Comparison of different calculation methods for estimating scale of fluctuation of design soil properties based on indirect measurement data

Mi Tian1,2*, Xiaotao Sheng3, Lihua Li1, Fan Zhang1, Henglin Xiao1

1School of Civil Engineering, Architecture and Environment, Hubei University of Technology, Wuhan, Hubei 430068, China
2Key Laboratory of Rock Mechanics in Hydraulic Structural Engineering of Ministry of Education, Wuhan University, Wuhan, Hubei 430072, China
3Key Laboratory of Geotechnical Mechanics and Engineering of Ministry of Water Resources, Yangtze River Scientific Research Institute, Wuhan, Hubei 430010, China
*Corresponding author’s e-mail: mitian0525@gmail.com

Abstract. Determination of scale of fluctuation (SOF) is a necessary prerequisite for describing the inherent spatial variability (ISV) of design soil properties (i.e., effective friction angle $\phi'$) using a stationary random field in geotechnical reliability analysis. It is very important to select an appropriate method to enable an accurate estimate of SOF. However, it has been a challenging task due to the fact that the true SOF is actually unknown. In addition, when the number of direct test data of design soil properties are too limited to generate meaningful statistics, the design soil properties are indirectly estimated through transformation models which involve transformation uncertainty. This paper investigates the validity and accuracy of three commonly-used calculation methods (e.g., Bayesian approaches, mean average method, and fitting sample autocorrelation function (ACF) method) for estimating SOF of $\phi'$ using indirect cone penetration test (CPT) data. It is found that Bayesian approaches provide more accurate estimates of $\lambda$ of $\phi'$ because it takes the transformation uncertainty from empirical regression models into reasonable consideration. The mean average method and fitting sample ACF method underestimate $\lambda$ of $\phi'$.

1. Introduction
The natural soil materials inherently vary from one site to another, which is known as inherent spatial variability (ISV) (e.g., Vanmarcke, 1977; Phoon and Kulhawy, 1999a and 199b). It can be probabilistically described using a stationary random field in geotechnical probabilistic analysis (e.g., Vanmarcke, 2010). Accurate determination of scale of fluctuation (SOF) is a necessary prerequisite for characterizing ISV of design soil properties (i.e., effective friction angle $\phi'$) using a stationary random field (e.g., Vanmarcke, 1977; Fenton, 1999).

Various kinds of calculation methods have been used in literature to estimate SOF of design soil properties, such as Bayesian approaches (e.g., Wang et al., 2010; Cao and Wang, 2013; Tian et al., 2016), mean average method (e.g., Vanmarcke, 1977), and fitting the sample autocorrelation function (ACF) (e.g., Jaksa 1995; Stuedlein et al., 2012). To enable an accurate estimate of SOF, an appropriate method should be used. However, the true SOF of design soil properties are actually unknown. The accuracy of these commonly-used methods is difficult to be determined.
In addition, it is not a trivial task to accurately estimate SOF of design soil properties due to various uncertainties (e.g., transformation uncertainty) existing in geotechnical practice and the fact that the number of direct site-specific data is often too sparse to accurately estimate SOF of the design soil properties. Considering that, for example, the sand effective friction angle $\varphi'$ can be directly obtained from laboratory triaxial tests. The number of test results obtained from laboratory triaxial tests is very limited because a large number of laboratory triaxial tests are costly and time-consuming. On the other hand, $\varphi'$ can be indirectly measured using fast and economical in-situ tests (e.g., cone penetration test (CPT)) through empirical regression functions (e.g., the empirical regression between normalized cone tip resistance $q$ from CPT and $\varphi'$) (e.g., Mayne et al., 2002). Empirical regression functions are often obtained from empirical data fitting, and hence they are associated with some uncertainties, namely “transformation uncertainty” (e.g., Wang et al., 2016). Transformation uncertainties can be propagated to geotechnical parameters which shall be rationally considered.

This paper explores the validity and accuracy of different methods for estimating SOF of $\varphi'$ using CPT data. It starts with random field model of sand effective friction angle and regression between cone tip resistance and the sand effective friction angle. Then, three commonly-used calculation methods for estimating SOF are presented. Finally, the validity and accuracy of these three calculation methods are explored and illustrated using simulated CPT data.

2. Random field model of sand effective friction angle

ISV of sand effective friction angle can be modeled as the sum of a spatially varying trend function $x(D)$ and a fluctuating component $w(D)$, as shown in Fig. 1, which is given by (e.g., Vanmarcke 1977; DeGroot and Baecher, 1993)

$$\varphi'(D) = x(D) + w(D)$$

(1)

Figure 1. Inherent spatial variability of soil properties

in which $\varphi'$ is the in-situ soil property, and $D$ is the depth. A one-dimensional stationary normal random field is used to probabilistically characterize ISV of $\varphi'$ within a statistically homogenous sand layer in this study. $\varphi'(D)$ is a normal random variable with a mean $\mu$ and standard deviation $\sigma$. The spatial correlation between values of $\varphi'(D)$ at different depths is characterized by the scale of fluctuation $\lambda$ and a correlation function. The correlation function is assumed as single exponential correlation function.

$$\rho[\varphi'(D_i), \varphi'(D_j)] = \exp\left(-\frac{2|D_i - D_j|}{\lambda}\right)$$

(2)

in which $\rho[\varphi'(D_i), \varphi'(D_j)]$ represents the correlation coefficient between $\varphi'(D_i)$ and $\varphi'(D_j)$ at respective depths of $D_i$ and $D_j$. The SOF is a separation distance, within which the soil property shows a relatively strong correlation from point to point. It was first proposed by Vanmarcke (1977), which is defined as
\[ \lambda = \lim_{h \to 0} \left( 1 - \frac{\tau}{h} \right) \rho(\tau) \, d\tau = 2 \int_{0}^{\infty} \rho(\tau) \, d\tau \]  

(3)

in which \( \rho(\tau) \) represents the correlation coefficient; \( h \) is the spatial distance; \( \tau \) is the relative distance between any two points in space.

Let \( \phi' = [\phi'(D_1), \phi'(D_2), \ldots, \phi'(D_n)]^T \) denote a vector of \( \phi'(D) \) at different depths. It is a Gaussian vector that can be written as:

\[ \phi'(D) = \mu_1 + \sigma \Sigma^T Z \]  

(4)

in which \( \mu_1 = [1, \ldots, 1]^T \), \( Z = [Z_1, \ldots, Z_n]^T \) is a standard Gaussian vector with independent components, and \( \Sigma \) is an \( n \)-by-\( n \) upper-triangular matrix obtained by Cholesky decomposition of the correlation matrix \( R \) of \( \phi'(D) \). The \((i, j)\) entry of \( R \) is given by the correlation function (i.e., Eq. (2)). Generally, the \( \phi' \) can be estimated from in-situ cone penetration tests by a semi-log empirical regression model (e.g., Kulhawy and Mayne 1990; Wang et al., 2010)

\[ \xi = \ln q = a \bar{q} + b + \sigma \varepsilon \]  

(5)

where \( \xi \) is measured cone tip resistance data in a log scale; \( q = (\bar{q} / p_a) / (\sigma_{\bar{q}} / p_0)^{0.5} \) = normalized cone tip resistance; \( \bar{q} \) = cone resistance measured at the tip of the cone; \( \sigma_{\bar{q}} \) and \( p_a \) = vertical effective stress and standard atmospheric pressure (i.e., 100kPa), respectively; \( a = 0.209; b = -3.684; \sigma = 0.586 \) (e.g., Kulhawy and Mayne, 1990); and \( \varepsilon \) = a standard Gaussian random variable that represents the transformation uncertainty associated with the regression equation.

3. Calculation methods for scale of fluctuation

Accurate estimation of SOF is the key parameter to define the correlation function and explicitly model the ISV of \( \phi' \) using Eq. (4). The SOF can be estimated from a number of direct observations. The most widely used methods for estimating SOF are Bayesian approaches, mean average method, and fitting sample autocorrelation function (ACF).

3.1. Bayesian approaches

Using Bayes’ Theorem (e.g., Ang and Tang, 2007), the posterior distribution of random field parameters of \( \phi' \), i.e., \( X = [\mu, \sigma, \lambda] \), for CPT data \( \hat{\xi} \) is written as:

\[ P(X | \hat{\xi}) = K^{-1} P(\hat{\xi} | X) P(X) \]  

(6)

in which \( K \) is a normalizing constant that is independent of \( X; P(\hat{\xi} | X) \) is the likelihood function for a given \( X \); and \( P(X) \) is the prior distribution of \( X \). The prior distribution is assumed to be a joint uniform distribution. The likelihood function is given by a joint normal distribution:

\[ P(\hat{\xi} | X) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2} \left[ \hat{\xi} - (a\mu + b) \right]^T C^{-1} \left[ \hat{\xi} - (a\mu + b) \right] \right\} \]  

(7)

Because of the page limit, details of the formulation and discussions on \( \hat{\xi} \) are referred to Wang et al. (2010) and Tian et al. (2016). The prior distribution and likelihood function are substituted into Eq. (6) to obtain the posterior distribution \( P(X | \hat{\xi}) \). To bypass the computational complexity in Eq. (6) (e.g., Zhang et al., 2009), Metropolis-Hastings algorithm is used to generate \( N_p \) random samples of \( X \) from Eq. (6) to describe the posterior distribution. Samples of \( X \) are subsequently used to obtain posterior knowledge on \( X \) through conventional statistical analyses.

3.2. Mean average method

Mean average method is a simple but approximate method of determining the SOF given by Vanmarcke (1977). It is derived based on the assumption that the correlation function follows the squared exponential decay model. The SOF is approximated as (e.g., Phoon and Kulhawy 1999a)
in which $d$ is the average distance between intersections of the fluctuating property $w(D)$ and its trend function $x(D)$.

### 3.3. Fitting sample autocorrelation function (ACF) method

SOF is estimated through calculation of the sample autocorrelation function (ACF) using the method of moments (e.g., Uzielli et al., 2005; Lloret-Cabot et al., 2014)

$$\rho(\tau=k\Delta D) \approx \frac{\sum_{i=1}^{n-k} \phi(D_i)\phi(D_{i+k})}{\hat{\sigma}^2(n-k-1)}$$

in which, $k=1, 2, \ldots, n$. $n$ is the number of measurement data of $\phi'$; $\Delta D$ is the sampling interval; $\hat{\sigma}^2$ is the standard deviation of $\phi'$ estimated from the measurement data. Subsequently, plausible theoretical autocorrelation function are fitted to the sample ACF (i.e., Eq. (9)). Four theoretical autocorrelation function are considered, such as single exponential correlation function (SECF), binary noise correlation function (BNCF), second-order Markov correlation function (SMCF), and squared exponential correlation function (SQECF). To increase the reliability of the estimated SOF, the theoretical autocorrelation functions are fitted only to the initial part of the sample ACF exceeding Bartlett’s limits $n_0 = 1.96\sqrt{n}$. The SOF is determined from the theoretical autocorrelation function with the maximum coefficient of determination $R^2$.

### 4. Comparison of different calculation methods

The validity and accuracy of different calculation methods for estimating SOF of $\phi'$ are explored in this section using simulated CPT data. Consider, for example a sampling interval of 0.05 m. The random field parameters of $\phi'$ are predefined as $\mu = 30.0^\circ$, $\sigma = 3.0^\circ$, and $\lambda = 0.5$ m. In this section, a relatively samll sampling interval (i.e., $\lambda/10$) is adopted to ensure that the effect of sampling interval is relatively minor. For the single exponential correlation function, 10 sets of $\zeta$ are simulated for a sampling depth $D = 50$ m (i.e., 100 $\lambda$). A relatively sufficient simulated CPT data are adopted to ensure that the CPT data contains the true SOF of $\phi'$ used in the simulation.

The Bayesian approaches, mean average method and fitting sample ACF method are applied to estimate the SOF of $\phi'$ using the 10 sets of simulated CPT data. A joint uniform prior distribution with a $\mu$ value ranging from 20$^\circ$ to 40$^\circ$, a $\sigma$ value ranging from 1$^\circ$ to 6$^\circ$, and a $\lambda$ value ranging from 0 m to 6 m is used in the Bayesian approaches (e.g., Phoon and Kulhawy, 1999a). 100,000 random samples of $\mu$, $\sigma$, and $\lambda$ are generated from the posterior distribution. For each set of CPT data, $\phi'$ is directly calculated from the regression function (i.e., Eq. (5)). Then, the mean average method and fitting sample ACF method are performed on the measured data of $\phi'$ to determine SOF of $\phi'$. Fig. 2 shows the results of SOF of $\phi'$ estimated from the three commonly used methods. For the 10 sets of CPT data, the $\lambda$ values of $\phi'$ obtained from the Bayesian approaches (i.e., the open squares) approximate the true value (i.e., 0.5 m), indicating that Bayesian approaches properly identify $\lambda$ of $\phi'$ using indirect CPT data. However, for the 10 sets of CPT data, the $\lambda$ values of $\phi'$ estimated from the mean average method (i.e., open circles in Fig. 2) are significantly smaller than the true value (i.e., 0.5 m). The $\lambda$ values obtained from the fitting sample ACF method (i.e., open triangles in Fig. 2) are also smaller than its true value.

Fig. 3(a)-(b) show the results of mean and standard deviation of $\phi'$ estimated from the three methods. Fig. 3(a) shows that all the three methods can obtain the accurate estimates of mean of $\phi'$. Fig. 3(b) shows that the $\sigma$ values (i.e., open squares in Fig. 3(b)) estimated from Bayesian approaches are close to the true value. However, the standard deviation values estimated from the mean average method and fitting the sample ACF method are obviously larger than the true value. The mean average method and fitting sample ACF method are performed on the measurements of $\phi'$ to determine the SOF of $\phi'$. The variability in the $\phi'$ values directly estimated CPT data using the empirical regression
function (i.e., Eq. (5)) contains not only the ISV of \( \phi' \) but also the transformation uncertainty, and it is, therefore, greater that the actual ISV of \( \phi' \) prescribed in simulation, which subsequently leads to overestimation in \( \sigma \) and underestimation in \( \lambda \).

![Figure 2](image_url)  
**Figure 2.** Estimates of scale of fluctuation of \( \phi' \) using the three commonly used methods

![Figure 3](image_url)  
**Figure 3.** Estimates of mean and standard deviation of \( \phi' \) using the three calculation methods

Table 1 summarizes the estimates of SOF of \( \phi' \) and the difference between the estimates and true values. It is shown that the estimated \( \lambda \) values (i.e., \( \lambda' = 0.5 \text{ m} \)) of \( \phi' \) using Bayesian approaches are the same with the true values. There is no difference between the estimated \( \lambda \) values and true value. However, there is a great difference (i.e., 0.4 m) between the \( \lambda \) values (i.e., \( \lambda' = 0.1 \text{ m} \)) estimated from the mean average method and true value (i.e., \( \lambda = 0.5 \text{ m} \)). The difference between the \( \lambda \) values (i.e., \( \lambda' = 0.3 \text{ m} \)) by fitting the sample ACF method and true value is 0.2 m. Compared with mean average method and fitting the sample ACF method, the Bayesian approaches can provide more accurate estimates of scale of fluctuation of \( \phi' \) using indirect CPT data and rationally consider various uncertainties (e.g., transformation uncertainty).

| Calculation methods | \( M_1 \): Bayesian approaches | \( M_2 \): mean average method | \( M_3 \): fitting sample ACF method |
|---------------------|-------------------------------|-------------------------------|----------------------------------|
| **Estimates of the SOF \( \lambda \) (m)** | 0.5                           | 0.1                           | 0.3                              |
| **Difference (m)**  | 0.0                           | 0.4                           | 0.2                              |

5. Summary and conclusions
This paper studied the validity and accuracy of different calculation methods for estimating scale of fluctuation of effective friction angle \( \phi' \) with simulated CPT data. The three commonly used
calculation methods, such as Bayesian approaches, mean average method and fitting sample ACF method were presented. Simulated CPT data were generated from predefined random field parameters and correlation function of $\phi'$, with sufficient data quantity. The three calculation method were used to estimate the SOF of $\phi'$ using simulated CPT data. It was shown that, compared with mean average method and fitting the sample ACF method, Bayesian approaches can provide more accurate estimates of SOF of $\phi'$ using indirect CPT data because it can rationally deal with various uncertainties (e.g., transformation uncertainty). The mean average method and fitting the sample ACF method usually underestimate the scale of fluctuation of $\phi'$. When determining the scale of fluctuation of design soil properties based on indirect measurement data, the Bayesian method is preferred.

References

[1] Ang, A.H.-S., Tang, W.H. (2007) Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering. John Wiley & Sons, New York.
[2] Cao, Z.J., Wang, Y. (2013) Bayesian approach for probabilistic site characterization using cone penetration tests. Journal of Geotechnical and Geoenvironmental Engineering, 139, 267-276.
[3] DeGroot, D.J., Baecher, G.B. (1993) Estimating autocovariance of in situ soil properties. Journal of Geotechnical Engineering (ASCE), 129(1): 147-166.
[4] Jaksa, M.B. (1995) The influence of spatial variability on the geotechnical design properties of a stiff overconsolidated clay. Ph.D. Thesis, Faculty of Engineering, The University of Adelaide.
[5] Kulhawy, F.H., Mayne, P.W. (1990) Manual on estimating soil properties for foundation design. Report EL 6800, Electric Power Research Inst., Palo Alto.
[6] Lloret-Cabot, M., Fenton, G.A., Hicks, M.A. (2014) On the estimation of scale of fluctuation in geostatistics. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 8(2): 129-140.
[7] Mayne, P.W., Christopher, B.R., DeJong, J. (2002) Subsurface Investigations-Geotechnical Site Characterization, No. FHWA NHI-01-031, Federal Highway Administration, U. S. Department of Transportation, Washington D. C.
[8] Phoon, K.K., Kulhawy, F.H. (1999a) Characterization of geotechnical variability. Canadian Geotechnical Journal, 36 (4), 612-624.
[9] Phoon, K.K., Kulhawy, F.H. (1999b) Evaluation of geotechnical property variability. Canadian Geotechnical Journal, 36 (4), 625-639.
[10] Stuedlein, A.W., Kramer, S.L., Arduino, P., et al. (2012) Geotechnical characterization and random field modeling of desiccated clay. Journal of Geotechnical and Geoenvironmental Engineering, 138(11): 1301-1313.
[11] Tian, M., Li, D.Q., Cao, Z.J., Phoon, K.K., Wang, Y. (2016) Bayesian identification of random field model using indirect test data. Engineering Geology, 210: 197-211.
[12] Uzielli, M., Vannucchi, G., Phoon, K.K. (2005) Random field characterization of stress normalised cone penetration testing parameters. Geotechnique, 55(1), 3-20.
[13] Vanmarcke, E.H. (1977) Probabilistic modeling of soil profiles. Journal of Geotechnical Engineering, 103(11), 1127-1246.
[14] Vanmarcke, E.H. (2010) Random fields: analysis and synthesis. Revised and expanded new edition. Beijing: World Scientific Publishing.
[15] Wang, Y., Au, S.K., Cao, Z.J. (2010) Bayesian approach for probabilistic characterization of sand friction angles. Engineering Geology, 114, 354-363.
[16] Wang, Y., Cao, Z.J., Li, D.Q. (2016) Bayesian perspective on geotechnical variability and site characterization[J]. Engineering Geology, 203: 117-125.
[17] Zhang, J., Zhang, L.M., Tang, W.H. (2009) Bayesian framework for characterizing geotechnical model uncertainty. Journal of Geotechnical and Geoenvironmental Engineering, 135(7), 932-940.