

Revisiting the Relationship between Scale of Fluctuation and Mean Cross Distance

Yu-Xin Zhu

MPhil Student, State Key Laboratory of Water Resources and Hydropower Engineering Science, Institute of Engineering Risk and Disaster Prevention, Wuhan University, Wuhan, China; State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu, China

Shuo Zheng

PhD Student, State Key Laboratory of Water Resources and Hydropower Engineering Science, Institute of Engineering Risk and Disaster Prevention, Wuhan, China

Zi-Jun Cao

Associate Professor, State Key Laboratory of Water Resources and Hydropower Engineering Science, Institute of Engineering Risk and Disaster Prevention, Wuhan University, Wuhan, China; State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu, China

Dian-Qing Li

Professor, State Key Laboratory of Water Resources and Hydropower Engineering Science, Institute of Engineering Risk and Disaster Prevention, Wuhan, China

ABSTRACT: Estimating scale of fluctuation is an intriguing issue, for which several methods have been developed, such as simple estimators (e.g., $0.8\bar{d}$ -estimator) based on the mean cross distance \bar{d} of a soil property profile, sample autocorrelation function method, maximum likelihood method, Bayesian method, etc. Among these methods, the $0.8\bar{d}$ -estimator is the simplest one and can be readily used by geotechnical practitioners whose training in probability theory and statistics is usually limited. It, however, shall be noted that the $0.8\bar{d}$ -estimator was derived from the normal random field with squared exponential correlation function, which is largely ignored in its practical applications. Effects of the distribution type (e.g., normal or lognormal) and correlation function on the performance of the $0.8\bar{d}$ -estimator remain unexplored and, hence, unknown to geotechnical practitioners, which potentially leads to misuse of the simple relationship. This paper aims to highlight the theoretical assumptions underlying the $0.8\bar{d}$ -estimator and to, systematically, explore the effects of these theoretical assumptions on its performance (i.e., unbiasedness and variability). It is found that the $0.8\bar{d}$ -estimator provides reasonably unbiased estimation of scale of fluctuation for the normal random field with squared exponential correlation function when there are, at least, two sampling data within a distance of scale of fluctuation. Whereas, results from the $0.8\bar{d}$ -estimator for other cases violating the assumptions are biased, and may lead to a significant underestimation of scale of fluctuation. It is also found that the variability of the $0.8\bar{d}$ -estimator increases as the sampling length decreases.

1. INTRODUCTION

Due to the variability of source materials, weathering patterns, transportation agents, stress and formation processes, etc. (Mitchell and Soga, 2005), the soil properties in situ exhibit a certain of heterogeneity inherently, which is known as

the inherent spatial variability (Phoon and Kulhawy, 1999). The inherent spatial variability is one primary source of geotechnical uncertainties and can be explicitly modeled by a random field model for a lack of site investigation data. Scale of fluctuation, λ , is an essential element in correlation function of

random field model (Vanmarcke, 1977), within which variations of geotechnical properties are considered strongly correlated. Proper estimation of λ is prerequisite for site characterization and the subsequent statistical probability analysis in civil engineering based on random field modeling.

The cone penetration test (CPT), which is fast, largely independent of operators and provides nearly continuous data profiles (Lunne et al., 1997), is usually used in random field characterization (Fenton, 1999). Several CPT based methods have been developed to estimate λ in literature, e.g., the maximum likelihood estimation (e.g., DeGroot and Baecher, 1993), sample autocorrelation function method (e.g., Lloret-Cabot et al., 2014) and Bayesian method (e.g., Cao and Wang, 2013), etc. However, the method of $0.8\bar{d}$ -estimator based on the mean cross distance of a soil property profile (e.g., The \bar{d} shown in Figure 1) is approximately the simplest one for objective and quantitative estimation of λ . Therefore, the $0.8\bar{d}$ -estimator is convenient for geotechnical practitioners to make a rapid estimation of λ at site. However, a note of caution, here, is that the $0.8\bar{d}$ -estimator was derived from normal random field with squared exponential correlation function, which is largely ignored in practical applications.

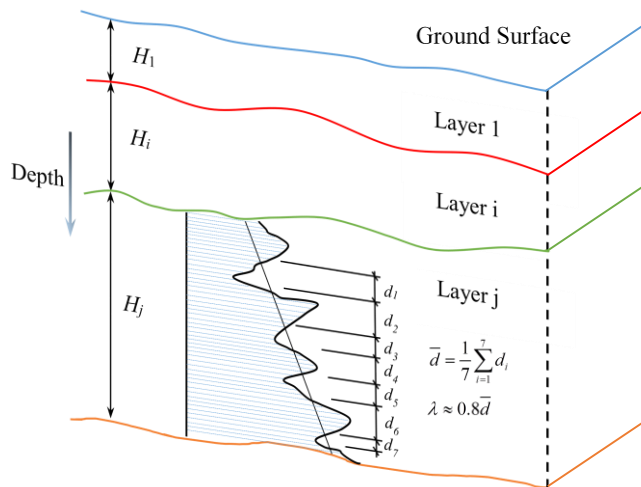


Figure. 1: Inherent soil variability and sketch map of $0.8\bar{d}$ -estimator method (Phoon and Kulhawy, 1999).

To enhance the application availability of $0.8\bar{d}$ -estimator for geotechnical practitioners, this paper clarifies the theoretical assumptions underlying the $0.8\bar{d}$ -estimator and systematically, explores the effects of these assumptions on its practical performances. This paper starts with the analytical derivation of the $0.8\bar{d}$ -estimator, followed by clarification of the underlying assumptions. Finally, the performance including unbiasedness and variability of the $0.8\bar{d}$ -estimator under different various cases are explored using data simulated from a virtual site.

2. DERIVATION OF $0.8\bar{d}$ -ESTIMATOR

Consider, for example, that the inherent spatial variability along the vertical direction of normalized cone tip resistance $q_N = q_c(\sigma'_v/p_a)^{0.5}/p_a$ can be represented by a one-dimensional random field model (Fenton and Griffiths, 2008). Herein, q_c is the cone tip resistance measured by cone penetration test; p_a is the standard atmospheric pressure and it is taken as equal to 100kPa; σ'_v is the vertical effective stress. Let x_N denotes the residual error of q_N after de-trending. Thus, x_N can be considered as a stationary random process $X(D)$ with a constant mean μ_X and standard deviation σ_X , where D is the sampling depth. Without loss of generality, the μ_X can be assumed to be equal to zero in this section.

The length (or distance) of the x_N profile stays above and below the zero-mean are taken as d_0^+ and d_0^- , and the mean up-crossing and down-crossing rate of zero-mean are taken as v_0^+ and v_0^- . The relationship between v_0^+ and d_0^+ (or v_0^- and d_0^-) can be obtained by the theory of recurrent events (Vanmarcke, 1970). The v_0^+ and v_0^- are theoretically identical for a sufficiently long profile of x_N . Then, the relationship between \bar{d} and v_0^+ can be obtained:

$$\bar{d} = \frac{E[d_0^+ + d_0^-]}{2} = \frac{1}{v_0^+ + v_0^-} = \frac{1}{2v_0^+} \quad (1)$$

where v_0^+ was given by Rice (1944, 1945) and commented by Rainal (1988):

$$v_0^+ = \frac{1}{2} \int_{-\infty}^{+\infty} |\dot{x}| f_{X,\dot{X}}(0, \dot{x}) d\dot{x} \quad (2)$$

where \dot{x} denotes the derivative of x ; $f_{X,\dot{X}}(0, \dot{x})$ is the joint probability density function (PDF) of $X(D)$ and its derivative process $\dot{X}(D)$ evaluated at $x=0$. In the context of random process (Bendat and Piersol, 2000), $X(D)$ and $\dot{X}(D)$ are independent and $\dot{X}(D)$ is also a normal random process with a zero mean if $X(D)$ is a stationary normal random process. Thus, Eq. (2) can be rewritten as:

$$v_0^+ = \frac{1}{2} f_X(0) \int_{-\infty}^{+\infty} |\dot{x}| f_{\dot{X}}(\dot{x}) d\dot{x} = \frac{1}{2} f_X(0) E[|\dot{X}|] \quad (3)$$

where $f_X(0)$ is the marginal PDF of $X(D)$ evaluated at $x=0$; $f_{\dot{X}}(\dot{x})$ is the marginal PDF of $\dot{X}(D)$; $E[|\dot{X}|]$ is the mean of the absolute value of the slope of $X(D)$.

$$E[|\dot{X}|] = 2 \int_0^{\infty} \frac{\dot{x}}{\sqrt{2\pi}\sigma_{\dot{X}}} \exp\left\{-\frac{\dot{x}^2}{2\sigma_{\dot{X}}^2}\right\} d\dot{x} = \sqrt{\frac{2}{\pi}} \sigma_{\dot{X}} \quad (4)$$

where $\sigma_{\dot{X}}$ is the standard deviation of the $\dot{X}(D)$. v_0^+ can be given by combining the expression $\sigma_{\dot{X}} = \sigma_X (M_2/M_0)^{1/2} / 2\pi$ (Vanmarcke 1983) and Eqs. (3)-(4):

$$v_0^+ = \frac{1}{2\pi} \frac{\sigma_{\dot{X}}}{\sigma_X} = \frac{1}{2\pi} \left(\frac{M_2}{M_0} \right)^{1/2} \quad (5)$$

where M_0 and M_2 are zero and second order spectral moment of $X(D)$. As indicated by Eq. (5), the estimated v_0^+ is independent of σ_X , so does the λ . The derivation of M_0 and M_2 are based on the spectral density function $s_X(\omega)$ of $X(D)$, where ω is the frequency.

Since the $0.8\bar{d}$ -estimator was derived from the squared exponential correlation function. The spectral density function of $X(D)$ can be obtained using the Wiener-Khinchine formula and Euler formula (Li and Chen, 2009).

$$s_X(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} R_X(\tau) \cos(\omega\tau) d\tau = \frac{\lambda}{2\pi} e^{-\frac{\omega^2 \lambda^2}{4\pi}} \quad (6)$$

Then, M_0 and M_2 can be calculated as follows:

$$M_0 = \int_{-\infty}^{+\infty} \frac{\lambda}{2\pi} e^{-\frac{\omega^2 \lambda^2}{4\pi}} d\omega = 1 \quad (7)$$

$$M_2 = \int_{-\infty}^{+\infty} \omega^2 \frac{\lambda}{2\pi} e^{-\frac{\omega^2 \lambda^2}{4\pi}} d\omega = \frac{2\pi}{\lambda^2} \quad (8)$$

The relationship between \bar{d} and λ can be obtained by combining Eqs. (1), (5), (7) and (8):

$$\bar{d} = \pi \left(\frac{M_2}{M_0} \right)^{1/2} = \sqrt{\frac{\pi}{2}} \lambda \quad (9)$$

Then, Eq. (9) can be approximately rewritten as:

$$\lambda \approx 0.8\bar{d} \quad (10)$$

Eqs. (1)-(10) give the analytical derivation of the classic $0.8\bar{d}$ -estimator, and it shows that the $0.8\bar{d}$ -estimator is valid under the following three assumptions:

1. Normal random field
2. Squared exponential correlation function
3. A sufficiently long sampling distance

The next section explores effects of these assumptions on the performance of $0.8\bar{d}$ -estimator of λ .

3. ILLUSTRATIVE EXAMPLE

In this section, CPT data is simulated from a virtual site for exploring the effects of assumptions on the performance of $0.8\bar{d}$ -estimator. The unbiasedness and variability of the λ value estimated from $0.8\bar{d}$ are two major criteria considered in this section for discussing the performance of the $0.8\bar{d}$ -estimator.

3.1. A virtual site for simulating CPT data

Random field model (Vanmarcke, 1983) is applied to representing a virtual site for simulating CPT data within a statistically homogeneous soil stratum. For example, the detrended normalized cone tip resistance x_N of a virtual site can be represented by a one-dimensional normal random field $X(D)$ with a mean μ_X and standard deviation σ_X , and the spatial correlation of $X(D)$ at a separation distance of τ can be characterized by correlation

Table 1: Five typical correlation functions in geostatistical analysis (Phoon et al, 2003).

Types	Expression
SQECF	$\rho(\tau) = \exp[-\pi(\tau /\lambda)^2]$
SMCF	$\rho(\tau) = (1 + 4 \tau /\lambda)\exp(-4 \tau /\lambda)$
SECF	$\rho(\tau) = \exp(-2 \tau /\lambda)$
BNCF	$\rho(\tau) = \begin{cases} 1 - \tau /\lambda, & \tau \leq \lambda \\ 0, & \tau > \lambda \end{cases}$
CECF	$\rho(\tau) = \cos(\tau /\lambda)\exp(- \tau /\lambda)$

function $\rho(\tau)$. In this study, the five typical correlation functions in geostatistical analysis (i.e., squared exponential correlation function (SQECF), second-order Markov correlation function (SMCF), single exponential correlation function (SECF), binary noise correlation function (BNCF), and cosine exponential correlation function (CECF)) are considered to represent the spatial correlation of $X(D)$ as shown in Table 1.

Let $\underline{X}=[X(D_1), X(D_2), \dots, X(D_n)]$ be $n \times 1$ vector of x_N data at n different sampling depths. Thus, \underline{X} follows multivariate normal distribution with a mean vector $\mu_X \underline{l}$ and covariance matrix $\underline{C} = \sigma_X^2 \underline{R}$, in which \underline{l} is a vector with n components that are all equal to one and \underline{R} is the correlation matrix of \underline{X} . Then, \underline{X} can be written as (e.g., Cao and Wang, 2014).

$$\underline{X} = \mu_X \underline{l} + \sigma_X \underline{L}^T \underline{N} \quad (11)$$

in which \underline{N} is a $n \times 1$ standard normal random vector; \underline{L} is a n by n upper-triangular matrix obtained from Cholesky decomposition of correlation coefficient matrix \underline{R} , the (i,j) -th entry $\rho_{i,j}$ of which represents the correlation coefficient between $X(D_i)$ and $X(D_j)$. The $\rho_{i,j}$ can be calculated using the five typical correlation functions shown in Table 1.

Eq. (11) is used to simulate x_N data (e.g., $\hat{\underline{X}} = [\hat{X}(D_1), \hat{X}(D_2), \dots, \hat{X}(D_n)]$) with different sampling depths and intervals in this section. Moreover, when the lognormal random field is of interest, the simulation is performed using the

following equations. Consider, for example, a stationary lognormal random field of $X(D)$ with squared exponential correlation function. The statistics of the logarithm of $X(D)$ (i.e., $Y(D) = \ln X(D)$) are calculated as follows (Fenton and Griffiths, 2008):

$$\mu_Y = \ln \mu_X - \sigma_Y^2 / 2 \quad (12)$$

$$\sigma_Y = \sqrt{\ln [1 + (\sigma_X / \mu_X)^2]} \quad (13)$$

$$\rho_Y = \frac{\ln [(\sigma_X / \mu_X)^2 \exp[-\pi(|\tau|/\lambda)^2] + 1]}{\sqrt{\ln [1 + (\sigma_X / \mu_X)^2]}} \quad (14)$$

where μ_Y and σ_Y are the mean and standard deviation of $Y(D)$, respectively; ρ_Y is correlation function of $Y(D)$. It is worth noting that ρ_Y is largely dependent on the coefficient of variation of $X(D)$, (i.e., $Cov = \sigma_X / \mu_X$), which means that ρ_Y is sensitive to the statistics of $X(D)$. Since there are various combinations of σ_X / μ_X , the method of simple estimator may not be suitable for the lognormal random field. In this section, for the limited page, the lognormal random field with correlation function of SQECF (see in Table 1) is taken as an example.

3.2. Simulated cases

Although there are various combinations of random field model parameters, it was proved and explained in previous section that the μ_X and σ_X have no effect on the performance of the simple estimator in analytical derivation, and the sensitivity studies using different combinations of random field model parameters (i.e., μ_X and σ_X) provide similar results. Consider, for example, $\mu_X = 150$ and $\sigma_X = 15$ in this section, which are consistent with typical values of normalized cone tip resistance reported in the literature (e.g., Uzielli et al., 2005). The range of λ is taken as $[0.1, 6]$ according to the previous studies on CPT-based spatial variability characterization (e.g., Cao et al., 2016), and thirteen typical values shown in the fourth row of Table 2 are selected to be the predetermined scale of fluctuation, λ_T .

To systematically explore the performances of $0.8\bar{d}$ -estimator under different cases, this section considers five correlation functions shown in Table 1 (i.e., SECF, BNCF, CECF, SMCF and SQECF), two types of random field (i.e., normal and lognormal random fields), three sampling intervals (i.e., 0.02m, 0.05m and 0.1m), and four sampling lengths (i.e., 20m, 30m, 40m and 50m) for simulating x_N data at a virtual site. This results in a total of 351 cases, which are summarized in Table 2.

For each case, 50 sets of x_N data are generated to account for the effect of statistical uncertainty, and the λ value of each set of simulated data is estimated by $0.8\bar{d}$. The mean value of the 50 sets of calculated λ is denoted as λ_M , and the ratio λ_M/λ_T indicates the unbiasedness of the λ value estimated from the $0.8\bar{d}$. The value of ratio $\rho=\lambda/\lambda_T$ is also calculated to explore the accuracy of the λ value estimated from the $0.8\bar{d}$.

3.3. Performance of the $0.8\bar{d}$ -estimator

Figure. 2 shows the performance of the unbiasedness of $0.8\bar{d}$ -estimator under different correlation functions and different types of random fields. Results of λ_M/λ_T under different sampling intervals (i.e., 0.02m, 0.05m and 0.1m) are shown by hollow symbols of squares, circles, and triangles, respectively. The solid horizontal line of $\lambda_M/\lambda_T=1$ provides the reference to examine the unbiasedness of the λ value estimated from the $0.8\bar{d}$. The closer the λ_M/λ_T values are to the line, the more unbiased the estimated λ values are.

For the normal random field, it is found that, for the correlation functions of SECF, BNCF and CECF (see in Figure. 2(a)-(c)), the λ_M/λ_T values of all decay exponentially, which indicates that using the $0.8\bar{d}$ -estimator leads to an underestimation of the λ , especially when the predetermined λ_T value is relatively large. Moreover, the performance of $0.8\bar{d}$ -estimator depends on the sampling interval. The smaller the sampling interval is, the lower the estimated λ value is.

As shown in Figure. 2(d)-(e), for the correlation functions of SMCF and SQECF, the λ_M/λ_T values are invariant for different sampling intervals under the unbiased cases, and there is an approximate linear relationship between \bar{d} and λ . More importantly, the $0.8\bar{d}$ -estimator is statistically unbiased for SQECF when there are, at least, two sampling data within a distance of λ_T . If there is only one sampling data within a distance of λ_T , using the $0.8\bar{d}$ -estimator leads to a significant overestimation of λ . The λ value estimated from $0.8\bar{d}$ for SMCF is somewhat underestimated.

For the lognormal random field with correlation function of SQECF, the performance of the $0.8\bar{d}$ -estimator is rather poor as shown in Figure. 2(f). Similar to the performance of the random fields with SECF, BNCF and CECF, using the $0.8\bar{d}$ -estimator for lognormal random fields generally leads to underestimation of λ , particularly as λ_T is relatively large. The performance under different sampling interval is also different. Hence, the $0.8\bar{d}$ -estimator is not suitable for the lognormal random field.

Table 2: Summary of simulated cases.

Factors	Simulated Cases												
Correlation Function	SECF			BNCF			CECF			SMCF		SQECF	
Random Field	Normal						Lognormal						
Predetermined SOF (m), λ_T	0.1	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
Sampling Interval (m), ΔD	0.02			0.05			0.10						
Sampling Length (m), D	50			40			30			20			

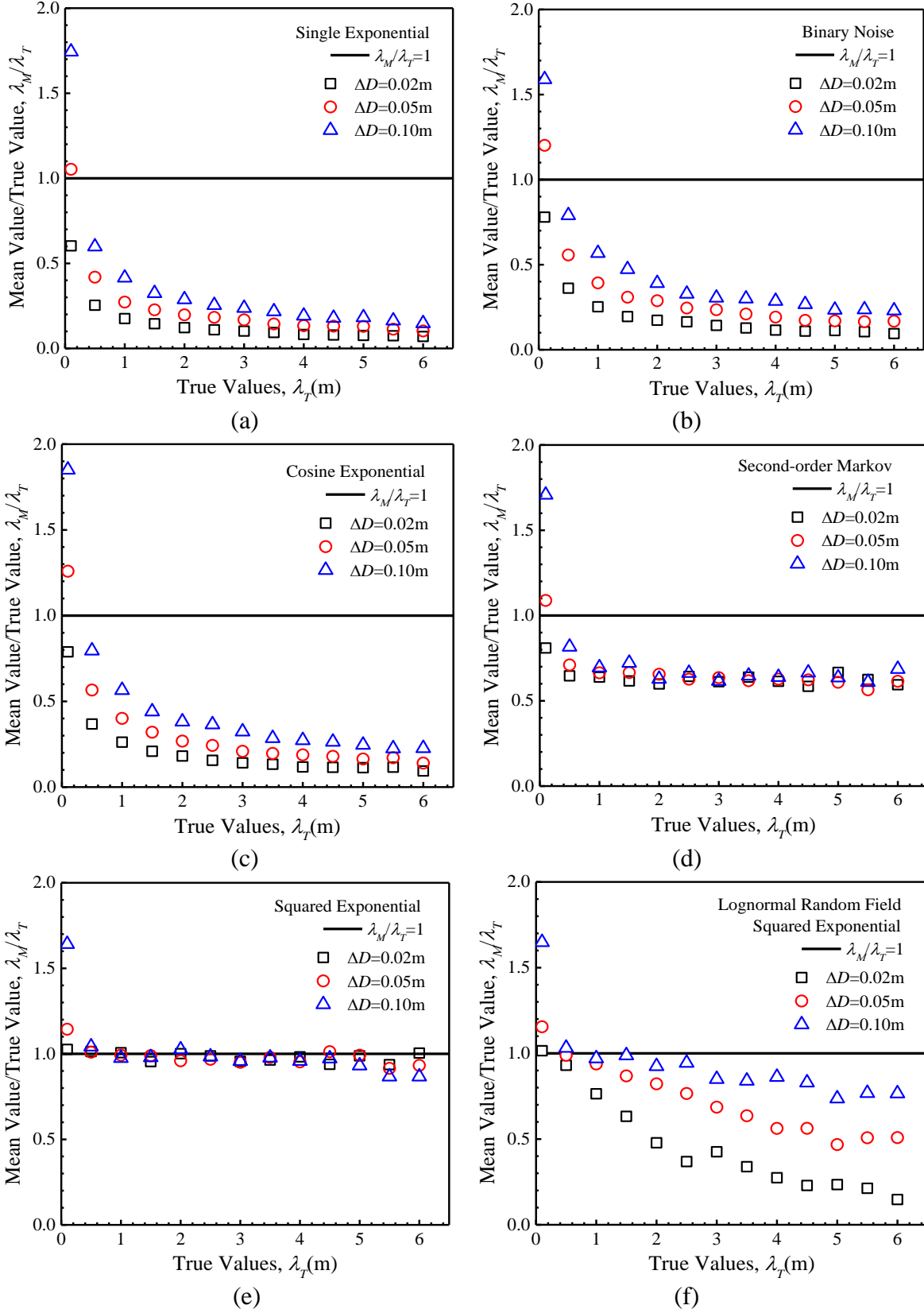


Figure 2: Unbiasedness of λ values estimated from the 0.8δ -estimator.

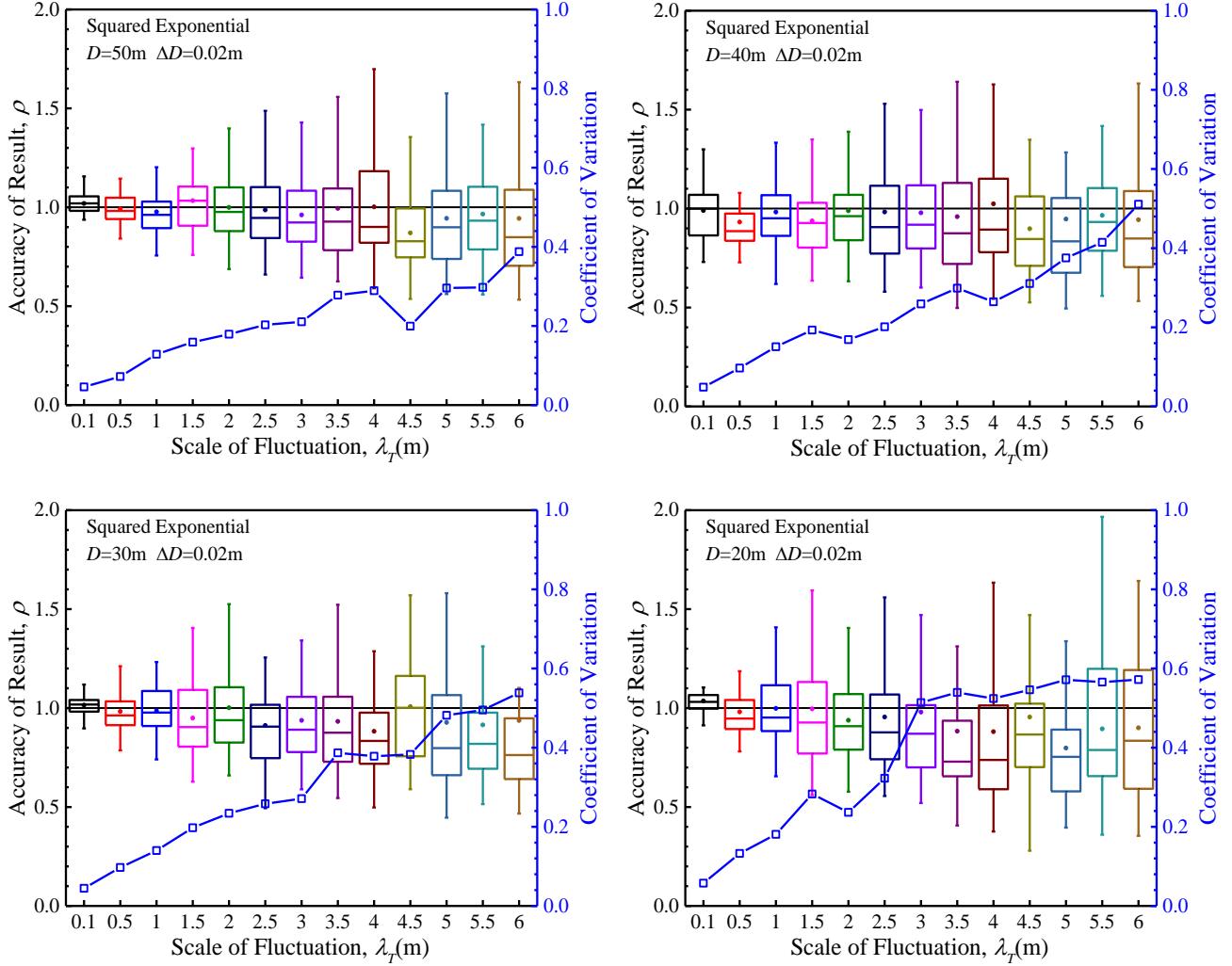


Figure 3: Variability of λ values estimated from $0.8\bar{d}$ -estimator for the normal random field with SQECF.

The performance of the $0.8\bar{d}$ -estimator in terms of variability for the normal random field with SQECF are also investigated by considering different sampling lengths (including 20m, 30m, 40m, and 50m). As shown in Figure. 3, the left axis indicates the accuracy of result ρ , which are shown by the box-plot, and the right axis colored by blue indicates the coefficient of variation (i.e., Cov) of the $\rho=\lambda/\lambda_T$ value, which is shown by the line with hollow squares symbols colored by blue. The solid horizontal line indicates the unbiased situation of $\lambda/\lambda_T=1$. The box-plot contains the maximum and minimum value, upper and lower quartiles, and the mean value of ρ . It can be found that when the sampling length is fixed, the variability of estimated ρ value increases as the λ_T increases. Moreover, for a given λ_T , the

variability increases as the sampling length decreases. Thus, the variability of the λ value estimated from the $0.8\bar{d}$ is significantly affected by the sampling length.

4. SUMMARY AND CONCLUSIONS

This paper clarifies theoretical assumptions underlying the $0.8\bar{d}$ -estimator, and systematically, explores effects of these assumptions on its performance using CPT data simulated from a virtual site. Results showed that the $0.8\bar{d}$ -estimator is only valid for the normal random field with squared exponential correlation function (SQECF) when there are, at least, two sampling points within a distance of λ . Using $0.8\bar{d}$ -estimator for other cases that violating the assumptions causes significant underestimation

of λ . In addition, it was also exhibited that the $0.8\bar{d}$ -estimator is not applicable to the lognormal random fields neither. The variability of the λ value estimated from the $0.8\bar{d}$ largely depends on the sampling length. Although the $0.8\bar{d}$ -estimator provides a simple and convenient way to estimate λ at site, it is only applicable when the underlying assumptions (i.e., normal random field with correlation function of SQECF and a sufficient amount of samples) are satisfied. These assumptions shall be bearing in mind when using it in practice.

5. ACKNOWLEDGMENTS

The work described in this paper was supported by grants from National Key R&D Program of China (Project No. 2017YFC1501300), and the National Natural Science Foundation of China (Project Nos. 51528901, 51679174, 51779189), and an open fund from State Key Laboratory Hydraulics and Mountain River Engineering, Sichuan University (Project No. SKHL1619). The financial support is gratefully acknowledged.

6. REFERENCES

- Bendat, J. S., and Piersol, A. G. (2000). *Random data analysis and measurement procedures*. John Wiley and Sons. Hoboken, New Jersey.
- Cao, Z. J., and Wang, Y. (2013). "Bayesian approach for probabilistic site characterization using cone penetration tests." *Journal of Geotechnical and Geoenvironmental Engineering*, 139(2): 267-276.
- Cao, Z. J., and Wang, Y. (2014). "Bayesian model comparison and selection of spatial correlation functions for soil parameters." *Structural Safety*, 49: 10-17.
- Cao, Z. J., Wang, Y., and Li, D. Q. (2016). "Quantification of prior knowledge in geotechnical site characterization." *Engineering Geology*, 203:107-116
- DeGroot, D. J., and Baecher, G. B. (1993). "Estimating Autocovariance of In-Situ Soil Properties." *Journal of Geotechnical Engineering*, 119(1): 147-166.
- Fenton, G. A. 1999. Random field modeling of CPT data. *Journal of Geotechnical and Geoenvironmental Engineering*, 126(12): 486-498.
- Fenton, G. A., and Griffiths, D. V. (2008). *Risk Assessment in Geotechnical Engineering*. John Wiley and Sons. Hoboken, New Jersey.
- Lloret-Cabot, M., Fenton, G. A., and 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.
- Li, J., and Chen, J. B. (2009). *Stochastic dynamics of structures*. John Wiley & Sons.
- Lunne, T., Robertson P. K., and Powell J. J. M. (1997). *Cone penetration testing in geotechnical practice*, Blakie Academic and Professional, Melbourne, Australia.
- Mitchell, J. K., and Soga, K. (2005). *Fundamentals of soil behavior*. John Wiley and Sons. Hoboken, New Jersey.
- Phoon, K. K., and Kulhawy, F. H. (1999). "Characterization of geotechnical variability." *Canadian Geotechnical Journal*, 36(4): 612-624.
- 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*, 129(7): 649-659.
- Rainal, A. J. (1988). "Origin of Rice's formula." *IEEE transactions on information theory*, 34(6): 1383-1387.
- Rice, S. O. (1944). "Mathematical analysis of random noise." *Bell System Technical Journal*, 23(3): 282-332.
- Rice, S. O. (1945). "Mathematical analysis of random noise." *Bell System Technical Journal*, 24(1): 46-156.
- Uzielli, M., Vannucchi, G., Phoon, K. K. (2005). "Random field characterisation of stress-normalised cone penetration testing parameters." *Géotechnique*, 55(1): 3-20.
- Vanmarcke, E. H. (1970). First passage and other failure criteria in narrow-band random vibration: a discrete state approach. Ph.D. thesis, Massachusetts Institute of Technology.
- Vanmarcke, E.H. (1977). Probabilistic modeling of soil profiles. *Journal of the geotechnical engineering division*, 103(11): 1227-1246.
- Vanmarcke, E. H. (1983). *Random fields: analysis and synthesis*. The MIT Press. Cambridge (MA), London.