculating the long-term settlement of natural soft clays, so it is vital to determine the creep index quickly and accurately. However, the prediction accuracy of the existing creep index models is low. This study presents seven gene expression programming (GEP) models by using different combinations of the liquid limit w_L, plasticity index I_p, void ratio e and clay content CI as input variables for the prediction of creep index. A total of 151 datasets were collected from the available literature for building and testing the GEP models. The proposed GEP models were compared with two machine learning (ML) models (i.e., back propagation neural network and random forest) and five conventional empirical models in terms of three statistical indicators. The research results showed that the prediction performances of the two proposed GEP models (i.e., with combinations $CI{-}w_L{-}e$ and $CI{-}I_p{-}w_L{-}e$ as input, respectively) surpass those of the five conventional empirical models and two ML-based models, recommended for predicting the creep index of natural soft clays in engineering practice.

Keywords Gene expression programming · Creep index · Plasticity index · Liquid limit · Void ratio · Clay content

1 Introduction

For the analysis and design of slope stability or the safety of tunneling or embankments, the long-term settlement calculation of these infrastructures is vital in order to control the post construction settlement within an allowed range (Shen et al. 2014; Meng et al. 2018; Yang et al. 2019; Zhang et al. 2020; Zhu et al. 2020). Currently, the most widely used methods to calculate the long-term settlement of soft clays is the standard or advanced elastic viscoplastic models (Yin et al. 2010, 2015; Tan et al. 2018). However, the determination of viscosity parameters requires a lot of time for researchers and engineers, which made it difficult to be used in practice.

Creep index C_α is one of the key parameters when using the standard or advanced elastic viscoplastic models to calculate the long-term settlement of soft clays (Karstunen and Yin 2010; Yin et al. 2017). Although the creep index is not an intrinsic property of intact clays, it can provide a help for understanding the creep behaviors of remoulded clays (Zhang et al. 2020). Therefore, it is very important to determine the creep index quickly and accurately.

Previous studies have proved that the microstructure of soft clays has an important effect on its creep property, and the calculation formula of creep index is usually determined by regression analysis technology based on experimental data (Yin et al. 2009, 2014a; b). For example, Nakase et al. (1988) developed an empirical model involving the plasticity index I_p and the creep index C_α; Zeng et al. (2012) proposed an empirical model describing the relationships between the creep index C_α and the void ratio at liquid limit e_L and the void ratio e; Yin et al. (2015) proposed an elastic-viscoplastic model of natural soft clay, which takes nonlinear creep into account; Zhu et al. (2016) further proposed an empirical model considering both the plasticity index I_p and liquid limit w_L. However, the models proposed by Nakase et al. (1988), Zeng et al. (2012), Yin (1999), Yin et al. (2015) and Zhu et al. (2016) have fewer influencing factors and lower prediction accuracy. It cannot

✉ Xinhua Xue
scuxxh@163.com

[1] State Key Laboratory of Hydraulics and Mountain River Engineering, College of Water Resource and Hydropower, Sichuan University, Chengdu 610065, People's Republic of China

provide a very accurate reference for the determination of creep index C_α.

In recent years, artificial intelligence (AI) methods have been widely used in the field of geotechnical engineering (Sharma et al. 2021; Jong et al. 2021; Zhang et al. 2021). For example, Gordan et al. (2016) investigated the seismic slope stability by using the hybrid model of artificial neural networks (ANNs) and particle swarm optimization (PSO). Koopialipoor et al. (2019) predicted the safety factor (SF) of slopes by using the PSO-ANN model. Fattahi (2017) evaluated the slope stability using the adaptive neuro-fuzzy inference system (ANFIS) model. Previous studies have shown that the predictive performance of empirical models is far inferior to that of models proposed based on AI techniques; however, AI techniques may face some issues like trapping in local minima (Xiong et al. 2004; Sun et al. 2016; Zhang et al. 2016). The gene expression programming (GEP) was first proposed by Ferreira (2001, 2006) to solve some problems in genetic programming (GP). The biggest difference between the two is GEP's use of linear fixed length expression tree (ET), which is the key to GEP's ability to solve relatively complex problems with high performance (Jafari and Mahini 2017; Murad et al. 2019).

Considering the lack of prediction accuracy of existing models and the advantages of AI technologies, the main purpose of this study is to propose a new creep index model based on GEP technique. The biggest difference between GEP technique and most regression technologies is that when establishing the functional relationship between creep index and various parameters, GEP technology only needs to consider the possible parameters in the relationship, and does not need to specify predefined functions. The main contributions of this study can be clarified as follows:

(1) To the best of our knowledge, this study is the first method using GEP technology to predict the creep index of soft clays in the literature;
(2) In this paper, the influence of different parameter combinations on the prediction accuracy of the model is studied, and seven GEP models are proposed according to the results of seven combinations;
(3) Compared with the literature models, the GEP model established in this paper considers more influencing factors for prediction of the creep index of soft clays;
(4) Compared with other ML-based models (i.e., back propagation (BP) neural network and random forest (RF)), GEP model has relatively simple calculation formula and high prediction accuracy, which is conducive to popularization and application. In addition, in order to facilitate the application of GEP model in engineering, we developed a convenient graphical user interface.

The rest of the paper is organized as follows: Database is presented in Sect. 2. Methodologies are explained in Sect. 3. Results and discussion are presented in Sect. 4. Finally, conclusions are introduced in Sect. 5.

2 Database

The choice of input variables is vital to the accurate prediction of creep index C_α. In this study, four physical parameters (i.e., liquid limit w_L, void ratio e, plasticity index I_p and clay content CI) of soft clays were taken as the input variables and the creep index C_α as the output variable. A total of 151 sets of data points collected from the literature (Shen et al. 2014; Meng et al. 2018; Tan et al. 2018; Yang et al. 2019; Zhu et al. 2020) were used to establish the GEP model. Figure 1 and Tables 1 and 2 show the frequency distribution histogram of the four input parameters, the statistical results of the database and the Pearson correlation analysis results between different parameters, respectively. As observed from Fig. 1 and Table 1, these parameters have a wide range of values, which is enough to make the proposed GEP model have better generalization and application. The 151 sets of data were randomly divided into two parts: training set (120 groups) and test set (31 groups), which were used for model establishment and evaluation, respectively.

3 Methodology

3.1 GEP

GEP, a machine learning algorithm based on genetic algorithm (GA) and genetic programming (GP), was first proposed by Ferreira (2001). Generally speaking, GEP is mainly composed of five parts: termination condition, fitness function, terminal set, control parameters and function set. In GEP, a genome or chromosome may contain one or more genes, and a gene can be divided into a head containing terminals and functions (e.g., variables, mathematical operators and functions) and a tail containing only terminals (e.g., variables and constants). In this study, GeneXproTools5.0 software was used to establish the GEP model. The main steps of GEP establishment are summarized as follows:

(1) Selecting an appropriate fitness function is conducive to the successful solution of the problem. In this study, the following equation was used as the fitness value:

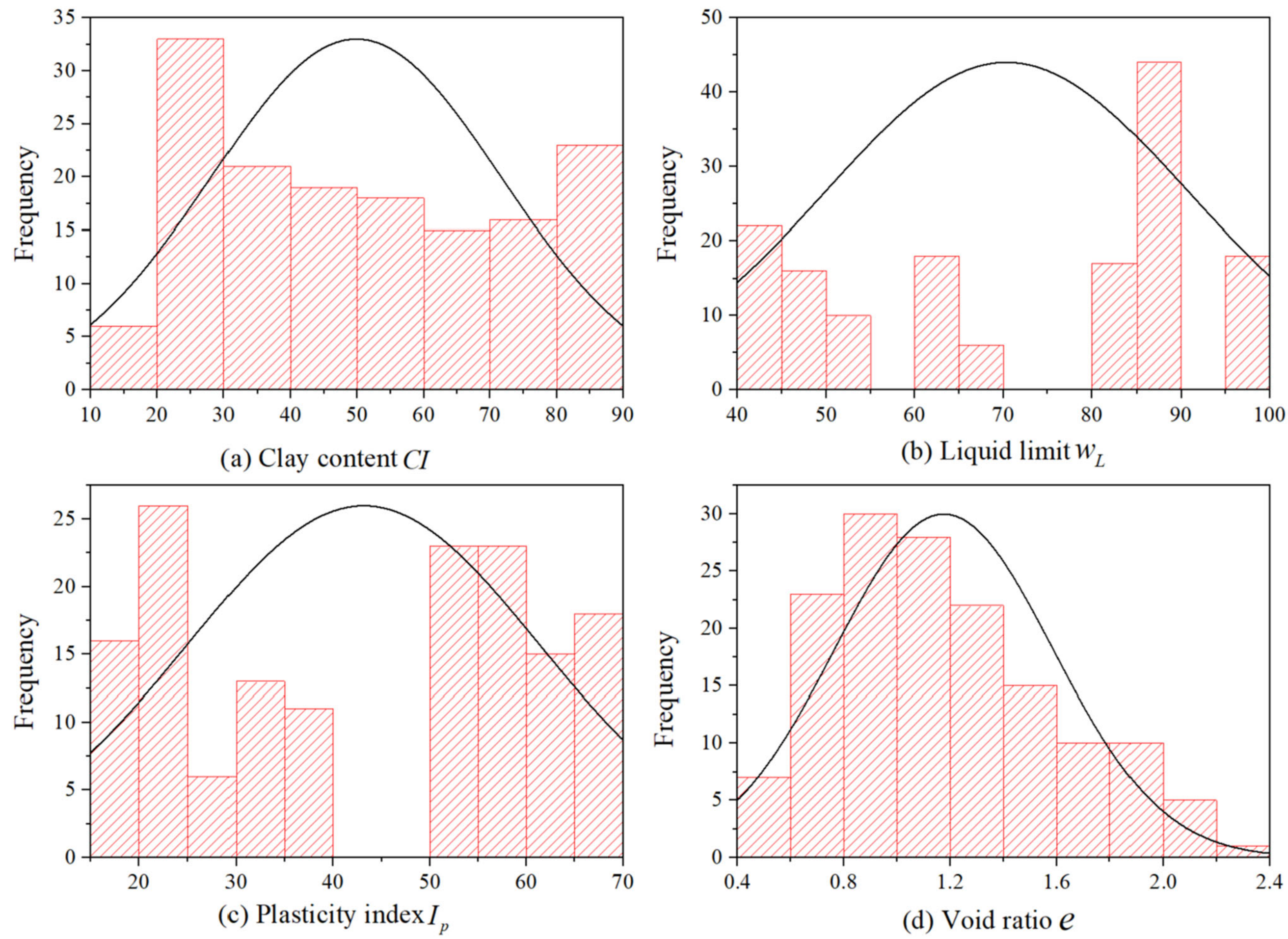


Fig. 1 Histograms of the four input parameters

Table 1 Statistics analysis of datasets

Parameter	Min	Mean	Max	Standard deviation
CI	11.5	49.86	83	21.69
w_L	40	70.35	98	20.32
I_p	19	43.16	68	18.11
e	0.466	1.18	2.28	0.41
C_α	0.055	0.019	0.0036	0.011

Table 2 Results of Pearson correlation analysis

	CI	w_L	I_p	e	C_α
CI	1				
w_L	− 0.014	1			
I_p	0.090	0.985**	1		
e	0.260**	0.716**	0.749**	1	
C_α	− 0.001	0.769**	0.769**	0.850**	1

**Correlation is significant at the 0.01 level (two-tailed)

$$f_i = 1000\left[1 + \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(Y_{i,j} - X_i\right)^2}\right]^{-1} \tag{1}$$

where f_i represents the fitness value and it ranges between 0 and 1000 (ideally, the fitness value is 1000); m represents the total number of chromosomes; $Y_{i,j}$ and X_i represent the value predicted by the individual chromosome i for fitness case j and the monitored value for fitness case i, respectively.

(2) The sets of functions F and terminals T need to be selected. Obviously, the set of functions F consists of all the function symbols that may appear in the formula, thus giving F = {Inv, X^2, 3Rt, *, /, +, −, exp, Ln}. The set of terminals T consists of input and output parameters, thus giving T = {w_L,e,I_p,CI,C_α}.

(3) In this study, the optimal value of each parameter in the GEP model is determined by the trial-and-error

Table 3 Optimal parameters of the first GEP model

Parameter	Value
Population size	50
Gene recombination rate	0.1
Chromosome length	60
Linking function	Multiplication (×)
Mutation rate	0.044
Gene transposition rate	0.1
Head size	12
IS transposition rate	0.1
One-point recombination rate	0.3
RIS transposition rate	0.1
Two-point recombination rate	0.3

strategy, and the selected parameters are listed in Table 3.

(4) Select the type of linking function. In the GEP model, there are many linking functions. (e.g., multiplication (×), subtraction (−), addition (+), division (/), Min, Max, CL2D {0,1}, CL2A {− 1,1}, CL3A {− 1,0,1}, CL3B {− 1,0,1}, CL3C {− 1,0,1} and AMin2 {0,1}). In this study, the linking functions of addition (+) and multiplication (×) are selected because they can provide better results than other linking functions.

(5) Select the genetic operators. In this study, the selection of genetic operators is mainly based on the research results of Ferreira (2006), and the selected genetic operators are listed in Table 3.

Figure 2 illustrates the flowchart of GEP model.

3.2 Empirical formulas

According to Zhang et al. (2020), some available empirical formulas are listed as follows:

(1) Developed by Nakase et al. (1998):

$$C_\alpha = 0.00168 + 0.00033 I_p \tag{2}$$

(2) Developed by Yin (1999):

$$C_\alpha = 0.000369 I_p - 0.00055 \tag{3}$$

(3) Developed by Zeng et al. (2012):

$$C_\alpha = \left(-0.0067 + 0.0115 e_L - 0.0016 (e_L)^2\right)(1 + e) \tag{4}$$

(4) Developed by Zhu et al. (2016):

$$C_\alpha = (0.0007 w_L - 0.0223)\left(\frac{w_L}{w}\right)^{0.23031 - 0.014978 w_L} \tag{5}$$

(5) Developed by Zhu et al. (2020):

$$C_\alpha = \left(-0.0274 + 0.0011 w_L - 0.00048 I_p\right) \times \left(\frac{w}{w_L}\right)^{0.7872 - 0.0369 w_L + 0.0619 I_p} \tag{6}$$

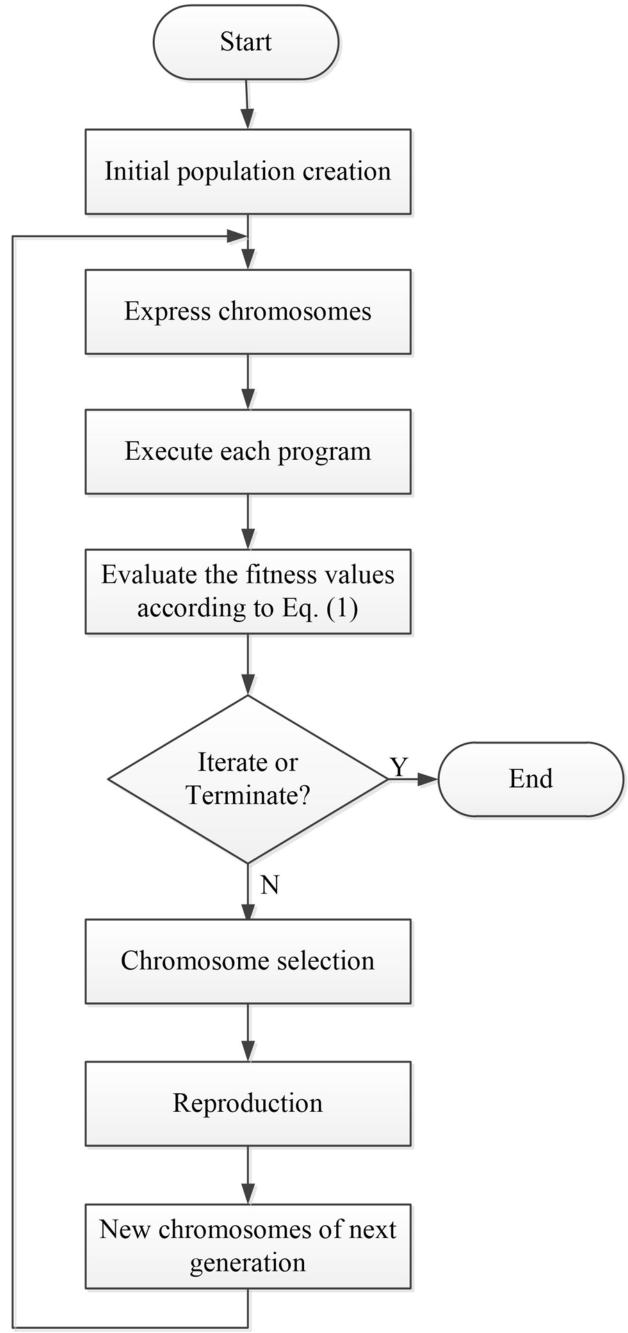


Fig. 2 Flowchart of GEP (Ferreira 2006)

3.3 RF

RF, first proposed by Breiman (2001), is a supervised ML algorithm composed of decision trees. The following is a brief introduction to the RF algorithm.

The training data is drawn randomly from the distribution of the random vector S and T and assuming that $\mathrm{h}_1(\mathrm{x}),\mathrm{h}_2(\mathrm{x}),\ldots,\mathrm{h}_k(\mathrm{x})$ are ensemble of classifiers, then the margin function $mg(S,T)$ can be expressed as (Breiman 2001):

$$mg(S,T) = I(\mathrm{h}_k(S) = T)v_k a - \max_{j \neq \mathrm{Y}} I(\mathrm{h}_k(S) = j)v_k a \tag{7}$$

where $I(\cdot)$ denotes the indicator function. It should be noted that the confidence in the classification is proportional to the margin.

The calculation formula of generalization error G_e is given as follows (Breiman 2001):

$$G_\mathrm{e} = G_{S,T}(mg(S,T) < 0) \tag{8}$$

where the subscripts S,T indicate that the probability is over the S,T space. More details of RF can be found in Breiman (2001).

3.4 BP

The BP neural network is generally composed of three layers of neurons, which are (1) the output layer, (2) the hidden layer, and (3) the input layer. The gradient descent algorithm is commonly used to train BP neural network by adjusting the weights to minimize the total error between the actual output and the target output. More details of BP neural network can be found in Xue (2017).

3.5 Evaluation of the proposed creep index models

In order to evaluate the prediction performance of each creep index model, three statistical indices, namely, the root mean squared error (RMSE), mean absolute percentage error (MAPE) and mean absolute error (MAE) are used in this study, and the calculation formulas are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(O_i - P_i)^2}{n}} \tag{9}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{O_i - P_i}{P_i}\right| \tag{10}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|O_i - P_i| \tag{11}$$

where O_i and P_i represent the actual and predicted results, respectively. n is the total number of data (n = 151). Obviously, the lower values of these three indicators, the better the prediction performance.

4 Results and discussion

In this study, different combinations of four input variables were used to obtain seven GEP models for predicting the creep index C_α of soft clays. The optimal parameters of the first GEP model are listed in Table 3. Because the setting parameters of the other six GEP models are fine-tuned on this basis, they are not listed separately in this study.

4.1 Combination of $CI - e$

The expression tree of the first established GEP model consists of five sub-expression trees, as shown in Fig. 3. In Fig. 3, the constants of the first sub-expression tree (gene) c_6 and c_9 are 10.69 and -9.365, respectively. The constant of the third sub-expression tree c_2 is 2.566, and the constants of the fourth sub-expression tree (gene) c_6 and c_8 are -0.899 and 4.90, respectively. The linking function of this model is multiplication ($\times$), and the expression of the model can be written as:

$$C_a = \frac{\left(\sqrt{CI} + CI + e \cdot CI + 2.566\right)\left[\sin\left(\frac{\sin CI + CI}{CI}\right)\right]^{-1}}{(\sin CI + CI + 98.415)(2\sin CI + CI + e)(e - \tan CI - 4.409)} \tag{12}$$

4.2 Combination of $CI - I_p - e$

The expression tree of the second established GEP model consists of two sub-expression trees, as shown in Fig. 4. In Fig. 4, the constants of the first sub-expression tree c_6 and c_7 are 1.17 and $-$ 9.67, respectively. The constants of the second sub-expression tree c_3 and c_9 are 8.69 and $-$ 16.94, respectively. The linking function of this model is addition ($+$), and the expression of the model can be written as:

$$C_a = \frac{e^3}{93.559 + 109.661e} + \frac{(CI + 16.934)I_p}{e^{8.679}CI} \tag{13}$$

4.3 Combination of $CI - w_L - e$

The expression tree of the third established GEP model consists of three sub-expression trees, as shown in Fig. 5. In Fig. 5, the constants of the first sub-expression tree c_0,c_1,c_5,c_8 and c_9 are $-$ 68.312, $-$ 88.91, 87.29, $-$ 59.56 and 75.13, respectively. The constants of the second sub-expression tree c_8 and c_9 are 50.21 and 35.14, respectively.

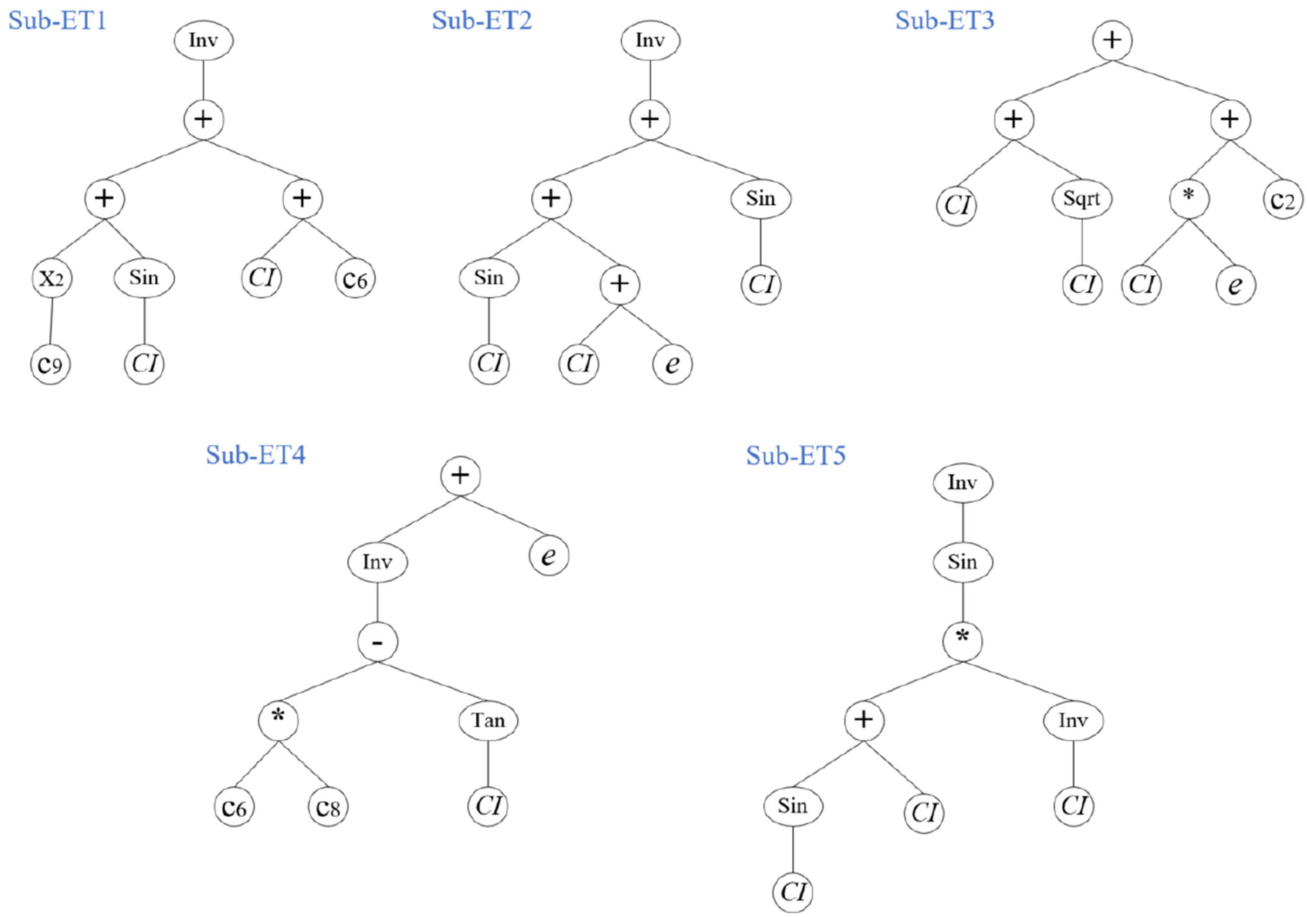
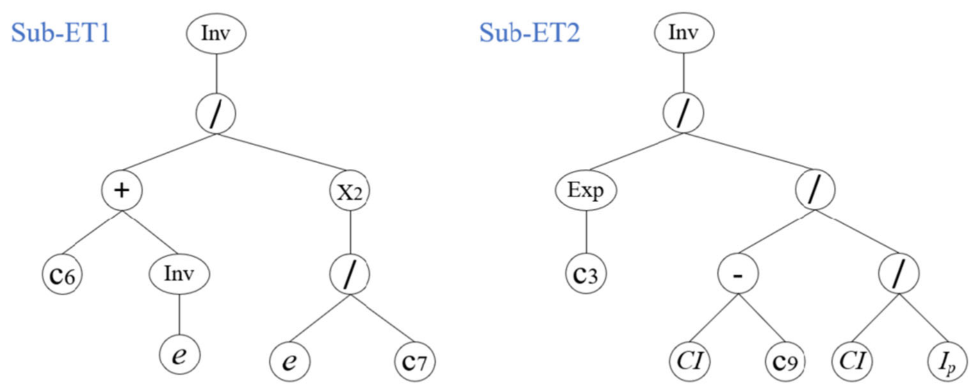


Fig. 3 Expression tree language (5 sub-expression trees) of the first model

Fig. 4 Expression tree language (2 sub-expression trees) of the second model



The constants of the third sub-expression tree c_0,c_6 and c_9 are − 87.24, − 15.48 and 33.71, respectively. The linking function of this model is addition (+), and the expression of the model can be written as:

$$C_a = \frac{1.663}{(62.967 - CI)w_L} + 0.173e^{1.33}(e + w_L) \times 10^{-3} + \frac{0.015w_L - 0.032e}{w_L} - 0.012 \tag{14}$$

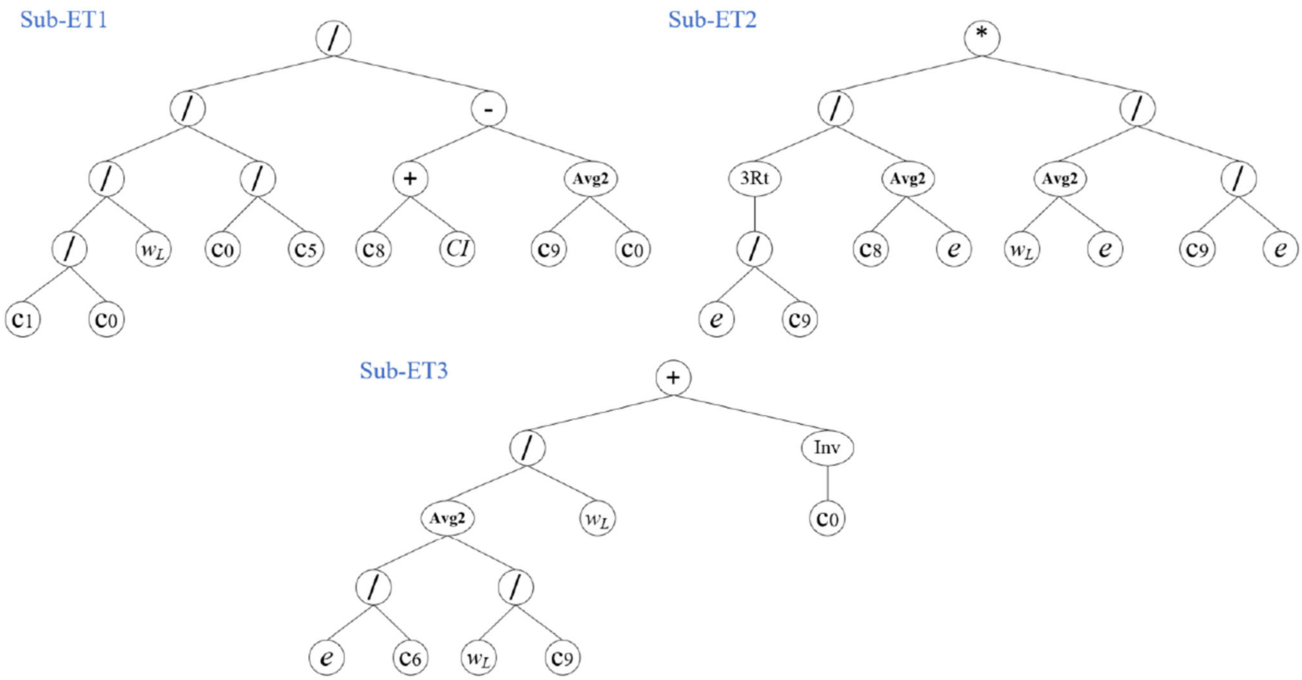


Fig. 5 Expression tree language (3 sub-expression trees) of the third model

4.4 Combination of $CI - w_L - I_p - e$

The expression tree of the fourth established GEP model consists of three sub-expression trees, as shown in Fig. 6. In Fig. 6, the constants of the first sub-expression tree c_0 and c_1 are − 3.03 and − 32.22, respectively. The constants of the second sub-expression tree c_5, c_6 and c_7 are − 63.03, 79.16 and 35.75, respectively. The constant of the third sub-expression tree c_1 is − 6.46. The linking function of this model is addition (+), and the expression of the model can be written as:

$$C_a = \frac{2ew_L + 158.32e + 35.752}{15892.85 - 4CI} + \mathrm{Tanh}\left[0.25I_p - 0.25e - 1.616 + \frac{1}{(CI - 6.462)e}\right] + \mathrm{Tanh}\left(0.5eCI - 0.25I_p - 0.25e - 0.5CI - 30.706\right) \quad (15)$$

4.5 Combination of $I_p - e$

The expression tree of the fifth established GEP model consists of three sub-expression trees, as shown in Fig. 7. In Fig. 7, the constant of the first sub-expression tree c_6 is 7.20. The constants of the second sub-expression tree c_2 and c_5 are − 20.10 and 98.72, respectively. The constant of the third sub-expression tree c_6 is 7.20. The linking function of this model is addition (+), and the expression of the model can be written as:

$$C_a = 0.193e - \frac{4.91e}{I_p} + \frac{0.019}{e^{1.5}I_p} \quad (16)$$

4.6 Combination of $w_L - e$

The expression tree of the sixth established GEP model consists of three sub-expression trees, as shown in Fig. 8. In Fig. 8, the constants of the first sub-expression tree c_2 and c_6 are − 52.99 and 72.32, respectively. The constants of the second sub-expression tree c_2 and c_6 are − 58.88 and 72.32, respectively. The constant of the third sub-expression tree c_6 is 72.32. The linking function of this model is addition (+), and the expression of the model can be written as:

$$C_a = \frac{52.991 - e}{5230.139w_L - 72.32w_L^2} - \frac{0.814e}{w_L + e} + \frac{e}{1.5e - e^2 + 36.16} \quad (17)$$

4.7 Combination of $w_L - I_p - e$

The expression tree of the seventh established GEP model consists of two sub-expression trees, as shown in Fig. 9. In Fig. 9, the constant of the first sub-expression tree c_1 is − 71.43. The linking function of this model is addition (+), and the expression of the model can be written as:

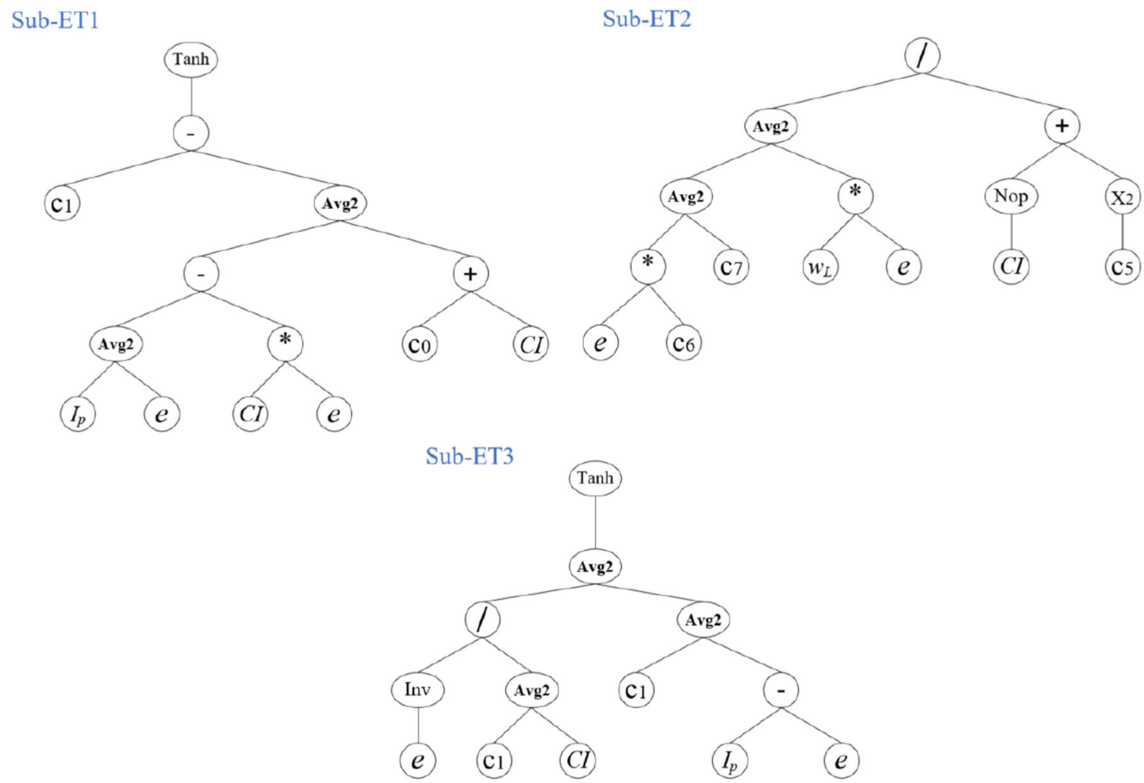


Fig. 6 Expression tree language (3 sub-expression trees) of the fourth model

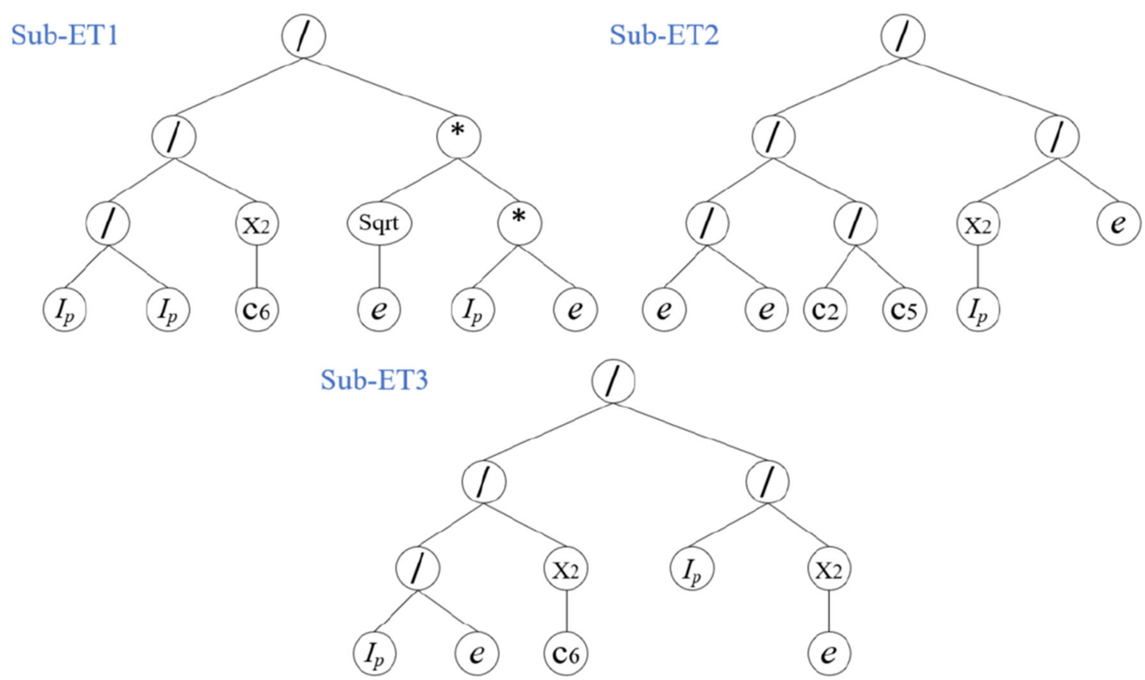


Fig. 7 Expression tree language (3 sub-expression trees) of the fifth model

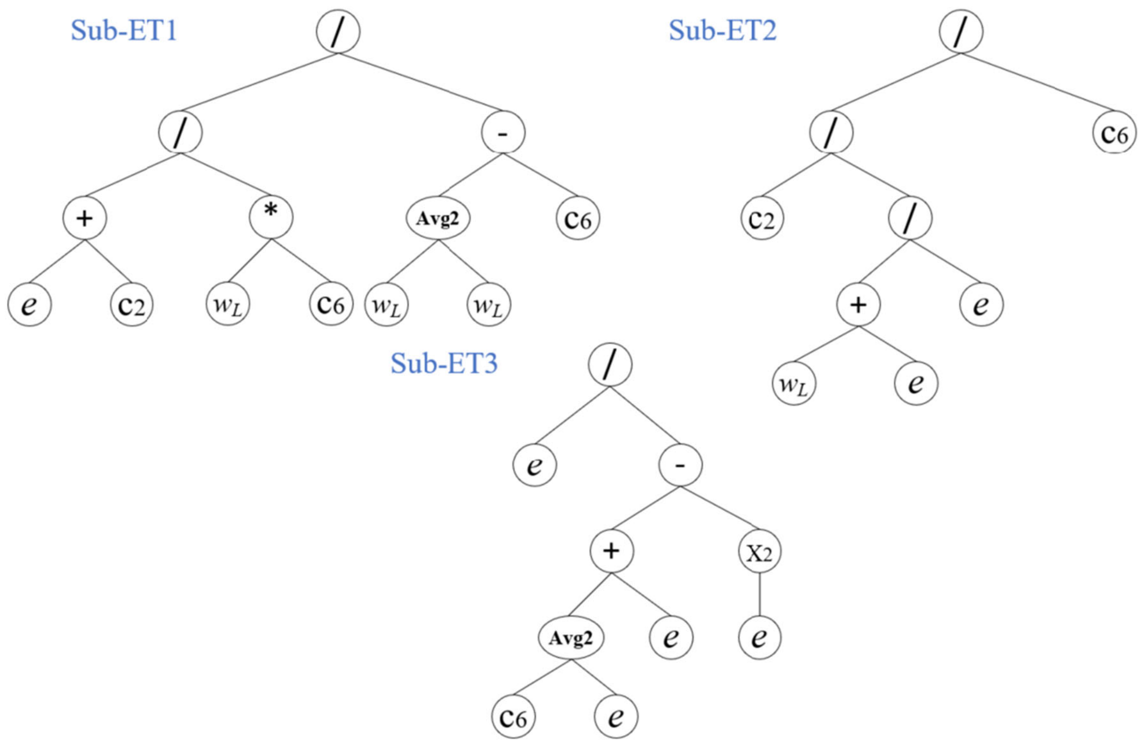


Fig. 8 Expression tree language (3 sub-expression trees) of the sixth model

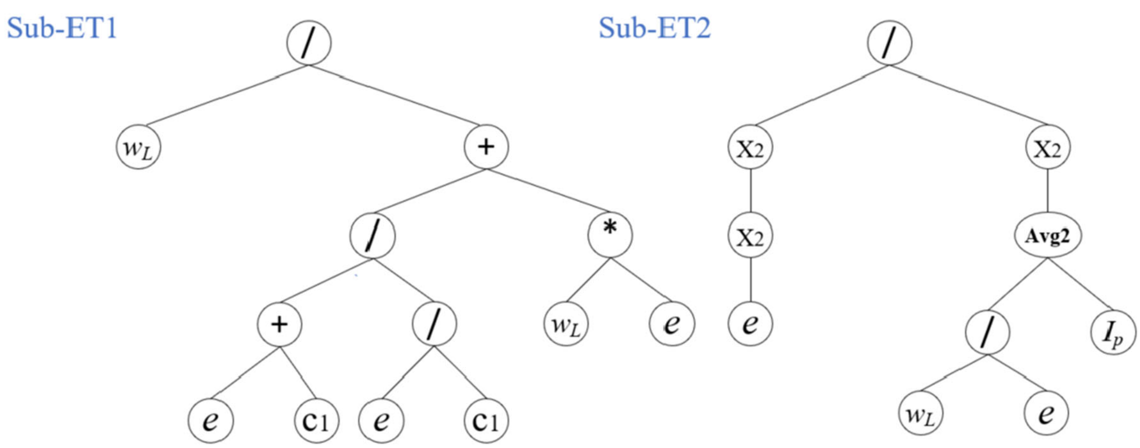


Fig. 9 Expression tree language (2 sub-expression trees) of the seventh model

$$C_\alpha = \frac{ew_L}{e^2 w_L - 71.433e - 71.433} + \frac{e^4}{4\left(I_p + \frac{w_L}{e}\right)^2} \quad (18)$$

4.8 Comparison of different GEP models

The prediction accuracy of these seven GEP models in all data sets, training sets and test sets was compared, as shown in Fig. 10. As observed from Fig. 10, regardless of all data or training or test sets, the RMSE, MAE and MAPE values of the two GEP models (i.e., with combinations $CI - w_L - e$ and $CI - w_L - I_p - e$ as input, respectively) are the lowest among these seven models. For example, the RMSE, MAE and MAPE values of the GEP models (with combinations $CI - w_L - e$ and $CI - w_L - I_p - e$ as input, respectively) for all data sets are 0.0047, 0.0032 and 0.1783; 0.0045, 0.0029 and 0.1757, respectively, and they are recommended for the prediction of the creep index C_α

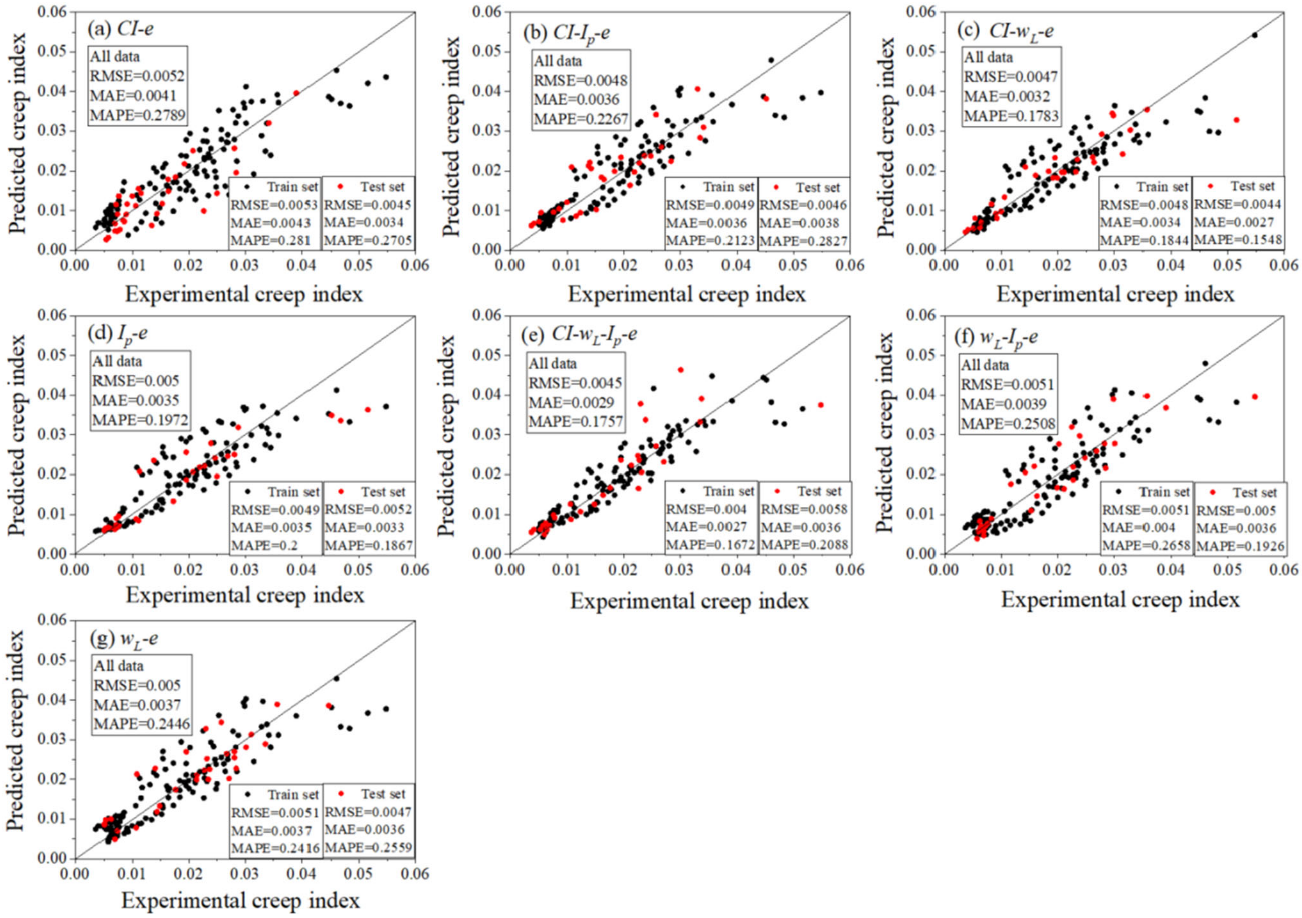


Fig. 10 Prediction results of different GEP models

in engineering practice. In addition, it can be seen from Fig. 10 that different parameter combinations have a significant impact on the prediction accuracy of the model. Therefore, it is necessary to consider the influence of parameter combinations when using GEP or other neural networks to build the model.

4.9 Comparison among the proposed models, ML-based models and existing empirical models

The prediction results of the existing five empirical models and two ML-based models (i.e., BP and RF) on all data samples are plotted in Fig. 11. The prediction performance comparisons of the two GEP models and the existing five empirical models and two ML-based models are shown in Table 4. As can be seen from Fig. 11 and Table 4, the RMSE, MAE and MAPE values of the two GEP models are 0.0045, 0.0029 and 0.1757; 0.0047, 0.0032 and 0.1783, respectively. However, the RMSE, MAE and MAPE values of the empirical models of Yin (1999), Nakase et al. (1998), Zeng et al. (2012), Zhu et al. (2020) and Zhu et al. (2016) are 0.0080, 0.0052 and 0.2497; 0.0079, 0.0053 and 0.2857; 0.0098, 0.0080 and 0.4803; 0.0097, 0.0077 and 0.5569; 0.0102, 0.0078 and 0.4801, respectively. The RMSE, MAE and MAPE values of BP and RF models are 0.0049, 0.0037 and 0.2483; 0.0056, 0.0032 and 0.2459, respectively. The above results show that the forecasting performances of the two GEP models developed in this study surpass those of the five empirical models and two ML-based models.

4.9.1 Sensitivity analysis of variables

To study whether the two developed GEP models can capture the functional relationship between the creep index C_α and various parameters, and also to compare with the results of sensitivity analysis in Zhang et al. (2020), a parametric study was carried out. As such, in the two developed GEP models, the desired independent variable

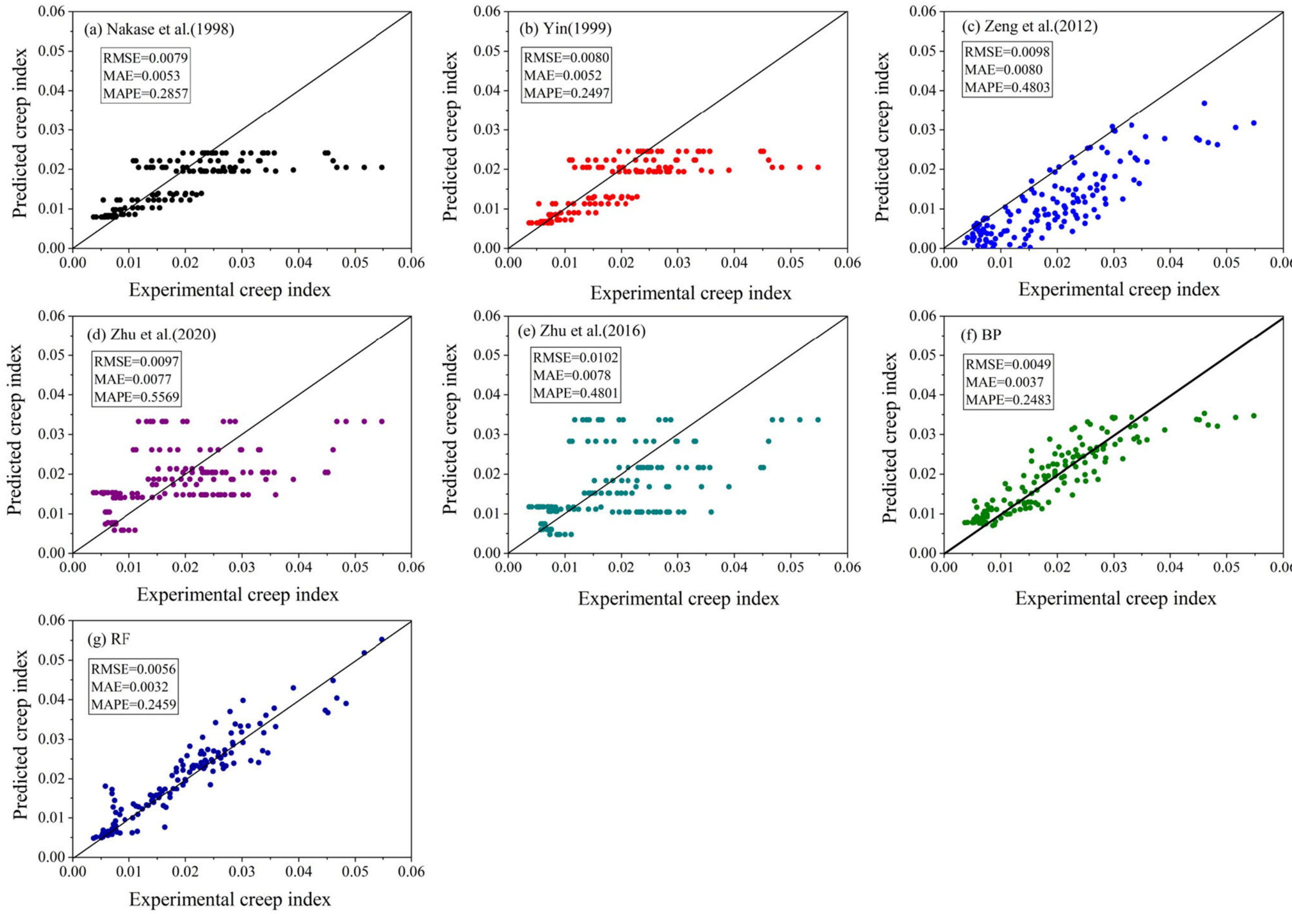


Fig. 11 Prediction results of different creep index models

Table 4 Prediction performance comparisons of different creep index models

Model	RMSE	MAE	MAPE
GEP (CI-wL-IP-e)	**0.0045**	**0.0029**	**0.1757**
GEP ($CI-wL-e$)	0.0047	0.0032	0.1783
Nakase et al. (1998)	0.0079	0.0053	0.2857
Yin (1999)	0.0080	0.0052	0.2497
Zeng et al. (2012)	0.0098	0.0080	0.4803
Zhu et al. (2020)	0.0097	0.0077	0.5569
Zhu et al. (2016)	0.0102	0.0078	0.4801
BP model	0.0049	0.0037	0.2483
RF model	0.0056	0.0032	0.2459

The bold values are used to highlight the model results

change within the scope of the database, while the values of other variables are the same as those in Zhang et al. (2020). The variations of input variables against the creep index C_α predicted by the two developed GEP models are shown in Figs. 12 and 13. As can be seen from Figs. 12 and 13, the creep index C_α increases nonlinearly and monotonically with the increases in the clay content CI, void ratio e and liquid limit w_L. In regard to plasticity index I_p, the predicted creep index C_α increases initially with an increase in plasticity index I_p, and when plasticity index I_p reaches its maximum, the creep index C_α stabilizes as the plasticity index I_p continues to increase. The evolution of the predicted value of creep index C_α with the change of the independent variable is similar at three points, which is similar to the results of the study in Zhang et al. (2020), except that the size of the creep index C_α differs. Nevertheless, from the sensitivity analysis results of Zhang et al. (2020), it can be seen that the prediction performance of the PSO-RF model largely depends on the quality and size of the database used, so it is difficult to obtain a completely smooth correlation between output and input parameters, which merely reflects a general trend (Zhang et al. 2020).

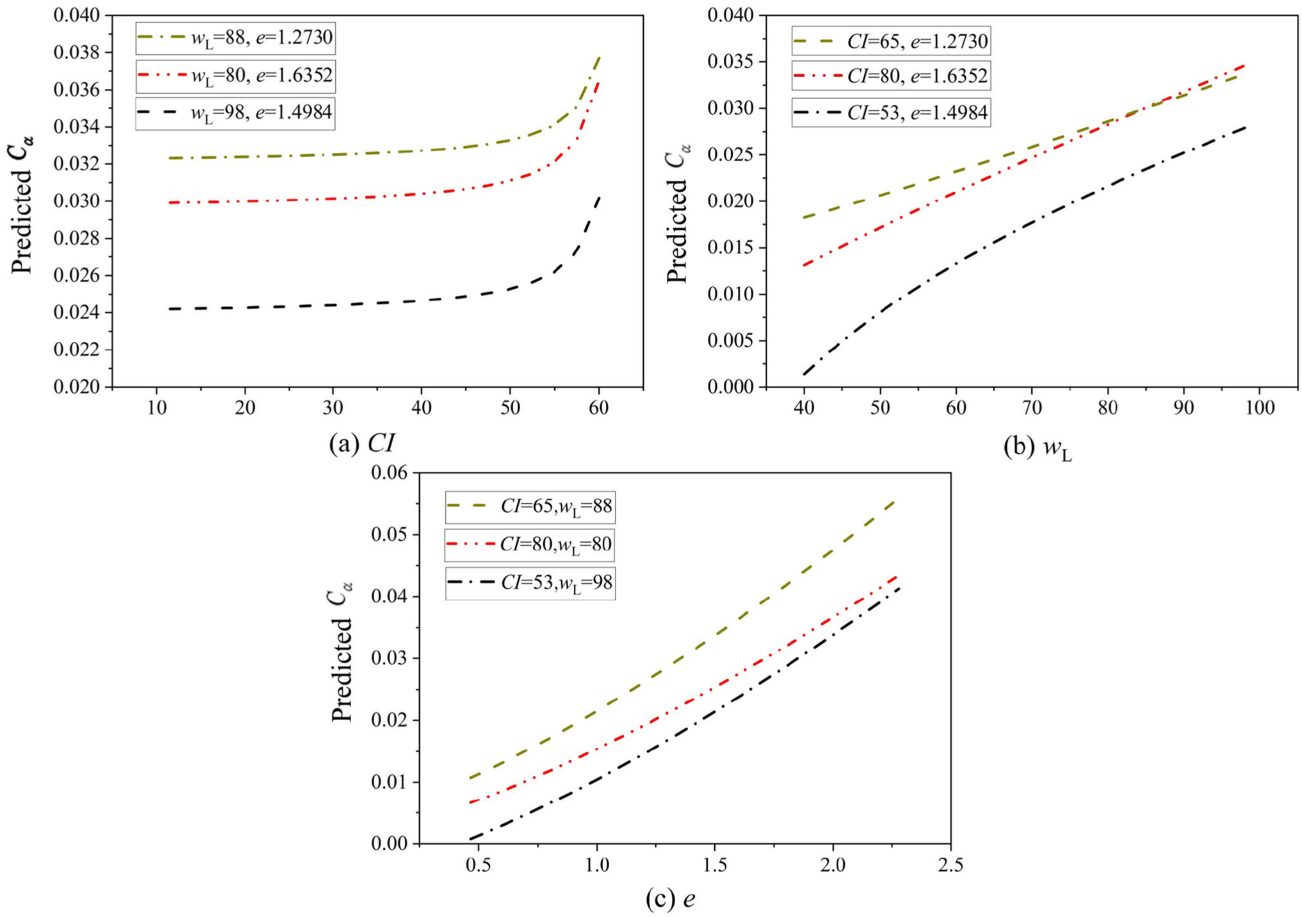


Fig. 12 Predicted C_α using GEP model combining $CI - w_L - e$ against **a** CI **b** w_L **c** e

Overall, the correlations presented in Figs. 12 and 13 are consistent with the physical explanation, which confirms the reasonableness of the developed GEP model. Therefore, the GEP models developed in this study can accurately reflect the internal mechanism between the creep index C_α and various parameters.

4.9.2 Graphical user interface

In order to promote the application of GEP model (Eq. (15)) in engineering, we developed a convenient graphical user interface (GUI) based on Visual Basic 6.0 software, as shown in Fig. 14. In Fig. 14, the input parameters are on the left and the formula calculation results are on the right.

5 Conclusions

In this study, different combinations of four input variables were used to predict creep index C_α and seven GEP models were established. The proposed GEP models are evaluated by using five empirical models and two ML-based models (i.e., BP and RF). The following conclusions can be drawn:

(1) The two developed GEP models (i.e., with combinations $CI - w_L - I_p - e$ and $CI - w_L - e$ as input, respectively) have higher prediction precision than the available five regression models in the literature and two ML-based models (i.e., BP and RF), with the RMSE, MAE and MAPE values of 0.0045, 0.0029 and 0.1757; 0.0047, 0.0032 and 0.1783, respectively.

(2) The results of parametric analysis of GEP model show that the creep index C_α increases nonlinearly and monotonically with the increases in the liquid limit w_L, clay content CI, and void ratio e. In regard to plasticity index I_p, the predicted creep index C_α increases initially with an increase in plasticity index I_p, and when plasticity index I_p reaches its maximum, the creep index C_α stabilizes as the plasticity index I_p continues to increase.

(3) The GEP models proposed in this study are only suitable for prediction of creep index of soft clays, not for rock-like materials. Therefore, the applicability of the proposed models and the richness of the

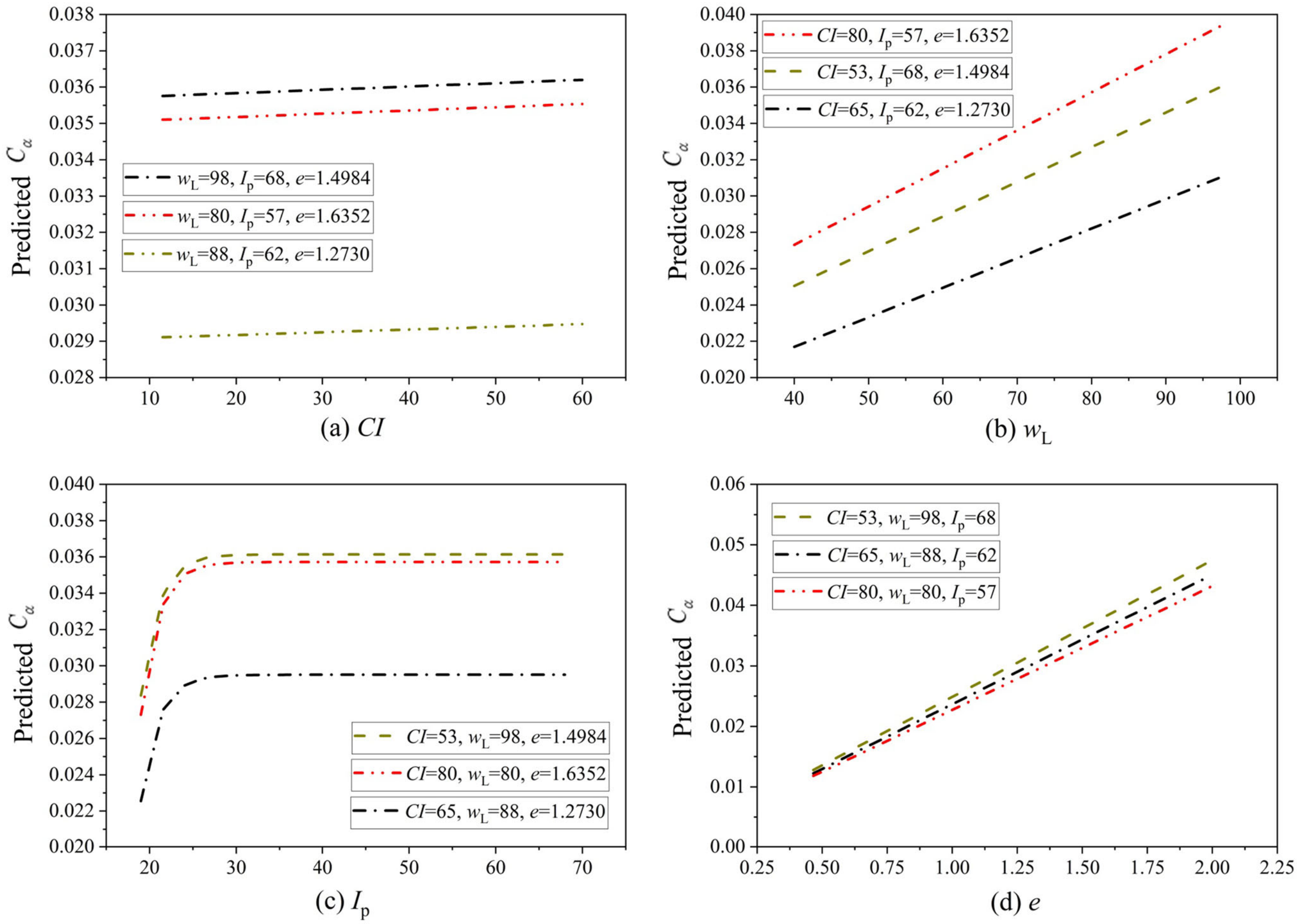


Fig. 13 Predicted C_α using GEP model combining $CI - w_L - I_p - e$ against **a** CI **b** w_L **c** I_p **d** e

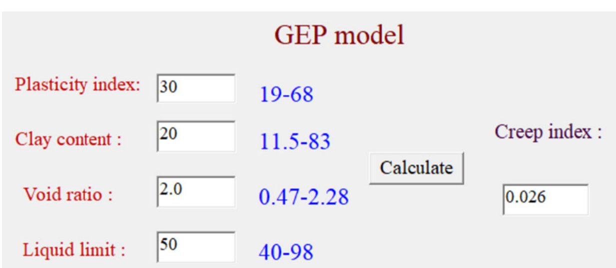


Fig. 14 Graphical user interface

database need to be further studied. In addition, GEP also has some problems, such as slow convergence rate, premature convergence and easy to fall into local extreme points, which should be studied in further research.

Author contributions XX: Methodology, Data acquisition, Writing—original draft. CD: Software, Numerical analysis.

Funding The authors have not disclosed any funding.

Data availability Data will be made available on reasonable request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Breiman L (2001) Random forests. Mach Learn 45:5–32

Fattahi H (2017) Prediction of slope stability using adaptive neuro-fuzzy inference system based on clustering methods. J Min Environ 8(2):163–178

Ferreira C (2001) Gene expression programming: a new adaptive algorithm for solving problems. Complex Syst 13(2):87–129

Ferreira C (2006) Gene expression programming: mathematical modeling by an artificial intelligence. Biswas Hope press, Canada, pp 223–225

Gordan B, Armaghani DJ, Hajihassani M, Wroblewski P (2016) Prediction of seismic slope s tability through combination of particle swarm optimization and neural network. Eng Comput Ger 32:85–97

Jafari S, Mahini SS (2017) Lightweight concrete design using gene expression programing. Constr Build Mater 139:93–100

Jong SC, Ong DEL, Oh E (2021) State-of-the-art review of geotechnical-driven artificial intelligence techniques in underground soil-structure interaction. Tunn Undergr Space Technol 113:103946

Karstunen M, Yin ZY (2010) Modelling time-dependent behavior of Murro test embankment. Geotechnique 60(10):735–749

Koopialipoor M, Armaghani DJ, Hedayat A, Maarto A (2019) Applying various hybrid intelligent systems to evaluate and predict slope stability under static and dynamic conditions. Soft Comput 23(14):5913–5929

Meng FY, Chen RP, Xin K (2018) Effects of tunneling-induced soil disturbance on the post-construction settlement in structured soft soils. Tunn Undergr Space Technol 80:53–63

Murad Y, Ashteyat A, Hunaifat R (2019) Predictive model to the bond strength of FRP-to concrete under direct pullout using gene expression programming. J Civ Eng Manag 25(8):773–784

Nakase A, Kamei T, Kusakabe O (1988) Constitutive parameters estimated by plasticity index. Chin J Geotech Eng 114(7):844–858

Nakase A, Kamei T, Kusakabe O (1998) Constitutive parameters estimated by plasticity index. J Geotech Eng 114:844–858

Sharma S, Ahmed S, Naseem M, Alnumay W, Singh S, Cho GH (2021) A survey on applications of artificial intelligence for pre-parametric project cost and soil shear-strength estimation in construction and geotechnical engineering. Sensors 21(2):463–506

Shen SL, Wu HN, Cui YJ, Yin ZY (2014) Long-term settlement behaviour of metro tunnels in the soft deposits of Shanghai. Tunn Undergr Space Technol 40:309–323

Sun YB, Wendi D, Kim DE (2016) Application of artificial neural networks in groundwater table forecasting-a case study in a Singapore swamp forest. Hydrol Earth Syst Sci 20:1405–1412

Tan F, Zhou WH, Yuen KV (2018) Effect of loading duration on uncertainty in creep analysis of clay. Int J Numer Anal Methods Geomech 42:1235–1254

Xiong LH, Kieran MO, Guo SL (2004) Comparasion of three updating schemes using artificial neural network in flow forecasting. Hydrol Earth Syst Sci 8(2):247–255

Xue XH (2017) Prediction of daily diffuse solar radiation using artificial neural networks. Int J Hydrog Energy 42:28214–28221

Yang B, Yin K, Lacasse S, Liu Z (2019) Time series analysis and long short-term memory neural network to predict landslide displacement. Landslides 16:677–694

Yin JH (1999) Properties and behavior of Hong Kong marine deposits with different clay contents. Can Geotech J 36:1085–1095

Yin ZY, Chang CS (2009) Microstructural modeling of stress-dependent behavior of clay. Int J Solids Struct 46:1373–1388

Yin ZY, Chang CS, Karstunen M, Hicher PY (2010) An anisotropic elastic-viscoplastic model for soft clays. Int J Solids Struct 47:665–677

Yin ZY, Xu Q, Yu C (2014a) Elastic-viscoplastic modeling for natural soft clays considering nonlinear creep. Int J Geomech 15(5):1–10

Yin ZY, Yin JH, Huang HW (2015) Rate-dependent and long-term yield stress and strength of soft Wenzhou marine clay: experiments and modeling. Mar Georesour Geotechnol 33:79–91

Yin ZY, Zhu QY, Yin JH, Ni Q (2014b) Stress relaxation coefficient and formulation for soft soils. Geotech Lett 4:45–51

Yin ZY, Zhu QY, Zhang DM (2017) Comparison of two creep degradation modeling approaches for soft structured soils. Acta Geotech 12:1395–1413

Zeng LL, Hong ZS, Liu SY, Chen FQ (2012) Variation law and quantitative evaluation of secondary consolidation behavior for remolded clays. Chin J Geotech Eng 34:1496–1500

Zhang CS, Ji J, Gui YL, Kodikara J (2016) Evaluation of soil-concrete interface shear strength based on LS-SVM. Geomech Eng 11(3):361–372

Zhang P, Yin ZY, Jin YF, Chan THT (2020) A novel hybrid surrogate intelligent model for creep index prediction based on particle swarm optimization and random forest. Eng Geol 265:1–12

Zhang WG, Li HR, Li YQ, Liu HL, Chen YM, Ding XM (2021) Application of deep learning algorithms in geotechnical engineering: a short critical review. Artif Intell Rev 54:5633–5673

Zhu QY, Yin ZY, Hicher PY, Shen SL (2016) Nonlinearity of one-dimensional creep characteristics of soft clays. Acta Geotech 11:887–900

Zhu QY, Jin YF, Yin ZY (2020) Modeling of embankment beneath marine deposited soft sensitive clays considering straightforward creep degradation. Mar Georesour Geotechnol 38(5):553–569