Quantification of statistical uncertainties of rock strength parameters using Bayesian-based Markov Chain Monte Carlo method

Liang Han\(^2\), Lin Wang\(^2\), Wengang Zhang\(^{1,2,3*}\)

1 Key Laboratory of New Technology for Construction of Cities in Mountain Area, Chongqing University, Chongqing 400045, China
2 School of Civil Engineering, Chongqing University, Chongqing 400045, China
3 National Joint Engineering Research Center of Geohazards Prevention in the Reservoir Areas, Chongqing University, Chongqing 400045, China

* corresponding author: zhangwg@cqu.edu.cn/zhangwg@ntu.edu.sg

Abstract. Although unconfined compressive strength (UCS) plays an important role in geotechnical design and analysis involving rock materials, how to quantify the statistical uncertainties underlying rock strength parameters is rarely reported. Based on a site investigation report in Bukit Timah Granite (BTG) formation in Singapore, this paper presents a set of database about UCS from four sites in BTG formation. Subsequently, Markov Chain Monte Carlo (MCMC) algorithm was applied to quantitatively evaluate the uncertainties of statistical parameters including the mean value, variance, and autocorrelation distance of UCS of BTG rocks making use of the available test data under the Bayesian framework. It was proven that the Bayesian-based MCMC method can effectively quantify the uncertainty of geo-mechanical parameters via a series of equivalent samples. The results indicate that the uncertainties of statistical parameters of UCS of BTG rocks are significant, and the magnitude to some extent relies on the selection of basic parameter in Bayesian framework. In terms of the basic parameters, the sensitive degree of uncertainties of three statistical parameters is different from each other.

1. Introduction
Strength parameters, i.e., unconfined compressive strength (UCS), play an important role in rock engineering practice. According to previous researches, the rock properties such as UCS and elastic modulus are of great uncertainty [1-3]. Based on references [4, 5], the sources of uncertainty of geo-mechanical materials can be attributed to three parts: (1) inherent uncertainty; (2) transformation uncertainty; (3) statistical uncertainty. The inherent uncertainty is related to the natural processes of geological formation, and the transformation uncertainty is related to the transformation model among several parameters. Due to the limitation of budget and time line, the data from site investigation is generally sparse, which will unavoidably lead to the statistical uncertainties of rock property parameters. However, the majority of previous literatures mainly focus on the inherent uncertainty and transformation uncertainty of rock parameters, while the statistical uncertainty is rarely explored [6-8]. Obviously, if the statistical uncertainty embedded in the database from a field investigation cannot be
figured out, the inherent and transformation uncertainty will be under/overestimated so that the probabilistic analysis results can be far away from the actual situation. In this paper, with a site investigation report in Bukit Timah Granite (BTG) formation in Singapore, the Bayesian-based Markov chain Monte Carlo method was employed to quantitatively evaluate the statistical parameters (mean value, variance, and autocorrelation distance) of UCS of BTG rocks. And the whole procedure is presented with the basis of four boreholes from four different sites in BTG formation. It was proven that the Bayesian-based MCMC method can effectively quantify the uncertainty of geo-mechanical statistical parameters [9-11] via a series of equivalent samples making use of the site-specific test data.

2. Database and pre-processing

2.1 Database introduction
This field site investigation was conducted in BTG formation in Singapore, which consists of four sites, namely Site A, Site B, Site C, and Site D, and BTG formation occupies about one-third area of Singapore land. In this site investigation, there are 165 boreholes and 742 cores, and each borehole has approximately 4.5 cores as shown in Table 1. Figure 1 shows that the sampling intervals among these boreholes are generally greater than 1.0 m, and some of them are greater than 2.5 m. From Figure 2, it can be seen that UCS ranges from 0 to 240 MPa, and the trend along the depth is not evident.

| Site    | Number of Borehole | Number of cores | Average number of cores per borehole |
|---------|-------------------|-----------------|-------------------------------------|
| Site A  | 48                | 191             | 4.0                                 |
| Site B  | 36                | 171             | 4.8                                 |
| Site C  | 48                | 201             | 4.2                                 |
| Site D  | 33                | 179             | 5.4                                 |
| Overall | 165               | 742             | 4.5                                 |

Figure 1. Sampling intervals among these boreholes
2.2 Data pre-processing

Firstly, the natural logarithm transformation is applied to UCS, because UCS is a non-negative parameter, and the ln(UCS) is denoted by Y to keep the notation concise. It can be seen that the trend is much less pronounced for ln(UCS) along the depth than it was for UCS, and the ln(UCS) has the mean value of 4.13 MPa, standard deviation of 0.64 MPa, and the coefficient of variance (COV) of 0.16. From Figure 3, it can be seen that statistical uncertainty is significant, and the correlation among the mean, the variance, and the autocorrelation distance is not evident.

To explore the statistical uncertainty, four boreholes (i.e., borehole A, borehole B, borehole C, and borehole D) in four sites are selected as the research objective, and the number of cores is 10 except the borehole of Site B (see Figure 4). The differences of mean values among these boreholes are slightly significant, while the differences of variance and COV are obvious (see Table 2). Based on Honjo and Kazumba (2002) [13] and Namikawa (2019) [11], the maximum-likelihood method can be used for analyzing the autocorrelation distance and generally produce satisfactory results, hence it is adopted in this study.
Geotechnical properties generally have been modelled to follow commonly used distribution model, in the view of references [14, 15], normal or lognormal distribution model can provide a good fitting for UCS. In this study, since the natural logarithm transformation has been conducted for UCS, the distribution characteristic of $Y$ can be described using normal distribution model. Besides, the spatial autocorrelation is a common phenomenon for geo-materials, so does the UCS or $Y$. There are several correlation structures in the literatures (e.g., [16]) for illustrating the spatial autocorrelation, including single exponential correlation function (SECF), binary noise correlation function (BNCF), second-order Markov correlation function (SMCF), and squared exponential correlation function (SQECF). For simplicity and generalization, the SECF (see equation (1)) is adopted in this study. Note that only the vertical spatial variability was considered in this study.

$$\rho_{ij} = \exp\left(-\frac{|h_i - h_j|}{\theta}\right)$$

Where $\rho_{ij}$ represents the correlation coefficient between $Y_i$ and $Y_j$ obtained from core $i$ and $j$, $h_i$ and $h_j$ are the depths of two cores in a borehole, $\theta$ is the autocorrelation distance.

It is assumed that $Y$ follows a multivariable normal distribution as shown in equation (2).

$$f(Y|\theta) = \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} \exp\left\{-\frac{1}{2}(Y - \mu_Y)^T \Sigma^{-1} (Y - \mu_Y)\right\}$$

$$Y = [Y(h_1), Y(h_2), ..., Y(h_n)]^T$$

$$\mu_Y = [\mu_{Y_1}, \mu_{Y_2}, ..., \mu_{Y_n}]$$

$$\Sigma = \sigma_Y^2 \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{bmatrix}$$

Where $Y$ is an any given borehole data vector of ln(UCS), $n$ is the number of data points, $\mu_Y$ is the mean value of $Y$, and $\Sigma$ is the variance matrix.

Equation (2) represents the likelihood function about $\theta$, and it can be transformed into logarithmic form as shown in equation (6). With equation (6), the autocorrelation distance can be obtained as shown in
Figure 5. Table 2 shows all autocorrelation distances for four sites, and it can be seen that the $\theta$ equals to 0, 0.95 m, 1.19 m, and 1.17 m for the four boreholes, respectively.

$$L(\theta) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln(\Sigma) - \frac{1}{2} (Y - \mu_Y)^\top \Sigma^{-1} (Y - \mu_Y)$$ (6)

| Borehole   | n  | mean | Variance | COV | $\theta$ |
|------------|----|------|----------|-----|----------|
| A          | 10 | 4.06 | 0.08     | 0.07| 0        |
| B          | 9  | 4.23 | 0.59     | 0.18| 0.95     |
| C          | 10 | 4.23 | 0.06     | 0.06| 1.19     |
| D          | 10 | 4.50 | 0.31     | 0.12| 1.17     |

Figure 5. An example of calculating $\theta$ of borehole C using equation (6)

3. Bayesian framework

3.1 Probability distribution

Generally, it is difficult to determine what type of distribution geotechnical materials belong to. Anyway, it has been proven that the multivariable normal probability density function (PDF) is an effective tool to describe the distribution characteristic of geotechnical materials. Therefore, the multivariable normal PDF is adopted in this study as follows:

$$P(Y|\mu_Y, \sigma_Y^2, \theta_Y) = N(Y|\mu_Y, C)$$

$$= \frac{1}{\sqrt{2\pi}^{\frac{n}{2}}(\sigma_Y^2)^{\frac{n}{2}}} \times \exp \left\{ -\frac{1}{2\sigma_Y^2} (Y - \mu_Y)^\top C^{-1} (Y - \mu_Y) \right\}$$ (7)

where $C$ represents the correlation matrix (see equation (8)), $\mu_Y$, $\sigma_Y^2$, are the statistical parameters inferred from the observed $Y$ values, $\theta_Y$ represents the autocorrelation distance of $Y$ along the vertical direction.

$$C = \Sigma./\sigma_Y^2$$ (8)

With the aid of Bayesian approach, the joint probability distribution of $\mu_Y$, $\sigma_Y^2$, and $\theta_Y$ can be expressed using equation (9):

$$P(\mu_Y, \sigma_Y^2, \theta_Y|Y) \propto P(Y|\mu_Y, \sigma_Y^2, \theta_Y)P(\mu_Y)P(\sigma_Y^2)P(\theta_Y)$$ (9)

where $P(\mu_Y)$, $P(\sigma_Y^2)$, and $P(\theta_Y)$ are the prior distributions of $\mu_Y$, $\sigma_Y^2$, and $\theta_Y$, respectively; $P(\mu_Y, \sigma_Y^2, \theta_Y|Y)$ is the posterior distribution that represents probability properties of the statistical parameters based on the observed data.
3.2 Markov chain Monte Carlo simulation (MCMCS)

The MCMCS is a numerical process which can be used for generating equivalent samples from an arbitrary PDF, and hence it is generally applied to evaluate the posterior distribution of parameters [2, 3, 9-11, 17]. In this study, the posterior distribution is a joint PDF (see equation 9) with respect to the three statistical parameters, namely \( \mu_y \), \( \sigma^2_y \), and \( \theta_y \). There are two commonly used algorithms for MCMCS, namely Metropolis-Hastings (MH) algorithm and Gibbs sampler (GS). For MH algorithm, the new state in a Markov chain is determined from the former state through the acceptance/rejection procedure. The basic idea of GS is to decompose the random parameters into several groups and to directly sample in each group sequentially without acceptance/rejection procedure. Which algorithm is suitable for analysis in this study depends on the form of posterior PDF.

In this study, the values of each statistical parameter are sequentially sampled from the conditional posterior distributions as follows:

\[
P(\mu_y | \sigma^2_y, \theta_y, Y) \propto P(Y | \mu_y, \sigma^2_y, \theta_y) \cdot P(\mu_y) \]
\[
P(\sigma^2_y | \mu_y, \theta_y, Y) \propto P(Y | \mu_y, \sigma^2_y, \theta_y) \cdot P(\sigma^2_y) \]
\[
P(\theta_y | \mu_y, \sigma^2_y, Y) \propto P(Y | \mu_y, \sigma^2_y, \theta_y) \cdot P(\theta_y) \]

(10)

Based on previous researches, when the posterior and prior distributions belong to the same distribution class, the parameter can be directly sampled, and such prior distribution is called the natural conjugate distribution. Since the \( \mu_y \) is assumed to follow the multivariable normal distribution, hence when the prior PDF of \( \mu_y \) is selected as normal distribution, the prior PDF of \( \mu_y \) is the natural conjugate PDF. The conditional posterior PDF of \( \mu_y \) is expressed as follows:

\[
P(\mu_y | \sigma^2_y, \theta_y, Y) \propto \frac{1}{\sqrt{2\pi \sigma^2_y}} \exp \left[ -\frac{1}{2\sigma^2_y} (Y - \mu_y)^T \Sigma^{-1} (Y - \mu_y) \right] \]

\[
\times \frac{1}{\sqrt{2\pi \sigma^2_{\mu_0}}} \exp \left[ -\frac{(\mu_y - \mu_{\mu_0})^2}{2\sigma^2_{\mu_0}} \right] \]

\[
\propto \exp \left[ -\frac{1}{2\sigma^2_y} (Y - \mu_y)^T \Sigma^{-1} (Y - \mu_y) - \frac{(\mu_y - \mu_{\mu_0})^2}{2\sigma^2_{\mu_0}} \right] \]

\[
\propto \exp \left[ -\frac{(\mu_y - \mu_{\mu_0})^2}{2\sigma^2_{\mu_0}} \right] \]

(11)

\[
\mu_{\mu} = \frac{2\mu_{\mu_0} \sigma^2_y + \sigma^2_{\mu_0} \sum_{i=1}^{n} \sum_{j=1}^{n} \zeta_{ij} (Y_i - Y_j)}{2\sigma^2_y + \sigma^2_{\mu_0} \sum_{i=1}^{n} \sum_{j=1}^{n} \zeta_{ij}} \]

\[
\sigma^2_{\mu} = \frac{1}{\sigma^2_{\mu_0} + \sum_{i=1}^{n} \sum_{j=1}^{n} \zeta_{ij}} \]

(12)

where \( \mu_{\mu_0} \) and \( \sigma^2_{\mu_0} \) denote the basic parameters, namely the mean and variance of \( \mu_y \), respectively; \( \zeta_{ij} \) denotes the elements of \( \Sigma^{-1} \); and \( Y_i \) and \( Y_j \) represent the \( Y \) at the depth \( h_i \) and \( h_j \).

When the inverse gamma distribution is selected as the prior distribution of \( \sigma^2_y \), this prior distribution is also the natural conjugate distribution, and the conditional distribution of \( \sigma^2_y \) is expressed as follows:
\[ P(\sigma^2_Y|\mu_Y, \theta_Y, Y) \propto \frac{1}{\sqrt{(2\pi)^n(\sigma^2_Y)^n}|C|} \exp\left[-\frac{1}{2\sigma^2_Y} (Y - \mu_Y)^\top C^{-1}(Y - \mu_Y)\right] \]

\[ \times (\sigma^2_Y)^{(n_{a+1})} \exp\left[-\frac{\beta_{a0}}{\sigma^2_Y}\right] \]

\[ \propto (\sigma^2_Y)^{(n + n_{\alpha} + n_{\beta} + 1)} \exp\left[-\frac{(Y - \mu_Y)^\top C^{-1}(Y - \mu_Y)}{2\sigma^2_Y}\right] \]

\[ \alpha_{\alpha} = \frac{n}{2} + a_{\alpha0} \]

\[ \beta_{\alpha} = \beta_{\alpha0} + \frac{1}{2} (Y - \mu_Y)^\top C^{-1}(Y - \mu_Y) \]

where \( a_{\alpha0} \) and \( \beta_{\alpha0} \) are the basic parameters, which denotes the parameters of inverse gamma distribution for \( \sigma^2_Y \).

There is not natural conjugate distribution for \( \theta_Y \), and the normal distribution is employed to describe the prior distribution of \( \theta_Y \). Therefore, the conditional posterior distribution of \( \theta_Y \) is expressed as follows:

\[ P(\theta_Y|\mu_Y, \sigma^2_Y, Y) \propto \frac{1}{\sqrt{(2\pi)^n(\sigma^2_Y)^n}|C|} \exp\left[-\frac{1}{2\sigma^2_Y} (Y - \mu_Y)^\top C^{-1}(Y - \mu_Y)\right] \]

\[ \times \frac{1}{\sqrt{2\pi\sigma^2_{\theta0}}} \exp\left[-\frac{(\theta_Y - \mu_{\theta0})^2}{2\sigma^2_{\theta0}}\right] \]

where \( \mu_{\theta0} \) and \( \sigma^2_{\theta0} \) are the basic parameters, namely the mean and variance of \( \theta_Y \), respectively.

Apparently, the statistical parameters \( \mu_Y \) and \( \sigma^2_Y \) can be directly sampled from the corresponding conditional posterior distributions using GS, while the statistical parameter \( \theta_Y \) should be sampled from its conditional posterior distribution using MH algorithm. Of course, the statistical parameters \( \mu_Y \) and \( \sigma^2_Y \) can also be sampled with the aid of MH algorithm, however, the acceptance/rejection procedure will bring much more computational costs.

### 3.3 Implementation procedure

According to the above analysis, the MCMCS of statistical parameters \( \mu_Y, \sigma^2_Y, \) and \( \theta_Y \) will be conducted by combing the GS and MH algorithm. The GS and MH iterates over the following steps:

1. Initializing the \((\mu_Y, \sigma^2_Y, \theta_Y)\) samples according to the observed \( Y \).
2. Drawing the \( \mu_Y^{(j)} \) sample from \( P(\mu_Y^{(j)}|\sigma^2_Y^{(j-1)}, \theta_Y^{(j-1)}, Y) \) using GS.
3. Drawing the \( \sigma^2_Y^{(j)} \) sample from \( P(\sigma^2_Y^{(j)}|\mu_Y^{(j)}, \theta_Y^{(j-1)}, Y) \) using GS.
4. Drawing the \( \theta_Y^{(j)} \) sample from \( P(\theta_Y^{(j)}|\mu_Y^{(j)}, \sigma^2_Y^{(j)}, Y) \) using MH algorithm.

In step (4), the candidate sample of \( \theta_Y^{(j)} \) is obtained from the selected proposal PDF \( f_i \), and the chance to accept the candidate sample \( \theta_Y^{(j)} \) as the \( j \)th sample \( \theta_Y^{(j)} \) depends on the “acceptance ratio”, \( r \), which is calculated by equation (16).
\[
    r = \frac{P\left(\theta_y^{(j)|Y}, \mu_y, \sigma_y^{(j)}, Y\right)}{P\left(\theta_y^{(j-1)|Y}, \mu_y, \sigma_y^{(j)}, Y\right)} \times \frac{f\left(\theta_y^{(j)|\theta_y^{(j)}}\right)}{f\left(\theta_y^{(j-1)|\theta_y^{(j)}}\right)}
\]

The sample \(\theta_y^{(j)}\) is determined by equation (17).

\[
    \theta_y^{(j)} = \begin{cases} 
        \theta_y^{(j-1)}, & \text{if } \text{rand()} < r \\
        \theta_y^{(j)}, & \text{Otherwise}
    \end{cases}
\]

(5) Cycling the steps (2) ~ (4) \(m\) times to obtain \(m\) samples for \((\mu_y, \sigma_y^2, \theta_y)\).

4. Results and discussion

4.1 Determination of basic parameters

Before the implementation, some basic parameters should be determined. Firstly, the parameters \(\mu_{\beta0}, \sigma_{\beta0}^2, \mu_{\var0}, \sigma_{\var0}^2, \mu_{\theta0}, \sigma_{\theta0}^2\) are obtained by analysing the database in Figure 2 (b), and the parameters \(\alpha_{\theta0}\) and \(\beta_{\theta0}\) in equation (13) and (14) are calculated by equation (18). Finally, the basic parameters from equation (11) to (15) are obtained as shown in Table 3, and it gives the values in each site and the overall values

\[
    \mu_{\var0} = \frac{\beta_{\theta0}}{\alpha_{\theta0} - 1} \quad \text{for } \alpha_{\theta0} > 1,
\]

\[
    \sigma_{\var0}^2 = \frac{\beta_{\theta0}^2}{\left(\alpha_{\theta0} - 1\right)\left(\alpha_{\theta0} - 2\right)} \quad \text{for } \alpha_{\theta0} > 2
\]

Where \(\mu_{\var0}\) and \(\sigma_{\var0}^2\) are the mean and variance of \(\sigma_y^2\).

| Site    | \(\mu_{\theta0}\) | \(\sigma_{\theta0}^2\) | \(\mu_{\var0}\) | \(\sigma_{\var0}^2\) | \(\alpha_{\theta0}\) | \(\beta_{\theta0}\) | \(\mu_{\theta0}\) | \(\sigma_{\theta0}^2\) |
|---------|------------------|-------------------|----------------|-------------------|-----------------|----------------|----------------|----------------|
| Over all| 4.10             | 0.22              | 0.24           | 0.07              | 2.85            | 0.44           | 0.80           | 2.09            |
| Site A  | 4.05             | 0.28              | 0.26           | 0.10              | 2.66            | 0.44           | 1.04           | 1.95            |
| Site B  | 4.07             | 0.16              | 0.24           | 0.07              | 2.91            | 0.47           | 1.02           | 5.74            |
| Site C  | 4.10             | 0.22              | 0.19           | 0.03              | 3.16            | 0.41           | 0.78           | 0.81            |
| Site D  | 4.20             | 0.18              | 0.26           | 0.06              | 3.06            | 0.53           | 0.48           | 0.30            |

4.2 Results based on the overall basic parameters

In this study, 12000 samples were drawn based the above algorithm, and the first 2000 samples are assumed as the samples in the burning period, hence the rest 10000 samples are considered to analyse the statistical uncertainty of \(Y\). Figure 6 and Figure 7 show the drawn equivalent samples of statistical parameters \(\mu_y, \sigma_y^2, \theta_y\) and it can be seen that the statistical uncertainty is evident. To intuitively reflect the uncertainty, the 0.05 quantile and 0.95 quantile are depicted in the above figures using the dash line, and the mean of equivalent samples is also provided in dot dash line, whose specific data are shown from Table 4 to Table 6. In Table 4, 5, and 6, COV of these characteristic values for four boreholes is also calculated to illustrate effect of specific site on the uncertainty, and besides, the differential value between 0.05 quantile and 0.95 quantile can quantify the magnitude of uncertainty. Apparently, the uncertainty of \(\mu_y\) and \(\theta_y\) is similar for the four boreholes, while the uncertainty of \(\sigma_y^2\) has significant differences among these four boreholes as shown in Figure 6 and 7. For the uncertainty of \(\sigma_y^2\), the borehole B has the biggest one as the value of \(\sigma_y^2,0.95,0.05\) is 0.33, the borehole A has the secondary one \((\sigma_y^2,0.95,0.05 = 0.18)\), and the borehole C and D have the lowest one \((\sigma_y^2,0.95,0.05 = \ldots\)
0.13) as shown in Table 5. Interestingly, even the measured autocorrelation distance is 0 for the borehole A, the drawn autocorrelation distance still can vary from about 0 m to about 6 m (see Figure 6 (e) and (f) and Figure 7 (e) and (f)), which is similar to the rest three boreholes.

According to COV, it can be found that the figures of equivalent samples is generally less than the measured ones except the COV of $\mu_Y, \sigma_Y$ (0.14 > 0.04), indicating that the effect of specific site on uncertainty does exist. It is worth noting that the basic parameters we used are the overall basic parameters, hence the only difference for these four sampling procedures is the input data and the initial $\sigma_Y^2$ and $\theta_Y$. Generally speaking, the initial data will not affect the sampling results especially for the results after the burning period. Therefore, it can be concluded that the statistical uncertainty of $\sigma_Y^2$ is mainly controlled by the input data, while the statistical uncertainty of $\mu_Y$ and $\theta_Y$ are mainly determined by the selected basic parameters.

Figure 6. Sampling results for the borehole A and B based on the overall basic parameters
Figure 7. Sampling results for the borehole C and D based on the overall basic parameters

Table 4. Characteristic value of equivalent $\mu_Y$ samples based on the overall basic parameters

| Site  | $\mu_{Y,0.05}$ | $\mu_{Y,0.95}$ | COV    |
|-------|----------------|----------------|--------|
| A     | 3.81           | 4.36           | 0.03   |
| B     | 3.77           | 4.48           | 0.02   |
| C     | 4.05           | 4.56           | 0.14   |
| D     | 4.05           | 4.56           | 0.02   |

Table 5. Characteristic value of equivalent $\sigma^2_Y$ samples based on the overall basic parameters

| Site  | $\sigma^2_{Y,0.05}$ | $\sigma^2_{Y,0.95}$ | $\sigma^2_{Y,0.95}-\sigma^2_{Y,0.05}$ | $\sigma^2_{Y,mean}$ | $\sigma^2_{Y,measured}$ | COV    |
|-------|---------------------|---------------------|---------------------------------------|---------------------|--------------------------|--------|
| A     | 0.08                | 0.35                | 0.26                                  | 0.18                | 0.08                     | 0.22   |
| B     | 0.11                | 0.44                | 0.33                                  | 0.23                | 0.59                     | 0.28   |
| C     | 0.06                | 0.23                | 0.17                                  | 0.13                | 0.06                     | 0.30   |
| D     | 0.06                | 0.23                | 0.17                                  | 0.13                | 0.31                     | 0.26   |


Table 6. Characteristic value of equivalent $\theta_Y$ samples based on the overall basic parameters

|                | Site A | Site B | Site C | Site D | COV |
|----------------|--------|--------|--------|--------|-----|
| $\theta_{Y,0.05}$ | 0.17   | 0.18   | 0.18   | 0.18   | 0.01|
| $\theta_{Y,0.95}$ | 3.45   | 3.49   | 3.39   | 3.38   | 0.01|
| $\theta_{Y,0.95-0.05}$ | 3.27   | 3.31   | 3.21   | 3.21   | 0.01|
| $\theta_Y$,mean  | 1.57   | 1.59   | 1.55   | 1.54   | 0.01|
| $\theta_Y$,measured | 0.00   | 0.95   | 1.19   | 1.17   | 0.59|

The relationship between variance and mean generally comprehensively reflects the statistical uncertainty of data, hence Figure 8 shows the variations of variance of $Y$ and the mean of $Y$, and it can be seen that there is no significant relationship between the variance and the mean, which is slightly similar to the results in Figure 3 (a). To describe the uncertainty intuitively, the confidence ellipse approach is applied under the significance level of 90%, which is denoted by 90% CE as shown in Figure 8. In this case, the uncertainty can be quantified by the area of 90% CE, and the greater area represents the more evident uncertainty. Based on the area of 90% CE, the statistical uncertainty of borehole B is the largest with the area of 0.0568, the borehole A second (0.0568), the borehole C and D lowest (0.0215), which is consistent to the inference from Figure 6 and 7. Apparently, the uncertainty of $\mu_Y$ is significantly greater than it of $\sigma^2_Y$ as the $\mu_Y$ varies in a relatively larger interval.

![Figure 8](image-url)

Figure 8. Variations of variance of sampled $Y$ with mean of sampled $Y$ based on the overall basic parameters

Figure 9 shows the relationship between $\sigma^2_Y$ as the $\theta_Y$, similarly, the correlation between $\sigma^2_Y$ as the $\theta_Y$ is weak. And it also can be observed that the statistical uncertainty of borehole B is greatest, the borehole at Site A second, and the borehole C and D lowest. In the view of scope of variation, the uncertainty of $\theta_Y$ is greater than that of $\sigma^2_Y$. 

![Figure 9](image-url)
4.3 Results based on the basic parameters of each site

According to the analysis above, it is found that the basic parameters play an important role in the evaluation of statistical uncertainty, hence the selection of basic parameters should be cautious. In geotechnical engineering, the effect of site on statistical characteristics of geo-materials properties is very significant, and it has been proven by previous relevant researches [10, 18-21]. As a result, this subsection will mainly focus on the statistical uncertainty of four boreholes under the basic parameters of each site.

Figure 10 and 11 show the sampling results for the four boreholes, and it also can be seen that the statistical uncertainty is evident. 0.05 quantile, 0.95 quantile, and mean are also plotted in Figure 10 and 11, and the specific data are given in Table 7, 8 and 9. Similarly, the difference of statistical uncertainty of $\mu_Y$ is not obvious under the basic parameters of each site as the COV of $\mu_Y$, $0.95 - \mu_Y$, $0.05$ is only 0.17. In view of the pre-mentioned conclusion, the statistical uncertainty of $\mu_Y$ and $\theta_Y$ is strongly related to the selection of basic parameters. From Table 3, it can be observed that the basic parameters $\mu_0$ and $\sigma^2_{\mu_0}$ of each site are close to each other, which may be the reason for that.

Differing from the results under the overall basic parameters, the borehole B, C and D have the comparative statistical uncertainty about $\sigma^2_{\mu_0}$ under the basic parameters of each site with the $\sigma^2_{Y,0.95} - \sigma^2_{Y,0.05}$ of around 0.40, and the borehole A shows relatively less uncertainty ($\sigma^2_{Y,0.95} - \sigma^2_{Y,0.05} = 0.28$), which demonstrates the effects of basic parameters (or site). Furthermore, the COV of $\sigma^2_{Y,0.95} - \sigma^2_{Y,0.05}$ is 0.15 (see Table 8) dramatically less than it (0.30) under the overall basic parameters (see Table 5), which reveals the effects of basic parameters on evaluation of statistical uncertainty.

The statistical uncertainty of $\theta_Y$ has the greatest difference under different basic parameters as shown in Figure 10 (e) and (f) and Figure 11 (e) and (f). Values of $\theta_{Y,0.95} - \theta_{Y,0.05}$ for Borehole A and B are 3.28 and 5.13, respectively, while they are 1.36 and 1.28 for Borehole C and D, respectively. Therefore, the statistical uncertainty of $\theta_Y$ of the borehole B is the most significant, the borehole A second, and the boreholes C and D the lowest, which is significantly different from the pattern under the overall basic parameters, further indicating the effects of basic parameters (or specific site).

Figure 12 shows the relationship between the mean of $Y$, and the variance of $Y$, and it can be seen that there is no evident correlation relationship between them. Similarly, the 90% CE is also employed for
quantifying the statistical uncertainty intuitively. With the aid of area of 90% CE, it can be observed that borehole B has the largest statistical uncertainty with the area of 0.0792, and borehole B has the lowest uncertainty as the area is 0.0381. On the whole, the statistical uncertainty under the basic parameters of each site is generally greater than that under the overall parameters. In addition, the difference of statistical uncertainty between \( \sigma_Y \) and \( \mu_Y \) become smaller in terms of the scope of interval under the basic parameters of each site.

Figure 13 shows the relationship between the variance of \( Y \) and the autocorrelation distance, and there is also no distinguish correlation relationship. Obviously, borehole B has the largest uncertainty, the borehole A second, and the boreholes C and D the lowest, which is mainly attributed to the reduction of the autocorrelation distance uncertainty of at the boreholes C and D. From Table 3, it is clear that the basic parameters \( \mu_0 \) and \( \sigma_{\theta_0} \) of the boreholes C and D are less than those of other boreholes and the overall basic parameters, which further illustrates the effects of sites.

Figure 10. Sampling results for the borehole A and B based on the basic parameters of each site
Figure 11. Sampling results for the borehole C and D based on the basic parameters of each site

Table 7. Characteristic value of equivalent $\mu_Y$ samples based on the basic parameters of each site

| Site   | $\mu_Y_{0.05}$ | $\mu_Y_{0.95}$ | $\mu_Y_{0.95} - \mu_Y_{0.05}$ | $\mu_Y$ mean | $\mu_Y$ measured | COV |
|--------|----------------|----------------|-------------------------------|--------------|------------------|-----|
| Site A | 3.79           | 4.36           | 0.57                          | 4.07         | 4.06             | 0.06|
| Site B | 3.71           | 4.48           | 0.77                          | 4.12         | 4.23             | 0.03|
| Site C | 4.09           | 4.68           | 0.59                          | 4.40         | 4.23             | 0.17|
| Site D | 4.25           | 4.75           | 0.49                          | 4.51         | 4.50             | 0.04|

Table 8. Characteristic value of equivalent $\sigma_Y^2$ samples based on the basic parameters of each site

| Site   | $\sigma_Y^2_{0.05}$ | $\sigma_Y^2_{0.95}$ | $\sigma_Y^2_{0.95} - \sigma_Y^2_{0.05}$ | $\sigma_Y^2$ mean | $\sigma_Y^2$ measured | COV |
|--------|---------------------|---------------------|------------------------------------------|-------------------|------------------------|-----|
| Site A | 0.08                | 0.36                | 0.28                                      | 0.19              | 0.08                   | 0.27|
| Site B | 0.12                | 0.53                | 0.36                                      | 0.27              | 0.59                   | 0.27|
| Site C | 0.17                | 0.41                | 0.24                                      | 0.33              | 0.06                   | 0.17|
| Site D | 0.17                | 0.40                | 0.23                                      | 0.32              | 0.31                   | 0.15|

**Notes:**

- $\mu_Y$: Mean of equivalent $Y$ samples
- $\sigma_Y^2$: Variance of equivalent $Y$ samples
- $\mu_Y_{0.05}$, $\mu_Y_{0.95}$: Lower and upper quantiles of $\mu_Y$
- $\mu_Y_{0.95} - \mu_Y_{0.05}$: Range of $\mu_Y$
- $\mu_Y$ mean, $\mu_Y$ measured: Mean and measured values of $\mu_Y$
- COV: Coefficient of variation

- $\sigma_Y^2_{0.05}$, $\sigma_Y^2_{0.95}$: Lower and upper quantiles of $\sigma_Y^2$
- $\sigma_Y^2_{0.95} - \sigma_Y^2_{0.05}$: Range of $\sigma_Y^2$
- $\sigma_Y^2$ mean, $\sigma_Y^2$ measured: Mean and measured values of $\sigma_Y^2$
- COV: Coefficient of variation
Table 9. Characteristic value of equivalent $\theta_Y$ samples based on the basic parameters of each site

|        | Site A | Site B | Site C | Site D | COV |
|--------|--------|--------|--------|--------|-----|
| $\theta_{Y,0.05}$ | 0.21   | 0.24   | 0.10   | 0.05   | 0.51|
| $\theta_{Y,0.95}$ | 3.49   | 5.37   | 1.46   | 1.33   | 0.57|
| $\theta_{Y,0.95-0.05}$ | 3.28   | 5.13   | 1.36   | 1.28   | 0.57|
| $\theta_{Y,\text{mean}}$ | 1.60   | 2.35   | 0.69   | 0.56   | 0.56|
| $\theta_{Y,\text{measured}}$ | 0.00   | 0.95   | 1.19   | 1.17   | 0.59|

Figure 12. Variations of variance of sampled $Y$ with mean of sampled $Y$ based on the basic parameters of each site

4.4 Discussions about the results

Firstly, the statistical parameters $\mu_Y$, $\sigma^2_Y$, $\theta_Y$ have significant uncertainty for the sampling in a borehole, even if the measure parameter is zero, it also can be very large by considering the uncertainty, which may be due to two reasons: (1) the limited number of samples in a borehole; (2) the measurement error. With the advance of instrumentations, the measurement error can be negligible, hence the limited number of samples may be the main reason resulting in the significant statistical uncertainty.

Then, according to comparison between the results in subsection 4.2 and 4.3, it can be found that the basic parameters can extensively influence the quantification of statistical uncertainty, indicating that the selection of basic parameters should be cautious. Based on the previous relevant researches, the basic parameters of site-specific should be the first choice for the quantification of uncertainty.

Lastly, the measured autocorrelation distance is 0 in borehole A, which may beyond the common sense. There may be two reasons for explanation of this phenomenon: (1) the result is correct namely the UCSs are totally unrelated to each other along the borehole; (2) the actual autocorrelation distance may be very large, while the borehole depth is far less than the actual autocorrelation distance, when the standard deviation is relatively small, the variation of UCSs along borehole will not be visible so that the UCS looks unrelated to each other. Therefore, to figure out this issue, it still needs the advance of theory and more detailed field investigation materials.
5 Conclusions
With the aid of site investigation report, this paper quantified the statistical uncertainty of UCS using Bayesian-based Markov chain Monte Carlo simulation. Since UCS is non-negative, the natural logarithm transformation was applied for UCS. The main conclusions arrived at in this study include:

(1) The statistical parameters $\mu_Y$, $\sigma^2_Y$, $\theta_Y$ have considerable uncertainty, and the quantification of these uncertainties to some extent relies on the selection of basic parameters.

(2) The uncertainties of $\sigma^2_Y$, $\theta_Y$ are more sensitive to the selection of basic parameters than the uncertainty of $\mu_Y$.

(3) The correlation relationship among $\mu_Y$, $\sigma^2_Y$, and $\theta_Y$ is not evident, while it reflects that $\sigma^2_Y$ has the minimum uncertainty in terms of the scope of uncertainty interval.

References
[1] Zhang W G, Zhang R H, Han L and Goh A T C 2019 Undergr. Sp., 4 98-108.
[2] Wang Y and Aladejare A E 2015 Int. J. Rock Mech. Min. Sci., 75 73-81.
[3] Wang Y and Aladejare A E 2016 Int. J. Rock Mech. Min. Sci., 85 10-9.
[4] Phoon K K and Kulhawy F H 1999 Can. Geotech. J., 36 612-24.
[5] Phoon K K and Kulhawy F H 1999 Can. Geotech. J., 36 625-39.
[6] Zhang W G, Han L, Gu X, Wang L, Chen F Y and Liu H L 2020 Undergr. Sp.
[7] Zeng P, Senent S and Jimenez R 2016 J. Comput. Civil. Eng., 30.
[8] Lü Q, Xiao Z, Zheng J and Shang Y 2018 Geosci. Front., 9 1619-29.
[9] Wang Y and Cao Z 2013 Eng. Geol., 159 106-18.
[10] Ching J Y and Phoon K K 2017 J. Eng. Mech., 143 04017028.
[11] Namikawa T 2019 Soils Found., 59 1228-40.
[12] Han L, Zhang W G, Wu C Z and Goh A T C 2018 Proc. Int. Conf. the 6th International Symposium on Reliability Engineering and Risk Management (6ISRERM), Singapore, 2018, pp. 249-254.
[13] Honjo Y and Kazumba S 2002 Proc. Int. Symp. 47th Geotechnical Engineering Symp, Japanese Geotechnical Society, Tokyo, Japan, 2002, pp. 279–286.
[14] Sari M and Karpuz C 2006 Int. J. Rock Mech. Min. Sci., 43 328-35.
[15] Martin C D ans Christiansson R 2009 *Int. J. Rock Mech. Min. Sci.*, 46 219-228.
[16] Cao Z J and Wang Y 2014 *Struct. Saf.*, 49 10-7.
[17] Beck J L and Au S K 2002 *J. Eng. Mech.*, 128 380-91.
[18] Ching J Y, Phoon K K, Beck J L and Huang Y 2017 *Asce-Asme Journal of Risk and Uncertainty in Engineering Systems Part a-Civil Engineering*, 3.
[19] Ching J Y and Phoon K K 2019 *J. Eng. Mech.*, 145 04018126.
[20] Ching J Y, Phoon K K 2019 *Asce-Asme Journal of Risk and Uncertainty in Engineering Systems Part a-Civil Engineering*, 6.
[21] Ching J Y, Phoon K K, Khan Z, Zhang D and Huang H 2020 *Comput. Geotech.*, 124 103623.