HOKKAIDO UNIVERSITY

Title	Effect of Fold Number K in Repeated K - Fold Cross - Validation on Deep - Learning - Based Consolidation Settlement Prediction
Author(s)	Oda, Kazuhiro
Relation	Proceedings of the 10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026), June 28 - July 1, 2026, Sapporo, Japan
Citation	Proceedings of the 10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026)
Issue Date	2026-07-01
Doc URL	https://hdl.handle.net/2115/100463
Type	conference paper
File Information	C000173_1.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026),
28 June-01 July 2026, Sapporo, Japan
Kitahara, T., Miyamori, Y. & Beer, M. (editors)

Effect of Fold Number K in Repeated K-Fold Cross-Validation on Deep-Learning-Based Consolidation Settlement Prediction

Kazuhiro Oda[1, *]

[1]*Department of Architecture and Environment Design, Osaka Sangyo University, Nakagaito 3-1-1 Daito, Japan.*
E-mail: oda@aed.osaka-sandai.ac.jp
**Corresponding Author*

Predicting consolidation settlement in large-scale land reclamation remains difficult because oedometer data are spatially sparse and natural marine clays exhibit strongly non-linear structured behavior. This study examines the influence of the fold number K in repeated K-fold cross-validation on uncertainty quantification and consolidation settlement prediction using a deep neural network framework. The framework was trained using 49 boreholes and 634 oedometer-test samples from Kobe Airport, Japan. The repeated K-fold partitioning was performed at the oedometer-test-sample level, and model uncertainty, σ_{model}, and predictive uncertainty for observed oedometer data, σ_{pred}, were evaluated separately. Independent blind tests at two boreholes excluded from training showed that more than 95% of the laboratory pressure-void ratio data points were included in the 95% prediction interval. The uncertainty of the predicted e–log p' relationships was then propagated to a coupled soil–water FEM analysis at monitoring point KC-1 through Monte Carlo simulation. Within the investigated range of K = 3–10, no clear dependence of the predicted void ratio uncertainty or final settlement uncertainty on K was observed. The simulated settlement reproduced the early-to-mid-term field monitoring data for the modeled Holocene clay layer, while later discrepancies were attributed to settlement of the underlying Pleistocene strata. These results suggest that moderate K values, such as K = 5, can be a practical compromise between computational cost and uncertainty evaluation for the present dataset and modeling conditions.

Keywords: Settlement, deep learning, consolidation, prediction, cross-validation.

1 Introduction

Accurate prediction of consolidation settlement is a central issue in large-scale land reclamation projects, as even minor discrepancies between predicted and actual settlement can significantly impact construction management and project costs. However, predicting consolidation settlement remains difficult because laboratory data are spatially sparse, and natural marine clays exhibit strongly non-linear structured behavior. Natural clays commonly exhibit apparent yield stresses greater than the in-situ effective overburden stress and a marked increase in compressibility post-yield. Such behavior is associated with the structure of natural clays, including aging, rate effects, and destructuration during compression, and has been widely discussed in previous studies on natural clay compressibility, (Burland 1990, Leroueil and Vaughan 1990, Hong et al. 2012). Therefore, accurate settlement prediction requires reconstructing the full non-linear compression response rather than relying on idealized bi-linear models.

While deep neural networks (DNNs) have emerged as powerful tools for predicting complex soil properties, predicting the non-linear e-log p' relationship heavily depends on the training data, necessitating cross-validation to ensure generalization performance. In this context, K-fold cross-validation is widely applied. However, the influence of hyperparameter selection, specifically the number of folds (K), on the predictive uncertainty and the ultimate settlement estimations has not been sufficiently investigated. Previous ANN-based studies have demonstrated the applicability of data-driven spatial interpolation and stochastic settlement estimation for Kobe Airport clays; however, the influence of the fold number K on uncertainty propagation to settlement has not been sufficiently examined (Oda 2013, Oda 2016). Therefore, the objective of this study is to evaluate how the fold number K in repeated K-fold cross-validation affects the uncertainty of predicted e–log p' relationships and the final settlement estimated by coupled consolidation analysis. The study specifically distinguishes model uncertainty, σ_{model}, from predictive uncertainty for observed oedometer data, σ_{pred}, and examines the sensitivity of these uncertainty measures for K = 3–10.

2 Probabilistic Data-Driven Framework for Consolidation Modeling

2.1 *DNN Architecture and Strategy*

To model the complex non-linear consolidation behavior, a fully connected Multi-Layer Perceptron (MLP) architecture was employed. Figure 1 shows the MLP architecture used in this study. The input variables were latitude, longitude, depth, and consolidation pressure p', and the output variable was the void ratio e. The network was designed with 16 hidden layers, each containing 250 neurons, to provide sufficient representational capacity for mapping complex spatial heterogeneity. The Rectified Linear Unit (ReLU) activation function was adopted to mitigate the vanishing gradient

problem. To prevent the high-capacity model from overfitting, an "Early Stopping" regularization strategy was implemented.

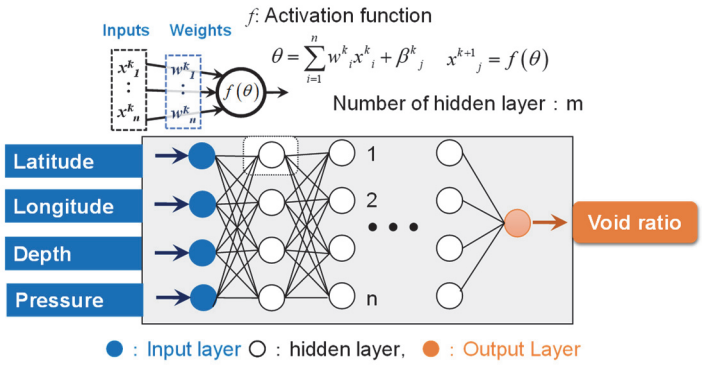

Figure 1. MLP architecture for predicting void ratio from spatial coordinates and consolidation pressure.

2.2 *Formulation of Predictive Uncertainty*

To quantify the epistemic uncertainty arising from limited data, Repeated K-fold cross-validation was introduced, (Nadeau 2003, Kim 2009). Figure 2 shows the conceptual view of K-fold cross-validation. In a standard K-fold process, the dataset is divided into K folds, where $K-1$ folds are used for training and the remaining fold is used for validation. In the Repeated K-fold approach, this process is repeated R times, set to 20 in this study, with randomly redefined partitions, yielding a total of $K \times R$ predictive models.

The repeated K-fold partitioning was performed at the oedometer-test-sample level. Therefore, all pressure–void-ratio data points obtained from the same oedometer specimen were assigned to the same fold.

The predictions obtained from the $K \times R$ models are averaged to obtain the ensemble-mean e–log p' relationship at a target location. In the figures, this ensemble mean is denoted by μ. This study uses two uncertainty measures. The first is the model uncertainty, σ_{model}, which represents the variability of the predicted e–log p' relationship caused by repeated data partitioning and model training:

$$\sigma_{model} = \sqrt{\sigma^2_{repeat} + \sigma^2_{fold}} \tag{1}$$

where σ_{repeat} represents the variability among repeat-wise average predictions, whereas σ_{fold} represents the variability among folds within each repeat. The resulting σ_{model} is evaluated at each location and consolidation pressure and represents the uncertainty associated with model construction and data partitioning. This uncertainty is used later as the input uncertainty for the Monte Carlo settlement analysis.

The second is the predictive uncertainty, σ_{pred}, which additionally includes the out-of-fold residual scatter, σ_{oof}:

$$\sigma_{pred} = \sqrt{\sigma^2_{model} + \sigma^2_{oof}} \tag{2}$$

The term σ_{oof} is estimated from out-of-fold residuals and includes observation scatter, measurement error, and unexplained variability in laboratory test data. Therefore, σ_{pred} is used to construct prediction intervals for observed oedometer data in the blind-test validation. It is not used directly as material uncertainty in the FEM settlement analysis, because it includes residual scatter of laboratory observations in addition to model uncertainty.

This distinction allows the uncertainty used for blind-test validation and that used for settlement analysis to be interpreted separately.

Fold 1	Train	Train	Train	Train	Validation
Fold 2	Train	Train	Train	Validation	Train
Fold 3	Train	Train	Validation	Train	Train
Fold 4	Train	Validation	Train	Train	Train
Fold 5	Validation	Train	Train	Train	Train

Figure 2. Conceptual view of K-fold cross-validation.

3 Site Characterization and Datasets

The study site is Kobe Airport (Figure 3), a large-scale reclaimed artificial island constructed on a soft marine clay deposit in Osaka Bay, Japan. The seabed consists of a thick Holocene clay layer overlying a Pleistocene clay formation. The initial void ratio decreases from approximately 3.0 near the seabed surface to about 1.3 at a depth of 30 m.

A comprehensive geotechnical database was compiled from site investigation reports, utilizing 49 boreholes, shown by blue solid circles, for model training. These boreholes provided 634 oedometer-test samples for training. Each sample consisted of multiple pressure–void-ratio data points. In the repeated K-fold procedure, the data were partitioned at the oedometer-test-sample level, whereas the validation coverage was evaluated using the individual pressure-void ratio data points. Two boreholes, Point A and Point B, were explicitly excluded from the training dataset and used as independent blind-test locations. The data from these boreholes were used as independent blind tests to evaluate the model's ability to predict soil properties at unmeasured locations. Point A and Point B contain 14 and 13 oedometer-test samples, respectively, covering depths up to 30 m. KC-1, a location with long-term settlement records but no site-specific borehole investigation data, was used to evaluate the practical applicability of the proposed methodology.

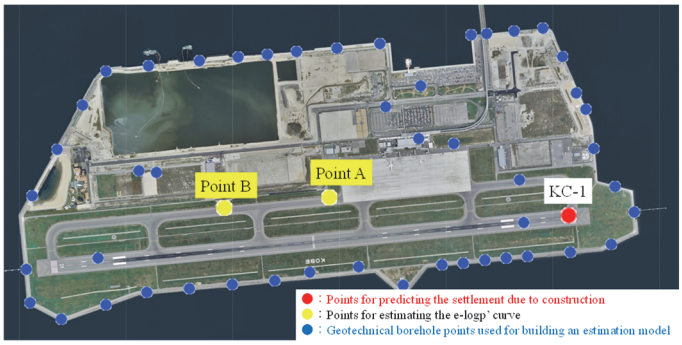

Figure 3. Kobe Airport site and dataset allocation: training boreholes, blind-test boreholes Points A and B, and settlement monitoring point KC-1.

4 Model Validation Using Independent Blind Tests

Figure 4 shows representative comparisons between the predicted e–log p' relationships and the oedometer test results for K = 5. The 95% bands were calculated as $\mu \pm 1.96\sigma_{pred}$ and are therefore interpreted as prediction intervals for observed oedometer data, not as confidence intervals of the ensemble mean. The ensemble-mean curves reproduced the overall trend of the experimental data in the representative blind-test examples. To quantitatively validate the model, independent blind tests were evaluated using data from two locations excluded from model training: Borehole Point A (14 samples) and Point B (13 samples), yielding a total of 217 pressure–void-ratio data points. Table 1 summarizes the empirical coverage of the laboratory pressure-void ratio data points within the 95% prediction interval derived from σ_{pred} in Eq. (2). Because σ_{pred} includes out-of-fold residual scatter, this interval is intended to evaluate the reproducibility of observed oedometer data. The empirical coverage exceeded 95% for all K values, indicating that the prediction interval was slightly conservative for the blind-test data. The empirical coverage exceeded 95% for all K values, indicating that the prediction interval was slightly conservative for the blind-test data. This result supports the use of σ_{pred} for evaluating the scatter of observed oedometer-test data.. However, the coverage probability alone does not fully characterize prediction accuracy; it should be interpreted together with point-wise prediction errors and interval width.

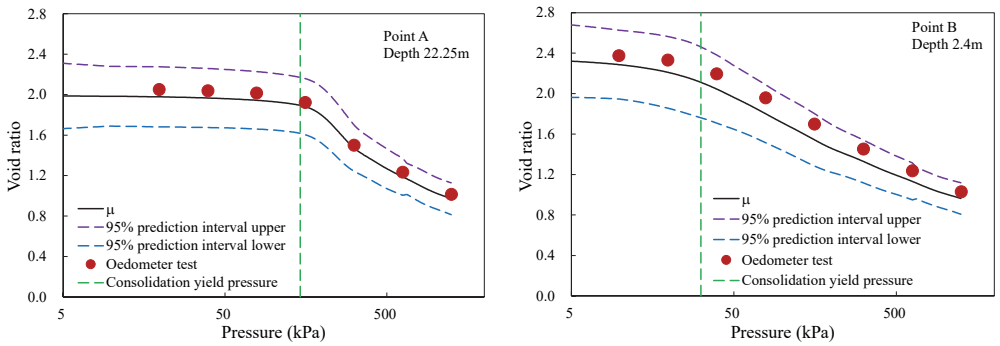

Figure 4. Representative blind-test comparisons between predicted e–log p' relationships and oedometer test results for K = 5: Point A at 22.25 m and Point B at 2.4 m.

Table 1. Empirical coverage of blind-test oedometer data within the 95% prediction interval.

K	3	4	5	6	7	8	9	10
Included / total	212/217	212/217	211/217	211/217	215/217	214/217	211/217	211/217
Coverage	0.977	0.977	0.972	0.972	0.990	0.986	0.972	0.972

5 Influence of Fold Number K on Settlement Prediction

5.1 *Reconstruction of e-log p′ Relationships at Uninvestigated Locations*

To assess engineering applicability, an axisymmetric coupled soil–water FEM analysis was conducted for monitoring point KC-1, taking into account the non-penetrating sand drain ground improvement. To represent radial drainage toward the sand drains, the model was constructed as an axisymmetric domain with the sand drain located along the central axis. Because the sand drains did not reach the lower drainage layer, the non-penetrating part was treated as an undrained boundary. The analytical model shown in Figure 5 was divided into 16 vertical and 9 horizontal divisions, resulting in 144 finite elements. In addition to the e–log p' relationships, the log k–e relationships required for pore-water flow were also predicted using the same repeated K-fold framework and assigned to the FEM model. Detailed results for the permeability relationship are omitted here because the present paper focuses on the influence of K on compression-related uncertainty and final settlement. For each of the 100 Monte Carlo simulations, the e–log p' relationship at the center depth of each layer was generated by adding normally distributed random fluctuations based on σ_{model} to the ensemble-mean prediction. The same sampling procedure was applied to the log k–e relationship when assigning permeability-related parameters. The uncertainty σ_{pred} was not used as the FEM input because it includes out-of-fold residual scatter associated with laboratory observations.

The predictive model reproduced the main characteristics of structured clay at this location, notably the increase in consolidation yield stress and the subsequent rise in compressibility in the 10th and 11th layers (Figure 6). This behavior is consistent with the reported consolidation characteristics of the Holocene clay layer at Kobe Airport (Hasegawa 2007). However, the evaluated model uncertainty σ_{model} varied by depth (Figure 7). For K = 5, the standard deviation was relatively large. in the 2nd layer due to scattered consolidation test results near the surface, and large in the 11th and 12th layers, reflecting the difficulty of predicting structured clay behaviors with sparse data points.

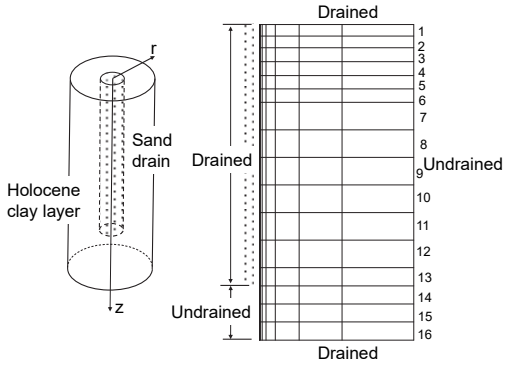

Figure 5. Axisymmetric FEM model for the non-penetrating sand-drain improved ground at KC-1.

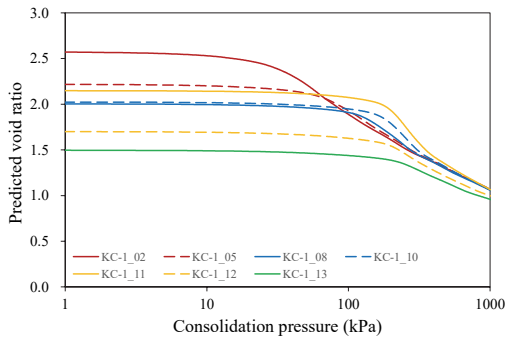

Figure 6. Representative predicted e–log p' relationships at KC-1.

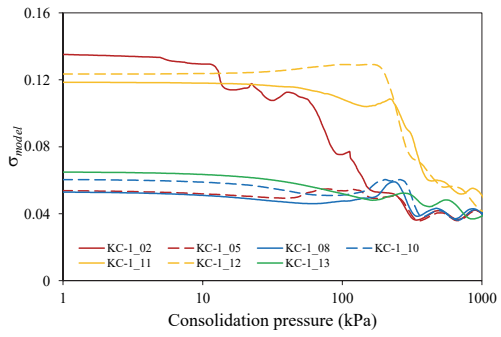

Figure 7. Variation of σ_{model} with consolidation pressure for K = 5.

5.2 *Sensitivity of the Fold Number K on Predicted Void-Ratio Uncertainty*

To investigate the sensitivity to the fold number, K was varied from 3 to 10. Figure 8 compares the relationship between σ_{model} and consolidation pressure for Layer 11, where the predicted uncertainty was relatively large. Although σ_{model} varied with consolidation pressure, the curves for different K values showed no systematic ordering. Therefore, within the investigated range of K = 3–10, no clear dependence of the predicted void-ratio uncertainty on K was observed.

This result should be interpreted with caution. Because the number of repetitions was fixed at R = 20, the total number of trained models increased from 60 to 200 as K increased from 3 to 10. The present comparison therefore reflects the combined effects of fold number, training-data fraction, validation-data fraction, and ensemble size. A larger number of repetitions would be required to detect more subtle K-dependent effects.

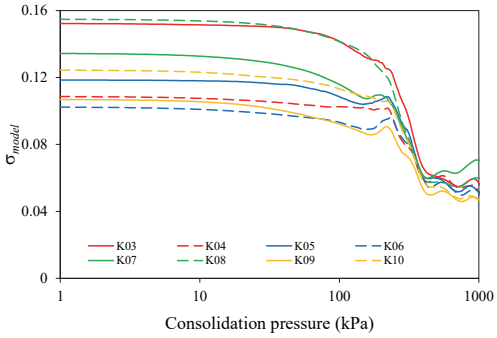

Figure 8. Effect of fold number K on the relationship between σ_{model} and consolidation pressure in Layer 11.

5.3 *Simulated Settlement and Field Monitoring Comparison*

Figure 9 shows the time history of settlement induced by land reclamation for K = 5. Up until approximately mid-to-late 2003, the simulated results were in good agreement with the field measurements. Beyond this point, however, the measured settlement begins to exceed the simulated values. The field monitoring data include not only the settlement of the Holocene clay layer Ma13, which is the target layer of the present analysis, but also the settlement of the underlying Pleistocene strata, including Ma12 and deeper layers. Therefore, the later-stage discrepancy is interpreted as the increasing contribution of these deeper deposits to the measured surface settlement. This interpretation is consistent with previous observations at Kobe Airport and other reclaimed islands in Osaka Bay (Hasegawa 2006, Furudoi 2010).

Figure 10 shows the effect of the fold number K on the mean and standard deviation of the predicted final settlement of the modeled Holocene clay layer. The mean final settlement ranged from 712.2 to 724.4 cm over K = 3–10, and the standard deviation ranged from 11.8 to 21.3 cm. The overall magnitude of the final settlement was approximately 718.8 cm. Within K = 3–10, no systematic trend was observed in either the mean settlement or its standard deviation. The corresponding 95% uncertainty interval width was approximately 61.07 cm, corresponding to about 8.5% of the mean final settlement.

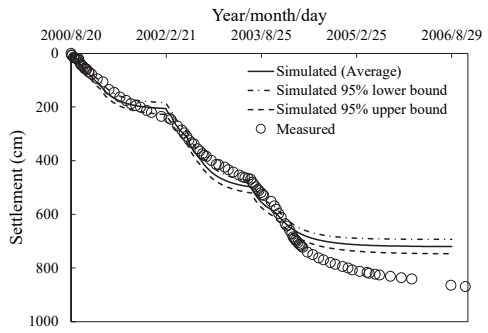

Figure 9. Time history of settlement induced by land reclamation for K = 5.

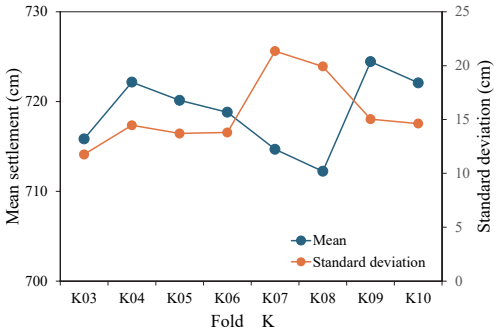

Figure 10. Effect of fold number K on the mean and standard deviation of predicted final settlement.

6 Discussions

6.1 *Observed Insensitivity to the Fold Number K*

The results in Figures 8 and 10 indicate that no clear dependence on the fold number K was observed for either σ_{model} of the predicted e–log p' relationship or the standard deviation of the final settlement. This suggests that, for the present

dataset and modeling conditions, the repeated K-fold ensemble was relatively insensitive to the choice of K within the range of 3–10. However, this result should not be interpreted as a general proof that the method is independent of K. Because R was fixed at 20, the ensemble size changed with K, and the training and validation data fractions also varied. Therefore, the observed behavior reflects the combined effects of fold number, ensemble size, and data partitioning. Additional analyses with a larger R and a fixed total number of trained models would be useful for isolating the effect of K itself.

6.2 *Practical Implications for Engineering Design*

From an engineering perspective, the results suggest that a moderate fold number such as $K = 5$ may be a practical choice for the present dataset because it reduces the computational burden while producing uncertainty estimates comparable to those obtained with larger K values. This implication is particularly relevant when the settlement analysis requires many Monte Carlo simulations coupled with FEM. Nevertheless, $K = 5$ should not be regarded as a universal recommendation. The appropriate value of K may depend on the number of oedometer-test samples, spatial data density, soil heterogeneity, network architecture, and the number of repetitions R. In addition, only 100 Monte Carlo simulations were conducted in this study; therefore, small differences in the estimated settlement standard deviation should be interpreted with caution. Future studies should increase both R and the number of Monte Carlo simulations and should include the underlying Pleistocene layers for long-term settlement prediction.

7 Conclusions

This study examined the influence of the fold number K in repeated K-fold cross-validation on the uncertainty of predicted e–log p' relationships and consolidation settlement. The following conclusions were obtained:

1. The DNN framework with repeated K-fold cross-validation reconstructed spatially continuous e–log p' relationships of structured marine clay from sparse oedometer-test data.
2. The repeated K-fold partitioning was performed at the oedometer-test-sample level, and model uncertainty σ_{model} and predictive uncertainty σ_{pred} were evaluated separately. Independent blind tests showed that more than 95% of the laboratory pressure-void ratio data points were included in the 95% prediction interval derived from σ_{pred}.
3. Within the investigated range of $K = 3$–10, no clear dependence of the predicted void-ratio uncertainty or final settlement standard deviation on K was observed. This suggests that $K = 5$ may be a practical compromise between computational cost and uncertainty evaluation for the present dataset and modeling conditions.
4. The coupled FEM settlement analysis reproduced the early-to-mid-term field monitoring data for the modeled Holocene clay layer. The later-stage discrepancy was attributed to settlement of the underlying Pleistocene strata, which were outside the primary modeling scope of this study.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 24K07992.

References

Burland, J. B., On the compressibility and shear strength of natural clays. Géotechnique, 40(3), 329–378, 1990.
Furudoi, T., The Second Phase Construction of Kansai International Airport Considering the Large and Long-Term Settlement of the Clay Deposits, *Soils and Foundations*, 50(6), 805-816, 2010.
Hasegawa, N., Matsui, T., Tanaka, Y., Takahashi, Y., and Nambu, M., In-situ consolidation behavior of Pleistocene clay below seabed at Kobe Airport, *Journal of the Japan Society of Civil Engineers C*, 62(4), pp. 780-792, 2006, (in Japanese).
Hasegawa, N., Matsui, T., Tanaka, Y., Takahashi, Y., and Nambu, M., Consolidation properties of Holocene clay below seabed at Kobe Airport, Journal of the Japan Society of Civil Engineers C, 63(4), 923–935, 2007, (in Japanese).
vHong, Z., Zeng, L. L., Cui, Y. J., Cai, Y. Q., and Lin, C., Compression behaviour of natural and reconstituted clays. Géotechnique, 62(4), 291–301, 2012.
Kim, J.-H., Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational Statistics & Data Analysis*, 53(11), 3735–3745, 2009. doi: 10.1016/j.csda.2009.04.009.
Leroueil, S., and Vaughan, P. R., The general and congruent effects of structure in natural soils and weak rocks. Géotechnique, 40(3), 467–488, 1990.
Nadeau, C., and Bengio, Y., Inference for the generalization error. Machine Learning, 52, 239–281, 2003.
Oda, K., Lee, M., and Kitamura, S., Spatial interpolation of consolidation properties of Holocene clays at Kobe Airport using an artificial neural network. International Journal of GEOMATE, 4(1), 423–428, 2013.
Oda, K., Yokota, K., and Bu, L. D., Stochastic estimation of consolidation settlement of soft clay layer with artificial neural network. Japanese Geotechnical Society Special Publication, 2(74), 2529–2534. doi:10.3208/jgssp.JPN-041, 2016.