Site Sampling: Assessing Residual Uncertainty

Échantillonnage du site : évaluation de l'incertitude résiduelle

Fenton G.A.
Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands
Department of Engineering Mathematics, Dalhousie University, Halifax, Nova Scotia, Canada

Hicks M.A.
Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands

ABSTRACT: Geotechnical design is plagued by the uncertainty associated with site characterization. Common questions are "How many samples should be taken?" and "How do these samples reduce my uncertainty?" Of considerable interest is the question "What site sampling plan will give the best cost to effectiveness ratio?" This papers looks specifically at the effect of the number of samples on residual uncertainty. The results can be used to quantitatively select the required number of samples needed to achieve a target maximum residual uncertainty level. To study this problem, a square domain is selected (the site) and a stationary Gaussian random field is simulated within the domain (the random soil properties). The random field is sampled at a series of locations and a trend is estimated from the samples. The trend is then removed from the random field and the residual random field is statistically analyzed to determine various measures of the effectiveness of the sampling scheme. These measures include: 1) the variance of the residual field average (i.e. does the estimate represent the average?), 2) the residual standard deviation (i.e. how much residual uncertainty remains?), and 3) the residual correlation length (i.e. how does trend removal affect the perceived correlation lengths?).

RÉSUMÉ : Le design géotechnique est traditionnellement affecté par des incertitudes associées à la caractérisation du site. Les questions les plus courantes sont : combien d'échantillons devraient être prélevés ? Comment ces échantillons peuvent réduire mon incertitude ? Un des intérêts les plus importants vient de cette question. Quel plan d'échantillonnage du site donnera le meilleur coefficient d'efficacité? Cet article examine spécifiquement l'effet du nombre d'échantillons sur des incertitudes résiduelles. Les résultats peuvent être utilisés pour quantifier et sélectionner le nombre demandé d'échantillons nécessaires pour atteindre un objectif d'incertitude maximal avec le niveau résiduel. Pour étudier ce problème, un domaine carré est sélectionné (le site) et un champ gaussien aléatoire stationnaire est simulé dans le domaine (les propriétés du sol aléatoires). Le champ aléatoire est échantillonné à une série d'emplacements et une tendance a été estimée à partir de l'échantillon. La tendance retirée du champ aléatoire et le champ résiduel aléatoire est statistiquement analysées afin de déterminer les mesures diverses de l'efficacité du plan d'échantillonnage. Ces mesures comprennent : 1) la variance de la moyenne de champ résiduel, c'est à dire comment la tendance estimée représentent la moyenne réelle sur le terrain ? 2) l'écart type résiduel, c'est-à-dire à quel degré d'incertitude résiduelle demeure, et 3) la valeur longueur résiduelle de corrélation, c'est-à-dire comment la suppression tendance affecte les longueurs de corrélation ?.

KEYWORDS: geotechnical design, site characterization, residual uncertainty, sampling, required number of samples, sampling plans.

1 INTRODUCTION

Site characterization is clearly an essential component of any geotechnical design and a great deal of effort has been devoted over recent decades on how to best perform such a characterization. How many samples should be taken? How should these samples be used in the design process?

The ground is one of the most complex of engineering materials, and yet is the most fundamental, in all senses of the word. While steel, concrete, and wood, for example, have fairly well established and relatively small uncertainties, the ground can vary by orders of magnitude from site to site, and even within a site.

As a result of the large uncertainty in the ground, all geotechnical designs must start with a geotechnical investigation so that the best "nominal" or "characteristic" ground parameters can be used in the design process. Traditionally, the intensity of the site investigation has not been particularly important, so long as a reasonable estimate of the characteristic design values can be estimated. However, recent impetus has been towards providing reasonable estimates of the reliability of designed geotechnical systems. In order to do so the ground used to provide the geotechnical resistance needs to be properly evaluated, in both the mean and the covariance.

In this paper, the ability of a soil sampling scheme to predict the actual mean, variance, and correlation length of the soil at a site is investigated. A key question is how does the number of samples affect the accuracy of the estimate? Or, put another way, how many samples are required to achieve a certain desired accuracy? The answer is found by considering a square site and using random field simulation to generate realizations of the soil properties over the site, sampling each realization, and then comparing the estimated mean, variance, and correlation length to the 'true' values. The goal here is to investigate the discrepancies between the estimated statistics and the true 'local' statistics, the latter obtained by sampling the field at all locations. Note that the 'local' statistics will differ from the population parameters, μ (mean), σ (standard deviation), and θ (correlation length), which are used by the random field generator, due to the fact that the local statistics are derived from a single realization. In detail, the soil is represented by a stationary Gaussian random field, $X(\mathbf{x})$, at spatial position $\mathbf{x}$, which is simulated within the domain and sampled at n_s locations. The samples are then used to estimate a mean trend, $\hat{\mu}(\mathbf{x})$, which can then be compared to the field realization to assess its ability to represent the actual mean trend. Defining the residual to be

$$X_r(\mathbf{x}) = X(\mathbf{x}) - \hat{\mu}(\mathbf{x}) \quad (1)$$

then $\hat{\mu}(\mathbf{x})$ is a good estimate of the mean trend if X_r is generally small. If the site is sampled at all locations, then $\hat{\mu}(\mathbf{x})$ can be taken to be equal to $X(\mathbf{x})$, in the event that a pointwise trend is assumed (as in Kriging), in which case $X_r(\mathbf{x}) = 0$ everywhere. Sampling at all locations is the best case since there is then minimum residual uncertainty (zero in the case of Kriging).

Sampling at all locations is, of course, prohibitively expensive and would also change the resulting field properties while measuring them (see, e.g., Heisenberg, 1927). In practice, soil properties are estimated from a relatively small number of samples so that $\hat{\mu}(\mathbf{x})$ will only ever approximate $X(\mathbf{x})$ in some way (i.e., via a trend).

In assessing the ability of $\hat{\mu}(\mathbf{x})$ to represent $X(\mathbf{x})$, it will also be useful to consider the average residual over the domain,

$$\bar{\mu}_r = \frac{1}{D\times D}\int_{D\times D} X_r(\mathbf{x})d\mathbf{x} = \frac{1}{n}\sum_{i=1}^{n}\left[X(\mathbf{x}_i)-\hat{\mu}(\mathbf{x}_i)\right] \qquad (2)$$

where D is the edge dimension of the $D\times D$ square domain. The domain is broken up into n cells in the simulation, resulting in the summation form on the right, in which $\mathbf{x}_i$ is the location of the center of the i'th cell.

The agreement between $\hat{\mu}(\mathbf{x})$ and $X(\mathbf{x})$ will be determined here by considering three measures; 1) the standard deviation of the residual field average, $\bar{\mu}_r$ (i.e., how well does the estimated trend represent the actual field average?), 2) the standard deviation of the residual, X_r (i.e. how much residual uncertainty remains?), and 3) the residual correlation length (i.e. how does the trend removal affect the perceived correlation lengths?).

Five sampling schemes are considered in the paper, ranging from a single sample taken at the field midpoint to nine samples taken over a 3 x 3 array at the quarter points of the field. In some cases a further 'maximum' sampling scheme is performed, where every point in the field is sampled, to see what the maximum attainable uncertainty reduction is.

For each sampling scheme, three types of trend removal are performed; a) removing the constant sample mean, b) removing a bilinear trend surface which is fit to the sample, and c) removing a Kriged surface fit to the sample. The residual statistics are determined by Monte Carlo simulation, with 2000 realizations for each case, where the field is discretized into 128 x 128 cells and the random fields generated using the Local Average Subdivision method (Fenton and Vanmarcke, 1990).

2 RESULTS

Consider first the average of the residual, $\bar{\mu}_r$, given by Eq. 2. It can be shown that the mean of $\bar{\mu}_r$ is zero, so that a measure of how accurately $\hat{\mu}(\mathbf{x})$ represents $X(\mathbf{x})$ can be obtained by looking at the standard deviation of $\bar{\mu}_r$ – small values of this standard deviation imply that $\hat{\mu}(\mathbf{x})$ remains close to the field average. Figure 1 illustrates how the standard deviation of $\bar{\mu}_r$, normalized by dividing by the standard deviation of the random field value, $X(\mathbf{x}_i)$, in the i'th cell (referred to as σ_{cell}), varies as a function of the number of samples taken from the domain, n_s, and the normalized correlation length, θ/D. Note that if only one sample is taken at the midpoint of the domain, $n_s = 1$, then a bilinear trend cannot be fit to the sample, nor is a Kriged surface removal attempted. Thus, parts b and c in Figure 1 do not have a curve corresponding to $n_s = 1$. In all plots it is apparent that as the number of samples increases, the accuracy improves (in agreement with the findings of Lloret-Cabot, et al., 2012). It can be seen, however, that for $n_s = 3$ to 9, there is very little difference between the detrending methods, so far as the field average is concerned. It is to be noted that the field average is a constant, not a trend, so it is not expected that the bilinear and Kriged surface trends will do any better than the sample mean, when compared to the field average.

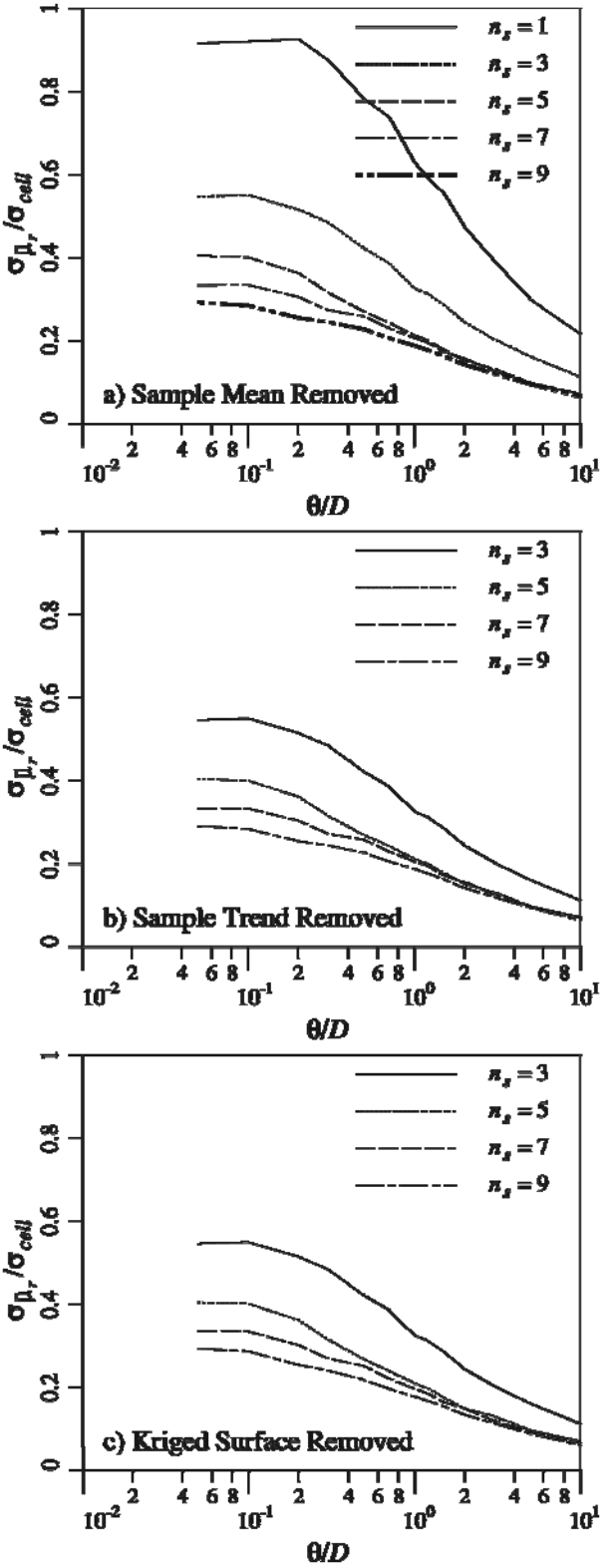

Figure 1. Standard deviation of the field average residual (eq. 2), normalized by the standard deviation of X, versus normalized correlation length.

In all cases in Figure 1, the agreement between $\hat{\mu}(\mathbf{x})$ and $X(\mathbf{x})$ improves as the correlation length increases. This is because the field becomes increasingly smooth, or flat, as the correlation length increases, so that all trends considered become closer to the flatter $X(\mathbf{x})$.

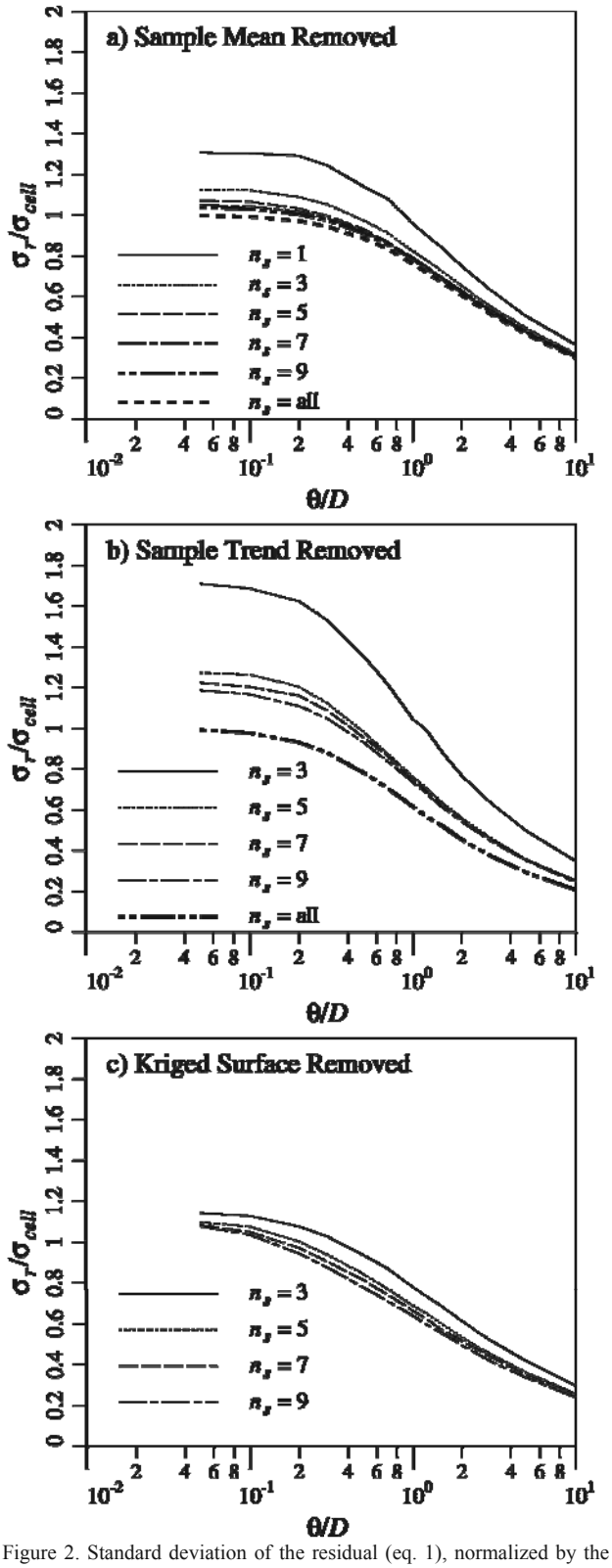

Figure 2. Standard deviation of the residual (eq. 1), normalized by the standard deviation of *X*, versus normalized correlation length.

A possibly better measure of how well $\hat{\mu}(\mathbf{x})$ represents the field is obtained by considering the standard deviation of the residual, $X_r(\mathbf{x})$ (see eq. 1), directly. This measure will include the effects of trend removal and is illustrated in Figure 2, again with the standard deviation of the residual, σ_r, divided by the standard deviation of X, σ_{cell}. In detail, the standard deviation of the residual is estimated as the square root of the variance,

$$\sigma_r^2 \approx \frac{1}{n-1}\sum_{i=1}^{n}\left[X(\mathbf{x}_i)-\hat{\mu}(\mathbf{x}_i)\right]^2 \qquad (3)$$

for each realization. The value of σ_r used in Figure 2 is averaged over all realizations. As in Figure 1, the $n_s = 1$ case only appears in Figure 2a, since bilinear trend and Kriging surfaces are not well defined for only one sample point. However, Figures 2a and b now include a limiting case where the entire simulation has been sampled ($n_s =$ all), representing the best site knowledge possible. This case was not included in Figure 1 since, when all values are sampled, $\overline{\mu}_r = 0$, that is, the average residual is zero. In Figure 1, this would have corresponded to a horizontal line at zero standard deviation. In Figure 2, the '$n_s =$ all' case corresponds to the classical case where both the estimated mean (trend) and the variance are computed from the same set of observations. As the correlation length decreases, these observations become increasingly independent, and the estimated standard deviation approaches the true standard deviation, so that $\sigma_r / \sigma_{cell} \to 1.0$ as seen in Figures 2 a and b when $n_s =$ all. In Figure 2 c, the case '$n_s =$ all' is not included in the Kriging surface case since, when the entire field is sampled, the residual is zero with zero variability, and so the curve corresponding to this case lies at zero.

As in Figure 1, Figure 2 also shows that the ability of $\hat{\mu}(\mathbf{x})$ to represent $X(\mathbf{x})$ improves as the correlation length increases, for all of the trends considered. In the limit, as $\theta / D \to \infty$, all random fields become uniform (under the assumed finite variance correlation structure), random from realization to realization, but constant within each realization. In this limiting case, the sample perfectly predicts the uniform field, and the residual becomes zero everywhere so that $\sigma_r = 0$. It is apparent in Figure 2 that all curves are heading towards 0, as $\theta / D \to \infty$.

One of the perhaps surprising results of Figure 2 is that the removal of a bilinear trend is not generally as good as the removal of the constant sample mean at smaller correlation lengths, and especially at a lower number of samples. The reason for this becomes apparent when, for example, the case where $n_s = 3$ is considered. If the correlation length is small, then the three samples will be largely independent, and the resulting fitted bilinear plane could (and often does) end up with quite an unrepresentative slope, leading to a high variability in the residual. Even when $n_s = 9$ the residual variability is higher at low correlation lengths than seen using the constant sample mean. At low correlation lengths, the Kriging surface performs about the same as the constant sample mean.

At large correlation lengths, e.g. $\theta / D = 10$, the bilinear trend performs better than the constant sample mean for all n_s except $n_s = 3$, where the relative standard deviation is 0.35 versus 0.32 for the constant sample mean. For higher number of samples, the relative standard deviation using the bilinear trend is 0.25, versus 0.31 for the constant sample mean. The Kriged surface performs the best out of the three methods (relative standard deviation of 0.30) when the number of samples is 3, and about the same as the bilinear trend for higher numbers of samples.

The last measure of the quality of the trend type used considered in this paper is how well the estimated correlation length agrees with the actual correlation length, Figure 3. Once $\hat{\mu}(\mathbf{x})$ has been established from the soil samples, the correlation length is estimated here using the following steps;

1. for each direction through the soil domain, $i = 1, 2$,
2. estimate the semi-variogram along all lines through the domain in direction i using the entire $X_r(\mathbf{x})$ field,
3. average the semi-variograms obtained in step 2 to obtain the final semi-variogram estimate in direction i,
4. fit a theoretical semi-variogram, having parameter θ (correlation length), to the semi-variogram estimated in step 3 by minimizing the sum of squared errors (i.e. regression).

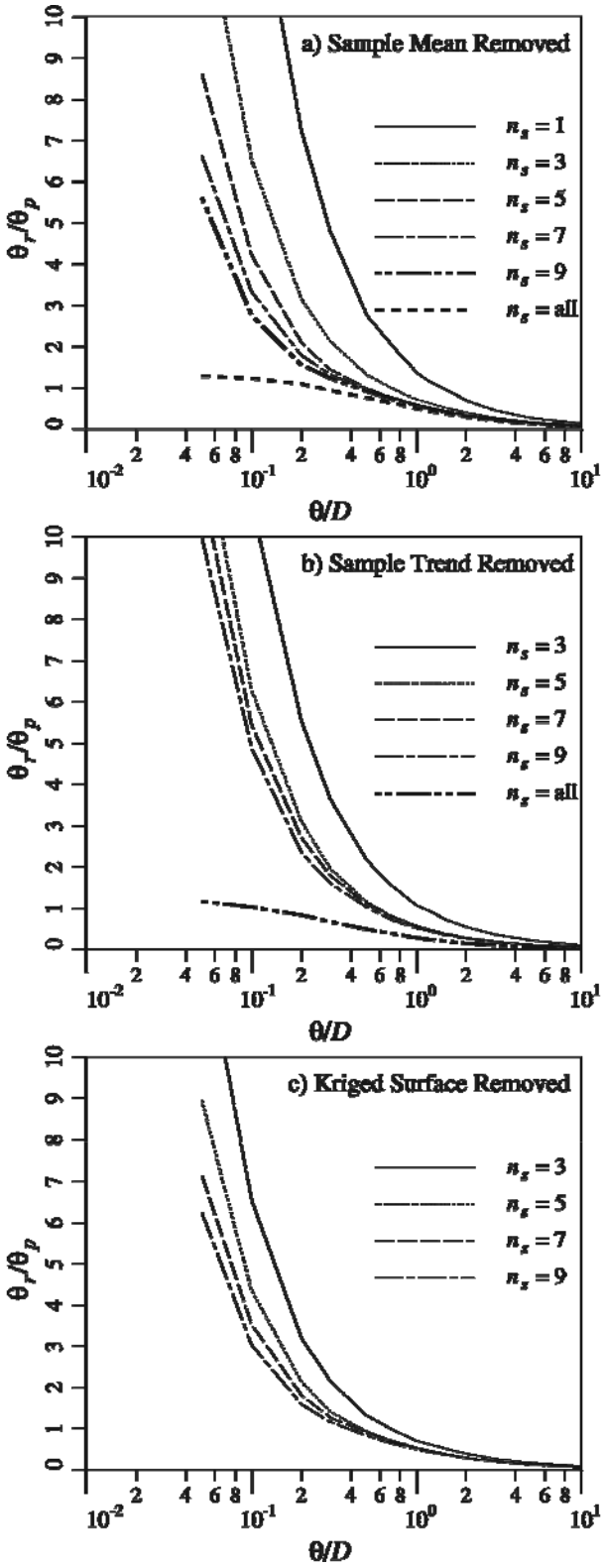

Figure 3. Estimated correlation length of the residual, normalized by the point correlation length, versus normalized actual correlation length.

The correlation length estimated from the residual, θ_r, will agree with the actual correlation length used in the simulation, θ, when the ratio $\theta_r / \theta = 1$. It can be immediately seen in Figure 3 that this only occurs in general when the entire field is sampled and the correlation length is relatively small (i.e. significantly less than D). That is, when the entire field is sampled ($n_s =$ all), so that the sample average is equal to the actual field average, the estimated correlation length becomes equal to the actual correlation length when the samples are relatively independent (small θ).

In general, when $\theta < D$ the estimated correlation length is overestimated, and often considerably overestimated, especially when the actual correlation length is small. This occurs because errors between the estimated trend (of any of the three types) and actual bilinear field trend (bilinear because correlation is a measure of the degree of linear dependence between random variables) are perceived in the estimation process to be caused by a strong lingering correlation (and not by an error in the original trend estimate) – hence a longer correlation length is estimated to account for the evident residual trend.

Of the three trend types considered, the best is the constant sample mean and the worst is the bilinear trend (except when $n_s =$ all). The Kriged surface is slightly worse than the constant sample mean. For example, when $n_s = 9$ and $\theta / D = 0.05$, then $\theta_r / \theta = 5.6$, 10.0, and 6.3 for the constant sample mean, bilinear trend, and Kriging surface, respectively. It should be noted that the best performer, the constant sample mean, may be so only because the simulated field is assumed stationary (i.e. constant mean).

At the other end of the plot, where $\theta > D$, the correlation length is underestimated ($\theta_r / \theta < 1$). In general, this is because the removal of a trend in a strongly correlated field is also removing the evidence of the strong correlation (strong correlation is evidenced by a trend having little variation off the trend) resulting in a residual field without strong correlation – hence a small correlation length. Of the three trend types considered the best performer at the large correlation length end is again the constant sample mean. For example, when $n_s = 9$ and $\theta / D = 10$, then $\theta_r / \theta = 0.08$, 0.05, and 0.06 for the constant sample mean, bilinear trend, and Kriging surface, respectively.

3 CONCLUSIONS

There is no difference between the accuracies of the trend type selected when matching the trend to the field average, $\bar{\mu}_r$. As expected, the accuracy improves as the number of samples and the correlation length increase. If a target standard deviation, $\sigma_{\bar{\mu}_r}$, equal to 20% of the random field standard deviation, σ_{cell}, is desired, then only one sample is required if $\theta / D \geq 10$, while 9 or more samples are required if $\theta / D \leq 1$.

In general, if the correlation length is small, the most accurate approach is to use a constant sample mean, which shows the best general results for all three measures of accuracy considered in this paper. Kriging is almost identical, only losing out slightly when considering the residual estimated correlation length. At the other end of the scale, when the correlation length is large, the bilinear trend is more accurate with respect to the residual standard deviation than is the constant sample mean, as expected.

In the absence of knowledge about the actual correlation length, it appears that the Kriging surface removal, although not generally the best in any one measure, is very competitive and is certainly a good overall choice.

4 REFERENCES

Fenton G.A. and Vanmarcke, E.H. 1990. Simulation of Random Fields via Local Average Subdivision, *ASCE Journal of Engineering Mechanics*, 116(8), 1733 – 1749.

Heisenberg W. 1927. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik, *Zeitschrift für Physik*, 43(3-4), 172 – 198.

Lloret-Cabot, M., Hicks, M.A., and Van Den Eijnden, A.P. 2012. Investigation of the reduction in uncertainty due to soil variability when conditioning a random field using Kriging, *Géotechnique Letters,* 2, 123 – 127.