The 2025 World Congress on
Advances in Structural Engineering and Mechanics (ASEM25)
BEXCO, Busan, Korea, August 11-14, 2025

Data-driven equation for in-situ undrained shear strength of normally consolidated clayey soils

*Yerim Yang[1], Hangseok Choi[2], and Kibeom Kwon[3]

[1] *Department of Civil, Environmental and Architectural Engineering, Korea University, Seoul 02841, Korea*
[2] *School of Civil, Environmental and Architectural Engineering, Korea University, Seoul 02841, Korea*
[3] *Future and Fusion Lab of Architectural, Civil and Environmental Engineering, Korea University, Seoul 02841, Korea*
[1] purunyr@korea.ac.kr

ABSTRACT

With the growing number of tunneling projects near marine environments, accurately predicting the undrained shear strength (S_u) of soft clay-containing soils has become critically important. This study employed symbolic regression to develop a predictive equation for in-situ S_u, utilizing a global database. The derived equation effectively captures the relationship between S_u and key soil parameters, achieving a low mean absolute error and a high coefficient of determination between measured and predicted values. These results highlight the potential of symbolic regression as a powerful tool for predicting the strength parameters of normally consolidated (NC) clayey soils.

1. INTRODUCTION

Tunneling in soft ground involves considerable geotechnical challenges, particularly when excavation is conducted below the groundwater level. Under such conditions, clayey soils are typically saturated and behave in an undrained manner, resulting in the undrained shear strength (S_u) as a key parameter for short-term stability (Chongzhi et al. 2021). Therefore, accurate prediction of S_u is essential for evaluating tunnel face stability, designing pressure-controlled tunneling systems such as earth pressure balance (EPB) or slurry shield TBMs, and preventing ground collapse or excessive deformation. In this context, this study proposes a predictive equation for in-situ S_u (denoted as $S_{u(mob)}$) of normally consolidated (NC) clayey soils based on a symbolic regression approach, aiming to capture the underlying relationships between $S_{u(mob)}$ and key soil parameters in a data-driven yet interpretable manner.

---

[1] Ph. D. Student
[2] Professor
[3] Researcher

2. MODEL DEVELOPMENT

2.1 Data preparation

This study utilized the CLAY/10/7490 database, a global dataset containing 7,490 instances compiled from 251 published studies (Ching and Phoon, 2012). The database includes ten soil parameters, encompassing three index properties (Atterberg limits and plasticity index), four mechanical properties (effective stresses, shear strength, and sensitivity), and three parameters obtained from piezocone penetration tests (CPTu).

To improve the simplicity and practical applicability of the proposed equation, two input features were selected based on a review of existing predictive equations in the literature (Table 1): plasticity index (PI) and vertical effective stress (σ'_v). In this study, soils with an overconsolidation ratio (OCR) ranging from 1.0 to 1.3 were classified as NC clayey soils. Sensitivity (S_t) was excluded from the model development due to limited data availability. Consequently, a dataset of 399 instances containing $S_{u(mob)}$, PI, and σ'_v was prepared for model training and evaluation. Fig.1 presents the correlation plots and corresponding Spearman correlation coefficients among these three variables.

Table 1 Existing equations for predicting S_u

Literature	Equation
Skempton (1957)	$\boldsymbol{S_u} = (0.11+0.0037\cdot\mathrm{PI})\boldsymbol{\sigma'_v}$
Mesri (1975, 1989)	$\boldsymbol{S_{u(mob)}} = 0.22\cdot\boldsymbol{\sigma'_p}$
Jamiolkowski et al. (1985)	$\boldsymbol{S_{u(mob)}} = (0.23\cdot\mathbf{OCR^{0.8}})\boldsymbol{\sigma'_v}$
Ching and Phoon (2012)	$\boldsymbol{S_{u(mob)}} = (0.229\cdot\mathbf{OCR^{0.823}}\cdot\boldsymbol{S_t^{0.121}})\boldsymbol{\sigma'_v}$

Note: σ'_p (= OCR·σ'_v); preconsolidation stress

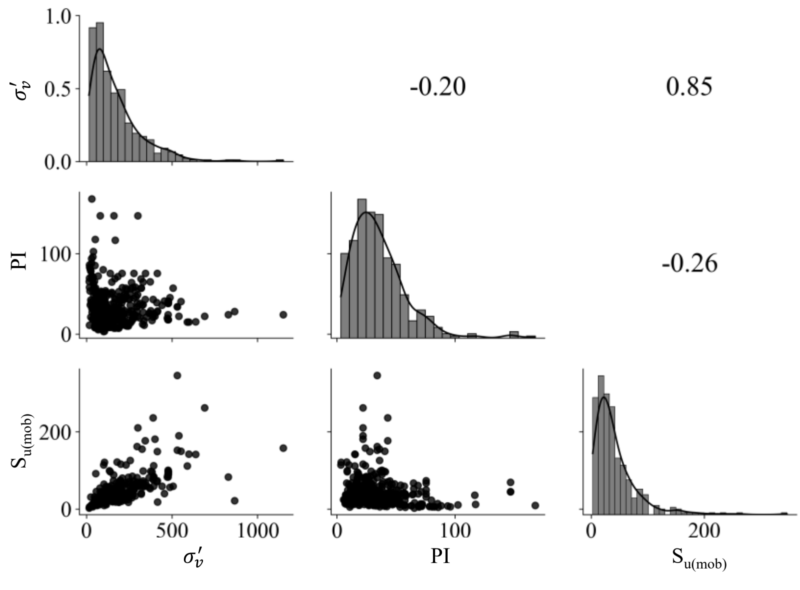


Fig. 1 Pairwise correlation plots with Spearman correlation coefficients

2.2 Symbolic regression technique

Symbolic regression is an emerging data-driven approach that aims to discover interpretable mathematical expressions capable of capturing complex relationships between input and output variables. This study employed PySR, a high-performance symbolic regression framework that improves search efficiency through simulated annealing, expression simplification, and complexity control. Further details can be found in Cranmer (2023).

2.3 Model implementation

The dataset of 399 instances was divided into a training set (319 instances) and a test set (80 instances) with an 80:20 ratio. To address skewness in the target variable ($S_{u(mob)}$), a logarithmic transformation was applied. Subsequently, PySR was employed to derive the optimal predictive equation from the training set, with hyperparameters tuned within a predefined search space. To ensure model simplicity and interpretability, the complexity of the equation was limited to 15 components. The resulting predictive equation was evaluated on both the training and test sets using four performance metrics: mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2).

3. PROPOSED EQUATION

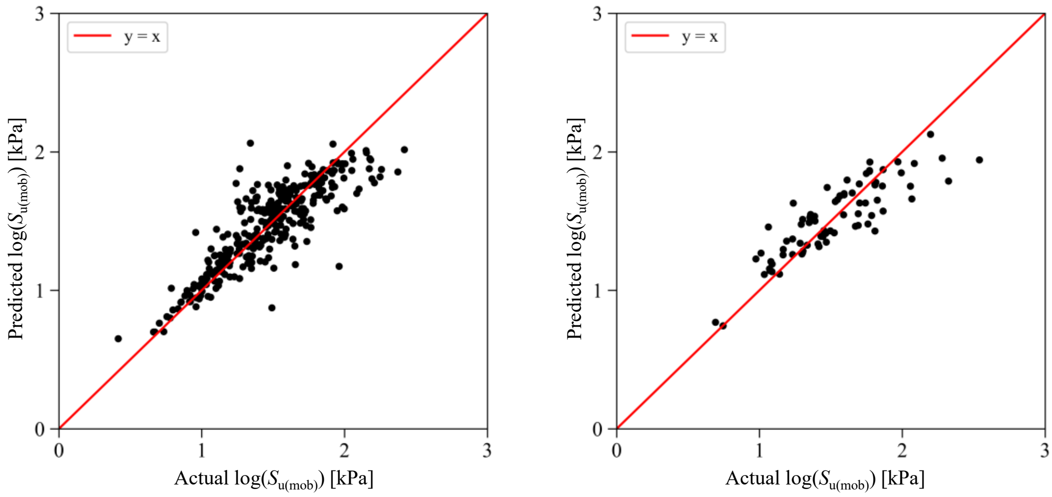
The proposed symbolic regression-based equation for predicting $S_{u(mob)}$ of NC clayey soils is presented in Eq. (1):

$$\log\left(S_{\mathrm{u(mob)}}\right) = 2.593 - \frac{19.049}{\sqrt{\sigma'_{\mathrm{v}} + 6.010}} + \frac{5.953}{\mathrm{PI}^2} \ , \qquad (1)$$

The derived equation indicates that $S_{u(mob)}$ is positively correlated with σ'_v while negatively correlated with PI, which is consistent with the Spearman correlation coefficients shown in Fig. 1. The overall performance metrics, summarized in Table 2, confirm the strong predictive capability of the proposed equation. Furthermore, comparison plots between the actual and predicted values are provided in Fig. 2.

Table 2 Predictive performance for the proposed equation

	MAE [kPa]	MAPE [%]	RMSE [kPa]	R^2
Training	0.125	8.662	0.176	0.753
Test	0.139	9.144	0.183	0.751



(a) Training (b) Test

Fig. 2 Comparison plots for the proposed equation

4. CONCLUSIONS

This study proposed a regression equation for estimating the in-situ undrained shear strength ($S_{u(mob)}$) of normally consolidated (NC) clayey soils. This equation was derived using a symbolic regression technique applied to a comprehensive global database. The key findings are as follows:

1) Data analysis revealed that $S_{u(mob)}$ exhibits a positive correlation with vertical effective stress (σ'_v) and a negative correlation with the plasticity index (PI).
2) Based on these correlations, an optimal symbolic equation was formulated with two independent variables, σ'_v and PI, expressed as follows:

$$\log\left(S_{\mathrm{u(mob)}}\right) = 2.593 - \frac{19.049}{\sqrt{\sigma'_{\mathrm{v}}} + 6.010} + \frac{5.953}{\mathrm{PI}^2}$$

3) The proposed equation exhibited reliable predictive performance, with a low mean absolute error (MAE) below 0.140 kPa and a high coefficient of determination (R^2) exceeding 0.750 for both the training and test sets.
4) The robust performance of the proposed model highlights the potential of symbolic regression as a powerful and interpretable approach for predicting critical geotechnical parameters in tunneling projects.

ACKNOWLEDGEMENT

This research was supported by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure and Transport (RS-2024-00410248 and RS-2022-00144188).

REFERENCES

Ching, J. and Phoon, K. K. (2012), “Modeling parameters of structured clays as a multivariate normal distribution,” *Can. Geotech. J.*, **49**(5), 522-545.

Chongzhi, W., Lin, W. and Zhang, W. (2021), “Assessment of undrained shear strength using ensemble learning based on Bayesian hyperparameter optimization,” *Modeling in Geotechnical Engineering, Academic Press*, 309-326.

Cranmer, M. (2023), “Interpretable machine learning for science with PySR and SymbolicRegression. jl,’ *arXiv preprint arXiv:2305.01582*.

Jamiolkowski, M. (1985), “New developments in field and laboratory testing or soils,” *Proc. 11th. Int. Conf. on SMFE., San Francisco, CA,* 1, 57-153.

Mesri, G. (1975), “Discussion on “New design procedure for stability of soft clays”,” *J. Geotech. Eng.*, *ASCE*, **101**(4), 409-412.

Mesri, G. (1989), “A re-evaluation of $S_{u(mob)} = 0.22\ \sigma_p$’ using laboratory shear tests,” *Can. Geotech. J.*, **26**(1), 162164.

Skempton, A. W. (1957), “Discussion: Further data on the c/p ratio in normally consolidated clays,” *Proc. Inst. Civ. Eng.*, **7**, 305-307.