NRC Research Press

ARTICLE

Statistical characterization of random field parameters using frequentist and Bayesian approaches

Jianye Ching, Shih-Shuan Wu, and Kok-Kwang Phoon

Abstract: Because information collected in a site investigation is limited, it is not possible to obtain actual values for the mean, standard deviation, and scale of fluctuation for a soil property of interest. The deviation between the estimated values and the actual values is called the statistical uncertainty. There are at least two schools of thought on how to model the statistical uncertainty: frequentist thought and Bayesian thought. The purpose of this paper is to discuss their philosophical difference, to show how to quantify the statistical uncertainty based on these two distinct schools of thought, and to compare their performances. To quantify the statistical uncertainty, the confidence interval will be used for the frequentist school of thought, whereas the posterior probability distribution will be used for the Bayesian school of thought. Examples will be presented to compare the performances of these two schools of thought in terms of their consistencies. The results show that, in general, the Bayesian thought performs better in terms of consistency. In particular, the Markov chain Monte Carlo method is recommended when the amount of information available is very limited.

Key words: statistical uncertainty, site investigation, frequentist, Bayesian, reliability.

Résumé : En raison du manque de données provenant d'études sur le terrain, il est impossible d'obtenir les valeurs réelles de la moyenne, de l'écart-type et de l'échelle de variation d'un paramètre du sol que l'on souhaite étudier. L'écart entre les valeurs estimées et les valeurs réelles est appelé « incertitude statistique ». Il existe au moins deux écoles de pensée en ce qui concerne la manière de modéliser l'incertitude statistique : l'école fréquentiste et l'école bayésienne. Dans le présent article, on examine la différence de philosophie entre ces deux écoles, on décrit la méthode de calcul de l'incertitude statistique utilisée par chacune des deux écoles et on compare les performances de ces méthodes. Dans le cas de l'école fréquentiste, l'intervalle de confiance servira à calculer l'incertitude statistique, alors que, dans le cas de l'école bayésienne, on utilisera la distribution des probabilités postérieures pour effectuer ce même calcul. On fournit des exemples permettant de comparer les performances des méthodes de calcul utilisées par les deux écoles en ce qui a trait à leurs consistances respectives. Les résultats montrent en général que les méthodes de l'école bayésienne sont plus performantes en termes de consistance. En particulier, on recommande d'utiliser la méthode de Monte-Carlo par chaînes de Markov lorsque la quantité de données disponibles est très limitée. [Traduit par la Rédaction]

Mots-clés : incertitude statistique, étude sur le terrain, fréquentiste, école bayésienne, fiabilité.

Introduction

One of the purposes of site investigation is to obtain information on the spatial distribution of geotechnical design parameters. Typically, the spatial distribution is expressed as a trend function and a zero-mean oscillating component about the trend. This oscillating component is called the spatial variability. The spatial variability has profound impact on the behavior of a geotechnical system, as illustrated in the literature. To name a few, the impact of spatial variability on the behavior of a footing has been studied by Fenton and Griffiths (2002, 2003), Srivastava and Sivakumar Babu (2009), and Ahmed and Soubra (2012, 2014). The impact on piles has been studied by Haldar and Sivakumar Babu (2008), Fan and Liang (2013), and Wang and Cao (2013*a*); on slope stability by Griffiths and Fenton (2004), Cho (2007), Low et al. (2007), Hicks and Spencer (2010), Griffiths et al. (2011), Ching and Liao (2013), and Li et al. (2014); and on retaining structure and excavation by Fenton et al. (2005), Griffiths et al. (2008), Luo et al. (2011), Wu et al. (2012), and Hu and Ching (2015). Owing to the significance of spatial variability, it is important to gather information on its statistical properties through site investigation. The spatial variability is typically characterized by its standard deviation (σ) and scale of fluctuation (δ). The standard deviation quantifies the magnitude of the oscillation around the trend. The scale of fluctuation (Vanmarcke 1977) is a measure of the distance within which soil properties are significantly correlated.

There are several techniques that can be employed to estimate σ and δ (in particular δ), such as the method of moments (Uzielli et al. 2005; Dasaka and Zhang 2012; Firouzianbandpey et al. 2014; Lloret-Cabot et al. 2014), the fluctuation function method (Wickremesinghe and Campanella 1993; Cafaro and Cherubini 2002), the maximum likelihood (ML) method (DeGroot and Baecher 1993), and the Bayesian method (Wang et al. 2010). Phoon and Fenton (2004) proposed a practical bootstrap technique to obtain a more robust estimate of the autocorrelation function and to gain an appreciation of the underlying variability in the estimate with minimal assumptions. More recently, Lloret-Cabot et al. (2014) developed an estimation method based on conditional simulations. They found that this method gains better accuracy in estimating the horizontal δ. A common issue of these techniques is that the resulting σ and δ estimates are not identical to their actual values. The deviation between the estimated and actual

Received 2 March 2015. Accepted 24 July 2015.

J. Ching and S.-S. Wu. Department of Civil Engineering, National Taiwan University, Taipei, Taiwan.
K.-K. Phoon. Department of Civil and Environmental Engineering, National University of Singapore, Singapore.
Corresponding author: Jianye Ching (e-mail: jyching@gmail.com).

Published at www.nrcresearchpress.com/cgj on 28 July 2015.

values is called the statistical uncertainty. This statistical uncertainty is different from other geotechnical uncertainties, such as the measurement error, model error, and transformation error. A simple framework to understand this distinction is to imagine spatial variability, measurement error, transformation error, model error, and others as components of a probability model. The complexity of this probability model depends on the needs of the problem. For example, if one is interested in estimating a design parameter from a field parameter, transformation error should be added to this probability model. If one is interested in estimating the bearing capacity of a footing from field data, model error should be added to this probability model. All probability models contain parameters such as mean (μ), coefficient of variation, autocorrelation, cross-correlation, and many others. It is not possible to estimate these probabilistic parameters exactly in the presence of limited data. Statistical uncertainty can be viewed as the degree of imprecision in this estimation process. It applies to all probabilistic parameters. In this paper, we only focus on three probabilistic parameters associated with the random field model, (μ, σ, δ).

Statistical uncertainty is present because the amount of information collected in a site investigation is always limited (Phoon and Kulhawy 1999). The only way to eliminate the statistical uncertainty is to investigate the entire population. For site investigation, this means that the entire subsurface domain of interest must be explored, which is obviously not practical. Because of statistical uncertainty, the σ and δ estimates based on different borehole or sounding data within the same soil layer will be different, although the actual σ and δ values are fixed numbers. The upshot is that the actual values of σ and δ cannot be known precisely in real-world applications. In the literature, case histories illustrating the statistical uncertainty can be found. One example is the clay site in Taranto, Italy (Cafaro and Cherubini 2002). The site consists of two fairly uniform clay layers (upper and lower layers). However, the δ estimates for the lower layer obtained from five cone penetration test (CPT) soundings range from 0.185 to 0.536 m. The actual δ value cannot be known precisely.

Although statistical uncertainty is well known, there are only a limited number of previous studies focusing on this subject. Honjo and Setiawan (2007) studied the statistical uncertainty for the average of a property field along a certain depth. They were not concerned with the statistical uncertainties in σ and δ: σ and δ are prescribed in their analysis. Instead, they were concerned with the unknown constant trend (or the mean value μ). The rationale for prescribing σ and δ is based on the argument that the number of data points in site investigation is typically insufficient to estimate these second-order statistics. To circumvent this practical difficulty, Honjo and Setiawan (2007) suggested that conservative values for σ and δ can be assumed based on previous studies and experiences. In other previous studies, the statistical uncertainty may have been treated implicitly by lumping with other sources of uncertainties. For instance, Wang et al. (2010) proposed a Bayesian method of characterizing the uncertainties in (μ, σ, δ). The resulting posterior probability distribution actually incorporates the statistical uncertainties in (μ, σ, δ). To our best knowledge, no study in the literature studied the statistical uncertainties in (μ, σ, δ) in an explicit way.

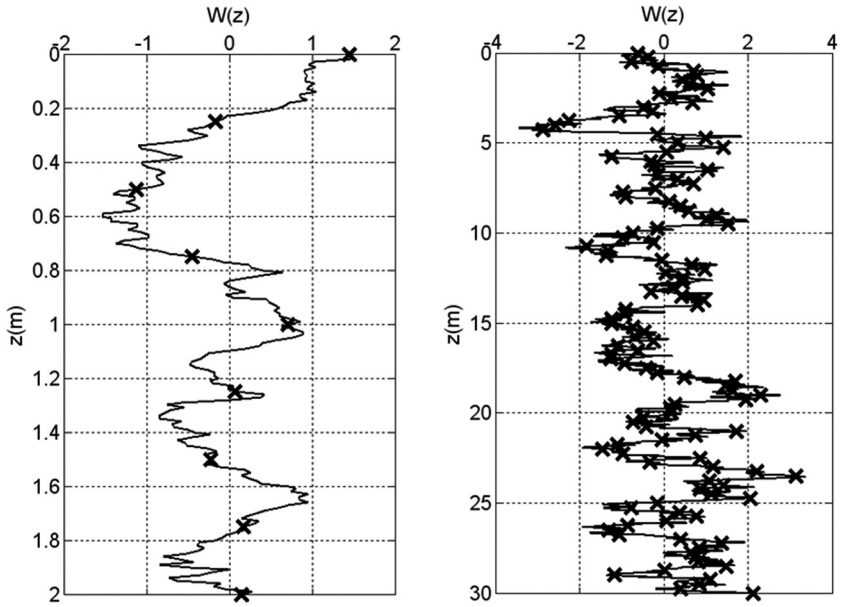
Although it is possible to assume conservative values for (σ, δ) based on previous studies and experiences, as suggested by Honjo and Setiawan (2007), there are practical difficulties for doing so. First of all, (σ, δ) values in the literature vary in a wide range. For instance, the coefficient of variation (COV = σ/μ) for the undrained shear strength of a clay varies from 0.1 to 0.6 (Phoon and Kulhawy 1999). Its vertical δ varies from 0.8 to 6.2 m. Its horizontal δ is known in a very limited way (only three studies collected by Phoon and Kulhawy 1999, ranging from 46 to 60 m; another three studies collected by El-Ramly et al. 2003, ranging from 22 to 40 m). It is not simple to select suitable conservative values for (σ, δ) based on the aforementioned wide ranges. Moreover, δ may depend on the problem scale (Fenton 1999), and (σ, δ) may depend on the adopted trend function and the sampling interval as well (Cafaro and Cherubini 2002). The scale considered in previous studies may not be similar to the scale applicable for the geotechnical project at hand. The trend function and sampling interval studied in the literature may not be applicable to the conditions in the project at hand. Jaksa et al. (2005) suggested using a "worst case" scale of fluctuation, which for the example of a three-storey, nine-pad footing building examined, is equal to the spacing between footings. This worst case strategy circumvents the need to estimate the scale of fluctuation from site investigation data, but it is problem dependent and it may be too conservative. Finally, the assumed (σ, δ) values may not be consistent with the site investigation data. This inconsistency will be clarified in this paper. There are strong reasons, either practical or analytical, to estimate (σ, δ) based on the local site investigation data. If local CPT data are available, it is possible to estimate (μ, σ, δ) simultaneously. It is also useful to note that many design soil parameters (such as undrained shear strength) can be expressed as a function of the CPT data (Mayne 1986; Robertson et al. 1986; Senneset et al. 1989; Chen and Mayne 1994, 1996; Ching and Phoon 2012; Ching et al. 2014).

The purpose of this paper is to quantify the statistical uncertainties for the three parameters (μ, σ, δ) that are required for the second-moment characterization of a random field. As discussed earlier, statistical uncertainties apply to all probabilistic parameters estimated from limited data. This study only focuses on estimation of random field parameters. This estimation problem is applicable to a CPT sounding, which is presented at the end of this paper. The measurement error for CPT is relatively small (Phoon and Kulhawy 1999) and can be neglected to avoid complicating our discussions. We only consider one field measurement, which is the cone tip resistance. Hence, there is no cross-correlation between different field measurements. This paper does not attempt to convert CPT data to a design parameter such as the undrained shear strength. Hence, there is no transformation error. This paper does not study the response of a geotechnical structure to a spatially variable soil mass. Hence, there is no model error. To quantify the statistical uncertainties for (μ, σ, δ), we will first demonstrate the statistical uncertainties in (μ, σ, δ) using simulated spatial variability examples. Second, two schools of thought, frequentist and Bayesian, will be adopted to interpret and quantify the statistical uncertainties. Analytical tools for quantifying the statistical uncertainties will be also introduced. For the frequentist school of thought, the tool is the confidence interval. For Bayesian school of thought, the tool is the posterior probability distribution. Third, these two schools of thought will be compared on the basis of consistency, namely, which school of thought can provide a confidence interval or posterior probability distribution that is more consistent with the actual values of (μ, σ, δ). Finally, the issue of estimating statistical uncertainties in the presence of very limited information (thin soil layers) will be addressed.

Simulated example

Vanmarcke (1977) proposed that the spatial variability can be modeled as a random field. Among random field models, stationary random fields are widely used due to their simplicity and possibly the only practical version that can be characterized from limited data (Phoon et al. 2003). Three parameters are required to characterize a second-order stationary random field model: (*i*) mean (μ); (*ii*) standard deviation (σ); and (*iii*) autocorrelation function. For a one-dimensional random field $W(z)$, where W is the

Fig. 1. Simulated **W** data points with (μ, σ, δ) = (0, 1, 1): (*a*) one realization with D = 2 m, dz = 0.25 m; (*b*) one realization with D = 30 m, dz = 0.25 m.



(a) D = 2 m & dz = 0.25 m (b) D = 30 m & dz = 0.25 m

property field and z is the location (e.g., depth), the autocorrelation function is defined as the correlation between two locations that are Δz apart:

$$(1) \qquad \rho(\Delta z) = \rho[W(z), W(z + \Delta z)] = \frac{\mathrm{CV}[W(z),\ W(z + \Delta z)]}{\sqrt{\mathrm{Var}[W(z)]}\sqrt{\mathrm{Var}[W(z + \Delta z)]}}$$

where Var( ) denotes variance; CV(,) denotes covariance. Variance of $W(z)$ is a measure of the spread of $W(z)$. Its square root is the standard deviation of $W(z)$. Covariance between $W(z)$ and $W(z + \Delta z)$ is a measure of how much $W(z)$ and $W(z + \Delta z)$ change together. The autocorrelation $\rho(\Delta z)$ can be viewed as covariance normalized with respect to the standard deviation of $W(z)$ and $W(z + \Delta z)$. The hypothesis of second-order stationarity allows ρ to be function of Δz only, rather than the absolute coordinates z and $z + \Delta z$ (see left-hand side of eq. (1)). Only second-order stationary fields are discussed in this paper. Hence, the qualifier "second-order" will be dropped hereafter. The most popular autocorrelation model is the single exponential model (Vanmarcke 1977, 1983):

$$(2) \qquad \rho(\Delta z) = \exp(-2|\Delta z|/\delta)$$

where δ is the scale of fluctuation. It is clear that the correlation decreases as Δz increases. This is commonly observed in natural soils: soil properties are strongly spatially correlated within a small interval and are weakly spatially correlated when located far apart.

A one-dimensional stationary normal random field with mean μ, standard deviation σ, and scale of fluctuation δ can be simulated by the following equation:

$$(3) \qquad \mathbf{W} = \mu\mathbf{1} + \sigma\mathbf{L}\mathbf{U}$$

where $\mathbf{W} = [W(z_1)\ \ W(z_2)\ \ \ldots\ \ W(z_n)\ ]^{\mathrm{T}}$ contains the point processes of $W(z)$ at the n locations of interest; $\mathbf{U} = [U_1\ \ U_2\ \ \ldots\ \ U_n\ ]^{\mathrm{T}}$ contains independent standard normal samples; $\mathbf{1} = [1\ \ 1\ \ \ldots\ \ 1\ ]^{\mathrm{T}}$; **L** is the Cholesky decomposition of the correlation matrix **R** ($\mathbf{R} = \mathbf{L} \times \mathbf{L}^{\mathbf{T}}$):

$$(4) \qquad \mathbf{R} = \begin{bmatrix} 1 & \exp(-2|z_1 - z_2|/\delta) & \cdots & \exp(-2|z_1 - z_n|/\delta) \\ & \ddots & & \vdots \\ & & 1 & \exp(-2|z_{n-1} - z_n|/\delta) \\ \mathrm{SYM} & & & 1 \end{bmatrix}$$

In this study, $\mathbf{W} = [W(z_1)\ \ W(z_2)\ \ \ldots\ \ W(z_n)\ ]^{\mathrm{T}}$ represents the vertical profile of a CPT sounding in a soil layer; d$z = z_{i+1} - z_i$ is the vertical sampling interval; and $D = z_n - z_1$ is the total depth of the CPT sounding. It is assumed that the CPT data have been transformed (e.g., detrended, or converted to the normalized cone tip resistance $Q = (q_t/P_a)/(\sigma'_v/P_a)^n$, where q_t is the cone tip resistance, P_a is the atmospheric pressure, and σ'_v is the overburden stress, and possibly taken logarithm log(Q), etc.) so that **W** is a realization of a stationary normal random field. Without loss of generality, the actual values of (μ, σ, δ) are taken to be (0, 1, 1). Two normalized parameters are used to characterize the total depth D and sampling interval dz: $n_D = D/\delta$ and Δ = dz/δ. Phoon et al. (2003) termed n_D the "normalized sampling length" and Δ is the reciprocal of "number of points in one scale of fluctuation". It is clear that the total number of data points $n = n_D/\Delta + 1$. Basically, n_D is the equivalent number of scales of fluctuation within the total depth D, whereas Δ is the normalized sampling interval (normalized with respect to δ). The following analysis results only depend on n_D and Δ.

Consider that D = 2 m and dz = 0.25 m (n_D = 2 and Δ = 0.25). The CPT readings are typically taken every 10–50 mm, i.e., dz = 0.01~0.05 m. A larger dz value is selected to illustrate the scenario of very limited data points, arguably less than what is available for the same depth of CPT sounding in practice. Later, we will see that further reducing dz to a smaller value (e.g., dz = 0.02 m) will not alter the main conclusions. Figure 1*a* shows the simulated **W** data points. Figure 1*a* may represent site investigation data in a thin soil layer. The crosses are the data points, whereas the curve represents the underlying continuous random field that generates the data points. The continuous curve is regarded as unknown in the following analysis. It is assumed that the measurement error is zero because the current paper focuses on the statistical uncer-

tainty only. It is also realistic for CPT data because its measurement error is low (Phoon and Kulhawy 1999). Figure 1*b* shows the simulated **W** data points for another case with $D = 30$ and $dz = 0.25$ ($n_D = 30$ and $\Delta = 0.25$). Figure 1*b* may represent site investigation data in a thick soil layer. The two sets of **W** data have very different numbers of data points ($n = 9$ and 121, respectively). It is desirable to accurately estimate (μ, σ, δ) using the **W** data. As mentioned earlier, estimation errors are inevitable due to limited amount of information, and the deviations between the estimated (μ, σ, δ) values and (0, 1, 1) (these correct answers are available in simulated data because they are inputs for the random field) are the statistical uncertainties.

Frequentist and Bayesian schools of thought

Frequentist school of thought

From the frequentist point of view, the **W** data are a *random realization* from the population. Central to the frequentist school of thought are the estimators of (μ, σ, δ), which are functions of **W**. For the reason that will be apparent later, it is easier to estimate [ln(σ), ln(δ)] rather than (σ, δ). There are various estimators. Among them, the maximum likelihood (ML) estimator has several desirable properties, including asymptotically unbiased, asymptotically normally distributed, asymptotically minimum variance, invariant, sufficient, and consistent (DeGroot and Baecher 1993). The ML estimators of (μ, ln(σ), ln(δ)) maximize the following likelihood function:

$$f[\mathbf{W}|\mu,\ \ln(\sigma),\ \ln(\delta)] = \frac{1}{\sqrt{2\pi}^n}\frac{1}{\sqrt{|\sigma^2\mathbf{R}|}} \times \exp\left[-\frac{1}{2\sigma^2}(\mathbf{W} - \mu\mathbf{1})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{W} - \mu\mathbf{1})\right] \tag{5}$$

where $|\sigma^2\mathbf{R}|$ is the determinant of the square matrix $\sigma^2\mathbf{R}$. The likelihood function follows the normal distribution because the simulated random field is a normal random field (see eq. (3)). The ML estimators of (μ, ln(σ), ln(δ)) are denoted by (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})). The determinant $|\sigma^2\mathbf{R}|$ can easily approach zero when the number of data points (n) is large. When this happens, the evaluation of $|\sigma^2\mathbf{R}|$ may suffer from round-off error. The solution is simple: first determine the eigenvalues (λ_1, λ_2, …, λ_n) for $\sigma^2\mathbf{R}$, then ln($|\sigma^2\mathbf{R}|$) can be determined as

$$\ln(|\sigma^2\mathbf{R}|) = \sum_{i=1}^{n}\ln(\lambda_i) \tag{6}$$

Frequentist asymptotic approximation method

Before the **W** data are obtained, (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})) must be regarded as random variables because they depend on **W**, and **W** is random. The multivariate probability density function (PDF) of (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})) is called the sampling distribution. As shown by Mardia and Marshall (1984), (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})) are asymptotically (multivariate) normally distributed with mean equal to the actual values of (μ, ln(σ), ln(δ)) and the covariance matrix $\mathbf{C}(\mu, \ln(\sigma), \ln(\delta))$ equal to the inverse of the Fisher information matrix (Mardia and Marshall 1984). The phrase "asymptotically" means that (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})) will be indeed normally distributed with the aforementioned mean and covariance if **W** contains a large amount of information, e.g., n_D is very large. The reason for adopting (ln(σ), ln(δ)) is now evident. It is more reasonable to assert that (ln(σ_{ML}), ln(δ_{ML})) are asymptotically normal because both negative and positive values can occur. In practice, the actual values of (μ, ln(σ), ln(δ)) are unknown, and the sampling distribution is approximated as a multivariate normal distribution with covariance matrix $\mathbf{C}(\mu_{ML}, \ln(\sigma_{ML}), \ln(\delta_{ML}))$.

For the simulated **W** data in Fig. 1*a*, the ML estimates are (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})) = (0.0852, −0.384, −1.588). Note that (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})) are not equal to their actual values (0, ln(1), ln(1)). The deviations are the statistical uncertainties. The statistical uncertainties are large because there are only nine data points. Moreover, these are spatially correlated data points. There are in fact only roughly $n_D = 2$ independent data points. As mentioned earlier, the confidence interval (region) plays a central role in the frequentist school of thought. The 95% confidence region for the actual values of (μ, ln(σ), ln(δ)) can be constructed based on the sampling distribution of (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})). This 95% confidence region is a three-dimensional (3D) ellipse defined as

$$\begin{bmatrix}\mu - \mu_{ML}\\ \ln(\sigma) - \ln(\sigma_{ML})\\ \ln(\delta) - \ln(\delta_{ML})\end{bmatrix}^{\mathrm{T}}\mathbf{C}[\mu_{ML},\ \ln(\sigma_{ML}),\ \ln(\delta_{ML})]^{-1} \times \begin{bmatrix}\mu - \mu_{ML}\\ \ln(\sigma) - \ln(\sigma_{ML})\\ \ln(\delta) - \ln(\delta_{ML})\end{bmatrix} \le 7.815 \tag{7}$$

where 7.815 is the 0.95 quantile of the chi-squared random variable with 3 degrees of freedom. Figure 2 shows the confidence interval for μ and confidence region for (ln(σ), ln(δ)) for the **W** data in Fig. 1*a*. The actual values of (μ, ln(σ), ln(δ)) = (0, ln(1), ln(1)) = (0, 0, 0) are shown as circles. They are within the confidence interval and confidence region. The confidence interval for μ is reasonably narrow: from about −0.4 to 0.55. However, the confidence region for (ln(σ), ln(δ)) is fairly wide. For ln(σ), it spans from about −1 to 0.2, meaning that σ spans from 0.37 to 1.22. For ln(δ), it spans from about −5.6 to 2.4, meaning that δ spans from 0.04 to 11. This confirms the observation made by Honjo and Setiawan (2007) that data in site investigation are typically insufficient to estimate second-order statistics such as σ and δ.

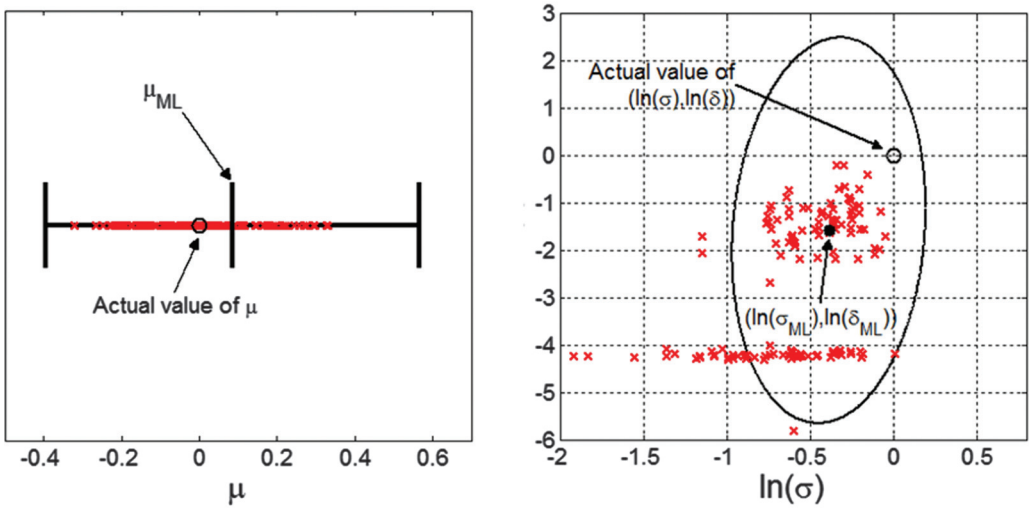
For the **W** data in Fig. 1*b*, the ML estimates are (0.129, 0.0583, 0.101) — much closer to (0, 0, 0) than the ML estimates for the **W** data in Fig. 1*a* because there are now roughly $n_D = 30$ independent data points. The new confidence interval and region are shown in Fig. 3. The confidence interval for μ in Fig. 3*a* is slightly narrower than that in Fig. 2*a*. It is interesting to see that the confidence region for (σ, δ) in Fig. 3*b* is much more clustered than that in Fig. 2*b*. This is reasonable because the result in Fig. 3*b* is based on more independent data points ($n_D = 30$). Moreover, it appears that the increase in n_D does not greatly improve the resolution for μ, although it greatly improves the resolution for (σ, δ).

Limited number of data points (frequentist bootstrap)

The multivariate normal sampling distribution and the resulting confidence interval discussed earlier are valid for an asymptotic case, which means n_D is large. However, for site investigation, it is common to have small n_D. With limited amount of data, the earlier asymptotic approximation is not accurate. The bootstrap method (Efron and Tibshirani 1993) is a general framework for obtaining (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})) samples from the sampling distribution, without adopting the asymptote assumption. There are various bootstrap methods for a stationary random field. In this study, the block bootstrap method (Künsch 1989; Politis and Romano 1992) is adopted. The steps for the block bootstrap are as follows:

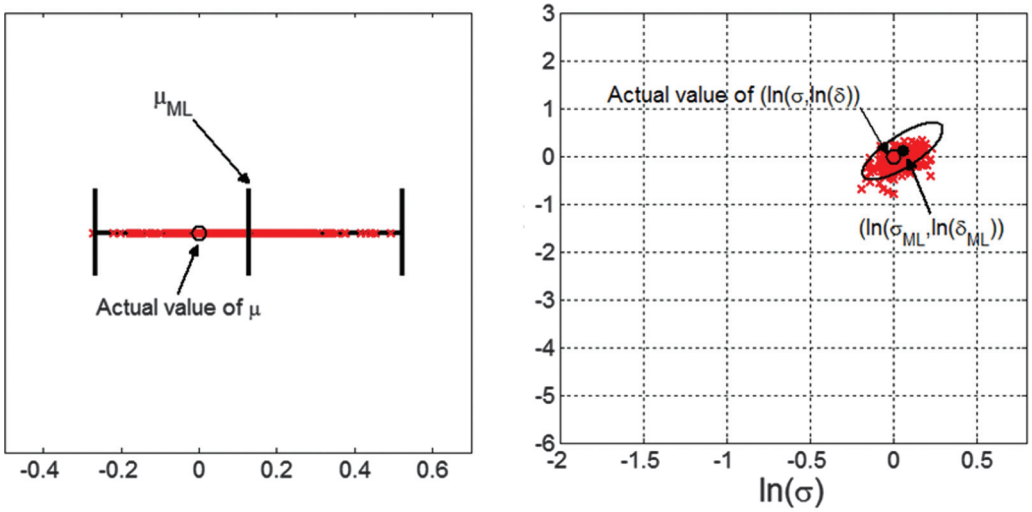
1. Randomly select blocks of length L from the original data [$W(z_1)$, $W(z_2)$, … $W(z_n)$]. For instance, consider a case with $n = 9$. Let the block length L be equal to 3. Three examples of such random selections can be $B_1 = [W(z_3), W(z_4), W(z_5)]$, $B_2 = [W(z_7), W(z_8), W(z_9)]$, and $B_3 = [W(z_5), W(z_6), W(z_7)]$. Note that the blocks are allowed to overlap.

Fig. 2. (*a*) Confidence interval for μ and (*b*) confidence region for (ln(σ), ln(δ)) for **W** data in Fig. 1*a*. Crosses are samples from block bootstrap method.



(a) 95% confidence interval for μ (b) 95% confidence region for (ln(σ),ln(δ))

Fig. 3. (*a*) Confidence interval for μ and (*b*) confidence region for (ln(σ), ln(δ)) for **W** data in Fig. 1*b*. Crosses are samples from block bootstrap method.



(a) 95% confidence interval for μ (b) 95% confidence region for (ln(σ),ln(δ))

2. Concatenate the randomly selected blocks together to form a re-sampled **W** of length n. Consider the example in step 1, the concatenated (re-sampled) **W** is simply $[B_1 \ B_2 \ B_3] = [W(z_3), W(z_4), W(z_5), W(z_7), W(z_8), W(z_9), W(z_5), W(z_6), W(z_7)]$.
3. Evaluate (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})) using the re-sampled **W** data points.
4. Repeat steps 1 and 2 to obtain n_B samples of (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})). Note that n_B is distinctive from n.

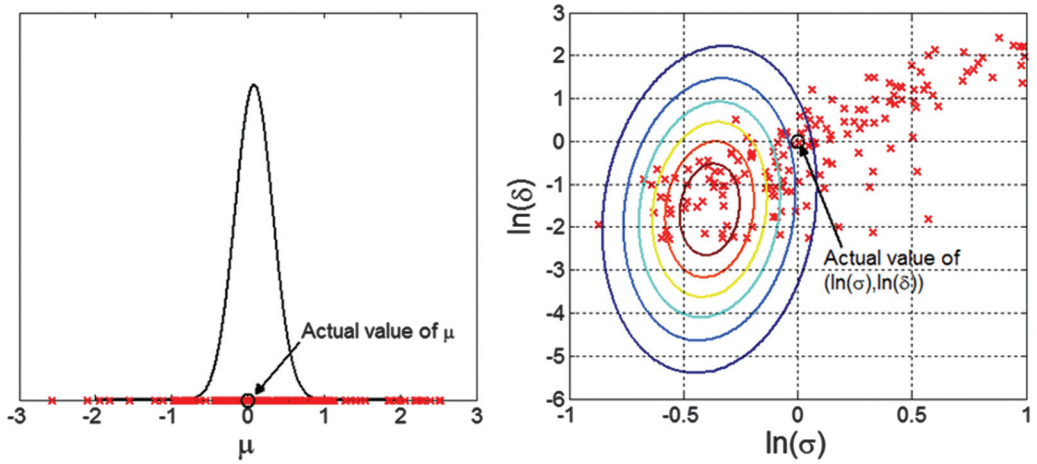
These n_B samples are therefore approximately from the sampling distribution of (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})). The block size L needs to be chosen with caution. Politis and White (2004) recommended L to be proportional to $n^{1/3}$. In this study, it is found that the block bootstrap method performs reasonably well if L is taken to be the nearest integer of $2n^{1/3}$.

The **W** data in Fig. 1 are re-analyzed using the bootstrap method (n_B = 200). The resulting (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})) samples are plotted in Figs. 2 and 3 as crosses. The bootstrap (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})) samples are broadly consistent with the asymptotic 95% confidence interval (region) for the case in Fig. 1*b* (see Fig. 3). For the case in Fig. 1*a*, the (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})) samples are less consistent with the asymptotic 95% confidence interval (region) (Fig. 2).

Bayesian school of thought

From the Bayesian point of view, the parameters (μ, ln(σ), ln(δ)) are *uncertain*. A *prior* PDF, denoted by $f(\mu, \ln(\sigma), \ln(\delta))$, is used to describe the a priori knowledge for (μ, ln(σ), ln(δ)). It specifies the *plausibility* of various combinations of (μ, ln(σ), ln(δ)) before the **W** data are obtained. For instance, a uniform prior PDF can be taken for (μ, ln(σ), ln(δ)) over the rectangular region formed by $\mu \in [-1, 1]$, ln(σ) $\in [-2, 2]$, and ln(δ) $\in [-3, 3]$. The likelihood function in eq. (5) describes the *plausibility* of **W**, given the full knowledge of (μ, ln(σ), ln(δ)). The *posterior* (updated) PDF can then be determined using the Bayes rule:

Fig. 4. Posterior PDF $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$ for **W** data in Fig. 1*a*: (*a*) $f(\mu|\mathbf{W})$; (*b*) contours for $f(\ln(\sigma), \ln(\delta)|\mathbf{W})$. Crosses are Markov chain Monte Carlo (MCMC) samples from $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$.



(a) f(μ|W) (b) f(ln(σ),ln(δ)|W)

$$f[\mu,\ \ln(\sigma),\ \ln(\delta)|\mathbf{W}] = \frac{f[\mathbf{W}|\mu,\ \ln(\sigma),\ \ln(\delta)]f[\mu,\ \ln(\sigma),\ \ln(\delta)]}{f(\mathbf{W})} \quad (8)$$

where $f(\mathbf{W})$ is a normalizing constant that does not depend on $(\mu, \ln(\sigma), \ln(\delta))$. The posterior PDF $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$ describes the plausibility of various combinations of $(\mu, \ln(\sigma), \ln(\delta))$ *after* the **W** data are incorporated. The $(\mu, \ln(\sigma), \ln(\delta))$ value that maximizes the posterior PDF $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$ is called the maximum a posteriori (MAP) estimate, denoted by $(\mu_{MAP}, \ln(\sigma_{MAP}), \ln(\delta_{MAP}))$.

In the frequentist point of view, it does not make sense to speak about the PDF of $(\mu, \ln(\sigma), \ln(\delta))$ because $(\mu, \ln(\sigma), \ln(\delta))$ are not random. In this point of view, one can only speak about the confidence interval (region) for $(\mu, \ln(\sigma), \ln(\delta))$ — an interval or region that contains the actual $(\mu, \ln(\sigma), \ln(\delta))$ with high confidence. In contrast, from the Bayesian point of view, it makes sense to speak about the PDF of $(\mu, \ln(\sigma), \ln(\delta))$ because $(\mu, \ln(\sigma), \ln(\delta))$ are uncertain. The posterior PDF $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$ is central to the Bayesian thought.

Bayesian asymptotic approximation method (Laplace approximation)

There is also an asymptotic normal approximation for $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$, called the Laplace approximation (Bleistein and Handelsman 1986). The normal PDF that approximates $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$ has mean equal to $(\mu_{MAP}, \ln(\sigma_{MAP}), \ln(\delta_{MAP}))$ and the covariance matrix equal to $\mathbf{C}(\mu_{MAP}, \ln(\sigma_{MAP}), \ln(\delta_{MAP})|\mathbf{W})$, where $\mathbf{C}(\mu_{MAP}, \ln(\sigma_{MAP}), \ln(\delta_{MAP})|\mathbf{W})$ is equal to the inverse of the Hessian matrix for $-\ln[f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})]$. The phrase "asymptotic" means that $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$ will be indeed normally distributed with the aforementioned mean and covariance if **W** contains a large amount of information, e.g., n_D is very large. In the Bayesian framework, there is a term called the "Bayesian confidence region", also called the "credible region". It is analogous to the frequentist 95% confidence region (Jaynes 1976; Lee 2012). For the asymptotic normal posterior PDF, the Bayesian 95% confidence region is a 3D ellipse described by an inequality similar to eq. (7), but the center of the ellipse is replaced by $(\mu_{MAP}, \ln(\sigma_{MAP}), \ln(\delta_{MAP}))$, and the covariance matrix replaced by $\mathbf{C}(\mu_{MAP}, \ln(\sigma_{MAP}), \ln(\delta_{MAP})|\mathbf{W})$.

Consider the simulated **W** data in Fig. 1*a* and consider a completely flat prior PDF $f(\mu, \ln(\sigma), \ln(\delta))$ equals constant. The MAP estimates are $(\mu_{MAP}, \ln(\sigma_{MAP}), \ln(\delta_{MAP})) = (0.0852, -0.384, -1.588)$. Figure 4 shows the contours of the asymptotic normal posterior PDF $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$. The posterior PDF for μ is narrow. However, the posterior PDF for $(\ln(\sigma), \ln(\delta))$ is fairly wide. This again confirms the observation made by Honjo and Setiawan (2007) that data in site investigation are typically insufficient to estimate second-order statistics such as σ and δ. For the **W** data in Fig. 1*b*, the MAP estimates are (0.129, 0.0583, 0.101). The (approximate) normal posterior PDFs are shown in Fig. 5. The posterior PDF for μ in Fig. 5*a* is slightly narrower than that in Fig. 4*a*. It is interesting to see that the posterior PDF for (σ, δ) in Fig. 5*b* is more compact than that in Fig. 4*b*. This is reasonable because the result in Fig. 5*b* is based on more independent data points ($n_D = 30$). Moreover, while the increase in n_D does not greatly improve the resolution for μ, it greatly improves the resolution for (σ, δ).

Limited number of data points

The Laplace approximation discussed in the previous subsection is valid only for asymptotic cases with large n_D. The Markov chain Monte Carlo (MCMC) method (Gilks et al. 1995; Wang and Cao 2013*b*; Cao and Wang 2014) is a general framework of obtaining samples from $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$ regardless of the amount of information. The steps for a simplified MCMC algorithm (Metropolis et al. 1953) are as follows:

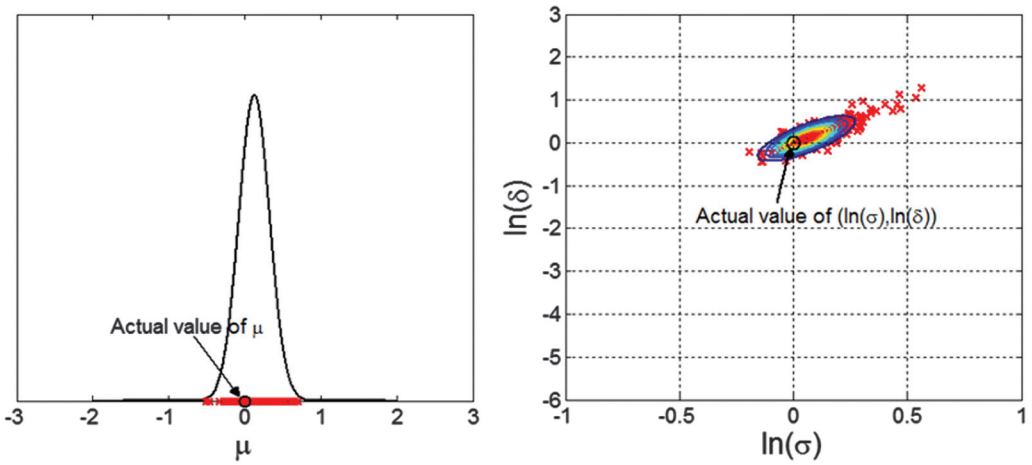
1. The initial time step $t = 0$, and the initial samples are $(\mu^0, \ln(\sigma^0), \ln(\delta^0))$. In this study, the MAP estimates $(\mu_{MAP}, \ln(\sigma_{MAP}), \ln(\delta_{MAP}))$ are taken to be the initial samples.
2. Draw the *candidate* sample $(\mu^c, \ln(\sigma^c), \ln(\delta^c))$ from a joint normal proposal PDF centering at $(\mu^0, \ln(\sigma^0), \ln(\delta^0))$ with a certain covariance matrix. In this study, $\mathbf{C}(\mu_{MAP}, \ln(\sigma_{MAP}), \ln(\delta_{MAP})|\mathbf{W})$ is found to be a satisfactory choice for this covariance matrix.
3. Compute the r ratio:

$$r = \frac{f[\mathbf{W}|\mu^c,\ \ln(\sigma^c),\ \ln(\delta^c)]f[\mu^c,\ \ln(\sigma^c),\ \ln(\delta^c)]}{f[\mathbf{W}|\mu^0,\ \ln(\sigma^0),\ \ln(\delta^0)]f[\mu^0,\ \ln(\sigma^0),\ \ln(\delta^0)]} \quad (9)$$

4. Accept $(\mu^1, \ln(\sigma^1), \ln(\delta^1)) = (\mu^c, \ln(\sigma^c), \ln(\delta^c))$ with probability $\min(1, r)$. If $(\mu^c, \ln(\sigma^c), \ln(\delta^c))$ is not accepted, repeat the previous sample: $(\mu^1, \ln(\sigma^1), \ln(\delta^1)) = (\mu^0, \ln(\sigma^0), \ln(\delta^0))$.
5. Incrementally cycle steps 2–4 for $t = 1, 2, \ldots, n_T$ times to obtain $(\mu^0, \ln(\sigma^0), \ln(\delta^0)) \ldots (\mu^{n_T}, \ln(\sigma^{n_T}), \ln(\delta^{n_T}))$ samples.

When t is sufficiently far away from the initial time $t = 0$, the aforementioned process will reach a *stationary* state distributed as $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$. The initial nonstationary period is called the

Fig. 5. Posterior PDF $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$ for **W** data in Fig. 1*b*: (*a*) $f(\mu|\mathbf{W})$; (*b*) contours for $f(\ln(\sigma), \ln(\delta)|\mathbf{W})$. Crosses are MCMC samples from $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$.



burn-in period. The samples within the burn-in period must be discarded.

The **W** data in Fig. 1 are re-analyzed using the MCMC method (n_T = 40 000). The burn-in period is very short. The resulting (μ, $\ln(\sigma)$, $\ln(\delta)$) samples, excluding those in the burn-in period, are plotted in Figs. 4 and 5 for the **W** data in Figs. 1*a* and 1*b*, respectively. Note that the MCMC samples of two time instants t_1 and t_2 are correlated if t_1 and t_2 are nearby. Therefore, only one (μ, $\ln(\sigma)$, $\ln(\delta)$) sample is taken for every 200 time steps. This minimizes the correlation between the MCMC samples. Also, this makes the number of the adopted MCMC samples equal to 40 000/200 = 200, consistent with the number of the bootstrap samples. For the **W** data in Fig. 1*a*, the MCMC samples are inconsistent with the asymptotic posterior PDF (Laplace approximation) (see Fig. 4). The (μ, $\ln(\sigma)$, $\ln(\delta)$) samples spread in a range that is well outside the main region of the asymptotic posterior PDF. This is probably because there are only two independent data points (n_D = 2), so the Laplace approximation is poor. For the **W** data in Fig. 1*b*, nevertheless, the MCMC samples are more consistent with the asymptotic posterior PDF (see Fig. 5).

Consistency issue for ML and MAP estimators

The estimators of (μ, $\ln(\sigma)$, $\ln(\delta)$), such as ML and MAP estimators, are not always consistent. In the following, we present two scenarios where the (μ, $\ln(\sigma)$, $\ln(\delta)$) estimators become inconsistent, meaning that (μ_{ML}, $\ln(\sigma_{ML})$, $\ln(\delta_{ML})$), or (μ_{MAP}, $\ln(\sigma_{MAP})$, $\ln(\delta_{MAP})$), will not be close to their actual values even if there are many data points:

1. When Δ (= dz/δ) is too large, the δ estimator becomes inconsistent. This scenario will be referred to as the "large-Δ" issue. For instance, for Δ = 2 and δ = 1 m, the sampling interval dz = 2δ = 2 m. The **W** data will contain nearly uncorrelated $W(z_1)$, $W(z_2)$, … $W(z_n)$ values because the distance between two adjacent **W** data are larger than δ. In this case, δ_{ML} = 1 m and δ_{ML} = 0 (δ = 0 means no correlation) are nearly equally plausible. However, the latter estimate (δ_{ML} = 0) is absurd. This is because the sampling interval dz is too large to detect the correlation. Increasing n_D (while keeping Δ constant) will not remove this issue.

   Random simulations are performed to understand how (μ_{ML}, $\ln(\sigma_{ML})$, $\ln(\delta_{ML})$) change with respect to Δ. Twenty random fields with (μ, σ, δ) = (0, 1, 1) and with total depth D = 30 m (so that $n_D = D/\delta$ = 30) are simulated. Then, these random fields are sampled at different interval dz (or Δ), where dz (or Δ) can be as small as 0.05 m and as large as 5 m. The resulting discrete data are then used to obtain (μ_{ML}, $\ln(\sigma_{ML})$, $\ln(\delta_{ML})$). Figure 6 shows how (μ_{ML}, $\ln(\sigma_{ML})$, $\ln(\delta_{ML})$) change with respect to Δ for the 20 realizations of random fields. It can be concluded that Δ must be less than 0.5 to avoid this issue. This criterion (Δ < 0.5) is general, applying to cases with various (μ, σ, δ). This criterion works well for both ML and MAP estimators. Note that this large-Δ issue does not make the (μ, $\ln(\sigma)$) estimators inconsistent (see Fig. 6). Increasing n_D (while keeping Δ constant) will let the (μ, $\ln(\sigma)$) estimators converge to their actual values.
2. When n_D (= D/δ) is too small, all (μ, $\ln(\sigma)$, $\ln(\delta)$) estimators become inconsistent. This scenario will be referred to as the "small-n_D" issue. For instance, for n_D = 0.5 and δ = 1 m, the sampling depth D = 0.5δ = 0.5 m. The **W** data only "see" one-half of the scale of fluctuation. In this case, it is impossible to accurately estimate δ because the sampling depth D is too small to see the correlation structure. The (σ, μ) estimators will be biased as well. Decreasing Δ (while keeping n_D constant) will not remove this issue.

   Random simulations are performed to understand how (μ_{ML}, $\ln(\sigma_{ML})$, $\ln(\delta_{ML})$) change with respect to n_D. First of all, a continuous random field with (μ, σ, δ) = (0, 1, 1) and with maximum total depth of 50 m and Δ = 0.25 (so that dz = 0.25 m) is simulated. Then, this random field is sampled using a constant dz = 0.25 m, but the resulting discrete data are truncated at various values of D (<50 m), where D (or n_D) can be as small as 0.5 m and as large as 50 m. In other words, when D < 50 m, only part of the random field with maximum total depth of 50 m is used. The discrete data are then used to obtain (μ_{ML}, $\ln(\sigma_{ML})$, $\ln(\delta_{ML})$). Figure 7 shows how (μ_{ML}, $\ln(\sigma_{ML})$, $\ln(\delta_{ML})$) changes with respect to n_D for 20 realizations of such random fields. It can be concluded that n_D must be no less than 10~20 to avoid the small-n_D issue. For a reason that will be clear later, the threshold n_D = 20 is taken. This criterion ($n_D \geq 20$) is general, applying to cases with various (μ, σ, δ). This criterion works well for both ML and MAP estimators.

In principle, both the large-Δ and small-n_D issues only affect the analysis results for the asymptotic approximations (frequentist and Bayesian). This is because the methods for limited amount of data (bootstrap and MCMC) do not rely on the ML or MAP estimators. However, we shall see in the section "Identifiability issue for $\ln(\sigma)$ and $\ln(\delta)$ and the solution" that the small-n_D issue also affects MCMC results in a practical, important way. In geotechnical engineering practice, the large-Δ issue can be overcome. The CPT typ-

Fig. 6. Relationship between (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})) and Δ for 20 cases with n_D = 30.

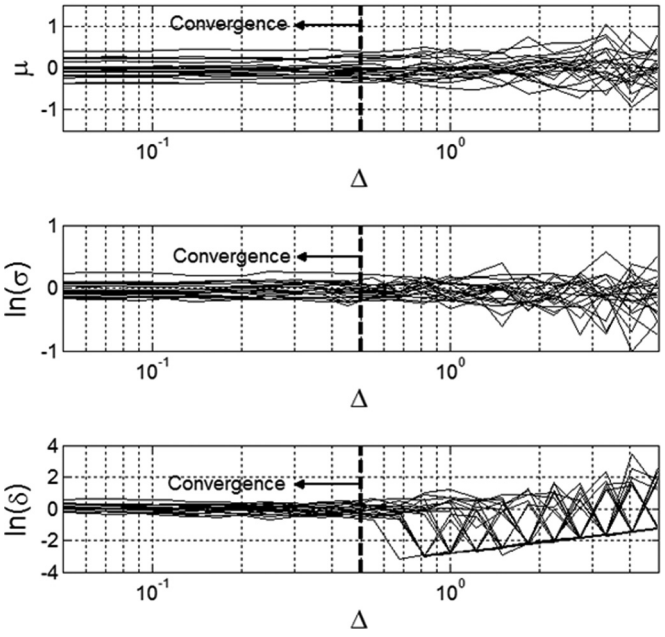
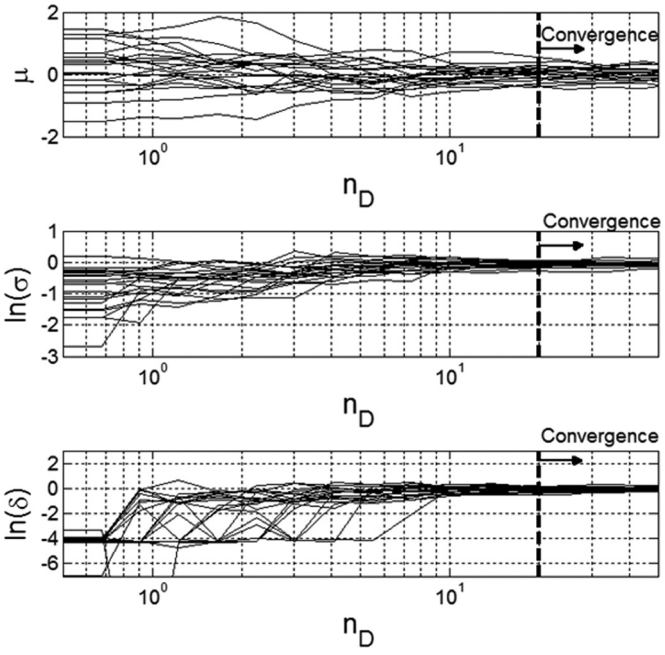


Fig. 7. Relationship between (μ_{ML}, ln(σ_{ML}), ln(δ_{ML})) and n_D for 20 cases with Δ = 0.25.



ically has a sampling depth interval dz = 0.01–0.05 m. Compared to the typical range of the vertical scale of fluctuation of 0.1–6 m (Phoon and Kulhawy 1999), Δ is sufficiently small. However, the small-n_D issue is difficult to overcome. It is common for n_D to be less than 20 in practice. Suppose the actual vertical scale of fluctuation is 1 m. This means that the total data depth must be at least 20 m within the soil layer of interest! It is common to have soil layers thinner than 20 m. Unfortunately, the small-n_D issue is more damaging than the large-Δ issue: the former makes all (μ, ln(σ), ln(δ)) estimators inconsistent. This finding suggests that we must be cautious to use the asymptotic approximations (frequentist and Bayesian) in practice because they may produce inconsistent results.

Comparison of consistencies

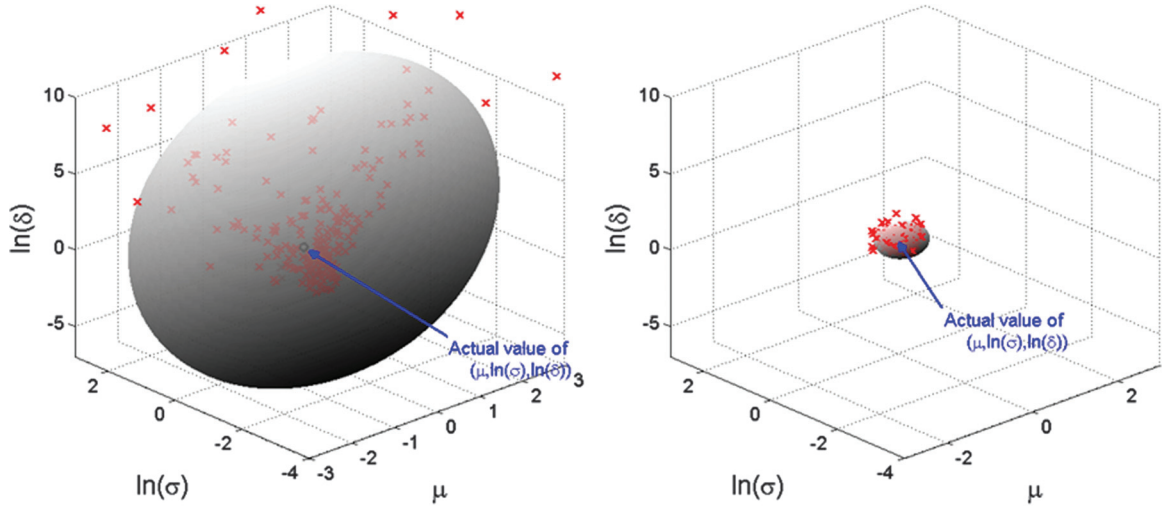
In this section, we compare the consistencies of the four methods (two frequentist and two Bayesian methods) introduced earlier. For all methods, the consistency is quantified based on how often the actual values of (μ, ln(σ), ln(δ)), namely (0, 0, 0), are within the 95% confidence region. The confidence regions for the four methods are summarized as follows:

1. (Frequentist asymptotic) The 95% confidence region based on the asymptotic approximation is adopted. This confidence region is the 3D ellipse described in eq. (7).
2. (Frequentist bootstrap) The 95% confidence region based on bootstrap samples is adopted. This confidence region is a 3D ellipse described by an inequality similar to eq. (7), but the center of the ellipse is replaced by the sample mean of the bootstrap samples, and the covariance matrix replaced by the sample covariance of the bootstrap samples. It must be understood that the "cloud" of the bootstrap samples may not resemble an ellipse. Here, the elliptical form is used for simplicity and also for comparison with the asymptotic confidence region.
3. (Bayesian asymptotic) The 95% Bayesian confidence region based on the asymptotic posterior PDF (Laplace approximation) is adopted. This confidence region is the 3D ellipse described by an inequality similar to eq. (7), but the center of the ellipse is replaced by (μ_{MAP}, ln(σ_{MAP}), ln(δ_{MAP})), and the covariance matrix replaced by $\mathbf{C}(\mu_{MAP}, \ln(\sigma_{MAP}), \ln(\delta_{MAP})|\mathbf{W})$.
4. (Bayesian MCMC) The 95% Bayesian confidence region based on MCMC samples is adopted. This confidence region is a 3D ellipse described by an inequality similar to eq. (7), but the center of the ellipse is replaced by the sample mean of the MCMC samples, and the covariance matrix replaced by the sample covariance of the MCMC samples. Again, it must be understood that the cloud of the MCMC samples may not resemble an ellipse. The elliptical form is used for simplicity and also for comparison with the asymptotic confidence region.

To illustrate the 95% confidence region, let us consider the 95% Bayesian confidence region based on MCMC samples (item 4 in the preceding text). Consider the two examples in Fig. 1. Figure 8 shows the MCMC samples for (μ, ln(σ), ln(δ)) in the 3D space. The 95% Bayesian confidence region is also shown. It is a 3D ellipsoid that roughly encloses 95% of the MCMC sample points, or the MCMC sample "cloud". The ellipsoid is large for the example in Fig. 1*a* because the statistical uncertainties (μ, ln(σ), ln(δ)) are large due to the very limited amount of data (n_D = 2), whereas it is small for the example in Fig. 1*b* because the statistical uncertainties are small due to the large amount of data (n_D = 30). It is desirable that the actual location of (μ, ln(σ), ln(δ)) = (0, 0, 0) (the locations indicated in Fig. 8) is within the ellipsoid. This means that the actual location is correctly covered by the MCMC cloud: a sign of consistency. A method is considered to perform ideally if the chance that the ellipsoid does not contain (0, 0, 0) is exactly 0.05 (1 – 0.95 = 0.05). Random simulations are performed to understand how this chance changes with respect to n_D. One hundred realizations of random fields with (μ, σ, δ) = (0, 1, 1) and with Δ = 0.25 (dz = 0.25) are simulated. Each random field realization is sampled with total depth D, and the sampled random field is used to construct the four ellipsoids described earlier. It is easy to check whether the four ellipsoids contain (0, 0, 0) using the inequalities discussed earlier. The chance that the ellipsoid does not contain (0, 0, 0) is simply [number of realizations where the ellipsoid does not contain (0, 0, 0)]/100.

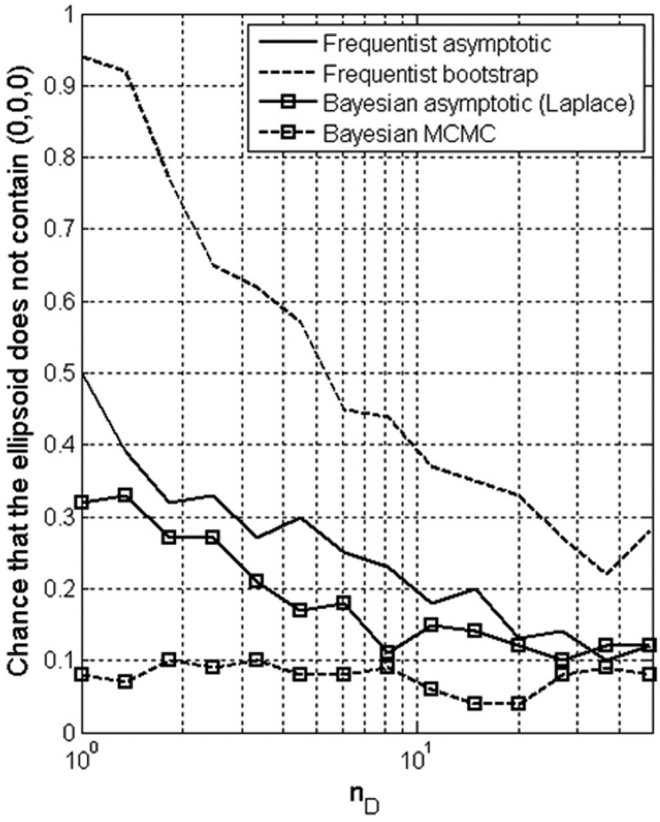
Figure 9 shows how this chance varies with respect to n_D for the four methods discussed earlier. It is evident the Bayesian MCMC

Fig. 8. MCMC samples for (μ, ln(σ), ln(δ)) in 3D space: (*a*) example in Fig. 1*a*; (*b*) example in Fig. 1*b*.



(a) Example in Figure 1a (b) Example in Figure 1b

Fig. 9. Chance that ellipsoid does not contain (0, 0, 0) versus n_D.



method outperforms the other three methods because the chance for the MCMC method is the closest to 0.05. The two asymptotic methods (frequentist and Bayesian) have similar performances. The bootstrap method has the poorest performance. Its basic assumption that the **W** data are sufficiently representative of the "population" is questionable. For large sample sizes, the adopted block bootstrap method may need unrealistically large sample sizes to converge to the correct probability of 0.05. The maximum sample size presented in Fig. 9 is $n_D = 50$. More research is needed on how to produce better bootstrap samples from a stationary random field. The earlier observation that the block bootstrap method performs poorly does not imply that other bootstrap methods would also perform poorly. When n_D is large ($n_D \geq 20$), the two asymptotic methods have performances comparable to the MCMC method. This is expected. However, when n_D is small, the two asymptotic methods are obviously outperformed by the MCMC method. This is significant: as mentioned earlier, it is quite common in geotechnical engineering practice to encounter cases with small n_D. This makes the MCMC method the most consistent method for geotechnical engineering practice.

Cases with small n_D

Cases with small n_D are important to geotechnical engineering practice because it is quite common to have thin soil layers. In this section, some aspects for cases with small n_D are further explored.

Why do asymptotic approximation methods fail?

It is insightful to understand why the asymptotic approximation methods do not perform well for cases with small n_D but perform well for cases with large n_D. Consider again the two examples in Fig. 1. Figures 10*a* and 10*b* show the actual posterior PDF $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$ together with the MCMC samples: Fig. 10*a* shows $f(\mu|\mathbf{W})$, whereas Fig. 10*b* shows $f(\ln(\sigma), \ln(\delta)|\mathbf{W})$. Here, $f(\mu|\mathbf{W})$ is obtained by numerical integration of $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$ with respect to $\ln(\sigma)$ and $\ln(\delta)$ (the marginalization theory). Similarly, $f(\ln(\sigma), \ln(\delta)|\mathbf{W})$ is obtained by numerical integration of $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$ with respect to μ. Figures 10*c* and 10*d* show the Laplace approximation for $f(\mu|\mathbf{W})$ and $f(\ln(\sigma), \ln(\delta)|\mathbf{W})$, respectively. It is clear that the Laplace approximation is a poor approximation for the actual posterior PDF. This is expected because it is an asymptotic approximation and should not perform well for nonasymptotic cases with small n_D. However, the MCMC samples correctly characterize the actual posterior PDF. This is why the MCMC method outperforms the Laplace approximation method for cases with small n_D. Figure 11 shows the results for the example in Fig. 1*b* ($n_D = 30$). Here the Laplace approximation is an acceptable approximation for the actual $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$. Nonetheless, the MCMC samples still correctly characterize the actual posterior PDF.

Effect of reducing Δ

It is also insightful to understand whether the performance for the Laplace approximation can be improved for cases with small n_D if we reduce the sampling interval Δ. As mentioned earlier, the

Fig. 10. Posterior PDF $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$ for **W** data in Fig. 1*a*: (*a*) actual $f(\mu|\mathbf{W})$; (*b*) contours for actual $f(\ln(\sigma), \ln(\delta)|\mathbf{W})$; (*c*) Laplace approximation for $f(\mu|\mathbf{W})$; (*d*) Laplace approximation for $f(\ln(\sigma), \ln(\delta)|\mathbf{W})$.

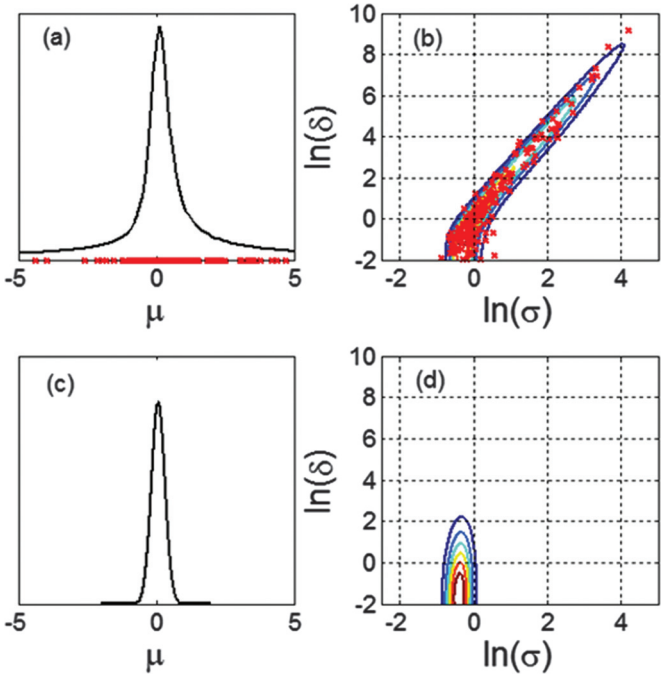
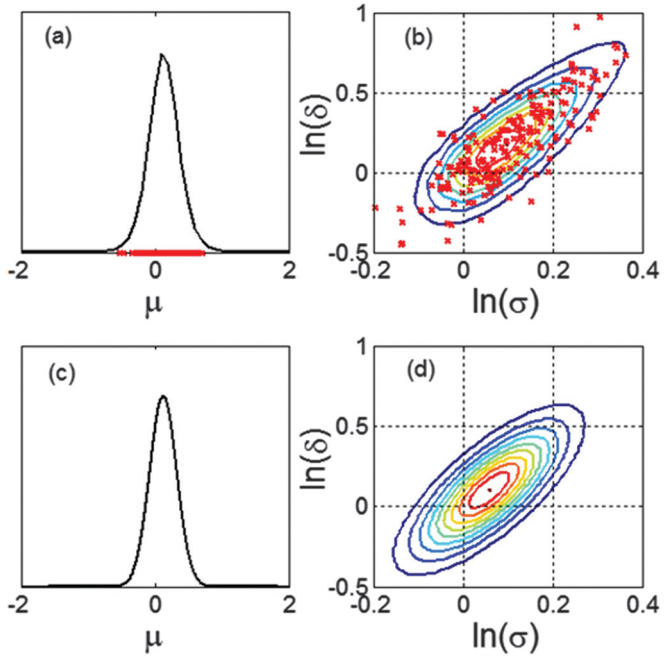
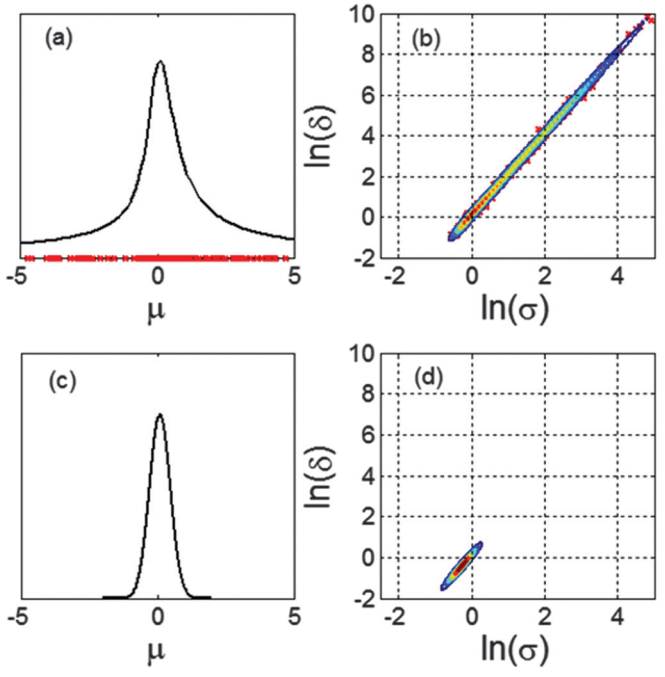


Fig. 11. Posterior PDF $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$ for **W** data in Fig. 1*b*: (*a*) actual $f(\mu|\mathbf{W})$; (*b*) contours for actual $f(\ln(\sigma), \ln(\delta)|\mathbf{W})$; (*c*) Laplace approximation for $f(\mu|\mathbf{W})$; (*d*) Laplace approximation for $f(\ln(\sigma), \ln(\delta)|\mathbf{W})$.



large-Δ issue is easier to overcome (e.g., via CPT) than the small-n_D issue. For cases with thin soil layers, the small-n_D issue is a fundamental issue that cannot be easily overcome. Figures 12*a* and 12*b* show the actual posterior PDF $f(\mu|\mathbf{W})$ and $f(\ln(\sigma), \ln(\delta)|\mathbf{W})$ together with the MCMC samples for the example in Fig. 1*a* ($n_D = 2$) but with a very small $\Delta = 0.02$. Figures 12*c* and 12*d* show the Laplace approximation for $f(\mu|\mathbf{W})$ and $f(\ln(\sigma), \ln(\delta)|\mathbf{W})$. It is clear that the Laplace approximation is still poor. This is consistent with our previous observation for the MAP estimator: reducing Δ (while keeping n_D constant) will not remove the inconsistency issue in the MAP estimators. Because the Laplace approximation heavily depends on the MAP estimators, reducing Δ does not help much in improving the performance of the Laplace approximation.

Fig. 12. Posterior PDF $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$ for **W** data in Fig. 1*a* with $\Delta = 0.02$: (*a*) actual $f(\mu|\mathbf{W})$; (*b*) contours for actual $f(\ln(\sigma), \ln(\delta)|\mathbf{W})$; (*c*) Laplace approximation for $f(\mu|\mathbf{W})$; (*d*) Laplace approximation for $f(\ln(\sigma), \ln(\delta)|\mathbf{W})$.



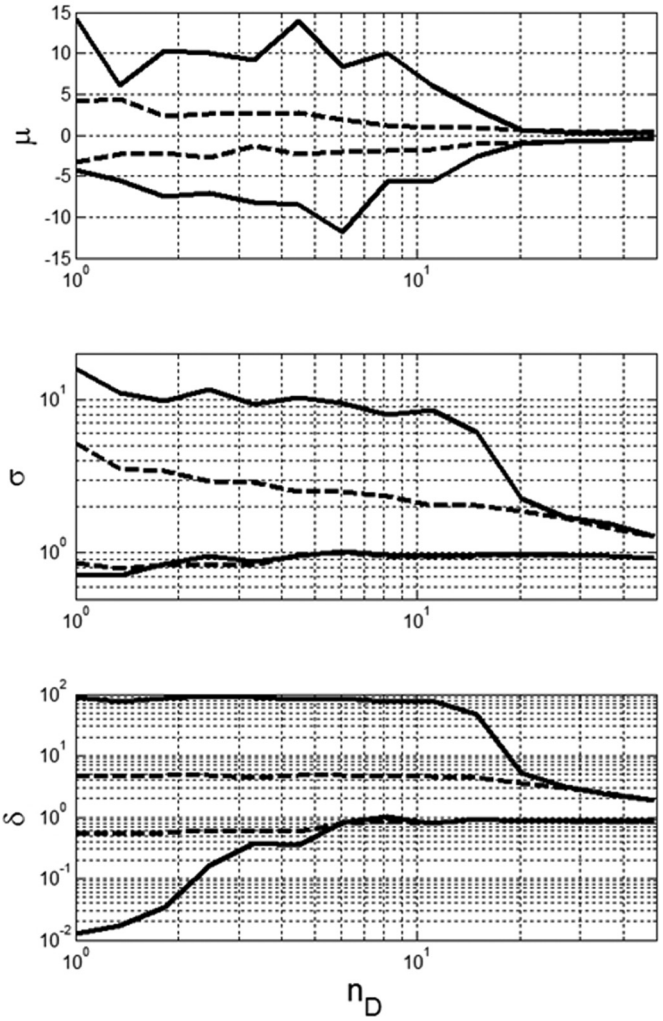
Identifiability issue for ln(σ) and ln(δ) and the solution

It is interesting to observe that there is a strong trade-off between $\ln(\sigma)$ and $\ln(\delta)$ for cases with small n_D. This can be seen from Figs. 10*b* and 12*b*. There seems to be a long and narrow region with high posterior probability density, and this region does not vanish but becomes narrower when Δ gets smaller (Fig. 12*b*). This means that when n_D is small, there are infinitely many combinations of $\ln(\sigma)$ and $\ln(\delta)$ that are all consistent with the data **W**, and these combinations roughly follow a straight line in the $\ln(\sigma)$–$\ln(\delta)$ space. Although weaker, there are also trade-offs between μ and $\ln(\sigma)$ and between μ and $\ln(\delta)$. The trade-off between $\ln(\sigma)$ and $\ln(\delta)$ is referred to as "not identifiable" in Katafygiotis and Beck (1998), meaning that there are infinitely many combinations of $\ln(\sigma)$ and $\ln(\delta)$ that are all equally consistent with the data **W**. There are two implications for the strong trade-off between $\ln(\sigma)$ and $\ln(\delta)$:

1. It is not always proper to assume conservative values for σ and δ based on previous studies and experiences, as suggested by Honjo and Setiawan (2007), even when n_D is small. It is quite likely that the assumed values of $\ln(\sigma)$ and $\ln(\delta)$ do not fall within the long and narrow region that is consistent with the data **W**.
2. When n_D is small, $\ln(\sigma)$ and $\ln(\delta)$ cannot be both estimated accurately.

The first implication prohibits us to prescribe σ and δ values simultaneously, whereas the second implication prohibits us from not prescribing them at all. However, if we prescribe the value for $\ln(\delta)$ and allow the MCMC to obtain $\ln(\sigma)$ samples, the problem will become identifiable. Conversely, one can prescribe

Fig. 13. Ranges of MCMC (μ, σ, δ) samples versus n_D. Solid lines, case where δ is bounded between 0.01 and 100; dashed lines, case where δ is bounded between 0.5 and 5.



ln(σ) and allow the MCMC to obtain ln(δ) samples. In the Bayesian framework, a range of ln(δ) can be prescribed, rather than a value of ln(δ). This can be done by adopting a uniform prior PDF for ln(δ), e.g., f(ln(δ)) is uniform from ln(0.5) to ln(5). The bounds can be specified based on previous studies and experiences. For instance, Phoon and Kulhawy (1999) showed that the vertical scale of fluctuation for CPT cone resistance q_c (or q_t) ranges from 0.1 to 2.2 m based on 17 previous studies. A range of ln(σ) can be prescribed by adopting a uniform prior PDF for ln(σ) as well based on previous studies and experiences.

The impact of adopting a prior PDF for ln(δ) is demonstrated in Fig. 13. The solid lines in Fig. 13 show how the ranges of the MCMC (μ, σ, δ) samples vary with n_D for a case where the prior PDF f(ln(δ)) is uniform from ln(0.01) to ln(100). A unbounded flat prior (uniform between $-\infty$ and ∞) is adopted for ln(σ). The upper solid lines are the 97.5 sample percentiles of the MCMC samples, whereas the lower solid lines are the 2.5 sample percentiles. One can see when n_D is large ($n_D \geq 20$), (σ, δ) samples are within narrow ranges. However, when n_D is small, (σ, δ) samples spread widely: they are not identifiable as a whole. As mentioned earlier, this identifiability issue cannot be resolved by reducing Δ. The wide range is not acceptable to geotechnical engineering practice, e.g., for the case with $n_D = 2$, σ ranges from 0.9 to 10 and δ ranges from 0.06 to 100. As another demonstration, a prior PDF f(ln(δ)) that is uniform between ln(0.5) and ln(5) is adopted. An unbounded flat prior is adopted for ln(σ). The dashed lines in Fig. 13 show how the ranges of the MCMC (μ, σ, δ) samples vary with n_D for the same random field realization. When n_D is small, not only is the range for δ significantly reduced, but the ranges for μ and σ are also significantly reduced. Now the range may be acceptable to geotechnical engineering practice, e.g., for the case with $n_D = 2$, σ ranges from 0.8 to 3 and δ ranges from 0.05 to 5. This shows the positive impact of adopting a prior PDF for ln(δ) for cases with small n_D. For cases with large n_D ($n_D \geq 20$), the prior PDF has no impact.

It is interesting that the small-n_D issue causing the inconsistency in the ML and MAP estimators also has a certain implication on the MCMC results here. It seems that $n_D = 20$ is not only a threshold for the ML and MAP estimators but also a threshold for whether the prior PDF is needed. For cases with $n_D \geq 20$, not only are the ML and MAP estimators consistent but also a strong prior PDF is not needed for the MCMC. However, for cases with $n_D < 20$, the ML and MAP estimators may become inconsistent, and a strong prior PDF is needed for the MCMC. In summary, cases with $n_D < 20$ should be treated differently from cases with $n_D \geq 20$. For the former ($n_D < 20$), it is recommended that the MCMC should be used together with a strong prior PDF for ln(δ). For the latter ($n_D \geq 20$), it is recommended that an asymptotic approximation method (either frequentist or Bayesian) can be used.

Recommended procedure for estimation of (μ, σ, δ)

Based on the earlier observations, the following procedure is proposed for the estimation of (μ, σ, δ):

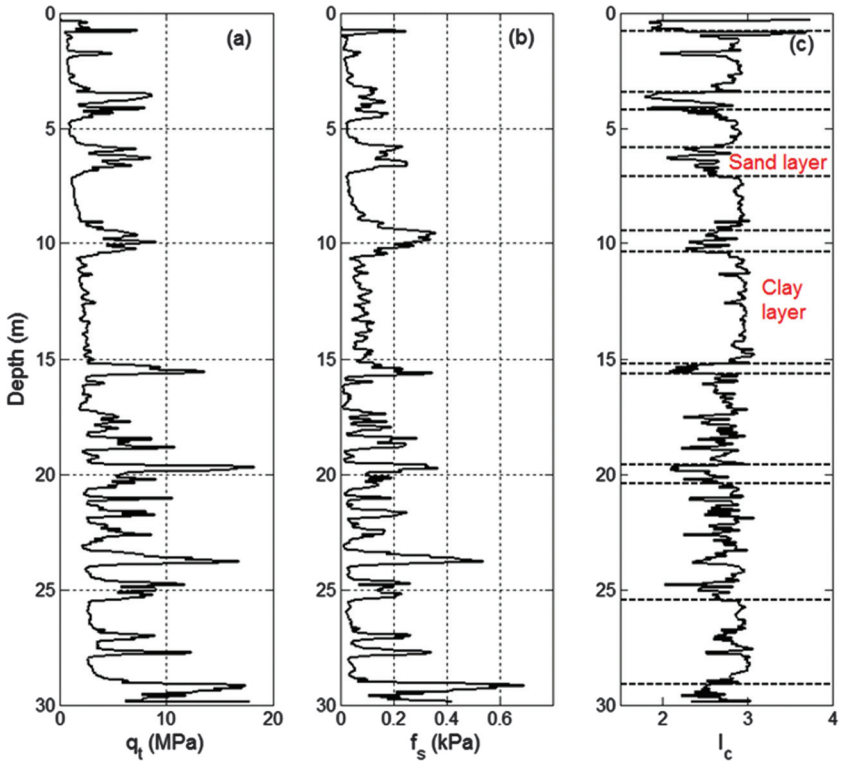
a. Given the **W** data, first find the MAP (μ_{MAP}, ln(σ_{MAP}), ln(δ_{MAP})). Then, determine **C**(μ_{MAP}, ln(σ_{MAP}), ln(δ_{MAP})|**W**) as the inverse of the Hessian matrix for $-\ln[f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})]$.
b. If D is larger than 20 times the resulting δ_{MAP} (namely, $n_D \geq 20$), the posterior PDF $f(\mu, \ln(\sigma), \ln(\delta)|\mathbf{W})$ is then approximated (Laplace approximation) by the multivariate normal distribution with mean (μ_{MAP}, ln(σ_{MAP}), ln(δ_{MAP})) and covariance matrix **C**(μ_{MAP}, ln(σ_{MAP}), ln(δ_{MAP})|**W**). The Bayesian confidence region (a 3D ellipsoid) can be constructed using eq. (7), but the center of the ellipse is replaced by (μ_{MAP}, ln(σ_{MAP}), ln(δ_{MAP})), and the covariance matrix replaced by **C**(μ_{MAP}, ln(σ_{MAP}), ln(δ_{MAP})|**W**). There is a high chance (≈ 0.95) that the actual values of (μ, ln(σ), ln(δ)) are within this ellipsoid.
c. If D is smaller than 20 times the resulting δ_{MAP} (namely, $n_D < 20$), implement the MCMC to obtain the (μ, ln(σ), ln(δ)) samples. We encourage the use of a prior PDF for ln(δ) based on past studies and experiences to limit its range and allow the MCMC to obtain the ln(σ) samples. One can use (μ_{MAP}, ln(σ_{MAP}), ln(δ_{MAP})) as the initial sample for the MCMC to minimize the burn-in period. One can also set the covariance matrix of the proposal PDF to be equal to **C**(μ_{MAP}, ln(σ_{MAP}), ln(δ_{MAP})|**W**) to minimize the dependency among the MCMC samples. Based on the MCMC samples, the Bayesian confidence region (also a 3D ellipsoid) can be constructed using eq. (7), but the center of the ellipse is replaced by the sample mean of the MCMC samples, and the covariance matrix replaced by the sample covariance of the MCMC samples. There is a high chance (≈ 0.95) that the actual values of (μ, ln(σ), ln(δ)) are within this ellipsoid.

In principle, the frequentist asymptotic approximation method can be used in step b. However, the Bayesian asymptotic approximation method is adopted there. The entire procedure follows the Bayesian frame work in this case.

Real example

A CPT sounding at Wufeng District in Taichung City (Taiwan) is analyzed. The geology of Wufeng consists of an alluvial plain crossed by several rivers underlain by Pliocene sandstones, shales, and mudstones. The Wufeng site suffered significant liquefaction

Fig. 14. CPT data at Wufeng site: (*a*) q_t; (*b*) f_s; (*c*) I_c.



during the 1999 Chi-Chi earthquake. The detailed documentation of this site can be found in Chu et al. (2004). The CPT sounding location is underlain by interlayers of silty sands and silty clays, and the groundwater table is about 1–2 m deep. Figures 14*a* and 14*b* show the CPT data (q_t and sleeve friction f_s), together with the soil behavior type index profile (I_c) (Robertson 2009) in Fig. 14*c*. The data interval dz is 0.05 m. The CPT-based stratification result based on I_c is also shown in Fig. 14*c* (layer boundaries shown as horizontal dashed lines). In this example, the estimation of (μ, σ, δ) will be demonstrated for two soil layers — one sand layer and one clay layer, shown in Fig. 14*c*. For the sand layer, the logarithm of the normalized cone resistance Q_{tn} (Robertson 2009) with n = 0.5 is analyzed, whereas for the clay layer, ln(Q_{tn}) with n = 1 is analyzed:

$$(10)\qquad Q_{tn} = [(q_t - \sigma_{v0})/P_a]\left(P_a/\sigma'_{v0}\right)^n$$

where q_t is the (corrected) cone resistance; P_a = 101.3 kN/m^2 is one atmosphere pressure; σ'_{v0} and σ_{v0} are the effective and total overburden stresses, respectively. Figure 15 shows the Q_{tn} profiles within these two soil layers in the logarithmic scale. It is visually plausible from Fig. 15 that the normalized cone resistance can be viewed as a realization from a stationary field. It is possible to achieve stationarity by de-trending q_t as well, but normalization may be more robust, as the trend function can change from location to location. The total depths (D) for the CPT records are 0.80 and 4.55 m for the sand and clay layers, respectively. The analysis follows the recommended procedure presented in the last section and assumes that the measurement errors in CPT are small; therefore, the uncertainties in (μ, σ, δ) are mainly statistical uncertainties. This assumption is justifiable because Phoon and Kulhawy (1999) reported the coefficient of variation of measurement error for CPT to be between 5% and 15%.

Sand layer

The MAP estimates (μ_{MAP}, ln(σ_{MAP}), ln(δ_{MAP})) are found to be (4.07, –1.20, –1.77), hence δ_{MAP} = exp(–1.77) = 0.17 m. Because D is smaller than 20δ_{MAP} (D = 0.8 m), this soil layer is considered as a thin soil layer, and the MCMC is implemented to obtain the (μ, ln(σ), ln(δ)) samples. The prior PDF for ln(δ) is taken to be uniform between ln(0.1) and ln(10). Based on the sample mean and sample covariance of the MCMC samples, the posterior PDF f(μ, ln(σ), ln(δ)|**W**) is approximated by the multivariate normal distribution with the following mean and covariance matrix:

$$(11)\qquad f[\mu,\ \ln(\sigma),\ \ln(\delta)|\mathbf{W}] = N\left(\begin{bmatrix} 4.10 \\ -0.60 \\ -0.52 \end{bmatrix}, \begin{bmatrix} 0.236 & \approx 0 & \approx 0 \\ & 0.316 & 0.652 \\ \text{SYM} & & 1.516 \end{bmatrix}\right)$$

There is a high confidence that the actual (μ, ln(σ), ln(δ)) are within the Bayesian confidence region of this multivariate normal PDF.

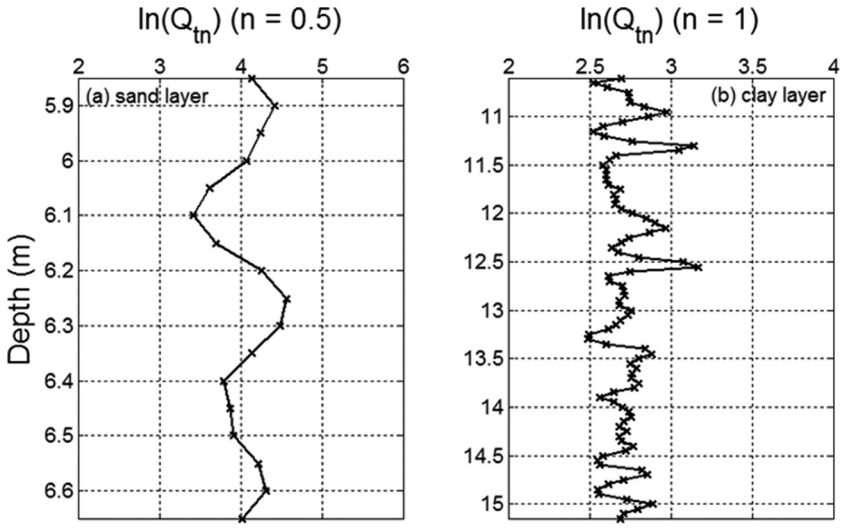
Clay layer

The MAP estimates (μ_{MAP}, ln(σ_{MAP}), ln(δ_{MAP})) are found to be (2.72, –2.04, –1.60), hence δ_{MAP} = exp(–1.60) = 0.20 m. Because D is larger than 20δ_{MAP} (D = 4.55 m), this soil layer is considered as a thick soil layer, and the Laplace approximation is adopted: the posterior PDF f(μ, ln(σ), ln(δ)|**W**) is approximated by the multivariate normal distribution with mean (μ_{MAP}, ln(σ_{MAP}), ln(δ_{MAP})) and covariance matrix **C**(μ_{MAP}, ln(σ_{MAP}), ln(δ_{MAP})|**W**):

$$(12)\qquad f[\mu,\ \ln(\sigma),\ \ln(\delta|\mathbf{W}] = N\left(\begin{bmatrix} 2.72 \\ -2.04 \\ -1.60 \end{bmatrix}, \begin{bmatrix} 7.27 \times 10^{-4} & 0 & 0 \\ & 0.0115 & 0.0209 \\ \text{SYM} & & 0.0724 \end{bmatrix}\right)$$

There is a high confidence that the actual (μ, ln(σ), ln(δ)) are within the Bayesian confidence region of this multivariate normal PDF.

Fig. 15. Q_{tn} profiles: (*a*) sand layer; (*b*) clay layer.



Conclusions

1. Based on the comparison results, it is concluded that the MCMC method (Bayesian) is the most consistent method for the estimation of (μ, σ, δ), in the sense that the actual values of (μ, σ, δ) are encompassed by the "cloud" of the MCMC samples with a large chance. This conclusion is based on a realistic upper bound of $n_D < 50$. In particular, for cases with small n_D (thin soil layers), the MCMC method is the only consistent method (among the four methods). For cases with large n_D ($n_D \geq 20$) (thick soil layers), both asymptotic approximation methods (frequentist and Bayesian) perform equally well to the MCMC method. The adopted block bootstrap method (frequentist) is not consistent, even up to a fairly large value of $n_D = 50$. It is possible that the adopted block bootstrap method may need unrealistically large sample sizes to converge to the prescribed confidence level, or the method's performance is strongly dependent on choice of block length L. More research is needed.
2. The two asymptotic approximation methods rely on the ML and MAP estimators. However, there are consistency issues for the ML and MAP estimators: the small-n_D and large-Δ issues. The ML and MAP estimators cease to be consistent if n_D is too small (<20) or Δ is too large (>0.5). The large-Δ issue is easier to overcome (e.g., via CPT), but the small-n_D issue is a fundamental issue (thin soil layers) and is difficult to overcome. The MCMC method is the only consistent method for cases with small n_D.
3. There are more issues for cases with small n_D: σ and δ are not identifiable as a whole, even if the MCMC method is applied. This is because there is a strong trade-off between σ and δ when n_D is small. This identifiability issue disappears when $n_D \geq 20$. The upshot is that it is inadequate to treat both σ and δ as unknowns because there are infinitely many combinations of σ and δ that are (roughly) equally plausible, and these combinations roughly follow a straight line in the ln(σ)–ln(δ) space. To resolve this identifiability issue, we partially agree with Honjo and Setiawan (2007) in that we need certain prior knowledge regarding σ and δ (e.g., previous studies and experiences). However, it may not be appropriate to prescribe both the σ and δ values as suggested by Honjo and Setiawan (2007) because the prescribed combination of σ and δ may not be close to the straight line in the ln(σ)–ln(δ) space, i.e., the prescribed combination cannot produce the observed data. A more appropriate strategy that will maintain consistency with the observed data is to adopt a prior PDF for ln(δ) to limit its range and allow the MCMC to obtain the ln(σ) samples.

Acknowledgements

The first two authors would like to express their gratitude to the National Taiwan University. The second author was supported by the "Aim for Top University" project of National Taiwan University (project No. 103R4000) during his postdoc study. We would also like to thank Prof. C.S. Ku of the I-Shou University (Kaohsiung, Taiwan) for providing the CPT data for the real example.

References

Ahmed, A., and Soubra, A.H. 2012. Probabilistic analysis of strip footings resting on a spatially random soil using subset simulation approach. Georisk, **6**(3): 188–201. doi:10.1080/17499518.2012.678775.

Ahmed, A., and Soubra, A.H. 2014. Probabilistic analysis at the serviceability limit state of two neighboring strip footings resting on a spatially random soil. Structural Safety, **49**: 2–9. doi:10.1016/j.strusafe.2013.08.001.

Bleistein, N., and Handelsman, R. 1986. Asymptotic expansions of integrals. Dover, New York.

Cafaro, F., and Cherubini, C. 2002. Large sample spacing in evaluation of vertical strength variability of clayey soil. Journal of Geotechnical and Geoenvironmental Engineering, ASCE, **128**(7): 558–568. doi:10.1061/(ASCE)1090-0241(2002)128:7(558).

Cao, Z., and Wang, Y. 2014. Bayesian model comparison and characterization of undrained shear strength. Journal of Geotechnical and Geoenvironmental Engineering, ASCE, **140**(6): 04014018. doi:10.1061/(ASCE)GT.1943-5606.0001108.

Chen, B.S.Y., and Mayne, P.W. 1994. Profiling the overconsolidation ratio of clays by piezocone tests. Report No. GIT-CEE/GEO-94-1. Submitted to the National Science Foundation by Georgia Institute of Technology, Atlanta, August 1994.

Chen, B.S.Y., and Mayne, P.W. 1996. Statistical relationships between piezocone measurements and stress history of clays. Canadian Geotechnical Journal, **33**(3): 488–498. doi:10.1139/t96-070.

Ching, J., and Liao, H.-J. 2013. Re-analysis of Freeway-3 dip slope failure case – a spatial variability view. Journal of GeoEngineering, **8**(1): 1–10.

Ching, J., and Phoon, K.K. 2012. Establishment of generic transformations for geotechnical design parameters. Structural Safety, **35**: 52–62. doi:10.1016/j.strusafe.2011.12.003.

Ching, J., Phoon, K.-K., and Chen, C.-H. 2014. Modeling piezocone cone penetration (CPTU) parameters of clays as a multivariate normal distribution. Canadian Geotechnical Journal, **51**(1): 77–91. doi:10.1139/cgj-2012-0259.

Cho, S.E. 2007. Effects of spatial variability of soil properties on slope stability. Engineering Geology, **92**: 97–109. doi:10.1016/j.enggeo.2007.03.006.

Chu, D.B., Stewart, J.P., Lee, S., Tsai, J.S., Lin, P.S., Chu, B.L., Seed, R.B., Hsu, S.C., Yu, M.S., and Wang, M.C.H. 2004. Documentation of soil conditions at liquefaction and non-liquefaction sites from 1999 Chi-Chi (Taiwan) earthquake. Soil Dynamics and Earthquake Engineering, **24**: 647–657. doi:10.1016/j.soildyn.2004.06.005.

Dasaka, S.M., and Zhang, L.M. 2012. Spatial variability of in situ weathered soil. Géotechnique, **62**(5), 375–384. doi:10.1680/geot.8.P.151.3786.

DeGroot, D.J., and Baecher, G.B. 1993. Estimating autocovariance of in-situ soil

properties. Journal of Geotechnical Engineering, ASCE, **119**(1): 147–166. doi:10.1061/(ASCE)0733-9410(1993)119:1(147).
Efron, B., and Tibshirani, R. 1993. An introduction to the bootstrap. Chapman and Hall/CRC, Boca Raton, Fla.
El-Ramly, H., Morgenstern, N.R., and Cruden, D.M. 2003. Probabilistic stability analysis of a tailings dyke on presheared clay–shale. Canadian Geotechnical Journal, **40**(1): 192–208. doi:10.1139/t02-095.
Fan, H., and Liang, R. 2013. Reliability-based design of laterally loaded piles considering soil spatial variability. *In* Foundation engineering in the face of uncertainty (ASCE GSP 229), pp. 475–486. doi:10.1061/9780784412763.036.
Fenton, G. 1999. Random field modeling of CPT data. Journal of Geotechnical and Geoenvironmental Engineering, ASCE, **125**(6): 486–498. doi:10.1061/(ASCE)1090-0241(1999)125:6(486).
Fenton, G.A., and Griffiths, D.V. 2002. Probabilistic foundation settlement on spatially random soil. Journal of Geotechnical and Geoenvironmental Engineering, ASCE, **128**(5): 381–390. doi:10.1061/(ASCE)1090-0241(2002)128:5(381).
Fenton, G.A., and Griffiths, D.V. 2003. Bearing capacity prediction of spatially random c–ϕ soils. Canadian Geotechnical Journal, **40**(1): 54–65. doi:10.1139/t02-086.
Fenton, G.A., Griffiths, D.V., and Williams, M.B. 2005. Reliability of traditional retaining wall design. Géotechnique, **55**(1): 55–62. doi:10.1680/geot.2005.55.1.55.
Firouzianbandpey, S., Griffiths, D.V., Ibsen, L.B., and Andersen, L.V. 2014. Spatial correlation length of normalized cone data in sand: Case study in the north of Denmark. Canadian Geotechnical Journal, **51**(8): 844–857. doi:10.1139/cgj-2013-0294.
Gilks, W.R., Richardson, S., and Spiegelhalter, D. 1995. Markov Chain Monte Carlo in practice. Chapman and Hall/CRC, Boca Raton, Fla.
Griffiths, D.V., and Fenton, G.A. 2004. Probabilistic slope stability analysis by finite elements. Journal of Geotechnical and Geoenvironmental Engineering, ASCE, **130**(5): 507–518. doi:10.1061/(ASCE)1090-0241(2004)130:5(507).
Griffiths, D.V., Fenton, G.A., and Ziemann, H.R. 2008. Reliability of passive earth pressure. Georisk, **2**(2): 113–121. doi:10.1080/17499510802178640.
Griffiths, D.V., Huang, J., and Fenton, G.A. 2011. Probabilistic infinite slope analysis. Computers and Geotechnics, **38**(4): 577–584. doi:10.1016/j.compgeo.2011.03.006.
Haldar, S., and Sivakumar, Babu, G.L. 2008. Effect of soil spatial variability on the response of laterally loaded pile in undrained clay. Computers and Geotechnics, **35**(4): 537–547. doi:10.1016/j.compgeo.2007.10.004.
Hicks, M.A., and Spencer, W.A. 2010. Influence of heterogeneity on the reliability and failure of a long 3D slope. Computers and Geotechnics, **37**(7–8): 948–955. doi:10.1016/j.compgeo.2010.08.001.
Honjo, Y., and Setiawan, B. 2007. General and local estimation of local average and their application in geotechnical parameter estimations. Georisk, **1**(3): 167–176. doi:10.1080/17499510701745960.
Hu, Y.G., and Ching, J. 2015. Impact of spatial variability in undrained shear strength on active lateral force in clay. Structural Safety, **52**: 121–131. doi:10.1016/j.strusafe.2014.09.004.
Jaksa, M.B., Goldsworthy, J.S., Fenton, G.A., Kaggwa, W.S., Griffiths, D.V., Kuo, Y.L., and Poulos, H.G. 2005. Towards reliable and effective site investigations. Géotechnique, **55**(2): 109–121. doi:10.1680/geot.2005.55.2.109.
Jaynes, E.T. 1976. Confidence intervals versus Bayesian intervals. *In* Foundations of probability theory, statistical inference, and statistical theories of science. *Edited by* W.L. Harper and C.A. Hooker. D. Reidel, Dordrecht.
Katafygiotis, L.S., and Beck, J.L. 1998. Updating models and their uncertainties. II: Model identifiability. Journal of Engineering Mechanics, ASCE, **124**(4): 463–467. doi:10.1061/(ASCE)0733-9399(1998)124:4(463).
Künsch, H.R. 1989. The jackknife and the bootstrap for general stationary observations. Annals of Statistics, **17**: 1217–1241. doi:10.1214/aos/1176347265.
Lee, P.M. 2012. Bayesian statistics: an introduction. John Wiley & Sons, Hoboken, New Jersey.
Li, D.Q., Qi, X.H., Phoon, K.K., Zhang, L.M., and Zhou, C.B. 2014. Effect of spatially variable shear strength parameters with linearly increasing mean trend on reliability of infinite slopes. Structural Safety, **49**: 45–55. doi:10.1016/j.strusafe.2013.08.005.
Lloret-Cabot, M., Fenton, G.A., and Hicks, M.A. 2014. On the estimation of scale of fluctuation in geostatistics. Georisk, **8**(2): 129–140. doi:10.1080/17499518.2013.871189.
Low, B.K., Lacasse, S., and Nadim, F. 2007. Slope reliability analysis accounting for spatial variation. Georisk, **1**(4): 177–189. doi:10.1080/17499510701772089.
Luo, Z., Atamturktur, S., Juang, C.H., Huang, H.W., and Lin, P.S. 2011. Probability of serviceability failure in a braced excavation in a spatially random field: Fuzzy finite element approach. Computers and Geotechnics, **38**(8): 1031–1040. doi:10.1016/j.compgeo.2011.07.009.
Mardia, K.V., and Marshall, R.J. 1984. Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika, **71**(1): 135–146. doi:10.1093/biomet/71.1.135.
Mayne, P.W. 1986. CPT indexing of in situ OCR in clays. *In* Use of In situ tests in geotechnical engineering (ASCE GSP 6), pp. 780–793.
Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. 1953. Equation of state calculations by fast computing machines. Journal of Chemical Physics, **21**: 1087–1092. doi:10.1063/1.1699114.
Phoon, K.K., and Fenton, G.A. 2004. Estimating sample autocorrelation functions using bootstrap. *In* Proceedings, Ninth ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability, Albuquerque, New Mexico, 26–28 July 2004 [CDROM].
Phoon, K.K., and Kulhawy, F.H. 1999. Characterization of geotechnical variability. Canadian Geotechnical Journal, **36**(4): 612–624. doi:10.1139/t99-038.
Phoon, K.K., Quek, S.T., and An, P. 2003. Identification of statistically homogeneous soil layers using modified Bartlett statistics. Journal of Geotechnical and Geoenvironmental Engineering, ASCE, **129**(7): 649–659. doi:10.1061/(ASCE)1090-0241(2003)129:7(649).
Politis, D.N., and Romano, J.P. 1992. A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. Annals of Statistics, **20**, 1985–2007. doi:10.1214/aos/1176348899.
Politis, D.N., and White, H.L. 2004. Automatic block-length selection for the dependent bootstrap. Econometric Reviews, **23**: 53–70. doi:10.1081/ETC-120028836.
Robertson, P.K. 2009. Interpretation of cone penetration tests — a unified approach. Canadian Geotechnical Journal, **46**(11): 1337–1355. doi:10.1139/T09-065.
Robertson, P.K., Campanella, R.G., Gillespie, D., and Greig, J. 1986. Use of piezometer cone data. *In* Use of In situ tests in geotechnical engineering (ASCE GSP 6), pp. 1263–1280.
Senneset, K., Sandven, R., and Janbu, N. 1989. Evaluation of soil parameters from piezocone tests. *In* In situ testing of soil properties for transportation. Transportation Research Record No. 1235, pp. 24–37.
Srivastava, A., and Sivakumar, Babu, G.L. 2009. Effect of soil variability on the bearing capacity of clay and in slope stability problems. Engineering Geology, **108**(1–2): 142–152. doi:10.1016/j.enggeo.2009.06.023.
Uzielli, M., Vannucchi, G., and Phoon, K.K. 2005. Random field characterisation of stress-normalised cone penetration testing parameters. Géotechnique, **55**(1): 3–20. doi:10.1680/geot.2005.55.1.3.
Vanmarcke, E.H. 1977. Probabilistic modeling of soil profiles. Journal of Geotechnical Engineering, ASCE, **GT11**: 1227–1246.
Vanmarcke, E.H. 1983. Random fields: analysis and synthesis. The MIT Press, Cambridge, Mass.
Wang, Y., and Cao, Z. 2013*a*. Expanded reliability-based design of piles in spatially variable soil using efficient Monte Carlo simulations. Soils and Foundations, **53**(6): 820–834. doi:10.1016/j.sandf.2013.10.002.
Wang, Y., and Cao, Z. 2013*b*. Probabilistic characterization of Young's modulus of soil using equivalent samples. Engineering Geology, **159**: 106–118. doi:10.1016/j.enggeo.2013.03.017.
Wang, Y., Au, S.K., and Cao, Z. 2010. Bayesian approach for probabilistic characterization of sand friction angles. Engineering Geology, **114**: 354–363. doi:10.1016/j.enggeo.2010.05.013.
Wickremesinghe, D., and Campanella, R.G. 1993. Scale of fluctuation as a descriptor of soil variability. *In* Probabilistic methods in geotechnical engineering. *Edited by* K.S. Li and S.-C.R. Lo. Balkema, Rotterdam, the Netherlands, pp. 233–239.
Wu, S.H., Ou, C.Y., Ching, J., and Juang, C.H. 2012. Reliability-based design for basal heave stability of deep excavations in spatially varying soils. Journal of Geotechnical and Geoenvironmental Engineering, ASCE, **138**(5): 594–603. doi:10.1061/(ASCE)GT.1943-5606.0000626.