Scale of Fluctuation for Geotechnical Probabilistic Analysis

Xinyao NIE[a], Jie ZHANG[a], Hongwei HUANG[a], Zhongqiang LIU[b] and Suzanne LACASSE[b]

[a] *Department of Geotechnical Engineering, Tongji University, China.*
[b] *Norwegian Geotechnical Institute, Oslo, Norway.*

Abstract. In the past few years, random field theory has been increasingly used to model the inherent soil variability. The scale of fluctuation is one of the important parameters describing a stationary random field. In this study, the factors affecting an accurate estimation of the scale of fluctuation were studied with numerical experiments to show how a proper sampling strategy can help improve the estimate of scale of fluctuation. Hypothetical data sets were generated from random field theory. Data were then sampled for different sampling strategies. The scale of fluctuation estimated from the sampling programs were compared with the predefined scale of fluctuation. The accuracy with which one can estimate scale of fluctuation depends on both the sampling intensity and extent of the sampling range. For the numerical example in this study, the sampling interval should be close enough such that 10 samples are measured within one scale of fluctuation, and the distance covered by the sampling should cover at least 100 scales of fluctuation.

Keywords. random field, scale of fluctuation, spatial variability, parameter estimation, sampling strategy

1. Introduction

The inherent variability of soil is one of the major sources of uncertainties in geotechnical engineering (e.g., Vanmarcke, 1977; Phoon and Kulhawy, 1999). In the past few years, random field theory has been increasingly used to model the inherent (aleatory) soil variability to reduce the uncertainty in soil characteristics (e.g., Keaveny et al., 1989; Lacasse and Nadim, 1996; Fenton and Griffiths, 2002; Dasaka and Zhang, 2012). A stationary random field is widely used in geotechnical engineering. It is often characterized by the mean, variance and scale of fluctuation (SoF). Generally, while the mean and variance can be determined conveniently, much more efforts are required to estimate the SoF of the random field (Onyejekwe and Ge, 2013). In particular, the quantity of data is often quite limited in geotechnical engineering. It is also largely unclear to the geotechnical profession how the samples should be taken in the field to ensure a reliable estimate of the SoF. The site exploration program is often planned without particular consideration of the random nature of the soil characteristics.

This paper studies the factors affecting the estimation of SoF with numerical experiments.

The most common autocorrelation models to estimate SoF is first presented. Then, the numerical experiments for assessing the factors affecting the estimation of SoF are described. The results of the numerical experiments are then presented and discussed.

2. Scale of fluctuation

Because of the complex geological and environmental processes involved, the soil characteristics *in situ* are rarely homogeneous. The soil characteristics can be highly variable and spatially correlated in the vertical and horizontal directions. As shown in Figure 1, a soil property $g(z)$ can be decomposed into a deterministic trend component $t(z)$ and a stationary random function $w(z)$ as follows (e.g., DeGroot and Baecher, 1993):

$$g(z) = t(z) + w(z) \qquad (1)$$

The concept of the SoF was first proposed by Vanmarcke (1977). Within the SoF, the values of $w(z)$ will tend to be either all above or all below zero, indicating that the soil property with-

in the SoF shows a relatively strong correlation. The SoF can be infinite when the soil is described by a fractal model (e.g., Fenton, 1999; Jaksa, 2013). This paper discusses the finite SoF only. Fenton (1999) suggested that there may be little difference between the finite and infinite SoF's if an appropriate finite-scale model and a fractal model over the finite domain are used.

In practice, the SoF should be estimated from a population of observations. Various approaches for estimating the SoF have been proposed. To apply these methods, the data may need to be transformed such that the stationary assumption is valid (e.g., Campanella et al., 1987). In the present study, the autocorrelation fitting method (ACFM) will be used for estimating the SoF. It appears to be one of the most widely used methods for estimating SoF (e.g., Uzielli et al., 2005; Lloret-Cabot et al., 2014).
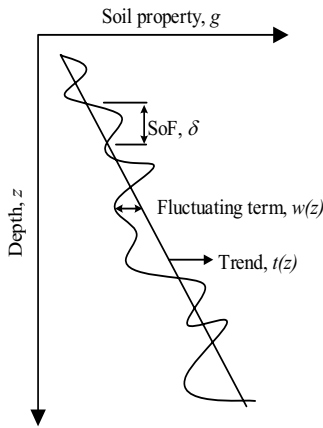
tion function can be obtained from the following equation:

$$\hat{\rho}(\tau) = \frac{\sum_{i=1}^{n(\tau)} \left[w(z_i) - \overline{w}\right]\left[w(z_i + \tau) - \overline{w}\right]}{\left[n(\tau) - 1\right]\hat{\sigma}^2} \qquad (2)$$

where $n(\tau)$ denotes the number of pairs that are separated by the distance τ. As τ increases, less number of pairs will be available for calculating $\hat{\rho}(\tau)$. Generally, the data for calculating $\hat{\rho}(\tau)$ are considered insufficient for obtaining a reliable estimate of the SoF when τ exceeds the distance of a quarter of the sampling space domain (Lumb, 1975; Box et al., 1994).

Table 1. Theoretical autocorrelation models (Vanmarcke, 1977 and 2010)

Model	Autocorrelation function
Triangular	$\rho(\tau) = \begin{cases} 1 - |\tau|/\delta, & |\tau| \le \delta \\ 0, & |\tau| > \delta \end{cases}$
Exponential	$\rho(\tau) = \exp\left(-\dfrac{2|\tau|}{\delta}\right)$
Squared exponential	$\rho(\tau) = \exp\left(-\pi\left(\dfrac{\tau}{\delta}\right)^2\right)$
Cosine exponential	$\rho(\tau) = \cos\left(\dfrac{\tau}{\delta}\right)\exp\left(-\dfrac{|\tau|}{\delta}\right)$
Second-order Markov	$\rho(\tau) = \left(1 + \dfrac{4|\tau|}{\delta}\right)\exp\left(-\dfrac{4|\tau|}{\delta}\right)$

Note: τ is separation distance and δ is scale of fluctuation.

Figure 1. Illustration of the soil inherent variability

3. Autocorrelation fitting method (ACFM)

The main idea of ACFM is to fit theoretical models to the sample autocorrelation function $\hat{\rho}(\tau)$ based on an ordinary least squares approach. Some common theoretical autocorrelation models are given in Table 1.

For a one-dimensional stationary random field $w(z)$, let's assume that the sequence of observations $\mathbf{w}^T = \{w(z_1), w(z_2), \ldots, w(z_n)\}$ is made. Let $\overline{w}$ and $\hat{\sigma}^2$ denote the sample mean and the sample variance of $w(z)$. The sample autocorrela-

4. Numerical experiments

To study the factors affecting the estimation of SoF, data were generated from random field theory with a given mean, standard deviation and SoF. The data were then extracted for different site exploration strategies and used for estimating SoF with the ACFM approach. By comparing the estimated values of SoF with the predefined value of SoF (*ASoF*), the accuracy of the estimation was assessed. In this study, the random field was assumed to be a one-dimensional Gaussian random field with a mean of 0, a variance of 1 and an SoF of 100. The exponential autocorrelation model shown in Table 1 was used. The

Cholesky decomposition method was implemented to generate the data from the random field (DeGroot and Baecher, 1993).

As an example, one can assume a length of sampling space domain (L) of 1000 and a size of sampling interval (D) of 10 in a site exploration program. Figure 2 shows two possible data sets generated from this site exploration program. The values of SoF estimated from datasets A and B are 90.9 and 61.9, respectively. Due to the random nature of the data, the estimated values of SoF from one given site exploration program can also vary.
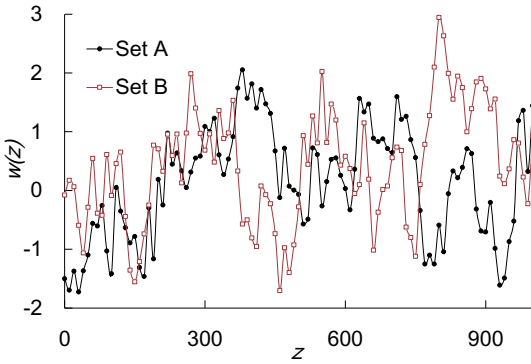

Figure 2. Two sets of sampled data (D=10, $ASoF$=100 and L=1000)

In the present study, Monte Carlo simulation was used to assess the effectiveness of a site exploration program. Specifically, N datasets from the site exploration program were considered, and the SoF was evaluated for each of the datasets. The values of the mean (μ_{SoF}) and standard deviation (σ_{SoF}) of the estimated SoFs were then calculated. Figures 3 and 4 show the mean and standard deviation of the estimated SoFs as N increases from 5 to 3000. The values of μ_{SoF} and σ_{SoF} gradually stabilize as N increases and it appears that N = 3000 is sufficient for estimating μ_{SoF} and σ_{SoF} without too much computation effort. In this study, a target of N = 3000 was employed.
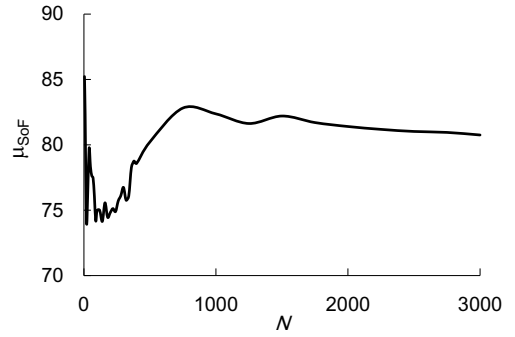

Figure 3. Mean SoF (μ_{SoF}) versus N (D=10, $ASoF$ = 100 and L = 1000)
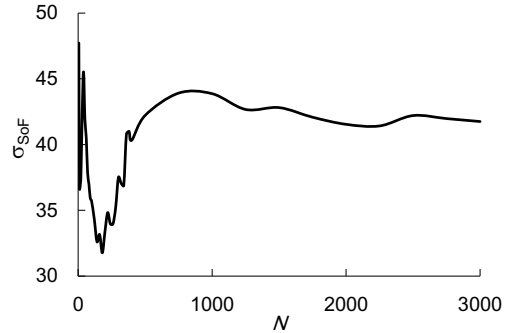

Figure 4. Standard deviaiotn of SoFs (σ_{SoF}) versus N (D=10, $ASoF$ = 100 and L = 1000)

To evaluate the effectiveness of the site exploration program, the relative error (ε_{SoF}) and coefficient of variation (δ_{SoF}) were defined as follows:

$$\varepsilon_{SoF} = \frac{|\mu_{SoF} - ASoF|}{ASoF} \qquad (3)$$

$$\delta_{SoF} = \frac{\sigma_{SoF}}{\mu_{SoF}} \qquad (4)$$

The site exploration program is more effective as both ε_{SoF} and δ_{SoF} get closer to 0.

5. Results and discussions

5.1. Distribution of estimators

Figure 5 shows the histogram of the estimators for the scale of fluctuation (SoF), as estimated by the aforementioned site exploration program and Monte Carlo simulations. The distribution of estimators is right-skewed and more than two thirds of the estimators are smaller than the pre-

defined scale of fluctuation implied in the analysis (*ASoF*) of 100. Note that in the example in Figure 5, the parameter ε_{SoF} is 19% and δ_{SoF} is 52%, indicating that this particular sampling strategy is biased and has rather high variability.

5.2. Sampling intensity (sampling interval)

In the present study, *L* and *D* are two parameters that define a sampling strategy, where *D* denotes the sampling interval and *L* denotes the length of the domain over which sampling is done (also called sampling range). To screen out this possible effect, normalized values were used to represent the sampling interval and sampling range.
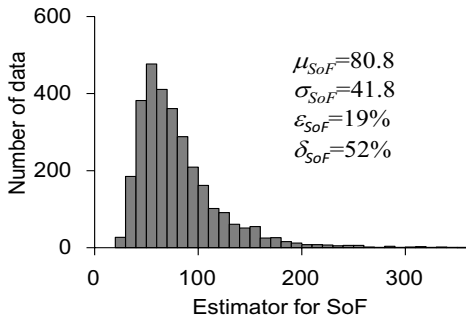

Figure 5. The histogram of estimators for SoF (*D* = 10, *ASoF*=100 and *L* = 1000)

Figures 6 and 7 show how the error and coefficient of variation of the SoF, ε_{SoF} and δ_{SoF}, change with the normalized sampling interval (*ASoF/D*) for the case *L/ASoF* of 10. A larger normalized sampling interval (*ASoF/D*) denotes "more intense" sampling. The two figures show that when *ASoF/D* exceeds 10, both ε_{SoF} and δ_{SoF} become nearly constant. The results indicate that a further increase in the sampling intensity (decrease in sampling interval *D*) does not increase the effectiveness of the sampling in terms of reducing the uncertainty due to natural variability. In practice, the sampling costs increase with the sampling intensity. Thus, from a cost-effectiveness point of view, the normalized sampling interval adopted in the site exploration program should not exceed 10, unless one wants to reduce the epistemic uncertainty due to lack of knowledge (e.g., Baecher and Christian, 2003).
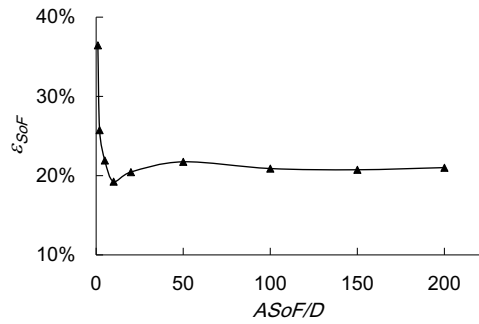

Figure 6. ε_{SoF} versus normalized sampling interval *ASoF/D* for *L/ASoF* = 10
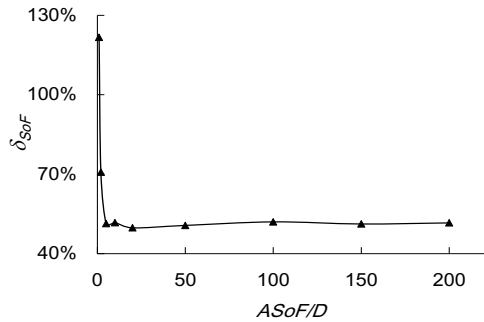

Figure 7. δ_{SoF} versus normalized sampling interval *ASoF/D* for *L/ASoF* = 10

5.3. Sampling range

As shown in the previous section, *ASoF/D*=10 implies a fairy dense sampling interval. To study the effect of sampling range or the length of the domain over which sampling is done, Figures 8 and 9 show how the parameters ε_{SoF} and δ_{SoF} change with *L/ASoF* for the case of *ASoF/D* = 10. Here, *L/ASoF* denotes the normalized sampling range. A larger value of *L/ASoF* denotes a wider sampling range.
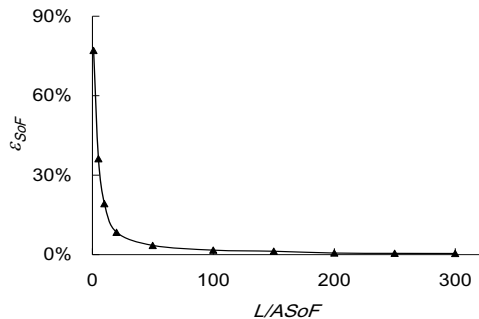

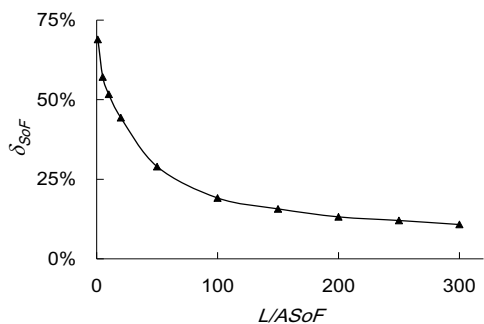**Figure 8.** ε_{SoF} versus normalized sampling range *L/ASoF* for *ASoF/D* = 10

Figure 9. δ_{SoF} versus normalized sampling range $L/ASoF$ for $ASoF/D = 10$

Table 2. Scale of fluctuation of CPT Cone Resistance (q_c)

Soil type	Direction	SoF (m)	Reference
Offshore soils	Horizontal	30	Høeg and Tang (1976); Tang (1979)
Silty clay	Horizontal	5-12	Lacasse and de Lamballerie (1995)
Offshore sand	Horizontal	14-38	Keaveny et al. (1989)
Sand	Vertical	0.13-0.71	Campanella et al. (1987)
Silty clay	Vertical	1	Lacasse and de Lamballerie (1995)
Clay	Vertical	0.4	Cafaro and Cherubini (2002)
Sand, clay	Vertical	0.13-1.11	Uzielli et al. (2005)

In general, both ε_{SoF} and δ_{SoF} decrease as the sampling range increase, indicating that increasing the sampling range helps to improve the effectiveness of the sampling program. As shown in Figure 8, when $L/ASoF$ exceeds 100, ε_{SoF} is very close to 0, indicating that a further increase in the sampling range is not necessary to better estimate the SoF. As shown in Figure 9, δ_{SoF} does not converge to a stable state even when $L/ASoF$ is larger than 300, but the variance is getting smaller and smaller. A sampling range wider than 300 is required to reduce δ_{SoF} to a value less than 10%. In the case analyzed in this paper, a normalized sampling range of 300 implies 3000 measurements and 3000 measurements are not sufficient to reduce the statistical error to zero. Thus, it seems that a very large quantity of data are required for an accurate estimation of SoF.

5.4. Scale of fluctuation

In practice, the soils are often sampled through a predetermined investigation scheme with given D and L. Figures 10 and 11 illustrate the effect of $ASoF$ on ε_{SoF} and δ_{SoF} for the case of $D = 1$ and $L = 1000$, respectively. Both ε_{SoF} and δ_{SoF} increase as $ASoF$ increases, indicating that the effectiveness of a sampling program may vary with the underlying value of SoF. Thus, when planning the site exploration program for estimating the SoF, it is important to first have some fair judgment about the actual possible range of the SoF. Table 2 lists some values of SoF of CPT cone resistance (q_c) reported in literatures.
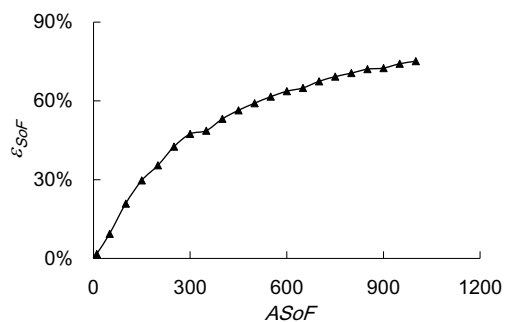

Figure 10. ε_{SoF} versus $ASoF$ ($D = 1$ and $L = 1000$)

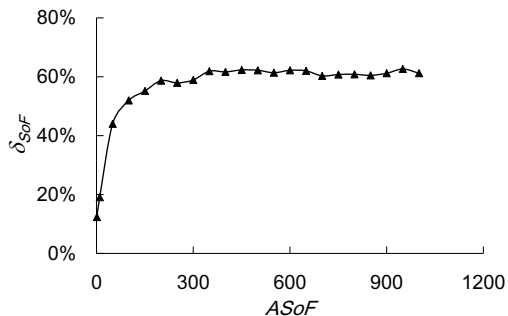**Figure 11.** δ_{SoF} versus $ASoF$ ($D = 1$ and $L = 1000$)

5.5. Implications for site exploration program

From the results of the analysis, an accurate estimation of SoF is very data demanding, which in most cases exceeds the data available in ordinary geotechnical engineering. Considering such data demand, tests such as the cone penetration test (CPT) that can produce a large number of measurements cost-effectively would be more useful for estimating the spatial variability of the soil properties than other non-continuous tests. However, if a soil layer is not thick enough, even

the CPT may not be able to produce sufficient data for an accurate estimation of SoF. The SoF values inferred in many geotechnical projects may only be an approximate estimate of the actual SoF and may not be the actual SoF, due to the limited quantity of data in geotechnical engineering.

The effectiveness of a sampling strategy is a function of the sampling intensity and sampling range. With the current level of efforts in site exploration, the effectiveness of a site exploration program could be enhanced by carefully selecting the sampling interval and the sampling range as a function of the expected scale of fluctuation.

6. Conclusion

The scale of fluctuation in a random field model characterizes the spatial variability of soil properties, which can be of great importance to geotechnical probabilistic analysis. Through numerical solutions, the paper investigated the factors affecting the cost-effectiveness of a site exploration program. The analyses showed that the reliability of the scale of fluctuation estimate depends on both the sampling interval and the extent of the sampling. For the numerical example assessed in this study, the sampling interval should be short enough such that at least 10 observations (measurements) are obtained within one scale of fluctuation, and the sampling extent should be wide enough such that it covers at least 100 times the scale of fluctuation.

7. Acknowledgement

References

Baecher, G. B., and Christian, J. T. (2003). *Reliability and Statistics in Geotechnical Engineering*. John Wiley & Sons.

Box, G. E., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. Englewood Cliffs, NJ: Prentice Hall.

Cafaro, F., and Cherubini, C. (2002). Large Sample Spacing in Evaluation of Vertical Strength Variability of Clayey Soil. *Journal of Geotechnical and Geoenvironmental Engineering*, 128(7), 558-568.

Campanella, R. G., Wickremesinghe, D. S., and Robertson, P. K. (1987). Statistical Treatment of Cone Penetrometer Test Data. *Proceedings of the 5th International Conference on Application of Statistics and Probability*, Rotterdam. 1011-1019.

Dasaka, S. M., and Zhang, L. M. (2012). Spatial Variability of in Situ Weathered Soil. *Géotechnique*, 62(5), 375-384.

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). Estimation for Stochastic Soil Models. *Journal of Geotechnical and Geoenvironmental Engineering*, 125(6), 470-485.

Fenton, G. A., and Griffiths, D. V. (2002). Probabilistic Foundation Settlement on Spatially Random Soil. *Journal of Geotechnical and Geoenvironmental Engineering*, 128(5), 381-390.

Høeg, K., and Tang, W. H. (1976). Probabilistic Considerations in the Foundation Engineering of Offshore Structures. *Proceedings of 2nd International Conference of Structural Safety and Reliability*, Aachen, Germany. 29.

Jaksa, M. B. (2013). Assessing Soil Correlation Distances and Fractal Behavior. *Foundation Engineering in the Face of Uncertainty: Honoring Fred H. Kulhawy*. 405-420.

Keaveny, J. M., Nadim, F., and Lacasse, S. (1989). Autocorrelation Functions for Offshore Geotechnical Data. *International Conference of Structural Safety and Reliability*, Perth, Australia. 263-270.

Lacasse, S., and de Lamballerie, J. Y. (1995). Statistical Treatment of CPT Data. *Proceedings of the International Symposium on Cone Penetration Testing*. 4-5.

Lacasse, S., and Nadim, F. (1996). Uncertainties in Characterising Soil Properties. *Uncertainty in the Geologic Environment: From Theory to Practice*. 49-75.

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.

Lumb, P. (1975). Spatial Variability of Soil Properties. *Proceedings of the second international conference on application of statistics and probability in soil and structural engineering,* Aachen, Germany. 397-421.

Onyejekwe, S., and Ge, L. (2013). Scale of Fluctuation of Geotechnical Parameters Estimated From CPTu and Laboratory Test Data. *Foundation Engineering in the Face of Uncertainty: Honoring Fred H. Kulhawy*. 434-443.

Phoon, K. K., and Kulhawy, F. H. (1999). Characterization of Geotechnical Variability. *Canadian Geotechnical Journal*, 36(4), 612-624.

Tang, W. H. (1979). Probabilistic Evaluation of Penetration Resistances. *Journal of the Geotechnical Engineering Division*, 105(10), 1173-1191.

Uzielli, M., Vannucchi, G., and Phoon, K.K. (2005). Random Field Characterisation of Stress-Normalised Cone Penetration Testing Parameters. *Géotechnique*, 55(1), 3-20.

Vanmarcke, E. H. (1977). Probabilistic Modeling of Soil Profiles. *Journal of the Geotechnical Engineering Division*, 103(11), 1227-1246.

Vanmarcke, E. (2010). *Random Fields: Analysis and Synthesis*. World Scientific.