Hierarchical Bayesian Approach to Estimating Variability of Liquefaction Resistance of Sandy Soils Considering Individual Differences in Laboratory Tests

Kyohei Ueda[1]

Abstract: Cyclic undrained triaxial tests are commonly used in research and practical design to evaluate the liquefaction resistance of sandy soils. This paper aims to propose a methodology to evaluate liquefaction resistance by considering the variability or uncertainty associated with experimentation, using Bayesian statistics with a Markov chain Monte Carlo technique. In addition to conventional nonhierarchical Bayesian modeling, hierarchical Bayesian modeling is adopted to properly incorporate the factor of variability caused by individual differences (e.g., difference between experimenters) into the liquefaction resistance evaluation. Findings show that the regression curves of the cyclic resistance ratio estimated by a nonhierarchical model for all experimenters' results in a cooperative triaxial test program are too generic and poorly applicable. In contrast, the curves estimated by a nonhierarchical model for each experimenter's results sometimes deviate from the overall trend, dragged by the individual characteristics. The hierarchical Bayesian modeling demonstrates that both the overall trend and each experimenter's individuality can be rationally considered in the regression results (e.g., posterior distributions of model input parameters) by referring to the other experimenters' results, even though the number of test cases is limited for the focal experimenter. Another advantage of the modeling is that, when a different experimenter newly performs similar laboratory tests, the posterior distribution based on the existing dataset can be used as a prior distribution to estimate model input parameters specific to the experimenter. The proposed methodology may also be used to estimate the variability of liquefaction resistance considering individual differences in laboratory tests that are difficult to quantify, e.g., differences in testing apparatus and specimen size. **DOI: [10.1061/(ASCE)GT.1943-5606.0002749](https://doi.org/10.1061/(ASCE)GT.1943-5606.0002749).** *This work is made available under the terms of the Creative Commons Attribution 4.0 International license, https://creativecommons.org/licenses/by/4.0/.*

Author keywords: Liquefaction resistance; Undrained cyclic triaxial test; Individual difference; Bayesian estimation; Markov chain Monte Carlo method; Hierarchical model.

Introduction

In the organization of observation and experiment results, both the median trend and the variability due to uncertainty should be evaluated, for example, to derive an empirical equation. According to Ching and Phoon (2019), the uncertainty encountered in geotechnical engineering can be classified into four categories: spatial variability, transformation uncertainty, statistical uncertainty, and measurement uncertainty. Because boring and sounding are generally performed at certain intervals, the spatial distribution of soil properties is unknown at unexplored locations; this kind of uncertainty is called spatial variability (Phoon and Kulhawy 1999a; Vanmarcke 1977). The spatial variability should be considered when laboratory tests are performed using in-situ sampled specimens but can be ignored when tests are done using industrial soil materials (e.g., standard sand, such as Toyoura sand in Japan). Transformation uncertainty arises in the correlation equation (or transformation model) for designing soil parameters (e.g., friction angle of sand) based on test indexes [e.g., standard penetration test (SPT) blow count or cone penetration test (CPT) results] (Phoon and Kulhawy 1999b). Statistical uncertainty is due to the lack of

quantity of observations and experiments. Measurement uncertainty comes from the precision and accuracy of measuring instruments, along with any other factor that might affect an experimenter's ability to measure.

The preceding variabilities and uncertainties are also inevitable when investigating the liquefaction strength and potential of the ground during earthquakes. In studies based on field observations, Liao et al. (1988) and Hwang et al. (2004) evaluated the reliability-based liquefaction potential considering the variability of SPT results. In addition, research has been conducted on liquefaction assessment considering the variability of CPT results (e.g., Juang et al. 1999; Moss et al. 2006).

While there have been many studies on the variability of liquefaction characteristics based on field observations, few studies have attempted to quantify statistically the liquefaction strength and potential in the laboratory. For example, Haldar and Miller (1984) proposed a statistical model for determining the cyclic shear strength of sand deposits based on large-scale laboratory shaking table tests. Similarly, Liu and Chang (1994) developed a statistical model for estimating the liquefaction resistance of fine sand based on cyclic triaxial tests. Polito (2009, 2011) also proposed a statistical model for estimating the cyclic shear resistance against the number of cyclic loads to liquefaction by regression analysis using the fines content, median grain size, and void ratio as parameters; the regression analysis was performed for three specimen preparation methods on a large database of cyclic triaxial tests to reduce the statistical uncertainty. The proposed model could evaluate the general trend (mean and variation coefficient of the liquefaction resistance) of the experimental results for each preparation method,

[1]Assistant Professor, Disaster Prevention Research Institute, Kyoto Univ., Gokasho, Uji, Kyoto 611-0011, Japan. ORCID: https://orcid.org/0000-0001-6202-6431. Email: ueda.kyohei.2v@kyoto-u.ac.jp

© ASCE   04021188-1   J. Geotech. Geoenviron. Eng.

J. Geotech. Geoenviron. Eng., 2022, 148(2): 04021188

including the various specimen sizes used by multiple laboratories and experimenters. However, the method cannot consider the effect of the individual differences between specimen sizes or experimenters on the liquefaction resistance. Therefore, if individual differences between experimenters cannot be ignored, the use of the general trend for individual experimental results may not be ideal. A regression model can be developed for individual experimental results, but this can lead to a statistical uncertainty problem due to the insufficient number of laboratory tests. Another drawback of the proposed model is that, when a new but similar laboratory test is performed, it is difficult to use the model coupled with the test results directly. This means that the regression analysis has to be repeated to include the new test results. Furthermore, the model input parameters estimated in the regression analysis may be uncertain due to the presence of statistical and measurement uncertainties, even if the transformation (or model) uncertainty could be minimized.

The variability involved in the transformation process of building a statistical model from experimental results can be modeled as random variables and estimated by applying the concept of Bayesian statistics in order to consider the aforementioned uncertainties (e.g., Gelman et al. 2013). The advantage of this process is that the distributions of both the model input parameters and model uncertainty variables are provided with the help of Markov chain Monte Carlo (MCMC) methods. This concept is Bayesian estimation using MCMC, which has been gaining recognition in the realm of geotechnical engineering (Ching et al. 2011; Park et al. 2012; Santoso et al. 2011; Wang et al. 2010; Zhang et al. 2014, 2012, 2009, 2010). While most of these studies employ nonhierarchical modeling, Zhang et al. (2014) propose a hierarchical probabilistic model to calibrate the cross-site variability (i.e., variability from site to site) for the reliability-based design of pile foundations.

However, to the best of the author's knowledge, hierarchical Bayesian modeling (e.g., Gelman et al. 2013) is not considered to be widely used in geotechnical engineering. In particular, no researcher has applied hierarchical or even nonhierarchical Bayesian modeling to assess liquefaction resistance via laboratory tests. Thus, the objective of this paper is to propose a methodology to evaluate liquefaction resistance (or estimate the distribution of model input parameters for defining the resistance) by considering the aforementioned uncertainties using the concept of Bayesian modeling. This work takes as an example the results of cyclic undrained triaxial tests conducted by multiple experimenters using Toyoura standard sand (Toki et al. 1986). In addition to conventional nonhierarchical Bayesian modeling, this study implements Bayesian estimation using a hierarchical model to properly incorporate the

factor of variability caused by individual differences (e.g., due to participation of different experimenters) into the liquefaction resistance evaluation.

Cooperative Test Program of Cyclic Undrained Triaxial Tests

As explained by Toki et al. (1986), five organizations (or experimenters) participated in a cooperative test program of cyclic undrained triaxial tests using saturated Toyoura sand with relative densities (D_r) of 50% and 80%. As described in an accompanying paper (Tatsuoka et al. 1986), various factors affected the test results, but effort was exerted to reduce the experimental variability of this cooperative campaign using common specifications as much as possible. The specimen preparation method used to achieve the target D_r, frequency of cyclic loading, and isotropic compression stage before cyclic loading were common among the experiments, and some specifications were different, as shown in Table 1. Results showed that the effect of the height-diameter ratio (H/D) was not excessively large, and the variation observed in the experiments could be attributed to the difference in membrane penetration due to the discrepancy in specimen diameter. In addition to the H/D and specimen diameter, the individual differences between experimenters were regarded as a factor, but they have not been quantitatively analyzed. However, because it is difficult for even the most skilled technicians to eliminate individual differences completely, laboratory test results should be discussed quantitatively considering individual differences. Therefore, this study attempts to analyze individual differences by applying the hierarchical Bayesian modeling described in the next section. Note that not all experimenters used exactly the same triaxial testing device, which may affect the experimental protocol and accuracy. However, this paper does not explicitly deal with the device difference, assuming that individual differences mainly arise from the experimenter's skill difference.

Hierarchical Bayesian Model Using MCMC Methods

Nonhierarchical Bayesian Model

In general, there exists variability or uncertainty in the results of observations and experiments. According to Ching and Phoon (2019), the four sources of uncertainty are spatial variability, transformation uncertainty, statistical uncertainty, and measurement error. The last three are likely to be included in the interpretation of experimental results. Hence, when high levels of skill and experience are

Table 1. Equipment and specimen characteristics used by each laboratory

| Test item | Laboratory (or experimenter) number | | | | | | | |
| | 1 | 2 | | 3 | 4 | 5 | | |
		2a	2b			5a	5b	5c
Specimen height, H (cm)	12.5	10	15	12.5	17	10	15	20
Specimen diameter, D (cm)	5	5	7.5	5	7	5	7.5	10
H/D	2.5	2.0	2.0	2.5	2.4	2.0	2.0	2.0
Thickness of latex rubber membrane (mm)	0.25	0.25	0.3	0.25	0.2	0.3	0.2	0.2
Diameter of loading piston (mm)	19.7	15	15	16	10	13	13	13
Load cell inside or outside the triaxial cell	In	In	Out	In	In	Out	Out	Out
Piston friction (in single amplitude, gf)	—	—	≈400	—	≈15	≈15	≈15	≈15
Correction for piston friction	Not needed	Not needed	Yes	Not needed	No, because piston friction is negligible			
Volume change of drainage line (cm^3/kgf/cm^2)	≈0.05	≈0.0	≈0.0	≈0.0	≈0.0	<0.01	<0.01	<0.01

Source: Data from Toki et al. (1986).

required in an experiment, e.g., cyclic undrained triaxial test of sand, obtaining an average trend (e.g., average values of lique-faction strength) is often insufficient. In addition to the average, the uncertainty should be precisely considered in result interpreta-tion. Therefore, this study uses the concept of Bayesian statistics (e.g., Gelman et al. 2013), which expresses the model parameter to be estimated as a probability distribution.

Bayesian statistics is the statistical science of making inferences in the form of the Bayesian formula

$$p(\theta_i|\mathbf{x}) = \frac{p(\mathbf{x}|\theta_i)p(\theta_i)}{\sum_{i=1}^{n}p(\mathbf{x}|\theta_i)p(\theta_i)} \propto p(\mathbf{x}|\theta_i)p(\theta_i) \qquad (1)$$

where $p(\theta_i|\mathbf{x})$ = probability distribution that the uncertain input parameter θ_i follows when the data $\mathbf{x}$ is obtained; $p(\theta_i|\mathbf{x})$ = pos-terior distribution in Bayesian statistics; $p(\mathbf{x}|\theta_i)$ = probability that the data $\mathbf{x}$ will be observed when the values of θ_i are fixed [or $p(\mathbf{x}|\theta_i)$ is called the likelihood function when regarded as a func-tion of θ_i for fixed $\mathbf{x}$]; $p(\theta_i)$ = probability distribution of θ_i when there is no data $\mathbf{x}$ (called the prior distribution in the Bayesian stat-istical model); the denominator of the right-hand side $p(\mathbf{x}) = \sum_{i=1}^{n}p(\mathbf{x}|\theta_i)p(\theta_i)$ = marginal probability density function of $\mathbf{x}$ and can be regarded as a constant for normalization.

In Eq. (1), the likelihood and prior distribution are easy to cal-culate, but the calculation of the posterior distribution, $p(\theta_i|\mathbf{x})$, is generally challenging, particularly when there are many variables to be updated. Therefore, the posterior distribution is estimated by generating a large number of random samples using MCMC meth-ods from a distribution proportional to the target posterior distribu-tion, while ignoring the normalization constant, $p(\mathbf{x})$ (e.g., Gelman et al. 2013). When the Markov chain reaches a state of equilibrium, the samples drawn from the Markov chain are the same as those of the target distribution; thus, the Markov chain samples can be used to investigate the target distribution properties. This concept is Bayesian estimation using MCMC, which has been gaining recog-nition in the field of geotechnical engineering (Ching et al. 2011; Santoso et al. 2011; Wang et al. 2010; Zhang et al. 2014, 2012, 2010). In this study, the No-U-Turn Sampler (NUTS) (Hoffman and Gelman 2011), an implementation of the hybrid (or also known as Hamiltonian) Monte Carlo (HMC) method (Betancourt and Girolami 2013; Duane et al. 1987), was used as the MCMC sam-pling method. NUTS is efficient, even when the number of param-eters is large.

The other advantages of Bayesian statistics are as follows: (1) they provide not only the posterior distribution of the model uncertainty variables but also the posterior distributions of the model input parameters; and (2) the posterior distribution estimated from existing datasets can be used as a prior distribution to update the posterior distribution when similar laboratory test results be-come newly available. The use of hierarchical Bayesian modeling, described in the next subsection, enables the estimation of the pos-terior distribution by considering the influence of different exper-imenters on the test results.

In this study, the likelihood of observing a measured cyclic re-sistance ratio (CRR) in the triaxial tests was assumed to follow a normal (or Gaussian) distribution as follows:

$$f(x_i|\mu_{\mathrm{CRR}}, \sigma_{\mathrm{CRR}}) = \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{CRR}}}\exp\left[-\frac{(x_i - \mu_{\mathrm{CRR}})^2}{2\sigma_{\mathrm{CRR}}^2}\right] \qquad (2)$$

where μ_{CRR} and σ_{CRR} = mean and standard deviation, respectively, of the normal distribution. Assuming that these observations are stat-istically independent, the chance to observe $\mathbf{x} = \{x_1, x_1, \ldots, x_n\}$, in which n is the number of data (i.e., CRR), is given as the product of the likelihoods of the following observations

$$f(\mathbf{x}|\mu_{\mathrm{CRR}}, \sigma_{\mathrm{CRR}}) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma_{\mathrm{CRR}}}\exp\left[-\frac{(x_i - \mu_{\mathrm{CRR}})^2}{2\sigma_{\mathrm{CRR}}^2}\right] \qquad (3)$$

A normal distribution is assumed for the following reasons: (1) the distribution of errors resulting from the addition of many random numbers is a normal (or Gaussian) random number, regard-less of the details of the causes according to the central limit theo-rem (Taylor 1997); and (2) the distribution of errors can be regarded as normal if there are at least a few error factors in actual measure-ments (Taylor 1997). No individual differences (e.g., between experimenters) were assumed to exist in the derivation of Eq. (3). Thus, the methods can be categorized in nonhierarchical modeling.

According to the empirical equation proposed by Tatsuoka et al. (1980) for estimating CRR, the mean was modeled as follows:

$$\mu_{\mathrm{CRR}} = \mathrm{CRR}_{20}\left(\frac{N_c}{20}\right)^b \qquad (4)$$

where N_c = number of cyclic loads; CRR_{20} represents the CRR at $N_c = 20$; and b = model input parameter. CRR_{20} was assumed to have a linear relationship with the relative density, D_r (%), as follows:

$$\mathrm{CRR}_{20} = a\frac{D_r}{100} \qquad (5)$$

where a = model input parameter. Although a higher-order equa-tion (e.g., a second-order equation) can be used for D_r in place of Eq. (5), this simple type of modeling was selected because it can be regarded as linear in a limited range of relative densities. The parameters a and b in Eqs. (4) and (5) are the parameters to be estimated by MCMC-based Bayesian modeling. The model equations in this study are relatively simple, but the methods illus-trated in this study can be applied to calibrate more complex mod-els incorporating other quantitative indicators as parameters. For example, the saturation degree (or B-value) may influence the test results in addition to D_r; however, the effect was not considered in this study because the B-value was 0.96 or higher in all experiments.

Given that there is no prior knowledge about the model input parameters (i.e., a and b) and the model uncertainty variable σ_{CRR}, the following noninformative prior distributions were adopted

$$f(a) \propto 1, \quad -\infty < a < +\infty \qquad (6)$$

$$f(b) \propto 1, \quad -\infty < b < +\infty \qquad (7)$$

$$f(\sigma_{\mathrm{CRR}}) \propto 1, \quad 0 < \sigma_{\mathrm{CRR}} < +\infty \qquad (8)$$

Eq. (8) indicates that σ_{CRR} has an equal chance to be any positive number. With the substitution of these prior distributions and the laboratory test data (i.e., CRR, N_c, D_r) into Eq. (3), the Bayesian formulation for calculating the posterior probability density function (PDF) of all the uncertain parameters can be given as follows:

$$f(\mu_{\mathrm{CRR}}, \sigma_{\mathrm{CRR}}|\mathbf{x}) = f(a, b, \sigma_{\mathrm{CRR}}|\mathbf{x})$$
$$= kf(a)f(b)f(\sigma_{\mathrm{CRR}})\prod_{i=1}^{n}f(x_i|a, b, \sigma_{\mathrm{CRR}}) \qquad (9)$$

where k = normalization constant intended to make the integration of the posterior PDF equal to unit. As mentioned earlier, the posterior samples were collected via hybrid MCMC simulation; the sample length of a Markov chain was set as 30,000 in this study according to preliminary sensitivity analyses, showing that the length is ad-equate for robust estimation of the posterior statistics. In practice, the number of samples to be discarded is determined by monitoring

© ASCE   04021188-3   J. Geotech. Geoenviron. Eng.

J. Geotech. Geoenviron. Eng., 2022, 148(2): 04021188

a trace plot of samples. In this study, according to Driscoll and Maki (2007), the second half of the samples in each chain was used for estimating the posterior of model input parameters by discarding the first-half samples.

Hierarchical Bayesian Model

The nonhierarchical Bayesian modeling in the previous subsection can quantitatively consider the effect of differences in D_r on the laboratory test results, as the mean value of CRR is a function of D_r, as shown in Eqs. (4) and (5). In the modeling, the parameters a and b in the equations were common across all the experimental results; this implies that all the experimental results were assumed to be independent replications (i.e., the results from different facilities or experimenters followed a common probability distribution). However, the cooperative test program of the cyclic undrained triaxial tests described earlier was considered a pseudoreplication, although all the participants followed common specifications as much as possible. Individual differences were believed to exist due to the differences between the experimenters (or facilities); the laboratory tests required high levels of skill and experience, and it may have been challenging to keep the experimenters' practices at the same level. In this case, taking an overall average while ignoring the individual characteristics may lead to an over or underestimation of the variation range for each experimenter; i.e., the range of one experimenter might be influenced considerably by the other experimenters' data. Likewise, the (nonhierarchical) Bayesian estimation based only on the results from an individual experimenter (ignores other experimenters' data) may be too sensitive to the individual characteristics. Typically, the data obtained from one experimenter are insufficient for statistical processing with a sufficient degree of accuracy, partly due to time and cost constraints. This trade-off between the poor applicability of a generic model and the imprecision of an individual-specific model has been noted (Ching et al. 2018, 2017).

Therefore, we introduce the concept of a hierarchical model (e.g., Gelman et al. 2013) to the MCMC-based Bayesian estimation in this study. In the hierarchical Bayesian model, the model input parameters that vary from experimenter (or bin) to experimenter are assumed to follow a hierarchical prior distribution instead of a noninformative prior distribution; overall trends (include the results from other experimenters) are considered to construct a more realistic statistical model that incorporates individual differences. The hierarchical prior distribution here means that a further prior distribution is set to the parameters of the prior distribution, and this further prior distribution is called a superprior distribution.

The formulation of the hierarchical Bayesian modeling is described subsequently. First, let $\mathbf{x}_i = \{x_{i1}, x_{i2}, \ldots, x_{in_i}\}$ denote the CRR data in the triaxial tests performed by the ith experimenter. The number of data, n_i, may vary from experimenter to experimenter. Similar to the formulation of nonhierarchical modeling given by Eq. (3), the chance to observe $\mathbf{x}_i$ is given as the product of the likelihoods [Eq. (2)] of observations

$$f(\mathbf{x}_i|\mu_i, \sigma_i) = \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(x_{ij} - \mu_i)^2}{2\sigma_i^2}\right] \qquad (10)$$

where μ_i and σ_i = mean and standard deviation, respectively, that are specific to the ith experimenter. As in the formulation of non-hierarchical modeling given by Eq. (4) with Eq. (5), the mean was modeled as

$$\mu_i = a_i \frac{D_r}{100} \left(\frac{N_c}{20}\right)^{b_i} \qquad (11)$$

where a_i and b_i = model input parameters specific to the ith experimenter.

Next, let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ denote the measured CRR data obtained from N experimenters. When the statistics on the CRR specific to each experimenter are known, the measurements of the CRR from different experimenters are statistically independent. Thus, the chance to observe $\mathbf{X}$ can be given as the product of the likelihoods of observing $\mathbf{x}_i(i = 1, 2, \ldots, N)$ as follows:

$$f(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\sigma}) = \prod_{i=1}^{N}\left\{\prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(x_{ij} - \mu_i)^2}{2\sigma_i^2}\right]\right\} \qquad (12)$$

where $\boldsymbol{\mu} = \{\mu_1, \mu_2, \ldots, \mu_N\}$ and $\boldsymbol{\sigma} = \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$.

A hierarchical prior distribution incorporating a superprior distribution should be introduced in the hierarchical Bayesian modeling to consider the overall tendency of all the experimenters in addition to their individual characteristics. In this study, the parameters a_i and b_i were assumed to follow a normal distribution instead of a noninformative prior distribution

$$f(a_i|\mu_a, \sigma_a) = \frac{1}{\sqrt{2\pi}\sigma_a} \exp\left[-\frac{(a_i - \mu_a)^2}{2\sigma_a^2}\right] \quad (i = 1, 2, \ldots, N) \qquad (13)$$

$$f(b_i|\mu_b, \sigma_b) = \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left[-\frac{(b_i - \mu_b)^2}{2\sigma_b^2}\right] \quad (i = 1, 2, \ldots, N) \qquad (14)$$

where the means μ_a and μ_b and the standard deviations σ_a and σ_b are common across all experimenters and are sometimes called hyperparameters. Similarly, σ_i is considered to follow a normal distribution

$$f(\sigma_i|\mu_\sigma, \sigma_\sigma) = \frac{1}{\sqrt{2\pi}\sigma_\sigma} \exp\left[-\frac{(\sigma_i - \mu_\sigma)^2}{2\sigma_\sigma^2}\right] \quad (i = 1, 2, \ldots, N) \qquad (15)$$

σ_i should be positive and modeled to follow a log-normal distribution to avoid a negative value, but a simpler modeling was adopted in this study because all samples of σ_i were positive, even under the assumption of a normal distribution, as illustrated in Fig. 7. Given the lack of prior knowledge about the hyperparameters, the following noninformative prior distributions were adopted

$$f(\mu_a) \propto 1, \quad -\infty < \mu_a < +\infty \qquad (16)$$

$$f(\mu_b) \propto 1, \quad -\infty < \mu_b < +\infty \qquad (17)$$

$$f(\mu_\sigma) \propto 1, \quad -\infty < \mu_\sigma < +\infty \qquad (18)$$

$$f(\sigma_a) \propto 1, \quad 0 < \sigma_a < +\infty \qquad (19)$$

$$f(\sigma_b) \propto 1, \quad 0 < \sigma_b < +\infty \qquad (20)$$

$$f(\sigma_\sigma) \propto 1, \quad 0 < \sigma_\sigma < +\infty \qquad (21)$$

From the preceding equations, Bayesian estimation using the hierarchical model gives the posterior PDF of all the uncertain parameters as follows (Gelman et al. 2013; Givens and Hoeting 2005):

$$f(\mu_a, \mu_b, \mu_\sigma, \sigma_a, \sigma_b, \sigma_\sigma, \boldsymbol{\mu}, \boldsymbol{\sigma}|\mathbf{X}) = kq(\mu_a, \mu_b, \mu_\sigma, \sigma_a, \sigma_b, \sigma_\sigma, \boldsymbol{\mu}, \boldsymbol{\sigma}) \qquad (22)$$
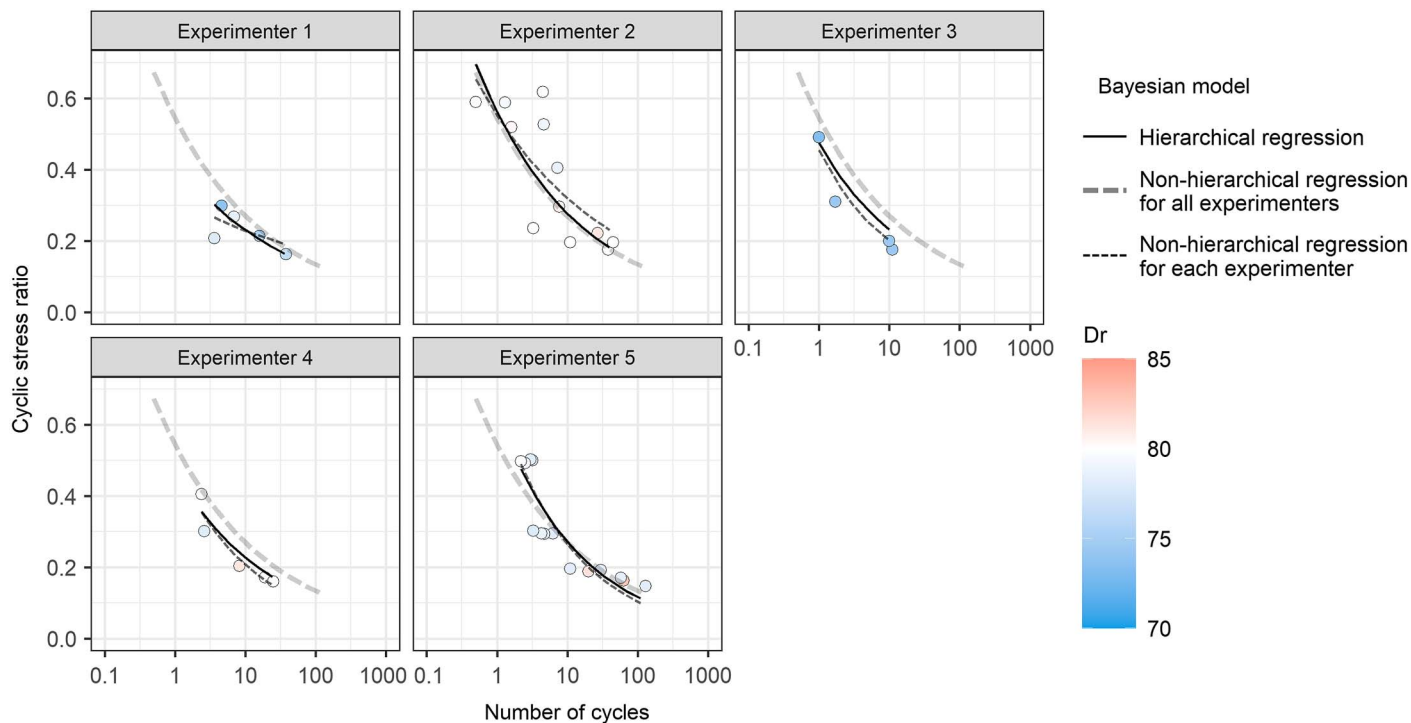
© ASCE

04021188-4

J. Geotech. Geoenviron. Eng.

Fig. 1. Comparison of the hierarchical model that considers the individual differences between experimenters with the nonhierarchical models.

where k = normalization constant intended to make the integration of the posterior PDF equal to unit. The unnormalized PDF, q, can be given as follows:

$$q(\mu_a, \mu_b, \mu_\sigma, \sigma_a, \sigma_b, \sigma_\sigma, \boldsymbol{\mu}, \boldsymbol{\sigma})$$
$$= q(\mu_a, \mu_b, \mu_\sigma, \sigma_a, \sigma_b, \sigma_\sigma, \mathbf{a}, \mathbf{b}, \boldsymbol{\sigma})$$
$$= f(\mathbf{X}|\mathbf{a}, \mathbf{b}, \boldsymbol{\sigma})f(\mu_a)f(\mu_b)f(\mu_\sigma)f(\sigma_a)f(\sigma_b)f(\sigma_\sigma)$$
$$\times \prod_{i=1}^{N} f(a_i|\mu_a, \sigma_a) \times \prod_{i=1}^{N} f(b_i|\mu_b, \sigma_b) \times \prod_{i=1}^{N} f(\sigma_i|\mu_\sigma, \sigma_\sigma) \quad (23)$$

where $\mathbf{a} = \{a_1, a_2, \ldots, a_N\}$; and $\mathbf{b} = \{b_1, b_2, \ldots, b_N\}$. As in the nonhierarchical Bayesian modeling, the sample length of a Markov chain was set as 30,000, of which the second half was used for estimating the posterior.

In Eq. (23), a total of $(3N + 6)$ uncertain variables should be updated. If the number of experimenters is $N = 5$, as shown in Table 1, the number of uncertain variables to be updated will become 21. Thus, Eq. (23) can be categorized as a high-dimensional updating problem. If the prior distributions are assumed to be conjugate (Ang and Tang 2007; Gelman et al. 2013), then conventional Monte Carlo simulation can be applied to solve the problem numerically. If the prior distributions are not conjugate, as is often the case in geotechnical engineering, MCMC-based Bayesian modeling becomes a powerful tool for solving such complex problems with no need for large sample approximation.

Bayesian Estimation of Liquefaction Resistance Considering Individual Differences between Experimenters

The hierarchical Bayesian model was applied to the series of cyclic undrained triaxial tests with $D_r = 80\%$ in the cooperative test program (Toki et al. 1986) in consideration of the individual differences between the five experimenters (Table 1). As listed in the table,

some specifications were not common between the experimenters, but it was assumed that these discrepancies would not significantly affect the test results. Toki et al. (1986) and Tatsuoka et al. (1986) concluded that the cyclic undrained triaxial strength is insensitive to the H/D ratios of specimens between 1.5 and 2.7. They also reported that the specimen diameter, D, has a nonnegligible effect on the test results, as discussed in the next section. Strictly speaking, individual variations due to the participation of different experimenters should be defined in a narrow sense under the ideal condition that all specifications are common between the experimenters, but experimenter differences may be interpreted broadly to include specifications that are not common.
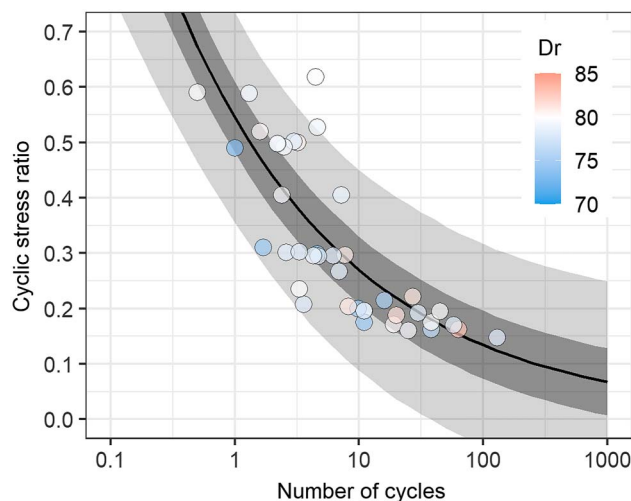

Fig. 2. Bayesian prediction intervals estimated by the nonhierarchical model for all experimenters (the dark and light gray areas represent 50% and 95% intervals, respectively).

The regression (or average) curves of CRR for $D_r = 80\%$ estimated by the hierarchical Bayesian model are drawn as solid lines in Fig. 1 for each experimenter along with the experimental plots. The figure also compares the nonhierarchical Bayesian estimation results for all bins (the nonhierarchical model was applied to the results of all experimenters, resulting in the presence of a common CRR curve regardless of the experimenter) and for each bin (the model was applied only to the results of each experimenter). As shown in the color legend, there is a slight deviation from the target D_r in the laboratory tests, which is quantitatively considered in Eqs. (4) and (11) for the nonhierarchical and hierarchical estimations, respectively.
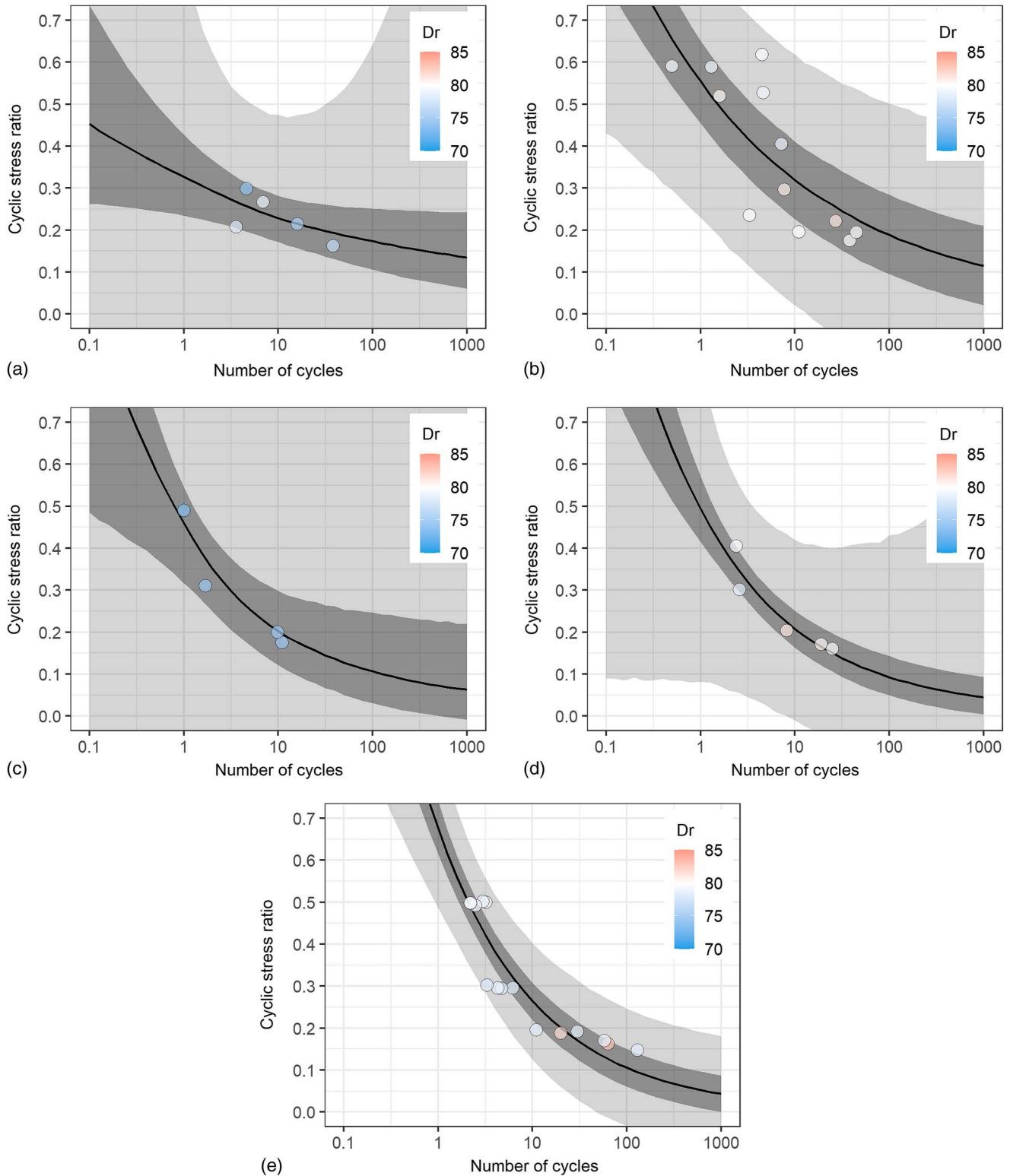

Fig. 3. Bayesian prediction intervals estimated by the nonhierarchical model for each experimenter (the dark and light gray areas represent 50% and 95% intervals, respectively): (a) Experimenter 1; (b) Experimenter 2; (c) Experimenter 3; (d) Experimenter 4; and (e) Experimenter 5.

The average CRR curve of the nonhierarchical Bayesian model for each experimenter (thin dashed lines in the figure) best represents each experimenter's results because other experimenters' results were not considered in the estimation. However, some of the estimated CRR curves deviate from the nonhierarchical Bayesian regression for all experimenters' results (thick dashed line in the figure), influenced considerably by the individuality of each experimenter. For example, the curves for Experimenters 1 and 3 differ from the overall average in terms of slope and position, respectively. This means that the regression for each bin is too sensitive to the
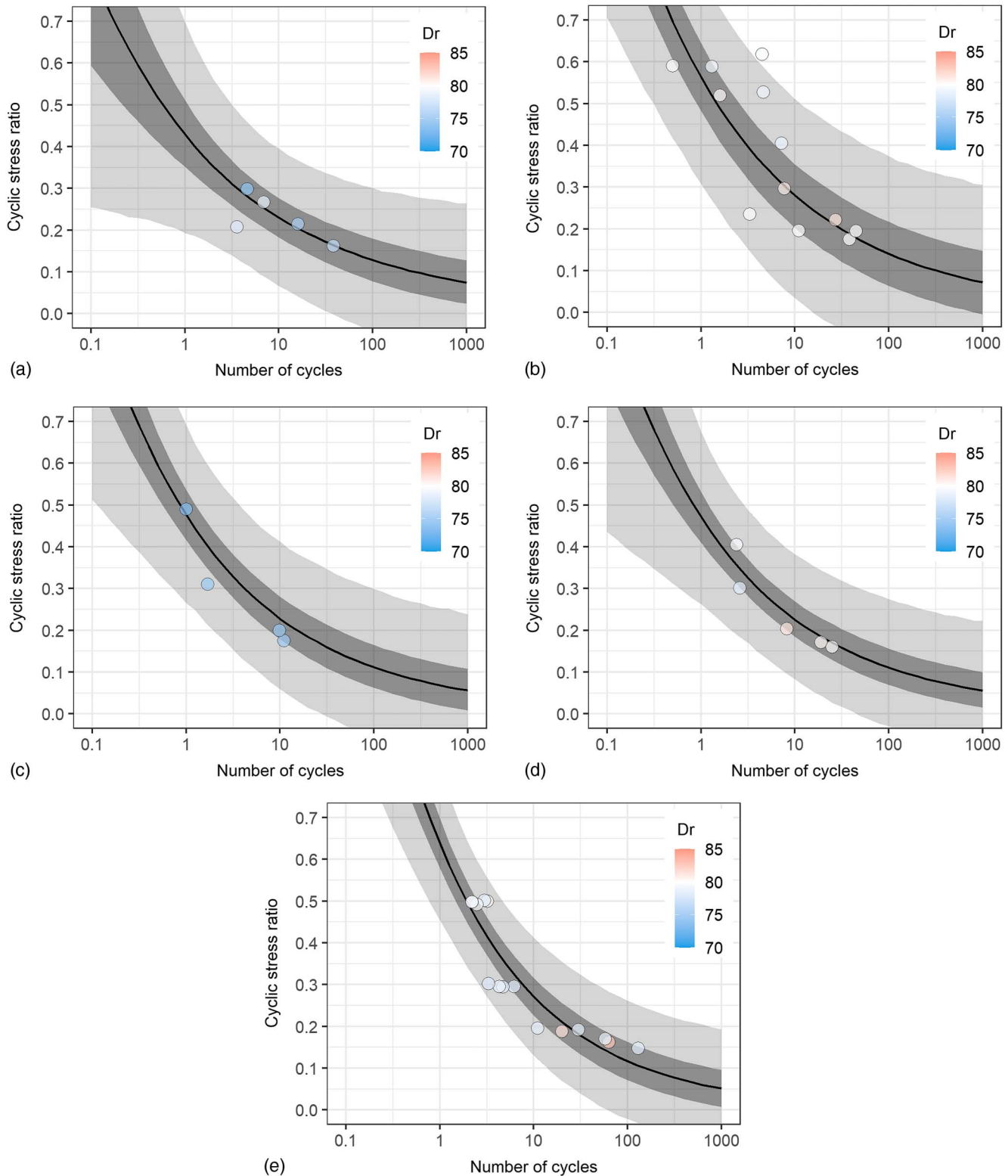

Fig. 4. Bayesian prediction intervals estimated by the hierarchical model that considers the individual differences between experimenters (the dark and light gray areas represent 50% and 95% intervals, respectively): (a) Experimenter 1; (b) Experimenter 2; (c) Experimenter 3; (d) Experimenter 4; and (e) Experimenter 5.

© ASCE   04021188-7   J. Geotech. Geoenviron. Eng.

J. Geotech. Geoenviron. Eng., 2022, 148(2): 04021188

individual characteristics, while the regression for all bins is too generic and thus poorly applicable. The CRR curves estimated by the hierarchical Bayesian model demonstrate that such trade-offs are well resolved. Even though the number of test cases is limited, as in the case of Experimenters 1 and 3, a CRR curve that considers individuality and does not stray too far from the overall trend (represented by the thick dashed line) can be obtained by referring to the other experimenters' results. When the number of test cases is relatively large and the variability is not so large, as in the case of Experimenter 5, the hierarchical regression result is approximately equal to the nonhierarchical regression result for each bin.

The Bayesian prediction intervals estimated by the nonhierarchical model for all experimenters' results are shown in Fig. 2. The average curve is represented by the solid line, and the dark gray and light gray areas represent 50% and 95% intervals, respectively. Moreover, Figs. 3 and 4 show the Bayesian prediction intervals for each experimenter that are obtained from the nonhierarchical model for each experimenter and from the hierarchical model, respectively. Figs. 3(a, c, d) indicate that, when the number of test cases is limited, the 95% prediction intervals are very wide, particularly in the region with no test data, compared with the intervals in Figs. 2 and 3(b and e), where the number of test cases is large. However, the very wide prediction intervals are suppressed in Fig. 4 where the hierarchical Bayesian model was used to consider the overall trend in Fig. 2 instead of ignoring other experimenters' results. This is the advantage of implementing a hierarchical Bayesian model; an
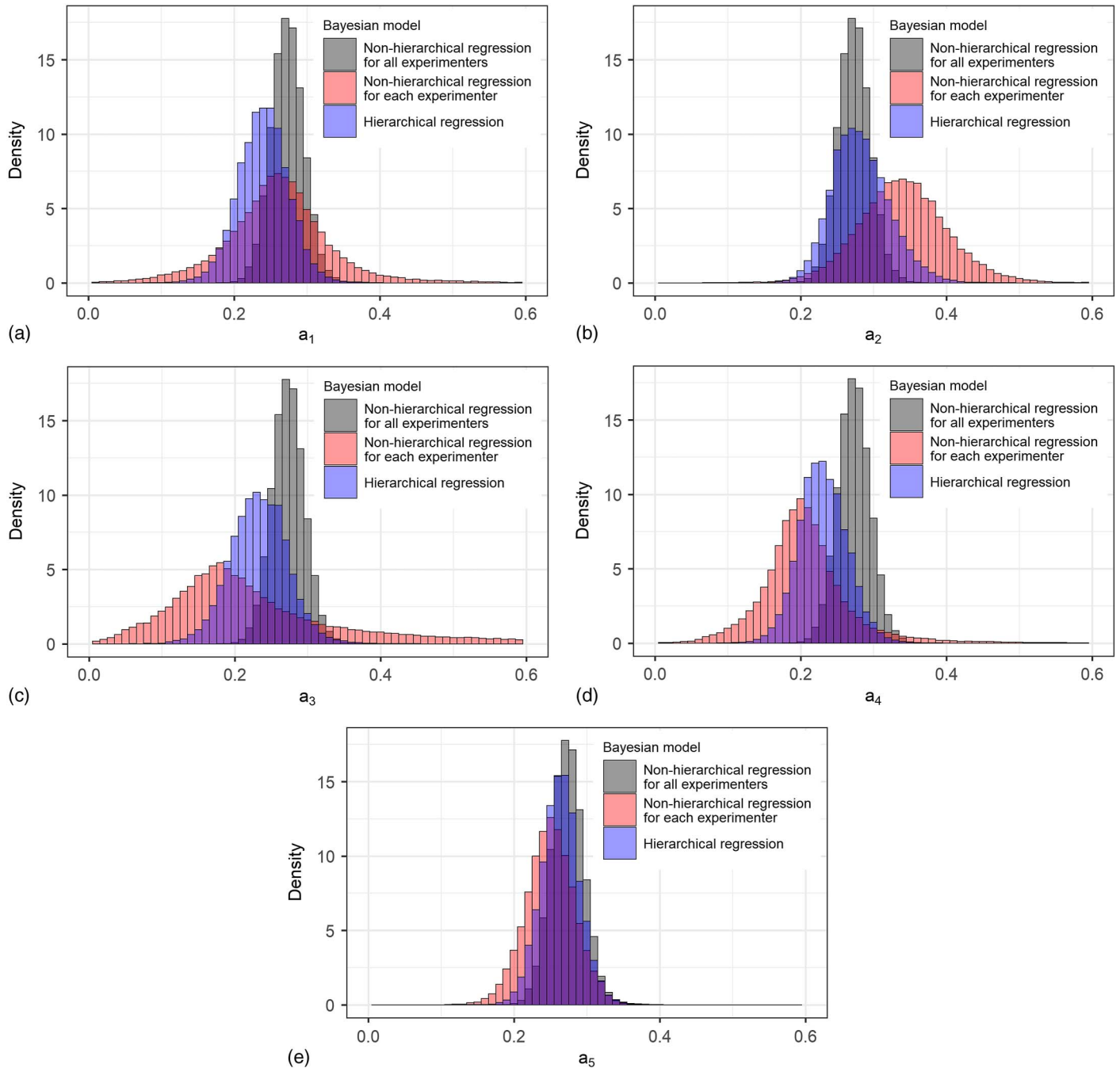

Fig. 5. Histograms of parameter a in the nonhierarchical and hierarchical models: (a) Experimenter 1; (b) Experimenter 2; (c) Experimenter 3; (d) Experimenter 4; and (e) Experimenter 5.

© ASCE 04021188-8 J. Geotech. Geoenviron. Eng.

J. Geotech. Geoenviron. Eng., 2022, 148(2): 04021188

experimenter can make reasonable predictions even with a limited number of test cases. Note that the test results used in this study do not have significant outliers, as shown in Fig. 2; because the presence of outliers may affect the robustness of statistical estimation methods, future research is needed on this point.

Figs. 5 and 6 compare the histograms (or posterior distributions) of the model input parameters, a_i (or a) and b_i (or b), respectively, for each experimenter among the three different Bayesian models. The gray histogram bars, which depict the nonhierarchical regression model for all bins, are common between all experimenters. As shown in the figures, the MCMC samples for these parameters follow a normal distribution regardless of the Bayesian model; the target distributions in the hierarchical model were set as shown

in Eqs. (13) and (14), which confirms that the MCMC sampling was successfully performed. However, the distribution spread differs from model to model, particularly for the parameter a_i, which determines the position of a CRR curve; the distribution is narrow in the nonhierarchical regression model for all bins, but it is broad in the nonhierarchical regression model for each bin, particularly for Experimenters 1 and 3, where the number of test cases is limited. As for Experimenter 3, the peak location (i.e., mean value) of the distribution is also different between the models, as shown in Fig. 5(c). Even when the number is relatively large, a similar tendency can be observed in Fig. 5(b) for Experimenter 2. Thus, reliable experimental data with low variability should be obtained to accurately estimate the parameter a_i using a nonhierarchical model
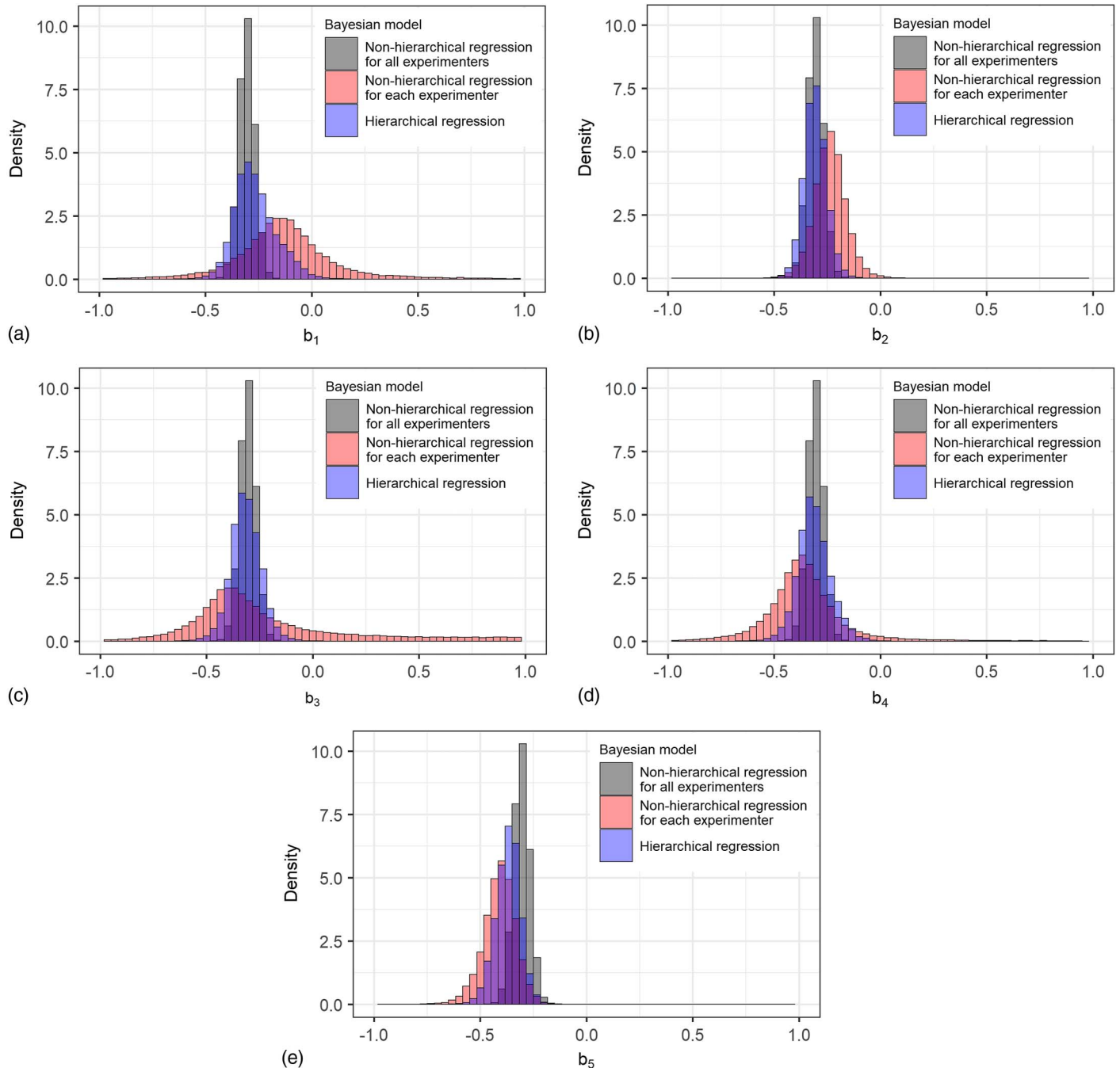

Fig. 6. Histograms of parameter b in the nonhierarchical and hierarchical models: (a) Experimenter 1; (b) Experimenter 2; (c) Experimenter 3; (d) Experimenter 4; and (e) Experimenter 5.

for each bin, even if the number of test data is sufficient. In other words, the hierarchical model is considered superior to the nonhierarchical model, except when there is enough reliable data in a particular bin. By contrast, the distributions of the parameter b_i, which determines the slope of a CRR curve from the nonhierarchical model for each bin, do not deviate significantly from the distribution of the nonhierarchical model for all bins when a sufficient number of data is available, as shown in Figs. 6(b and e). For any of the model input parameters, the hierarchical Bayesian model can be applied to reasonably estimate the parameter distribution considering the individual differences of the experimenters in addition to the overall trend. For example, the peak location and width of the

blue histogram bars in Fig. 5(c) are affected by both the gray and red histogram bars.

The histograms (or posterior distributions) of the parameter σ_i in Eq. (10) [or σ_{CRR} in Eq. (3)], which show the variability of the proposed models [i.e., Eqs. (11) and (4)], are compared in Fig. 7 for each experimenter. The MCMC samples obtained from the nonhierarchical regression model for all bins (gray histogram bars) follow a normal distribution, whereas the samples obtained from the nonhierarchical model for each bin (red histogram bars) are close to a log-normal distribution. As a result, the distribution of the hierarchical model is a mixture of both, as shown by the blue histogram bars. Nonetheless, the distribution does not deviate considerably
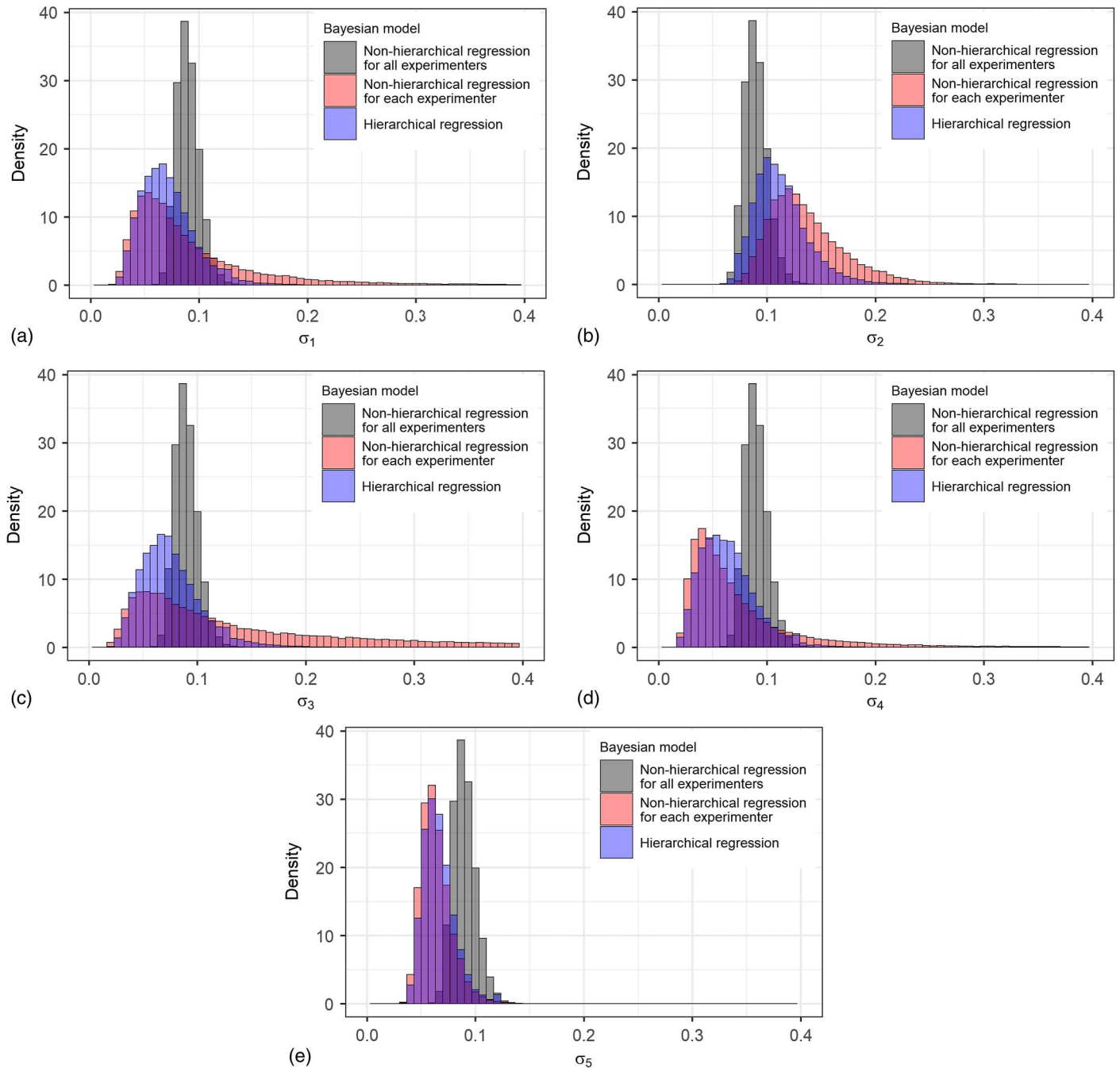

Fig. 7. Histograms of parameter σ in the non-hierarchical and hierarchical models: (a) Experimenter 1; (b) Experimenter 2; (c) Experimenter 3; (d) Experimenter 4; and (e) Experimenter 5.

© ASCE 04021188-10 J. Geotech. Geoenviron. Eng.

J. Geotech. Geoenviron. Eng., 2022, 148(2): 04021188

from the normal distribution given by Eq. (15), as the right-hand side tail of the mixture distribution is not as heavy as that of the log-normal distribution.

Fig. 8 shows the histograms (or posterior distributions) of the means (i.e., μ_a, μ_b, and μ_σ) and standard deviations (i.e., σ_a, σ_b, and σ_σ) of the model parameters a_i, b_i, and σ_i in the hierarchical Bayesian model. As explained in Eqs. (13)–(21), these prior distributions were assumed noninformative. The figure indicates that the posterior distributions of the means can be modeled as a normal distribution, while those of the standard deviations follow a log-normal distribution. When a different experimenter newly performs similar laboratory tests, these posterior distributions can be used as prior distributions, instead of Eqs. (16)–(21), to estimate

the posterior distribution of the model input parameters specific to the experimenter (a_i, b_i, and σ_i). This is an advantage of Bayesian statistics using a hierarchical model. The model parameters for other experimenters that have already been estimated, as shown in Figs. 5–7, are also updated by adding new experimental results.

Discussion

The previous section introduces a methodology for estimating the variability (or distribution) of CRR and model input parameters using an MCMC-based hierarchical Bayesian model, with focus on the individual differences between experimenters. Other individual
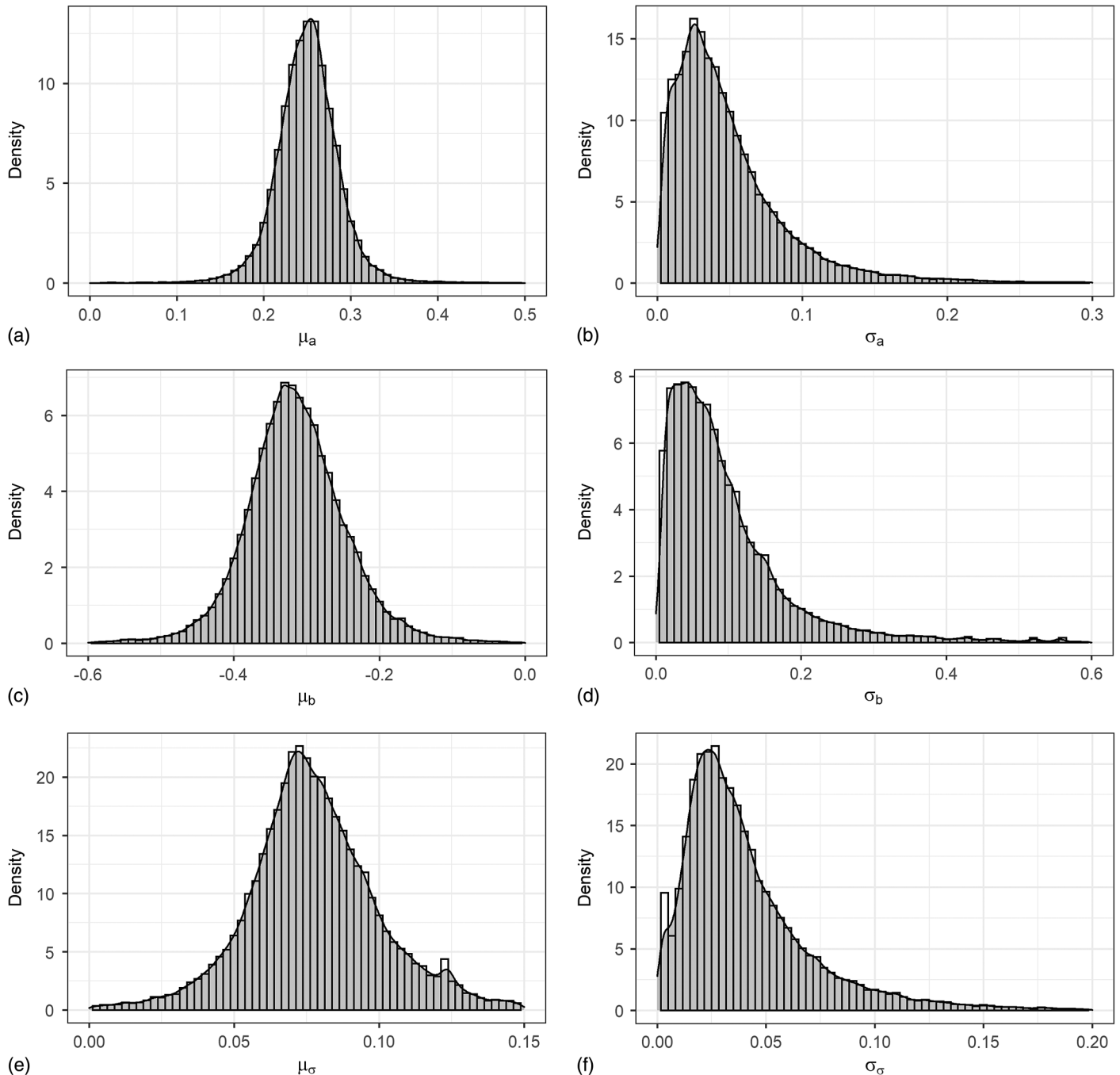

Fig. 8. Histograms of different variables in the hierarchical model: (a) μa; (b) σa; (c) μb; (d) σb; (e) $\mu\sigma$; and (f) $\sigma\sigma$.
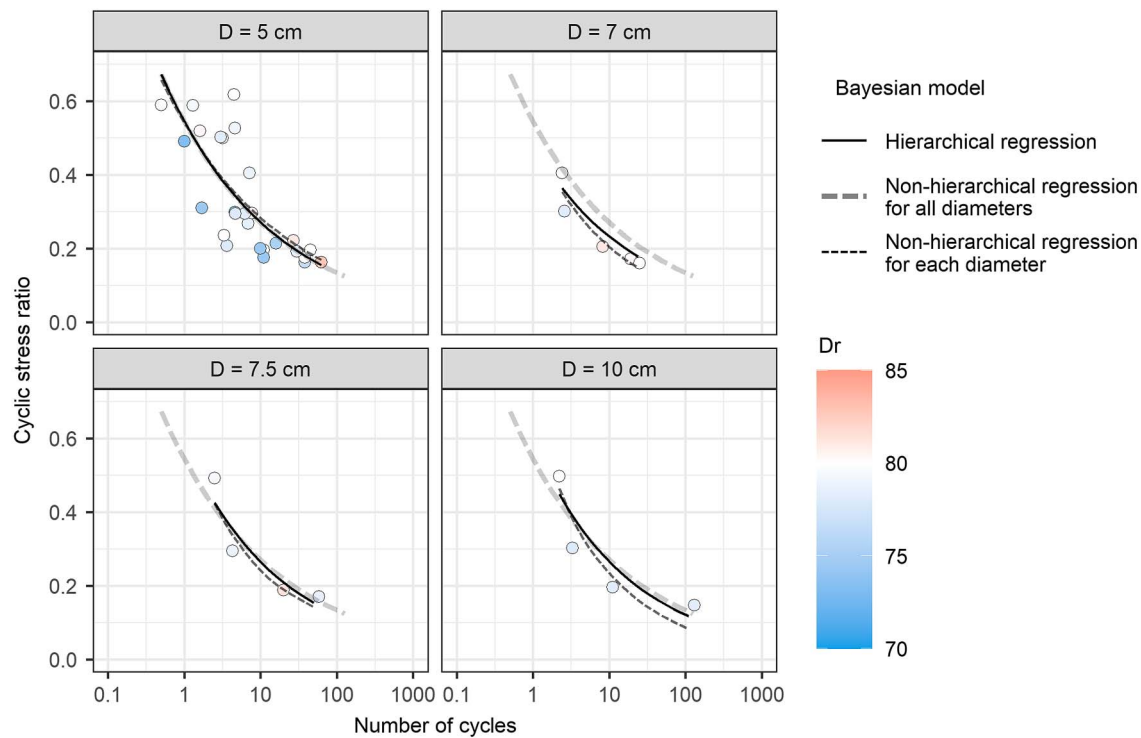
Fig. 9. Comparison of the hierarchical model that considers the difference in the specimen diameter with the nonhierarchical models.

differences that are difficult to quantify in laboratory tests, e.g., differences in testing apparatus and specimen size, can also be handled using the same methodology. In this section, to illustrate another example, we quantify the variation in CRR by Bayesian estimation using a hierarchical model, considering the specimen diameter as an individual difference. The specimen diameter is taken into account because Toki et al. (1986) and Tatsuoka et al. (1986) reported that any difference in membrane penetration due to the difference in specimen diameter has a nonnegligible effect on test results.

The average curves of CRR estimated by the hierarchical Bayesian model are compared with those by the nonhierarchical Bayesian models in Fig. 9, along with the experimental plots. Comparisons are made for each diameter, but the nonhierarchical regression results for all bins (thick dashed lines) are common regardless of the specimen diameter. At $D = 5$ cm, the estimated CRR curves from the three Bayesian models are approximately equal, and this can be attributed to the fact that the number of test cases is sufficient. In the cases of $D = 7.5$ cm and $D = 10$ cm, the CRR curves estimated using the non-hierarchical regression model for each bin are slightly dragged by the results for each diameter, but the application of the hierarchical model produces estimation curves that are almost identical to that for $D = 5$ cm. Meanwhile, the CRR curve estimated for $D = 7$ cm using the hierarchical Bayesian model differs from the curves for the other diameter and an effect of specimen diameter initially appears. However, Table 1 shows that all the experiments with $D = 7$ cm were conducted by Experimenter 4, and the CRR curve estimated for the experimenter by the hierarchical model in Fig. 1 is smaller than the overall average curve. This means that the result of the hierarchical model for $D = 7$ cm, shown in Fig. 9, is influenced by the individual differences of the experimenters. Thus, it is difficult to conclude that the difference in specimen diameter affects the experimental results, as described by Toki et al. (1986) and Tatsuoka et al. (1986), in the Bayesian estimation using a hierarchical model. For the Bayesian prediction intervals obtained by the nonhierarchical model, shown in Fig. 10,

the individual differences between specimen diameters also show only a small effect.

Conclusions

Cyclic undrained triaxial tests are commonly used in research and practical design to evaluate the liquefaction resistance of sandy soils. This paper proposes a methodology to evaluate liquefaction resistance by considering the variability or uncertainty associated with experimentation, using the concept of Bayesian statistics. The results of five research institutes' cooperative test program involving cyclic undrained triaxial tests on saturated Toyoura sand were selected in this study as experimental data. In addition to conventional nonhierarchical Bayesian modeling, Bayesian estimation using a hierarchical model was introduced to properly incorporate the variability caused by individual discrepancies (e.g., due to participation of different experimenters) into the liquefaction resistance evaluation. In both the nonhierarchical and hierarchical modeling, a procedure based on hybrid MCMC simulation was adopted to solve the Bayesian updating equation.

The regression curve of CRR estimated by the nonhierarchical model for all experimenters' results was too generic and poorly applicable, while the curve estimated by the nonhierarchical model for each experimenter's results sometimes deviated from the overall trend, influenced by the individual characteristics. The CRR curves estimated by the hierarchical Bayesian model considered the individuality of each experimenter and did not stray far from the overall trend by referring to the other experimenters' results, even when an experimenter had a limited number of test cases.

The wide posterior distributions of the model input parameters and model variability estimated by the nonhierarchical Bayesian model for each experimenter can be suppressed by considering the other experimenters' results (i.e., overall trend) using the hierarchical model, thereby allowing an experimenter to make reasonable
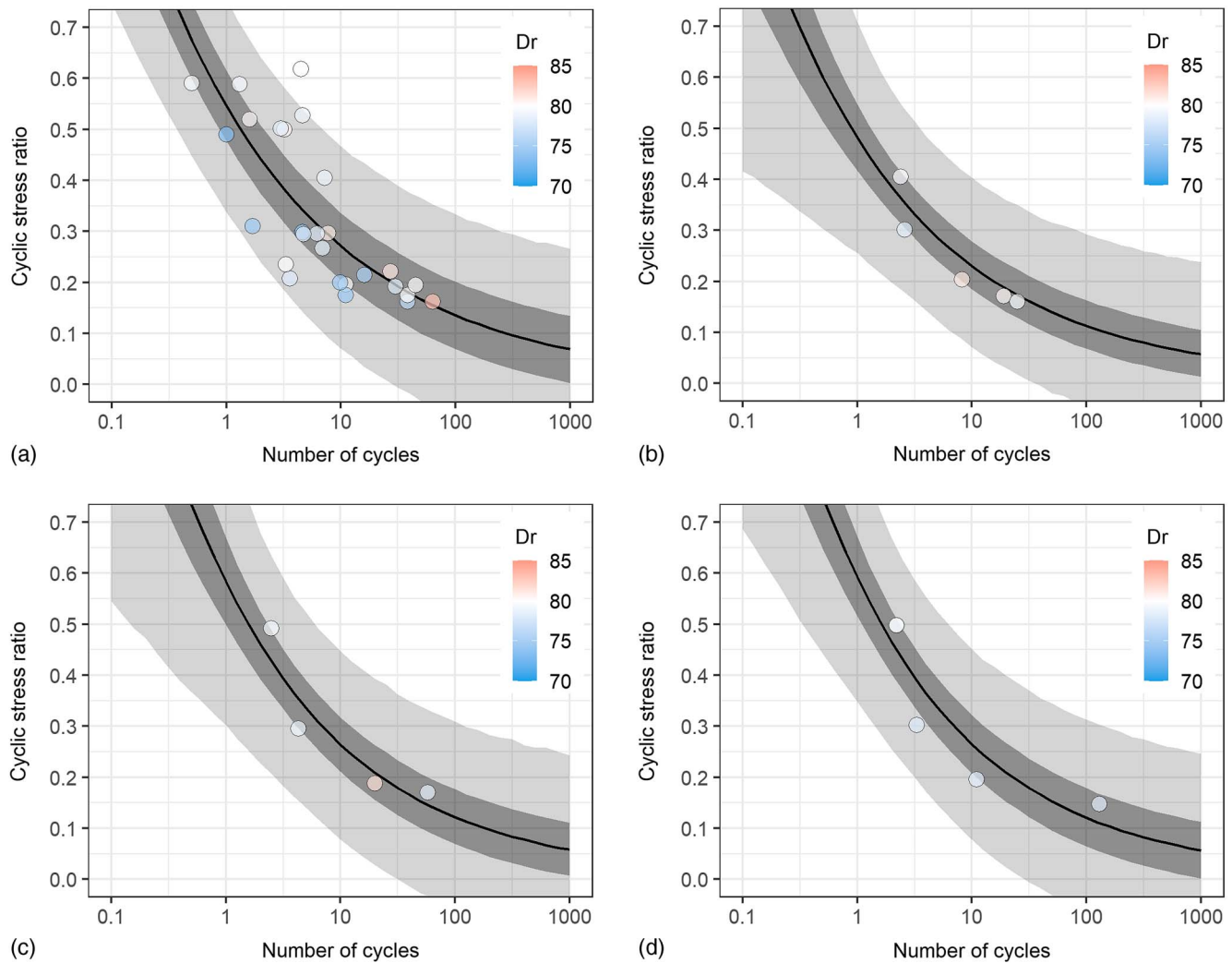
Fig. 10. Bayesian prediction intervals estimated by the hierarchical model that considers the difference in the specimen diameter (the dark and light gray areas represent 50% and 95% intervals, respectively): (a) D = 5 cm; (b) D = 7 cm; (c) D = 7.5 cm; and (d) D = 10 cm.

predictions even with a limited number of test cases for an experimenter. The hierarchical Bayesian model also estimated the posterior distributions of the means and standard deviations of the experimenter-specific model input parameters. These posterior distributions can be used as prior distributions instead of noninformative priors to update the posterior distribution of model input parameters when a different experimenter newly performs similar laboratory tests.

The proposed methodology can also be used to estimate the variability of liquefaction resistance considering other individual differences in laboratory tests, e.g., differences in testing apparatus and specimen size. The advantage of hierarchical Bayesian modeling is that it can reasonably consider factors that affect the experimental results but are difficult to quantify. This may allow for a more rational evaluation of the response of liquefiable ground by means of numerical analysis method, reflecting the realistic variability of model parameters obtained from laboratory test results.

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

References

Ang, A. H.-S., and W. H. Tang. 2007. *Probability concepts in engineering planning and design: Emphasis on application to civil and environmental engineering*. Hoboken, NJ: Wiley.

Betancourt, M. J., and M. Girolami. 2013. "Hamiltonian Monte Carlo for hierarchical models." Accessed December 1, 2013. https://ui.adsabs.harvard.edu/abs/2013arXiv1312.0906B.

Ching, J., K.-H. Li, K.-K. Phoon, and M.-C. Weng. 2018. "Generic transformation models for some intact rock properties." *Can. Geotech. J.* 55 (12): 1702–1741. https://doi.org/10.1139/cgj-2017-0537.

Ching, J., G.-H. Lin, J.-R. Chen, and K.-K. Phoon. 2017. "Transformation models for effective friction angle and relative density calibrated based on generic database of coarse-grained soils." *Can. Geotech. J.* 54 (4): 481–501. https://doi.org/10.1139/cgj-2016-0318.

Ching, J., H.-D. Lin, and M.-T. Yen. 2011. "Calibrating resistance factors of single bored piles based on incomplete load test results." *J. Eng. Mech.* 137 (5): 309–323. https://doi.org/10.1061/(ASCE)EM.1943-7889.0000230.

Ching, J., and K.-K. Phoon. 2019. "Constructing site-specific multivariate probability distribution model using Bayesian machine learning." *J. Eng. Mech.* 145 (1): 04018126. https://doi.org/10.1061/(asce)em.1943-7889.0001537.

Driscoll, T. A., and K. L. Maki. 2007. "Searching for rare growth factors using multicanonical Monte Carlo methods." *SIAM Rev.* 49 (4): 673–692. https://doi.org/10.1137/050637662.

Duane, S., A. D. Kennedy, B. J. Pendleton, and D. Roweth. 1987. "Hybrid Monte Carlo." *Phys. Lett. B* 195 (2): 216–222. https://doi.org/10.1016/0370-2693(87)91197-X.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. *Bayesian data analysis*. 3rd ed. Boca Raton, FL: CRC Press.

Givens, G., and J. Hoeting. 2005. *Computational statistics (wiley series in computation statistics)*. Hoboken, NJ: Wiley.

Haldar, A., and F. J. Miller. 1984. "Statistical evaluation of cyclic strength of sand." *J. Geotech. Eng.* 110 (12): 1785–1802. https://doi.org/10.1061/(ASCE)0733-9410(1984)110:12(1785).

Hoffman, M. D., and A. Gelman. 2011. "The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo." Accessed November 1, 2011. https://ui.adsabs.harvard.edu/abs/2011arXiv1111.4246H.

Hwang, J. H., C. W. Yang, and D. S. Juang. 2004. "A practical reliability-based method for assessing soil liquefaction potential." *Soil Dyn. Earthquake Eng.* 24 (9): 761–770. https://doi.org/10.1016/j.soildyn.2004.06.008.

Juang, C. H., D. V. Rosowsky, and W. H. Tang. 1999. "Reliability-based method for assessing liquefaction potential of soils." *J. Geotech. Geoenviron. Eng.* 125 (8): 684–689. https://doi.org/10.1061/(ASCE)1090-0241(1999)125:8(684).

Liao, S. S. C., D. Veneziano, and R. V. Whitman. 1988. "Regression models for evaluating liquefaction probability." *J. Geotech. Eng.* 114 (4): 389–411. https://doi.org/10.1061/(ASCE)0733-9410(1988)114:4(389).

Liu, H. C., and N. Y. Chang. 1994. "Statistical modeling of liquefaction resistance of a fine sand containing plastic fines." In *Proc., 5th U.S. Conf. on Earthquake Engineering*, 251–260. Oakland, CA: Earthquake Research Institute.

Moss, R. E., R. B. Seed, R. E. Kayen, J. P. Stewart, A. D. Kiureghian, and K. O. Cetin. 2006. "CPT-based probabilistic and deterministic assessment of in situ seismic soil liquefaction potential." *J. Geotech. Geoenviron. Eng.* 132 (8): 1032–1051. https://doi.org/10.1061/(ASCE)1090-0241(2006)132:8(1032).

Park, J. H., D. Kim, and C. K. Chung. 2012. "Implementation of Bayesian theory on LRFD of axially loaded driven piles." *Comput. Geotech.* 42 (May): 73–80. https://doi.org/10.1016/j.compgeo.2012.01.002.

Phoon, K.-K., and F. H. Kulhawy. 1999a. "Characterization of geotechnical variability." *Can. Geotech. J.* 36 (4): 612–624. https://doi.org/10.1139/t99-038.

Phoon, K.-K., and F. H. Kulhawy. 1999b. "Evaluation of geotechnical property variability." *Can. Geotech. J.* 36 (4): 625–639. https://doi.org/10.1139/t99-039.

Polito, C. 2009. "Linear regression models for predicting liquefaction during cyclic triaxial testing." *Geotech. Test. J.* 32 (4): 335–346. https://doi.org/10.1520/GTJ101752.

Polito, C. 2011. "Regression models for validating cyclic triaxial test results." *Geotech. Test. J.* 34 (2): 155–166. https://doi.org/10.1520/GTJ103249.

Santoso, A. M., K.-K. Phoon, and S.-T. Quek. 2011. "Effects of soil spatial variability on rainfall-induced landslides." *Comput. Struct.* 89 (11): 893–900. https://doi.org/10.1016/j.compstruc.2011.02.016.

Tatsuoka, F., S. Toki, S. Miura, H. Kato, M. Okamoto, S.-I. Yamada, S. Yasuda, and F. Tanizawa. 1986. "Some factors affecting cyclic undrained triaxial strength of sand." *Soils Found.* 26 (3): 99–116. https://doi.org/10.3208/sandf1972.26.3_99.

Tatsuoka, F., S. Yasuda, T. Iwasaki, and K. Tokida. 1980. "Normalized dynamic undrained strength of sands subjected to cyclic and random loading." *Soils Found.* 20 (3): 1–16. https://doi.org/10.3208/sandf1972.20.3_1.

Taylor, J. R. 1997. *An introduction to error analysis: The study of uncertainties in physical measurements*. Sausalito, CA: Univ. Science Books.

Toki, S., F. Tatsuoka, S. Miura, Y. Yoshimi, S. Yasuda, and Y. Makihara. 1986. "Cyclic Undrained triaxial strength of sand by a cooperative test program." *Soils Found.* 26 (3): 117–128. https://doi.org/https://doi.org/10.3208/sandf1972.26.3_117.

Vanmarcke, E. H. 1977. "Probabilistic modeling of soil profiles." *J. Geotech. Eng. Div.* 103 (11): 1227–1246. https://doi.org/10.1061/AJGEB6.0000517.

Wang, Y., Z. Cao, and S.-K. Au. 2010. "Efficient Monte Carlo simulation of parameter sensitivity in probabilistic slope stability analysis." *Comput. Geotech.* 37 (7): 1015–1022. https://doi.org/10.1016/j.compgeo.2010.08.010.

Zhang, J., J. P. Li, L. M. Zhang, and H. W. Huang. 2014. "Calibrating cross-site variability for reliability-based design of pile foundations." *Comput. Geotech.* 62 (Oct): 154–163. https://doi.org/10.1016/j.compgeo.2014.07.013.

Zhang, J., W. H. Tang, L. M. Zhang, and H. W. Huang. 2012. "Characterising geotechnical model uncertainty by hybrid Markov Chain Monte Carlo simulation." *Comput. Geotech.* 43 (Jun): 26–36. https://doi.org/10.1016/j.compgeo.2012.02.002.

Zhang, J., L. M. Zhang, and W. H. Tang. 2009. "Bayesian framework for characterizing geotechnical model uncertainty." *J. Geotech. Geoenviron. Eng.* 135 (7): 932–940. https://doi.org/10.1061/(ASCE)GT.1943-5606.0000018.

Zhang, L. L., J. Zhang, L. M. Zhang, and W. H. Tang. 2010. "Back analysis of slope failure with Markov chain Monte Carlo simulation." *Comput. Geotech.* 37 (7): 905–912. https://doi.org/10.1016/j.compgeo.2010.07.009.

© ASCE 04021188-14 J. Geotech. Geoenviron. Eng.

J. Geotech. Geoenviron. Eng., 2022, 148(2): 04021188