Proc. of the 6th Intl. Symposium on Reliability Engineering and Risk Management (6ISRERM)
31 May – 1 June 2018, Singapore
Editor(s) Xudong Qian, Sze Dai Pang, Ghim Ping Raymond Ong, Kok-Kwang Phoon
Copyright © 2018 Author(s). All rights reserved.

RELIABILITY ASSESSMENT FOR PILE FOUNDATIONS USING CONSTRUCTION INFORMATION BASED ON INCOMPLETE DATA

YU OTAKE[1] and SHINYA WATANABE[1]

[1]*Department, of Civil Engineering and architecture, NIIGATA UNIVERSITY*
E-mail: y_ohtake@eng.niigata-u.ac.jp

In recent years, problems related to pile foundation construction quality (i.e. inclination and unequal settlement after completion) have frequently been reported. Consequently, we have developed a method for real-time reliability evaluations of pile bearing capacity that take into account construction conditions by using construction information such as extent of penetration, rotational torque, and so on. This study focuses on rotary penetration steel piles, in which the rotation torque and penetration amount during rotation can be observed, to try to develop a real-time reliability updating system. The challenges in this study are to model the relationship with loading tests for pile structures, SPT-N values, rotation torque and extent of penetration per rotation, with consideration for any differences in the amount of data for each index (i.e. partial lack of data). Finally, the effectiveness of the proposed method was demonstrated based on application to a location for which full data was held.

Keywords: *Reliability analysis*, Construction information, Incomplete data, Bayes' theorem

1 Introduction

In recent years, the importance of utilizing information in construction obtained during the construction process has been increasing (Yu(2015)). This research focuses on the rotary penetration steel pile (RPS-pile), in which the rotational torque and extent of penetration per rotation can be observed, to try to develop a real-time reliability updating system. Specifically, we developed a method to confirm and update construction quality sequentially using not only the N value, but also construction information (i.e. rotational torque and extent of penetration) based on extensive data. The challenges in this study are to model a relationship with loading tests for pile structures, Standard Penetration Test (SPT)-N values, rotational torque, and extent of penetration per rotation, with due consideration for any differences in the amount of data for each index (i.e. partial lack of data).

2 Data base

2.1 *Outline*

The data used in this study consists of vertical loading test data, information during construction, and ground survey data (i.e. SPT-N values). Specifically, the pile head torque T (kNm), extent of penetration S (m) per pile revolution, N value, soil classification (sand/clay/gravel), and maximum skin-friction force f_i (kN/m^2) has been collected. The list of indices measured at each site is shown in Table 1.

In addition, construction information is observed as continuous data against depth, and SPT-N values are basically measured at one-meter intervals. On the other hand, since the strain gauge for measuring the maximum skin-friction force is discretely installed on a pile, the number of observed data measurements originally is small.

Table 1. List of Indices

Source of data	Item	Symbol	unit
Loading test	Maximum skin-friction force	f_i	kN/m^2
Soil investigation	SPT-N value	N	-
	Soil classification	-	-
Construction information	Pile head torque	T	kNm
	Extent of penetration	S	m

Note here all indices were measured from only 13 sets on three sites. Consequently, the amount of f_i data in the database is extremely small, while the number of data of indices such as T and N is substantial. In this research, how to overcome the problem of insufficient data and to create an estimation formula is a problem for statistical analysis.

• If corresponding with SPT-N value: Since the SPT-N value is measured at one-meter intervals, the average value included within the range of ± 0.5 m from the measurement point of the SPT-N value.

• If corresponding with fi: Since fi is calculated at the location of the strain gauges, the average value included within the intervals of gauges (Figure. 1(a)).

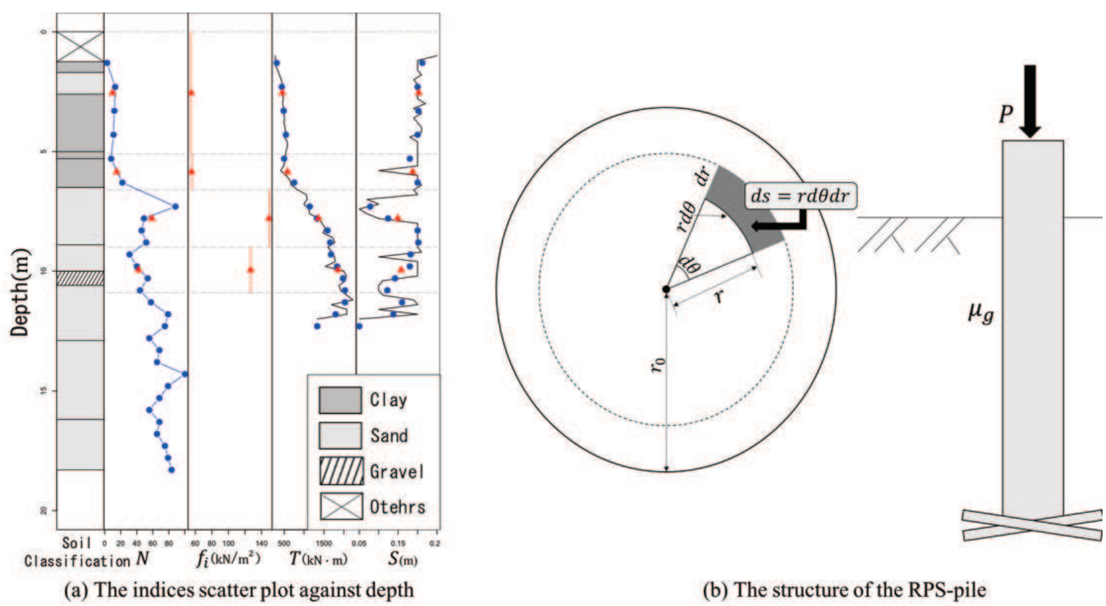

Figure 1. The indices scatter plot against depth and the structure of the RPS-pile

In Figure 1(a), the gray line is the continuous value for each piece of construction information, while the blue circle and the red triangle indicate the discrete values of the indices corresponding to the N value and the measurement depth of fi, respectively. In the vertical loading test, f_i adopts the measurement value when the settlement of the rotating pile reaches 10% of the pile tip diameter. Herein, if the SPT-N value exceeds 50, we decided not to include the data set in this study.

Furthermore, T is normalized by the cube of the tip radius based on the assumption that T is mainly caused by the frictional force between the bottom of the blade installed at the tip of the pile and the ground.

The minute pile head torque dT is calculated by integration of the bottom surface of the blade:

$$dT = \int_0^{r_0} \int_0^{2\pi} r^2 \mu_g P d\theta dr = \frac{2}{3}\pi\mu P r_0^3 \tag{1}$$

From the above, the relationship with the bottom skin-friction F and T can be described as follows:

$$F = \mu_g P = \frac{3}{2\pi} \times \frac{T}{r_0^3} = C \times \frac{T}{r_0^3} \tag{2}$$

From Equation. (2), it is considered that F is strongly affected by r_0. Consequently, T' is defined as the normalized value by the cube of r_0.

2.2 *Derived flow of estimation equation*

The estimation equation is derived based on the concept of conditional distribution of multivariate normal distribution and used not only datasets including all indices, but also all datasets including those partially lacking indices. The relationship between conditional distribution of multivariate normal distribution and linear regression analysis is expressed as follows(Jianye and Kok-Kwang(2010, 2012)):

$$\boldsymbol{x} = (X_1, X_2 \cdots X_D)^T \tag{3}$$

Where $\boldsymbol{x}$ is set as the stochastic variable vector, and the mean value vector and co-variable matrix are described as follows:

$$\boldsymbol{\mu} = \{\mu_1, \mu_2, \cdots \mu_D\}^T \quad \boldsymbol{\Sigma} = \{\sigma_{ij}\}\ i, j = 1,2 \cdots, D \tag{4}$$

Note here that $\boldsymbol{\Sigma}$ must be a symmetric and nonspecific matrix. Then, $\boldsymbol{x}$ is denoted as a decoupled vector.

$$\boldsymbol{x} = [\boldsymbol{x}_1, \boldsymbol{x}_2]^T \tag{5}$$

Where $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, are partial vectors of $\boldsymbol{x}_1$ is $M \times 1$, $\boldsymbol{x}_2$ is $D - M$ respectively. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, are also separated respectively in relation to $\boldsymbol{x}$.

$$\boldsymbol{\mu} = [\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T]^T \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \tag{6}$$

Where, $\boldsymbol{\Sigma}_{11}$ is $M \times M$, $\boldsymbol{\Sigma}_{12}$ is $M \times (D - M)$, $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}^T$, $\boldsymbol{\Sigma}_{22}$ is $(D - M) \times (D - M)$ respectively, and $\boldsymbol{\Sigma}_{22}$ is a nonspecific matrix, then conditional distribution $f(\boldsymbol{x}_1 | \boldsymbol{x}_2)$ is also normal distribution, and it's expectation is denoted as follows:

$$E[\boldsymbol{x}_1 | \boldsymbol{x}_2] = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2) \tag{7}$$

The equation is equivalent to multiple regression analysis, and the intercept and gradient coefficients are expressed as follows:

$$\text{Intercept: } \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2 \qquad \text{Gradient coefficients: } \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \tag{8}$$

Furthermore, valiance of the estimation equation is expressed as follows:

$$\text{var}[\boldsymbol{x}_1 | \boldsymbol{x}_2] = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \tag{9}$$

From the above, the estimation equation can be derived using the concept of conditional distribution of multivariate normal distribution if the mean value vector and co-variance matrix of the dataset is computed. However, since the probability distribution of each index is assumed to have a normal distribution, the probability distribution for each index collected in this study must be converted to normal distribution using Box-Cox transformation or log transformation.

3 Updating method of regression equation

3.1 *Bayes' theorem*

The probability distribution $\pi(\boldsymbol{\theta})$ of each index's parameter vector $\boldsymbol{\theta}$ is prior probability distribution on Bayes' theorem. $\pi(\boldsymbol{\theta})$ can be updated to posterior probability distribution $\pi(\boldsymbol{\theta}|\boldsymbol{D})$ based on the likelihood function corresponding to the data sample $\boldsymbol{D}$ measured at the sites.

$$\pi(\boldsymbol{\theta}|D) = \frac{f(D|\boldsymbol{\theta})}{\int f(D|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\, d\theta}\, \pi(\boldsymbol{\theta}) \tag{10}$$

The indices that can be updated are $\{N, T', S\}$, which are measured at the sites. In order to assume the indices are normally distributed, the likelihood is calculated based on three-dimensional normal distribution. The parameters that can be updated are mean value, standard deviation, and the correlation coefficients of $\{N, T', S\}$.

3.2 *Prior distribution and updating method*

In Bayes' theorem, prior distributions of each index are set subjectively. In this research, the prior distributions are modeled from collected data, while giving as much engineering meaning to the prior distributions as possible.
The brief analysis process in modeling of the prior distributions is as follows:
Step1: Bootstrap method sampling for each site is carried out.
Step2: The parameters (i.e. mean value, standard deviation and correlation coefficient) are computed for each iteration.
Step3: The prior distributions of each parameter are modeled as a mixture distribution based on the bootstrap method sampling at all sites as shown in table 2.

Table 2. The profiles of the parameter for multivariate normal distribution

Parameter of multivariate normal distribution	Range of parameter	Distribution	Parameter of prior dist.
μ_X	-	Normal	μ, σ
σ_X	σ_X >0	Lognormal	μ, σ
ρ_{XY}	-1< ρ_{XY} <1	range extended beta	p, q

Furthermore, the domain extended beta distribution (DEB) used to fit the correlation coefficient in Table 2 will be denoted, as the probability density function (PDF) of the normal beta distribution is represented by the following equation:

$$f(x) = \frac{x^{p-1}(1-x)^{q-1}}{B(p,q)} \tag{11}$$

Where, $0 < x < 1$, $p > 0$, $q > 0$, while $B(p,q)$ is the normalization constant for setting the integral value to 1, and is defined as follows:

$$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}dx \tag{12}$$

DEB represents a modified beta distribution in which the domain condition $0 < x < 1$ is modified to $\beta < x < \alpha$, and the PDF of DEB is expressed as follows:

$$f(x) = \frac{g(x)^{p-1}\left(1-g(x)\right)^{q-1}}{B'(p,q)} \tag{13}$$

Where, $g(x) = (x - \beta)/(\beta - \alpha)$, $\beta < x < \alpha$, $p > 0$, $q > 0$, in Equation. (13),

$$B'(p,q) = \int_\alpha^\beta g(x)^{p-1}\left(1-g(x)\right)^{q-1}dx \tag{14}$$

Generally, the domain for the correlation of coefficient is $-1 \leq \rho_{XY} \leq 1$, but is in such a position that it is impossible to have a perfect correlation between each index in reality. Then, $\alpha = 1$, $\beta = -1$ in Equation. (14), the domain of DEB is $-1 < x < 1$ for the prior distribution of correlation coefficient.
Furthermore, the updating method for the probability distribution in Bayes' theorem adopts a particle filter. Updating is carried out in five-meter pitches by using the measured $\{N, T', S\}$ with considerations for application to real-time updating.

4 Results

Figures 2(a) shows histograms versus scatter plots which represented joint distribution of the indices. According to Figure 2(a), the distribution of the indices is characterized by a long tail and is similar to log-normal distribution. Figures 2(a) shows histograms versus scatter plots which are transformed to log-normal distribution. It can be seen that the distribution of indices can be transferred to normal distribution based on a simple log transformation. We considered applying the Box-Cox transformation method for high precision transformation to normal distribution, but log transformation was adopted for ease of understanding the formula meanings. For that reason, $\boldsymbol{x} = (\ln f_i, \ln N, \ln T', \ln S)^T$ is adopted as the stochastic variable vector.

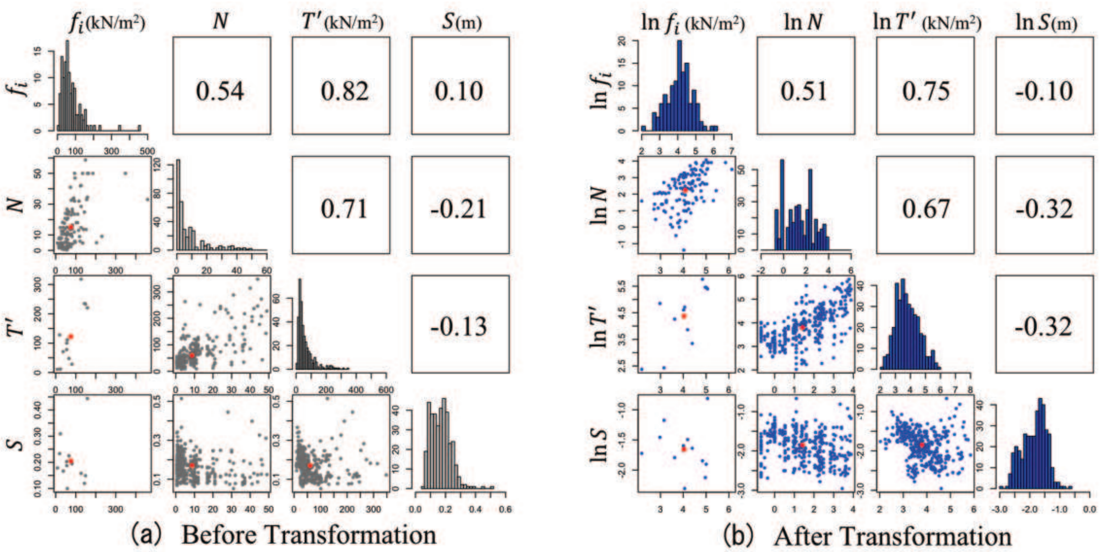

Figure 2. The pairs scatter plot on before and after transformation

$$\boldsymbol{\mu} = [4.09, 1.42, 3.80, -1.85]^T \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.48 & 0.45 & 0.38 & -0.03 \\ 0.45 & 1.63 & 0.65 & -0.16 \\ 0.38 & 0.65 & 0.58 & -0.10 \\ -0.03 & -0.16 & -0.10 & 0.16 \end{bmatrix} \tag{15}$$

To substitute the mean value vector and co-variance matrix into Equation. (7), (9), a skin-friction force f_i estimation equation is derived. The equation is called Prior Equation (not updating) in the paper.

$$\ln(f_i) = 0.02\ln(N) + 0.70\ln(T') + 0.28\ln(S) + 1.90 \tag{16}$$

$$\mathrm{var}[\ln(f_i)] = 0.20 \tag{17}$$

The green line in Figure 3 shows the estimated maximum skin-friction force fi based on the Prior Equation scatter plot against depth, red points in Figure 3 show the actual measured f_i on the vertical loading test. In addition, the blue line in Figure 3 shows Post Equation (Updated), which has been updated at each five meters in depth, while the range of estimation error is also indicated.

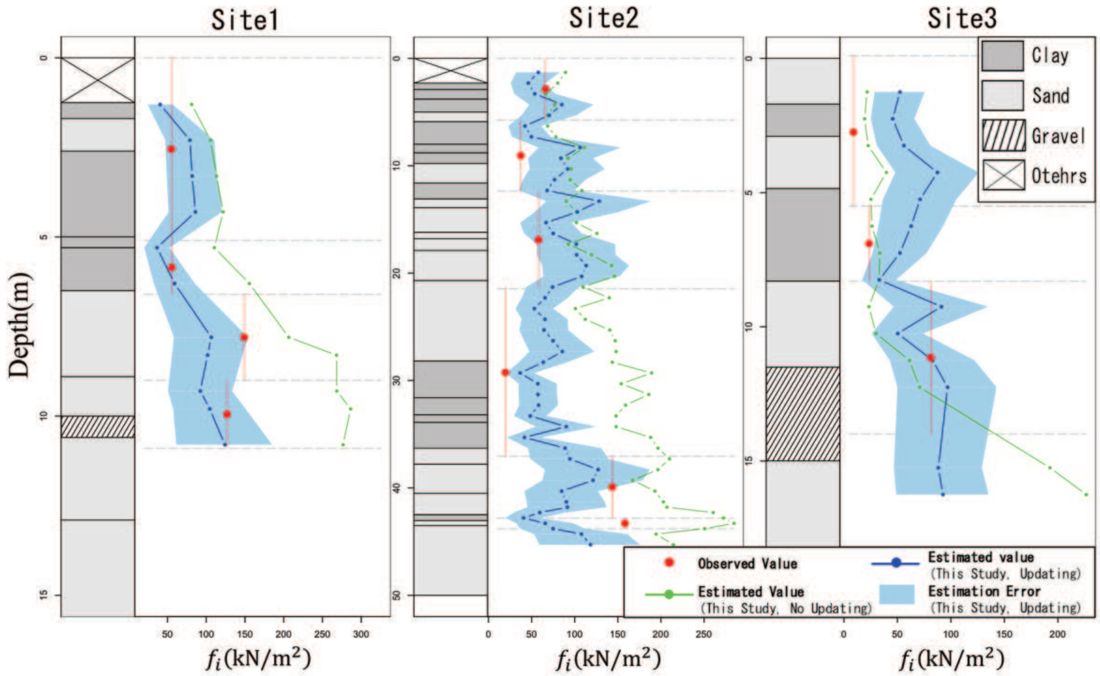

Figure 3. The estimated maximum skin-friction force considering construction information

First of all, the Prior Estimation (green line) according to Equation (16) shows a bias in the estimation accuracy for Site 1, but it shows that it represents the average characteristic of the measured values for Sites 2 and 3.

As can be seen, the Post Estimation (blue line) upon updating every five meters provides a good estimate for all sites. Specifically, on the measured value at around 30 meters depth, Post Estimation upon updating is seen to be close to the measured *fi* value in comparison with the Prior Estimation.

5 Remarks

This paper develops a reliability assessment method for pile foundations using construction information based on incomplete data for rotary penetration steel piles (RPS-piles). The main task in the data analysis was to include omitted datasets in the database. We proposed a method to derive regression equations based on the concept of conditional distribution of multivariate normal distributions, while also showing how to update equations corresponding to measured records at the site. It was demonstrated that the proposed method provides accurate estimates in comparison with equations that are not updated. In the future, the proposed method also needs to be verified based on additional data, and should be considered from the standpoint of geotechnical engineering to implement actual structural design and construction management.

References

Jianye, C. and Kok-Kwang, P. : Modeling parameters of structured clays as a multivariate normal distribution, Can. Geotech. J. 49:522-545, 2012.

Jianye, C. and Kok-Kwang, P. : Reducing shear strength uncertainties in clays by multivariate correlations, Can. Geotech. J. 47:16-33, 2010.

Yu Otake, Y. Honjo and T. Kusano, A Procedure to Determine Resistance Factors for a Newly Developed Rotation Steel Pile, Proc. of Geotechnical Safety and Risk V, pp.338-343, 2015.