Environ Earth Sci (2016)75:262
DOI 10.1007/s12665-015-4889-2

ORIGINAL ARTICLE

Deriving an intelligent model for soil compression index utilizing multi-gene genetic programming

Danial Mohammadzadeh S[1] · Jafar Bolouri Bazaz[1] · S. H. Vafaee Jani Yazd[2] · Amir H. Alavi[3]

Received: 9 February 2015 / Accepted: 6 August 2015

Abstract Multi-gene genetic programming (MGGP) is a new nonlinear system modeling approach that integrates the capabilities of standard GP and classical regression. This paper deals with the prediction of compression index of fine-grained soils using this robust technique. The proposed model relates the soil compression index to its liquid limit, plastic limit and void ratio. Several laboratory test results for fine fine-grained were used to develop the models. Various criteria were considered to check the validity of the model. The parametric and sensitivity analyses were performed and discussed. The MGGP method was found to be very effective for predicting the soil compression index. The prediction coefficients of determination were 0.856 and 0.840 for the training and testing data, respectively. A comparative study was further performed to prove the superiority of the MGGP model to the existing soft computing and traditional empirical equations.

Keywords Multi-gene genetic programming · Soil compression index · Soil engineering properties · Prediction

✉ Amir H. Alavi
alavi@msu.edu; ah_alavi@hotmail.com
http://www.egr.msu.edu/∼alavi/

[1] Civil Engineering Department, Ferdowsi University of Mashhad, Mashhad, Iran

[2] Department of Civil Engineering, The Institute of Higher Education of Eqbal Lahoori, Mashhad, Iran

[3] Department of Civil and Environmental Engineering, Michigan State University, East Lansing, MI 48824, USA

Introduction

Soil compressibility is the volume reduction under load due to the drainage of pore water. Precise determination of this property is important for the analysis of the soil layer settlements (Tiwari and Ajmera 2012; Singh and Noor 2012). This issue becomes more challenging for fine-grained soils due to their low permeability compared to coarse-grained soils. Compression index (C_c) is one of the main soil compressibility indicators. This parameter is usually used for the measurement of the soil layer settlement (Carter and Bentley 1991; Gulhati and Datta 2005; Singh and Noor 2012; Mohammadzadeh et al. 2014). The settlement is typically obtained using the logarithm of the normal compressive stress (σ') against soil void ratio (e) curve (Gregory et al. 2006). C_c is considered as the slope of the virgin region of the compression curve. This curve can be obtained from a standard consolidation (oedometer) test on an undisturbed sample (Gregory et al. 2006; Mohammadzadeh et al. 2014). To avoid cumbersome consolidation tests, various empirical equations have been developed to predict C_c (Skempton 1944; Nishida 1956; Cozzolino 1961; Terzaghi and Peck 1967; Sowers 1970; Azzouz et al. 1976; Wroth and Wood 1978; Mayne 1980; Park and Lee 2011). The existing models are mainly developed based on traditional regression analyses that have notable modeling drawbacks such as large uncertainties or oversimplification of the complicated mechanism of the consolidation process. Also, the regression method assumes the structure of the model in advance by merely a limited number of linear or nonlinear equations (Alavi and Gandomi 2011; Mohammadzadeh et al. 2014). Consequently, more sophisticated methods are needed for the analysis of the complex behavior of C_c.

Published online: 25 January 2016

In this context, soft computing techniques may be considered as efficient alternatives to traditional methods. These techniques automatically specify the structure of a prediction model by learning from data (Salehi and Taghikhany 2012; Salehi et al. 2014; Alavi et al. 2016a, b). Different well-known branches of soft computing have been successfully applied to engineering problems (e.g., Das and Basudhar 2008; Samui 2008a, b; Kayadelen et al. 2009; Günaydın 2009; Kolay et al. 2010; Das et al. 2010, 2011a, b, c; Yilmaz 2010a, b; Akgun and Türk 2010; Kaunda et al. 2010; Azamathulla and Wu 2011; Azamathulla et al. 2012; Mert et al. 2011; Yilmaz et al. 2012; Sattari et al. 2012; Tasdemir et al. 2013; Ocak and Seker 2012, 2013; Isik and Ozden 2013; Wu et al. 2013; Salehi et al. 2015; Park et al. 2013; Ceryan et al. 2013). As a robust soft computing tool, artificial neural networks (ANNs) are widely utilized to predict the C_c of soil layers (Desai et al. 2009; Jianping et al. 2010; Farkhonde and Bolouri 2010; Daryaee et al. 2010; Kumar and Rani 2011; Park and Lee 2011; Kumar et al. 2012; Rani et al. 2013). A limitation of the ANNs and many of the other soft computing methods is that they do not usually generate prediction equations (Alavi and Gandomi 2011). To deal with this issue, genetic programming (GP) was established by Koza (1992). GP is a soft computing method that works based on the Darwinian evolution theory for creating computer programs. The computer programs evolved by GP can be converted to highly nonlinear prediction equations (Javadi et al. 2006; Guven 2009; Guven et al. 2009; Gandomi et al. 2010; Gandomi and Alavi 2011; Rezania et al. 2011; Azamathulla et al. 2011; Alavi et al. 2011; Mousavi et al. 2012; Azamathulla 2012; Mahmood et al. 2013).

Multi-gene GP (MGGP; Searson et al. 2007, 2010) is a new subset of GP. While classical GP output is merely a single program as the optimal solution, the MGGP optimal solution includes several genes that are each a computer program (Searson et al. 2010; Gandomi and Alavi 2012a, b). As opposed to classical GP and other soft computing techniques, application of MGGP in the engineering domain is new (Gopalakrishnan et al. 2010; Desai and Shaikh 2012; Gandomi and Alavi 2012a, b; Muduli et al. 2013a, b; Muduli and Das 2013a, b; Mohammadi et al. 2014). This paper proposes the MGGP technique to develop a predictive equation for the compression index of soils. A comprehensive set consolidation test results was established for the development of the models. The proposed model was verified through different validation phases.

Multi-gene genetic programming

GP generates computer models by simulating the biological evolution of living organisms (Koza 1992). The genetic operators of genetic algorithm (GA) and GP are almost the same. GA gives the optimal solution as a string of numbers while the solutions generated by GP are computer programs provided in a tree-like form (Koza 1992; Gandomi and Alavi 2012a, b). In GP, the main goal is to obtain a program connecting inputs to the corresponding output. This process is started by creating a population of programs. The tree-like programs created by GP include functions and terminals chosen randomly. Figure 1 illustrates a typical tree-like program evolved by GP. Further details about GP are given in Koza (1992), Alavi and Gandomi (2011) and Gandomi and Alavi (2012a, b).

MGGP (Searson et al. 2007, 2010; Searson 2009) is a new variant of GP. MGGP evolves a number of genes (sub-programs) that are each individual tree-like programs (Searson et al. 2007, 2010; Searson 2009). Then, it assigns weights to each of these genes to generate the final model as a linear combination of the sub-programs. Figure 2 shows a typical program evolved by MGGP. The inputs of the model are x_1, x_2 and x_3 and the functions used for the evolution process are $\times$, $-$, $+$, log, and $\sqrt{}$. The model is linear in the parameters regarding the coefficients c_0, c_1 and c_2 despite using nonlinear terms (Searson et al. 2007, 2010; Gandomi and Alavi 2012a, b). To obtain the linear coefficients, an ordinary least squares analysis is performed on the training data. It is possible to use a partial least squares method to this aim (Searson et al. 2007, 2010; Gandomi and Alavi 2012a, b). The important MGGP parameters that need notable control are the number of genes and tree depth. Restricting the tree depth mostly results in generating more compact models (Searson et al. 2007, 2010; Gandomi and Alavi 2012a, b).

The initial population generated by MGGP contains GP trees with different randomly generated genes. MGGP employs a tree crossover operator as well as the classical recombination operators (Searson et al. 2007, 2010; Gandomi and Alavi 2012a, b). As an example, assume that two parent programs evolved by MGGP contain two

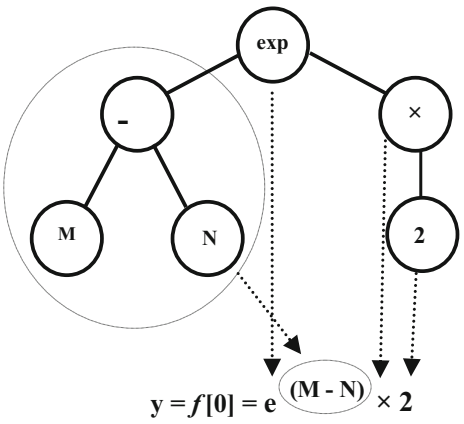
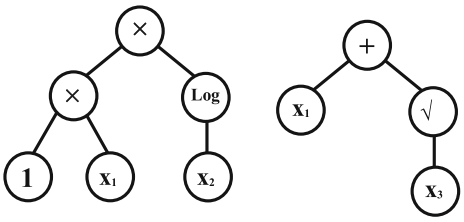


Fig. 1 Tree representation of a GP model



$$y = c_0 + c_1(1 \times x_1 \times \mathrm{Log}(x_2)) + c_3(x_1 + \sqrt{x_3})$$

Fig. 2 A typical program evolved by MGGP

genes (G1 G2) and three genes (G3 G4 G5). The genes subjected to crossover are shown by [] as given herein: (G1 [G2]) and (G3 [G4 G5]). Thus, during the crossover operation, the genes are exchanged to create two new programs: (G1 G4 G5) and (G3 G2). Besides, there are different types of mutation in MGGP such as sub-tree mutation and mutation of constants using an additive Gaussian perturbation. Further details about MGGP and its operators can be found in Searson et al. (2007, 2010) and Gandomi and Alavi (2012a, b).

Modeling of C_c of fine-grained soils

Plasticity is one of the major physical properties of fine-grained soils. It is measured by Atterberg limits, i.e., liquid limit (LL) and plastic limit (PL). These parameters represent a threshold of water content at which soil is expected to have certain behavior types (e.g., viscous flow, plasticity). LL denotes the lower limit of viscous flow, or the water content, at which the soil behavior changes from plastic to liquid. On the other hand, PL represents the lower limit of the plastic state (Kim et al. 2013). The structure of the existing models (Skempton 1944; Nishida 1956; Cozzolino 1961; Terzaghi and Peck 1967; Sowers 1970; Azzouz et al. 1976; Mayne 1980; Park and Lee 2011; Mohammadzadeh et al. 2014) clearly indicates that C_c of fine-grained soils can be considered as a function of percentage of LL and PL, as well as initial void ratio (e_o). The same strategy was considered in this study to develop a simplified prediction model for C_c using merely its basic physical properties. Determination of LL, PL and e_o is much more easier than performing the consolidation tests to find C_c. Hence, the adapted strategy would result in a notable cost savings for geotechnical engineering design. As it is, all these parameters have a rational relationship with the natural water content for saturated soils (Bartlett and Lee 2004; Kim et al. 2013; Mohammadzadeh et al. 2014).

To develop the MGGP-based model, a comprehensive database containing soil properties (LL, PL and e_o) and corresponding C_c was developed. The database was divided into three sub-divisions of learning, validation and testing subsets as will be described in "MGGP-based formulation of C_c" section. The MGGP algorithm was run on the learning data and checked on the other two subsets. The optimal model was selected considering its simplicity as well as its performance on the learning and validation data.

Determination coefficient (R^2), root mean squared error (RMSE) and mean percent error (MAE) were used to assess the performance of the models (Alavi and Gandomi 2011):

$$R^2 = \frac{\left(\sum_{i=1}^{n}(h_i - \bar{h}_i)(t_i - \bar{t}_i)\right)^2}{\sum_{i=1}^{n}(h_i - \bar{h}_i)^2 \sum_{i=1}^{n}(t_i - \bar{t}_i)^2}, \quad (1)$$

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

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|h_i - t_i|, \quad (3)$$

where h_i is the measured C_c, t_i is the predicted C_c, $\bar{h}_i$ is the average of the measured C_c, and $\bar{t}_i$ is the average of the predicted C_c.

Laboratory test

Several laboratory test results were performed on samples obtained from different locations in Khorasan Razavi Province, Iran to develop the database. The soil samples were picked up from test pits at 0.5–1.0 m depth. The physical and plastic characteristics of soil were determined through geotechnical laboratory tests. Conventional oedometer tests were carried out to determine C_c. The testing procedure was as described in Abbasi et al. (2012) and Mohammadzadeh et al. (2014). A conventional oedometer apparatus having brass ring, 75 mm in diameter and 20 mm in height was used to perform the tests. The vertical dead load was applied using a loading device until there was no change in dial gauge reading for two consecutive hours. Other details of the test were performed in general accordance with ASTM D2435 procedure (2000). The variation of void ratio versus pressure was plotted for each specimen on a semilogarithmic scale to obtain the C_c values (Abbasi et al. 2012). One-dimensional consolidation and basic geotechnical characterization tests are very well-known tests. Thus, for brevity, some basic information was provided herein. Table 1 presents the descriptive statistics of the parameters (LL, PL, e_0, C_s) used in this study. There were a total of 101 laboratory tests on fine-grained soil samples. To enrich the database, seven consolidation test results done by Malih (2006) were added to the available data. Different soil types used in this study were silty clay with sand (CL–ML), gravelly lean clay with sand (CL), and silty, clayey sand (SC–SM).

Table 1 Descriptive statistics of the variables used for the model development

Parameters	e_0	LL (%)	PL (%)	C_c
Mean	0.745	36.156	22.605	0.171
Standard deviation	0.123	12.786	5.639	0.047
Range	0.516	52.600	29.200	0.176
Minimum	0.514	19.400	14.800	0.077
Maximum	1.030	72.000	44.000	0.253

Table 2 Parameter considered for the implementation of the MGGP algorithm

Parameters	Settings
Population size	200, 400, 800, 1200, 1500, 2000, 3000
Generations	200, 400, 800, 1000
Maximum number of genes	4, 6, 10, 12
Maximum tree depth	4, 6, 10, 12
Tournament size	25
Crossover events	0.85
Mutation events	0.1
Function set	+, −, ×, /, ^2, ^3, tanh, sin, cos, exp

To avoid overfitting, the data were classified into three subsets: (1) learning, (2) validation, and (3) test subsets. The models were trained using the learning data. Then, their performance was checked on the validation data. The model with the highest R^2 and lowest RMSE and MAE values on the learning and validation data was chosen as the optimal model. Eventually, the optimal model was tested on un-seen testing data to guarantee that it has generalization capacity. Since both the learning and validation sets were included in the model selection, they were grouped together as "training data" (Alavi et al. 2011; Mohammadzadeh et al. 2014). The training, validation and testing data are commonly taken as 50–70, 15–25 and 15–25 of all data, respectively (Shahin and Jaksa 2005; Alavi et al. 2011). Herein, about 65 % (71 sets), 15 % (16 sets) and 20 % (21 sets) of the data were considered for the training, validation and testing processes, respectively.

MGGP-based formulation of C_c

To find the optimal model, the MGGP algorithm was run several times with different combinations of the parameters. This process was based on both an extensive trail study and referring to the suggested values in previous studies (Searson et al. 2010; Gopalakrishnan et al. 2010; Desai and Shaikh 2012; Gandomi and Alavi 2012a, b). Table 2 presents the parameters considered for different runs. Large values were taken for the population and generations to ensure that the best models were derived. As it is seen in Table 2, there were $7 \times 4 \times 4 \times 4 = 448$ parameter arrangement. All these arrangements were tested with two replications for each. Thus, about $448 \times 2 = 896$ optimal individual runs were performed. GPTIPS toolbox (Searson 2009) coded in MATLAB was utilized to run the algorithm. The RMSE function was adapted for error minimization during runs.

Table 3 Individual genes/model terms for the prediction of C_c

Terms	Values
Bias	−0.176
Gene 1	$-0.0227x_2 - 0.0113x_3$
Gene 2	$-(0.162\ (x_3 + x_2\tanh(x_3)))/x_2$
Gene 3	$0.016x_2 - 0.194x_3 + (0.016x_3)/x_1$
Gene 4	$0.00468x_1x_2$
Gene 5	$0.87x_1$
Gene 6	$-(0.0191x_2)/(x_1 + x_3 + \tanh(\tanh(\exp(x_2 - \exp(x_1)))))$
Gene 7	$0.187x_3$
Gene 8	$(0.014x_3)/x_2$
Gene 9	$7.33 \times 10\text{^}(-4)x_2 + 7.33 \times 10\text{^}(-4)\ \sin(x_2) + 7.33 \times 10\text{^}(-4)\ \tanh(x_2)^3 - (4.23 \times 10\text{^}14x_2 + 4.23 \times 10\text{^}14x_3 + 4.23 \times 10\text{^}14\tanh(x_1))/(1.15 \times 10\text{^}18x_1 + 1.15 \times 10\text{^}18x_2 + 5.76 \times 10\text{^}17\exp(\tanh(x_1x_3)))$

$x_1 = e_0$; $x_2 = \text{LL}$; $x_3 = \text{PL}$

Table 3 shows the individual genes/model terms for the best models that were obtained during the conducted runs. Each gene includes its weighting coefficient. As can be observed from this table, the derived model is composed of complicated array of operators, variables, and constants to estimate C_c. Besides, Fig. 3 shows the expression trees (ETs) of the MGGP model. In these figures, x_1, x_2, and x_3 represent e_0, LL and PL, respectively.

To facilitate the use of the developed model, it was transformed into a functional form and thereafter simplified as follows:

$$\begin{aligned} C_c = {} & 0.87e_0 - 0.00591\text{LL} - 0.0186\text{PL} + 7.33 \times 10^{-4}\sin(\text{LL}) \\ & - 0.162\tanh(\text{PL}) + 7.33 \times 10^{-4}\tanh(\text{LL})^3 + 0.00468e_0\text{LL} \\ & - (1.38 \times 10^{15}\text{LL})/(7.21 \times 10^{16}e_0 + 7.21 \times 10^{16}\text{PL} \\ & + 7.21 \times 10^{16}\tanh(\tanh(\exp(\text{LL} - \exp(e_0))))) \\ & - (4.22 \times 10^{14}\text{LL} + 4.22 \times 10^{14}\text{PL} \\ & + 4.22 \times 10^{14}\tanh(e_0))/(1.15 \times 10^{18}e_0 + 1.15 \times 10^{18}\text{LL} \\ & + 5.76 \times 10^{17}\exp(\tanh(e_0\text{PL}))) + (0.016\text{PL})/e_0 \\ & - (0.148\text{PL})/\text{LL} - 0.176. \end{aligned} \quad (4)$$

Figure 4 shows the experimental against predicted C_c values. The numbers of population, generations, genes and tree depth for the model were equal to 1500, 800, 10 and 10, respectively. The performance of the model is good on

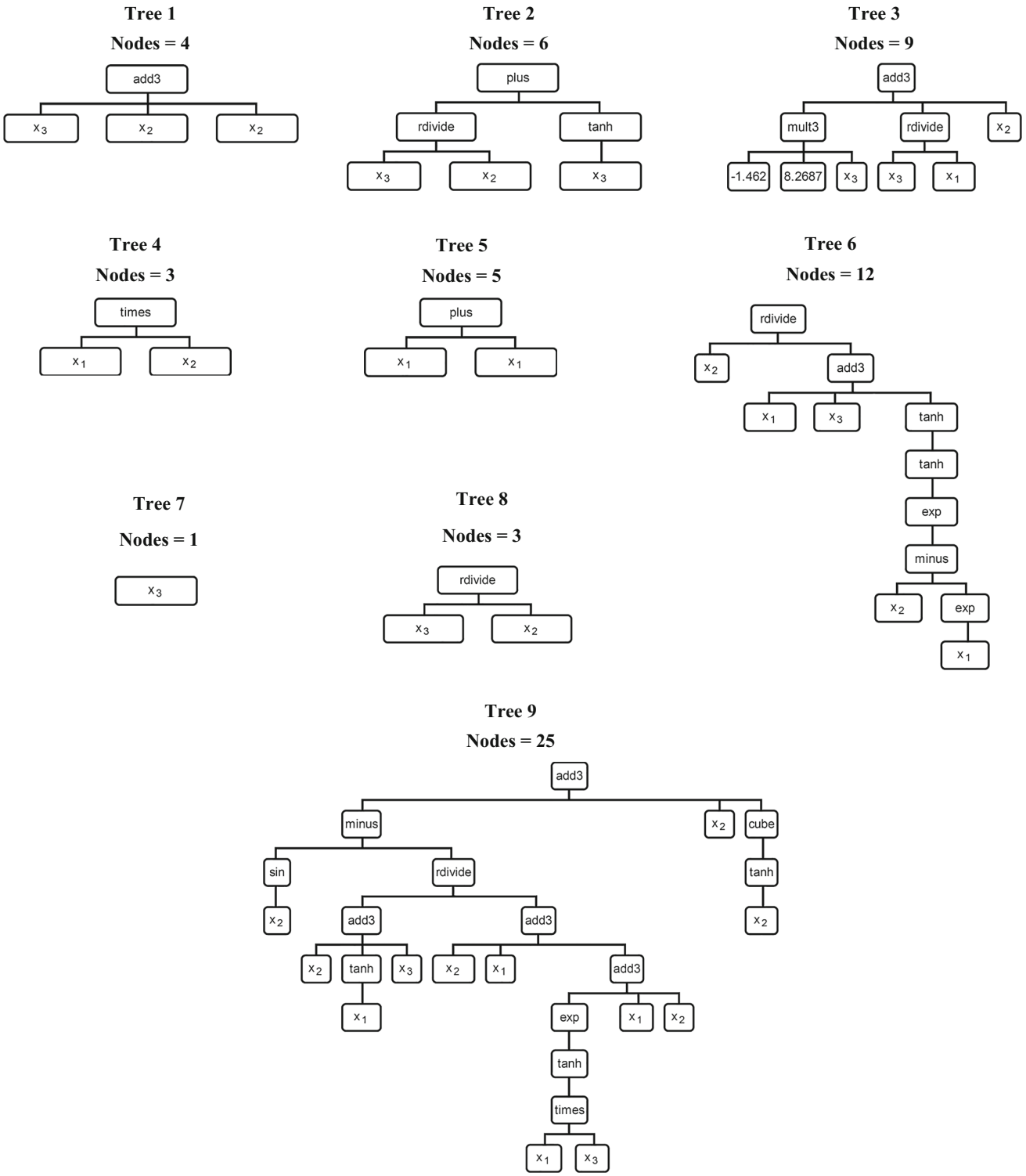


Fig. 3 Expression trees of the best models for the prediction of C_c (ET $= \sum$sub-ET$_i$)

both the training and testing data. The importance of each of the nine genes of the derived model is shown in Fig. 5. It is seen that the weight of the genes (sub-programs) 5, 7, 2 and the bias terms are higher than the other genes. This means that they have higher contribution to the prediction of C_c. Besides, Fig. 6 presents the accuracy against the complexity of the evolved models. As it is seen, the models with good performance seem to be much less complex. The

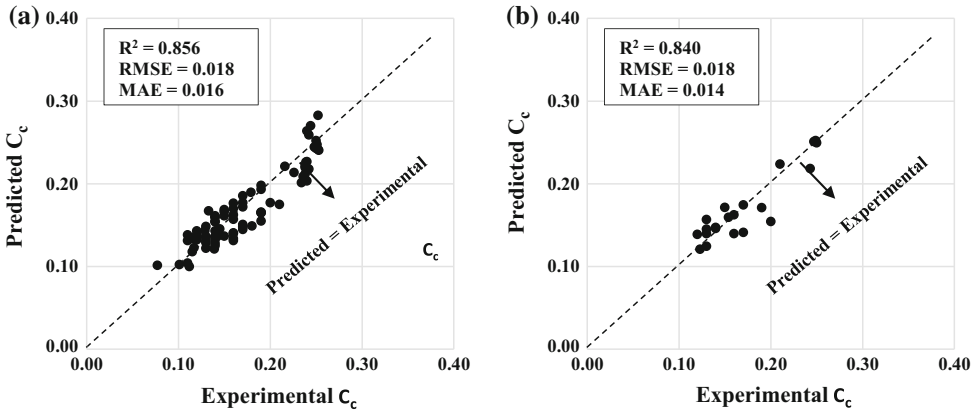


Fig. 4 Performance of the MGGP model on: **a** training data, and **b** testing data

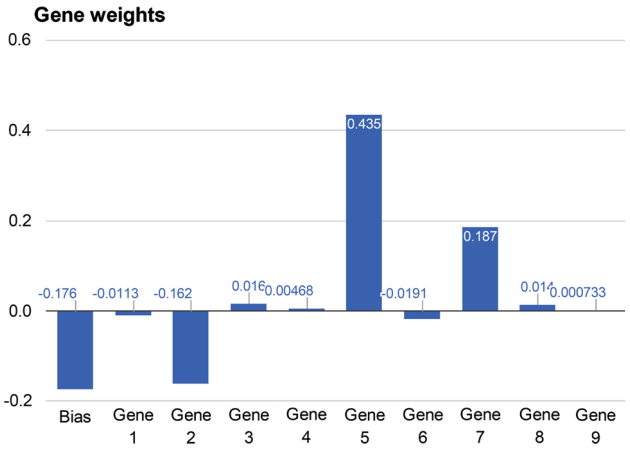


Fig. 5 Weights of the genes (sub-programs) of the MGGP model

red circle in Fig. 6 designates the best model presented herein.

Validation verification

Smith (1986) argues that for correlation coefficient ($|R|$) > 0.8, a powerful correlation exists between the predicted and actual values and, therefore, the model is suitable. In addition, the error values should be minimum in all cases (Alavi et al. 2011). Figure 4 indicates that the MGGP model has high R and low RMSE and MAE values. Besides, Golbraikh and Tropsha (2002) defined new statistical parameters of k and k' to validate the models on the testing data. Also, Roy and Roy (2008)

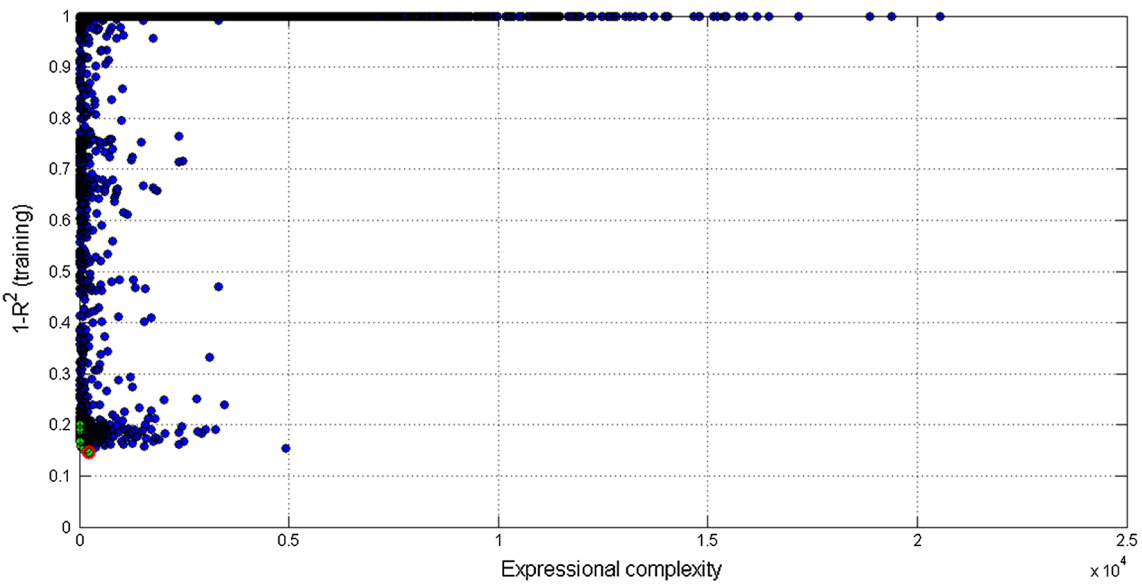


Fig. 6 Complexity against accuracy of the models evolved by MGGP

Table 4 Verifying the validity of the MGGP model on the testing data

Items	Formulas	Conditions	MGGP
1	$\frac{\sum_{i=1}^{n}(h_i-\bar{h}_i)(t_i-\bar{t}_i)}{\sqrt{\sum_{i=1}^{n}(h_i-\bar{h}_i)^2\sum_{i=1}^{n}(t_i-\bar{t}_i)^2}}$	$0.8 < R$	0.916
2	$k=\frac{\sum_{i=1}^{n}(h_i\times t_i)}{h_i^2}$	$0.85 < k < 1.15$	1.001
3	$k'=\frac{\sum_{i=1}^{n}(h_i\times t_i)}{t_i^2}$	$0.85 < k' < 1.15$	0.989
4	$R_{\mathrm{m}}=R^2\times(1-\sqrt{\lvert R^2-R_{\mathrm{O}}^2\rvert})$	$0.5 < R_{\mathrm{m}}$	0.503
Where	$R_{\mathrm{O}}^2=1-\frac{\sum_{i=1}^{n}(t_i-h_i^{\mathrm{o}})^2}{\sum_{i=1}^{n}(t_i-\bar{t}_i)^2},\ h_i^{\mathrm{o}}=k\times t_i$	Should be close to 1	1.000
	$R_{\mathrm{O}}'^2=1-\frac{\sum_{i=1}^{n}(h_i-t_i^{\mathrm{o}})^2}{\sum_{i=1}^{n}(h_i-\bar{h}_i)^2},\ t_i^{\mathrm{o}}=k'\times h_i$	Should be close to 1	0.998

h_i Actual output, t_i predicted output

Table 5 Comparison of the predictions provided by different models

References	R^2	RMSE	MAE
Skempton (1944)	0.367	0.072	0.056
Nishida (1956)	0.752	0.301	0.285
Cozzolino (1961)	0.752	0.105	0.103
Terzaghi and Peck (1967)	0.367	0.110	0.077
Sowers (1970)	0.752	0.055	0.042
Azzouz et al. (1976)	0.752	0.036	0.032
Mayne (1980)	0.367	0.102	0.073
Park and Lee (2011)	0.752	0.089	0.085
MEP (Mohammadzadeh et al. 2014)	0.811	0.019	0.016
ANN (Mohammadzadeh et al. 2014)	0.859	0.017	0.014
MGGP (this study)	0.840	0.018	0.014

proposed a predictability indicator parameter called R_{m}, as well as two other related parameters (R_{O}^2 and $R_{\mathrm{O}}'^2$). k, k', R_{m}, R_{O}^2, and $R_{\mathrm{O}}'^2$ are defined based on actual (h_i) and predicted (t_i) outputs. It is suggested that at least one slope of regression lines (k or k') through the origin should be close to 1. k is the slope of the regression line in plot of h_i against t_i values. k' is the slope of the regression line in plot of t_i against h_i values (Golbraikh and Tropsha 2002). Either the squared correlation coefficient (through the origin) between predicted and experimental values (R_{O}^2), or the coefficient between experimental and predicted values ($R_{\mathrm{O}}'^2$) should be close to 1. R^2 and R_{O}^2 are used for calculating R_{m}. For $R_{\mathrm{m}} > 0.5$, a prediction model can be considered valid. The above criteria were checked for the developed model and given in Table 4. This table indicates that the derived model meets the mentioned criteria and, therefore, is credible.

Furthermore, the MGGP model was compared with two other powerful soft computing methods, called multi-expression programming (MEP) and ANN (Mohammadzadeh et al. 2014) and several other well-known empirical models by other researchers (Skempton 1944; Nishida 1956; Cozzolino 1961; Terzaghi and Peck 1967; Sowers 1970; Azzouz et al. 1976; Mayne 1980; Park and Lee 2011). The MEP and ANN models were established upon the same data used for the development of the MGGP-based model. The comparison study was conducted on the testing data and presented in Table 5. According to this table, the proposed model has a comparable performance with ANN. The MGGP formula outperforms the MEP model and has a notable better performance than the existing regression-based models. In particular, the prediction error values (RMSE and MAE) for the existing empirical models are much higher than those for the MGGP model. It is notable that the regression analysis used for developing the C_{c} empirical models assumes the structure of the model by a limited number of linear or nonlinear equations. Hence, such models cannot consider the highly nonlinear interactions between the soil parameters and C_{c} (Alavi et al. 2011). Conversely from the empirical and analytical methods, MGGP models the mechanical behavior directly from the experimental data with no prior assumptions. The best MGGP-based equation is chosen among numerous linear and nonlinear preliminary models (Alavi et al. 2011; Alavi and Gandomi 2011).

Parametric and sensitivity analyses

To ensure the validity of MGGP-based model, a comparative parametric analysis was performed. Figure 7 shows the results of the parametric analysis. As it is seen, C_{c} increases with increasing LL up to about 32 % and then it decreases for higher LL values. Figure 7b indicates that C_{c}

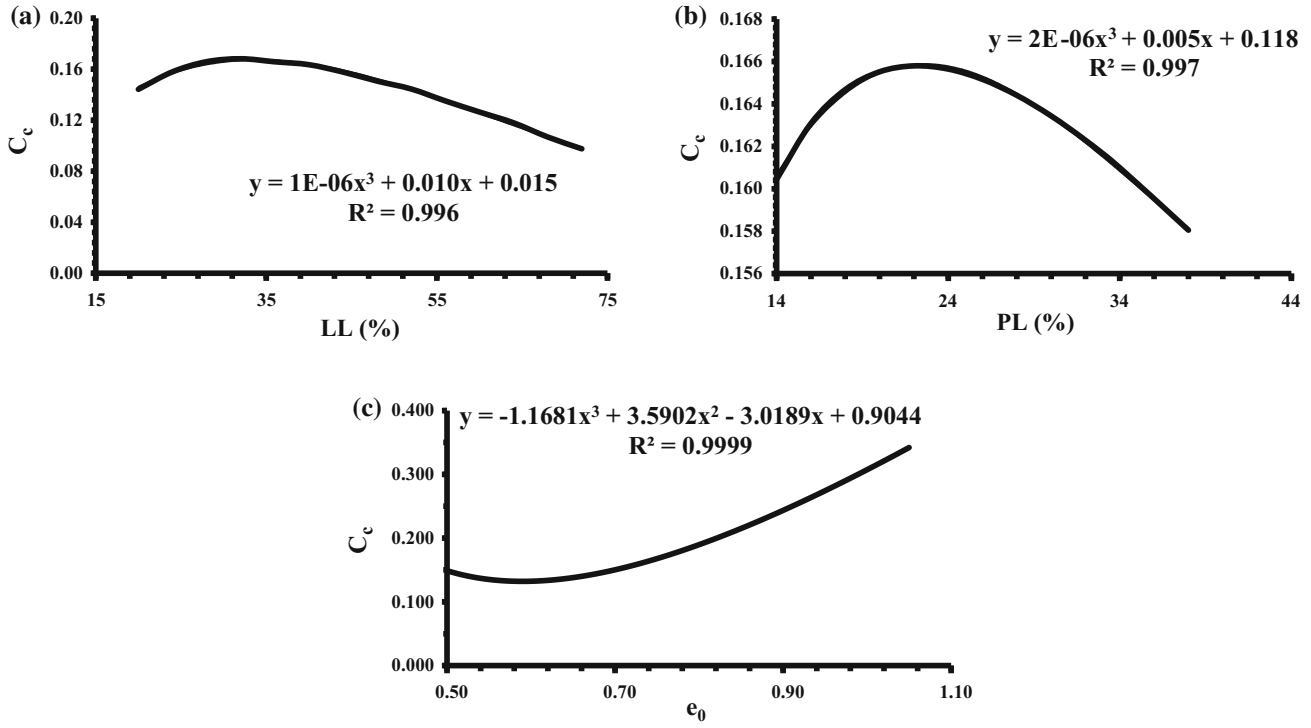


Fig. 7 Parametric analysis of C_c in the MGGP model

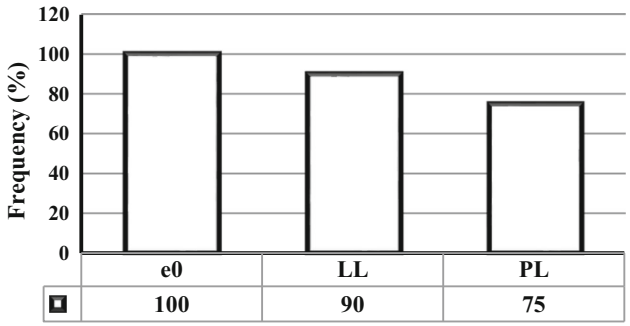


Fig. 8 Contributions of the predictor variables

is positively correlated with PL up to 22 % and for higher PL values C_c decreases. Referring to Fig. 7c, C_c decreases with increasing e_0 up to about 0.6 and thereafter it starts increasing.

As discussed before, the effect of all the considered parameters (i.e., LL, PL and e_0) on C_c is well understood. Excluding any of these three parameters from the modeling process decreased the accuracy of the derived models. Herein, a sensitivity analysis was conducted to provide a better understanding of the contribution of these important parameters to the prediction of C_c. A common approach for the sensitivity analysis in the GP-based modeling is to obtain the frequency values of the input parameters (Francone 1998–2004; Alavi et al. 2011; Gandomi et al. 2011a, b; Mohammadzadeh et al. 2014). A parameter with a frequency equal to 100 % has been included in 100 % of the best 30 programs generated by MGGP. The sensitivity analysis results are shown in Fig. 8. As it is seen, the frequency of e_0, LL and PL is equal to 100, 100, and 85 %, respectively. Thus, the MGGP-based model is more sensitive to e_0 and LL compared to PL. This is in agreement with the results reported by other researchers (Mohammadzadeh et al. 2014; Daryaee et al. 2010).

Besides, Fig. 9 shows the ratios of the experimental C_c values to those predicted by the MGGP model with respect to LL, PL and e_0. As the scattering increases in these figures, the model accuracy consequently decreases. It can be observed from these figures that the predictions obtained by the proposed model have a very good accuracy with no significant trend with respect to the design parameters. In the case of LL and e_0, the scattering slightly decreases with increasing these parameters.

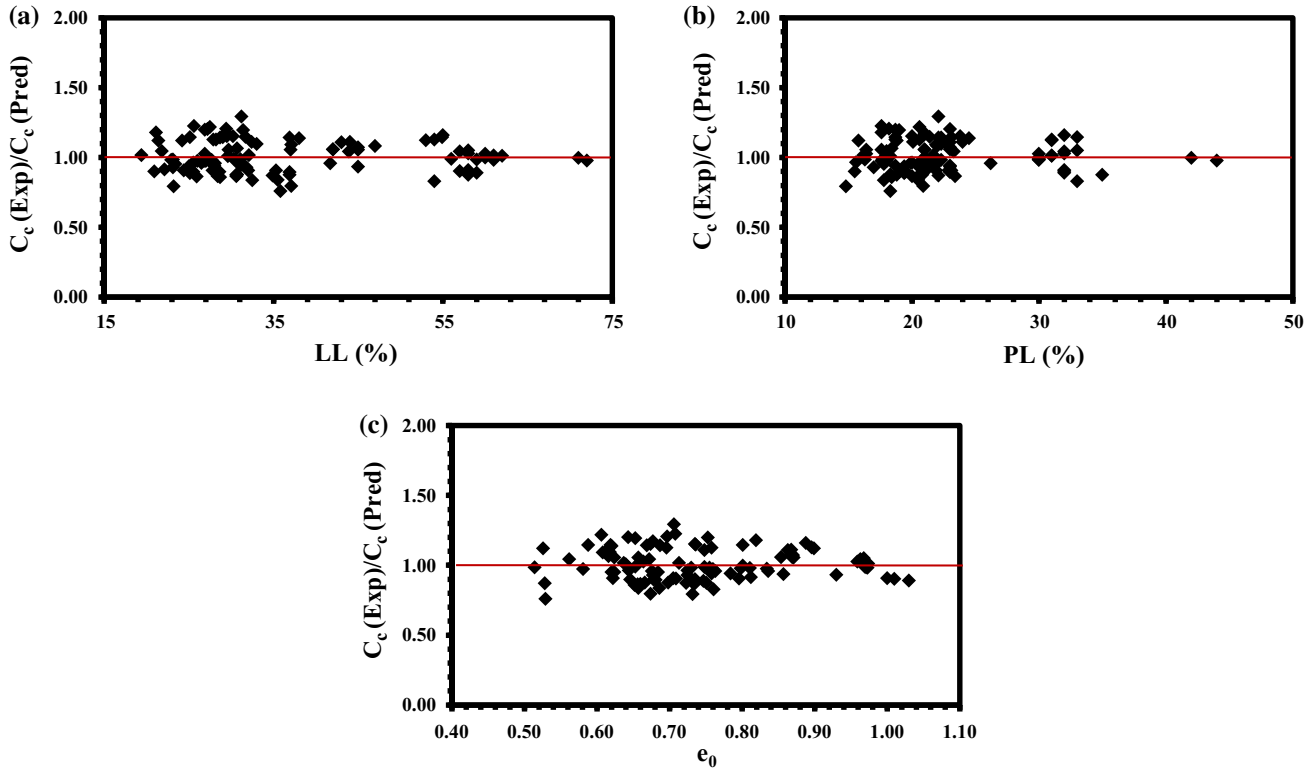


Fig. 9 The ratio of the experimental to the predicted C_c values with respect to the input parameters

Conclusion

This paper presented the MGGP method for the prediction of C_c of fine-grained soils. C_c was formulated in terms of basic properties of soil, i.e., LL, PL and e_0. The model was established upon a comprehensive experimental database. Extensive simulations were carried out to find the optimal model. To verify the validity of the model, various criteria were adapted. The proposed model has a good performance and outperforms the existing soft computing and empirical models. Contrary to the classical models for C_c, the MGGP model is highly nonlinear and includes complex combination of the predictor variables. For further evaluation of the behavior of the model, parametric and sensitivity analyses were performed. It was found that C_c increases with increasing LL and PL up to 32 and 22 %, respectively. In general, C_c was positively correlated with e_0. More, LL and e_0 seem to be more influencing on C_c than PL. Within its calibrated range, the proposed model can be considered as an efficient alternative to cumbersome testing procedures for the determination of C_c.

References

Abbasi N, Javadi AA, Bahramloo R (2012) Prediction of compression behaviour of normally consolidated fine-grained soils. World Appl Sci J 18(1):6–14

Akgun A, Türk N (2010) Landslide susceptibility mapping for Ayvalik (Western Turkey) and its vicinity by multicriteria decision analysis. Environ Earth Sci 61(3):595–611

Alavi AH, Gandomi A (2011) A robust data mining approach for formulation of geotechnical engineering systems. Eng Comput 28(3):242–274

Alavi AH, Ameri M, Gandomi AH, Mirzahosseini MR (2011) Formulation of flow number of asphalt mixes using a hybrid computational method. Constr Build Mater 25(3):1338–1355

Alavi AH, Hasni H, Lajnef N, Chatti K, Faridazar F (2016a) An intelligent structural damage detection approach based on self-powered wireless sensor data. Automation Constr 62:24–44

Alavi AH, Hasni H, Lajnef N, Chatti K, Faridazar F (2016b) Damage detection using self-powered wireless sensor data: an evolutionary approach. Measurement. doi:10.1016/j.measurement.2015.12.020

ASTM D2435 (2000) Standard test methods for one-dimensional consolidation properties of soils using incremental loading, vol 04.08. American Society for Testing Materials, Philadelphia, PA

Azamathulla HM (2012) Gene expression programming for prediction of scour depth downstream of sills. J Hydrol 460–461:156–159

Azamathulla HM, Wu F (2011) Support vector machine approach for longitudinal dispersion coefficients in natural streams. Appl Soft Comput 11(2):2902–2905

Azamathulla HM, Guven A, Demir YK (2011) Linear genetic programming to scour below submerged pipeline. Ocean Eng 38(8–9):995–1000

Azamathulla HM, Ghani AA, Fei SY (2012) ANFIS-based approach for predicting sediment transport in clean sewer. Appl Soft Comput 12(3):1227–1230

Azzouz AS, Krizek RJ, Corotis RB (1976) Regression analysis of soil compressibility. Soils Found 16(2):19–29

Bartlett SF, Lee HS (2004) Estimation of compression properties of clayey soils, Salt Lake Valley, Utah, University of Utah. Report No. UT-04.28. Utah Department of Transportation

Carter M, Bentley S (1991) Correlations of soil properties. Pentech Press, London

Ceryan N, Okkan U, Kesimal A (2013) Prediction of unconfined compressive strength of carbonate rocks using artificial neural networks. Environ Earth Sci 68(3):807–819

Cozzolino VM (1961) Statistical forecasting of compression index. Proc Fifth Int Conf Soil Mech Found Eng Paris 1:51–53

Daryaee M, Kashefipour M, Ahadiyan J, Ghobadiyan R (2010) Modeling the compression index of fine soils using artificial neural network and comparison with the other empirical equations. J Water Soil 24(4):659–667

Das SK, Basudhar PK (2008) Prediction of residual friction angle of clays using artificial neural network. Eng Geol 100(3–4):142–145

Das SK, Samui P, Sabat AK, Sitharam TG (2010) Prediction of swelling pressure of soil using artificial intelligence techniques. Environ Earth Sci 61(2):393–403

Das SK, Biswal RK, Sivakugan N, Das B (2011a) Classification of slopes and prediction of factor of safety using differential evolution neural networks. Environ Earth Sci 64(1):201–210

Das SK, Samui P, Khan SZ, Sivakugan N (2011b) Machine learning techniques applied to prediction of residual strength of clay. Cent Eur J Geosci 3(4):449–461

Das SK, Samui P, Sabat AK (2011c) Application of artificial intelligence to maximum dry density and unconfined compressive strength of cement stabilized soil. Geotech Geol Eng 29(3):329–342

Desai CK, Shaikh A (2012) Prediction of depth of cut for single-pass laser micro-milling process using semi-analytical, ANN and GP approaches. Int J Adv Manuf Technol 60:865–882

Desai VGM, Desai V, Rao DH (2009) Prediction of compression index using artificial neural networks. In: Indian geotechnical conference (IGC-2009), Guntur, India, pp 614–617

Farkhonde S, Bolouri J (2010) Estimation of compression index of clayey soils using artificial neural network. In: 5th National conference on civil engineering, Mashhad, Iran, Paper ID 1151

Francone F (1998–2004) Discipulus owner's manual. Register Machine Learning Technologies, Littleton

Gandomi AH, Alavi AH (2011) Multi-stage genetic programming: a new strategy to nonlinear system modeling. Inf Sci 181(23):5227–5239

Gandomi AH, Alavi AH (2012a) New multi-gene genetic programming approach to nonlinear system modeling. Part I: materials and structural engineering problems. Neural Comput Appl 21:171–187

Gandomi AH, Alavi AH (2012b) A new multi-gene genetic programming approach to nonlinear system modeling. Part II: geotechnical and earthquake engineering problems. Neural Comput Appl 21:189–201

Gandomi AH, Alavi AH, Sahab MG (2010) New formulation for compressive strength of CFRP confined concrete cylinders using linear genetic programming. Mater Struct 43(7):963–983

Gandomi AH, Alavi AH, Yun GJ (2011a) Formulation of uplift capacity of suction caissons using multi expression programming. KSCE J Civ Eng 15(2):363–373

Gandomi AH, Alavi AH, Yun GJ (2011b) Nonlinear modeling of shear strength of SFRCB beams using linear genetic programming. Struct Eng Mech 38(1):1–25

Golbraikh A, Tropsha A (2002) Beware of q2. J Mol Graph Model 20:269–276

Gopalakrishnan K, Kim S, Ceylan H, Khaitan S (2010) Natural selection of asphalt stiffness predictive models with genetic programming. In: Dagli CH et al (eds) Proceedings of the artificial neural networks in engineering (ANNIE), St. Louis, Missouri, 1–3 November 2010

Gregory AS, Whalley WR, Watts CW, Hallett PD, Whitmore AP, Bird NRA (2006) Calculation of the compression index and precompression stress from soil compression test data. Soil Tillage Res 89:45–57

Gulhati SK, Datta M (2005) Geotechnical engineering. Tata McGraw Hill Publishing Company Limited, New Delhi

Günaydın O (2009) Estimation of soil compaction parameters by using statistical analyses and artificial neural networks. Environ Geol 57(1):203–215

Guven A (2009) Linear genetic programming for time-series modelling of daily flow rate. J Earth Syst Sci 118(2):137–146

Guven A, Azamathulla HM, Zakaria NA (2009) Linear genetic programming for prediction of circular pile scour. Ocean Eng 36(12–13):985–991

Isik F, Ozden G (2013) Estimating compaction parameters of fine- and coarse-grained soils by means of artificial neural networks. Environ Earth Sci 69(7):2287–2297

Javadi AA, Rezania M, Mousavi Nezhad M (2006) Evaluation of liquefaction induced lateral displacements using genetic programming. Comput Geotech 33(4–5):222–233

Jianping I, Yangsong Z, Changhong Y, Guangyun G (2010) Application of BP neural network in prediction of compression index of soil. J Cent South Univ (Sci Technol) 41(2):722–727

Kaunda RB, Chase RB, Kehew AE, Kaugars K, Selegean JP (2010) Neural network modeling applications in active slope stability problems. Environ Earth Sci 60(7):1545–1558

Kayadelen C, Taşkıran T, Günaydın O, Fener M (2009) Adaptive neuro-fuzzy modeling for the swelling potential of compacted soils. Environ Earth Sci 59(1):109–115

Kim H-S, Cho G-C, Lee JK, Kim S-J (2013) Geotechnical and geophysical properties of deep marine fine-grained sediments recovered during the Second Ulleung Basin Gas Hydrate Expedition, East Sea, Korea. Mar Pet Geol 47:56–65

Kolay E, Kayabali K, Tasdemir Y (2010) Modeling the slake durability index using regression analysis, artificial neural networks and adaptive neuro-fuzzy methods. Bull Eng Geol Environ 69(2):275–286

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

Kumar VP, Rani CHS (2011) Prediction of compression index of soils using artificial neural networks (ANNs). Int J Eng Res Appl 1(4):1554–1558

Kumar Y, Venkatesh K, Kumar V (2012) Prediction of compression index of cohesive soil using neural network approach. In: International conference on advances in architecture and civil engineering, pp 363–366

Mahmood MT, Majid A, Han J, Choi YK (2013) Genetic programming based blind image deconvolution for surveillance systems. Eng Appl Artif Intell 26(3):1115–1123

Malih SK (2006) Fuzzy model for predicting fine-grained soil consolidation test parameters. Master Thesis, University of Guilan

Mayne PW (1980) Cam-clay predictions of undrained strength. J Geotech Eng Div ASCE 106(11):1219–1242

Mert E, Yilmaz S, Inal M (2011) An assessment of total RMR classification system using unified simulation model based on artificial neural networks. Neural Comput Appl 20:603–610

Mohammadi A, Wang G, Bolandi H, Alavi AH, Gandomi AH (2014) Multigene genetic programming for estimation of elastic modulus of concrete. Math Probl Eng. doi:10.1155/2014/474289

Mohammadzadeh D, Bolouri Bazaz J, Alavi AH (2014) An evolutionary computational approach for formulation of compression index of fine-grained soils. Eng Appl Artif Intell 33:58–68

Mousavi SM, Aminian P, Gandomi AH, Alavi AH, Bolandi H (2012) A new predictive model for compressive strength of high

performance concrete using gene expression programming. Adv Eng Softw 45(2011):105–114
Muduli PK, Das SK (2013a) CPT-based seismic liquefaction potential evaluation using multi-gene genetic programming approach. Indian Geotech J (Online). doi:10.1007/s40098-013-0048-4
Muduli PK, Das SK (2013b) SPT-based probabilistic method for evaluation of liquefaction potential of soil using multi-gene genetic programming. Int J Geotech Earthq Eng 4(1):42–60
Muduli PK, Das SK, Das MR (2013a) Uplift capacity of suction caisson in clay using artificial intelligence techniques. Mar Georesour Geotechnol 31(4):375–390
Muduli PK, Das SK, Bhattacharya S (2013b) CPT-based probabilistic evaluation of seismic soil liquefaction potential using multi-gene genetic programming. Georisk Int J. doi:10.1080/17499518.2013.845720
Nishida Y (1956) A brief note on compression index of soils. J Soil Mech Found Div ASCE 82(SM3):1027-1–1027-14
Ocak I, Seker S (2012) Estimation of elastic modulus of intact rocks by artificial neural network. Rock Mech Rock Eng 45(6):1047–1054
Ocak I, Seker SE (2013) Calculation of surface settlements caused by EPBM tunneling using artificial neural network, SVM, and Gaussian processes. Environ Earth Sci 70(3):1263–1276
Park H, Lee SR (2011) Evaluation of the compression index of soils using an artificial neural network. Comput Geotech 38:472–481
Park S, Choi C, Kim B, Kim J (2013) Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. Environ Earth Sci 68(5):1443–1464
Rani CHS, Kumar VP, Togati VK (2013) Artificial neural networks (ANNS) for prediction of engineering properties of soils. Int J Innov Technol Explor Eng 3(1):123–130
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
Roy PP, Roy K (2008) On some aspects of variable selection for partial least squares regression models. QSAR Comb Sci 27:302–313
Salehi H, Taghikhany T (2012) Application of robust-optimum algorithms in semi-active control strategy for seismic protection of equipment. 15th world conference on earthquake engineering, Lisbon, Portugal, 24–28 September, pp 11–20
Salehi H, Taghikhany T, Fallah AY (2014) Seismic protection of vulnerable equipment with semi-active control by employing robust and clipped-optimal algorithms. Int J Civil Eng 12(4):413–428
Salehi H, Das S, Chakrabartty S, Biswas S, Burgueno R (2015) Structural assessment and damage identification algorithms using binary data. ASME 2015 conference on smart materials, adaptive structures and intelligent systems, pp 1–10
Samui P (2008a) Support vector machine applied to settlement of shallow foundations on cohesionless soils. Comput Geotech 35(3):419–427
Samui P (2008b) Slope stability analysis: a support vector machine approach. Environ Geol 56(2):255–267
Sattari MT, Apaydin H, Ozturk F (2012) Flow estimations for the Sohu Stream using artificial neural networks. Environ Earth Sci 66(7):2031–2045
Searson DP (2009) GPTIPS: genetic programming and symbolic regression for MATLMATLAB. Manual Software. https://sites.google.com/site/gptips4matlab/
Searson DP, Willis MJ, Montague GA (2007) Co-evolution of nonlinear PLS model components. J Chemom 2:592–603
Searson DP, Leahy DE, Willis MJ (2010) GPTIPS: an open source genetic programming toolbox for multigene symbolic regression. In: Proceedings of the international multiconference of engineers and computer scientists, Hong Kong
Shahin M, Jaksa M (2005) Neural networks prediction of pullout capacity of marquee ground anchors. Comput Geotech 32(3):153–163
Singh A, Noor S (2012) Soil compression index prediction model for fine grained soils. Int J Innov Eng Technol 1(4):34–37
Skempton AW (1944) Notes on the compressibility of clays. Q J Geol Soc Lond 100:119–135
Smith GN (1986) Probability and statistics in civil engineering. Collins, London
Sower GB (1970) Introductory soil mechanics and foundation, 3rd edn. The Macmillan Company of Collier-Macmillan Ltd., London
Tasdemir Y, Kolay E, Kayabali K (2013) Comparison of three artificial neural network approaches for estimating of slake durability index. Environ Earth Sci 68(1):23–31
Terzaghi K, Peck RB (1967) Soil mechanics in engineering practice. Wiley, New York
Tiwari BT, Ajmera B (2012) New correlation equations for compression index of remoulded clays. J Geotech Geoenviron Eng 138(6):757–762
Wroth CP, Wood DM (1978) The correlation of index properties with some basic engineering properties of soils. Can Geotech J 15:137–145
Wu X, Niu R, Ren F, Peng L (2013) Landslide susceptibility mapping using rough sets and back-propagation neural networks in the Three Gorges, China. Environ Earth Sci 70(3):1307–1318
Yilmaz I (2010a) The effect of the sampling strategies on the landslide susceptibility mapping by conditional probability and artificial neural networks. Environ Earth Sci 60(3):505–519
Yilmaz I (2010b) Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. Environ Earth Sci 61(4):821–836
Yilmaz I, Marschalko M, Bednarik M, Kaynar O, Fojtova L (2012) Neural computing models for prediction of permeability coefficient of coarse-grained soils. Neural Comput Appl 21(5):957–968