A probabilistic data-driven framework for modelling clay consolidation and settlement behavior

Kazuhiro Oda[1]*

[1] Department of Architecture and Environment Design, Osaka Sangyo University, Osaka, Japan

Abstract. Predicting consolidation settlement in large-scale reclamation projects remains difficult because laboratory consolidation test data are sparse in space and natural marine clays show strongly non-linear structured behavior. This study develops a probabilistic data-driven framework to reconstruct spatially continuous e-log p' and log k-e relationships from sparse consolidation test data. A deep neural network is combined with repeated K-fold cross-validation to estimate the ensemble mean response and its 95% confidence interval. The framework was trained using consolidation test data from 49 boreholes at Kobe Airport and was evaluated using two independent blind-test boreholes that were not used in model development. The predicted mean curves reproduced the main features of the observed compression and permeability responses, including depth-dependent yield behavior and post-yield changes in compressibility. The engineering applicability of the framework was examined through settlement analysis at monitoring point KC-1, where no site-specific borehole data were available. In this analysis, soil deformation was treated as one-dimensional, whereas pore-water flow was modeled as two-dimensional. The predicted material relationships were used as input. The calculated settlement history reproduced the main observed trend, while the late-stage difference from the measurements showed the importance of deeper strata outside the present modeling scope. The proposed framework provides a practical way to interpolate consolidation behavior in space with quantified ML-related uncertainty for large reclamation projects with similar geological and data conditions.

1 Introduction

1.1 Background and challenges

Accurate prediction of consolidation settlement is a key issue in large-scale land reclamation. Even small differences between predicted and actual settlement can affect fill volume, construction control, and project cost over a wide area. The consolidation properties of marine soft clays also vary in space because of depositional conditions. In addition, site investigation data include uncertainty from sampling, specimen preparation, and laboratory testing. Because high-quality consolidation tests cannot be performed at every location, engineers must estimate consolidation behavior at unmeasured locations from sparse site investigation data by using spatial estimation methods [1-3].

1.2 Complexity of natural clay behavior

Many natural marine clays are structured soils. Their compressibility cannot be represented well by the idealized bilinear e-log p' relationship often used in basic soil mechanics. Compared with reconstituted clay, natural clay often has a compression curve above the intrinsic compression line and an apparent yield stress larger than the in-situ effective overburden stress. This behavior is caused by structure formed during deposition, aging, and bonding. After yielding, destructuration progresses and may cause a large increase in compressibility over a transitional stress range. Therefore, accurate settlement prediction requires the full non-linear compression response of structured clay, not only a small number of scalar indices [4-6].

1.3 Limitations of existing methods

Data-driven site characterization often uses kriging, random fields, and Gaussian process regression to estimate spatial distributions of soil properties and uncertainty from sparse data. However, most studies predict stratigraphy or scalar engineering properties at arbitrary locations. They do not directly predict an entire compression curve such as the e-log p' relationship. Recent data-driven methods for reclamation projects have included stratigraphic uncertainty and spatial variability of scalar consolidation parameters in spatial and time-dependent settlement analysis. Constitutive studies of structured clay have mainly focused on reproducing material behavior itself. To the best of the authors' knowledge, few practical frameworks directly reconstruct the full compression response of structured clay over a heterogeneous three-dimensional domain from sparse site data [3,7].

* Corresponding author: oda@aed.osaka-sandai.ac.jp

 https://doi.org/10.1051/e3sconf/202673003004

1.4 Objectives of this study

This study develops a probabilistic data-driven framework for reconstructing spatially continuous e-log p' and log k-e relationships of natural marine clay from sparse consolidation test data. The framework combines a deep neural network with repeated K-fold cross-validation to quantify predictive uncertainty and is validated using independent blind-test boreholes. Its engineering applicability is examined by using the predicted relationships in a consolidation settlement analysis at an unmeasured monitoring point.

2 Probabilistic neural network framework

2.1 Deep neural network architecture

To model the non-linear consolidation behavior of clay, this study used a fully connected multilayer perceptron (MLP) as the deep neural network. The network was designed to capture the sharp change in compressibility that occurs just after the consolidation yield stress. This feature can be smoothed by shallow networks or conventional regression methods.

Figure 1 shows the neural network used in this study. The architecture consists of an input layer, 16 hidden layers with 250 neurons in each hidden layer, and an output layer. This deep and wide structure was selected to provide enough capacity to represent the spatial heterogeneity of soil properties in the three-dimensional domain, namely longitude, latitude, and depth. The 16 hidden layers also help the model represent highly non-linear features, such as the rapid increase in compression index Cc after yielding and the gradual decrease at higher stress levels. The rectified linear unit (ReLU) was used as the activation function in the hidden layers to reduce the vanishing-gradient problem and to support efficient training.

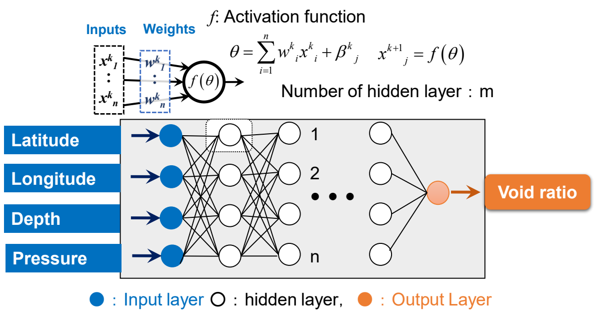

Fig. 1. Schematic view of neural network used in this study.

2.2 Training strategy and regularization

A high-capacity model can overfit the training data. In that case, the model learns noise in the training data rather than general trends. To reduce this risk, early stopping was used for regularization. The model kept a sufficiently complex architecture so that it could represent sharp yield behavior, but the iterative training process was stopped before the validation error began to increase. This strategy aimed to balance model flexibility and generalization performance.

2.3 Quantifying uncertainty via repeated K-fold cross validation

The relationship between consolidation pressure p' and void ratio e is first considered. This relationship is the consolidation curve,

$$e = e(p') \qquad (1).$$

Conventional deterministic prediction methods have difficulty representing epistemic uncertainty caused by limited data. To quantify this uncertainty, this study used repeated K-fold cross-validation [8,9].

Figure 2 shows the concept of K-fold cross-validation. In this method, the dataset is divided into K folds. For each split, K - 1 folds are used for model training, and the remaining fold is used for validation. By rotating the validation fold and repeating this procedure K times, K predictive models ($\hat{f}_{r,\ell}$) are obtained from one K-fold cross-validation run.

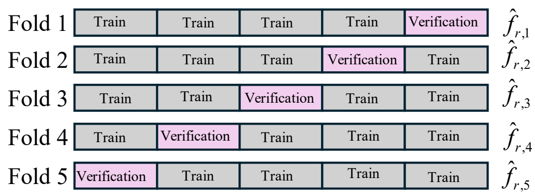

Fig. 2. Conceptual view of K-fold cross validation.

Repeated K-fold cross-validation was grouped by consolidation test, not by individual data point. Each standard consolidation test includes multiple loading-increment data points on a single e-log p' curve. To avoid information leakage between neighboring loading stages from the same specimen, all data points from one consolidation test were assigned to the same fold. Therefore, out-of-fold predictions and residuals were evaluated for consolidation tests that were completely excluded from the corresponding training fold. They were not evaluated for isolated loading-increment data points whose neighboring stages were already included in the training data.

In this study, the dataset consisted of 634 consolidation tests from 49 boreholes. These tests provided 5,359 loading-increment data points for neural-network training. Fold assignment was therefore based on the 634 consolidation tests, not on the 5,359 individual data points.

In repeated K-fold cross-validation, the K-fold procedure is repeated R times, and the fold partition is randomly redefined for each repeat. As a result, $K \times R$ predictive models are obtained.

Let

$$\hat{e}_{r,\ell}(p') \qquad (2)$$

denote the predicted void ratio at consolidation pressure p' obtained from the model trained in repeat r ($r = 1, \ldots, R$) and fold ℓ ($\ell = 1, \ldots, K$). The final prediction $\mu(p')$ at pressure p' is then defined as the ensemble average of all model predictions:

https://doi.org/10.1051/e3sconf/202673003004

$$\mu(p) = \frac{1}{KR}\sum_{r=1}^{R}\sum_{\ell=1}^{K}\hat{e}_{r,\ell}\,(p') \tag{3}$$

The predicted consolidation curve from repeated K-fold cross-validation is therefore an ensemble prediction obtained by averaging models trained with different data partitions.

The predictive uncertainty is represented by two main components. The first component is the model uncertainty caused by the variation among individual model predictions, denoted by

$$\sigma_{\mathrm{model}}(p'). \tag{4}$$

If the fold-averaged prediction in repeat r is defined as

$$\mu_r(p') = \frac{1}{K}\sum_{\ell=1}^{K}\hat{e}_{r,\ell}\,(p)', \tag{5}$$

then the model uncertainty is given by

$$\sigma^2_{\mathrm{model}}(p') = \mathrm{Var}_r\big(\mu_r(p')\big) + \frac{1}{R}\sum_{r=1}^{R}\mathrm{Var}_\ell\left(\hat{e}_{r,\ell}(p')\right). \tag{6}$$

Accordingly,

$$\sigma_{\mathrm{model}}(p') = \sqrt{\mathrm{Var}_r\big(\mu_r(p')\big) + \frac{1}{R}\sum_{r=1}^{R}\mathrm{Var}_\ell\left(\hat{e}_{r,\ell}(p')\right)}. \tag{7}$$

The second component is the variability estimated from validation residuals. For each data point i, the prediction obtained from a fold in which that point is not used for training is referred to as an out-of-fold (OOF) prediction. Let $\hat{e}_i^{\mathrm{OOF},(r)}$ denote the OOF prediction in repeat r. The average OOF prediction is then defined as

$$\mu_i^{\mathrm{OOF}} = \frac{1}{R}\sum_{r=1}^{R}\hat{e}_i^{\mathrm{OOF},(r)}. \tag{8}$$

Based on this quantity, the OOF residual scale is defined as

$$\sigma_{\mathrm{res,OOF}}(p') = \sqrt{\mathbb{E}\left[\left(e_i - \mu_i^{\mathrm{OOF}}\right)^2 \middle| p'_i = p'\right]}. \tag{9}$$

This quantity may be interpreted as a residual scale that includes observation error, unexplained variability, and model misspecification. Thus, $\sigma_{\mathrm{res,OOF}}(p')$ represents the residual scale for predicting unseen consolidation-test curves, rather than the interpolation error of individual loading increments within a test that was already partly used for training.

Accordingly, the predictive uncertainty associated with an observed consolidation test result is expressed as

$$\sigma_{\mathrm{pred}}(p') = \sqrt{\sigma^2_{\mathrm{model}}(p') + \sigma^2_{\mathrm{res,OOF}}(p')}. \tag{10}$$

Here, $\sigma_{\mathrm{model}}(p')$ represents uncertainty associated with model construction, whereas $\sigma_{\mathrm{res,OOF}}(p')$ represents the variability estimated from validation residuals.

In this way, repeated K-fold cross-validation estimates not only a single predicted value $\mu(p')$ for each consolidation pressure p', but also the associated uncertainty through $\sigma_{\mathrm{model}}(p')$ and $\sigma_{\mathrm{pred}}(p')$. The framework can evaluate both the average trend of the consolidation curve and the reliability of the prediction.

The relationship between void ratio e and the coefficient of permeability k is also required, as expressed by Eq. (11).

$$k = k(e) \tag{11}$$

The same scheme was applied to this relationship by replacing consolidation pressure with void ratio as the input and using $\log k$ as the output.

In this study, the number of folds K and the number of repeats R were set to 5 and 20, respectively.

3 Site characterization and database

3.1 Project overview and geological setting

The study site is the Kobe Airport (Figure 3), which is located at the northern side of Osaka Bay in Japan. Kansai International Airport is also located in Osaka Bay and is known for long-term residual settlement. Like Kansai International Airport, Kobe Airport is a large reclaimed artificial island constructed on soft marine clay deposits. The typical geological profile consists of a thick Holocene clay layer overlying Pleistocene clay layers. As shown in Figure 4, the seabed soil has high compressibility and high water content. The initial void ratio (e_0) is about 3.0 near the seabed surface and decreases to about 1.3 at a depth of 30 m, near the Ds-1 layer. Based on these void ratio characteristics, the coefficient of permeability k of the clay layers is estimated to be in the range of 10^{-8} to 10^{-9} m/s. The seabed was subjected to the time-dependent reclamation loading history shown in Figure 5, which caused significant consolidation settlement.

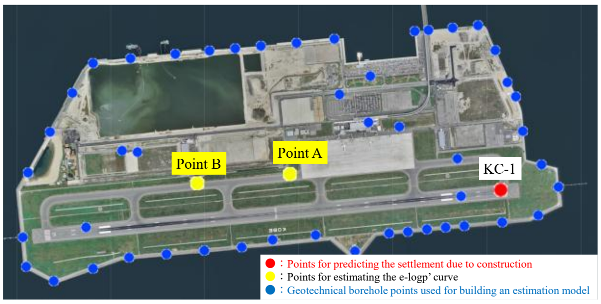

Fig. 3. Reclaimed artificial island examined in this study.

3.2 Geotechnical database compilation

To construct the data-driven model, a geotechnical database was compiled from site investigation reports. The dataset includes standard incremental loading consolidation tests on undisturbed samples. A total of 49 boreholes were used for model training. These boreholes provided 634 consolidation test samples, and the raw test results were discretized into 5,359 individual data points to train the neural networks.

For the *e*-log*p*' prediction model, the input variables were spatial coordinates (latitude, longitude, and depth) and consolidation pressure *p*', and the target variable was void ratio *e*. For the permeability prediction model, the input variables were spatial coordinates and void ratio *e*, and the target variable was coefficient of permeability *k*.

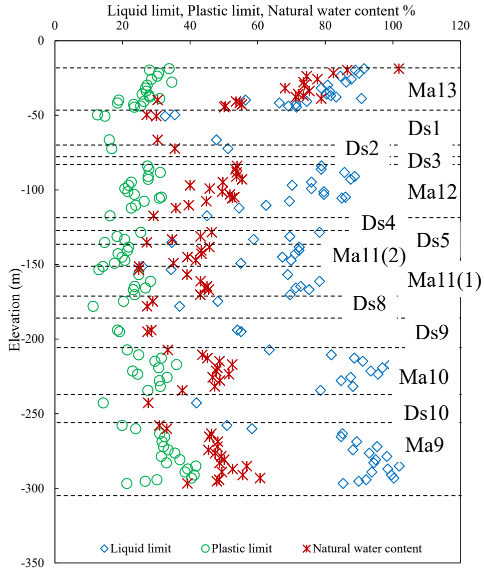

Fig. 4. Geological profile of the targeted reclamation site.

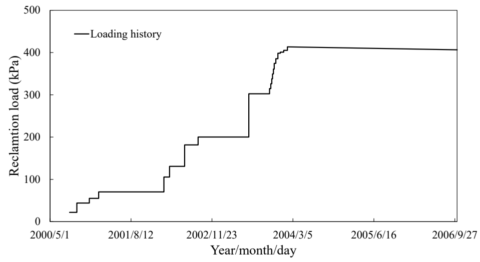

Fig. 5. Time history of the reclamation load.

3.3 Strategy for validation and testing

The validation strategy was designed to assess the model's ability to interpolate soil properties in space. The dataset was divided into three groups, as shown in Figure 3. The first group was the training locations. Data from 49 boreholes were used to train the neural network and to perform repeated K-fold cross-validation grouped by consolidation test. The second group was the independent validation sites, Points A and B. These two boreholes were completely excluded from the training dataset and were used as blind tests to evaluate prediction at unmeasured locations. Points A and B contain 14 and 13 consolidation test results, respectively. The third group was settlement monitoring point KC-1. This location has long-term settlement records measured during and after reclamation, but it has no site-specific borehole data. It therefore represents a practical application case in which settlement behavior must be predicted from learned spatial patterns.

4 Model validation and spatial prediction

4.1 Parametric study of the DNN architecture

To examine the suitability of the neural network architecture used in this study, a parametric study was carried out for the number of hidden layers and the number of neurons in each hidden layer. Table 1 shows how these two parameters affect the RMSE for predicting the *e*-log *p*' relationship. The minimum RMSE in the parametric study was 0.119. This value is close to the RMSE of 0.122 obtained with the neural network used in this study, which has 16 hidden layers and 250 neurons in each hidden layer. Thus, the selected architecture did not give the absolute minimum RMSE, but it was within a stable performance range.

Table 1. Effect of the number of hidden layers and neurons on the RMSE for the *e*- log*p'* relationship.

Neurons	Hidden layers		
	4	8	16
8	0.152	0.188	0.241
16	0.123	0.130	0.127
32	0.120	0.119	0.122
64	0.120	0.120	0.123
128	0.122	0.120	0.121
250	0.121	0.121	0.122
512	0.123	0.122	0.148

Table 2 summarizes the corresponding MAE, MAPE, and R^2 values. The MAPE is low, and the R^2 value is very high.

Table 2. Summary of MAE, MAPE, and R^2 values.

MAE	MAPE	R^2
0.088	0.061	0.930

Table 3 shows how the number of hidden layers and neurons affects the RMSE for predicting the log *k*–*e* relationship. The minimum RMSE in the parametric study was 0.680, which is not substantially different from the RMSE of 0.692 obtained with the neural network used in this study.

Table 3. Effect of the number of hidden layers and neurons on the RMSE for log *k-e* relationship.

Neurons	Hidden layers		
	4	8	16
8	0.757	0.856	0.980
16	0.697	0.718	0.735
32	0.692	0.697	0.709
64	0.683	0.685	0.691
128	0.683	0.685	0.689
250	0.684	0.689	0.692
512	0.680	0.689	0.909

Table 4 summarizes the corresponding MAE, MAPE, and R^2 values. The prediction performance is slightly lower than that for the *e*-log *p*' relationship, but the MAPE is still low and the R^2 value remains high.

Table 4. Summary of MAE, MAPE, and R² values.

MAE	MAPE	R^2
0.512	0.033	0.811

These results do not identify a unique optimal architecture. However, they show that the neural network with 16 hidden layers and 250 neurons in each hidden layer gives accuracy close to the best values for both relationships. This architecture was therefore retained as a stable choice with sufficient capacity for the present data.

The purpose of this study is not to compare different machine-learning algorithms, but to examine whether the proposed data-driven framework can reconstruct full consolidation and permeability curves and use them in settlement analysis. Therefore, a full benchmark comparison with other algorithms, such as random forests or Gaussian process regression, is outside the scope of this paper. Instead, the robustness of the adopted DNN architecture was examined through a parametric study of the number of hidden layers and neurons.

4.2 Performance on independent "blind" test data

Figures 6 and 7 compare the predicted *e*-log *p*' relationships at the two blind-test boreholes with the laboratory consolidation test results. Overall, the ensemble mean curves reproduced the main shape of the observed compression response, including the depth-dependent appearance of consolidation yielding and the post-yield change in compressibility.

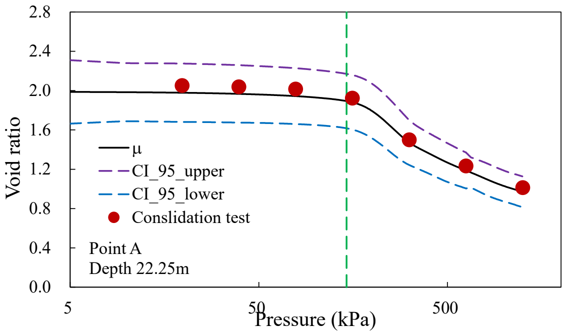

Fig. 6. Predicted *e*-log *p*' relationship (Point A, depth 22.25m).

Table 5 summarizes the blind-test on *e*-log *p*' relationships. Because $\sigma_{res,OOF}$ is much larger than σ_{model}, it provides the main contribution to σ_{pred}. For Points A and B combined, 27 consolidation tests were available, comprising 217 data points in total. Of these data points, 201, or about 93%, were within the 95% confidence interval. This high empirical coverage indicates that the proposed method can reproduce the *e*-log *p*' relationship with high accuracy at the blind-test boreholes.

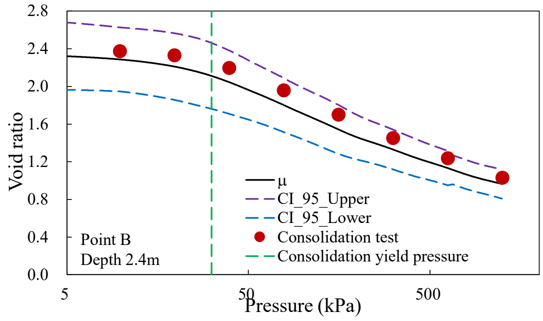

Fig. 7. Predicted *e*-log *p*' relationship (Point B, depth 2.4m).

Table 5. Summary of blind-test results for the *e*-log *p*' relationships.

	Point A	Point B	All
No. of tests	14	13	27
No. of data points	112	105	217
RMSE	0.121	0.110	0.116
MAE	0.090	0.082	0.087
Bias	-0.017	-0.039	-0.028
R^2	0.931	0.932	0.932
Coverage of 95% CI	102/112	99/105	201/217

4.3 Prediction of permeability

Figures 8 and 9 show the predicted log *k*-*e* relationships at the same blind-test locations. The predicted mean curves broadly captured the decrease in the coefficient of permeability with decreasing void ratio. The 95% confidence intervals also showed depth-dependent differences among the four examples. The scatter and local fluctuations were larger than those for the *e*-log *p*' relationships, which is consistent with the higher variability generally associated with permeability measurements.

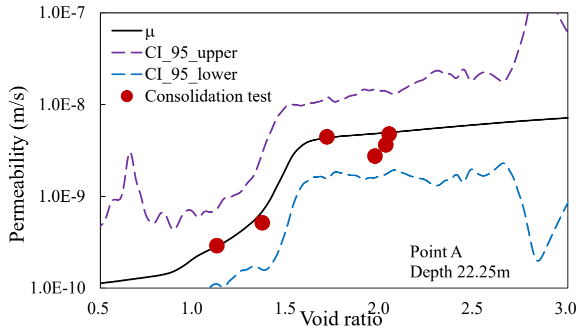

Fig. 8. Predicted relationship between coefficient of permeability and void ratio (Point A, depth 22.25m).

Table 6 summarizes the blind-test results for the log *k*-*e* relationships. Although σ_{model} is approximately 40% of $\sigma_{res,OOF}$, it makes only a small contribution to σ_{pred}. For the combined dataset from Points A and B, 190 data points were available. Of these data points, 150, or about 79%, were within the 95% confidence interval. These results indicate that the proposed method captures the main trend of the log *k*–*e* relationship. However, the empirical coverage of the 95% confidence interval was lower than the nominal 95% level. This suggests that uncertainty in the permeability prediction is underestimated to some extent. Therefore, the log *k*–*e* prediction should be interpreted with more caution than the e–log p’ prediction.

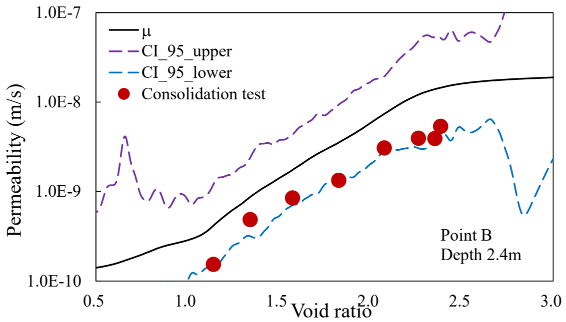

Fig. 9. Predicted relationship between coefficient of permeability and void ratio (Point B, depth 2.4m).

Table 6. Summary of blind test results for the log k-e relationships.

	Point A	Point B	All
No. of tests	14	13	27
No. of data points	98	92	190
RMSE	0.705	0.764	0.734
MAE	0.528	0.596	0.561
Bias	-0.031	0.329	0.144
R^2	0.772	0.773	0.773
Coverage of 95% CI	79/98	71/92	150/190

4.4 Applicability of proposed methodology

Taken together, the results show that the proposed framework can estimate both e–log p’ and log k–e relationships continuously over the reclamation area while providing an uncertainty measure around the predicted mean response. In practical use, the confidence interval can be used as a model-derived indicator of prediction uncertainty. However, it should be interpreted together with the distance from training boreholes, geological continuity, and the ranges of pressure and void ratio covered by the training data. It should not be regarded as a complete measure of spatial uncertainty.

5 Application: coupled finite element analysis

5.1 Numerical modelling strategy

To examine the engineering applicability of the proposed framework, a coupled consolidation analysis was carried out for settlement monitoring point KC-1, where no borehole investigation data were available. In the analysis, soil deformation was treated as approximately one-dimensional, whereas pore-water flow was modeled as two-dimensional. All constitutive inputs for the target clay layer were assigned from the spatially predicted e–log p’ and log k–e relationships. The analytical model shown in Figure 10 was constructed to represent the soil layering at KC-1. The time-dependent reclamation loading sequence, including the placement of fill materials, was applied as shown in Figure 5. For each finite element layer, the e–log p’ and log k–e relationships were assigned through Monte Carlo simulation. In this simulation, normally distributed random perturbations with the estimated standard deviation were added to the predictions obtained by repeated K-fold cross-validation. The standard deviations used for these perturbations were derived from the cross-validation grouped by consolidation test, as described in Section 2.3. They were not obtained by data-point-level splitting. In this Monte Carlo analysis, uncertainty in the compression and permeability relationships was propagated through the finite element model. Possible correlation between the errors in the e–log p’ and log k–e predictions was not explicitly modeled.

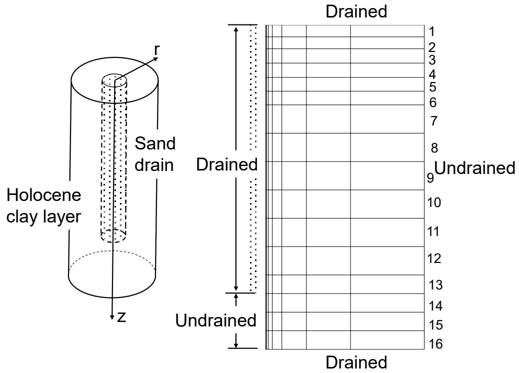

Fig. 10. Analytical model for FEM.

5.2 Interpretation of predicted parameters

The neural network reconstructed the depth-dependent variation of soil properties, which is important for settlement prediction. Figure 11 shows representative e–log p’ curves for KC-1. In shallow layers (for example, Layers 1 – 7), the curves show a gradual transition without a clear yield stress overshoot. This behavior is typical of remolded or less structured clay near the seabed. In middle layers (for example, Layers 9–10), a clear consolidation yield stress pc is observed, followed by a sharp increase in compressibility. This behavior represents structural breakdown of the clay skeleton. In deep layers (for example, Layers 13–16), the post-yield compression curves lie below those of the surface layers, reflecting the lower initial void ratio caused by the overburden pressure.

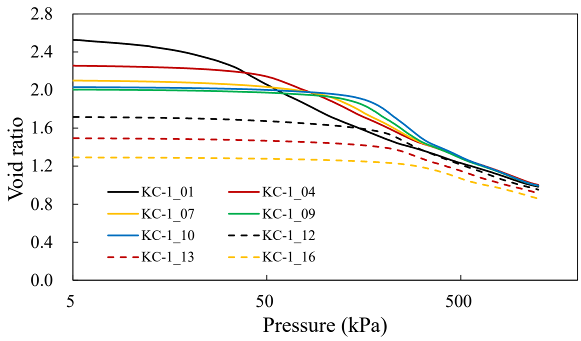

Fig. 11. Examples of e–log$p$’ curve used for the consolidation settlement simulation.

Figure 12 shows the predicted coefficient of permeability inputs. The model assigned initial void ratios ranging from $e_0 \approx 3.0$ at the surface to $e_0 \approx 1.3$ at a depth of 30m. The corresponding coefficient of permeability values were in the range of 10^{-8} to 10^{-9} m/s. These values are physically consistent with the depositional environment of the marine clay.

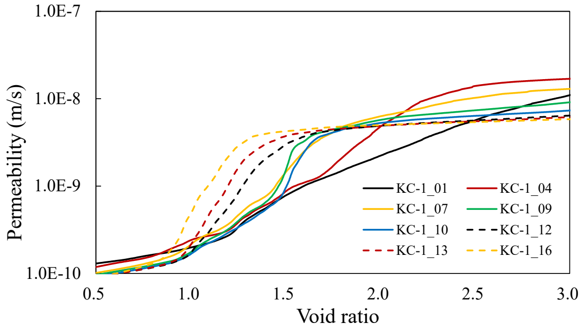

Fig. 12 Examples of log k–e curve used for the consolidation settlement simulation.

5.3 Reproduction of field settlement behavior

Figure 13 shows the time history of settlement caused by land reclamation. Until about July 2003, the simulated results agree well with the field measurements. After that time, the measured settlement becomes larger than the simulated settlement. The field measurements were obtained mainly for runway maintenance. Therefore, the recorded settlement includes not only consolidation of the Holocene clay layer (Ma13), which is the main target of this analysis, but also settlement of the underlying Pleistocene strata (the Ma12 layer and deeper layers). The difference between the simulated and observed values increases as the settlement contribution from these deep Pleistocene deposits becomes larger [10]. Settlement in the Pleistocene strata is common on reclaimed islands in Osaka Bay. The well-known settlement problem at Kansai International Airport is an example [11].

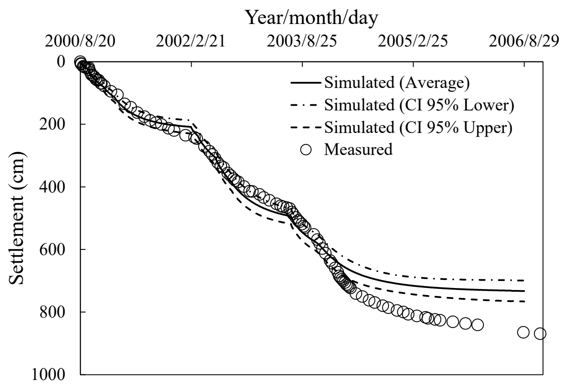

Fig. 13 Comparison between the simulated time-settlement curve with the actual field monitoring records at point KC-1.

The mean final settlement from the analysis is about 7.32 m, with a 95% confidence interval from 7.00 m to 7.66 m. The interval width of 0.66 m corresponds to 9.1% of the predicted final settlement. This confidence interval represents only the uncertainty propagated from the predicted e–log p' and log k–e relationships for the Holocene clay layer. It does not include uncertainty in the reclamation loading history, drainage and boundary conditions, stratigraphic geometry, constitutive assumptions, finite element idealization, or settlement of the underlying Pleistocene strata. Therefore, the interval should not be interpreted as the total uncertainty of the field settlement. Within this limited scope, the result suggests that the ML-related material-parameter uncertainty is relatively small.

6 Discussions

6.1 Efficacy of ensemble averaging

A key finding of this study is the advantage of the probabilistic ensemble over a single deterministic model. In the validation results, individual neural-network predictions generated from random data subsets showed some fluctuations. This variation occurs because a single model may overfit local noise or may be biased by a specific training data split. In contrast, the ensemble mean from repeated K-fold cross-validation smoothed these fluctuations and converged toward the laboratory-measured compression curves. Therefore, although individual models may have high variance, their average provides a smoother and more stable estimate of the observed consolidation behavior. This result supports the use of repeated K-fold cross-validation, because a single training run cannot provide the same level of reliability.

6.2 Engineering implications of uncertainty quantification

In contrast to using a single deterministic curve, the proposed framework quantifies ML-related prediction uncertainty using upper and lower bounds based on the estimated standard deviation. The width of the prediction band can be used as a model-derived indicator of prediction uncertainty, but it should be interpreted together with geological conditions and data coverage. In the KC-1 simulation, the mean curve was used to reproduce the most probable settlement. For design purposes, the upper-bound compressibility curve could be used to estimate a conservative upper bound of settlement. This can support a more rational assessment of settlement risk for infrastructure.

6.3 Limitations and future work

Repeated K-fold cross-validation was grouped by consolidation test to avoid leakage among loading-increment data from the same consolidation test. This procedure provides a residual estimate for unseen consolidation test curves. However, it is still not equivalent to leave-one-borehole-out validation, because other consolidation tests from the same borehole may remain in the training data. Therefore, the borehole-scale spatial interpolation capability was evaluated separately using the two independent blind-test boreholes, Points A and B. The validation at Points A and B provides an independent check of spatial interpolation, but the number of blind-test boreholes is still limited. Therefore, the present results should be regarded as a proof of concept for the Kobe Airport reclamation site rather than a complete demonstration of general spatial transferability. The present neural network does not explicitly impose monotonicity on the log k–e relationship, and the empirical coverage for permeability was lower than the nominal 95% level. Future work should examine monotonic neural networks or post-processing methods, recalibration of σ_{pred} for permeability, explicit treatment of correlation between compression and permeability errors, and validation using additional blind boreholes or leave-one-borehole-out cross-validation.

7 Conclusions

This study developed a probabilistic data-driven framework for predicting the spatial distribution of clay consolidation properties and examined its applicability through a large-scale reclamation case study. The main conclusions are as follows:

1. The proposed framework reconstructed spatially continuous e–log p' and log k–e relationships of marine clay from sparse consolidation test data. It also quantified ML-related uncertainty by providing 95% confidence intervals around the predicted mean response.
2. The ensemble mean obtained from repeated K-fold cross-validation reproduced the main features of the compression and permeability behavior observed at two independent blind-test boreholes. The empirical coverage was high for the e–log p' relationship, whereas the log k–e relationship required more cautious interpretation because of lower coverage.
3. When the predicted relationships were used in the settlement analysis at KC-1, the analysis reproduced the main settlement trend of the target clay layer. The late-stage difference between measured and simulated settlement also showed the need to include deeper strata in future applications.

Overall, the proposed framework provides a practical tool for interpolating consolidation behavior within a geologically consistent reclamation area, provided that the target depth, stress range, and soil conditions are covered by the training boreholes. Application to other deposits or to extrapolation outside the training domain requires retraining and independent validation. Remaining issues include comparison with simpler baseline models, additional borehole-level validation, calibration and monotonic constraints for the permeability model, and uncertainty sources outside the ML-derived material relationships. The present discussion is therefore not exhaustive, and these issues require further work.

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

References

1. K. Oda, M. Lee, S. Kitamura, Spatial interpolation of consolidation properties of Holocene clays at Kobe Airport using an artificial neural network. Int. J. GEOMATE 4, 423–428 (2013)
2. K. Oda, K. Yokota, L.D. Bu, Stochastic estimation of consolidation settlement of soft clay layer with artificial neural network. Jpn. Geotech. Soc. Spec. Publ. 2, 2529–2534 (2016). https://doi.org/10.3208/jgssp.JPN-041
3. Y. Tsuda, I. Yoshida, S. Nishimura, Spatial distribution estimation by considering the cross-correlation between components with indirect data using Gaussian process regression. Soils Found. 65, 101624 (2025). https://doi.org/10.1016/j.sandf.2025.101624
4. J.B. Burland, On the compressibility and shear strength of natural clays. Géotechnique 40, 329–378 (1990). https://doi.org/10.1680/geot.1990.40.3.329
5. S. Leroueil, P.R. Vaughan, The general and congruent effects of structure in natural soils and weak rocks. Géotechnique 40, 467–488 (1990). https://doi.org/10.1680/geot.1990.40.3.467
6. Z. Hong, L.L. Zeng, Y.J. Cui, Y.Q. Cai, C. Lin, Compression behaviour of natural and reconstituted clays. Géotechnique 62, 291–301 (2012). https://doi.org/10.1680/geot.10.P.046
7. I. Yoshida, Y. Tomizawa, Y. Otake, Estimation of trend and random components of conditional random field using Gaussian process regression. Comput. Geotech. 136, 104179 (2021). https://doi.org/10.1016/j.compgeo.2021.104179
8. J.-H. Kim, Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap. Comput. Stat. Data Anal. 53, 3735–3745 (2009). https://doi.org/10.1016/j.csda.2009.04.009
9. C. Nadeau, Y. Bengio, Inference for the generalization error. Mach. Learn. 52, 239–281 (2003)
10. N. Hasegawa, T. Matsui, Y. Tanaka, Y. Takahashi, M. Nambu, In-situ consolidation behavior of Pleistocene clay below seabed at Kobe Airport. J. Jpn. Soc. Civ. Eng. C 62, 780–792 (2006) (in Japanese)
11. T. Furudoi, The second phase construction of Kansai International Airport considering the large and long-term settlement of the clay deposits. Soils Found. 50, 805–816 (2010)