KSCE Journal of Civil Engineering (2013) 17(1):106-116
DOI 10.1007/s12205-013-1847-y

Geotechnical Engineering

www.springer.com/12205

A Methodology to Build a Groutability Formula via a Heuristic Algorithm

Chien-Lin Huang*, Jen-Chen Fan, Kuo-Wei Liao***, and Tsung-Hsun Lien******

Received December 18, 2011/Accepted April 17, 2012

Abstract

The goal of this study is to provide a methodology to develop a groutability (N) formula of sandy silt soils using microfine cement grouts in a permeation grouting. Because the Fines Content (FC) of the sandy silt soils studied is relatively high, and the size of the grouts used is significantly smaller than the Portland cement, the existing empirical formulas cannot deliver a promising prediction of N. Support Vector Machines (SVMs) is an alternative tool used to predict N. However, SVMs do not provide an explicit formula, which creates an obstacle for practical engineers. Thus, a heuristic algorithm (the Tabu search, TS) was used to build the prediction formula. A total of 240 in-situ data samples were analyzed to ensure the accuracy of the proposed formula. The format of the existing empirical formula was adopted in the proposed TS-based formula. Four parameters were considered in our TS models: the effective soil particle size (D_{10}), the soil particle size (D_{15}), the water-to-cement ratio (*w/c*) and the FC. The prediction accuracy of the TS-based formula was approximately 94.17%, indicating that the proposed formula is a suitable tool. Because the proposed formula has a similar format to that of formulas that are typically used, the proposed approach can be implemented readily in practical engineering settings. Note that the proposed formula was only verified by the collected data samples, the suitability of applying the built formula to other conditions needs more investigation.

Keywords: *Tabu search, groutability, microfine cement, permeation grouting, SVMs*

1. Introduction

With the development of microfine cement grouts and low-pressure grouting technology, permeation grouting has been successfully applied to engineering problems: rehabilitation projects for bridge structures (Khayat *et al*., 1997; Perret *et al*., 2002), environmental projects to isolate polluted sandy soil (Schwarz and Krizek 2006), remedial works of the defected earth fill dam (Chun *et al*., 2006) and retrofit projects to improve liquefaction resistance in sandy silt soils (Huang *et al*., 2007). One of the major challenges of using microfine cement grouts is estimating the groutability, N, within a reasonable level of error. Another challenge is whether an approach provides an easy evaluation process for practical engineers. Thus, this study aimed to develop an explicit prediction formula for sandy silt soils with a focus on microfine cement grouts used in permeation grouting.

Predicting the N with cement-based grouts is not an easy task. Many parameters affect N and must be evaluated systematically. For practical reasons, most researchers used the relative grain-size ratio between soils and grouts to predict N (Huang *et al*., 2007; Burwell 1985; Krizek *et al*., 1992; Incecik and Ceren 1995; Axelsson *et al*., 2009), as described in Eqs. (1) and (2).

$$N_1 = \frac{(D_{15})_{soil}}{(d_{85})_{grout}} \quad (1)$$

$$N_2 = \frac{(D_{10})_{soil}}{(d_{95})_{grout}} \quad (2)$$

A successful grouting should have N_1 larger than 15 and N_2 larger than 8 (Krizek *et al*., 1992), where D_{15} is the diameter through which 15% of the total soil mass passes and d_{85} is the diameter through which 85% of the total grout passes. Similar regulations are applied for D_{10} and d_{95}.

The existing empirical formulas may fail to deliver a promising prediction of N when microfine cement grouting with a high water-to-cement ratio (*w/c*) is used for sandy silt soils with high fines content (FC, size of soil particle < 0.074 mm) for several reasons. First, considering the effect of grout size, 70% of the microfine cements studied here have a grain size less than 1 μm, which is much smaller than that in traditional cement-based grouts (e.g., the Portland cement). Second, the FC of the sandy silt soils in this study is relatively high; that is, the effective soil particle size (D_{10}) and the soil particle size (D_{15}) have a very high probability of being less than 0.074 mm. This D_{10} or D_{15} will cause the relative grain-size ratio between soil and grout in the

*Manager, TRICORNTECH Corporation, 5941 Optical Court, San Jose, CA 95138, USA (E-mail: cheneyhuang@yahoo.com.tw)
**Professor, Dept. of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 106, Taiwan (E-mail: jcfan@ntu.edu.tw)
***Assistant Professor, Dept. of Construction Engineering, National Taiwan University of Science & Technology, Taipei 106, Taiwan (Corresponding Author, E-mail: kliao@mail.ntust.edu.tw)
****Former Research Assistant, Dept. of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 106, Taiwan (E-mail: b93602036 @ntu.deu.tw)

empirical formulas to be smaller than expected. Third, when *w/c* is larger than 3 like in this study, grouts often exhibit better flow properties, such as lower viscosity and higher infiltration rate. Thus, the empirical formulas may not be suitable to use only the relative grain-size ratio to determine N.

Thus, in addition to the relative ratio between the soils and grouts, there are many other important factors that should be considered when predicting N. For example, Zebovitz *et al.*, (1989) stated that the FC in the sand may control the grouting operation. In some cases, even though the relative grain-size ratio between soil and grout was satisfied, the sample was not groutable when the fine sands contained 5% fines by weight. Ozgurel and Vipulanandan (2005) also concluded that the N of sand with acrylamide grout was influenced by the amount of FC. More factors should be considered when constructing a prediction model. Thus, Akbulut and Saglamer (2002) included the FC and *w/c* in their prediction formulas.

In addition to the existing formulas introduced earlier, Artificial Neural Networks (ANNs) or Support Vector Machines (SVMs) are potential tools that can be used to build a prediction model of N. ANNs have been successfully applied to many geotechnical engineering problems, such as predicting the bearing capacity of strip footing (Kuo *et al.*, 2009), predicting the lateral load capacity of piles in clay (Das and Basudhar, 2006), analyzing the slope stability analysis (Cho, 2009) and predicting the rock fragmentation due to blasting (Bahrami *et al.*, 2011). ANNs also have been applied to the groutability problem. For example, the N of granular soils with cement-based grouts was estimated using an ANN model based on a database of 87 laboratory observations (Tekin and Akbas, 2011). The N of sandy silt soils with high FC was predicted via a Radial Basis Function Neural Network (RBFNN) (Liao *et al.*, 2011). Both ANN models provided a good correlation between the observed and predicted outcomes.

The SVM is a useful technique for data classification. Constructing a SVM is often considered to be easier than building an ANN model. SVMs also have been applied to many engineering problems, such as predicting the blast-induced ground vibration (Khandelwal, 2011), detecting local damages in a building structure (Mita and Hagiwara, 2003), identifying the lateral flow occurrence (Lee and Kim, 2010), analyzing hand writing, which require pattern classification or are regression-based applications.

Neither ANNs nor SVMs provide a specific formula for engineers, although they have the potential to deliver a promising prediction. The concept and operation of an ANN or an SVM model is a black box for some practical engineers. On the other hand, an explicit empirical formula is often used in geotechnical engineering. Thus, this study aimed to develop a formula with a similar format to that in existing formulas that provides an easy evaluation procedure for predicting N.

As mentioned earlier, the existing formulas, as described in Eqs. (1) and (2), may not be able to deliver a good prediction of N due to the size effects of soils and grouts. To improve the prediction accuracy, new threshold values for N_1 and N_2 in Eqs. (1) and (2) are needed. As discussed, other factors such as FC and *w/c*, if considered, can provide better prediction accuracy. In the latter case, at least four new thresholds need to be determined. Assuming that there are 10^3 possibilities for each threshold, the size of the search domain is 10^{12}. For such a large design domain, the Tabu Search (TS) is a good solution tool because a TS can reduce the computational cost using a Tabu list. Moreover, TS usually avoids a local optimum and can be applied to both discrete and continuous solution spaces (Tung and Chou, 2002). Thus, TS was adopted to build a prediction formula in this study. Please note that the results of permeation grouting performed in the field and in the laboratory may be different; only data samples from the field were considered in this study.

The 240 in-situ data samples of permeation grouting with microfine cement grouts were collected in the cities of Taipei and Kaohsiung, Taiwan and were analyzed by the proposed TS-based models. Two TS models were investigated. The first model has the same format as that used in Eqs. (1) and (2). The second model includes two additional parameters, the FC and the *w/c*. The effects of FC and *w/c* on N were examined thoroughly by comparing the results of these two TS models.

The results from the TS-based model were compared with those of some existing methods. The accuracy rates of our model were 91.25% and 94.17% for the first and the second model, respectively; these accuracy rates are much higher than the existing methods, which were only 45% to 68% accurate. These results indicate that the proposed method improves the prediction outcomes significantly. It is evident that the proposed TS-based model can be used for future soil improvement projects to avoid unnecessary or inefficient grouting, which reduces the engineering cost.

Although the SVM does not provide a formula for N, it is often considered to be a robust tool for predicting N. Thus, a SVM was conducted, and its results were used to validate the results obtained from the TS-based models.

2. Existing Approaches

2.1 The Empirical Formulas

Several relationships between soil grain-size and cement grout particle size have been defined to build a groutability prediction formula. For example, Burwell (1985) proposed Eqs. (1) and (2) to predict the N. Grouting only succeeds when N_1 is greater than 25 and N_2 is greater than 11. Krizek *et al.* (1992) proposed another similar approach; they suggested the identical equations (i.e., Eqs. (1) and (2)) but with different thresholds: grouting only succeeds if N_1 is greater than 15 and N_2 is greater than 8. Similar groutability rules for soils in field grouting were also observed (Huang *et al.*, 2007). They used $N_1 > 9$ or $N_2 > 4$ as a criterion to predict the N when microfine cement (MFC-GM8000) was used for the sandy silt soils.

Incecik and Ceren (1995) proposed an alternate equation as follow:

$$N_1 = \frac{(D_{10})_{soil}}{(d_{90})_{grout}} \tag{3}$$

Based on Eq. (3), grouting succeeds if N_1 is greater than 10. Axelsso *et al*. (2009) defined N_1 as follow

$$N_1 = \frac{b_{fic}}{(d_{95})_{grout}} = \frac{0.15D_{50}}{(d_{95})_{grout}} \tag{4}$$

where b_{fic} is the fictitious aperture. b_{fic} is determined through the following equation:

$$b_{fic} = 0.15 \times D_{50} \tag{5}$$

The threshold for N_1 depends on the *w/c*, taking *w/c* = 2.0 for example, grouting succeeds if $N_1 > 4.2$, while grouting cannot succeeds if $N_1 < 3$. Between the thresholds, grouts cannot be sufficiently injected into the soils.

The above formulas are mostly based on the relative ratio of the soil grain-size to the grout particle size. Akbulut and Saglamer (2002) conducted a series of tests in the laboratory and based on the results, they proposed a multi-parameter formula to predict the N as follow:

$$N_1 = \frac{(D_{10})_{soil}}{(d_{90})_{grout}} + k_1\frac{w/c}{FC} + k_2\frac{P}{D_r} \tag{6}$$

in which, P is in kPa, and D_r is the relative density of soil samples. $k_1 = 0.5$ and $k_2 = 0.01$ in 1/kPa are constants. The soil is groutable if $N_1 > 28$. Note that Eq. (6) provides a reasonable prediction only when the following conditions are satisfied: $0 < FC < 6\%$, $0.8 < w/c < 2$ and $50 < P < 200$.

2.2 The ANN Models

Tekin and Akbas (2011) developed an ANN model to estimate the N using a database of 87 laboratory results. In their model, the *w/c*, the relative density of the soil, the *P*, the D_{15} and the d_{85} were used as the input parameters. This ANN model provided a very accurate prediction.

Liao *et al*. (2011) used the Radial Basis Function Neural Network (RBFNN) to develop a prediction model for the same data samples studied here. In this earlier research, RBFNN provided a prediction with an accuracy of 94.97%. Several parameters were considered in the network; the grain-size of the soil (D_{10}/D_{15}), the void ratio (e), the FC, the uniformity coefficient (Cu), the coefficient of gradation (Cz) and the *w/c*. The result from this earlier study and from the SVM conducted in the current study were both used to validate the prediction performance of the proposed TS-based formulas as described in Section 5.

3. Concepts of the Built TS & SVM

3.1 The Basic Framework of the Proposed TS

TS is a heuristic optimization method that was originally proposed by Glover (1986). TS keeps moving from one point to another point to find the optimal solution. Solutions near the current best point are evaluated. The one with the best solution is the next point. To prevent any new search from moving back to the earlier points visited or getting trapped at a local optimum, a Tabu list is designed in which the better solutions from past searches are recorded. A new search point cannot be on the Tabu list. The length of Tabu list used here was 7 which is suggested by Glover (1986). If a Tabu list only records parts of design variables, TS often allows a move, even if it is recorded on the Tabu list. Such regulation is called aspiration criterion. TS is stopped by certain criteria, such as a predefined number of iterations, the convergence of the objective functions in the consecutive iterations or the desired amount of computational time (Hillier and Lieberman 2005). The step-by-step process of the proposed TS is described in Section 4.4.

3.2 The Basic Framework of the Proposed SVM

SVMs were first introduced by Boser *et al.* (1992). The goal of SVM is to separate the collected data samples into two classes using a linear or nonlinear boundary (classifier). A suitable kernel function can increase the chance of separating the nonlinear data. A Gaussian Radial Basis Function (RBF) kernel, as depicted in Eq. (7) is often considered as a reasonable choice and therefore was adopted in this study.

$$K(Y, Y_i) = e^{-\gamma(\|Y - Y_i\|)^2} \tag{7}$$

where Y is the input vector, γ is the parameter of the kernel function and Y_i are the support vectors. SVM uses the principle of the Structural Risk Minimization (SRM), as described in Eq. (8), to determine a classifier that has maximum separation distance.

$$\begin{aligned} &\text{minimize } \frac{1}{2}w^T w \\ &\text{subject to } y_i(w^T x_i + b) \ge 1 \quad i = 1, 2, \dots, n \end{aligned} \tag{8}$$

where w is a vector that defines a direction perpendicular to the hyper-plane; y_i is a label representing the two classes; $w^T x_i + b$ (=1) is the classifier; and n is the number of training data samples. In addition, SVM uses the slack variables (ξ_i) to obtain a smooth, not highly curved classifier. The mathematical formulation of the SVM including slack variables is described in Eq. (9).

$$\begin{aligned} &\text{minimize } \frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i \\ &\text{subject to } \begin{cases} y_i(w^T x_i + b) \ge 1 - \xi_i \\ \xi_i \ge 0, i = 1, 2, \dots, n \end{cases} \end{aligned} \tag{9}$$

where C is the error penalty and $\sum_{i=1}^{N}\xi_i$ represents the sum of the distances of the points on the wrong side.

Based on the concepts described above, two parameters, the kernel parameter (γ) and the error penalty (C), need to be determined in the proposed SVM. The following provides a detailed procedure of the proposed SVMs.

1. Normalization: each considered factor (D_{10}, D_{15}) was normalized to a suitable range, such as 0-1, to average the influence of each factor in the SVM model.
2. Mapping: a Gaussian RBF kernel function was used to map the original data into a nonlinear feature space.

3. Cross validation: To avoid the over-fitting, a cross validation was adopted. A total of 240 data samples were collected in this study. Of these, 195 were used in the training task, and 45 were used in the testing task. In the cross validation, the 195 data sample data were further divided into 13 subgroups.
4. Grid search: A grid search was used to find the values of the γ and C. The values of γ ranged from 0 to 50 with a step size of 1. The values of C ranged from 0 to 100 with a step size of 1.
5. Training: the training was conducted by using the best γ and C obtained from the cross validation and the grid search. Note that the data number for training was 195 which was different with the number (180) in the cross validation.
6. Testing: with the SVM classifier obtained in the training process, the testing can be performed directly, and the prediction accuracy was obtained.

4. Application of TS to the groutability problem

The overall procedure used in this study is provided in Fig. 1. As shown, the N prediction performance of four different approaches, the empirical formulas, the ANN, the SVM and the TS-based formulas, were compared. In this section, a detailed description of applying TS to the groutability problem is provided.

4.1 The Data Samples

A total of 240 on-site permeation grouting data samples were collected in the cities of Taipei and Kaohsiung, Taiwan. These 240 data were collected from 25 grouting holes. In each grouting hole, data samples (e.g., between 5 and 11) were recorded from the boring log. 21 of 25 grouting holes were located in the city of Kaohsiung, and a total of 220 grouting data samples were collected; a total of 20 grouting data samples from 4 grouting holes were collected in Taipei city. The use of permeation grouting in Kaohsiung city was intended to retrofit the bearing capacity of the soil foundation. The permeation grouting used in Taipei city was meant to prevent soil liquefaction. A mix of microfine cement (MFC-GM8000A) and micro-slag (MFC-GM8000B) in equal proportions was used as the injected grouts. The diameters through which 95%, 90% and 85% of the total grout passes are 7.4 μm, 6.4 μm and 4.5 μm, respectively. Furthermore, the diameter through which 70% of the total grout passing is less than 1 μm.

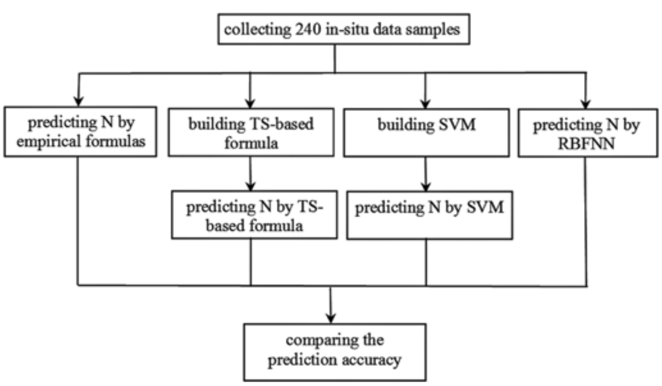

Fig. 1. The Flowchart of the Groutability Study

In the collected data, the parameters $(D_{10})_{soil}$, $(D_{15})_{soil}$, FC and w/c were considered in the proposed TS-based formulas. The corresponding information for each parameter is provided in Table 1. The statistical information for e, C_z and C_u are also provided, even though they were not included in the proposed formula. Note that the values of $(d_{85})_{\text{grout}}$ and $(d_{95})_{\text{grout}}$ were also included in the proposed TS-based formula. Because only one type of grout was used in this study, these two values were deterministic numbers and did not affect the optimization search.

4.2 The Preparation of Training and Testing Data

The idea of the proposed TS approach is to use a portion of the data samples to determine the values of the parameters in the proposed formula; this process is called *training*. Training delivers a deterministic formula. This formula can be used to predict the N of the remaining data samples; this process is called *testing*. The number of data samples for the training and testing used in TS is the same as that used in the SVM introduced in Section 3.2. That is, 195 and 45 data samples were used for training and testing, respectively. The details regarding the number of data samples used in training and testing procedures are shown in Table 2. As shown in Table 2, to create a generic TS-based formula, for each w/c value, the ratio of holes used in the training procedure should be as similar as possible. For example, the ratio is 0.75, 0.83 and 0.78 for w/c values of 3.34, 4.0 and 4.65, respectively. Using Table 2 to separate data samples into two groups (training and testing) results in multiple possibilities. An arbitrary selection among many possibilities was used to complete the construction of the TS-based formula. An additional 9 possible separations were analyzed using the constructed TS formula to ensure that the result was generic and that the data had not been overfit, which is discussed further in Section 5.4.

Table 1. Statistics of the Collected Data Samples

Factors	Max. value	Min. value	Mean	Standard deviation
D_{10}	85 μm	0.1 μm	20.07 μm	20.83μm
D_{15}	125 μm	0.2 μm	32.01 μm	27.54μm
e	1.04	0.35	0.71	0.13
FC	99.6%	6.9%	41.67%	29.94%
C_z	27.84	0.02	2.44	2.42
C_u	581.82	2.11	26.38	63.12
w/c	4.65	3.34	4.0*	0.53

*median

Table 2. The Number of Grouting Data Samples used in the Training and Testing Procedures

	MRT (Taipei)	No. 1 highway (Kaohsiung)		Total
	w/c=3.34	w/c=4.0	w/c=4.65	
No. of data (Holes) for training	13(3)	107(10)	75(7)	195(20)
No. of data (Holes) for testing	7(1)	20(2)	18(2)	45(5)

4.3 The Proposed Prediction Formulas

Two formulas were built based on TS. The first formula considered $(D_{10})_{soil}$, $(D_{15})_{soil}$, $(d_{85})_{grout}$ and $(d_{95})_{grout}$. The second formula considered FC and *w/c* as well. The format of the first formula is the same as that in Eqs. (1) and (2). That is, TS finds the best threshold values (*a* and *b*) of N_1 and N_2 for microfine cements in the sandy silt soils. If $N_1 > a$ and $N_2 > b$, the soil is considered to be groutable. If one of the conditions above $(N_1 > a$ and $N_2 > b)$ is not satisfied, the soil is considered to be non-groutable. The variables *a* and *b* should be a positive constant number in nature. The mathematical formulation of the Tabu optimization is described as follows:

$$\text{minimize } \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}-\boldsymbol{Y})^2}$$

$$Subject\ to: a>0, b>0 \tag{10}$$

where $\boldsymbol{X}$ is the N observed from the field; n is the number of training samples; $\boldsymbol{Y}$ is the groutability predicted by the proposed formula as described in Eqs. (1) and (2), $\boldsymbol{Y}$ is further defined as follows:

$$\boldsymbol{Y}\begin{cases}1, \text{ if } N_1>a, N_2>b \\ 0, otherwise\end{cases} \tag{11}$$

$\boldsymbol{X}$ is defined as the volume of injected grouts is at least two times the volume of void space in a given soil under the split pressure. There are 195 training samples in this study. Note that the groutability problem has been converted into an optimization problem by introducing the auxiliary variable, Y. The objective of this optimization is to minimize the Root Mean Square Error (RMSE) of the predicted N. The design variables are the thresholds of N_1 and N_2, that is *a* and *b*. A lower bound constraint for the design variables is added (must be a positive number) when formulating the optimization problem for practical situations.

To investigate the effects of FC and *w/c* on the prediction of N, another formula was constructed, as displayed in Eq. (12).

$$N_1 = \frac{(D_{15})_{soil}}{(d_{85})_{grout}} + k_1\frac{w/c}{FC}$$

$$N_2 = \frac{(D_{10})_{soil}}{(d_{95})_{grout}} + k_2\frac{w/c}{FC} \tag{12}$$

A similar groutability rule is also applied to Eq. (12). That is, if $N_1 > a$ and $N_2 > b$, then the soil is considered to be groutable ($Y = 1$). Otherwise, the soil is considered to be non-groutable ($Y = 0$). In the current formula, the TS needs to determine four unknown numbers (*a, b,* k_1 and k_2) instead of two. k_1 and k_2 are a constant number that indicates the relative importance of *w/c* and FC compared to the ratio of the grains. The mathematical formulation of the Tabu optimization is revised as follows:

$$Minimize\ \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}-\boldsymbol{Y})^2}$$

$$Subject\ to: a>0, b>0, k_1>0, k_2>0 \tag{13}$$

Some important parameters used in the TS were provided as follows. The length of the Tabu list was 7, and the number of neighborhoods was 8 and 80 for Eqs. (10) and (13), respectively. Other important issues, such as the termination rules for the search and step size for each movement, are described in the following section.

4.4 The Step-by-step Procedure of the Proposed TS

Based on the above description, the procedure of the proposed TS is summarized as follows.

Step 1: Choose an arbitrary starting point and evaluate its objective function (Eq. (10) or Eq. (13)).

Step 2: Use the predefined step size and the search direction to determine the subsequent points and evaluate their objective functions. The step size is first set to be 2 at the initial loop to increase the search area as large as possible. If the objective is not converged, the subsequent step size is reduced to half of the previous size. For each design variable, the current value, the current value with one unit of positive and negative step sizes is the search directions. For example, there are two design variables in the first proposed formula. Thus, there are eight search points in each loop as shown in Fig. 2. Similar regulation is used for the second proposed formula. As a result, there are eighty search points for each loop.

Step 3: Move to the next point. The next point is first chosen by selecting the point with the smallest objective value. This chosen point cannot be on the Tabu list. If the chosen point is on the Tabu list, then the point with smallest value among the remaining points is selected as the candidate for the next point. This chosen point again cannot be on the Tabu list. Otherwise, another candidate should be chosen until all the regulation is met. The number of design variable in both proposed optimization problem is not a big number. In such situation, including every design variable in the Tabu list will not create a computational burden and is affordable. Thus, this study does not use the aspiration criteria in TS.

Step 4: Update the Tabu list. Once the search moves to the next point, the Tabu list should be modified accordingly.

Step 5: Check the convergence of the objective function and step size. If both quantities are converged, then the final solu-

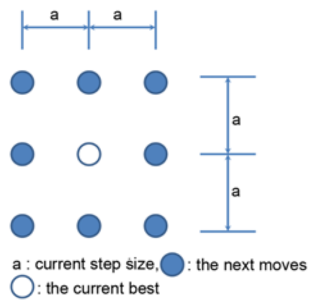

Fig. 2. Illustration of Search Direction for the First Proposed Formula

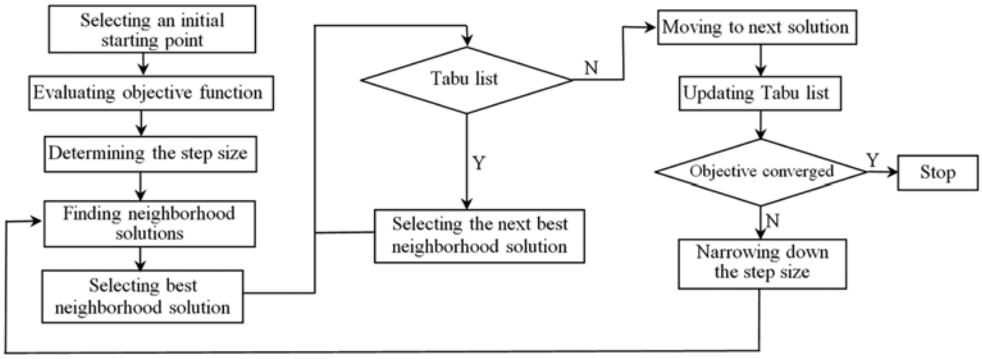

Fig. 3. The Flowchart of the Proposed TS

tion is found. Otherwise, proceed to the next step.

Step 6: Revise the step size to a smaller number. After modifying the step size, repeat step 2 to step 5 until the final solution is found.

A detailed flowchart is displayed in Fig. 3. It is seen that the TS tries to locate the optimal solution by exploring the neighborhood solutions. The Tabu list is used to avoid redundant evaluation at trial points. The search stops when the objective value converges.

5. Results and Discussions

5.1 Results of the Empirical Formulas

The predicted results of the collected data using empirical formulas are summarized in Table 3. The correct percentages for non-groutable soils were very high, ranging from 97.75% to 100%. However, the correct percentages for groutable soils were relatively low, ranging from 1.03% to 49.33%. This result indicates that most existing empirical formulas only provide accurate predictions when the soil is non-groutable. However, the probability that a particular soil is groutable ranges from 46% (76/166) to 59% (131/221) when the formulas predict that the soil was non-groutable. That is, the probability of inaccurate prediction is relatively high. Overall, the accuracy of the empirical formulas (ranging from 45% to 68%) is unsatisfactory.

Table 3. Prediction Results of the Existing Formulas

Empirical formula	Results from field		Results from formula ✓[1]	Results from formula X[1]	Accuracy	Overall accuracy
Burwell *et al.* (1985)[2]	✓		1	96	1.03%	(90+1)/187 = 48.66%
	X		0	90	100%	
Incecik and Ceren (1995)	✓		19	131	12.67%	(19+90)/240 = 45.42%
	X		0	90	100%	
Krizek *et al.* (1992)	✓		20	130	13.33%	(20+90)/240 = 45.83%
	X		0	90	100%	
Huang *et al.* (2007)	N_1[3]	✓	74	76	49.33%	(74+90)/240 = 68.33%
		X	0	90	100%	
	N_2[3]	✓	54	96	36.00%	(54+90)/240 = 60.00%
		X	0	90	100%	
Axelsson *et al.* (2009)[4]	✓		19	102	15.70%	(19+87)/210 = 50.48%
	X		2	87	97.75%	

[1] X denotes non-groutable, ✓ denotes groutable

[2] 53 of 240 data samples were not defined and were therefore excluded from this evaluation

[3] Defined in Eqs. (1) and (2)

[4] 30 of 240 data samples were not defined and were therefore excluded from this evaluation

5.2 Results of the TS

5.2.1 Results of the First TS-based Formula

The best values of the constant number (a, b) were found to be 3.125 and 1.500, respectively. Detailed results of the TS are displayed in Table 4. The time histories of the objective function are shown in Fig. 4. It was found that five loops were needed for the solution to converge. A loop refers to a complete optimization calculation with a given step size, that is, Step 2 to Step 4 that are described in Section 4.4. Also, it is seen that the best solution was found at iterations 4, 9 and 15 for the first, second and third loop, respectively. The total number of iterations in each loop is predefined to be 100. The results indicated that the best solution was found in an early stage of a TS search. Fig. 5 displayed the explored area in the search for the threshold values (a and b) for each loop. It is seen that the proposed TS initially used a larger step size to avoid getting trapped in a local minimum. Consecutive optimization was initialized at the previous optimal point using a finer step size. For example, the starting point, the best point and the last point were (6, 7), (2, 1) and (20, 25) in the first loop, which are represented by right, up and left triangles, respectively (Fig. 5, top left). The best point (2, 1) was used as the initial point in the second loop, which is shown as a right triangle in Fig. 5 (top right).

Based on the optimization results obtained from the TS, the proposed TS-based formula stated in Eqs. (1) and (2) was updated as follows:

Table 4. Optimization Results of the First TS-based Formula

Parameters	a	b
initial solution	6	7
Initial step size	2.0	
optimal solution	3.125	1.500
length of TS list	7	
accuracy in training	91.79%	
Accuracy in testing	88.89%	
Overall accuracy	91.25%	

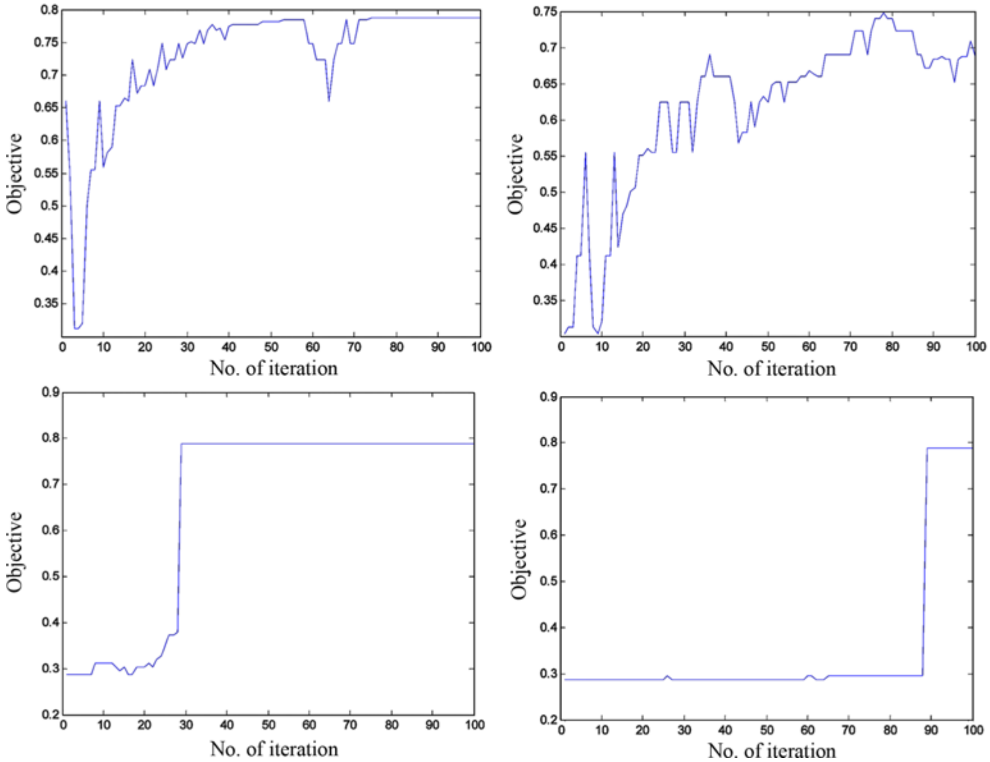

Fig. 4. Time History of the Objective Function (Top left: the 1st loop, top right: the 2nd loop, bottom right: the 3rd loop, bottom right: the last loop)

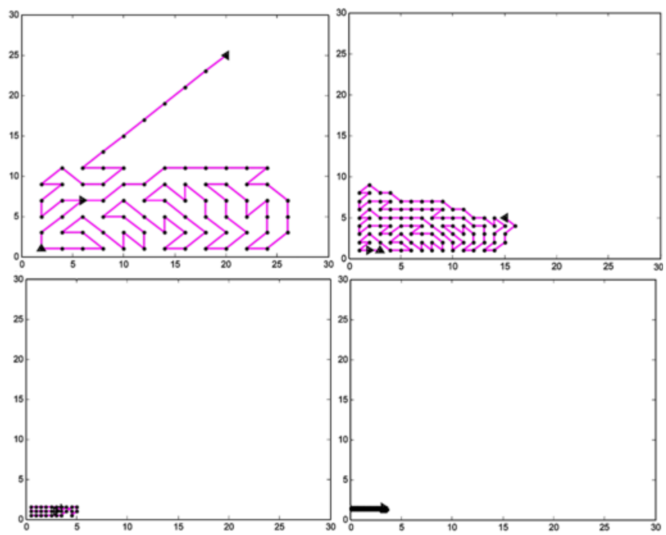

Fig. 5. The Area Explored in the Search for a (x-axis) and b (y-axis) (The right triangle, up triangle and left triangle denote the initial, best and last point (top left: the 1st loop, top right: the 2nd loop, bottom right: the 3rd loop, bottom right: the last loop))

$$\text{if}\begin{cases} N_1 = \dfrac{(D_{15})_{soil}}{(d_{85})_{grout}} > 3.125 \\ N_2 = \dfrac{(D_{10})_{soil}}{(d_{95})_{grout}} > 1.500 \end{cases} N \text{ is groutable}$$
$$\text{else} \qquad N \text{ is non-groutable} \tag{14}$$

Based on Eq. (14), the correct percentage for the training and testing procedure was 91.79% and 88.89%, respectively (Table 4). The prediction accuracy from the first TS-based formula is apparently at a different performance level compared to those of existing formulas, which indicates that the TS is a suitable tool for building the prediction formula. However, this result is not as accurate as results obtained from SVM (Section 5.3) or RBFNN (Liao *et al.*, 2011). For example, the SVM delivered a prediction accuracy of 93.33%. Thus, a further investigation of TS was conducted by including *w/c* and FC. Detailed results are provided in the next section.

5.2.2 Results of the Second TS-based Formula

The best values of the constant numbers (a, b, k_1 and k_2) were found to be 11.125, 5.500, 56.500 and 31.125, respectively. Six loops were needed in the current case. Detailed results of the TS are displayed in Table 5. The objective function time histories of the first and second loop are shown in Fig. 6. Like before, the best solution was found at the early iterations. Because the search domain is a four-dimension space, no illustration can be displayed here.

Based on the optimization results obtained from the TS, the proposed TS-based formula stated in Eq. (12) is updated as follows:

$$\text{if}\begin{cases} N_1 = \dfrac{(D_{15})_{soil}}{(d_{85})_{grout}} + 56.5\dfrac{w/c}{FC} > 3.125 \\ N_2 = \dfrac{(D_{10})_{soil}}{(d_{95})_{grout}} + 31.125\dfrac{w/c}{FC} > 1.500 \end{cases} N \text{ is groutable}$$
$$\text{else} \qquad N \text{ is non-groutable} \tag{15}$$

Table 5. Optimization Results of the Second TS-based Formula

Parameters	a	b	k_1	k_2
initial solution	9	6	25	25
Initial step size	2			
optimal solution	11.125	5.5	56.5	31.125
length of TS list	7			
accuracy in training	93.85%			
Accuracy in testing	95.56%			
Overall accuracy	94.17%			

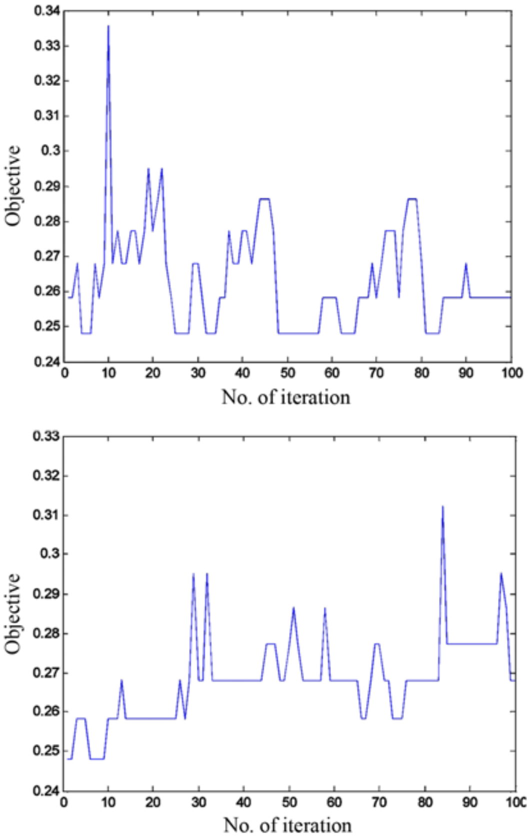

Fig. 6. Time History of the Objective Function (Upper: the first loop, Lower: the second loop)

Based on Eq. (15), the percentage correct for training and testing procedure was 93.85% and 95.56%, respectively (Table 5). The prediction accuracy was improved by introducing *w/c* and FC, as expected. Seven data points were incorrectly classified by the first TS-based formula but correctly predicted by the second TS-based formula (in the training stage). However, three data points were incorrectly classified by the second TS-based formula but correctly classified by the first TS-based formula. Although including two more factors increased the prediction accuracy, the data points from a real world data set were extremely complicated, and using a single rule to predict them perfectly was not an easy task.

Additionally, the obtained overall accuracy (94.17%, Table 5) was better than that of SVM (93.33%, Table 7) and close to that of RBFNN (94.97%, Liao *et al.,* 2011). It is reasonable to terminate the search for a better formula at this point.

5.3 Results of SVM & RBFNN

The input number for a SVM model has a significant impact on its performance. To investigate all possible factors for groutability studied here, 6 SVM models with different combinations of factors were investigated, as shown in Table 6. As mentioned in Section 3.2, to find a suitable pair of γ and C, cross-validation and grid search were needed for each SVM model. Fig. 7 provides an example of the prediction contours in the cross-validation and grid search (case 4). Based on these contours, the best values of γ and C were 5.1 and 14.1, respectively. The γ and C of each model and their resulting prediction accuracy are displayed in Table 7. It is seen that the overall accuracy increased as more factors were considered. Although the increasing rate was not significant, it is clear that case 6 dominated all other cases. This observation validated the results from the two TS-based formulas.

The prediction accuracy using RBFNN was 94.97%. The details can be found in Liao *et al*. (2011).

Table 6. Different SVM Models Considered

SVM models	Considered factors
Case 1	D_{10}
Case 2	D_{15}
Case 3	D_{10} & D_{15}
Case 4	D_{10}, D_{15} & *w/c*
Case 5	D_{10}, D_{15} & FC
Case 6	D_{10}, D_{15}, *w/c* & FC

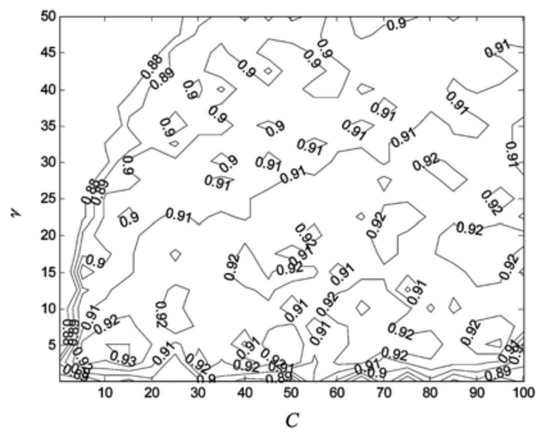

Fig. 7. The Contours of the Prediction Accuracy for Different Values of γ and C (case 4)

Table 7. γ, C and the Resulting Prediction Accuracies of Each SVM Model

SVM models	γ	C	accuracy (training)	accuracy (testing)	overall accuracy
Case 1	0.91	11.1	91.28%	91.11%	91.25%
Case 2	3.1	19.1	91.28%	91.11%	91.25%
Case 3	1.1	18.1	91.79%	88.89%	91.25%
Case 4	5.1	14.1	92.82%	88.89%	92.08%
Case 5	12.5	55.5	91.79%	93.33%	92.08%
Case 6	40.1	12.1	93.33%	93.33%	93.33%

5.4 Discussion

Compared to the results of SVM and RBFNN, the TS-based formulas could predict N with similar accuracy, especially the second proposed formula. Moreover, based on the results of the TS and SVM, it was confirmed that *w/c* and FC should be included in the prediction model. Some other issues are discussed further in this section.

As mentioned in Section 4.2, more than one separation is possible for the 240 collected samples. Here, an additional 9 separations were used to train and test the first TS-based formula. The results are displayed in Table 8 and Fig. 8. The mean values of prediction accuracy for the ten separations were 91.33%, 89.78% and 91.03% for training, testing and overall performance, respectively. Compared to the previous numbers (91.79%, 88.89% and 91.24%), which were calculated by a single separation (No.5 in Table 8), the previous conclusions for the TS approach were not biased by separating the collected data. To validate this conclusion, analysis of variance (ANOVA) was conducted on the training and testing data of these 10 separations. The hypothesis used in the ANOVA was that the means of each considered factor (D_{10}, D_{15}, *w/c* and FC) in the ten separation training data samples were the same. A similar hypothesis was applied to the testing case. The results are shown in Tables 9 and 10 for training and testing, respectively. Based on the results, the 10 training data samples were apparently not affected by the separations. However, the 10 testing data samples were affected by the separations. This outcome verified that the prediction variance in testing was larger than that of training, as displayed in Fig. 8 and Table 8. However, the coefficient of variance (cov) in testing was only 4.4%; thus, it is reasonable to conclude that the impact of separation on the prediction model (the TS-based formula) and accuracy was moderate.

The results of using these 10 separations for the second TS-

Table 8. The Prediction Accuracies of Different Separations (The first TS-based formula)

Separation No.	Training	Testing	Overall
1	89.74%	95.56%	90.83%
2	91.79%	86.67%	90.83%
3	90.77%	91.11%	90.83%
4	92.82%	84.44%	91.24%
5	91.79%	88.89%	91.24%
6	89.74%	97.78%	91.24%
7	91.28%	91.11%	91.24%
8	91.79%	86.67%	90.83%
9	91.79%	88.89%	91.27%
10	91.79%	86.67%	90.83%
Mean (cov)	91.33% (1%)	89.78% (4.4%)	91.03% (0.2%)

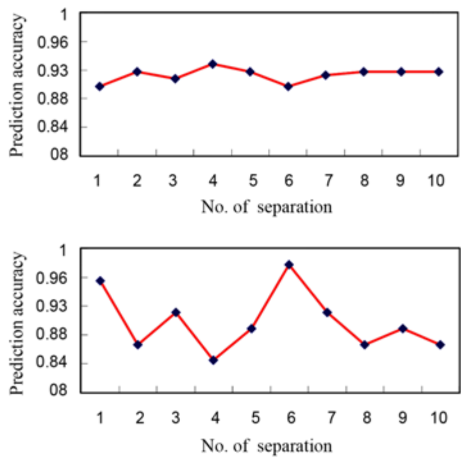

Fig. 8. The Variance in Prediction Accuracies of Ten Different Separations (Upper: training, Lower: testing, the first TS-based formula)

Table 9. The ANOVA for Data Samples in Training

D_{10}	SS	df	MS	F	Prob>F	*w/c*	SS	df	MS	F	Prob>F
Groups	1316.849	9	146.3166	0.33081	0.96507	Groups	0	9	0	1.15E-30	1.0
Error	858047.8269	1940	442.2927			Error	290.753	1940	0.14987		
Total	859364.6759	1949				Total	290.753	1949			
D_{15}	SS	df	MS	F	Prob>F	FC	SS	df	MS	F	Prob>F
Groups	1823.0024	9	202.5558	0.26198	0.9844	Groups	1440.59	9	160.065	0.18	0.9963
Error	1499957.0318	1940	773.1737			Error	1749739.95	1940	901.928		
Total	1501780.0342	1949				Total	1751180.53	1949			

*SS is Sum of Squares, df is degree of freedom, MS is Mean Squares

Table 10. The ANOVA for Data Samples in Testing

D_{10}	SS	df	MS	F	Prob>F	*w/c*	SS	df	MS	F	Prob>F
Groups	5706.3	9	634.038	1.63	0.1029	Groups	0	9	0	1.3E-31	1.0
Error	170651.4	440	387.844			Error	95.4028	440	0.21682		
Total	176357.8	449				Total	95.4028	449			
D_{15}	SS	df	MS	F	Prob>F	FC	SS	df	MS	F	Prob>F
Groups	7899.7	9				Groups	6242.5	9	693.615	0.89	0.6209
Error	300317.5	440				Error	383789.5	440	872.249		
Total	308217.2	449				Total	390032	449			

*SS is Sum of Squares, df is degree of freedom, MS is Mean Squares based formula were similar to those observed in the first TS-based formula, as shown in Table 11. The mean values of prediction accuracy for the ten separations were 94.36%, 94.00% and 94.29% for training, testing and overall performance, respectively. In comparison with the calculations by a single separation (No. 5 in Table 11), only the testing prediction accuracy was shown to have a moderate difference (94.00% vs. 95.56%); no significant difference was observed in the training and overall accuracies of the two analyses (94.36% vs. 93.85% and 94.29% vs. 94.17%). Thus, the impact of separation was not a critical issue when predicting N.

It is of interest to investigate which data samples were incorrectly predicted by the TS-based formulas and the SVM. The prediction accuracy was 91.79%, 93.85% and 91.79% for the first, second formulas and the SVM (case 3), respectively (for the training stage). The corresponding number of samples that were incorrectly classified was 16, 12 and 16 for the first, second formulas and the SVM, respectively. The ranges of incorrectly classified samples for these three approaches are provided in Table 12. It is seen that although the second TS-based formula had a better prediction accuracy, its ranges of incorrectly classified samples were larger than those of the other two approaches. This result indicates that the higher accuracy of the second formula was partially contributed from the *w/c* and FC. In addition, the ranges for the first TS-based formula and the SVM are exactly the same. These results indicate that the SVM and the TS provided consistent analyses and classification for the classification of N.

The support vectors, produced by the SVM, were points used to classify the raw data into two groups. These vectors included the points that were close to the classified boundary. Some of them may be on the wrong side of the boundary because the slack variable was introduced. Using D_{10} and D_{15} in training data

Table 11. The Prediction Accuracies for Different Separations (The second TS-based formula)

Separation No.	Training	Testing	Overall
1	94.36%	93.33%	94.17%
2	94.36%	93.33%	94.17%
3	93.85%	95.56%	94.17%
4	95.38%	88.89%	94.17%
5	93.85%	95.56%	94.17%
6	93.33%	97.78%	94.17%
7	93.33%	97.78%	94.17%
8	94.87%	93.33%	94.58%
9	94.87%	93.33%	94.58%
10	96.41%	91.11%	94.58%
Mean (cov)	94.36% (0.76%)	94.00% (2.78%)	94.29% (0.2%)

Table 12. The Ranges of Incorrectly Classified Samples for Three Different Approaches

	The first TS-based formula	The second TS-based formula	The SVM*
Ranges for D_{10}	6.3-13 (μm)	2.8-13 (μm)	6.3-13 (μm)
Ranges for D_{15}	10-27 (μm)	7.2-27 (μm)	10-27 (μm)

*case 3

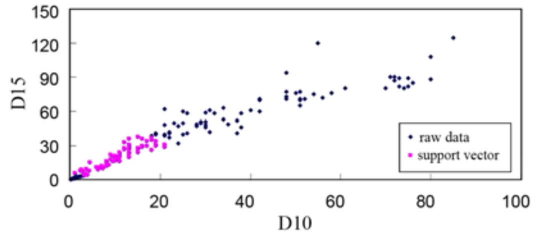

Fig. 9. The Support Vectors in SVM (Case 3)

samples as an example, there were 84 support vectors (D_{10} and D_{15}), as displayed in Fig. 9 (case 3). The ranges of D_{10} and D_{15} were 0.1-85 (μm) and 0.2-125 (μm), respectively, as shown in Table 1. The ranges of D_{10} and D_{15} in support vectors were 1-21 (μm) and 3.9-38 (μm), respectively. These ranges were the regions in which SVM may not be able to successfully predict N. Compared to the ranges of incorrectly classified samples by the TS-based formulas and the SVM, it is found that the support vectors included all data samples that were incorrectly classified. Thus, one can approximately estimate the N through the lower and upper bounds of the support vectors.

6. Conclusions

This study proposed a TS-based groutability prediction formula, in which the format of the existing empirical formula was adopted. A total of 240 in-situ data samples were collected and analyzed by both of the existing and TS-based formulas to demonstrate the improved classification of N. In addition, an SVM was used to validate the results of the proposed formula. The accuracies of the existing formulas ranged from 45% to 68%. However, the prediction accuracy of the TS-based formula was approximately 94.17%, which indicates that the TS-based formula is a good tool for predicting the groutability (N) of sandy silt soils for which microfine cement grouts were used in a permeation grouting.

Because the grain-size is relatively small and the FC is relatively high, many groutable data were judged as non-groutable using the existing empirical formulas. This result indicates that the empirical formulas are not suitable for microfine-cement grouts, and the threshold values in the formulas should be modified accordingly.

To modify the threshold values in the existing empirical formulas, this study reformulated the groutability problem as an optimization problem. TS was chosen as the optimizer. It was evident that TS was suitable for locating the best threshold values in the proposed groutability prediction formula. It was also found that including FC and *w/c* can increase the prediction accuracy. The modified formula delivered promising prediction results and can be readily implemented in practical engineering applications.

It should be noted that the proposed formula might be applicable only to sandy silt soils having similar grain-size ratio between soils and grouts, FC and *w/c*. More investigation is needed when applying this formula to other conditions (e.g., other countries).

Acknowledgements

This study was supported by the National Science Council of Taiwan under grant number NSC 99-2221-E-011-037. The support is gratefully acknowledged.

References

Akbulut, S. and Saglamer, A. (2002). "Estimating the groutability of granular soils: A new approach." *Tunn. Undergr. Space Technol.*, Vol. 17, No. 4, pp. 371-380.

Axelsson, M., Gustafson, G., and Fransson, A. (2009). "Stop mechanism for cementitious grouts at different water-to-cement ratio." *Tunn. Undergr. Space Technol.*, Vol. 24, No. 4, pp. 390-397.

Bahrami, A., Monjezi, M., Goshtasbi, K., and Ghazvinian, A. (2011). "Prediction of rock fragmentation due to blasting using artificial neural network." *En. Comput.*, Vol. 27, No. 2, pp. 177-181.

Boser, B., Guyon, I., and Vapnik, V. (1992). "An training algorithm for optimal margin classifiers." *Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh, ACM, pp. 144-152.

Burwell, E. B. (1985). "Cement and clay grouting of foundations: Practice of the corps of engineering." *J. Soil Mech. Found Div.*, Vol. 84, No. 1, pp. 1551/1-1551/22.

Cho, S. E. (2009). "Probabilistic stability analyses of slopes using the ANN-based response surface." *Comput. Geotech.*, Vol. 36, No. 5, pp. 787-797.

Chun, B. S., Lee, Y. J., and Chung, H. I. (2006). "Effectiveness of leakage control after application of permeation grouting to earth fill dam." *KSCE J. Civ. Eng.*, Vol. 10, No. 6, pp. 405-414.

Das, S. K. and Basudhar, P. K. (2006). "Undrained lateral load capacity of piles in clay using artificial neural network." *Comput. Geotech.*, Vol. 33, No. 8, pp. 454-459.

Glover, F. (1986). "Future paths for integer programming and links to artificial intelligence." *Comput. Oper. Res.*, Vol. 13, No. 5, pp. 533-549.

Hillier, F. S. and Lieberman, G. J. (2005). *Introduction to operations research*, McGraw-Hill, New York.

Huang, C. L., Fan, J. C., and Yang, W. J. (2007). "A study of applying microfine cement grout to sandy silt soil." *Sino-Geotech.*, Vol. 111, No. 1, pp. 71-82.

Incecik, M. and Ceren, I. (1995). "Cement grouting model tests." *Bull. Tech. Univ. Istanbul*, Vol. 48, No. 2, pp. 305-317.

Khandelwal, M. (2011). "Blast-induced ground vibration prediction using support vector machine." *En. Comput.*, Vol. 27, No. 3, pp. 193-200.

Khayat, K. H., Ballivy, G., and Gaudreault, M. (1997). "High-performance cement grout for underwater crack injection." *Can. J. Civil Eng.*, Vol. 24, No. 3, pp. 405-418.

Krizek, R. J., Liao, H. J., and Borden, R. H. (1992). "Mechanical properties of microfine cement/sodium silicate grouted sand." *Soil Improv. Geosynth.,* ASCE Special Technical Publication on Grouting, pp. 688-699.

Kuo, Y. L., Jaksa, M. B., Lyamin, A. V., and Kaggwa, W. S. (2009). "ANN-based model for predicting the bearing capacity of strip footing on multi-layered cohesive soil." *Comput. Geotech.,* Vol. 36, No. 3, pp. 503-516.

Lee, J. J. and Kim, Y. S. (2010). "Development of advanced pattern recognition model for evaluation of lateral displacement on soft ground using support vector machine." *KSCE J. Civ. Eng.,* Vol. 14, No. 2, pp. 173-182.

Liao, K. W., Fan, J. C., and Huang, C. L. (2011). "An artificial neural network for groutability prediction of permeation grouting with microfine cement grouts." *Comput. Geotech.,* Vol. 38, No. 8, pp. 378-386.

Mita, A. and Hagiwara, H. (2003). "Quantitative damage diagnosis of shear structures using support vector machine." *KSCE J. Civ. Eng.,* Vol. 7, No. 6, pp. 683~689.

Ozgurel, H. G. and Vipulanandan, C. (2005). "Effect of grain size and distribution on permeability and mechanical behavior of acrylamide grouted sand." *J. Geotech. Geoenviron. Eng.,* Vol. 131, No. 12, pp. 1457-1465.

Perret, S., Khayat, K. H., Gagnon, E., and Rhazi, J. (2002). "Repair of 130-year old masonry bridge using high-performance cement grout." *J. Bridge Eng.,* Vol. 7, No. 1, pp. 31-38.

Schwarz, L. G. and Krizek, R. J. (2006). "Hydrocarbon residuals and containment in microfine cement grouted sand." *J. Mater. Civil Eng.,* Vol. 18, No. 2, pp. 214-228.

Tekin, E. and Akbas, S. O. (2011). "Artificial neural networks approach for estimating the groutability of granular soils with cement-basis grouts." *Bull. Eng. Geol. Environ.,* Vol. 70, No. 1, pp. 153-161.

Tung, C. P. and Chou, C. A. (2002). "Application of tabu search to groundwater parameter zonation." *J. Am. Water Resour.,* Assoc. Vol. 38, No. 4, pp.1115-1126.

Zebovitz, S., Krizek, R. J., and Atmatzidis, D. K. (1989). "Injection of fine sands with very fine cement grout." *J. Geotech. Eng. Div.,* Vol. 115, No. 12, pp. 1717-1733.