Statistical Sample Size for Quality Control Programs of Cement-Based Solidification/Stabilization

Gordon A. Fenton[1], Rukhsana Liza[2], Craig B. Lake[3], W. Todd Menzies[4] and D.V. Griffiths[5]

[1]*Department of Engineering Mathematics, Dalhousie University, Halifax, Canada: Phone: (902) 494-6002, Fax: (902) 423-1801, e-mail: Gordon.Fenton@dal.ca*

[2]*Department of Civil and Resource Engineering, Dalhousie University, Halifax, Canada: Phone: (902) 293-5492, Fax: (902) 494-3108, e-mail: rk962901@dal.ca*

[3]*Department of Civil and Resource Engineering, Dalhousie University, Halifax, Canada: Phone: (902) 494-3220, Fax: (902) 494-3108, e-mail: Craig.Lake@dal.ca*

[4]*Stantec Consulting Limited, Dartmouth, Canada: Phone: (902) 468-7777, Fax: (902) 468-9009, e-mail: todd.menzies@stantec.com*

[5]*Department of Civil and Environmental Engineering, Colorado School of Mines, Golden, Colorado, USA: Phone: (303) 273-3669, Fax: (303)273-3602, e-mail: D.V. Griffiths@mines.edu*

Statistical Sample Size for Quality Control Programs of Cement-Based Solidification/Stabilization

Rukhsana Liza, Gordon A. Fenton, Craig B. Lake, William T. Menzies and D.V. Griffiths

Abstract

Sampling requirements for the quality control (QC) of cement-based solidification/stabilization (S/S) construction cells do not currently specify the sample size considering the accuracy of the estimated effective hydraulic conductivity of the cells from the samples, nor considering the risk associated with drawing the wrong conclusions about the acceptability of the cells. In this paper, probabilistic simulations are performed to examine the influence of a soil-cement material's mean, variance, and correlation length on sampling requirements for a QC program of cement-based S/S construction cells. The sampling requirements are determined by considering a hypothesis test, having null that the constructed material is unacceptable, and targeting acceptable probabilities of making an erroneous decision. Two types of errors can be made: 1) concluding that the material is acceptable when it actually is not, or 2) failing to conclude that the material is acceptable when it actually is. The paper investigates how many samples are required in order to keep the probabilities of making these errors acceptably small. Plots are provided which can be used to estimate required number of samples. The paper concludes by discussing how the simulation-based results compare to current sampling requirements for the QC of an actual set of cement-based S/S construction cells.

Key words: geotechnical quality control, sampling error, hypothesis test, groundwater contamination, solidification, stabilization

Introduction

Cement-based solidification/stabilization (S/S) is a source-controlled remediation technology in which cement is mixed with contaminated media, such as contaminated soil, sediment, sludge, and industrial waste, to minimize the migration of contaminants and thereby to limit the contamination of ground and/or surface waters. A major concern with such remediation efforts is to decide how reliable the efforts have been at meeting performance objectives set by a regulator. To assess the success of a cement-based S/S, a quality control (QC) program is typically undertaken which involves sampling the site to estimate its final effective hydraulic conductivity. If the sample suggests that the final effective hydraulic conductivity is sufficiently small, the cement-based S/S is considered to be successful. Such QC programs are usually performed by dividing the entire S/S site into a number of cells over the plan area (which will be referred to here as *construction cells*) and sampling each cell individually. The subdivision into a sequence of cells allows each cell to be assessed individually, which reduces the expense of additional remediation/replacement in the event that the effective hydraulic conductivity is suggested to be too high and thus unacceptable – only the unacceptable cell needs to be further remediated or replaced. Since the further remediation of unacceptable cells can be quite expensive, it is important that the cell sampling scheme be suitably accurate to avoid both 1) ground and/or surface water contamination by missing unacceptable construction cells and 2) having to further remediate construction cells deemed to be unacceptable but which are actually acceptable. The goal of this paper is to determine the number of samples required to allow a QC program to properly minimize the probabilities of the negative outcomes of water contamination and/or unnecessary further remediation costs.

Attention is restricted in this paper to S/S sites where contaminant migration occurs predominately via advection in the horizontal direction (i.e. no diffusion) and where the contaminated layer thickness is small relative to its areal extent. This allows the site to be modeled as two-dimensional, which basically means that the soil properties over the soil layer thickness are taken to be constant. Traditionally, the equation governing the total advective flow, Q, through a saturated S/S construction cell is given by Darcy's law as follows,

[1] $$Q = k_{eff} i A$$

where k_{eff} is the effective hydraulic conductivity of the construction cell, i is the hydraulic gradient across the cell and A is the area perpendicular to the direction of flow. The effective hydraulic conductivity, k_{eff}, is defined as the single value of hydraulic conductivity which yields the same total flow through the cell as does the actual spatially varying hydraulic conductivity field (see Fenton and Griffiths, 1993). In order to ensure that the construction cell will perform effectively in restricting contaminant migration via advection, samples are collected and tested during construction in a QC program to estimate the cell's effective hydraulic conductivity. If the estimated effective hydraulic conductivity is less than or equal to the regulatory hydraulic conductivity, k_{crit}, then the construction cell is considered to be acceptable. Otherwise it is deemed unacceptable and must be repaired or replaced. The question is: How many samples should be taken in order to reliably make this decision? In common practice samples are collected based on the sample density method (USACE 2000), which requires a certain number of samples per unit volume. The number of samples required by the USACE Method is independent of the statistics of the sampled field and makes no attempt to assess the probability of making an error in deciding about the acceptability of a cell. Since different levels of spatial variability of a constructed S/S system will certainly affect the accuracy of the estimate of k_{eff} , it is clear that using a fixed number of samples (e.g. per unit volume) will result in different decision error probabilities as the level of spatial variability of the hydraulic conductivity field changes. This paper aims to examine the influence of the mean, variance, and correlation length of a cell's hydraulic conductivity field on the number of samples required to achieve acceptably small decision error probabilities.

Random fields are commonly used to model spatially variable engineering properties (Fenton and Griffiths 2008) and they will be used here to model the hydraulic conductivity field. The sampling problem will be investigated by simulating possible realizations of the 2-D hydraulic conductivity field, virtually sampling each realization at selected locations and then deciding whether the realization is acceptable or not on the basis of the sample results. An error in the decision is made if either the cell is deemed to be acceptable when it is not (Type I error), or if the cell is deemed to be unacceptable when it is actually acceptable (Type II error). As will be shown, the probability of making a decision error reduces as the number of samples increases, not surprisingly, and the task is to determine just how many samples are required to reduce the error probabilities to acceptable levels, which will be assumed to be 5% in this paper.

The random conductivity field realizations will be simulated using a method called Local Average Subdivision (LAS) (Fenton and Vanmarcke 1990). The LAS algorithm preserves the spatial correlation, over the ensemble, between local averages of the property. Correlation between points can be characterized by a parameter called the correlation length, θ, which is the distance within which the property of interest is significantly correlated and beyond which is largely uncorrelated. One of the motivations of LAS arises from the fact that instead of point measurements, engineering properties are usually measured over some volume, thus representing the average property over that volume. Thus, LAS directly simulates realizations of such 'local' averages. Local averaging reduces the variance of the random field. In the 2-D model considered here, the final variance depends on the area selected for local averaging, decreasing as the local averaging area increases (Fenton and Griffiths 2008). Further details regarding the correlation structure and variance reduction used in the random field model can be found in the "Parametric Study" section of this paper.

Research relating to the sampling requirements for a QC program of cement-based S/S construction cells is not available in literature, so far as the authors are aware. Some research has been conducted on the sampling requirements for soil liner systems, which is similar to the requirements for cement-based S/S construction cells, as discussed next.

Benson et al. (1994) presented a method to select the number of samples that should be collected and tested during the construction of compacted soil liners in order to ensure reliable liners at some confidence level. Not surprisingly, they found that the accuracy of the estimate increases as the sample size increases and also showed that samples should be collected at higher frequency for soils having highly variable hydraulic properties as well as for soils with mean hydraulic conductivity close to the regulatory value. In their investigation, simulations were performed using a three-dimensional stochastic model with varying hydraulic conductivity mean, variance, and liner thickness. However, they did not explicitly consider the random field nature of the liner, that is independence between adjacent elements in their model was assumed for simplicity, i.e., they ignored the correlation between hydraulic conductivity values.

Menzies (2008) examined the influence of the correlation length on sampling requirements of soil liner systems in order to achieve target reliability against excessive flow through the liner. Influences of the hydraulic conductivity mean and variance on sampling requirements were investigated using a two-dimensional stochastic model to perform simulations. In Menzies' study, two types of hypothesis test errors were considered, i.e., Type I where the sample data led to the conclusion that the liner was acceptable when it was not, and Type II where the sample data suggested that the liner was unacceptable when it actually was acceptable. It was found that a "worst case" correlation length existed, which was about 5%-10% and 2%-3% of the liner size in any direction, that maximized the probabilities of Type I and Type II errors, respectively. Menzies (2008) also found that for a particular sample size, both types of error probabilities reached a maximum value when the mean hydraulic conductivity of the liner was close to the regulatory value, requiring more samples in this case to achieve the same reliability as obtained when the mean hydraulic conductivity is farther away from the regulatory value. In his stochastic model, Menzies used the arithmetic average of the hydraulic conductivity field to be the effective hydraulic conductivity. He also assumed the correlation structure to be isotropic. This work extends that of Menzies' to a case where the flow is in-plane so that geometric averaging is required.

Background on Sampling Theory

The overall objective of QC sampling of cement-based S/S construction cells is to ensure that the cell will be acceptable, i.e., that its effective hydraulic conductivity, k_{eff}, will be less than the regulatory hydraulic conductivity, k_{crit}. The decision about whether a construction cell is acceptable is made on the basis of a set of samples taken from the cell. This decision making process is essentially a hypothesis test where the null hypothesis (H_0) is that the cell is unacceptable, so that the burden of proof is on showing that the alternative hypothesis (H_a) is true, at an appropriate level of confidence.

[2]
$$\begin{aligned} H_o &: k_{eff} \geq k_{crit} \\ H_a &: k_{eff} < k_{crit} \end{aligned}$$

As mentioned previously, two types of errors may result in making this decision about the acceptability of the cell. These are 1) concluding that the S/S construction cell is acceptable when it is not (Type I), or 2) failing to conclude that the S/S construction cell is acceptable when it actually is (Type II). The challenge is to determine how many samples should be collected to ensure that the probability of making either type of error will be acceptably small.

Taking an infinite number of samples from the construction cell will eliminate any chances of making a decision error, but this is neither physically nor economically feasible. This means that some chance of error will always exist and so it is necessary to relate the error probabilities with the number of samples taken in order to determine the number of samples required.

Analytical results exist for the sample size required to ensure that the probabilities of Type I and II errors are sufficiently small (see, e.g., chapter 8 of Devore, 2008). These results, however, assume that the samples are independent. Since the construction cell hydraulic conductivity values are generally correlated, existing analytical results cannot be used to determine required sample sizes for the quality control of construction cells. The goal of this paper is to investigate how the probabilities of Type I and Type II errors change as a function of the number of samples within construction cells.

Probabilistic Simulations

The construction cells investigated in this paper are designed to provide a barrier against horizontal flow and are thin (vertically) relative to their planar dimension, as shown in Figure 1. Because the cell is relatively thin, the flow is largely in the plane and a two-dimensional flow model is acceptably accurate. Since a two-dimensional flow model is also much faster, computationally, than a three-dimensional model, the two-dimensional model will be used here. It is to be noted, however, that the resulting model can therefore only investigate the sampling requirements per unit area, not per unit volume. As will be seen later, this leads to some difficulties in comparing recommendations here to current practice.

The hypothesis test problem is studied here using Monte Carlo simulations employing a modified version of the two-dimensional random finite element method (RFEM) program, mrflow2d (Fenton and Griffiths 2008). The original program was designed to analyze stochastic fluid flow problems and is described in Fenton and Griffiths (1993). The program is modified in this study to enable the sampling of the random field at prescribed locations. The mesh discretization used in the simulations is as shown in Fig.1.

The flow regime assumes that an impervious boundary exists on the top and bottom, and on the left and right, faces of Fig.1. A uniform unit pressure head was applied on the front face which directs the flow, Q, in the x direction. The inputs to the model are the mean and standard deviation of point-scale hydraulic conductivity, correlation lengths (assumed isotropic), the number of elements in each direction, the element size, and the number and locations of the samples to be taken. Given these inputs, the RFEM model generates a random field of log-normally distributed hydraulic conductivity. The steps followed in the simulations are as follows:

1. Given the mean, standard deviation and correlation length of the hydraulic conductivity at the point-scale, generate a realization of the local averages, G_i, for $i = 1,2,......,m$, where m is the specified number of elements in the model, using the Local Average Subdivision (LAS) algorithm (Fenton and Vanmarcke 1990). Each local average, G_i, is the arithmetic average of a standard Gaussian field, G over the ith element.
2. The lognormally distributed hydraulic conductivity value, k_i, is assigned to the ith element through the transformation $k_i = \exp\{\mu_{\ln k} + \sigma_{\ln k} G_i\}$, where $\mu_{\ln k}$ and $\sigma_{\ln k}$ are the mean and standard deviation of the logarithm of k obtained from the specified mean and standard deviation μ_k and σ_k via the transformations:

[3a] $$\sigma_{\ln k}^2 = \ln\left(1 + v_k{}^2\right)$$

[3b] $$\mu_{\ln k} = \ln \mu_k - \frac{1}{2}\sigma_{\ln k}^2$$

where $v_k = \sigma_k / \mu_k$ is the coefficient of variation.

3. Sample the field at the specified element locations. This is done simply by recording the value of k_j generated for the j'th sampled element. Measurement error is assumed to be zero.
4. Compute the geometric average, k_G, of the sample and the effective hydraulic conductivity of the entire conductivity field, k_{eff} as follows,

[4] $$k_G = \exp\left\{\frac{1}{n}\sum_{j=1}^{n}\ln k_j\right\}$$

[5] $$k_{eff} = \exp\left\{\frac{1}{m}\sum_{i=1}^{m}\ln k_i\right\}$$

where

$n =$ number of samples taken from the random field,

$k_j =$ hydraulic conductivity of the j th sampled element of the random field,

$m =$ number of elements of the random field, and

$k_i =$ hydraulic conductivity of the i th element of the random field.

Fenton and Griffiths (1993) demonstrated that the geometric average was the best estimate of the effective hydraulic conductivity for relatively square flow regimes, where the effective hydraulic conductivity was defined by them to be the single value of hydraulic conductivity which yields the same total flow through the cell as does the actual spatially varying hydraulic conductivity field. Hence, geometric averages of the element hydraulic conductivities and the samples are used to obtain the actual and the predicted effective hydraulic conductivity of the random field, respectively. In other words, the effective hydraulic conductivity, k_{eff} , used in this paper, closely approximates the uniform (spatially constant) hydraulic conductivity value which yields the same total flow as computed through the actual spatially random hydraulic conductivity field. If $k_{eff} > k_{crit}$ then the total flow through the cell exceeds the regulatory limit and the cell is unacceptable.

The geometric average, k_G , is the sample estimate of the effective hydraulic conductivity, k_{eff}. If $k_G < k_{crit}$ then the cell is *deemed* to be acceptable, even though it may not be (Type I error). Alternatively, if $k_G > k_{crit}$ then the cell is deemed to be unacceptable, even though it may actually be acceptable (Type II error). For each realization, the sample geometric average, k_G, and the effective hydraulic conductivity, k_{eff} are compared to the regulatory hydraulic conductivity, k_{crit}. This comparison results in one of the following four outcomes being recorded for each realization:

- Both k_G and the actual effective hydraulic conductivity of the random field are below the regulatory value $\left(k_G < k_{crit} \cap k_{eff} < k_{crit}\right)$. This is a favorable outcome.

- Both k_G and the actual effective hydraulic conductivity of the random field are above the regulatory value $\left(k_G > k_{crit} \cap k_{eff} > k_{crit}\right)$. This outcome will result in the cell being deemed to be unacceptable but is a favorable outcome since it is predicted by the sample.
- k_G is less than the regulatory hydraulic conductivity, while the actual effective hydraulic conductivity of the field exceeds the regulatory value $\left(k_G < \mathrm{k}_{crit} \cap \mathrm{k}_{eff} > k_{crit}\right)$. This is an unfavorable Type I error (cell is assumed acceptable when it is not) resulting in the worst outcome of this hypothesis test, where an unsafe cell is deemed to be safe.
- k_G is greater than the regulatory value, while the actual effective hydraulic conductivity of the field is less than the regulatory value $\left(k_G > k_{crit} \cap k_{eff} < k_{crit}\right)$. This is an unfavorable Type II error (cell is assumed unacceptable when it is actually acceptable) which would require some unnecessary work, such as excavating the treated material and reapplication of the S/S process for the construction cell, resulting in a higher project cost.

Of the two types of errors, the Type I error is the worst from an environmental protection standpoint since it results in an unacceptable cell being accepted. The above steps are repeated over n_{sim} realizations for each parameter set (as discussed in the next section) to estimate the probabilities of Type I (p_1) and Type II (p_2) errors, according to:

[6] $$p_1 = \frac{n_1}{n_{sim}}$$

[7] $$p_2 = \frac{n_2}{n_{sim}}$$

where n_1 is the number of realizations where $k_G < k_{crit}$ while $k_{eff} > k_{crit}$, n_2 is the number of realizations where $k_G > k_{crit}$ while $k_{eff} < k_{crit}$, and n_{sim} is the total number of realizations considered.

Parametric Study

In order to enable the results to be scaled to any desired regulatory hydraulic conductivity, k_{crit} , the mean of the point-scale hydraulic conductivity of the input distribution, μ_k , and the effective hydraulic conductivity, k_{eff} , can be normalized by the regulatory hydraulic conductivity, k_{crit}.

[8] $$\mu'_k = \frac{\mu_k}{k_{crit}}$$

[9] $$k'_{eff} = \frac{k_{eff}}{k_{crit}}$$

where μ'_k is the normalized mean hydraulic conductivity and k'_{eff} is the normalized effective hydraulic conductivity.

The correlation length, $\theta_{\ln k}$, can also be non-dimensionalized by dividing by the effective dimension of the construction cell, D, where $D = \sqrt{XY}$ and X and Y are the planar dimensions of the construction cell;

[10] $$\theta'_{\ln k} = \frac{\theta_{\ln k}}{D}$$

Non-dimensionalizing the correlation length allows the results to be scaled to any construction cell size so long as it has same (or similar) aspect ratio (X/Y) as used in this study, which is $X/Y = 1$.

Parametric variations considered in the simulations are presented in Table 1.

Table 1: Parametric variations considered in the simulations

Parameter	Variation
Normalized mean hydraulic conductivity, μ'_k	0.01 to 10.0.
Coefficient of variation, $v_k = \sigma_k / \mu_k$	0.1, 1.0, 2.0, and 5.0.
Normalized correlation length, $\theta'_{\ln k}$	0.01, 0.05, 0.1, 0.5, 1.0, 5.0, and 10.0.
Number of samples, n	1, 4, 9, 16, 25, and 49 (see Fig 2)

The lognormally distributed random hydraulic conductivity field is fully specified by its mean, its variance, and its correlation structure. In this study, the correlation between pairs of $\ln k$ values is assumed to be Markovian having the following separable correlation function (which is a product of two directional correlation functions – see, e.g., Vanmarcke 1984, for more details.),

$$\rho_{\ln k}(\tau_1, \tau_2) = \exp(-2|\tau_1|/\theta_1)\exp(-2|\tau_2|/\theta_2) \qquad [11]$$

in which τ_i is the distance between points in each coordinate direction, $i=1$ and 2. The decay rate parameters θ_i, for $i=1$ and 2, are the directional correlation lengths. In this study, the correlation lengths are assumed to be equal; $\theta_1 = \theta_2 = \theta_{\ln k}$.

Since the correlation function is separable, its corresponding variance reduction function (see Vanmarcke 1984) is also separable and can be written explicitly as the product:

$$\gamma_{\ln k}(X,Y) = \gamma(X)\gamma(Y) \qquad [12]$$

where

$$\gamma(X) = \frac{\theta_{\ln k}^2}{2X^2}\left[\frac{2|X|}{\theta_{\ln k}} + \exp\left\{\frac{-2|X|}{\theta_{\ln k}}\right\} - 1\right] \qquad [13]$$

and similarly for $\gamma(Y)$.

Regarding the finite element model, a sensitivity analysis was performed in order to examine the influence of the element size on the output quantities of interest (i.e., the probabilities of Type I and Type II errors). A domain of size (1×1) was discretized into 32×32, 64×64, 72×72, 80×80, 88×88, and 128×128 elements. All mesh resolutions gave similar results for the approximately 'worst case' correlation length (see discussion below) of $\theta'_{\ln k} = 0.5$ and using $n_{sim} = 25000$. For example, Type I error probabilities ranged from 0.0239 at a resolution of 32×32 to 0.0230 at a resolution of 128×128. Some of the intermediate mesh resolutions actually yielded higher discrepancies. For example, the 88×88 resolution yielded a Type I error probability of 0.0262, which was 14% higher than that found at the 128×128 resolution. Since a sample size of $n_{sim} = 25000$ results in a standard error on the estimated error probability of about 4% (see below), it is believed that the high probability given by the 88×88 resolution field is an outlier, due to an unresolved modeling problem. All of the other resolutions were within 3% of the 128×128 resolution. The effect of mesh resolution on the estimated probabilities of Type II errors was very similar. The Type II error probability was estimated to be 0.09472 for the 32×32 mesh and 0.09468 for the 128×128 mesh. Ignoring the 88×88 mesh results, all other results were within 3% of the 128×128 mesh results. Based also on reasonable computing time, a 64×64 element density was selected for all simulations. The number of realizations selected was $n_{sim} = 25000$ for all parameter sets considered. This means that the standard deviation of any probability estimate is $\sqrt{\hat{p}(1-\hat{p})/n_{sim}}$, where $\hat{p}$ is the estimated probability, which, for small $\hat{p}$ is approximately $0.0063\sqrt{\hat{p}}$. In other words, the Monte Carlo simulation can reasonably accurately estimate $\hat{p}$ down to about 1/10000.

Results

Influence of Correlation Length on Error Probabilities

It is instructive to first consider the probabilities of Type I and II errors at the limiting values of $\theta_{\ln k}$. At the lower limit, when $\theta_{\ln k}$ is equal to 0, points within the field will have no correlation with each other, which means that the $\ln k$ field is white noise (Fenton and Griffiths 2008). In this case, any local average of $\ln k$ will consist of an infinite number of independent values whose average is a non-random constant (equal to the median) so that one (local average) sample is sufficient to completely specify the effective hydraulic conductivity of the field. That is, the probability of making any type of error (i.e., either Type I or Type II) will be zero on the basis of one or more samples if $\theta_{\ln k} = 0$. At the other extreme, when $\theta_{\ln k} \rightarrow \infty$, points within the random field are perfectly correlated with each other which means that they are all equal if the field is stationary, as assumed here. In this case, the field can be represented by a single (random) hydraulic conductivity value so that one sample is sufficient to predict the actual effective hydraulic conductivity of the entire field, resulting in error probabilities again being equal to 0. At intermediate correlation lengths (i.e., between zero and infinity), the probabilities of Type I and II errors are non-zero and will be affected by the number of samples taken – fewer samples will result in larger error probabilities.

Figure 3 shows the influence of the normalized correlation length on the probability of a Type I error for different numbers of samples ($n =$ 1, 4, 9, 16, 25, and 49) for $\mu_k' = 1.0$ and $v_k = 1.0$. Each point on the plot is obtained using 25000 realizations and indicates that, for given number of samples, as the correlation length increases the probability of a Type I error at first increases and then decreases, as expected. For example, when $\mu_k' = 1.0$, $v_k = 1.0$, and $n = 4$, the probability of a Type I error increases from close to 0 at a normalized correlation length of 0.1 to a maximum value of 0.036 at a normalized correlation length of 1.0, and then decreases to 0.019 when the normalized correlation length reaches 10.0. The probability continues to decrease thereafter to 0 as $\theta_{\ln k}' \rightarrow \infty$ (not shown). The highest error probability occurs at a "worst case" correlation length, in this case at about $\theta_{\ln k}' = 1.0$. Since the actual correlation length is rarely, if ever, known at any site, the practical importance of the existence of a "worst case" correlation length is that it can be used to produce sampling plans which are conservative, that is, guaranteed to have error probabilities no higher than specified in the sampling design.

Figure 3 also shows that for given correlation length, the probability of a Type I error decreases as the number of samples increases. For example, when $\mu'_k = 1.0, v_k = 1.0$ and $\theta'_{\ln k} = 0.5,$ the probability of a Type I error decreases from 0.032 when $n = 4$ to 0.009 when $n = 49$.

Figure 4 illustrates the influence of the normalized correlation length on the probability of a Type II error for various numbers of samples ($n = 1$, 4, 9, 16, 25, and 49) for $\mu'_k = 1.0$ and $v_k = 1.0$. Similar to Fig. 3, a "worst case" correlation length occurs at an intermediate correlation length, in this case at around 10% to 50% of the field dimension. For example, when $\mu'_k = 1.0$, $v_k = 1.0$, and $n = 4$, the probability of a Type II error starts at 0.03, increases to 0.18, and then drops back down to 0.02 for $\theta'_{\ln k} =$0.01, 0.1, and 10.0, respectively.

Figure 4 also shows that an increase in the number of samples decreases the probability of a Type II error. For example, when $\mu'_k = 1.0, v_k = 1.0$ and $\theta'_{\ln k} = 0.5,$ the probability of a Type II error decreases from 0.15 when $n = 4$ to 0.03 when $n = 49$. The converging nature of the plots on both sides of the worst case indicates that at very low and high correlation lengths, the probability of a Type II error tends to 0, which is as expected.

Similar trends to those shown in Figs. 3 and 4 are seen for all other parameter set combinations considered and so are not repeated here. The "worst case" correlation lengths occur at about 1 to 5 times the field dimension for Type I errors and at about 0.1 to 10 times the field dimension for Type II errors. In general, the "worst case" correlation is somewhere between 0.1 and 1.0 times the field dimension. If it is more important to minimize the probability of committing a Type I error, then choosing the correlation length to be equal to the field dimension would be appropriate. For most of the following comparisons, an intermediate worst case correlation length of $\theta'_{\ln k} = 0.5$ has been selected.

Influence of Mean on Error Probabilities

When the mean hydraulic conductivity of the random field is much less than the regulatory hydraulic conductivity, both the effective hydraulic conductivity and the sample geometric average will almost always be less than the regulatory value so that the probabilities of Type I and II errors will be small. Similarly, when the mean hydraulic conductivity is much higher than the regulatory value, both the effective hydraulic conductivity and the sample geometric average will almost always be higher than the regulatory value so that, again, the probabilities of Type I and II errors will be small. The highest decision error probabilities will occur when the mean hydraulic conductivity is close to the regulatory value. Figures 5 and 6 illustrate the influence of the mean on the probabilities of Type I and Type II errors, respectively, for $v_k = 1.0$, $\theta'_{\ln k} = 0.5$, and $n = 4$, 16 and 49. For given number of samples, the highest probability of a Type I error in Fig. 5 occurs when the mean hydraulic conductivities are about 1.7 times the regulatory value. For example, in the case where $v_k = 1.0$, $\theta'_{\ln k} = 0.5$ and $n = 4$, the probability of a Type I error reaches a maximum of about 0.15 when $\mu'_k \approx 1.7$.

Similarly the highest probabilities of a Type II error (Fig. 6) are observed when $\mu'_k \approx 1.1$. For example, for $\nu_k = 1.0$, $\theta'_{\ln k} = 0.5$, and $n = 4$, the probability of a Type II error reaches a maximum of about 0.15 when $\mu'_k = 1.1$.

Influence of Coefficient of Variation on Error Probabilities

Figures 7 and 8 illustrate the influence of the coefficient of variation on the probabilities of Type I and II errors, respectively, for $\mu'_k = 1.0$, $\theta'_{\ln k} = 0.5$, and varying n. Points on the plots are obtained using 25000 realizations. The figures show that both Type I and Type II error probabilities (mostly) decrease with increasing coefficients of variation. For example, for $\mu'_k = 1.0$, $\theta'_{\ln k} = 0.5$, and $n = 4$, probabilities of Type I and Type II errors decrease from 0.03 to 0.01 and from 0.15 to 0.12, respectively, when the coefficient of variation increases from 1 to 2. However, the probability of Type II errors does tend to show a maximum at around a coefficient of variation of 1.0, so that this value of ν_k seems to be a "worst case" for the probability of Type II errors.

Influence of Number of Samples on Error Probabilities

It is expected that in a QC program of a cement-based S/S construction cell, increasing n decreases the chances of making an error in the decision about the approval of the construction cell. When the entire cell is sampled at every point, the probability of making a decision error will be zero. Figures 9 and 10 show the influence of the number of samples on probabilities of making a Type I and a Type II error, respectively, for different normalized means (i.e., $\mu'_k = 0.01$, 0.1, 0.9, 1.0, 1.1, and 10.0), $\nu_k = 1.0$ and $\theta'_{\ln k} = 0.5$. These figures indicate that as the number of samples increase, the probabilities of Type I and Type II errors decrease as expected. Also as expected, the probabilities of both types of errors are very close to zero when the normalized mean is far from 1.0.

Application of the Simulation Results

Simulations are performed for an example to illustrate the scalability the simulation results presented in the previous sections (which considered a 1×1 cell). The example construction cell size is 10 m×10 m which is modeled using 160×160 elements each of size 0.0625 m×0.0625 m. The normalized mean and coefficient of variation of the point-scale hydraulic conductivity specified in the simulations are both 1.0, the normalized correlation lengths considered are 0.01, 0.05, 0.1, 1.0, and 10.0, and the number of samples used are 1, 4, 9, 16, and 25. Figures 11 and 12 present the comparison between the simulation results for probabilities of a Type I and a Type II error for the example problem and the case considered in this paper. Good agreements are obtained for both error probabilities between the two cases for all normalized correlation lengths, which illustrates the scalability of the simulation results presented in this paper.

The authors are currently developing a follow-up paper to present the results of a statistical analysis of an existing cement-based S/S system. The details can be found in Liza (2014), but the basics are summarized as follows: the S/S site is roughly peanut shaped in plan, having a treated area of about 114,000 m^2 with average thickness of 3.9 m. Over the site, 2086 hydraulic conductivity samples have been taken, allowing for reasonably accurate estimation of the hydraulic mean (normalized by a regulatory value of 1×10^{-8} m/s), and coefficient of variation, which were found to be 0.468 and 1.679, respectively. Liza (2014) also performed a goodness-of-fit test and found that the lognormal distribution gave a very reasonable fit to the hydraulic conductivity data. To estimate the correlation length, a relatively densely sampled rectangular subset of the site, having plan dimensions 55 m by 85 m, was selected over which a regular observational grid was interpolated. The estimated directional correlation lengths ranged from 9 m to 15 m, with an 'isotropic' correlation length estimate of 11 m. The remainder of this discussion concentrates on the 55 m by 85 m sub-site, since it has an estimated correlation length, but uses the μ_k' and v_k values estimated from the entire site.

The question now becomes: How do the results presented in this paper compare to the current sampling requirements specified by the USACE (2000) of 1 sample for every 500 m^3 of S/S material. First of all, the results presented herein are based on a 2-D model, and so the sampling requirements are given per unit area, not per unit volume. However, if the contaminated soil is in a layer which is thin relative to its areal extent, the 2-D specification is deemed to be quite reasonable. In any case, the actual volume of S/S material at the sub-site is approximately $55 \times 85 \times 3.9 = 18,233 \text{ m}^3$, requiring 18232/500 = 36 samples, according to USACE.

To use the results of Figures 11 and 12, the rectangular sub-site must be approximated by a square of dimensions $D \times D = 55 \times 85 = 4,675 \text{ m}^2$, so that $D = 68 \text{ m}$. For this square area, Figures 11 and 12 remain exactly the same except that the site size is now $68 \text{ m} \times 68 \text{ m}$ and when $\theta_{\ln k}' = 1$ it means that $\theta_{\ln k} = 68 \text{ m}$. If the error probabilities are to be restricted to being less than 5% at $\theta_{\ln k}' = 1$, it can be seen that $n = 25$ seems to be sufficient (giving a maximum probability of a Type I error of around 2% and a probability of a Type II error of around 4%). This number of samples is in the same ballpark as that recommended by USACE.

The maximum probabilities given in Figures 11 and 12 are approximately "worst case" since μ'_k is selected to be 1.0 with $v_k = 1$, the latter being approximately the worst case for the Type II error (see Figure 8). If the actual statistics for the site are used, $\mu'_k = 0.47$, $v_k = 1.7$, and $\theta'_{\ln k} = 11/68 = 0.16$, the results change as follows;

1. The reduced correlation length will reduce the Type I error, suggesting that fewer samples could be taken, but corresponds to the worst case Type II error, so that $n = 25$ would still seem reasonable,
2. The coefficient of variation of 1.7 leads to a reduction in the probabilities of both types of errors (see Figures 7 and 8), suggesting that this could lead to a reduction in the required number of samples,
3. The reduction in the normalized mean leads to a significant reduction in the probabilities of both types of errors, suggesting again that a reduction in the required number of samples would be appropriate.

It is to be noted that in general the actual statistics of an S/S field will not be known prior to the sampling (generally, if it were, there would be no need for the sampling), so that worst case results are recommended and conservative. In this light, it appears that the USACE (2000) sampling recommendations are quite reasonable for this site, leading to probabilities of Type I and II errors which are below 5%.

Summary and Conclusions

In this study, Monte Carlo simulations are performed using a modified version of the two-dimensional random finite element method (RFEM) program, mrflow2d, to examine the influence of the correlation length, hydraulic conductivity mean and coefficient of variation on sampling requirements for a QC program of cement-based S/S construction cells. The modification made to the program enables the sampling of the random field at prescribed locations.

Based on the results obtained in this study, the following conclusions can be drawn:

- For a specific number of samples in the QC program, the greatest probability of making an error in the hypothesis test occurs at a "worst case" correlation length, indicating that more samples are required at this correlation length. The "worst case" correlation lengths are found to be 1 to 5 times the effective construction cell dimension (square root of the construction cell area) for the probability of a Type I error and 0.1 to 10 times the effective construction cell dimension for the probability of a Type II error. If a single "worst case" value were to be recommended, it would be to set the correlation length equal to the effective construction cell dimension. The worst case correlation length leads to conservative sampling requirements to achieve target hypothesis error probabilities.
- For a specific number of samples, the greatest error probabilities occur when μ_k' is approximately 1.7 for Type I errors and 1.1 for Type II errors. This suggests that more samples are required when the normalized mean hydraulic conductivity

is in the range 1.1 to 1.7 in order to ensure that cells are properly identified as being unacceptable or acceptable (note that although the population mean μ_k may be above k_{crit}, individual cells may very well have $k_{eff} < k_{crit}$). For a constant number of samples, the probabilities of Type I and Type II errors rapidly approach zero when the mean hydraulic conductivity deviates significantly from the regulatory value (e.g. μ_k' = 0.01, 0.1, and 10.0). This, of course, implies that targeting the mean hydraulic conductivity well below the regulatory value is desirable, although possibly more expensive. Note that targeting a lower mean hydraulic conductivity may have no benefits with respect to the required number of QC samples, since the worst case must always be assumed prior to sampling.

- Increasing the number of samples is effective in decreasing both Type I and Type II error probabilities, which, of course, agrees with statistical theory.
- For a specific number of samples, an increase in the hydraulic conductivity coefficient of variation, v_k, generally results in an decrease in probabilities of Type I and Type II errors, at least when $\mu_k' = 1.0$ and $v_k > 1$. This reduction in error probability is largely because the resistance to flow increases as v_k increases, due to downstream blockages, so that the value of k_{eff} decreases with increasing v_k. The general implication is that when μ_k' is approximately 1.0, more samples will be required to achieve acceptably small error probabilities when v_k ; 1 or less.
- When an actual S/S site is considered, it appears that the USACE (2000) sampling recommendation is quite similar to the recommendations made in this study to

achieve error probabilities of less than 5% under reasonably "worst case" statistical assumptions. Work is ongoing to determine how to best refine the sampling requirements suggested by this research for general use in practice.

Acknowledgements

The authors acknowledge the financial support of the National Sciences and Engineering Research Council of Canada and the STEWARD program for the research presented in this paper.

References

Benson, C.H., Zhai, H., and Rashad, S.M. 1994. Statistical sample size for construction of soil liners. ASCE Journal of Geotechnical Engineering, **120**(10): 1704-1724.doi:10.1061/(ASCE) 0733-9410(1994)120:10(1704).

Devore, J.L. 2008. Probability and Statistics for Engineering and the Sciences. Seventh ed. Books/Cole, Pacific Grove, USA.

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.doi:10.1061/(ASCE)0733-9399(1990)116:8(1733).

Fenton, G.A., and Griffiths, D. V. 1993. Statistics of block conductivity through a simple bounded stochastic medium. Water Resources Research, **29**(6), 1825–1830.doi:10.1029/93WR00412.

Fenton, G.A., and Griffiths, D.V. 2008. Risk Assessment in Geotechnical Engineering. John Wiley & Sons, New York.

Liza, R. 2014. Statistical sample size for quality control programs of cement-based solidification/stabilization. Ph. D. Thesis, Dalhousie University, Civil Engineering Dept., Halifax, Canada.

Menzies, W.T. 2008. Reliability assessment of soil liners. M. A. Sc. Thesis, Dalhousie University, Civil Engineering Dept., Halifax, Canada.

USACE. 2000. Solidification/Stabilization (S/S) of contaminated material. Section 02160 A, Washington, D.C.

Vanmarcke, E.H. 1984. Random Fields: Analysis and Synthesis. The MIT press, Cambridge, Massachusetts, USA.

List of Symbols

The following symbols are used in this paper:

A	=	Construction cell area perpendicular to flow
D	=	Effective dimension of the construction cell
G	=	Standard normal random field
G_i	=	Local average of G over the i th element
H_0	=	Null hypothesis
H_a		Alternative hypothesis
i	=	Hydraulic gradient across the construction cell

k_{crit} = Regulatory hydraulic conductivity

k_{eff} = Effective hydraulic conductivity

k'_{eff} = Normalized effective hydraulic conductivity

k_G = Sample geometric average

k_i = Hydraulic conductivity of the i th element

k_j = Hydraulic conductivity of the j th sample

$\ln k$ = Log-hydraulic conductivity field

l = Number of samples in each of the x and y directions

m = Number of elements

n = Number of samples

n_1 = Number of realizations where $k_G < k_{crit}$ while actual $k_{eff} > k_{crit}$

n_2 = Number of realizations with $k_G > k_{crit}$ while actual $k_{eff} < k_{crit}$

p_1 = Probability of a Type I error

p_2 = Probability of a Type II error

Q = Total flow through the construction cell

X = Planar dimension of the construction cell in the x direction

Y = Planar dimension of the construction cell in the y direction

X/Y = Aspect ratio of the construction cell

θ_i = Correlation length in the i th direction of the $\ln k$ random field, $i = 1, 2$

θ_k = Random field correlation length for hydraulic conductivity

$\theta_{\ln k}$ = Correlation length of the $\ln k$ random field

$\theta'_{\ln k}$ = Normalized correlation length of the $\ln k$ random field

μ_k = Mean of the hydraulic conductivity field

$\mu_k^{'}$ = Normalized mean of the hydraulic conductivity field

$\mu_{\ln k}$ = Mean of the log-hydraulic conductivity field, $\ln k$

$\sigma_{\ln k}$ = Standard deviation of the log-hydraulic conductivity field, $\ln k$

σ_k = Standard deviation of the hydraulic conductivity field

v_k = Coefficient of variation of the hydraulic conductivity field

$\rho_{\ln k}$ = Correlation coefficient between points in the $\ln k$ random field

$\gamma_{\ln k}$ = Variance reduction function when $\ln k$ is averaged over some volume

γ = Same as $\gamma_{\ln k}$

τ_i = Distance between points in the ith direction of the random field, $i = 1, 2$

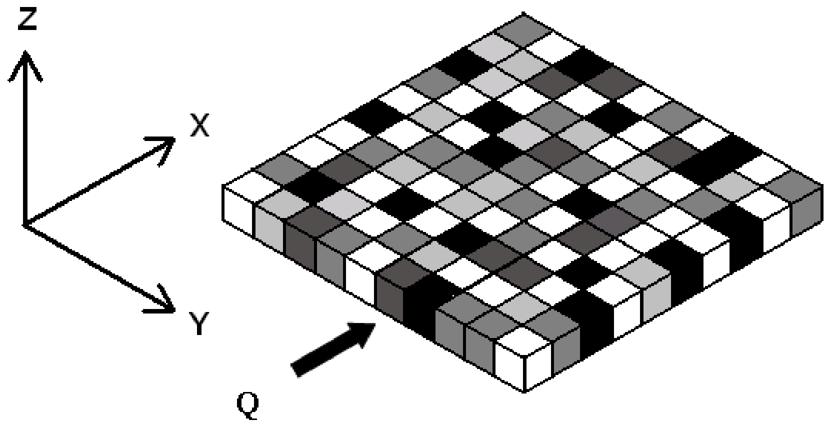

Fig.1. Illustration of mesh discretization used in the simulations

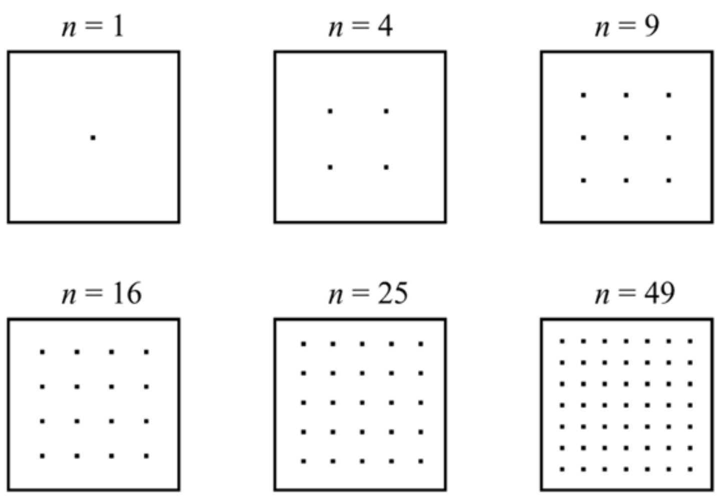

Fig. 2. Sampling locations shown as small black squares

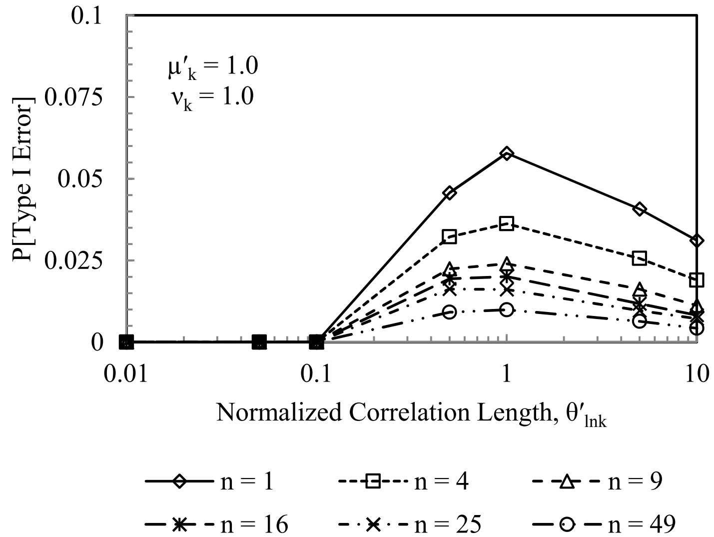

Fig. 3. Influence of correlation length on the probability of a Type I error for mean and coefficient of variation of 1.0

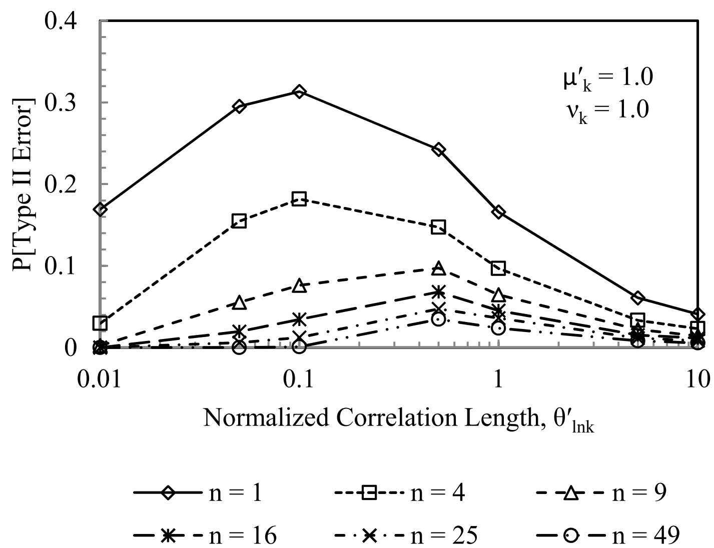

Fig. 4. Influence of correlation length on the probability of a Type II error for mean and coefficient of variation of 1.0

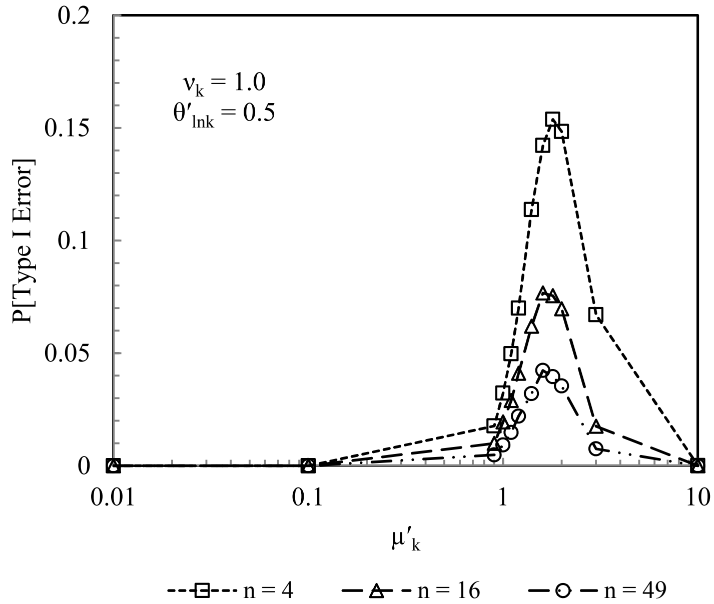

Fig. 5. Influence of mean on the probability of a Type I error

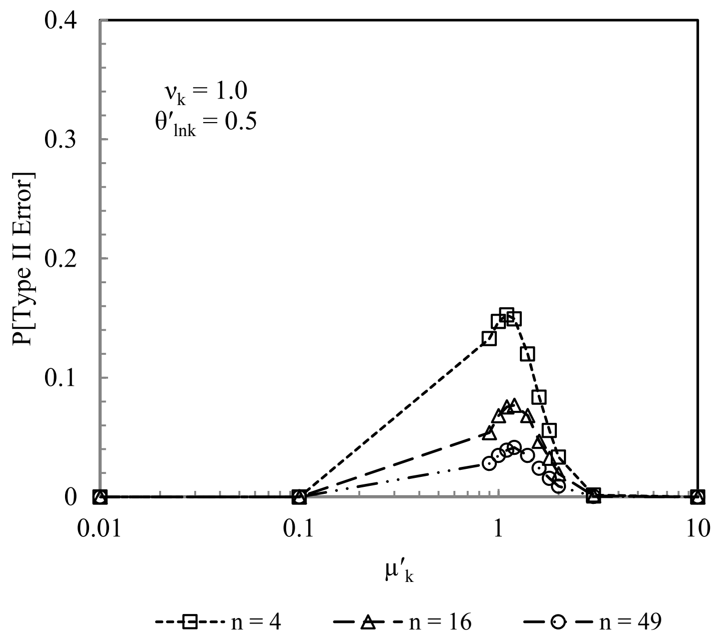

Fig. 6. Influence of mean on the probability of a Type II error

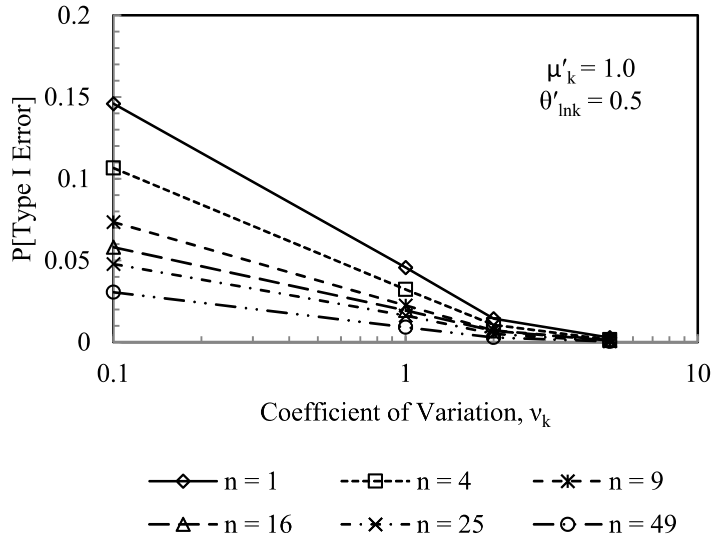

Fig. 7. Influence of coefficient of variation on the probability of a Type I error

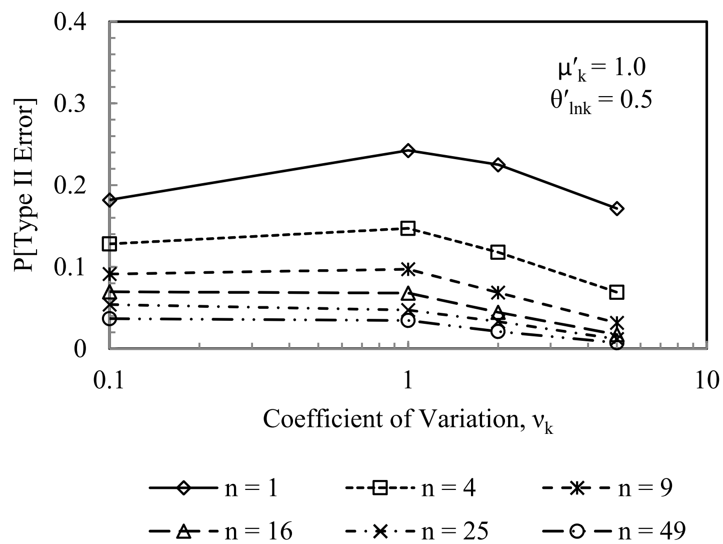

Fig. 8. Influence of coefficient of variation on the probability of a Type II error

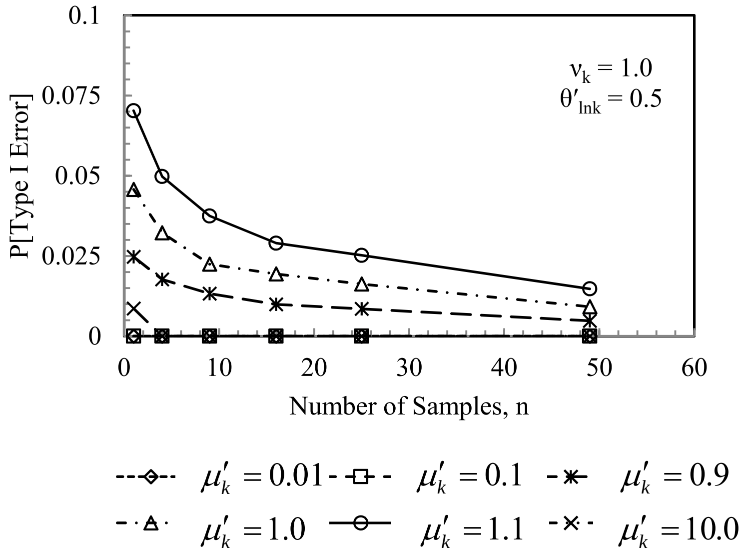

Fig. 9. Influence of number of samples on the probability of a Type I error

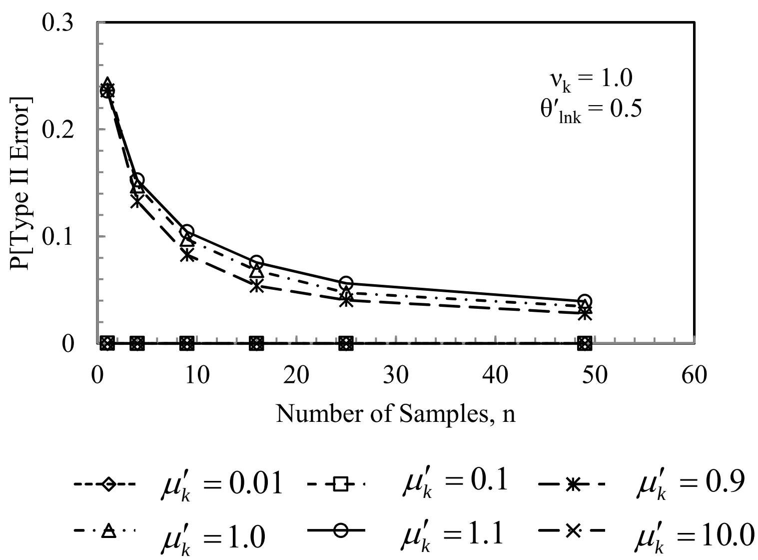

Fig. 10. Influence of number of samples on the probability of a Type II error

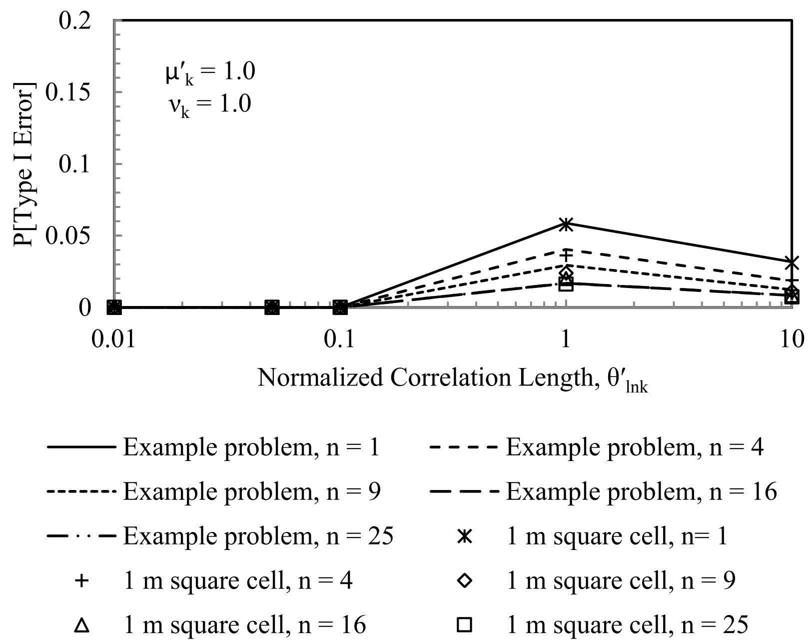

Fig. 11. Comparison of the simulation results for the probability of a Type I error between a (10 m × 10 m) and a 1 × 1 cell

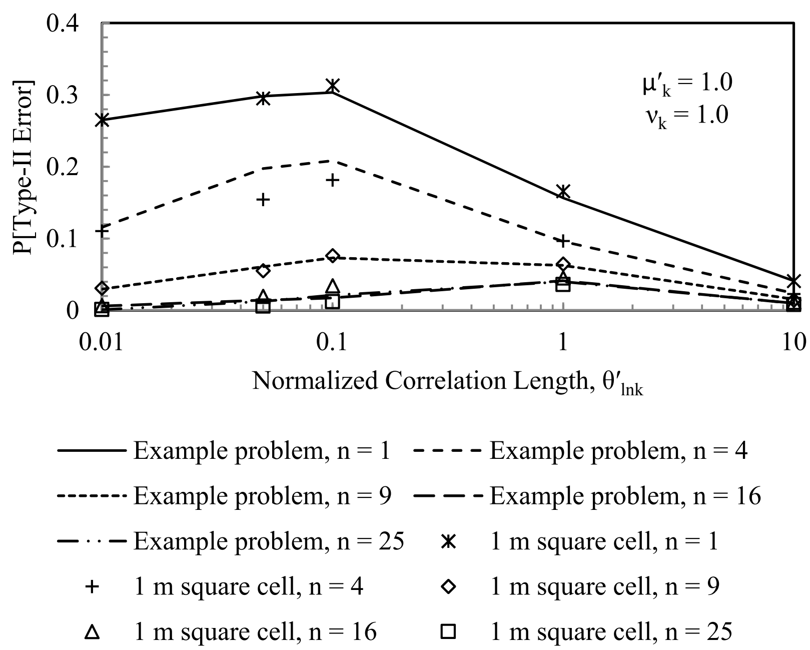

Fig. 12. Comparison of the simulation results for the probability of a Type II error between a (10 m × 10 m) and a 1 × 1 cell