Improvement of Side Resistance Prediction for Pile Foundation Using Construction Information

| Journal:          | Canadian Geotechnical Journal |
|-------------------|-------------------------------|
| Manuscript ID     | cgj-2018-0835.R3              |
| Manuscript Type:  | Article                       |
| Date Submitted by the Author: | 11-Mar-2020              |
| Complete List of Authors: | Otake, Yu; Niigata University, Civil Engineering and Architecture Watanabe, Shinya; Niigata University, Civil Engineering and Architecture Mizutani, Taisaku; Japanese Technical Association for Steel Pipe Piles and Sheet Piles |
| Keyword:          | Bayes’ theorem, construction information, incomplete data, multidimensional normal distribution, bearing capacity |
| Is the invited manuscript for consideration in a Special Issue?: | Not applicable (regular submission) |
Improvement of side resistance prediction for pile foundation using construction information

Yu Otake
Department of Civil Engineering and Architecture,
Niigata University, Japan
(y_ohtake@eng.niigata-u.ac.jp)

Shinya Watanabe
Department of Civil Engineering and Architecture,
Niigata University, Japan

Taisaku Mizutani
Japanese Technical Association for Steel Pipe Piles and Sheet Piles, Japan

May 7, 2020

Abstract

Problems such as the inclination and settlement of buildings after construction completion have been reported in recent years. Quality control during construction and kieruka ("visualizing" the work quality) are increasingly demanded. In light of this, the present study addresses execution with a "rotary penetration steel pipe pile (RPS-pile)", which is a system that is able to collect information in real time during the piling process. When a RPS-pile is placed, high push-in and pull-out bearing capacities are expected because the spiral blade at the pile tip resists at the bottom of the borehole against an external force. Additionally, continuous real-time data on the pile-head torque and the auger penetration depth per revolution are obtained, as these mechanical indicators are necessary for piling. This study aims at developing a method for the real-time confirmation of work quality at construction sites by utilizing construction information (information obtained during piling). Specifically, a method is proposed for the sequential updating of reliability in the estimation of the side friction acting on a drilled pile. In this method, real-time information on the pile-head torque and the auger penetration depth per revolution obtained during piling are used in addition to N-values that are observed in a standard penetration test conducted in advance as part of subsurface exploration.
Contents

1 Introduction .......... 3
   1.1 Background of the study .................. 3
   1.2 Importance of this study in light of previous studies ............ 3

2 Data used in this study .......... 6
   2.1 Development and outline of a database ............... 6
   2.2 Clean-up and screening of data .................. 8
      2.2.1 Discretization of continuous data corresponding to the observed depth ............. 8
      2.2.2 Definition of the normalized pile-head torque $T'$ ............. 8
      2.2.3 Criteria for the data screening .................. 9

3 Study method .......... 9
   3.1 Flow of the study .................. 9
   3.2 Method for creating an estimation equation by utilizing the conditional normal distribution .................. 10
   3.3 Approximation of an estimated covariance matrix by a positive definite matrix .................. 12
   3.4 Methods for estimating a mean vector and a variance-covariance matrix ............. 13
      3.4.1 Estimation Method 1 .................. 13
      3.4.2 Estimation Method 2 .................. 14
   3.5 Method for generating a prior distribution .................. 15
   3.6 Method for updating a prior parameter distribution based on observation data .................. 16

4 Study results .......... 17
   4.1 Data processing results .................. 17
   4.2 Estimation results obtained by Estimation Method 1 ............. 20
   4.3 Estimation results obtained by Estimation Method 2 ............. 20

5 Conclusions and future issues .......... 24

A Data used for the analyses .......... 24

B Changes in simultaneous probability density resulting from updating by using piling data (Estimation Method 2) .......... 24
1 Introduction

1.1 Background of the study

Because significant inclination and settlement of buildings after the completion of construction has been reported in recent years, the importance of quality control during construction has come to be widely recognized. In planning repairs and reinforcement that are required as a result of the deterioration of buildings or non-conformity caused by changes in laws and regulations, it is necessary to have a sound understanding of the initial conditions of the buildings after the completion of construction. Particularly with regard to subsurface structures, which are addressed in this study, piling is conducted without sufficient understanding of the geomaterials' characteristics, so efforts have been made to build structures in a reliable manner by means of an observational construction control system that utilizes gathered information. However, the information collected during construction is usually utilized for detecting abnormalities and not for quantitative assessment of the work quality or the structure performance. Therefore quality control during construction and mierekka ("visualizing" the work quality) are increasingly demanded.

This problem is addressed in this study. Specifically, by utilizing rotary penetration steel pipe piles (RPS-piles), methods are examined for collecting basic information on the structure performance that is necessary for observational construction control and for decision-making regarding repairs and reinforcements after construction. For this purpose, information collected during construction is related to structure performance. A RSP-pile is a steel pipe pile with a spiral blade at the pile tip. Fig. 1 shows a schema of a RSP-pile. Torque given to the pile head helps the pile auger penetrate the ground. Because this piling method ensures low vibration and low noise and does not require earth removal, it can be implemented even on small plots of land, and the effects on the surrounding area are limited. The characteristics of RPS-piles are as follows: large push-in and pull-out bearing capacities are expected because the spiral blade helps to resist at the bottom of the borehole against an external force; and continuous data on the pile-head torque and the auger penetration depth, which are mechanical indicators necessary for piling, are obtained in real time (Mori (2003), Saeki and Ohki (2003), Shioi et al. (2013) and Kawai et al. (2018)). In this paper, a basic concept is examined regarding the real-time estimation of the vertical bearing capacity of RSP-piles. Then, a method is proposed for the sequential updating of reliability in the estimation of the side friction acting on a drilled pile at various depths. In this method, real-time information on the pile-head torque and the auger penetration depth per revolution obtained during piling is used in addition to $N$-values that are observed in a standard penetration test in advance.

1.2 Importance of this study in light of previous studies

In the Specifications for Highway Bridges Part IV (JRA 1996), a major design standard in Japan, the ultimate push-in bearing capacity is calculated by using the design model shown by Eq. (1)

$$R_u = q_d A_p + U \sum L_i f_i = R_{tip}^u + R_{side}^u$$  \hfill (1)

where, $R_u$ = the ultimate bearing capacity of the pile (kN); $R_{tip}^u$ = the ultimate bearing capacity of the pile tip; $R_{side}^u$ = the ultimate bearing capacity of the pile side friction;
Figure 1: Schema of a RSP-pile

Table 1: Coefficient of the tip resistance of the pile ($\alpha_{qd}$) according to $D_w/D_p$ and the type of soil

| Soil Type | $D_w/D_p$ | $\alpha_{qd}$ |
|-----------|-----------|---------------|
| Sand      | 1.5       | 120           |
|           | 2.0       | 100           |
| Gravel    | 1.5       | 130           |
|           | 2.0       | 115           |

$q_d =$ the ultimate bearing capacity per unit area at the pile tip (kN/m$^2$) (hereinafter, tip resistance); $A_p =$ the area of the pile tip (m$^2$); $U =$ the circumferential length of the pile (m); $L_i =$ the thickness of the layer for which side friction acting on the drilled pile is taken into account (m); and $f_i =$ the ultimate unit side friction acting on a drilled pile in the layer for which the side friction acting on the drilled pile is taken into account (kN/m$^2$) (hereinafter, side resistance). The tip resistance ($q_d$) and the side resistance ($f_i$) vary depending on piling method and soil type. In Fig. 1c, Modeling of the ultimate bearing capacity is shown.

When a RSP-pile is utilized, $q_d$ is calculated by using Estimation Eq. (2):

$$q_d = \alpha_{qd} N$$

In this equation, $\alpha_{qd}$ is a coefficient of tip resistance (Table 1). This coefficient is determined according to $D_w/D_p$ (the two types of piles, $D_w/D_p = 1.5$ and $2.0$, are in the market as the factorial products), the ratio of the blade diameter ($D_w$) to the pile diameter ($D_p$), and also on the basis of the type of soil that the pile tip reaches. $N$ is $N$-values of standard penetration test.

Estimation equations [Eq. (3) and Eq. (4)] give $f_i$. $f_i$ is estimated by $N$-values and regression coefficient $\alpha_f$.

- Sandy soil : $f_i = \alpha_f N (\leq 150)$
- Cohesive soil : $f_i = c$ or $\alpha_f N (\leq 100)$
where, $c$ is the soil cohesion (kN/m²). Also, a regression coefficient $\alpha_f$ is 3 for sandy soil and 10 for cohesive soil.

Okahara et al. (1991) collected data from many vertical pile loading tests that were conducted in Japan, and they quantified the modeling error of the design equation [Eq.(2), Eq.(3) and Eq.(4)] shown above. In their study, the ultimate bearing capacity used for analysis was defined as the load which was applied when the settlement depth was 10% of the pile diameter. They also conducted statistical analysis of the bearing capacity ratio ($A/B$). Here, $A$ is a value of the ultimate bearing capacity observed in a vertical loading test. $B$ is the ultimate bearing capacity calculated by using a bearing capacity estimation equation on the basis of $N$-values. The analysis results were first divided into those for cast-in-place concrete piles (CCP) and those for steel pipe piles (SSP), and the data were further divided into those for bearing piles and those for friction piles. Consequently, the data were stratified, and statistics are shown for each stratum (Okahara et al. 1991). According to their analysis, the coefficient of variation of the ultimate bearing capacity of the pile tip ($COV_{R_{tip}}$) and of the ultimate bearing capacity of the pile side friction ($COV_{R_{side}}$) are 0.36-0.46 and 0.53-0.63 respectively, the latter being slightly greater. Bias, which is defined as the mean value of $A/B$, is 1.01-1.21. This empirical method suggests that the design model slightly underestimates the ultimate bearing capacity and has large uncertainty.

Furthermore, empirical methods for checking the pile foundation stability that are similar to the methods described in 4-studies have been used in major design standards in Japan and elsewhere (AIJ 2000; OCDI 2009) as well as by AASHTO (AASHTO 1989, 2000, 2004). For the purpose of revising the AASHTO standards, Paikowsky (2004) developed a large database of vertical pile loading test results that were chiefly collected from the East Coast of the USA and also from abroad, and analyzed the modeling error of a pile design equation. Paikowsky (2004) collected and analyzed dynamic loading test results in addition to static loading test results, but the static loading test results alone are shown below. The ultimate bearing capacity is defined by using Davisson’s criterion. In Paikowsky’s analysis, modeling error correction is done by considering the soil type alone, without separating the side resistance and the tip resistance. Pile types and piling methods are not taken into consideration. Paikowsky reported that $bias_{R_{tip}}$ was 1.12 and $COV_{R_{tip}}$ was 0.59 for sandy soil, and that $bias_{R_{tip}}$ was 0.83 - 0.84 and $COV_{R_{tip}}$ was 0.35 - 0.39 for cohesive soil and alternating strata. These results indicate that the calculated values are slightly lower than the observed values for sandy soil and are higher than the observed values for cohesive soil. The coefficient of variation varies depending on the soil type. Furthermore, in recent years, the modeling error of the ultimate bearing capacity of piles in the vertical direction has been discussed, and it has been reported that the magnitude and the uncertainty of the ultimate bearing capacity varies depending on the soil type (Tang and Phoon (2018) and Tang and Phoon (2019)).

A piling technique that utilizes RSP-piles was developed around 2000; thus, the aforementioned studies did not deal with this relatively new piling technique. Fig. 2 shows the results of loading tests (PWRC 2009, 2011; JICE 2004) which were performed in the process of developing this piling technique. Values of $f_i$ and $N$-values observed in standard penetration tests at piling sites are shown for each soil type and $D_o/D_p$. Estimation Eqs. (3) and (4), which are included in the current design standards of Japan, for $f_i$ of RSP-piles are also presented in Fig. 2. Fig. 3 show the histograms of $\alpha_{f_i}$ depending on the soil type. It is shown that the current design equation has large uncertainty on the relationship between $f_i$ and $N$-values. Regarding the statistics of $A/B$, $bias_{f_i} = 1.63$ and $COV_{f_i} = 0.97$ for sandy soil, and $bias_{f_i} = 2.06$
Figure 2: $f_i$ vs. SPT-N value determined by the specification’s calculation equation and $COV_{f_i} = 0.97$ for cohesive soil. In estimating an $f_i$ value from an $N$-value for a RSP-pile as well as for a conventional pile foundation, simple estimation equations involve large uncertainty.

As mentioned above, the empirical method of the ultimate bearing capacity of the pile in the vertical direction generally has large uncertainty. Although the magnitude of $bias_{f_i}$ and $COV_{f_i}$ are also big problem, it should be noted that the degree of the side friction varies depending on the soil type. At the piling sites, the soil layer composition of a certain point will be measured by some soil investigations such as standard penetration test etc., but the soil layer composition of the enforcement points of all piles will be not necessarily clarified. The risk caused by this information uncertainty has the potential to cause major problems such as significant inclination and settlement of building after the completion of construction.

This study aims at developing a method for the real-time confirmation of work quality at construction sites by utilizing information obtained during piling against this large uncertainty. Specifically, a method is proposed for the sequential updating of $f_i$ taking into account the characteristics of the piling site.

2 Data used in this study

2.1 Development and outline of a database

The database used in this study consists of the following: datasets contained in written reports on the review and certification of this piling technique (PWRC 2009, 2011; JICE 2004) (vertical loading test results available); and piling data collected at many piling sites in Japan (vertical loading test results unavailable). The database includes vertical loading test results, construction information (piling data), and soil investigation data (i.e., soil types and standard penetration test results). Because this study addresses real-time data updating for the side friction acting on a drilled pile, the vertical loading test results in the database are those of $f_i$ alone.

Table 2 shows the basic specifications of the pile used for analysis and the indices
used in various tests. The specifications of RSP-piles relevant to the database and the observations of each index are shown in Table 6 in Appendix-A.

As shown in Table 6 in Appendix-A, the piling sites to be able to measure all indices were only three (#1 - #3). This means that the number of datasets that contain observation data of all indices is limited and that the majority of the datasets are incomplete, lacking observation data for some indices. In particular, the amount of data obtained in loading tests regarding $f_i$ is much smaller than the amount of data on other indices. Construction information (piling data) and soil investigation data (i.e., $T$, $S$, and $N$) are always collected in the piling process and thus are included in many datasets. Additionally, in loading tests, the interval between strain gauges varies depending on the test; thus, the amount of data available per test differs from index to index. The database used in this study is incomplete with regard to some indices, which poses a challenge that needs to be addressed in conducting the analysis.

In addition, $f_i$ is evaluated with reference to Okahara et al. (1991). Specifically, the ultimate bearing capacity is defined as the vertical load that is observed when the settlement depth is 10% of the blade diameter of a RSP-pile in this study.

![Figure 3: Histograms of $\alpha f_i$ depending on the soil type](image_url)
2.2 Clean-up and screening of data

2.2.1 Discretization of continuous data corresponding to the observed depth

Piling data such as $T$ (pile-head torque) and $S$ (auger penetration depth) are regarded as continuous data because the interval between observations is short, whereas $N$ and $f_i$ are discrete data that are taken at a longer interval. Because the observation depth is not consistent for continuous data nor for discrete data, continuous data and discrete data need to be correlated in terms of the observation depth in order for us to correctly understand the relationships between indices.

For this purpose, continuous data taken from a certain range of depths are averaged so that each index is understood according to the observation depth of discrete data. Fig. 4 shows an example of observation data. Black lines show continuous data of each index, blue dots and red triangles respectively indicate discrete data in accordance with the observation depths of $N$ and $f_i$. In a standard penetration test, $N$-values are observed every 1m. Thus, as shown in Fig. 4 the mean value of the continuous data distributed in a range of $\pm 0.5$m of the observation depth of an $N$-value is used as the value corresponding to the observation depth of that $N$-value. Regarding the value of $f_i$, the mean value of observations taken in the range of depth between two strain gauges is used as the value corresponding to the observation depth of $f_i$. The depths of strain gauges are shown by solid gray lines.

2.2.2 Definition of the normalized pile-head torque $T'\prime$

On the assumption that $T$ is mainly generated by friction between the pile tip and the bottom surface of the borehole, $T$ is normalized by using the cube of the blade radius as follows.

The blade radius is $r_w$ (m), the facial pressure on the soil-pile interface of the pile tip is $\tilde{p}^{tip}$ (kN/m$^2$), and the friction coefficient of the soil-pile interface is $\mu_f$ (Fig. 1a of the manuscript)
and Fig. 1b). The differential torque \(dT\) acting on the derivative area of the pile tip \((ds)\) is expressed by Eq. (5).

\[
dT = r df = r \mu g \tilde{p}^{tip} ds = r^2 \mu g \tilde{p}^{tip} d\theta dr
\]

(5)

Where, the differential area of the pile tip \((ds)\) and the differential friction \((df)\) acting on that area. Thus, the torque \((T)\) applied to the pile-head is given by Eq. (6).

\[
T = \int_{r_w}^{r_w} \int_0^{2\pi} r^2 \mu g \tilde{p}^{tip} d\theta dr = \frac{2}{3} \pi \mu g \tilde{p}^{tip} r_w^3
\]

(6)

Then, \(r_w\), which is the pile structural parameter, is transferred to the left side as below.

\[
\frac{T}{r_w^3} = \frac{2}{3} \pi \mu g \tilde{p}^{tip} = C \mu g \tilde{p}^{tip} = T'\]

(7)

In this equation, \(C\) is a proportionality constant, \(T'\) is the normalized torque by using the cube the blade radius. Incidentally, \(f_i\) is given by Eq. (8).

\[
f_i = \mu g \tilde{p}^{side}
\]

(8)

Here, \(\tilde{p}^{side}\) is the facial pressure on the soil-pile interface of the pile side. From Eq. (7) and Eq. (8), the relationship between \(T'\) and \(f_i\) is estimated as a linear relationship as below.

\[
T' = C \frac{\tilde{p}^{tip}}{\tilde{p}^{side}} f_i = C' f_i
\]

(9)

Here, on the simple assumption that \(\tilde{p}^{tip} / \tilde{p}^{side}\) is constant according to the location (depth) like the earth pressure coefficient, the relationship between \(f_i\) and \(T'\) is considered to be in a linear relationship. Therefore, in this study, this normalized torque \(T'\) is used as the index for deriving the estimation equation of \(f_i\) (\(a_{f_i}\)).

### 2.2.3 Criteria for the data screening

Screening is applied to the database on the conditions that (1) \(T\) and \(S\) are non-negative, and (2) \(N\) is non-negative and less than 50. Regarding (1), some values fluctuate on a short cycle in the direction of depth, and some are negative. These values are regarded as observation errors and are excluded from the database. As for (2), it is probable that \(N\)-values observed as 50 were interpreted differently at various piling sites or by different measures. At one piling site, for example, all \(N\)-values observed as 50 or greater were recorded as 50. Thus, \(N\)-values recorded as 50 or greater are excluded, as these are not reliable data.

### 3 Study method

#### 3.1 Flow of the study

Ching and Phoon (2012, 2013, 2014) effectively solved various estimation problems by using multivariate normal distributions for the purpose of modeling multidimensional soil data. In Ching and Phoon (2012), by using Bayes’ theorem in a skillful manner, a multivariate estimation equation for undrained shear strength of clay was successfully derived from various empirical equations and databases. This example
Figure 5: The two estimation methods for $f_i$ in the flow of this study is similar to the work in this study, in that an incomplete dataset is used to derive a rational estimation equation. In Ching and Phoon (2013, 2014), diverse information on physical and strength tests of clay was used to create a database, and the validity of the authors’ modeling approach was suggested. Soil data do not necessarily follow a normal distribution. Ching and Phoon (2013, 2014) proved that by converting soil data distributions into multidimensional normal distributions, linear relationships between indices are obtained and that multidimensional data having complex non-linear relationships can be interpolated and estimated in a simple and rational manner.

Based on these previous studies, this study utilizes the aforementioned incomplete database in order to create a method for estimating $f_i$. The flow of this study is shown in Fig. 5. After a database is created, two methods for estimating $f_i$ are proposed.

3.2 Method for creating an estimation equation by utilizing the conditional normal distribution

As mentioned above, observations for all indices were taken at only three of the piling sites (#1 - #3.) The least squares method (LSM) in regression analysis that can be used for deriving an equation for estimating $f_i$ require that data be available for all indices that correspond to the variables in the equation. Thus, only the data collected at the three piling sites can be used for deriving an estimation equation based on LSM.

In view of the above, for the purpose of making the best possible use of the collected data and of proposing a rational and valid equation, a conditional multivariate normal distribution (Bishop 2006) was utilized in order to derive an estimation equation for $f_i$. The relationship between the conditional multivariate normal distribution and the linear model is explained as follows.

First, $x$ is defined by the $D$-dimensional random variable vector [Eq. (10)].

$$x = (x_1, x_2, \cdots, x_D)^T$$

(10)

When $x$ follows normal distribution, its probability density function is expressed by
Eq. (11) (i.e., a $D$-dimensional multivariate normal distribution).

$$f(x|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^D |\Sigma|}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}$$ (11)

where, $\mu$ and $\Sigma$ are parameters of the $D$-dimensional multivariate normal distribution. $\mu$ is a $D$-dimensional mean vector [Eq. (12)].

$$\mu = (\mu_1, \mu_2, \cdots, \mu_D)^T$$ (12)

$\Sigma$ is the covariance matrix for a $D \times D$ symmetric matrix in Eq. (13).

$$\Sigma = \begin{bmatrix} \sigma^2_1 & \sigma_{12} & \cdots & \sigma_{1D} \\ \sigma_{21} & \sigma^2_2 & \cdots & \sigma_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{D1} & \sigma_{D2} & \cdots & \sigma^2_D \end{bmatrix}$$ (13)

The covariance matrix ($\Sigma$) needs to be a positive definite matrix.

Next, $x$ is divided into the two subsets $x_1$ and $x_2$, as expressed in Eq. (14).

$$x = (x_1^T; x_2^T)^T$$ (14)

$x_1$ is an $M$-dimensional vector consisting of the elements from the first to the $M$-th elements (unknown parameters: herein objective variables), and $x_2$ is a $(D - M)$-dimensional vector consisting of the remaining elements (known parameters: herein explanatory variables). Based on these subsets, the mean vector ($\mu$) and the covariance matrix ($\Sigma$) are respectively defined by Eqs. (15) and (16).

$$\mu = (\mu_1^T; \mu_2^T)^T$$ (15)

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$ (16)

$\mu_1$ is an $M$-dimensional vector and $\mu_1$ is an $(D - M)$-dimensional vector. $\Sigma_{11}$ is an $M \times M$ matrix, $\Sigma_{12}$ is an $M \times (D - M)$ matrix, and $\Sigma_{22}$ is a $(D - M) \times (D - M)$ matrix. Because $\Sigma$ is a symmetric matrix, $\Sigma_{11}$ and $\Sigma_{22}$ are also symmetric matrices, and $\Sigma_{21} = \Sigma_{22}^T$. Under the condition that $x_2$, a subset of random variables, is given, $f(x_1|x_2)$, the conditional distribution of $x_1$, is considered to be as follows.

$$f(x_1|x_2, \mu_{1|2}, \Sigma_{1|2}) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_{1|2}|}} \exp \left\{ -\frac{1}{2} (x_1 - \mu_{1|2})^T \Sigma^{-1}_{1|2} (x_1 - \mu_{1|2}) \right\}$$ (17)

The mean vector $\mu_{1|2}$ and the covariance matrix $\Sigma_{1|2}$ are expressed by Eqs. (18) and (19), respectively.

$$\mu_{1|2} = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2)$$ (18)

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$ (19)

Eq. (18) is an example of the linear model defined by Eq. (20).

$$y = \beta_1 + \beta_2 x$$ (20)
By using the regression coefficients $\beta_1$ and $\beta_2$ in the linear model, Eq. (18) is expressed in the form of Eqs. (21) and (22).

\[
\text{Intercept} : \beta_1 = \mu_1 - \Sigma_{12} \Sigma_{22}^{-1} \mu_2 \quad (21)
\]

\[
\text{Gradient} : \beta_2 = \Sigma_{12} \Sigma_{22}^{-1} \quad (22)
\]

Thus, the creation of an estimation equation having $f_i$ as an objective variable is possible when the mean vector and the covariance matrix, the parameters of the multivariate normal distribution, can be estimated. In order to apply this idea for the creation of an estimation equation, because $x$ does not necessarily follow a normal distribution, the distribution of $x$ must be converted to a normal distribution based on Box-Cox transformation (Box and Cox 1964) or Johnson transformation (Johnson 1949).

### 3.3 Approximation of an estimated covariance matrix by a positive definite matrix

In order for a covariance matrix to be a parameter of multivariate normal distribution, the covariance matrix must be a positive definite matrix. However, an estimated covariance matrix is not always a positive definite matrix when the covariance matrix is calculated by using an incomplete dataset, or when each element of the covariance matrix is independently estimated for estimating the parameters, as in this study.

Thus, techniques for approximating an estimated covariance matrix by a positive definite matrix are explained below. First, assume that $\Sigma_{\text{est}}$ is the estimated $D \times D$ symmetric covariance matrix. Through eigenvalue decomposition, $\Sigma_{\text{est}}$ can be expressed by Eq. (23).

\[\Sigma_{\text{est}} = U \Lambda U^T\]  

(23)

$\Lambda$ is a diagonal matrix (Eq. (24)) in which the eigenvalues of $\Sigma_{\text{est}}$ are diagonal components when these eigenvalues are $\{\lambda_1, \cdots, \lambda_D\}$ in descending order of the values.

\[\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & \ddots \\ 0 & \ddots & \lambda_D \end{bmatrix}\]  

(24)

$U = (u_1, \cdots, u_D)$ is the orthogonal matrix in which unit eigenvectors $\{u_1, \cdots, u_D\}$ corresponding to the eigenvalues $\{\lambda_1, \cdots, \lambda_D\}$ make up columns.

The eigenvalues $\{\lambda_1, \cdots, \lambda_D\}$ are divided into a subset $\{\lambda_1, \cdots, \lambda_w\}$ in which each eigenvalue is greater than 0 and a subset $\{\lambda_{w+1}, \cdots, \lambda_D\}$ in which each eigenvalue is 0 or smaller. According to this division, $\Lambda$ can be expressed by Eq. (25).

\[\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ \vdots & \ddots & 0 \\ \lambda_w & \lambda_{w+1} \\ 0 & \ddots & \lambda_D \end{bmatrix}\]  

(25)
In $\Lambda$, eigenvalues of 0 or smaller $\{\lambda_{w+1}, \cdots, \lambda_D\}$ are replaced by positive real numbers ($\gamma$). $\Lambda'$ is defined by Eq. (26).

$$
\Lambda' = \begin{bmatrix}
\lambda_1 & & & \\
& \ddots & & \\
& & \lambda_w & \\
0 & & & \gamma
\end{bmatrix}
$$

(26)

By using $\Lambda'$ and $U$, a positive definite matrix $\Sigma_{est}'$ is defined by Eq. (27).

$$
\Sigma_{est}' = U\Lambda'U^T
$$

(27)

Because all eigenvalues are positive real numbers, $\Sigma_{est}'$ is a positive definite matrix. When the estimated variance-covariance matrix $\Sigma_{est}$ is not a positive definite matrix, the positive definite matrix $\Sigma_{est}'$ is used as an approximate matrix of $\Sigma_{est}$ for developing an estimation equation.

With the increase in the value of $\gamma$ that replaces a negative eigenvalue, the difference between the estimated $\Sigma_{est}$ and the approximated $\Sigma_{est}'$ increases. Thus, the value of $\gamma$ is determined through trial and error such that the value of $\gamma$ will not significantly affect the estimation accuracy of the estimation equation. In this paper, the value of $\gamma$ is 90% of the smallest eigenvalue among positive eigenvalues (Eq. (28)).

$$
\gamma = 0.90\lambda_w
$$

(28)

### 3.4 Methods for estimating a mean vector and a variance-covariance matrix

The estimation equation developed in this study has an objective variable ($f_i$) and explanatory variables $N$, $T'$, and $S$. The mean vector and the covariance matrix of the multivariate normal distribution for estimation are expressed by Eqs. (29) and (30).

$$
\mu = (\mu_{f_i}, \mu_N, \mu_{T'}, \mu_S)^T
$$

(29)

$$
\Sigma = \begin{bmatrix}
\sigma_{f_i}^2 & \sigma_{f_i,N} & \sigma_{f_i,T'} & \sigma_{f_i,S} \\
\sigma_{N,f_i} & \sigma_N^2 & \sigma_{N,T'} & \sigma_{N,S} \\
\sigma_{T',f_i} & \sigma_{T',N} & \sigma_{T'}^2 & \sigma_{T',S} \\
\sigma_{S,f_i} & \sigma_{S,N} & \sigma_{S,T'} & \sigma_S^2
\end{bmatrix}
$$

(30)

Two methods for estimating the mean vector and the covariance matrix are used, as explained below.

#### 3.4.1 Estimation Method 1

First, the database mentioned above is used for calculating the mean value, the standard deviation value, and the correlation coefficient of each index. On the basis of these values, the mean vector and the covariance matrix are created. Because the amount of data varies depending on the index, mean values and standard deviation values are calculated by using the data available for each index, and correlation coefficients are computed by pairwise elimination.
Fig. 6 outlines the calculation of correlation coefficients by pairwise elimination. In this method, as shown in Fig. 6, when there are missing data in each dataset of the variables $x_1$, $x_2$, and $x_3$, correlation coefficients are calculated by using simultaneously available observed data. For example, $\rho_{x_1x_2}$, the correlation coefficient for $x_1$ and $x_2$, is calculated by using the observed values in the dotted-line region of Fig. 6a, while $\rho_{x_2x_3}$, the correlation coefficient for $x_2$ and $x_3$, is calculated by using the observed values in the dotted-line region in Fig. 6b. Then, the mean value, the standard deviation value, and the correlation coefficient that are calculated for each index are used to create a mean vector and a covariance matrix.

### 3.4.2 Estimation Method 2

First, a prior distribution is created for the parameters of each index. The prior distribution is updated to a posterior distribution by utilizing Bayes’ theorem with a likelihood determined on the basis of observed data. The mean value of the posterior distribution is a representative value which is used as an estimated value of a parameter.

The basic idea behind using Bayes’ theorem to update an estimation equation is as follows. A parameter vector $\theta$ is a random variable, and the probability distribution of $\theta$ is $\pi(\theta)$. The probability distribution $\pi(\theta)$ is a prior distribution in Bayesian updating. The prior distribution $\pi(\theta)$ is updated to a posterior distribution $\pi(\theta|D)$ on the basis of a likelihood $f(D|\theta)$ that corresponds to observed data $D$. The posterior distribution $\pi(\theta|D)$ of the parameter $\theta$ is expressed by Eq. (31).

$$
\pi(\theta|D) = \frac{f(D|\theta)}{\int f(D|\theta)\pi(\theta)d\theta}\pi(\theta)
$$

(31)

The data $D$ used for updating are $\{N, T', S\}$ which are observed at piling sites. Because these three indices are assumed to follow normal distribution, the likelihood $f(D|\theta)$ is calculated by a trivariate normally distribution. The parameters for updating are the mean value ($\mu$), the standard deviation value ($\sigma$), and the correlation coefficient ($\rho$) for each of $\{N, T', S\}$; thus, the parameter vector ($\theta$) is a nine-dimensional
The parameters relating to \( f_i \), or \( f_i, f_i, f_i \), \( f_i, f_i \), \( f_i, f_i \), are calculated by using Estimation Method 1 above. It should be noted that Estimation Method 2 is interpreted as an application of empirical Bayes using the posterior mean of the estimated parameters.

### 3.5 Method for generating a prior distribution

From an engineering viewpoint, prior distributions are generated by the database for the following nine parameters which are elements of \( \theta \): mean values \( \mu_N, \mu_T, \mu_S \), standard deviation values \( \sigma_N, \sigma_T, \sigma_S \), and correlation coefficient \( \rho_{NT}, \rho_{TS}, \rho_{SN} \).

In generating a prior distribution, the bootstrap method (Efron and Tibshirani 1993) is used. In this method, a sample distribution of parameters is created by subjecting sample data to “random sampling with replacement”. In the example shown in Fig 7, bootstrap samples (1) to (B) \( \{V_1^{1*}, V_2^{1*}, \ldots, V_B^{1*}\} \) are created by applying random sampling with replacement \( B \) times, with possible duplication, to the sample data \( V_1^1 \) of Site 1.

\( \theta \) is one of the elements of the parameter vector \( \theta \). A sample parameter \( \hat{\theta}_1 \) of Site 1 is calculated for each bootstrap sample, and estimated parameters \( \{\hat{\theta}_1(V_1^{1*}), \hat{\theta}_1(V_2^{1*}), \ldots, \hat{\theta}_1(V_B^{1*})\} \) are obtained. This set of estimated parameters is a sample distribution of the estimated parameters for Site 1. The same procedure is followed for each piling site in order to create a sample distribution of the estimated parameters \( \hat{\theta} \) for each site. Then, all sample distributions of \( \hat{\theta} \) obtained at each site are synthesized into one sample distribution.

The operation iterated for each element of \( \theta \) to create nine synthetic sample distributions. Based on the variable range of each parameter, the parameter’s synthesized sample distribution are modeled by the probability density function shown in Table 3 by using a maximum likelihood method.

---

**Figure 7:** Calculation procedures for the prior distribution based on the bootstrap method

random vector [Eq. (32)].

\[
\theta = (\mu_N, \mu_T, \mu_S, \sigma_N, \sigma_T, \sigma_S, \rho_{NT}, \rho_{TS}, \rho_{SN})^T
\]  

(32)
Table 3: Probability density function for prior distributions (PDF)

| Elements of \( \theta \) | Range | PDF | Parameters of PDF |
|--------------------------|-------|-----|------------------|
| \( \mu_X \)              | Normal distribution | \( \mu, \sigma \) |
| \( \sigma_X > 0 \)       | Lognormal distribution | \( \mu, \sigma \) |
| \( \rho_{XY} \)          | \( -1 < \rho_{XY} < 1 \) | Four-parameter beta distribution | \( p, q \) |

The sample distribution is formulated and is used as a prior distribution (\( \pi(\theta) \)) of the parameter \( \theta \).

In here, the most characteristic point is that the beta distribution (Eq. (33)) is adopted as the probability density function of the prior distribution for the correlation coefficient, as shown in Table 3. Furthermore, as shown below, the variable domain is expanded and applied in order to correspond to the variable domain for the correlation coefficient (Four-parameter beta distribution).

\[
f(x) = \frac{x^{p-1}(1-x)^{q-1}}{B(p, q)}
\]

(33)

In Eq. (33), \( p \) and \( q \) are parameters of a beta distribution, \( p > 0 \) and \( q > 0 \). The condition of a variable range is \( 0 < x < 1 \). \( B(p, q) \) is a normalization constant that is defined by Eq. (34). This constant is used for obtaining a value of 1 by integration.

\[
B(p, q) = \int_0^1 x^{p-1}(1-x)^{q-1} dx
\]

(34)

In Four-parameter beta distribution, the variable range condition of a beta distribution expressed by Eq. (33) is extended from \( 0 < x < 1 \) to \( \alpha_1 < x < \alpha_2 \). The probability density function of the four-parameter beta distribution is expressed by Eq. (35).

\[
f(x) = \frac{(x - \alpha_1)^{p-1}(\alpha_2 - x)^{q-1}}{B(p, q)(\alpha_2 - \alpha_1)^{p-1}}
\]

(35)

Usually, the variable range of a correlation coefficient is \( -1 \leq \rho_{XY} \leq 1 \). But in this study, it is assumed that indices do not perfectly correlate. Therefore, \( \alpha_1 = -1 \) and \( \alpha_2 = 1 \) are given to Eq. (35), and the probability density function of the four-parameter beta distribution, in which \( 0 < x < 1 \) is extended to \( \alpha_1 < x < \alpha_2 \), is used as the prior distribution of a correlation coefficient.

3.6 Method for updating a prior parameter distribution based on observation data

A prior parameter distribution is updated to a posterior parameter distribution by means of particle filtering (Ristic et al. 2004). Particle filtering is a Sequential Monte Carlo method (Doucet et al. 2001). In this method, a posterior distribution is estimated on the basis of observation data by using Bayes’ theorem. As mentioned above, the observation data are the observed values of the three indices \( \{N, T', S_g\} \).

These observation data are defined as a data matrix \( D \) in Eq. (31). Posterior distribution \( \pi(\theta|D) \) is estimated by resampling from the prior distribution \( \pi(\theta) \), with the
likelihood \( f(D|\theta) \) obtained the data matrix \( D \) as the weight. The prior distribution \( \pi(\theta) \) is updated at an interval of 5m depth in order to obtain a posterior distribution \( \pi(\theta|D) \).

The mean value of the posterior distribution is used as a new parameter of the multivariate normal distribution, and a sequential estimation equation is updated by using Eq. (18) for estimating \( f_i \). The updating interval of 5m depth was chosen by trial and error in order to obtain a desirable estimate.

Points to note are explained below regarding the representative values of estimates and the methods for updating posterior distributions. Strictly speaking, the updating of a prior distribution \( \pi(\theta) \) needs to be done by using the likelihood \( f(D|\theta) \) based on a trivariate normal distribution. However, this method results in a large number of parameters whose distributions need to be updated at one time, which makes it difficult to estimate posterior distributions by particle filtering. Therefore, we simplify this issue by updating the bivariate normal distribution for each combination of two parameters (i.e., \( N - S, N - T' \) and \( S - T' \)). This simplified method has the problem on the updating operation regarding that two different posterior distributions are obtained for the mean values \( \{\mu_N, \mu_T, \mu_S\} \) and standard deviations \( \{\sigma_N, \sigma_T, \sigma_S\} \), which are subelements of parameter \( \theta \). The mean values of the representative values (i.e., the mean values and standard deviations) of these two distributions are used for creating the estimation equation for each depth.

The updating method is based on the simple assumptions mentioned above. Additionally, particle filtering requires large periods of time for processing reliability updating and is not suited to the real-time estimation of reliability. Because this study is a basic study on the validity of reliability updating in which real-time piling data are used, improvements in the strictness of updating techniques and in the calculation cost will be addressed in future studies.

4 Study results

4.1 Data processing results

Fig. 8 shows pair scatterplots, histograms, correlation coefficient and normal Q-Q plots after data for each index are processed. Data processing means normalization of pile-head torque \( T \) and screening. In these pair scatterplots, the mean value of each index is shown in red.

The histograms show that each index is distributed asymmetrically and extends toward the right. In the Q-Q plots, none of the indices follows a normal distribution. To acquire the data followed normal distribution, logarithmic conversion is applied to each index because this is a simple technique for transforming variables as well as for interpreting transformed variables. Fig. 9 shows pair scatterplots, histograms, correlation coefficient and normal Q-Q plots for each index that was subjected to logarithmic conversion. It is shown in Fig. 9 that the distributions of indices after logarithmic conversion are mostly similar to normal distributions. The pair scatterplots show that a joint distribution of two indices has the mean value in the center of the distribution, which suggests that the joint distribution is also similar in shape to a normal distribution after the indices are subjected to logarithmic conversion. Although there are various conversion methods to normal distribution, we adopt logarithmic conversion in consideration with the small number of data. It is one of the reasons that the equation can be transformed to a simple equation that can be easily understood by the engineers,
as shown in Eq. (36).

$$\ln f_i = \ln(\beta_0) + \beta_N \ln N + \beta_{T'} \ln T' + \beta_S \ln S$$  \hspace{1cm} (36)

Where, $\beta_0$, $\beta_N$, $\beta_{T'}$ and $\beta_S$ are regression coefficients. Note that, considering the subsequent exponential transformation [Eq. (39)], we denote the constant part as $\ln(\beta_0)$.

To create the estimation equation, a mean vector [Eq. (40)] and a covariance matrix [Eq. (41)], which are parameters of each index to which logarithmic conversion was applied, are estimated.

$$\mu = \left(\mu_{\ln f_i}, \mu_{\ln N}, \mu_{\ln T'}, \mu_{\ln S}\right)^T$$  \hspace{1cm} (37)

$$\Sigma = \begin{bmatrix}
\sigma_{\ln f_i}^2 & \sigma_{\ln f_i \ln N} & \sigma_{\ln f_i \ln T'} & \sigma_{\ln f_i \ln S} \\
\sigma_{\ln N \ln f_i} & \sigma_{\ln N}^2 & \sigma_{\ln N \ln T'} & \sigma_{\ln N \ln S} \\
\sigma_{\ln T' \ln f_i} & \sigma_{\ln T'} & \sigma_{\ln T'}^2 & \sigma_{\ln T' \ln S} \\
\sigma_{\ln S \ln f_i} & \sigma_{\ln S \ln N} & \sigma_{\ln S \ln T'} & \sigma_{\ln S}^2
\end{bmatrix}$$  \hspace{1cm} (38)

Eq. (36) becomes Eq. (39) when the former is converted to a linear axis. The simplicity of Eq. (39) is one of the reasons for using logarithmic conversion to acquire normal distributions of indices.

$$f_i = \beta_0 \times N^{\beta_N} \times T'^{\beta_{T'}} \times S^{\beta_S}$$  \hspace{1cm} (39)
Figure 9: Characteristics of the database after conversion (pair scatterplot, histogram, correlation coefficient and normal Q-Q plot)
4.2 Estimation results obtained by Estimation Method 1

First, the results of estimation by Estimation Method 1 are explained below regarding the equation for estimating $f_i$. The mean vector and the covariance matrix that are estimated by using the database are shown by Eqs. (40) and (41).

\[ \mu = (4.09, 1.42, 3.80, -1.85)^T \]  \hspace{1cm} (40)

\[ \Sigma = \begin{bmatrix}
0.48 & 0.45 & 0.39 & -0.03 \\
0.45 & 1.63 & 0.65 & -0.16 \\
0.39 & 0.65 & 0.58 & -0.10 \\
-0.03 & -0.16 & -0.10 & 0.16
\end{bmatrix} \]  \hspace{1cm} (41)

By using Eqs. (40) and (41), Estimation Eqs. (42) and (43) are derived on the basis of Eqs. (18) and (19).

\[ \ln f_i = 1.90 + 0.02 \ln N + 0.70 \ln T' + 0.28 \ln S \]  \hspace{1cm} (42)

\[ \tilde{\sigma}_{\ln f_i} = 0.45 \]  \hspace{1cm} (43)

Where, $\tilde{\sigma}_{\ln f_i}$ represents the estimation error on the derived equation calculated by Eq. (19). The coefficient of variation $COV_{f_i}$ is given by Eq. (44).

\[ COV_{f_i} = \sqrt{\exp(\tilde{\sigma}_{\ln f_i}^2) - 1} = 0.47 \]  \hspace{1cm} (44)

Fig. 10 shows the results of estimation conducted by using Estimation Method 1 for $f_i$. Red dots indicate the values of $f_i$ that were observed in loading tests, and black lines are the values of $f_i$ that are estimated by using Eq. (42), which was created by means of Estimation Method 1. Although the estimated values of $f_i$ show the characteristics of observed values, these estimated values seem to be relatively heavily biased.

4.3 Estimation results obtained by Estimation Method 2

In estimation method 2, $x = (\ln f_i, \ln N, \ln T', \ln S)^T$ obtained based on logarithmic transformation to each index is also used as an explanatory variable following the multivariate normal distribution. Thus, the parameter vector $\theta$ for creating the prior distribution $\pi(\theta)$ is Eq. (45).

\[ \theta = (\mu_{\ln N}, \mu_{\ln T'}, \mu_{\ln S}, \sigma_{\ln N}, \sigma_{\ln T'}, \sigma_{\ln S}, \rho_{\ln N \ln T'}, \rho_{\ln T' \ln S}, \rho_{\ln S \ln N})^T \]  \hspace{1cm} (45)

Fig. 11 shows the histogram of each index generated by means of the bootstrap method by using the data collected at piling sites. The data are contained in the database that is used for modeling. Parameters of prior distributions are identified by using a maximum likelihood method. Identified parameters of prior distributions are shown in Table 4. In Fig. 11, generated prior distributions are also superimposed (black line). Table 5 shows a regression coefficient $\{\beta_0, \beta_N, \beta_T, \beta_S\}$, a estimation error of an estimate on logarithmic axis $\tilde{\sigma}_{\ln f_i}$, and a coefficient of variance $COV_{f_i}$ for each depth of estimation in the $f_i$ estimation equation created by using Estimation Method 2. Because the estimation equation is updated for every 5m of depth, the results are shown for each 5-m section.
Figure 10: Estimation results calculated by Estimation Method 1

Figure 11: Prior distributions generated by the bootstrap method with the original database data
Table 4: Identified parameters of prior distributions

| Elements of $\theta$ | PDF                | Parameters of PDF |
|----------------------|--------------------|-------------------|
| $\mu_{\ln N}$        | Normal distribution | $\mu = 1.92$      |
| $\mu_{\ln T'}$       |                    | $\mu = 3.96$      |
| $\mu_{\ln S}$        |                    | $\mu = -1.73$     |
| $\sigma_{\ln N}$     | Lognormal distribution | $\mu = 0.01$  |
| $\sigma_{\ln T'}$    |                    | $\mu = -0.56$     |
| $\sigma_{\ln S}$     |                    | $\mu = -1.34$     |
| $\rho_{\ln N \ln T'}$| Four-parameter      | $p = 10.23$       |
| $\rho_{\ln T' \ln S}$| beta distribution  | $p = 1.55$        |
| $\rho_{\ln S \ln N}$ |                    | $p = 1.72$        |

Table 5: Regression coefficient and the accuracy of the regression equation

| site | Depth (m) | $\beta_N$ | $\beta_{T'}$ | $\beta_S$ | $\beta_0$ | $\sigma_{\ln f_i}$ | $COV_{f_i}$ |
|------|-----------|-----------|--------------|-----------|-----------|--------------------|-------------|
| 1    | 0-5       | 0.21      | 0.86         | -0.21     | 0.73      | 0.39               | 0.40        |
|      | 5-10      | 0.29      | 0.40         | -0.13     | 2.62      | 0.50               | 0.54        |
|      | 11-15     | 0.23      | 0.30         | -0.11     | 6.18      | 0.55               | 0.60        |
| 2    | 0-5       | 0.23      | 0.69         | -0.24     | 1.97      | 0.45               | 0.48        |
|      | 5-10      | 0.30      | 0.53         | -0.22     | 2.74      | 0.48               | 0.50        |
|      | 11-15     | 0.50      | 0.45         | -0.16     | 1.56      | 0.44               | 0.46        |
|      | 16-20     | 0.42      | 0.35         | -0.11     | 2.93      | 0.48               | 0.51        |
|      | 21-25     | 0.34      | 0.66         | -0.22     | 0.89      | 0.41               | 0.43        |
|      | 26-30     | 0.41      | 0.47         | -0.20     | 2.18      | 0.46               | 0.49        |
|      | 31-35     | 0.36      | 0.78         | -0.25     | 0.53      | 0.37               | 0.39        |
|      | 36-40     | 0.30      | 0.45         | -0.22     | 1.65      | 0.51               | 0.55        |
|      | 41-45     | 0.38      | 0.33         | -0.17     | 2.45      | 0.55               | 0.59        |
|      | 46-50     | 0.23      | 0.33         | -0.13     | 5.43      | 0.54               | 0.59        |
| 3    | 0-5       | 0.32      | 0.46         | -0.17     | 8.37      | 0.48               | 0.51        |
|      | 5-10      | 0.24      | 0.48         | -0.23     | 8.64      | 0.52               | 0.55        |
|      | 11-15     | 0.26      | 0.34         | -0.10     | 7.28      | 0.53               | 0.56        |
|      | 16-20     | 0.27      | 0.40         | -0.16     | 3.06      | 0.51               | 0.55        |
It should be noted that the proposed method does not necessarily reduce the uncertainty of the estimation equation $\sigma_{\ln f_i}$. Unlike in a general direct estimation for coefficients of an estimation equation, the reason is because the proposed method is to be the estimation method for parameters of the multi-dimensional normal distribution taking into account the characteristics for each depth. In addition, in the proposed method at present, the estimation error in these parameter themselves estimations are not take into account in calculating of $\sigma_{\ln f_i}$.

Fig. 12 shows the estimates of $f_i$ that are estimated by using the estimation equation created by means of Estimation Method 2. Red dots indicate observed values of $f_i$ that were obtained in vertical loading tests. Green lines show the results estimated by using prior distributions of parameters. Blue lines show estimates $f_i$ obtained by updating the estimation equation through Bayesian updating. Light blue regions indicate estimation error ranges for Estimation Method 2. Here, the estimation error ranges in Fig. 12 show the ranges simply conducted exponential transformation for $\ln f_i \pm \sigma_{\ln f_i}$ of an estimate on the logarithmic axis. For every piling site, the results of Estimation Method 2 better close to the values observed in vertical loading tests than do the results of Estimation Method 1. Therefore, by using Estimation Method 2, which is a method that is based on reliability updating that utilizes real-time piling data, bearing capacity can be estimated in real-time at the same time an RPS-pile is being placed. Furthermore, data for computing reliability analysis can also be acquired.
5 Conclusions and future issues

The study described in this paper aims at developing a method for real-time updating and verification of the bearing capacity reliability of RSP-piles. For that purpose, piling data observed onsite were effectively utilized. First, vertical loading test data of RSP-piles, soil investigation data, and piling data were collected for the purpose of creating a database consisting of observed values of side resistance \( f_i \), \( N \)-values obtained in standard penetration tests, and piling data (i.e., pile-head torque \( T \) and auger penetration depth per revolution \( S )\).

The database was incomplete because the available vertical loading test data were limited and a majority of datasets did not include observation data for some indices. Thus, the utilization of the incomplete database was a central topic of the study before an equation was derived for estimating the ultimate bearing capacity by effectively using piling data. Regarding this topic, a conditional multivariate normal distribution was used for testing the following two estimation methods, and the estimation results were examined.

1. Estimation Method 1: A pairwise elimination method was applied to sample data in order to set a mean vector and a covariance matrix. Then an estimation equation for estimating \( f_i \) was derived on the basis of \( N \), \( T' \), and \( S \) by using conditional multivariate normal distribution.

2. Estimation Method 2: Prior distributions of parameter vectors were estimated. A prior distribution was created by using the bootstrap method for each site. In this method, a reliability estimate of ultimate bearing capacity is updated in real time on the basis of Bayes’ theorem by using piling data that are observed at depth intervals of 5m.

This suggests that real-time updating of the reliability of bearing capacity estimates onsite is possible. The practical application of the estimation method verified above will be investigated in the future. For that purpose, it is necessary to create more accurate estimation equations and prior distributions, and the database needs to be expanded towards ensuring the validity of the analyses of the estimation results. Studies will be also conducted from the viewpoint of geotechnical engineering. Specifically, statistical methods such as principal component analysis will be utilized for analyzing correlations.

A Data used for the analyses

For the purpose of indicating the reliability of the statistical analyses described in this paper, the dimensions of the rotary steel pipe pile used for analyses are listed in Tab. 6. Regarding the datasets that include piling data, a depth distribution of each index is shown in Fig. 13.

B Changes in simultaneous probability density resulting from updating by using piling data (Estimation Method 2)

The diagrams in Fig. 14 show changes in the probability density of each variable. In Estimation Method 2, these changes are caused by updates to the parameter distribu-
tions by means of Bayes’ theorem, with the likelihoods determined on the basis of piling data. Due to space constraints, changes in the probability density are shown only for Site 1. The estimation equation is updated by using a conditional distribution. For this purpose, the probability density (shown in blue) according to the estimated posterior distribution of each parameter is used.

References

American Association of State Highway and Transportation Officials (AASHTO) (1989), Guide Specifications for Strength Evaluation of Existing Steel and Concrete Bridges, American Association of State Highway and Transportation Officials, Washington, DC.

American Association of State Highway and Transportation Officials (AASHTO) (2000), Manual for Condition Evaluation of Bridges, 2 edn, American Association of State Highway and Transportation Officials, Washington, DC.

American Association of State Highway and Transportation Officials (AASHTO) (2004), AASHTO LRFD Bridge Design Specifications, 3 edn, American Association of State Highway and Transportation Officials, Washington, DC.

Architectural Institute of Japan (AIJ) (2000), Recommendations for Limit State Design of Buildings, Architectural Institute of Japan.

Bishop, C. M. (2006), Pattern Recognition and Machine Learning, Springer.

Box, G. E. P. and Cox, D. R. (1964), ‘An analysis of transformations’, Journal of the Royal Statistical Society, Series B 26(2), 211–252.

Ching, J. and Phoon, K.-K. (2012), ‘Modeling parameters of structured clays as a multivariate normal distribution’, Canadian Geotechnical Journal 49(5), 522–545.

Ching, J. and Phoon, K.-K. (2013), ‘Multivariate distribution for undrained shear strengths under various test procedures’, Canadian Geotechnical Journal 50(9), 907–923.

Ching, J. and Phoon, K.-K. (2014), ‘Correlations among some clay parameters—the multivariate distribution’, Canadian Geotechnical Journal 51(6), 686–704.

Doucet, A., de Freitas, N. and Gordon, N. (2001), Sequential Monte Carlo Methods in Practice, Springer.

Efron, B. and Tibshirani, R. J. (1993), ‘An introduction to the bootstrap’, Chapman and Hall.

Japan Institute of Country-ology and Engineering (JICE) (2004), Construction Technology Review and Certification of Screwed Steel Pile [NS ECO-PILE], Japan Institute of Country-ology and Engineering, Japan.

Japan Road Association (JRA) (1996), Specifications for Highway Bridges Part IV, Sub-structures (SHB), Maruzen Publications.

Johnson, N. L. (1949), ‘Bivariate distributions based on simple translation systems’, Biometrika 36(3/4), 297–304.
Kawai, M., Ichikawa, K. and Kono, K. (2018), Development of new type of screwed pile with large bearing capacity and ecological driving method “tsubasa pile tm ”, in ‘Proceedings of the 4th CIGOS’, Vol. 8 of Lecture Notes in Civil Engineering, Springer, pp. 411–425.

Mori, G. (2003), Development of the screw steel pipe pile with toe wing “tsubasa pile”, in W. F. V. Impe, ed., ‘Proceedings of the 4th International Geotechnical Seminar on Deep Foundations on Bored and Auger Piles’, Vol. 171-176, Millpress, Rotterdam, Netherlands.

Okahara, M., Takagi, S., Nakatani, S. and Kimura, Y. (1991), ‘A study on bearing capacity of a single pile and design method of cylinder shaped foundations’, Technical Memorandum of PWRI.

Paikowsky, S. G. (2004), Load and resistance factor design (LRFD) for deep foundations, The National Academies Press, Washington, DC.

Public Works Research Center (PWRC) (2009), Construction Technology Review and Certification of Rotary Penetration Steel Pipe Pile with Toe Wing [Geowing Pile II], Public Works Research Center, Japan.

Public Works Research Center (PWRC) (2011), Construction Technology Review and Certification of Rotary Penetration Steel Pipe Pile with Toe Wing [Tsubasa Pile], Public Works Research Center, Japan.

Ristic, B., Arulampalam, S. and Gordon, N. (2004), Beyond the Kalman Filter: Particle Filters for Tracking Applications Particle Filters for Tracking Applications, Artech House.

Saeki, E. and Ohki, H. (2003), A study of the screwed pile—the results of installation and roading tests and analysis of penetration mechanisms, in W. F. V. Impe, ed., ‘Proceedings of the 4th International Geotechnical Seminar on Deep Foundation on Bored and Auger Piles’, Millpress, Rotterdam, Netherlands, pp. 259–266.

Shioi, Y., Oowada, A. and Sawaishi, M. (2013), Reinforcement method to reduce settlement for existing pile foundations on soft ground, in ‘Proceedings of the Second International Conference-Geotechnics for Sustainable Development (Geotec Hanoi 2013)’.

Tang, C. and Phoon, K.-K. (2018), ‘Statistics of model factors in reliability-based design of axially loaded driven piles in sand’, Canadian Geotechnical Journal 55(11), 1592–1610.

Tang, C. and Phoon, K.-K. (2019), ‘Characterization of model uncertainty in predicting axial resistance of piles driven into clay’, Canadian Geotechnical Journal. accepted.

The overseas coastal area development institute of Japan (OCDI) (2009), Technical standards and commentaries for port and harbour facilities in Japan, The overseas coastal area development institute of Japan, Japan.
Table 6: List of the rotary penetration steel pile dimensions and indices in the database

| Site No | Pile No | Pile diameter $D_p$ (mm) | Blade diameter $D_w$ (mm) | $D_D / D_p$ | Length $L$ (m) | $T$ (kN•m) | $S$ (m) | $N$ | Soil type | $f_s$ (kN/m²) |
|---------|--------|------------------------|------------------------|-------------|----------------|----------|--------|-----|-----------|--------------|
| 1       | 1      | 1200                   | 1800                   | 1.5         | 13.0           | □        | □      | □   | □         | □            |
| 2       | 2      | 1000                   | 1500                   | 1.5         | 46.3           | □        | □      | □   | □         | □            |
| 3       | 3      | 1600                   | 2400                   | 1.5         | 17.0           | □        | □      | □   | □         | □            |
| 4       | 4-A    | 1100                   | 1650                   | 1.5         | 36.0           | □        | □      | □   | □         | □            |
|         | 4-B    | 1100                   | 1650                   | 1.5         | 36.0           | □        | □      | □   | □         | □            |
|         | 4-C    | 1100                   | 1650                   | 1.5         | 35.0           | □        | □      | □   | □         | □            |
|         | 4-D    | 1100                   | 1650                   | 1.5         | 35.5           | □        | □      | □   | □         | □            |
| 5       | 5-A    | 1000                   | 1500                   | 1.5         | 49.0           | □        | □      | □   | □         | □            |
|         | 5-B    | 1000                   | 1500                   | 1.5         | 49.0           | □        | □      | □   | □         | □            |
|         | 5-C    | 1000                   | 1500                   | 1.5         | 49.0           | □        | □      | □   | □         | □            |
|         | 5-D    | 1000                   | 1500                   | 1.5         | 49.0           | □        | □      | □   | □         | □            |
|         | 5-E    | 1000                   | 1500                   | 1.5         | 49.0           | □        | □      | □   | □         | □            |
| 6       | 6-A    | 800                    | 1200                   | 1.5         | 15.7           | □        | □      | □   | □         | □            |
|         | 6-B    | 609.6                  | 914.4                  | 1.5         | 12.6           | □        | □      | □   | □         | □            |
|         | 6-C    | 609.6                  | 914.4                  | 1.5         | 14.7           | □        | □      | □   | □         | □            |
|         | 6-D    | 900                    | 1350                   | 1.5         | 13.7           | □        | □      | □   | □         | □            |
| 7       | 7      | 800                    | 1200                   | 1.5         | 50.3           | □        | □      | □   | □         | □            |
| 8       | 8      | 900                    | 1350                   | 1.5         | 55.2           | □        | □      | □   | □         | □            |
| 9       | 9      | 800                    | 1200                   | 1.5         | 18.7           | □        | □      | □   | □         | □            |
| 10      | 10     | 1200                   | 1800                   | 1.5         | 32.6           | □        | □      | □   | □         | □            |
| 11      | 11     | 1600                   | 2400                   | 1.5         | 51.0           | □        | □      | □   | □         | □            |
| 12      | 12     | 600                    | 1200                   | 2.0         | 35.7           | □        | □      | □   | □         | □            |
| 13      | 13     | 1200                   | 2400                   | 2.0         | 76.8           | □        | □      | □   | □         | □            |
| 14      | 14     | 800                    | 1200                   | 1.5         | 57.7           | □        | □      | □   | □         | □            |
| 15      | 15     | 900                    | 1350                   | 1.5         | 51.6           | □        | □      | □   | □         | □            |
| 16      | 16     | 800                    | 1200                   | 1.5         | 25.1           | □        | □      | □   | □         | □            |
| 17      | 17     | 500                    | 1000                   | 2.0         | 15.0           | □        | □      | □   | □         | □            |
| 18      | 18     | 508                    | 1016                   | 2.0         | 13.0           | □        | □      | □   | □         | □            |
| 19      | 19     | 508                    | 1016                   | 2.0         | 21.0           | □        | □      | □   | □         | □            |
| 20      | 20     | 508                    | 1016                   | 2.0         | 11.9           | □        | □      | □   | □         | □            |
| 21      | 21-A   | 318.5                  | 637                    | 2.0         | 19.9           | □        | □      | □   | □         | □            |
|         | 21-B   | 318.5                  | 637                    | 2.0         | 25.9           | □        | □      | □   | □         | □            |
| 22      | 22-A   | 318.5                  | 637                    | 2.0         | 22.9           | □        | □      | □   | □         | □            |
|         | 22-B   | 318.5                  | 637                    | 2.0         | 31.9           | □        | □      | □   | □         | □            |
| 23      | 23     | 318.5                  | 637                    | 2.0         | 36.9           | □        | □      | □   | □         | □            |
| 24      | 24     | 508                    | 1016                   | 2.0         | 56.7           | □        | □      | □   | □         | □            |
| 25      | 25     | 508/800                | 1016                   | 2.0         | 52.0           | □        | □      | □   | □         | □            |
| 26      | 26-A   | 1200                   | 2400                   | 2.0         | 61.0           | □        | □      | □   | □         | □            |
|         | 26-B   | 800                    | 1600                   | 2.0         | 60.6           | □        | □      | □   | □         | □            |
| 27      | 27     | 800                    | 1600                   | 2.0         | 43.0           | □        | □      | □   | □         | □            |
Figure 13: Samples of data observed at the sites
Figure 14: Transition of multi-dimensional normal distribution corresponding to each depth at site1 (Estimation Method 2)