Examination of the influence of data aggregation and sampling density on spatial estimation

Slobodan Vucetic[1], Tim Fiez[2] and Zoran Obradovic[3],
svucetic@eecs.wsu.edu, tfiez@wsu.edu, zoran@cis.temple.edu

[1] School of Electrical Engineering and Computer Science
[2] Department of Crop and Soil Sciences
Washington State University, Pullman, WA 99164-2752, USA
[3] Center for Information Science and Technology
Temple University, Philadelphia, PA 19122, USA

Preliminary version of paper:

Vucetic, S., Fiez, T., and Obradovic, Z., (2000) "Examination of the Influence of Data Aggregation and Sampling Density on Spatial Estimation," *Water Resources Research*, Vol. 36 , No. 12 , pp. 3721-3731.

Abstract.

Spatial processes may be sampled by point sampling or by aggregate sampling. If aggregate samples are collected over a regular grid and used to represent the central point of each aggregation area, the aggregate sampling functions as a low-pass filter and may eliminate aliasing during spatial estimation. To assess potential accuracy improvements, a numerical procedure for calculating the estimation error variance was developed. Analysis of point and block sampling techniques for kriging and inverse distance interpolation showed that for the same sampling density, block sampling provides better estimation. To achieve the same error levels, over 30%-50% more point samples were required than block samples. Furthermore, interpolation of block sampled data resulted in lower error variability and surfaces with more visual appeal.

1. Introduction

Representations of most spatial processes are based on sampling at sparse and discrete locations. Remote sensing and machine mounted sensors can provide very dense or essentially continuous sampling, but the applicability of these procedures is still limited for most biological and physical processes. Traditional sampling can be used to collect point samples or aggregate samples. For point sampling, a single sample or multiple samples from a very small spatial extent are collected per location. Sample analysis yields the true value of the underlying spatial attribute plus any analytical error at each location [*Starks*, 1986]. The term aggregation sampling is used to describe situations where samples are extracted from a larger area and averaged to represent that area or its central point. Examples of spatial aggregation techniques include block sampling in soil science [*Webster and Burgess*, 1984], throughfall sampling in hydrology [*Hill et al.*, 1999], marine surveys [*Pennington and Volstad*, 1991], and regionalized data in economics [*Anselin*, 1988].

One of the primary purposes of sampling is to collect data for spatial interpolation procedures. The correctness of an interpolated surface depends on several factors such as the sampling design and density [*Burgess et al.*, 1981; *Olea*, 1984; *Rouhani*, 1985; *Bogaert and Russo*, 1999] and the interpolation procedure [*Laslett et al.*, 1987; *Weber and Englund*, 1992]. Interpolation procedures include geostatistical-based kriging [*Krige*, 1963; *Cressie*, 1993], and deterministic methods such as inverse distance interpolation [*Isaaks and Srivastava*, 1989] and splining [*Hutchinson and Gessler*, 1994].

The type of sampling, point or aggregated, may also play an important role. Generally, point sampling should provide more accurate spatial estimation near sampling locations while aggregation sampling should allow better overall estimation at the expense of losing some of the accuracy near sampling locations. Signal processing theory can help in analyzing this difference more formally. If one collects aggregate samples over a regular grid and uses the data to represent the central point of the aggregation area (hereafter referred to as block sampling), block sampling can be regarded as point sampling of filtered underlying spatial process since the within cell aggregation functions as a low-pass filter.

The possible benefits of low-pass filtering are explained by the sampling theorem [*Proakis and Manolakis*, 1996]. In a two-dimensional case, the theorem implies that a square sampling grid with spacing D=1/2F is sufficient for perfect reconstruction of a two-dimensional signal with a maximum spectral bandwidth of F. In practice, the sampling theorem is only a guide, since an ideal reconstruction procedure assumes an infinite spatial domain and so exists just in theory. Connected with the sampling theorem is the problem of aliasing [*Proakis and Manolakis*, 1996]. Every frequency component of a signal above 1/2D (half the sampling frequency), also called the cut-off frequency, converts to noise after its reconstruction. To prevent this, signals are typically low-pass filtered (anti-aliased) prior to sampling to cancel out the part of the signal above the cut-off frequency and prevent it from becoming noise. From this perspective, the low-pass filtering resulting from block sampling may be desirable. Anti-aliasing filters used in signal processing are near perfect with respect to cutting-off all components of a signal with frequencies above 1/2D. Such filters are not applicable to block sampling since their coefficients follow fairly complex functions while practical block sampling patterns correspond to filters with few equal coefficients. Therefore, block sampling is a crude approximation of an optimal anti-aliasing filter.

If the benefits of block sampling suggested by signal processing theory result in improved spatial estimation, block sampling may be a way to reduce sample analytical costs while maintaining prediction accuracy or to achieve better prediction accuracy without increasing costs. Concerns over sampling and analysis costs often result in less than optimal sampling densities, which leads to poor spatial estimates in disciplines such as agriculture [*Mallarino et al.*, 1999]. Thus, the objective of this paper is to determine if data aggregation can improve spatial estimation by kriging and inverse distance interpolation as compared to traditional point sampling. We present both estimation accuracy and economic analyses. In addition, we examined (i) the influence of different covariogram models on spatial estimation and (ii) the influence of sampling density on kriging and inverse distance interpolation.

The paper is organized as follows. Section 2 provides a review of spatial statistics, the statistically based estimation procedure of kriging, and the non-statistical procedure of inverse distance interpolation, as well as more formal treatment of point and block sampling. In Section 3, we propose a procedure for the analysis of the influence of block sampling on spatial interpolation that includes (i) fast generation of artificial spatial layers satisfying desired covariograms, and (ii) automatic estimation of covariograms of aggregated spatial data from known covariograms of original spatial data. Finally, the experimental results and discussion are presented in Section 4.

2. Preliminaries

Data points close together in space are often more alike than those that are far apart. Non-statistical spatial estimators, such as inverse distance interpolation, implicitly use this notion of spatial continuity and often prove very successful at predicting data values at non-sampled locations. Alternatively, spatial statistics model spatial continuity. From these models, a statistical spatial estimation procedure called kriging has been developed. This section summarizes background information for spatial statistics, kriging and inverse distance interpolation needed in the rest of the paper. Also, point and block sampling are formally presented.

2.1. Spatial Continuity and Covariogram

In this paper it is assumed that spatial data over a given region, denoted $\{z(\mathbf{s}), \mathbf{s} \in D\}$, are a realization of a second-order stationary and isotropic random process, $\{Z(\mathbf{s}), \mathbf{s} \in D\}$, where D is a fixed subset of R^2 covering the region. Second-order stationarity implies that $Z(\mathbf{s})$ has constant mean with covariance that satisfies

$$\mathrm{Cov}(\, Z(\mathbf{s}_1), Z(\mathbf{s}_2)\,) = C_{ZZ}(\mathbf{s}_1-\mathbf{s}_2) = C_{ZZ}(\mathbf{h}), \quad \mathbf{s}_1, \mathbf{s}_2 \in D, \mathbf{h} \in R^2, \tag{1}$$

where $C_{ZZ}(\mathbf{h})$ is the covariogram. In the rest of the paper we assume, without loss of generality, that the mean of $Z(\mathbf{s})$ is zero. A scaled version of the covariogram, $\rho_{ZZ}(\mathbf{h})=C_{ZZ}(\mathbf{h})/C_{ZZ}(\mathbf{0})$, defines an autocorrelation function called the correlogram. The process is isotropic if the covariogram depends only on the distance between points,

$$C_{ZZ}(\mathbf{s}_1-\mathbf{s}_2) = C_{ZZ}(\,||\,\mathbf{s}_1-\mathbf{s}_2\,||_2) = C_{ZZ}(h), \qquad \mathbf{s}_1, \mathbf{s}_2 \in D,\ h \geq 0, \tag{2}$$

where h is the Euclidean distance between $\mathbf{s}_1$ and $\mathbf{s}_2$.

To guarantee the validity of statistical spatial estimation which demands positive variance of any linear combination of random variables from Z(**s**), a covariogram (or correlogram) model should be positive-definite,

$$\mathrm{Var}\left(\sum_i \lambda_i Z(\mathbf{s}_i)\right) = \sum_i \sum_j \lambda_i \lambda_j C_{ZZ}(\mathbf{s}_i - \mathbf{s}_j) \geq 0 . \qquad (3)$$

Covariance C(h) can be expressed as

$$C(h) = c_0\, \delta(h) + c_1\, \phi(h; R) , \qquad (4)$$

where $\phi(h; R)$ is a positive-definite function, $\delta(h)$ is a Dirac function, c_0 is the nugget effect, c_0+c_1 is the sill, and R is the range. Checking positive-definiteness of covariogram (or correlogram) models is a cumbersome procedure [*Chung*, 1968; *Christakos*, 1984], and as a result just a few functions, $\phi(h; R)$, that are known to be positive-definite are used in practice. Some of the most popular are

1. Spherical: $$\phi(h;R) = \begin{cases} 1 - 1.5h/R + 0.5(h/R)^3, & \text{if} \quad 0 \leq h \leq R \\ 0, & \text{otherwise} \end{cases}$$

2. Exponential: $$\phi(h;R) = \begin{cases} \exp\left(-\dfrac{3h}{R}\right), & \text{if} \quad 0 \leq h \leq R \\ 0, & \text{otherwise} \end{cases} \qquad (5)$$

3. Gaussian: $$\phi(h;R) = \begin{cases} \exp\left(-\dfrac{3h^2}{R^2}\right), & \text{if} \quad 0 \leq h \leq R \\ 0, & \text{otherwise} \end{cases} .$$

When common models do not properly fit observed covariance it is necessary to check the positive-definiteness of alternative models. Using Bochner's theorem [*Bochner*, 1959], it is possible to check the positive-definiteness of $C_{ZZ}(\mathbf{h})$ in the spectral domain. The Fourier transform of $C_{ZZ}(\mathbf{h})$ for a second-order stationary process Z(**s**) is defined as a real function

$$S_{ZZ}(\boldsymbol{\omega}) = \int_{R^2} C_{ZZ}(\mathbf{h}) e^{-j\langle \omega, \mathbf{h} \rangle} d\mathbf{h} , \qquad (6)$$

where $\langle \boldsymbol{\omega}, \mathbf{h} \rangle = \omega_1 h_1 + \omega_2 h_2$ in a two-dimensional case and $S_{ZZ}(\boldsymbol{\omega})$ is called the power spectrum of Z(**s**) [*Papoulis*, 1991]. If the power spectrum $S_{ZZ}(\mathbf{h})$ is positive [*Yao and Journel*, 1998] $C_{ZZ}(\mathbf{h})$ is positive-definite. The *Fast Fourier transform* (FFT) can be used as a discrete approximation of a Fourier transform allowing computationally efficient transformations between spatial and spectral domains [*Brigham*, 1988]. The FFT for a two-dimensional case is computed as

$$S_{ZZ}(m_1, m_2) = \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} C_{ZZ}(k_1, k_2) \cdot \exp\left[-2\pi j (k_1 m_1 + k_2 m_2)\right], \qquad (7)$$

where $C_{ZZ}(k_1,k_2)$ and $S_{ZZ}(m_1,m_2)$ are properly discretized functions $C_{ZZ}(\mathbf{h})$ and $S_{ZZ}(\boldsymbol{\omega})$. Extending Bochner's theorem to the discretized case [*Yao and Journel*, 1998], if $C_{ZZ}(k_1,k_2)$ is positive definite $S_{ZZ}(m_1,m_2)$ is a positive real function and vice-versa.

2.2. Ordinary Kriging

The problem of spatial estimation is to estimate the values of $z(\mathbf{s})$, $\mathbf{s}\in D$, within some region D, given the values $z(\mathbf{s}_1)$, $z(\mathbf{s}_2)$,…$z(\mathbf{s}_n)$, observed at n known spatial locations, where $z(\mathbf{s})$ is a realization of an underlying random process $Z(\mathbf{s})$. Ordinary kriging [*Krige*, 1963; *Matheron*, 1971] is the best linear unbiased estimator under the assumptions that $Z(\mathbf{s})$ is second-order stationary with known second-order statistics. In practice, the assumption of second-order stationarity is sometimes violated, and there are kriging approaches for these situations [*Cressie*, 1993]. In ordinary kriging $Z(\mathbf{s}_0)$, $\mathbf{s}_0\in D$, is expressed as a weighted sum of values from sampling locations, $Z(\mathbf{s}_i)$,

$$\hat{z}(\mathbf{s}_0)=\sum_{i=1}^{n}\lambda_i z(\mathbf{s}_i),\quad \sum_{i=1}^{n}\lambda_i=1, \tag{8}$$

with the coefficients λ_i chosen to minimize the error variance,

$$\begin{aligned}\sigma^2(\mathbf{s}_0)&=\mathrm{Var}\left(Z(\mathbf{s}_0)-\hat{Z}(\mathbf{s}_0)\right)=\mathrm{Var}\left(Z(\mathbf{s}_0)-\sum_{i=1}^{n}\lambda_i Z(\mathbf{s}_i)\right)=\\&=C_{ZZ}(\mathbf{0})+\sum_i\sum_j\lambda_i\lambda_j C_{ZZ}(\mathbf{h}_{ij})-2\sum_i\lambda_i C_{ZZ}(\mathbf{h}_{i0}),\end{aligned} \tag{9}$$

where $\mathbf{h}_{ij}=\mathbf{s}_i-\mathbf{s}_j$, and $\mathbf{h}_{i0}=\mathbf{s}_i-\mathbf{s}_0$. Therefore, covariogram estimation must precede kriging. The expression for weights λ_i can be derived through the use of Lagrange multipliers [*Matheron*, 1971; *Davis and Grivet*, 1984] and is

$$\boldsymbol{\lambda}=C^{-1}\mathbf{c}+\frac{(1-\mathbf{e}'C^{-1}\mathbf{c})}{\mathbf{e}'C^{-1}\mathbf{e}}\cdot C^{-1}\mathbf{e}, \tag{10}$$

where $\boldsymbol{\lambda}$ is a vector of weights, $\mathbf{e}$ is a vector of ones, C is a covariance matrix with elements $C_{ZZ}(\mathbf{h}_{ij})$, and $\mathbf{c}$ is a vector with elements $C_{ZZ}(\mathbf{h}_{i0})$.

Usually, only points in a certain neighborhood of $\mathbf{s}_0$ are used in the estimation of $Z(\mathbf{s}_0)$ instead of all n sample points. The benefits of this are twofold: the calculation of coefficients $\boldsymbol{\lambda}$ is faster without much loss in estimation accuracy, and it allows for the use of kriging if $Z(\mathbf{s})$ is not stationary but stationarity can be assumed for small regions.

2.3. Inverse Distance Interpolation

The other type of spatial estimator considered in our analysis is inverse distance interpolation. It does not require statistical analysis and covariogram fitting, which allows for very fast estimations at unknown locations. Although it does not achieve statistically optimal estimation like kriging, errors from inverse distance interpolation are reported to be fairly similar to that of kriging, and sometimes, surprisingly, even better [*Weber and Englund*, 1992**]**. These features make the inverse distance method popular among practitioners. The estimated value at an unknown location $\mathbf{s}_0$, $z(\mathbf{s}_0)$, is calculated as a linear combination of values at known locations according to (8) with weights λ_i calculated as

$$\lambda_i = \frac{1/d_{i0}^{\,p}}{\sum_{j=1}^{n} 1/d_{j0}^{\,p}}\,, \qquad (11)$$

and $d_{i0} = \|\mathbf{s}_i - \mathbf{s}_0\|_2$ is distance between $\mathbf{s}_i$ and $\mathbf{s}_0$, while p is an arbitrarily chosen parameter. Observe that for p=0, the inverse distance estimator is a simple mean estimator, and for p=∞ it is equivalent to the nearest neighbor estimator. In practice, p is typically chosen to be in the interval [1/2, 2], with larger values assigning larger weight to the closest samples and with p=2 being the most popular choice. If Z(**s**) is second-order stationary, equation (9) can be used to estimate the error variance of the inverse distance estimator. Similar to ordinary kriging, use of a moving neighborhood instead of a global one can be beneficial for inverse distance interpolation.

2.4. Point and Block Grid Sampling

In point sampling, samples are taken at the intersections of an imaginary grid laid over the surface. We denote point sampled data as $\{z(\mathbf{s}_i)\}$, i=1,2,…n, where $\mathbf{s}_i$ is the location of sample i. In block sampling, multiple sub-samples are taken within each block (encompassing each grid intersection) and combined to form one aggregate sample for analysis. We denote block sampled data as $\{y(\mathbf{s}_i)\}$, i=1,2,…m, where $\mathbf{s}_i$ is the location of the center of block i.

In practical applications of point sampling, as shown in Figure 1.a, several sub-samples are sometimes taken within a very small area around each sampling location instead of taking just one sample. This procedure can decrease the nugget effect and, therefore, result in better spatial estimation. Since in our analysis we assume that the nugget effect in the spatial data is zero, both types of point sampling are equivalent. In Figure 1.b we show an example of block sampling on a regular square grid, preferable under assumptions of second-order stationarity and isotropy [*Olea*, 1984], with squared aggregation area of the side equal to the sampling distance.

Assuming a uniform block sampling pattern, block sampled data $\{y(\mathbf{s}_i), \mathbf{s}_i \in D\}$ represent a finite sample from the random process Y(**s**) obtained by low-pass filtering the original random process Z(**s**),

$$Y(\mathbf{s}) = Z(\mathbf{s}) * b(\mathbf{s})\,, \qquad (12)$$

where $b(\mathbf{s})$ is a deterministic linear filter corresponding to the block sampling pattern. In general, block sampling can correspond to continuous or discrete filtering. In continuous block sampling, Y(**s**) can be expressed analytically as

$$Y(\mathbf{s}) = \int_{\mathbf{s}_i \in R^2} Z(\mathbf{s}-\mathbf{s}_i)\,b(\mathbf{s}_i)\,d\mathbf{s}_i\,. \qquad (13)$$

In discrete block sampling, which is of more practical interest, a small finite number of sub-samples are averaged and therefore $b(\mathbf{s})$ is a discrete function. Thus, Y(**s**) can then be expressed as

$$Y(\mathbf{s}) = \sum_{i=1}^{m_B} Z(\mathbf{s}-\mathbf{s}_i)\,b(\mathbf{s}_i)\,, \qquad (14)$$

where m_B is the number of sub-samples taken within the block and $\sum b(\mathbf{s}_i) = 1$ as the sub-sample aggregation is a simple averaging of $Z(\mathbf{s})$ over a block. If all sub-samples are weighted equally, all filter coefficients, $b(\mathbf{s}_i)$, are the same, $b(\mathbf{s}_i)=1/m_B$.

3. Calculating Estimation Error Variance

To analyze the influence of data aggregation and sampling density on spatial estimation, a proper procedure for the calculation of estimation error variance is needed. When interpolating point sampled data taken from an original process, $Z(\mathbf{s})$, the estimation error variance, $\sigma_P^2(\mathbf{s}_0)$, $\mathbf{s}_0 \in D$, depends only on the applied estimation method and can be expressed as

$$\sigma_P^2(\mathbf{s}_0) = \mathrm{Var}\left(Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)\right), \tag{15}$$

which for both kriging and inverse distance interpolation can be calculated using equation (9).

The estimation error variance with block sampled data can be expressed as

$$\sigma_B^2(\mathbf{s}_0) = \mathrm{Var}\left(Z(\mathbf{s}_0) - \hat{Y}(\mathbf{s}_0)\right) = \mathrm{Var}\left(Z(\mathbf{s}_0) - \sum_{i=1}^{n} \lambda_i(\mathbf{s}_0) Y(\mathbf{s}_i) \right), \tag{16}$$

where $\hat{Y}(\mathbf{s}_0)$ is the spatial estimation of the filtered process $Y(\mathbf{s}_0)$, and $\lambda_i(\mathbf{s}_0)$, i=1,… n, are the coefficients obtained for the estimation of point $\mathbf{s}_0$. Since $Y(\mathbf{s})$ can be considered as a linear unbiased estimator of $Z(\mathbf{s})$ and $\hat{Y}(\mathbf{s}_0)$ is also a linear unbiased estimator of $Y(\mathbf{s}_0)$, from (16) it follows that

$$\sigma_B^2(\mathbf{s}_0) = C_{ZZ}(\mathbf{0}) - 2\sum_{i=1}^{n} \lambda_i(\mathbf{s}_0) C_{ZY}(\mathbf{s}_i - \mathbf{s}_0) + \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i(\mathbf{s}_0)\lambda_j(\mathbf{s}_0) C_{YY}(\mathbf{s}_0 - \mathbf{s}_i - \mathbf{s}_j), \tag{17}$$

where C_{ZY} is the cross-covariance of processes $Z(\mathbf{s})$ and $Y(\mathbf{s})$, and C_{YY} is the covariance of the filtered process $Y(\mathbf{s})$. Both C_{ZY} and C_{YY} can be expressed as convolutions of the covariance of the original process $Z(\mathbf{s})$ and a linear filter $b(\mathbf{h})$ [*Papoulis*, 1991],

$$C_{ZY}(\mathbf{h}) = C_{ZZ}(\mathbf{h}) * b(-\mathbf{h}), \quad C_{YY}(\mathbf{h}) = C_{ZZ}(\mathbf{h}) * b(-\mathbf{h}) * b(\mathbf{h}). \tag{18}$$

As can be seen, (17) resembles (9), and therefore, procedures for the calculation of $\sigma_P^2(\mathbf{s}_0)$ and $\sigma_B^2(\mathbf{s}_0)$ are similar. The only difference is that (17) requires using C_{ZY} and C_{YY}, and in subsection 3.1 we discuss different methods for their calculation. Since some of our comparisons of point and block sampling are based on generated layers, in subsection 3.2 we present an efficient method for generating spatial layers with desired covariograms.

3.1. Calculating Cross-covariance C_{ZY} and Covariance C_{YY}

If $b(\mathbf{h})$ is a continuous filter, equation (18) can be solved analytically using equation (13). However, except for the case of simple radial filters, analytical solutions can be difficult to compute. If the problem is transformed to the Fourier domain, convolution can be performed by simple multiplication. In this approach C_{ZY} and C_{YY} can be calculated as

$$C_{ZY}(\mathbf{h}) = F^{-1}\left\{S_{ZZ}(\boldsymbol{\omega}) \cdot H_b^*(\boldsymbol{\omega})\right\}, \quad C_{YY}(\mathbf{h}) = F^{-1}\left\{S_{ZZ}(\boldsymbol{\omega}) \cdot |H_b(\boldsymbol{\omega})|^2\right\}, \tag{19}$$

where $H_b(\boldsymbol{\omega})$ is the Fourier transform of filter $b(\mathbf{h})$, and F^{-1} represents the inverse Fourier transform. However, this approach involves the calculation of Fourier integrals, which can also be difficult.

Therefore, for most practical cases, numerical approximation of C_{ZY} and C_{YY} using equation (14) is an acceptable solution. For the experiments presented in section 4, C_{ZZ}, C_{ZY} and C_{YY} were calculated along a regular grid with $N \times N = 161 \times 161$ points spanning an area of size $(-2D,2D) \times (-2D,2D)$, where D is sampling distance. Applying equation (14) in the case of sparse block sub-sampling requires $O(N^2 m_B)$ operations for C_{ZY} and $O(N^2 m_B^2)$ operations for C_{YY}. If the block sub-sampling is dense, $m_B = O(N)$ and the calculation of C_{ZY} and C_{YY} requires $O(N^3)$ and $O(N^4)$ operations, respectively.

In cases with dense block sub-sampling, numerical approximation of equation (19) using an FFT can be more efficient. Since an FFT of K elements takes $O(K \cdot \log K)$ operations [*Brigham*, 1988], $O(N^2 \log N)$ operations would be needed for both C_{ZY} and C_{YY}, which is a significant improvement for large N. While FFT-based covariance estimation has been successfully used to automatically compute covariance tables from available spatial data [*Yao and Journel*, 1998], we used equation (14) in our experiments for calculating both C_{ZY} and C_{YY}.

3.2. Fast Generation of Spatial Layers

Stationary Gaussian fields meeting a specified covariance model can be generated using a moving averages method [Journel, 1974; Oliver, 1995]. To generate a Gaussian random field, covariance should be expressed as the convolution of a function and its transpose, $C_{ZZ}(\mathbf{h}) = g * g^T$. A spatial layer, $z(\mathbf{s})$, with covariance function $C_{ZZ}(\mathbf{h})$ and mean 0 can be generated by the convolution of g with a two-dimensional stationary random field with a Dirac covariance measure, $x(\mathbf{s})$, as

$$z(\mathbf{s}) = \int g(\mathbf{s} - \mathbf{s}')x(\mathbf{s}')d\mathbf{s}' \quad . \tag{20}$$

Generated data are usually discrete nodes on a grid, so the integral in (5) is replaced by a sum. In this case, assigning random numbers from a Gaussian distribution to the nodes of a grid can generate the discrete random field $x(k_1,k_2)$.

The main problem with this approach is calculating $g(\mathbf{s})$, and it is usually reduced into two steps [*Oliver*, 1995]:

1. Calculating the Fourier transform of the covariance function denoted by $S(\boldsymbol{\omega})$;
2. Calculate the inverse Fourier transform of $\sqrt{S(\boldsymbol{\omega})}$ to obtain $g(\mathbf{s})$.

An analytical solution for $g(\mathbf{s})$ that represents a corresponding moving average filter has been derived in *Oliver* [1995] for exponential and Gaussian isotropic two-dimensional covariance models. After this step, discretized values of $g(\mathbf{s})$, denoted $g(k_1,k_2)$, can be used to generate the desired layer from $x(k_1,k_2)$.

Here, we propose a computationally efficient FFT-based procedure that finds the coefficients of a moving average filter, $g(k_1,k_2)$, and allows the generation of Gaussian layers with any valid covariance, $C(\mathbf{h})$, including anisotropic ones. Given a grid distance, D, and a square of K^2 grid nodes, our procedure for generating layer with desired covariance $C(\mathbf{h})$ is as follows:

1. Assign K^2 randomly chosen values from a Gaussian distribution with mean 0 and variance 1 to the grid thus forming $x(k_1,k_2)$, $k_1,k_2 = 1,2,\ldots K$. To speed up the procedure in the following steps, K should be a power of 2.
2. Calculate the FFT of the desired and properly discretized two-dimensional covariance, $C(k_1,k_2)$, to obtain its power spectrum, $S(m_1,m_2)$, $m_1,m_2 = 1,2,\ldots K$. Since $C(\mathbf{h})$ is positive-definite, $S(m_1,m_2)$ will be real and positive.
3. Find the coefficients of the discrete moving average filter, $g(k_1,k_2)$, by calculating the inverse FFT of $\sqrt{S(m_1,m_2)}$.
4. Apply filter $g(k_1,k_2)$ to random layer $x(k_1,k_2)$ to obtain $z'(k_1,k_2)$.
5. Finally, scale $z'(k_1,k_2)$ to derive layer $z(k_1,k_2)$ with specified mean and variance.
6. Depending on the size of filter $g(\mathbf{s})$, boundary points of the generated layer $z(\mathbf{s})$ should be removed, since its spatial variability will be slightly higher than desired. Our approach was to remove all points at distances less than desired range, R, from the edges of the generated layer.

A comparison of the correlogram of a simulated layer and the desired correlogram (spherical with parameters D = 5m, K = 256, and range R = 200m) is shown in Figure 2.

4. Experiments and Discussion

The following assumptions were used for all experiments:

1. The two-dimensional process describing spatial variability was second-order stationary and isotropic.
2. The sampling grid was regular and square.
3. Both kriging and inverse distance interpolation were performed using the nearest 16 samples to estimate the unknown location as shown in Figure 3. The use of local samples instead of all samples is more robust to non-stationarity and is computationally faster.
4. The nugget effect was assumed to be zero. The nugget effect can be considered as the sum of measurement error and microscale processes [Cressie, 1993], causing estimation error that can not be avoided by any kind of interpolation. In our experiments the goal was to compare different estimation and sampling techniques, and adding a nugget effect would distract from the analysis.

According to these assumptions we have used equations (9) and (17) to compute the estimation error variance, $\sigma^2(\mathbf{s}_i)$, $\mathbf{s}_i \in$ *block*, for 1600 regularly spaced points within an inner square, *block*, of size D×D from Figure 3, where D is the grid distance. Therefore, as mentioned in 3.1, to apply equations (9) and (17), C_{ZZ}, C_{ZY} and C_{YY} should be calculated along a regular grid with N×N = 161×161 points spanning an area of size (−2D,2D)×(−2D,2D). Averaging the 1600 values allowed us to accurately estimate the average error variance for a grid block, $\overline{\sigma}^2$. Also, using $\sigma^2(\mathbf{s}_i)$, we were able to compare the performance of the estimators across a *block*.

4.1. The Influence of Covariogram Models

First, we examined the influence of grid distance on the error variance of kriging and inverse distance interpolation for point sampled data. Exponential, spherical and Gaussian covariograms, all with range R=100m, nugget effect zero, and sill one were examined. Error variances were examined for grid distances D = 20, 40,… 300m. Normalized average error variances of kriging over a block, $\bar{\sigma}^2 / C_{ZZ}(\mathbf{0})$, are shown in Figure 4a as a function of grid density. Observe that instead of D we report D/R, which is invariant to scaling of D and R, and that $\bar{\sigma}^2 / C_{ZZ}(\mathbf{0})$ is in fact equal to 1- $\mathbf{r}^2$, with $\mathbf{r}^2$ denoting the coefficient of determination of spatial estimation.

Similar normalized average error variances resulting from inverse distance interpolation with p = 2 are shown in Figure 4b. For a given D/R, the estimation error variance of the exponential covariogram was the largest, while the estimation error variance of the Gaussian covariogram was the smallest for both kriging and inverse distance interpolator. This is because, for the same range, the exponential covariogram has the weakest spatial dependence and Gaussian the strongest. It is important to observe that error variance of both kriging and inverse distance interpolation is significant for D = R, which in practice is often assumed as an acceptable sampling density.

4.2. Comparison at Different Sampling Densities

Error variances of kriging and inverse distance interpolators with p = 0.5, 1, 2, for a spherical covariogram, which is the model most often used in practice are compared in Figure 5. Consistent with theory [*Cressie*, 1993] and given our statistical assumptions, kriging performed better than inverse distance interpolation for all sampling densities. For inverse distance interpolation, larger p values were better for dense sampling while p = 0.5 or 1 produced better results than p = 2 when D/R was greater than 1.5. The use of inverse distance interpolation with p = 2 (inverse distance squared) is very popular although these results show that the interpolated values are very unreliable when data are undersampled which is common with real-life data. Furthermore, the results indicate that statistical range analysis can be beneficial for determining the proper power, *p,* for inverse distance interpolation.

It has been observed that in some cases, inverse distance interpolation can produce better estimates than kriging [*Weber and Englund*, 1992]. Given the small difference between the best inverse distance interpolator and kriging shown in Figure 5, it is possible that errors in covariogram estimation might cause these observations. There is considerable literature [*Diamond and Armstrong*, 1984] examining the influence of covariogram estimation using sampled data. Generally, each error in the estimation of the true parameters of the covariogram causes an increase in the estimation error variance. Thus, for certain grid densities and coefficient p, the inverse distance interpolator can be the best interpolator if there are errors in covariogram estimation caused by sparse sampled data.

4.3. Comparison of Point and Block Sampled Data at Different Grid Densities

To examine block sampling, the following aggregation patterns were considered:

1. Dense within-block sampling where a large number of sub-samples are taken equally spaced and separated by D/40 within a whole block of size D×D. Practically, this represents almost continuous averaging over a block.

2. Sparse within-block sampling where just 9 sub-samples were taken within a block of size D×D at ad-hoc locations as shown in Figure 6.

Sparse within-block sampling can potentially be applied to many sampling situations without adding much cost as compared to point sampling, while dense within-block sub-sampling corresponds to continuous aggregate sampling. One-dimensional transects of the spectral characteristics of the proposed block sampling pattern assignments are shown in Figure 7. These patterns were compared to the optimal anti-aliasing filter that cancels out all frequencies of the underlying spatial signal above 1/2D. Frequency responses were calculated for distance D=100m. As can be seen, neither of the aggregation patterns were successful approximations of the optimal anti-aliasing filter.

To test the ability of the block sampling patterns to function as anti-aliasing filters, the estimation error variance on point sampled data and both dense and sparse within-block sampled data was compared. The estimation error variance $\sigma^2(\mathbf{s}_i)$ was computed for 1600 regularly spaced points, $\mathbf{s}_i$, within a *block*, and the corresponding normalized average error for block sampled data was computed using the procedure described in Section 3. The kriging error variance for spherical, exponential and Gaussian underlying spatial processes with range R=100m using point and block sampled data is shown in Figure 8. Results using inverse distance interpolation with p = 2 are shown in Figure 9.

Even though the analyzed block sampling patterns are crude approximation of optimal anti-aliasing filters, both block-sampling patterns resulted in lower prediction errors than point sampling. Furthermore, the difference between dense within-block sampling and sparse (9 sub-samples) within-block sampling was fairly small indicating that a sparse within-block sampling technique may be practical.

The block size of D×D used in the experiments is the optimal size suggested by theory. However, it might happen that size D is too large for a given application. In the following experiment we analyzed the influence of different block sizes on estimation error. In Figure 10 we plot the estimation error variance of kriging on dense within-block sampled data with the block size varying from zero (point sampling) to 1.4D×1.4D using a spherical covariogram and sampling density D equal to the range R. In accordance to the theory, the best estimation was achieved with block sizes between 0.8D×0.8D and D×D, although all block sizes in the examined range were superior to point sampling. Therefore, if a block size of D×D is considered too large, a smaller block size can still result in improved estimation over point sampling.

4.4. Economic Analysis

While block sampling results in lower prediction errors as compared to point sampling, more effort is required to collect a single block sample because multiple sub-samples must be collected to form each block sample. Thus on a per sample basis, block sampling would be more expensive than point sampling. However, to obtain a given error level (a horizontal line across Figure 8 or 9), more point samples are required than block samples so the point sampling procedure would result in higher sample analysis costs. For example, to obtain a kriging error variance of 60% with a spherical covariogram, point samples must be collected on an 81 m grid while block samples based on 9 sub-samples could be collected on a 95 m grid (Figure 11). Table 1 presents the grid distances needed to achieve different levels of accuracy, together with the ratio of the number of samples needed with point versus the two types of block sampling

(sparse and dense) for the same accuracy computed as α = (block sampling grid distance)^2/(point sampling grid distance)^2.

The overall costs to obtain a given error level for the two methods is a function of analysis costs and sample collection costs. As indicated in Table 1, to achieve a predetermined error level within a range $\bar{\sigma}^2 / C_{ZZ}(\mathbf{0}) \in [0.2, 0.8]$, kriging requires over 30% more point samples than block samples with sparse within-block sampling, and over 55% more point samples than block samples with dense within-block sampling. Consequently, block sampling based on 9 sub-samples is economically preferable if the costs of collecting the 8 additional sub-samples needed to form one block sample are less than 30% of the analytical expenses per sample.

From the presented analysis of block sampling, it is clear that its advantage over point sampling comes from discarding the highly variable part of the spatial information before kriging through aggregating sub-samples within a block. It has also been indicated that the proposed dense and sparse within-block sampling are crude approximation of the optimal anti-aliasing filtering which would provide perfect cancellation of the highly variable part of the spatial data. As indicated in Figure 10, aggregating soil cores within an area smaller than what was used for block sampling also discards some highly variable information before kriging and results in an estimation error variance between that obtained from point and optimal-sized block sampled data. Thus, if traveling expenses are too high, a similar economic analysis can be performed to determine the optimal aggregation area for a given application.

4.5. Sensitivity to Distance from Sampling Locations and Visualization of Kriging on Point and Block Sampled Data

In addition to the possible economic advantages of block sampling over point sampling there is also an important difference between the two sampling techniques in point-by-point estimation error variance within a *block*. As mentioned in the introduction, point sampling provides better estimation of points close to the sampling locations, while block sampling sacrifices some of the accuracy at these locations to improve estimation at points that are farther away from sampling locations. In Figure 12 we compare kriging error variances, $\sigma^2(\mathbf{s}_0)/C_{ZZ}(\mathbf{0})$, on point sampled and the two types of block sampled data (dense and sparse) with a spherical covariogram along a line AB from Figure 3. As can be seen, both types of block sampling leads to smoother kriging error variance over the whole spatial area as compared to point sampling. Although the overall estimation error variance, $\bar{\sigma}^2 / C_{ZZ}(\mathbf{0})$, is a good measure of the confidence of spatial estimation from sampled data, there are applications where the variability of estimation error can be important such as combining data interpolated from low resolution samples with data derived from high resolution sampling.

Finally, we show the results of a kriging experiment on a generated layer. A layer satisfying an exponential correlogram with range 250 m, and nugget effect 0 was generated using the method described in Section 3.2. These correlogram parameters were derived from data collected for soil nitrate N measured in a potato field in southeast Idaho with the exception that the nugget was set to 0. The simulated layer consisted of 81×81 points on a 10×10 m square grid. We simulated point and block sampling with 9 cores at 4 different sampling densities – 40, 60, 100, and 200 m. Ordinary kriging with a global neighborhood was performed on the sampled data. The coefficient of determination, $\mathbf{r}^2$, of kriging on point sampled data and block sampled data with 9 cores at 4 sampling densities versus the original data is reported in Table 2.

As can be seen, kriging on block sampled data was superior to point sampled data at all four sampling densities, as consistent with previous results. Figure 13 shows the original layer and the layers obtained by kriging on point and sparse within-block sampled data at a sampling distance of 100 m. Aside from the smaller interpolation error when block sampled data are used, the estimated layer using block sampled data has better visual appeal, allowing better insight in the spatial variability of the underlying spatial signal.

5. Conclusions

Spatial data available in practice are true or aggregated values extracted from an underlying spatial process. Data aggregation can be considered as spatial data filtering, which allows using the power spectrum representation obtained by Fourier transformation and an application of the sampling theorem in the analysis of sampling influence. Experiments to examine the influence of different covariogram models on spatial estimation and to compare kriging and inverse distance interpolation at different sampling densities provided useful insight in the capabilities of these spatial prediction methods. Kriging outperformed inverse distance interpolation but the margin was not large. The study on the influence of two sampling techniques, called point and block sampling, on kriging and inverse distance interpolation showed that block sampling provides better estimation at the majority of sampling densities and that in applications such as soil sampling it could be more economical than point sampling. An additional practical advantage of block sampling not considered in this study is that it is likely to provide more accurate estimation of the covariogram needed for kriging, resulting in a larger margin between kriging error on point and block sampled data.

In the broader context, block sampling can be regarded as a sampling technique that provides information on the characteristics of an underlying spatial process averaged across regions around sampling locations within the area of interest. Our analysis shows that increasing the aggregation region when block sampling up to a radius equal to half of the sampling distance leads to more accurate spatial estimation than with data obtained from an equal number of point samples. Further research is needed to determine the optimal sampling radius for various application types using the proposed analysis procedure.

Acknowledgements

We wish to thank anonymous reviewers for the many useful comments on the first version of this paper. Partial support by the Idaho National Engineering and Environmental Laboratory University Research Consortium project No. C94-175936 awarded to T. Fiez and Z. Obradovic is gratefully acknowledged.

References

Anselin, L., *Spatial Econometrics: Methods and Models*, Kluwer, Dordrecht, 1988.

Bochner, S., *Harmonic analysis and the theory of probability,* Univ. of California Press, Berkeley and Los Angeles, 1955.

Bogaert, P., Russo, D., Optimal spatial sampling design for the estimation of the variogram based on a least squares approach, *Water Resour. Res., 35*, 1275-1289, 1999

Brigham, E.O., *The Fast Fourier Transform and Its Applications*, Prentice Hall Signal Processing Series, 1988.

Burgess, T.M., Webster, R., McBratney, A.B., Optimal interpolation and isarithmic mapping of soil properties, IV, Sampling strategy, *J. Soil Sci., 32*, 643-659, 1981

Christakos, G., On the problem of permissible covariance and variogram models, *Water Resour. Res., 20*, 251-265, 1984.

Chung, K.L., *A Course in Probability Theory*, Harcourt, Brace, and World, New York, 1968.

Cressie, N.A.C., *Statistics for Spatial Data,* John Wiley & Sons, Inc., New York, 1993.

Davis, M.W., Grivet, C., Kriging in a global neighborhood, *Math. Geol., 16*, 249-265, 1984.

Diamond, P., Armstrong, M., Robustness of variograms and conditioning of kriging matrices, *Math. Geol., 16*, 809-822, 1984.

Hill, A.R., Kemp, W.A., Buttle, J.M., Goodyear, D., Nitrogen chemistry of subsurface from runoff on forested Canadian Shield hillslopes, *Water Resour. Res., 35*, 811-821, 1999.

Hutchinson, M.F., Gessler, P.E., Splines – more than just a smooth interpolator, *Geoderma, 62,* 45-67, 1994

Isaaks, E.H., Srivastava, R.H., *An Introduction to Applied Geostatistics,* Oxford University Press, New York, 1989.

Journel, A.G., Geostatistics for conditional simulation of ore bodies, *Econ. Geol., 69*, 673-687, 1974.

Laslett, G.M., McBratney, A.B., Pahl, P.J., Hutchinson, M.F., Comparison of several spatial prediction methods for soil pH, *J. Soil Sci., 38*, 325-341, 1987.

Mallarino, A.P., Oyarzabal, E.S., Hinz, P.N., Interpreting within-field relationships between crop yields and soil and plant variables using factor analysis, *Prec. Ag., 1*, 15-225 1999.

Matheron, G., The theory of regionalized variables and its applications, *Cahiers du Centre de Morphologie Mathematique, 5*, Fontainebleau, France, 1971.

Olea, R.A., Sampling design optimization for spatial functions, *Math. Geol., 16*, 369-392, 1984.

Oliver, D.S., Moving averages for Gaussian simulation in two and three dimensions, *Math. Geol., 27*, 939-960, 1995.

Papoulis, A., *Probability, Random Variables, and Stochastic Processes,* 3rd ed., McGraw Hill College Div, 1991.

Pennington, M., Volstad, J.H., Optimum size of sampling unit for estimating the density of marine populations, *Biometrics, 47*, 717-723, 1991.

Proakis, J.G., Manolakis, D.G., *Digital Signal Processing: Principles, Algorithms, and Applications*, 3rd ed., Prentice Hall, 1996.

Rouhani, S., Variance reduction analysis, *Water Resour. Res., 21*, 837-846, 1985.

Starks, T.H., Determination of support in soil sampling, *Math. Geol., 18*, 529-537, 1986.

Weber, D., Englund, E., Evaluation and comparison of spatial interpolators, *Math. Geol., 24*, 381-391, 1992.

Webster, R., Burgess, T.M., Sampling and bulking strategies for estimating soil properties in small regions, *J. Soil Sci., 35*, 127-140, 1984.

Yao, T., Journel, A.G., Automatic modeling of (cross) covariance tables using Fast Fourier Transform, *Math. Geol., 30*, 589-615, 1998.

Table 1. Normalized grid distances (grid distance / range) for point and two types of block sampling needed to achieve different levels of kriging estimation error variance and the ratio of the number of samples needed for point sampling versus two types of block sampling.

Normalized Estimation Error Variance	Grid Distance / Range			α	
	Point	Sparse Block	Dense Block	Sparse Block	Dense Block
0.2	0.31	0.36	0.41	1.35	1.75
0.3	0.44	0.53	0.58	1.45	1.74
0.4	0.58	0.66	0.72	1.29	1.54
0.5	0.69	0.79	0.86	1.31	1.55
0.6	0.81	0.95	1.02	1.38	1.58
0.7	0.95	1.14	1.25	1.44	1.73
0.8	1.13	1.43	1.61	1.60	2.05

Table 2. Kriging estimation results on point and block sampled data for different sampling densities for a representative soil fertility scenario.

Sampling Distance	Number of Samples	r^2 for Point Sampling	r^2 for Block Sampling - 9 Cores
40 m	441	0.801	0.850
60 m	196	0.709	0.777
100 m	81	0.577	0.675
200 m	25	0.268	0.435

Figure captions

Figure 1. White surfaces are the areas where soil cores are taken and aggregated to form a single sample for (a) point sampling and (b) block sampling.

Figure 2. Theoretical spherical covariogram with R = 100m, c_0=0 and c_1=1, and the covariogram of a generated layer.

Figure 3. Sixteen neighboring samples were used to estimate points within a block.

Figure 4. Influence of grid distance on the error variance of (a) kriging and (b) inverse distance interpolation with p = 2 obtained for exponential, spherical and Gaussian covariograms

Figure 5. Estimation error variance comparison of kriging and inverse distance interpolators with p = 0.5, 1 and 2 as a function of (grid distance / range).

Figure 6. Block sampling pattern for sparse within-block sampling with 9 cores.

Figure 7. Frequency responses of block sampling patterns as a function of standardized frequency, f/2D.

Figure 8. Kriging estimation error variance on point sampled data and block sampled data satisfying (a) spherical, (b) exponential, and (c) Gaussian correlograms.

Figure 9. Inverse distance interpolation with p = 2 estimation error variance on point sampled data and block sampled data satisfying (a) spherical, (b) exponential, and (c) Gaussian correlograms.

Figure 10. Normalized error variance on spatial data with a spherical covariogram and sampling distance D = R as a function of block size for dense within-block sampling.

Figure 11. Sampling grid distances necessary to achieve kriging normalized errors less than 0.6 for point and sparse block sampling. The covariogram is spherical with a range of 100 m and a nugget effect of 0.

Figure 12. Kriging error variance for point sampled data and two types of block sampled data (dense and sparse) along a line AB from Figure 3, at sampling density D = R.

Figure 13. (a) Original layer and kriged layer on (b) point and (c) block sampled data with sampling frequency of 100 m.

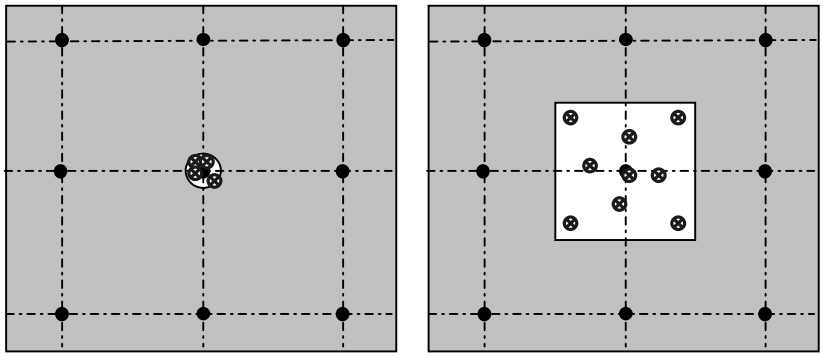

Figure 1.

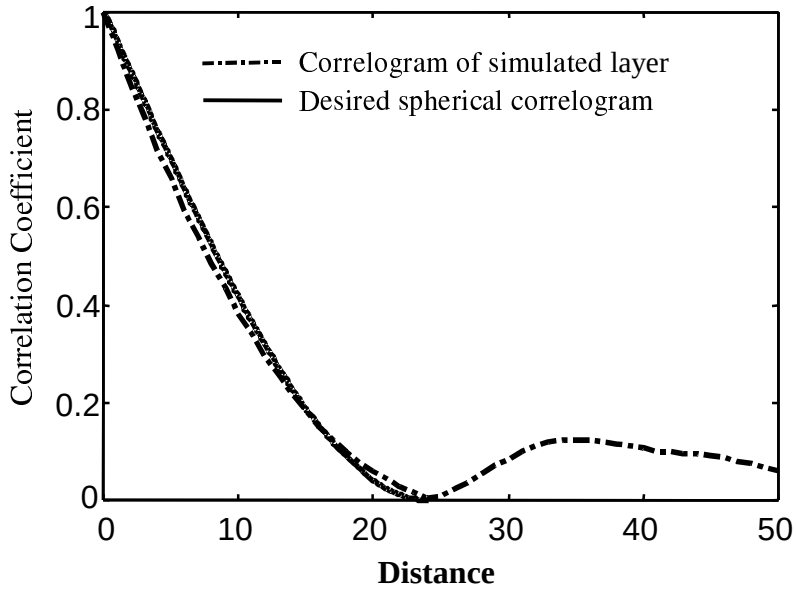
Correlogram of simulated layer
Desired spherical correlogram
Correlation Coefficient
Distance
0
0.2
0.4
0.6
0.8
1
0
10
20
30
40
50

Figure 2.

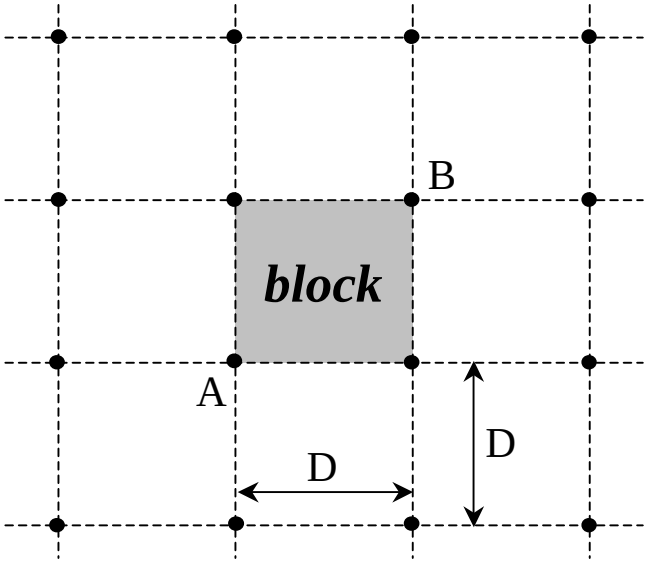

Figure 3.

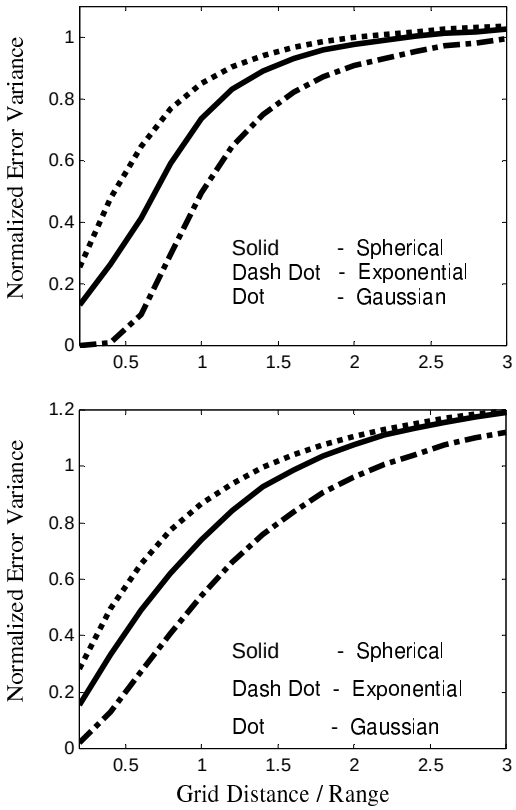
Normalized Error Variance
Solid - Spherical
Dash Dot - Exponential
Dot - Gaussian
Normalized Error Variance
Solid - Spherical
Dash Dot - Exponential
Dot - Gaussian
Grid Distance / Range

Figure 4.

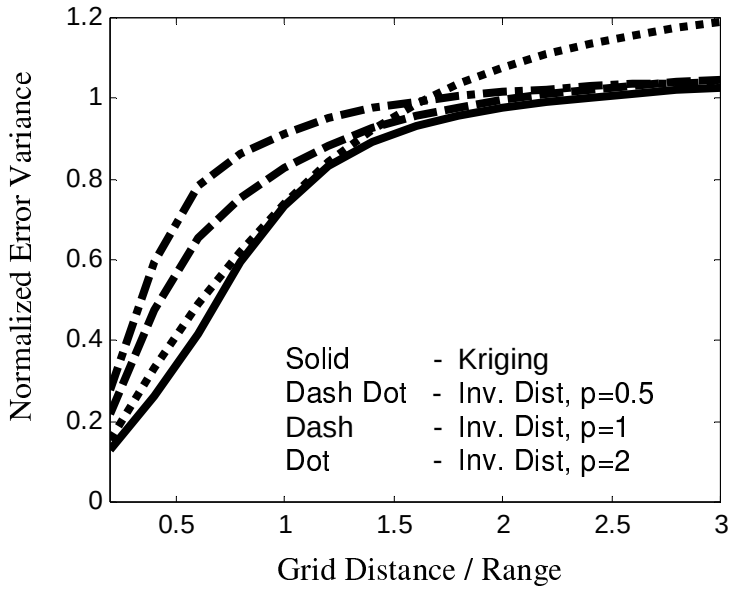
Normalized Error Variance
Grid Distance / Range
Solid - Kriging
Dash Dot - Inv. Dist, p=0.5
Dash - Inv. Dist, p=1
Dot - Inv. Dist, p=2
0
0.2
0.4
0.6
0.8
1
1.2
0.5
1
1.5
2
2.5
3

Figure 5.

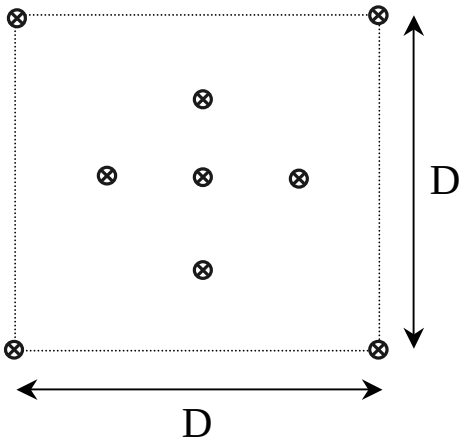

Figure 6.

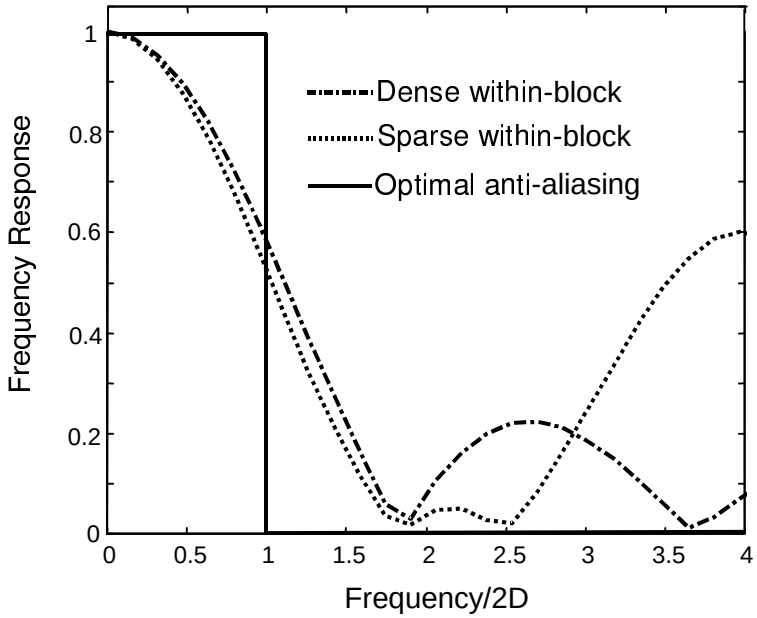
Frequency Response
Frequency/2D
Dense within-block
Sparse within-block
Optimal anti-aliasing
0
0.2
0.4
0.6
0.8
1
0
0.5
1
1.5
2
2.5
3
3.5
4

Figure 7.

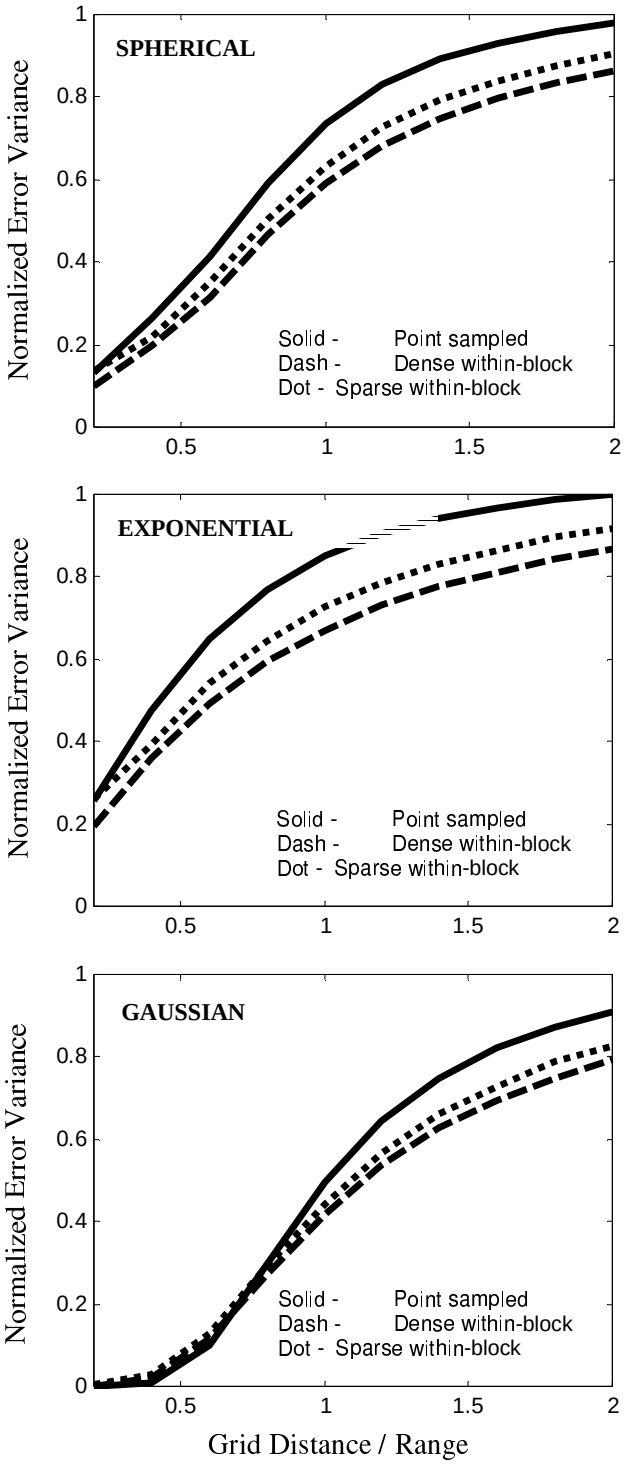
SPHERICAL
EXPONENTIAL
GAUSSIAN
Normalized Error Variance
Solid - Point sampled
Dash - Dense within-block
Dot - Sparse within-block
0
0.2
0.4
0.6
0.8
1
0.5
1
1.5
2
Grid Distance / Range

Figure 8.

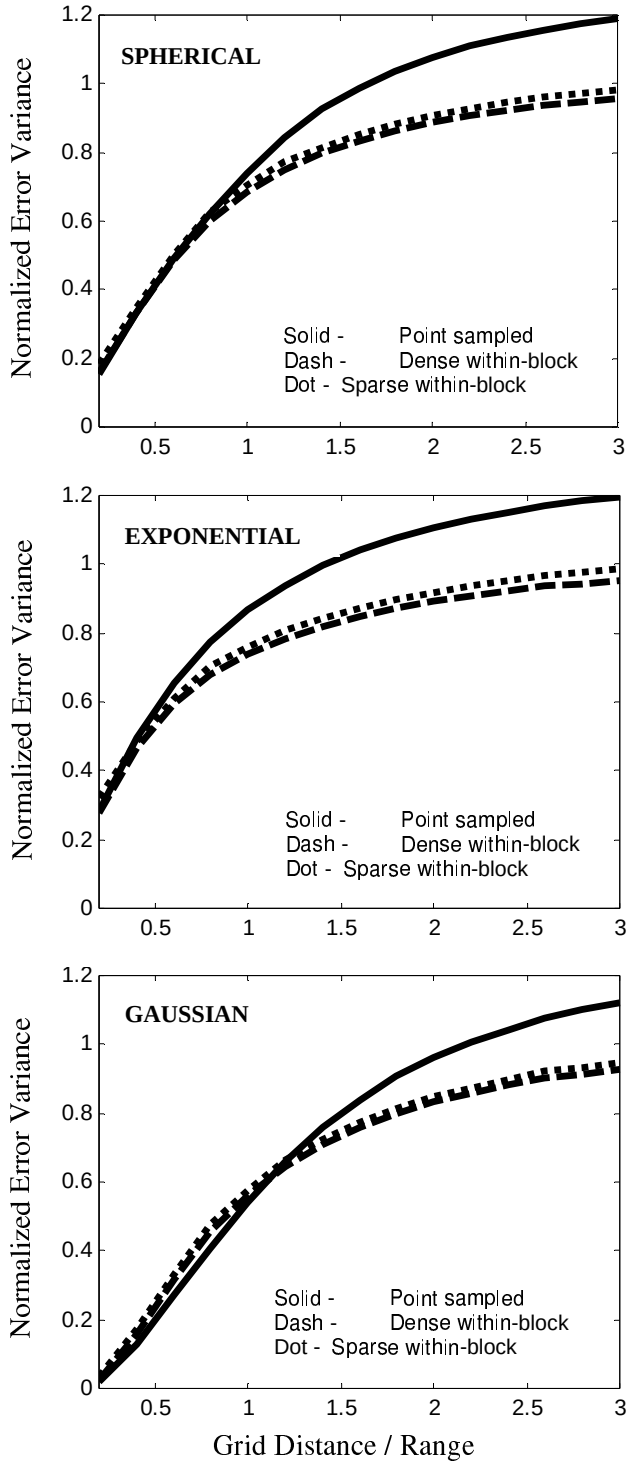
SPHERICAL
Normalized Error Variance
Solid - Point sampled
Dash - Dense within-block
Dot - Sparse within-block
EXPONENTIAL
Normalized Error Variance
Solid - Point sampled
Dash - Dense within-block
Dot - Sparse within-block
GAUSSIAN
Normalized Error Variance
Solid - Point sampled
Dash - Dense within-block
Dot - Sparse within-block
0 0.2 0.4 0.6 0.8 1 1.2
0.5 1 1.5 2 2.5 3
Grid Distance / Range

Figure 9.

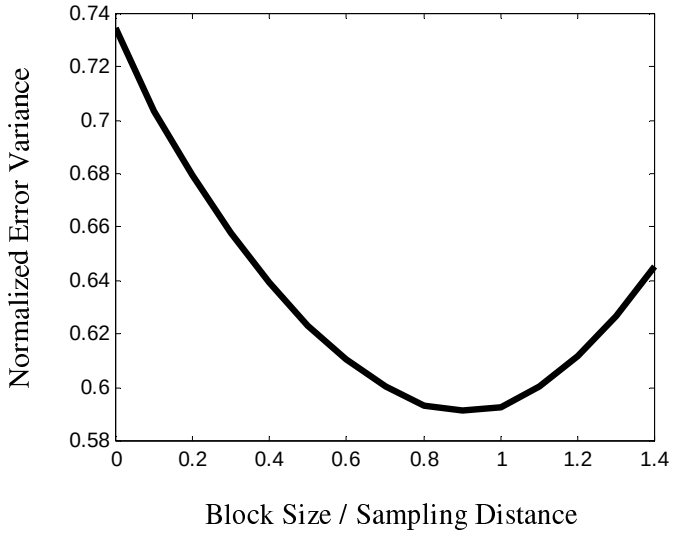
Normalized Error Variance
0.74
0.72
0.7
0.68
0.66
0.64
0.62
0.6
0.58
0
0.2
0.4
0.6
0.8
1
1.2
1.4
Block Size / Sampling Distance

Figure 10.

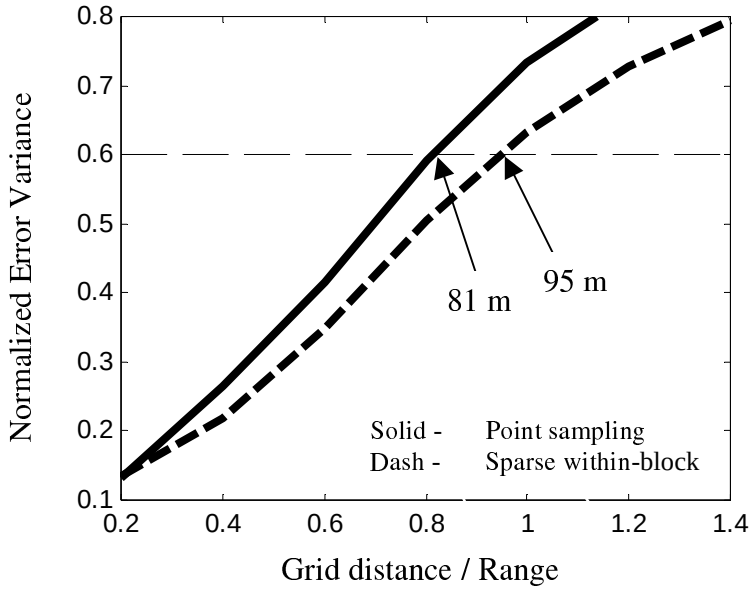
Normalized Error Variance
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.2
0.4
0.6
0.8
1
1.2
1.4
Grid distance / Range
81 m
95 m
Solid - Point sampling
Dash - Sparse within-block

Figure 11.

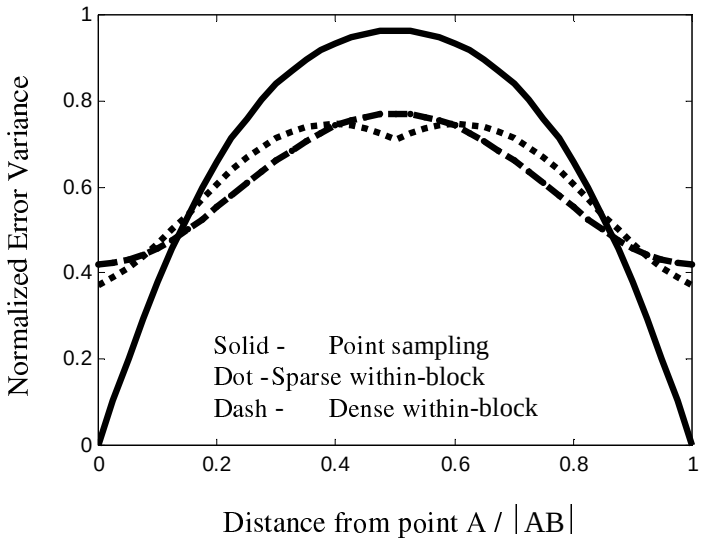
Normalized Error Variance
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
Solid - Point sampling
Dot - Sparse within-block
Dash - Dense within-block
Distance from point A / |AB|

Figure 12.

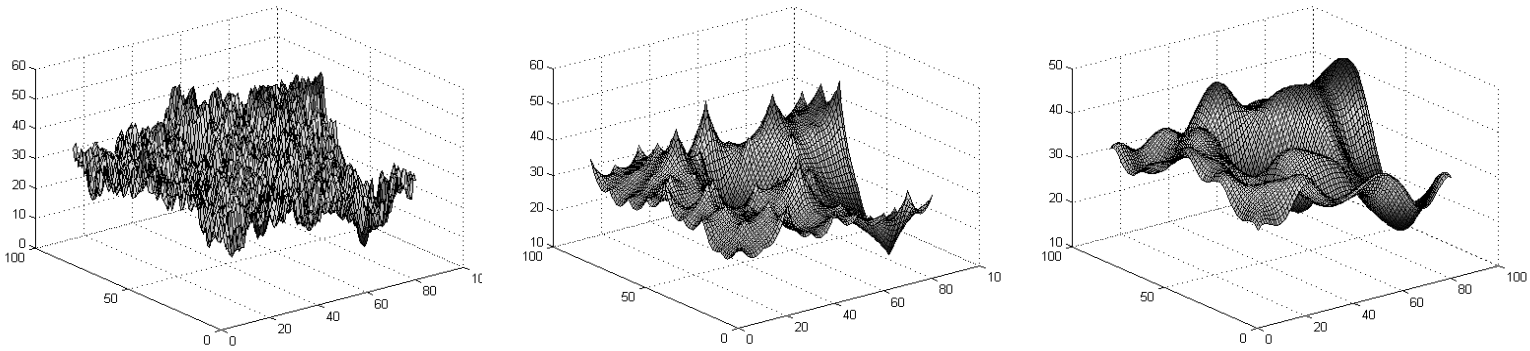
60
50
40
30
20
10
0
100
50
0
0
20
40
60
80
10
60
50
40
30
20
10
100
50
0
0
20
40
60
80
10
50
40
30
20
10
100
50
0
0
20
40
60
80
100

Figure 13.