ISGSR2007 First International Symposium on Geotechnical Safety & Risk
Oct. 18~19, 2007 Shanghai
Tongji University, China

Model Selection Issue in Calibrating Reliability-based Resistance Factors Based on In-situ Test Data

J. Ching, M.-T. Yen, H.-D. Lin
National Taiwan University of Science and Technology

ABSTRACT: This paper addresses the model selection issue often encountered in the process of calibrating reliability-based resistance factors. As well known, a predictive model must be assumed for the purpose of calibrating resistance factors based on in-situ test data. A critical question is raised by this research: which predictive model should we choose? What type of probability distribution model should we pick to model the model uncertainties? Those are important questions to ask because the calibration results seriously depend on the assumed predictive and probabilistic models. A full probabilistic framework is proposed in this research to resolve the model selection issue as well as to calibrate the resistance factors. Two examples of real dataset are used to illustrate the model selection issue and to demonstrate the use of the proposed methods. The proposed methods lead to reasonable conclusions and may contribute code calibration based on in-situ test data.

1 INTRODUCTION

Traditionally, geotechnical designs are usually achieved by the safety-factor approach, which accounts for uncertainties with an empirical basis. However, safety factor alone is not a consistent measure of uncertainties: it does not rigorously tell us how reliable the designed system is. As a simple example, two different geotechnical designs with the same factor of safety are usually not equally safe. Mathematically, safety factor is not a consistent indicator of safety status: the safety factor of the same limit state may change depending on the mathematical expression of the limit state function (Phoon 1995).

More recently, reliability-based design approaches have emerged as a new design paradigm because reliability is a consistent measure of uncertainties. For reliability-based designs, one designs a geotechnical system so that its reliability is in an acceptable range. Moreover, a single factor, called the resistance factor, is often used to quantify resistance uncertainties to facilitate reliability-based designs.

A common way of calibrating resistance factors based on in-situ load test data is described as follows. Given the observed resistances of m independent in-situ load tests $\{C_i: i=1,\ldots,m\}$ and the corresponding predicted resistances $\{R_i: i=1,\ldots,m\}$, the m resistance ratios $\{C_i/R_i : i=1,\ldots,m\}$ are computed, and their mean value and coefficient of variation (c.o.v.) are estimated. By assuming lognormality or Gaussianity for the ratio, resistance factors can be calibrated with a simple probability analysis. This process of calibrating resistance factors have been implemented in Barker et al. (1991), McVay et al. (2000), Zhang et al. (2001), and Paikowsky et al. (2004).

A necessary step for the above procedure is to select a deterministic predictive resistance model, denoted by $R(Z)$, where Z contains all parameters necessary to compute the resistance, e.g.: soil properties, model parameters, etc. The predictive models reflect our belief on the behavior of the target geotechnical system. For instance, when a resistance factor for vertical bearing capacity of a pile is calibrated, a model capable of predicting the vertical capacity of the pile, e.g.: a SPT-N-based model, must be chosen a priori.

1.1 Selection of Predictive Model – A Real Database of Pile Proof Load Tests

Not surprisingly, the chosen predictive model is critical for calibrating resistance factors. However, the common procedure of calibrating resistance factors does not rigorously address the model selection issue: how do we know the chosen predictive model is plausible? A heuristic argument to resolve this issue is to compare the variability of the resistance ratios of various predictive models, and the model with the lowest variability is considered to be the best model. However, this heuristic argument causes counter-intuitive conclusions when we studied a real database of pile proof load tests, described as follows.

A real database of 33 pile proof load tests in the west-coast region in Taiwan was studied. Each load test dataset contains the following information: (a) the actual ultimate vertical resistance of the pile determined by the Davisson's method, (b) the configuration of the pile, including the diameter and length, and (c) the soil profile and measured soil properties nearby the pile location, e.g.: measured unit weight, measured SPT-N value, measured undrained shear strength, etc. Based on the pile configurations, soil profile, and measured soil properties for the 33 test piles, the ultimate vertical resistances of the piles are predicted using three different predictive models, listed in Table 1 (Lin 2006).

The axial capacity of the pile (pile resistance) can be calculated as:

$$Q_u = Q_s + Q_b = \sum f_s A_s + q_b A_b \tag{1}$$

where Q_u = ultimate pile resistance, Q_s = skin resistance of the pile shaft, Q_b = point resistance at pile tip, f_s = unit skin resistance stress, q_b = unit point resistance stress, A_s = surface area of pile shaft, A_b = cross-section area of pile tip. There are two common methods of calculating f_s and q_b: the static method and the SPT-N value method. These three models listed in Table 1 are developed primarily based on the Taiwan's Foundation Design Code for Building (TGS 2001). The first two models are the recommended practice in the current design code. The third model adopts a combination of design values recommend by the American Petroleum Institute (API 1993), Kulhawy et al. (1983), and Meyerhof (1976).

Table 1 The three deterministic predictive models adopted for the pile load test dataset

Model	Skin friction	End bearing	Resistance ratio statistics	
			average	c.o.v.
SPT-N	Clay: $f_s = \alpha S_u$ (1) Sand: $f_s = N/3 \le 15T/m^2$	Clay: $q_b = 9S_u$ Sand: $q_b = 30\overline{N} \le 1500T/m^2$	0.99	0.19
Static #1	Clay: $f_s = \alpha S_u$ (1) Sand: $f_s = k\sigma_v^{'} \tan\delta \le f_{\max}$ (2)	Clay: $q_b = 9S_u$ Sand: $q_b = \sigma_v^{'} N_q^* \le q_{\max}$ (3)	1.16	0.23
Static #2	Clay: $f_s = \alpha S_u$ (4) Sand: $f_s = k\sigma_v^{'} \tan\delta \le f_{\max}$ (5)	Clay: $q_b = 9S_u$ Sand: $q_b = \sigma_v^{'} N_q^* \le q_{\max}$ (6)	0.85	0.23

Remarks: (1) α suggested by Tomlinson (1994)
(2) f_{max} suggested by DM7-2 (1982); k and δ=0.67ϕ' suggested by Taiwan code (TGS 2001)
(3) N_q^* and q_{max} suggested by DM7-2 (1982)
(4) α suggested by API (1993)
(5) f_{max} and k suggested by API (1993); δ=0.9ϕ' suggested by Kulhawy (1983)
(6) N_q^* suggested by Meyerhof (1976); q_{max} suggested by API (1993)

Notice that the predicted resistances $\{R_i$: i=1,…,$m\}$ are computed based on the ***measured*** soil properties. Compared with the actual resistances $\{C_i$: i=1,…,$m\}$, the mean value and coefficient of variation (c.o.v.) of the resistance factors are listed in the table. It is seen that the SPT-N model results in the smallest resistance ratio variability, which is rather counter-intuitive. However, should the SPT-N model be the best model among the three models? Intuitively, this should not be the case because predicting resistances based on SPT-N values will result in large uncertainties, and a model with large uncertainties should not be the best one. Later, it will be seen that the counter-intuitive, or even contradictory, results stem from the following issue: the predicted resistances $\{R_i: i=1,\ldots,m\}$ should have been computed based on the ***actual*** soil properties rather than the ***measured*** ones.

1.2 Selection of Probabilistic Model of Resistance Ratios

Another serious issue regarding model selection happens as we choose the probabilistic model for the resistance ratios. As it will be seen in the examples, the choice of the type of the probability density function (PDF) of the resistance ratios has a major effect on the calibrated resistance factors. In most literature, the lognormal distribution is often chosen as the PDF of the resistance ratios. Without verification, the adequacy of this choice is questionable.

1.3 Focus of This Paper

The focus of this paper is to provide a consistent and rigorous framework of selecting models for the purpose of calibrating resistance factors based on in-situ test data. The proposed full probabilistic framework is able to resolve the model selection issue for both the deterministic predictive models and probabilistic models. Moreover, the proposed framework can not only select the best model but also estimate the relationship between the target reliability index and resistance factor, i.e. accomplish the calibration of resistance factor at the same time. Our goal is to provide new technical guidelines for geotechnical code calibration based on in-situ test data.

The structure of the paper is as follows: First the main idea of model selection will be discussed in details. An indicator called "evidence" will be proposed to quantify the plausibility of a chosen model. Then, algorithms of estimating evidence will be proposed. At the same time, a method of estimating the relationship between the target reliability index and resistance factor will be presented. Finally, two examples of resistance factor calibration will be used to demonstrate the model selection issues.

2 MODEL SELECTION BASED ON IN-SITU TEST DATA

Let the in-situ test data be $\{C_i: i=1,\ldots,m\}$, where $C_i \in R$ is the observed resistance of the i-th test, and let $Z_i \in R^{n_i}$ be the measured soil parameters at the i-th test site (n_i is the number of soil parameters, and let $Z_{i,j} \in R$ be the j-th measured soil parameter in the i-th site). Let $\{ M^{(n)}: n=1,\ldots, N_M\}$ be the N_M chosen models for the dataset. To ease the notation, we will constantly use the notation $C_{1:m}$ to denote the dataset $\{C_i: i=1,\ldots,m\}$, and similar notations will be taken for other variables.

2.1 Evidence as Model Plausibility

One can quantify the plausibility of each model with $P(M^{(n)}|C_{1:m})$, the probability that $M^{(n)}$ is true conditioning on the data $C_{1:m}$. The Bayes' rule reveals that

$$P\left(M^{(n)} \mid C_{1:m}\right) = \frac{f\left(C_{1:m} \mid M^{(n)}\right) P\left(M^{(n)}\right)}{\sum_{n=1}^{N_M} f\left(C_{1:m} \mid M^{(n)}\right) P\left(M^{(n)}\right)} \qquad (2)$$

where $P(M^{(n)})$ is the probability that $M^{(n)}$ is true before the data is available. Note that

$$\sum_{n=1}^{N_M} P\left(M^{(n)} \mid C_{1:m}\right) = 1 \qquad (3)$$

because of the basic property of probability. It is usually the case that $P(M^{(n)})$ is taken to be uniform and equal to $1/ N_M$, i.e.: before data is available, all models are assumed to be equally likely. Under such assumption, one can easily see that $P(M^{(n)}| C_{1:m})$ is proportional to $f(C_{1:m}| M^{(n)})$, called the evidence of $M^{(n)}$ for the data $C_{1:m}$. In other words, determining the probability that $M^{(n)}$ is true is equivalent to estimating the evidence $f(C_{1:m}| M^{(n)})$.

2.2 Detailed Structure of a Probabilistic Model M

Before the method of estimating the evidence $f(C_{1:m}|\ M)$ is presented, it is essential to explain in more details the structure of a probabilistic model M.A probabilistic model M usually consists of a deterministic predictive resistance model and the prior probability distributions for all the uncertainties.

The following predictive resistance model is proposed in this paper as the deterministic predictive model for the results of in-situ proof load tests:

$$C_i = \rho_i \cdot R\left(Z_i^A, \theta\right) \tag{4}$$

where $C_i \in R$ is the observed resistance of the i-th test; $Z_i^A \in R^{n_i}$ contains the ***actual*** soil properties of the i-th test site; $R(Z_i^A, \theta)$ stands for the predicted resistance under the M model: it typically depends on the actual soil properties Z_i^A and model parameters θ; $\rho_i \in R$ is simply the ratio between the actual resistance C_i and the predicted resistance $R(Z_i^A, \theta)$.

Probably the most worth-mentioning feature in Eqn. (4) is that $R(Z_i^A, \theta)$ depends on the actual soil properties Z_i^A rather than on the measured soil properties Z_i.This arrangement can be justified by the following argument: if the actual soil properties Z_i^A are known, the predicted resistance should be independent of the measurement errors (the difference between Z_i^A and Z_i). As a consequence, it cannot be the case that the predicted resistance depends on the measured soil properties Z_i. This way of modeling the predicted resistance is very different from the traditional way, where the predicted resistance depends on the measured soil properties Z_i.

It will be seen later that the contradictory results presented in Introduction, i.e. the SPT-N model is the best model, stems from the fact that traditional methods use the measured soil properties Z_i to compute the predictive resistance. Although the new way of predicting resistance seems more reasonable, the resulting probabilistic model is much more complex and becomes much more challenging to handle. Compared to the traditional way, the number of uncertainties grows dramatically since the actual soil properties are uncertain rather than being fixed at the measured values.

To complete the probabilistic model M, the joint prior PDFs of all uncertainties must be specified. These prior PDFs reflect our degree of belief in the possible values of the uncertain variables before the in-situ test data is available. Suppose there are m sets of in-situ test data, the uncertain variables in Eqn. (4) then include the model parameters θ, the actual soil properties $Z_{1:m}^A$, and the resistance ratio $\rho_{1:m}$.

It is quite common to assume $\rho_{1:m}$ are independent identically distributed (i.i.d.) as some chosen prior PDF, e.g.: lognormal or Gaussian, whose mean value is μ and c.o.v. is δ. Independence is a sensible assumption because given the resistance ratio of one test, it does not seem to provide information of the resistance ratios for other tests. Also, the assumption that $\rho_{1:m}$ are identically distributed is routinely taken in the literature (e.g.: Barker et al. 1991; McVay et al. 2000). Note that knowing the prior PDF of ρ_1 is equivalent to knowing the conditional PDF $f(Ci|Z_i^A, \theta, \mu, \delta, M)$.

The parameters μ and δ, called the hyper-parameters, are in principle unknown. It is necessary to specify their prior PDFs $f(\mu|M)$ and $f(\delta|M)$. The model parameters θ are also unknown. It is also necessary to specify their PDF $f(\theta|M)$. The example choices of the prior PDFs $f(\mu|M)$, $f(\delta|M)$, and $f(\theta|M)$ will be demonstrated when the two real case studies are presented.

The total number of (uncertain) actual soil parameters $Z_{1:m}^A$ is $n_1+n_2+\ldots+n_m$. Their measured values are $Z_{1:m}$, which are also certain before the site investigation is carried out. Let $Z_{i,j}$ and $Z_{i,j}^A$ be the j-th measured and actual soil parameters, respectively, of the i-th in-situ test site. It is broadly accepted that the measured value $Z_{i,j}$ be the contaminated version of $Z_{i,j}^A$. It is then sensible to assume that $Z_{i,j}^A$ is distributed as some chosen prior PDF, denoted by $f(Z_{i,j}^A|M)$ (most often lognormal or Gaussian), whose mean value is $Z_{i,j}$ and c.o.v. is chosen as recommended by literature, e.g.: Phoon (1995), for measurement errors. In this study, all of $Z_{1:m,1:n_i}^A$ are assumed to be independent since measurement errors of different tests are usually independent. Also, measured soil parameters $Z_{i,j}$ in the condition will be omitted whenever there is no confusion.

Once the afore-mentioned prior PDFs are all specified, the joint prior PDF of all uncertainties can be written as ($C_{1:m}$ are not uncertain but are treated uncertain for the time being)

$$
\begin{aligned}
& f\left(C_{1:m}, Z^A_{1:m,1:n_i}, \theta, \mu, \delta \mid M\right) \\
& = f(\theta \mid M) f(\mu \mid M) f(\delta \mid M) \prod_{i=1}^{m} \left[f\left(C_i \mid Z^A_{i,1:n_i}, \theta, \mu, \delta, M\right) \prod_{j=1}^{n_i} f\left(Z^A_{i,j} \mid M\right) \right]
\end{aligned} \tag{5}
$$

where if $\rho_{1:m}$ is assumed to be Gaussian,

$$
f\left(C_i \mid Z^A_{i,1:n_i}, \theta, \mu, \delta, M\right) = \frac{1}{\sqrt{2\pi} \cdot \mu \cdot \delta} e^{\frac{-1}{2\mu^2 \cdot \delta^2}\left(C_i - \mu \cdot R\left(Z_i^A, \theta\right)\right)^2} \tag{6}
$$

if $\rho_{1:m}$ is assumed to be lognormal,

$$
f\left(C_i \mid Z^A_{i,1:n_i}, \theta, \mu, \delta, M\right) = \frac{1}{\sqrt{2\pi \cdot \log\left(1+\delta^2\right) \cdot C_i}} e^{\frac{-1}{2\log\left(1+\delta^2\right)}\left(\log[C_i] - \log\left[\mu \cdot R\left(Z_i^A, \theta\right) \big/ \sqrt{1+\delta^2}\right]\right)^2} \tag{7}
$$

and if $\rho_{1:m}$ is assumed to be Gumbel,

$$
\begin{aligned}
& f\left(C_i \mid Z^A_{i,1:n_i}, \theta, \mu, \delta, M\right) \\
& = \frac{1.2825}{\mu \cdot \delta \cdot R\left(Z_i^A, \theta\right)} \cdot e^{\frac{-1.2825\left(C_i - (1-0.4504\delta) \cdot \mu \cdot R\left(Z_i^A, \theta\right)\right)}{\mu \cdot \delta \cdot R\left(Z_i^A, \theta\right)}} \cdot e^{-\exp\left[\frac{-1.2825\left(C_i - (1-0.4504\delta) \cdot \mu \cdot R\left(Z_i^A, \theta\right)\right)}{\mu \cdot \delta \cdot R\left(Z_i^A, \theta\right)}\right]}
\end{aligned} \tag{8}
$$

3 ESTIMATION OF EVIDENCE

For notational simplicity, let us denote all uncertainties $Z_{1:m,1:n_i}{}^A$ $,\theta, \mu, \delta$ by X.The estimation of the evidence $f(C_{1:m}| M)$ seems simple because

$$
f\left(C_{1:m} \mid M\right) = \int f\left(C_{1:m}, X \mid M\right) \cdot dX = \int \prod_{i=1}^{m} \left[f\left(C_i \mid X, M\right) \right] \cdot f\left(X \mid M\right) \cdot dX \tag{9}
$$

Where

$$
f\left(X \mid M\right) = f\left(\theta \mid M\right) f\left(\mu \mid M\right) f\left(\delta \mid M\right) \prod_{i=1}^{m} \prod_{j=1}^{n_i} f\left(Z^A_{i,j} \mid M\right) \tag{10}
$$

is the joint prior PDF of all uncertainties, and

$$
f\left(C_i \mid X, M\right) = f\left(C_i \mid Z^A_{i,j}, \theta, \mu, \delta, M\right) \tag{11}
$$

is called the likelihood function of the i-th in-situ test result. Note that once the prior PDFs of all uncertainties are specified and the test data $C_{1:m}$ is available, all terms in the integrant in Eqn. (9) can be evaluated.

This result shows that the determination of the evidence can be achieved by computing the high dimensional integral in Eqn. (9). However, the direct calculation of the integral is infeasible due to the high dimensionality. One way of resolving this issue is to estimate the integral by the Law of Large Number with Monte Carlo simulations:

$$f\left(C_{1:m} \mid M\right) \approx \frac{1}{N}\sum_{n=1}^{N}\left[\prod_{i=1}^{m} f\left(C_i \mid X^{*(n)}, M\right)\right] = \frac{1}{N}\sum_{n=1}^{N}\left[\prod_{i=1}^{m} f\left(C_i \mid Z_{i,1:n_i}^{*A(n)}, \theta^{*(n)}, \mu^{*(n)}, \delta^{*(n)}, M\right)\right] \quad (12)$$

where $Z^{*}_{i,1:n_i}{}^{A(n)}$ { $Z^{*}_{i,1:n_i}{}^{A(n)}, \theta^{*(n)}, \mu^{*(n)}, \delta^{*(n)}$} are the n-th Monte Carlo samples drawn from their prior PDFs. However, the main support region of their prior PDFs is usually very different from the main support region of the likelihood function $\prod^{m}_{i=1} f(Ci|Z_{i,1:n_i}{}^{A}, \theta, \mu, \delta, M)$. The consequence is that most Monte Carlo samples fall into low likelihood region, so the variability of the estimator in (12) can be extremely large.

The above difficulty can be avoided by considering the following equality:

$$\begin{aligned} &\ln\left[f\left(C_{1:m} \mid M\right)\right] \\ &= \int\left(\sum_{i=1}^{m} \ln\left[f\left(C_i \mid X, M\right)\right] + \ln\left[f\left(X \mid M\right)\right]\right) f\left(X \mid C_{1:m}, M\right) dX + H\left[X \mid C_{1:m}, M\right] \end{aligned} \quad (13)$$

Where

$$H\left[X \mid C_{1:m}, M\right] = -\int \ln\left[f\left(X \mid C_{1:m}, M\right)\right] f\left(X \mid C_{1:m}, M\right) dX \quad (14)$$

is the differential entropy of $f(X|C_{1:m},M)$. Therefore,

$$\ln\left[f\left(C_{1:m} \mid M\right)\right] \approx \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{i=1}^{m}\ln\left[f\left(C_i \mid \hat{X}^{(n)}, M\right)\right] + \ln\left[f\left(\hat{X}^{(n)} \mid M\right)\right]\right) + \hat{H}\left[X \mid C_{1:m}, M\right] \quad (15)$$

where $X^{*(n)}$ is the n-th sample from $f(X|C_{1:m},M)$; $H^{*}[X|C_{1:m},M]$ is the estimated entropy of $f(X|C_{1:m},M)$. Note that the estimation of the first term at the right-hand-size of the equation is robust since the main support region of $f(X|C_{1:m},M)$ is the same as the important region of $\sum^{m}_{i=1} ln[f(C_i|X,M)]+ln[f(X|M)]$. Therefore, this approach resolves the issue encountered in Eqn. (12). The only two remaining questions are (a) how do we effectively draw samples from $f(X|C_{1:m},M)$? and (b) how do we estimate the differential entropy? Those issues are addressed in Ching et al. (2007), where detailed algorithms of resolving those two issues are presented.

4 CALIBRATION OF RESISTANCE FACTORS

Given the in-situ test data $C_{1:m}$ and the chosen probabilistic model M, it is possible to calibrate the reliability-based resistance factor for the same type of geotechnical designs. Let us assume that N samples of $f(X|C_{1:m},M)$, denoted by {$Z^{*}_{1:m,1:n_i}{}^{A(1:N)}, \theta^{*(1:N)}, \mu^{*(1:N)}, \delta^{*(1:N)}$}, have been obtained by the procedure presented in Ching et al. (2007). The reliability-based design can be achieved by restricting the failure probability of a future design conditioning on the past data, i.e. consider the following equation:

$$P\left(C_{new} \le C^{*} \mid C_{1:m}, M\right) = P_F^{*} \quad (16)$$

where C_{new} is the uncertain capacity of a new design, and C^{*} is the target design load; $C_{new} \le C^{*}$ is the failure event; $P_F{}^{*}$ is the target failure probability; $P(C_{new} \le C^{*}|C_{1:m},M)$ is the failure probability conditioning on the test data and the chosen model. This probability can be, in principle, computed with Bayesian analysis. Equation (16) can be rewritten as

$$P\left(C_{new} \le C^{*} \mid C_{1:m}, M\right) = P\left(\rho_{new} R\left(Z^{A}_{new}, \theta\right) \le C^{*} \mid C_{1:m}, M\right) = P_F^{*} \quad (17)$$

where ρ_{new} is the model factor of the new design, whose distribution is assumed to be identical to that of $\rho_{1:m}$, but ρ_{new} is independent of $\rho_{1:m}$; Z_{new}^A contains all uncertain soil parameters of the new site. Note that $R(Z_{new}^A, \theta)$, the predicted resistance of the new design, is uncertain because Z_{new}^A and θ are uncertain.

Let us assume that site investigation will be taken to observe Z_{new}^A, and the measured value is Z_{new}. Let us also assume that the prior PDF of Z_{new}^A is chosen as some PDF whose mean value is Z_{new} and c.o.v. chosen according to recommendations of literature. Therefore, the relationship between the resistance factor and target failure probability can be derived as follows:

$$\begin{aligned} P_F^* &= P\left(\rho_{new} R\left(Z_{new}^A, \theta\right) \leq C^* \mid C_{1:m}, M\right) \\ &= P\left(\rho_{new} R\left(Z_{new}^A, \theta\right) / R\left(Z_{new}, \theta^*\right) \leq C^* / R\left(Z_{new}, \theta^*\right) \middle| C_{1:m}, M\right) \\ &= P\left(\rho_{new} R\left(Z_{new}^A, \theta\right) / R\left(Z_{new}, \theta^*\right) \leq \eta \middle| C_{1:m}, M\right) \\ &= P\left(G\left(\rho_{new}, \theta, Z_{new}^A\right) \leq \eta \middle| C_{1:m}, M\right) \end{aligned} \tag{18}$$

Where θ^* is some designated (nominal) value of the model parameter; $G(\rho_{new},\theta,Z_{new}^A) \equiv \rho_{new} R(Z_{new}^A, \theta)/R(Z_{new},\theta^*)$, and $\eta \equiv C^*/R(Z_{new},\theta^*)$ is exactly the resistance factor. One can see that the resistance factor can be simply estimated as the $100 \cdot P_F^*$ percentile of $G(\rho_{new},\theta,Z_{new}^A)$ conditioning on the data $C_{1:m}$ and model M. If one can obtain stochastic simulation samples of $G(\rho_{new},\theta,Z_{new}^A)$ conditioning on the data $C_{1:m}$ and model M, denoted by $\{G^1,\ldots,G^N\}$, the relationship between resistance factor and target failure probability can be estimated as the following according to the Law of Large Number:

$$P_F^* \approx \frac{1}{N} \sum_{k=1}^{N} 1\left(G^k \leq \eta\right) \tag{19}$$

where $1(\cdot)$ is the indicator function: it is unity if the inside statement is true and is zero otherwise. Equivalently, η can be estimated as the $100 \cdot P_F^*$ largest number in $\{G^1,\ldots,G^N\}$. Once this relationship is obtained, the calibration of the resistance factor is achieved since one can determine the required resistance factor from the target failure probability.

Let us now turn to the question of how to draw samples of $\{G^1,\ldots,G^N\}$ conditioning the data $C_{1:m}$ and model M. Since $G(\rho_{new},\theta,Z_{new}^A)$ is a function of ρ_{new}, θ, and Z_{new}^A, it suffices if we know how to draw samples of ρ_{new}, θ, and Z_{new}^A conditioning on the data $C_{1:m}$ and model M, i.e. draw samples from $f(\rho_{new},\theta,Z_{new}^A \mid C_{1:m},M)$. First of all, Z_{new}^A is independent of $C_{1:m}$, so drawing Z_{new}^A samples conditioning on $C_{1:m}$ is the same as drawing Z_{new}^A samples from its prior PDF. Recall that the samples $\{Z^*_{1:m,1:n_i}{}^{A,1:N},\theta^{*1:N},\mu^{*1:N},\delta^{*1:N}\}$ obtained by the procedure described in Appendix A are distributed as $f(Z_{1:m,1:n_i}^A, \theta,\mu,\delta|C_{1:m},M)$, so $\theta^{*1:N}$ are distributed as $f(\theta|C_{1:m},M)$ and $\{\mu^{*1:N},\delta^{*1:N}\}$ are distributed as $f(\mu,\delta|C_{1:m},M)$. For each sample pair of $\{\mu^{*k},\delta^{*k}\}$, a sample of ρ_{new}, denoted by ρ_{new}^k, can be drawn from the chosen PDF whose mean value is μ^{*k} and c.o.v. is δ^{*k}. It is then clear that $\{\rho_{new}^{1:N},\theta^{1:N}\}$ are distributed as $f(\rho_{new},\theta \mid C_{1:m},M)$.

Previously we have seen the procedure of estimating evidence $f(C_{1:m}|M)$ and calibrating resistance factor when there is only one chosen probabilistic model M. When there are multiple models $\{M^{(n)}:n=1,\ldots,N_M\}$, it is then of interest to know how the resistance factor v.s. failure probability relationship can be determined. Let us denote the resistance model for the n-th model by $R^{(n)}(Z_{new}^A, \theta)$. The following discussion considers two scenarios: (a) the deterministic resistance models of $\{M^{(n)}:n=1,\ldots,N_M\}$ are the same, i.e.: $R^{(1)}(Z_{new}^A, \theta)=\ldots= R^{(N_M)}(Z_{new}^A,\theta)= R\,(Z_{new}^A,\theta)$, only the probability distribution models of the resistance ratio ρ are different; (b) the deterministic resistance models of $\{M^{(n)}:n=1,\ldots,N_M\}$ are different.

In the former case, let the evidence of the n-th model be estimated as f*(C1:m|M(n)). According to Bayes' rule, the following equation describes the resistance factor v.s. failure probability relationship:

$$P_F^* = \left[\sum_{n=1}^{N_M} P\left(\rho_{new}^{(n)} R\left(Z_{new}^{A}, \theta \right) \Big/ R\left(Z_{new}, \theta^* \right) \le \eta \middle| C_{1:m}, M^{(n)} \right) \cdot f\left(C_{1:m} \mid M^{(n)} \right) \right] \Big/ \sum_{n=1}^{N_M} f\left(C_{1:m} \mid M^{(n)} \right) \quad (20)$$

where the superscript in the resistance ratios is to highlight the fact that the assumed distribution of various models for the resistance ratios are different. Note that the resistance factor is not superscripted because the definition of $\eta^{(n)} \equiv C^*/R^{(n)}(Z_{new},\theta^*)$ stays the same among all the models. The following estimator can be used to obtain the relationship with Monte Carlo samples:

$$P_F^* \approx \left[\sum_{n=1}^{N_M} \left(\frac{1}{N_n} \sum_{k=1}^{N_n} 1\left(\rho_{new}^{(n),k} R\left(Z_{new}^{A(n),k}, \theta^{(n),k} \right) \Big/ R\left(Z_{new}, \theta^* \right) \le \eta \right) \right) \cdot \hat{f}\left(C_{1:m} \mid M^{(n)} \right) \right] \Big/ \sum_{n=1}^{N_M} \hat{f}\left(C_{1:m} \mid M^{(n)} \right) \quad (21)$$

where $\{Z_{new}^{A(n),1:Nn}, \rho_{new}^{(n),1:Nn}, \theta^{(n)1:Nn}\}$ are the Monte Carlo samples for the n-th model; N_n is the number of Monte Carlo samples for the n-th model.

In the latter case, a unique relationship between resistance factor and failure probability similar to Eqn. (21) is not available because the definition of η(n) ≡C*/R(n)(Znew,θ*) vary with different models. A possible solution is to just take the relationship obtained from the best model, the one with the largest evidence, as the final resistance factor v.s. failure probability relationship.

5 REAL EXAMPLE

A real example of in-situ test dataset is studied in this section to demonstrate the issue of model selection when calibrating resistance factors as well as the analysis results. The same example presented in Introduction is herein analyzed in details. In Introduction, the conclusion of the traditional approach is that the SPT-N model results in the least variability of the resistance ratio, so it seems that the SPT-N model prevails. Here, nine different probabilistic models, as listed in Table 2, are studied with the new approach. Basically, each deterministic predictive model in Table 1 is used to create three different probabilistic models by assuming the resistance ratio ρ to be Gaussian, lognormal and Gumbel. Evidences of the nine models will be estimated, and the relationship between resistance factor and failure probability will be found.

Table 2 The assumed nine models for the pile load test database and the logarithms of their evidences.

	Gaussian	Lognormal	Gumbel
SPT-N	Model 1	model 2	model 3
	-236.2	-252.52	-235.95
Static #1	Model 4	model 5	model 6
	-214.52	-213.93	-213.85
Static #2	Model 7	model 8	model 9
	-212.03	-212.51	-213.06

Fig.1 shows the soil profiles of the 33 pile locations ($m = 33$), where soil layers are numbered from I to V. The numbers right above the soil profiles are the observed ultimate vertical resistances (in tons) of the test piles interpreted by the Davisson’s method. Site investigations were taken to determine the soil classification, mean value and c.o.v. of unit weights and some shear strength parameters of each soil layer, as listed in Table 3.

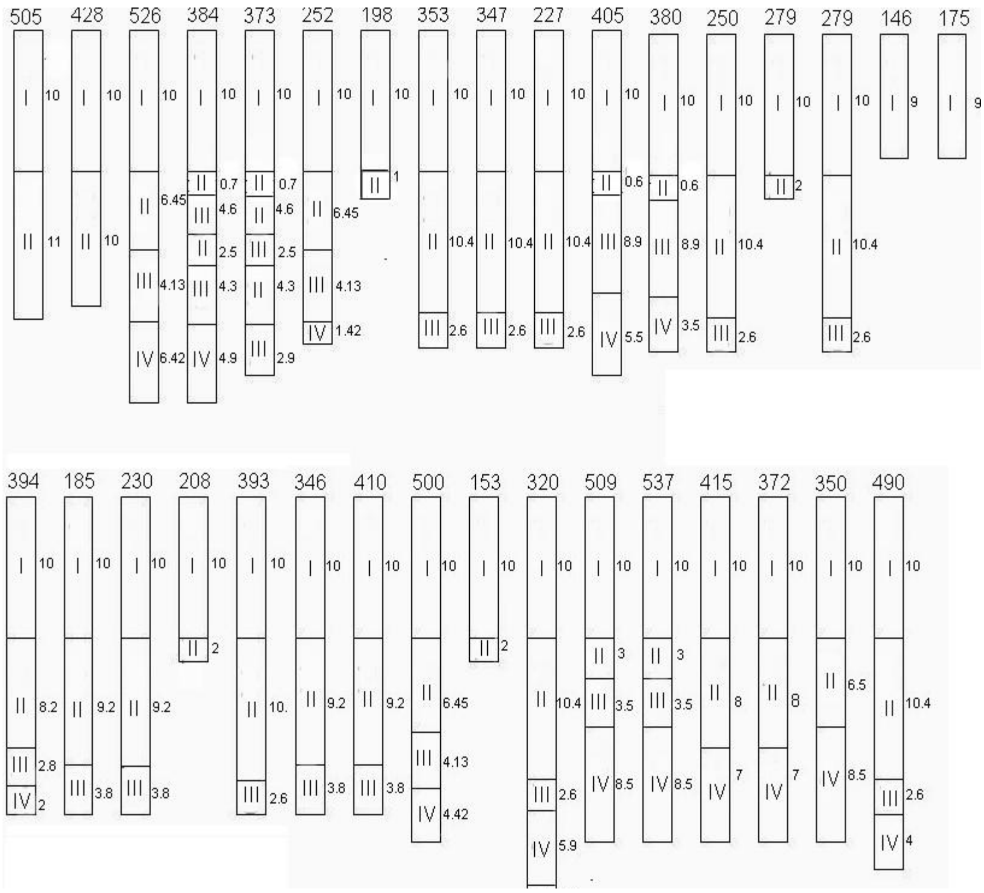

Fig.1 The soil profiles for the 33 piles in the pile load test database. The numbers right above each profile is the observed bearing capacity of the pile (in tons); the letters I to V are the indices of the soil layers (see Table 3 for details); the numbers at the right-hand side of each soil columns are the thicknesses of the layers (in meters).

Table 3 The classifications and properties of the soils in the pile load test database

Layer Index	Soil Type	γ (T/m^3) [c.o.v.]	SPT-N	S_u (T/m^2) [c.o.v.]	ϕ (°) [c.o.v.]
I	SM	1.96 [0.015]	20 [0.20]	–	34.8 [0.04]
II	SM	1.96 [0.015]	14 [0.21]	–	32.8 [0.07]
III	ML/CL	1.96 [0.02]	18 [0.28]	7.42 [0.47]	–
IV	SM	1.97 [0.015]	22 [0.18]	–	34.2 [0.07]
V	ML/CL	1.95 [0.021]	21 [0.19]	7.18 [0.42]	–

In the analysis, the actual values of these soil parameters, i.e. $Z_{1:m,1:n_i}{}^{A}$, are taken as uncertain. Their prior PDFs are taken to be independent lognormal with mean values and c.o.v.s identical to those in Table 3. It is the goal of this example to see if this new treatment of uncertainties will change the conclusion of model selection or not.

The only model parameter θ in all nine models is the depth of the water table. Based on the site investigation result, the prior PDF of the depth is taken to be uniform over [0.5m, 3.8m]. The resistance ratio ρ is assumed to be either Gaussian, lognormal, or Gumbel with unknown mean value equal to μ and unknown c.o.v. equal to δ. The prior PDFs of μ and δ are taken to be lognormal with mean values respectively equal to 1 and 0.2 and c.o.v.s respectively equal to 0.5 and 1.

For each probabilistic model, ten thousand samples of $\{Z^{*}_{1:m1:n_i}{}^{A,1:N},\theta^{*1:N},\mu^{*1:N},\delta^{*1:N}\}$ ($N = 10000$) are drawn. With the samples, the logarithms of the evidences of the nine models are estimated and listed in

Table 2. It is clear that the SPT-N model is the worst model regardless the assumed PDF type of the resistance ratio. This conclusion is completely different from that of the traditional method. To verify the underlying cause of this dramatic change of conclusion, a new set of similar analyses are taken to estimate the evidences of the nine models where the soil parameters $Z_{1:m,1:n_i}{}^A$ are held at their measured values; the results show that the SPT-N models (models 1-3) prevails again. These observations suggest that the way of treating soil parameters $Z_{1:m,1:n_i}{}^A$ is critical when model selection issues are of concern. We argue that it is plausible to treat $Z_{1:m,1:n_i}{}^A$ as uncertain rather than to hold them at the measured values.

According to the estimated evidence in Table 2, the models originated from the second static model, models 7-9, seem to be the best models among the nine models. Furthermore, the Gaussian assumption on the resistance ratio (model 7) seems slightly better than the lognormal and Gumbel assumptions (models 8 and 9).

Fig..2 shows the relationship between resistance factor and reliability index for models 7-9 for several imaginary new designs of piles. It is clear that the assumed PDF type of the resistance ratio will greatly affect the resulting resistance factor. For instance, corresponding to target reliability index of 3, the required resistance factor is among [0.35,0.41] for the lognormal assumption, among [0.25,0.32] for the Gaussian assumption, and among [0.36,0.42] for the Gumbel assumption. Without estimating the evidence, there is hardly any information for us to determine which assumed PDF type is more reasonable. This highlights the issue of selecting the PDF type of the resistance ratio.

The lower-right plot in Fig..2 contains the average result obtained with Eqn. (21) from the results of models 7-9. According to this average result, the required resistance factor corresponding to target reliability index of 3 is roughly in the range of [0.3, 0.35]. This range of resistance factor may be used to design the vertically-loaded piles in the west-coast region of Taiwan.

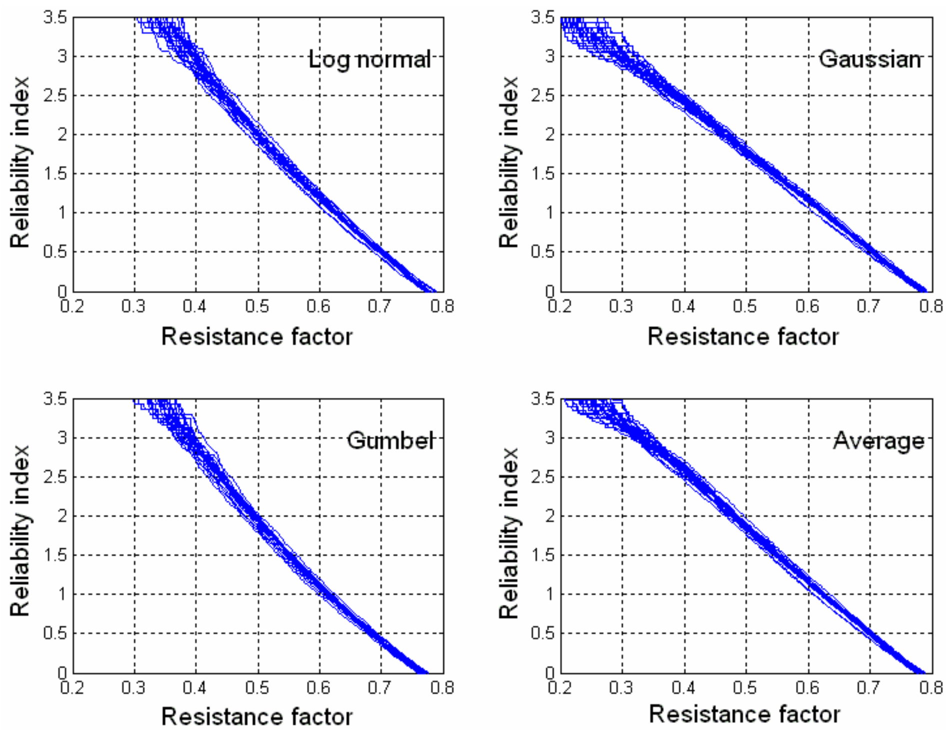

Fig.2 The relationship between resistance factor and reliability index for models 7-9 for imaginary new designs of piles.

6 DISCUSSION AND CONCLUSION

A real-data example is used to demonstrate the issue of model selection encountered in the process of calibrating resistance factors based on in-situ test data. It is proposed that the "evidence" of a model can reasonably quantify the plausibility of the chosen model. Compared to the common intuition that the best model is the one resulting in the least variability in the resistance ratio, the new approach seems to provide more rational choice of the model. This conclusion is demonstrated in the real-data example, where it is seen that the SPT-N pile-capacity model (a model with large uncertainties) results in the least resistance ratio variability, and yet its evidence is the smallest among the candidate models. It is argued that the key to appropriately conduct model selection is to correctly treat the soil properties as uncertain rather than to falsely treat them as fixed numbers.

It is shown that the assumed probability density function for the resistance ratio has a major effect on the calibrated resistance factors. This highlights the issue of selecting the PDF type of the resistance ratio: the resulting resistance factor may be misleading if the PDF type assumption is taken without further verification. The proposed framework is valuable in providing a way of rationally determining the plausibility of each candidate model. Through the proposed methods, the plausibility of each model can be quantitatively evaluated, so the model issue can be resolved rigorously.

REFERENCES

API, American Petroleum Institute (1993). Section G: foundation design. Recommended Practice for Planning, Design, and Constructing Offshore Platforms – Load and Resistance Factor Design, First Ed. API, Washington, DC.

Barker, R. M., Duncan, J. M., Rojiani, K. B., Ooi, P. S. K., Tan, C. K. and Kim. S. G. (1991). Manuals for the Design of Bridge Foundations. NCHRP Report 343, TRB, National Research Council, Washington, DC.

Ching, J., Yen, M.-T., and Lin, H.-D. (2007). Model selection issue in calibrating reliability-based resistance factors based on geotechnical in-situ test data, submitted to Structural Safety.

DM 7-2 (1982). Foundations and Earth Structures – Design Manual 7.2, Department of the Navy Naval Facilities Engineering Command, Alexandria, VA.

Kulhawy, F.H., Trautmann, C.H., Beech, J.F., O'Rourke, T.D., McGuire, W., Wood, W.A., and Capano, C. (1983). Transmission Line Structure Foundations for Uplift-Compression Loading, Report EL-2870, Electric Power Research Inst., Palo Alto.

Lin, H. D. (2006). A preliminary study on load and factors for foundation piles in Taiwan. Invited Lecture, Proceeding, International Symposium on New Generation Design Codes for Geotechnical Engineering Practice-Taipei 2006 (with CD-ROM), World Scientific Publishing Company.

McVay, M. C., Birgisson, B., Zhang, L. M., Perez, A., and Putcha, S. (2000). Load and resistance factor design (LRFD) for driven piles using dynamic methods—A Florida perspective. Geotech. Test. J., 23(1), 55–66.

Meyerhof, G.G. (1976). Bearing capacity and settlement of pile foundation. ASCE Journal of Geotechnical Engineering, 102(3), 195-228.

Paikowsky, S. G., Birgisson, B., McVay, M., Nguyen, T., Kuo, C., Baecher, G., Ayyab, B., Stenersen, K., O'Malley, K., Chernauskas, L., and O'Neill, M. (2004). Load and Resistance Factor Design (LRFD) for Deep Foundations. NCHRP Final Report 507, Transportation Research Board, Washington, DC.

Phoon, K.K. (1995). Reliability-based design of foundations for transmission line structures, Ph.D. Dissertation, Cornell University.

Tomlinson, M.J. (1994). Pile Design and Construction Practice. Fourth Ed., E & FN Spon, London.

TGS, Taiwan Geotechnical Society (2001). Foundation design code for building. Taiwan Geotechnical Society (TGS), Taipei.

Zhang, L., Tang, W. H., and Ng, C. W. W. (2001). Reliability of axially loaded driven pile groups. ASCE Journal of Geotechnical and Environmental Engineering, 127(12), 1051-1060.