WATER RESOURCES RESEARCH, VOL. 32, NO. 6, PAGES 1643–1652, JUNE 1996

A fast and exact method for multidimensional Gaussian stochastic simulations: Extension to realizations conditioned on direct and indirect measurements

C. R. Dietrich

Faculty of Environmental Sciences, School of Environmental Engineering, Griffith University
Brisbane, Queensland, Australia

G. N. Newsam[1]

Information Technology Division, Defence Science and Technology Organisation, Salisbury, South Australia

Abstract. Recently *Dietrich and Newsam* [1993] derived a fast and exact method for generating unconditional realizations of a stationary, multidimensional Gaussian random field on a rectangular sampling grid. The method is based on embedding the random field covariance matrix in a larger positive definite matrix with circulant/block circulant structure. The circulant structure of the embedding matrix means that a square root of this matrix can be efficiently computed by the fast Fourier transform; realizations are then generated by multiplying vectors of white noise by this square root. This paper extends the method to generating realizations conditioned on direct and/or indirect measurements of the field collected at an arbitrary set of scattered data points.

1. Introduction

Recently, *Dietrich and Newsam* [1993] (hereafter referred to as DN) presented a fast and exact method for generating unconditional realizations of a stationary, multidimensional Gaussian random field on a rectangular sampling grid. The method is based on embedding the random field covariance matrix in a larger positive definite matrix with circulant/block circulant structure. As circulant matrices are diagonalized by the fast Fourier transform (FFT), a square root of this matrix can be easily constructed. Simulations of the field can then be efficiently generated by multiplying vectors of white noise by this square root, with each matrix-vector product being rapidly calculated via one FFT. The speed of the method makes it competitive with other approaches, such as the spectral and turning bands methods, while the fact that it produces simulations with exactly the required structure gives it an edge over these methods as they are only approximate.

In many practical problems, however, realizations must be conditioned on available direct measurements of the actual field. Furthermore, in some cases the realizations must also be conditioned on indirect measurements that are linear functionals of the underlying field. Thus in the general case simulation involves generating realizations of a stationary Gaussian field on a rectangular sampling grid Ω_1 of n_1 nodes that are conditioned on n_2 direct measurements of the field collected on a grid Ω_2 and on m_3 indirect measurements collected on a grid Γ_3 that are linear combinations of n_3 unknown values of the field on a grid Ω_3. Owing to the cost of measurement data, in practice n_2 and m_3 are always much smaller than n_1, particularly in two- and three-dimensional simulations.

One example of indirect measurements would be a collection of local averages of the field measured at a few arbitrarily chosen points. Another example occurs in groundwater simulations where realizations are sought of the log transmissivity field conditioned on direct log transmissivity measurements and indirect piezometric head measurements. In this case, after linearization of the groundwater flow equation and removal of mean components, the piezometric head measurements are linearly related to the unknown log transmissivity values [*Hoeksema and Kitanidis*, 1984; *Ahmed and de Marsily*, 1987].

As we shall see, conditional realizations can be easily derived from unconditional realizations of the field over the grid $\Omega_1 \cup \Omega_2 \cup \Gamma_3$, if these are available. The original algorithm presented in DN required that the covariance matrix R of the vector z of values in the realization be a block Toeplitz matrix with Toeplitz blocks. In the absence of indirect data, i.e., $m_3 = 0$, this condition will be satisfied if the field is stationary and the grid $\Omega_1 \cup \Omega_2$ is a rectangular mesh. In practice, the sampling grid Ω_1 is generally rectangular, but the observation grid Ω_2 is usually arbitrary. Thus the whole grid $\Omega_1 \cup \Omega_2$ is not rectangular, although, since $n_2 \ll n_1$, it is "nearly rectangular." This implies that the covariance matrix R is only "nearly Toeplitz." Given the advantages of the approach of DN, the question of interest is therefore whether it can be extended to compute z efficiently on nearly rectangular grids.

One way to overcome this problem is to move each node in Ω_2 to the nearest node of Ω_1 [*Davis*, 1987b]. The perturbed grid is now rectangular; this allows the algorithm of DN to be applied without modification. However, as shown in the next section, such a perturbation may introduce unacceptable errors. Moreover, if, in addition to direct field data, the simulations are also conditioned on indirect field measurements then it is no longer possible to make R exactly block Toeplitz with Toeplitz blocks by just perturbing the grid. Therefore this paper extends the original algorithm of DN to matrices R that

[1]Also at the Cooperative Research Centre for Sensor Signal and Information Processing, Salisbury, South Australia.

Paper number 94WR02977.
0043-1397/96/94WR-02977$09.00

are "nearly Toeplitz." The extension will efficiently generate conditional simulations that have exactly the required covariance structure on the given grid with a setup cost of $O[n_1(n_2 + m_3)$ max $\{\log n_1; n_2 + m_3\}]$ floating point operations (flops) followed by a cost of $O[n_1$ max $\{\log n_1; n_2 + m_3\}]$ flops to compute each realization. Storage requirements are only $O[n_1(n_2 + m_3)]$.

The extension is based on the observation that conditioning on the direct and indirect observations essentially induces a low rank perturbation in the covariance matrix $\bar{R}_{11}$ of the unconditional distribution on the periodic extension $\bar{\Omega}_1$ of the domain Ω_1. Thus after using the algorithm of DN to find a square root $\bar{R}_{11}^{1/2}$ of $\bar{R}_{11}$, we can account for the conditioning by a low rank update of $\bar{R}_{11}^{1/2}$. The result is a fast algorithm that is competitive with alternative methods already proposed for this problem.

The structure of the paper is as follows. The next section briefly reviews the definition and construction of simulations of a Gaussian random field conditioned on direct measurements. Section 3 then shows how to extend the algorithm of DN to include conditioning on direct observations, and section 4 provides numerical examples in which the algorithm has been used to compute some conditioned realizations. Section 5 considers the more general case where the realizations are also conditioned on indirect measurements that are linear functionals of the underlying field. Finally, section 6 summarizes the algorithm and discusses its advantages and disadvantages when compared with other approaches for generating simulations, such as the turning bands method and Cholesky factorizations.

2. Simulations Conditioned on Direct Measurements

We shall only consider here the problem of generating two-dimensional simulations; simulations in three or more dimensions can be produced by a straightforward extension of the two-dimensional algorithm. Suppose that realizations of a stationary Gaussian field $Z(x, y)$ are to be generated on the $n_1 = (p + 1)(q + 1)$ nodes of the rectangular grid $\Omega_1 = \{(x_i, y_j) = (i\delta x, j\delta y)\colon 0 \le i \le p,\ 0 \le j \le q\}$, where δx and δy are the length and breadth of individual cells in the mesh, respectively. Now suppose that the realizations are to be conditioned on the vector z_2^* of n_2 measurements of Z on the grid $\Omega_2 = \{(x_k, y_k)\colon 1 \le k \le n_2\}$ (i.e., $[z_2^*]_k = Z(x_k, y_k)$). Without loss of generality, we may assume that Ω_1 and Ω_2 have no common nodes and Ω_2 lies within the boundary of Ω_1.

As Z is a stationary Gaussian random field, it is completely determined by its mean and covariance function. Since Z is stationary, the mean $\mu = E[Z(x, y)]$ is a constant: Without loss of generality we may assume that $\mu = 0$. Thus the covariance function $r(x, y)$ is given by

$$r(x - x', y - y') = E[Z(x, y)Z(x', y')] \tag{1}$$

Now if z_1 and z_2 are vectors of samples of Z on the grids Ω_1 and Ω_2 respectively, the random vector

$$z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \tag{2}$$

has zero mean and covariance given by the following matrix:

$$R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \tag{3}$$

Here the R_{ij} are matrices of covariances between field values at points in Ω_i and Ω_j, so that

$$[R_{ij}]_{mn} = r(x_m - x_n, y_m - y_n) \tag{4}$$

where $(x_m, y_m) \in \Omega_i$ and $(x_n, y_n) \in \Omega_j$ (note that $R_{21} = R_{12}^T$). Because Ω_1 is rectangular, the block R_{11} is quadrantally Toeplitz, i.e., it is symmetric block Toeplitz, with each block being itself symmetric and Toeplitz (see DN for details).

It is well known [e.g., *Anderson*, 1984] that the distribution of z_1 conditioned on an observation z_2^* of z_2 on Ω_2 is normal with mean

$$\mu_{z_1|z_2^*} = R_{12}R_{22}^{-1}z_2^* \tag{5}$$

and covariance

$$R_{z_1|z_2^*} = R_{11} - R_{12}R_{22}^{-1}R_{21} \tag{6}$$

Consequently, it is easy to check that if z in (2) is a random variable that is unconditionally distributed as $\mathcal{N}(0, R)$ on the grid $\Omega_1 \cup \Omega_2$, then the random variable

$$z_1|z_2^* = \mu_{z_1|z_2^*} + z_1 - R_{12}R_{22}^{-1}z_2 \tag{7}$$

has mean given by (5) and covariance given by (6), and so has the required distribution for a conditional simulation on Ω_1. Thus a method for efficient computation of unconditional simulations on $\Omega_1 \cup \Omega_2$ can easily be adapted to give a method for generating conditional simulations on Ω_1. This is in essence the approach adopted by *Journel and Huijbregts* [1978].

In practice, many methods for efficiently generating unconditional simulations (e.g., the popular spectral method of *Shinozuka and Jan* [1972]) only work on rectangular grids. Therefore in order to use such methods, as described in the introduction the observation grid Ω_2 is perturbed to coincide with the regular grid Ω_1. As the covariance function $r(x, y)$ is continuous, except possibly at the origin, a small change in Ω_2 will produce only a small change in R_{12} and R_{22}. Equation (7), however, shows that the conditional simulations depend on the inverse R_{22}^{-1}. Thus if R_{22} is poorly conditioned, a small perturbation in Ω_2 may still lead to a large perturbation in R_{22}^{-1} and so affect significantly the accuracy of the conditional mean (5) (see *Cressie and Zimmerman* [1992] for an analysis of this sensitivity). One situation where this is likely to occur is when two or more points of Ω_2 fall within one cell of Ω_1; the perturbation could end up moving them all onto the same grid point in Ω_1 and so severely distort any local information in the measurements. Such sensitivity may degrade the accuracy of any conditional simulations calculated from (7) using a perturbed grid. For this reason we present here an extension to DN's algorithm that does not require perturbation of the measurement grid.

The extended algorithm for computing an unconditional realization on a nearly rectangular mesh is detailed in the next section. Before closing this section, however, we note two short points.

First, given an unconditional realization, we briefly identify the additional steps needed to compute a conditional realization and estimate their cost. Given an unconditional realization z, to calculate a conditional realization we must first compute the conditional mean $\mu_{z_1|z_2^*} = R_{12}R_{22}^{-1}z_2^*$ (note that this need only be done once in computing a collection of realizations), and then compute $\mu_{z_1|z_2^*} + z_1 - R_{12}R_{22}^{-1}z_2$. The computational costs of this are essentially those of solving a linear system with matrix R_{22} (i.e., $O(n_2^3)$ flops), of computing the product of the matrix R_{12} with a vector of dimension n_2 (*i.e.*, $O(n_1n_2)$ flops), and of adding together vectors of dimension

n_1 (i.e., $O(n_1)$ flops). If $n_2 \ll n_1$, then this cost is effectively $O(n_1 n_2)$ flops.

Second, in practice, measurements z_2^* of Z on Ω_2 are likely to be noisy. This is usually modeled by assuming that $[z_2^*]_k = Z(x_k, y_k) + \epsilon_k$ where the errors ϵ_k are zero mean Gaussian random variables which may be correlated among themselves with correlation matrix Σ, but are independent of the vectors z_1 and z_2. This extension can be easily accommodated in the results presented above by simply replacing the matrix R_{22} in (3) and subsequent equations by $\bar{R}_{22} = R_{22} + \Sigma$.

3. Extension of the Circulant Embedding Approach

As in DN, our approach to generating unconditional realizations $z \sim \mathcal{N}(0, R)$ relies on the fact the quadrantally Toeplitz block R_{11} in (3) can be embedded in a matrix $\bar{R}_{11}$ with circulant structure. We shall give a slightly different presentation of the construction of the embedding to that in DN in order to simplify inclusion of the contributions from the observations. The embedding is now carried out as follows.

Let $P = p\delta x$ and $Q = q\delta y$ be the length and breadth of Ω_1 and consider the function $\bar{r}(x, y)$ defined as follows

$$\begin{aligned}
\bar{r}(x, y) &= r(x, y) && 0 \le x \le P,\ 0 \le y \le Q \\
\bar{r}(x, y) &= r(2P - x, y) && P \le x \le 2P,\ 0 \le y \le Q \\
\bar{r}(x, y) &= r(x, 2Q - y) && 0 \le x \le P,\ Q \le y \le 2Q \\
\bar{r}(x, y) &= r(2P - x, 2Q - y) && P \le x \le 2P,\ Q \le y \le 2Q
\end{aligned} \tag{8}$$

Observe that $\bar{r}(x, y)$ has been obtained by extending that part of $r(x, y)$ in $[0, P] \times [0, Q]$ to $[0, 2P] \times [0, 2Q]$ by wrap-around in both the x and y directions. We now periodically extend $\bar{r}(x, y)$ to the entire plane in the expectation that this extension will be the covariance function of a zero mean periodic stationary random field, denoted $\bar{Z}(x, y)$, in which a single period is contained in the cell $[0, 2P] \times [0, 2Q]$. We assume for the moment that this is indeed the case: Conditions under which this is true will be discussed later in the section.

Next we enlarge Ω_1 to the grid

$$\begin{aligned}
\bar{\Omega}_1 &= \{(x_i, y_j) \\
&= (i\delta x, j\delta y):\ 0 \le i \le 2p - 1,\ 0 \le j \le 2q - 1\}
\end{aligned} \tag{9}$$

Thus $\bar{\Omega}_1$ is composed of $\bar{n}_1 = 4pq$ nodes: Clearly $\Omega_1 \subset \bar{\Omega}_1$. Furthermore the vector

$$\bar{z} = \begin{pmatrix} \bar{z}_1 \\ z_2 \end{pmatrix} \tag{10}$$

denotes the vector of samples of $\bar{Z}(x, y)$ on $\bar{\Omega}_1 \cup \Omega_2$. Since we have assumed that $\bar{r}$ is a covariance function, the covariance matrix $\bar{R}$ of $\bar{z}$ will be positive definite. Moreover the domain partitioning $\bar{\Omega}_1 \cup \Omega_2$ induces a corresponding partitioning of $\bar{R}$ into

$$\bar{R} = \begin{pmatrix} \bar{R}_{11} & \bar{R}_{12} \\ \bar{R}_{21} & R_{22} \end{pmatrix} \tag{11}$$

with $\bar{R}_{11}$ being the covariance matrix among pairs of points of $\bar{\Omega}_1$, $\bar{R}_{12}$ being the covariance matrix between pairs of points of $\bar{\Omega}_1$ and Ω_2, and $\bar{R}_{21} = \bar{R}_{12}^T$. Now as the restriction of $\bar{r}$ to $[0, P] \times [0, Q]$ is exactly r, the restriction of $\bar{z}$ to $\Omega_1 \cup \Omega_2$ has R as its covariance matrix. Thus once a realization $\bar{z}$ is computed over $\bar{\Omega}_1 \cup \Omega_2$, its restriction to $\Omega_1 \cup \Omega_2$ immediately yields a realization of z.

Following DN, to generate a realization $\bar{z}$ we first find an efficient square root $\bar{R}^{1/2}$ of $\bar{R}$. That is, we construct a matrix $\bar{R}^{1/2}$ with the property that $\bar{R}^{1/2}(\bar{R}^{1/2})^H = \bar{R}$ (here superscript H denotes the conjugate transpose) such that the matrix-vector product $\bar{R}^{1/2}\epsilon$ can be efficiently computed for any complex vector ϵ. If $\epsilon \sim \mathcal{N}(0, I)$ is taken to be a vector of independent zero mean complex normal random variables with unit variance, then the realization $\bar{z} = \bar{R}^{1/2}\epsilon$ will have exactly the desired distribution; in particular it will have $\bar{R}$ as its covariance matrix.

To find such a matrix $\bar{R}^{1/2}$, first note that since $\bar{R}$ is positive definite, the upper rectangular block $\bar{R}_{11}$ is also positive definite. Moreover, $\bar{R}_{11}$ is circulant: This is due to the periodicity of $\bar{r}$, the manner in which $\bar{\Omega}_1$ has been defined, and the stationarity of $\bar{Z}(x, y)$. Hence $\bar{R}_{11}$ is uniquely determined by its first column vector $\bar{\rho}$ and admits the decomposition

$$\bar{R}_{11} = W\Lambda W^H \tag{12}$$

where W is the normalized two-dimensional discrete Fourier transform matrix with entries $w_{jukv} = \exp\{[2\pi i(ju/2p + kv/2q)]/(\bar{n}_1)^{1/2}\}$ for $0 \le j, u \le 2p - 1$ and $0 \le k, v \le 2q - 1$, and Λ is a diagonal matrix whose diagonal entries are nonnegative and are given by the vector $(\bar{n}_1)^{1/2}W\bar{\rho}$. Again, a more detailed description of this embedding technique can be found in DN: There $\bar{R}_{11}$ was directly derived as the smallest circulant embedding of the quadrantally Toeplitz matrix R_{11} (the matrices R_{11} and $\bar{R}_{11}$ correspond to the matrices R and S on p. 2863 of DN). We emphasize again that in this note we have presented the circulant embedding strategy in a different, though equivalent, fashion in order to take into account the extra interactions with the conditioning data.

Since the normalized FFT matrix W is unitary (i.e. $WW^H = W^HW = I_{\bar{n}_1}$ where $I_{\bar{n}_1}$ is the identity matrix of dimension $\bar{n}_1$), simple multiplication of block matrices shows that the matrix

$$\bar{R}^{1/2} = \begin{pmatrix} W\Lambda^{1/2} & 0 \\ K & L \end{pmatrix} \tag{13}$$

where $K = \bar{R}_{21}W\Lambda^{-1/2}$ and L is any matrix (e.g., a Cholesky factor) such that $LL^T = R_{22} - KK^H$, is indeed a square root $\bar{R}^{1/2}$ of $\bar{R}$. Note that $KK^H = \bar{R}_{21}\bar{R}_{11}^{-1}\bar{R}_{12}$ so that $R_{22} - KK^H$ has real entries and is the inverse of the lower right square block of dimension n_2 of $\bar{R}^{-1}$ [*Schweppe*, 1973, p. 495]. Since $\bar{R}$ is positive definite, $\bar{R}^{-1}$ is positive definite. In turn, the inverse of any lower right square block in $\bar{R}^{-1}$ is positive definite. This ensures the existence of L.

Given $\bar{R}^{1/2}$, if $\epsilon \sim \mathcal{N}(0, I)$ is a complex white noise vector, then the real and imaginary parts of $\bar{z} = \bar{R}^{1/2}\epsilon$ are two independent realizations of $\bar{z}$. The restriction of the real and imaginary parts of $\bar{z}$ to $\Omega_1 \cup \Omega_2$ thus yields two independent realizations of z. Rewriting the equality $\bar{z} = \bar{R}^{1/2}\epsilon$ in terms of the partitioning of $\bar{R}^{1/2}$ in (13) shows that

$$\bar{z} := \begin{pmatrix} \bar{z}_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} W\Lambda^{1/2} & 0 \\ K & L \end{pmatrix} \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$$

or

$$\bar{z}_1 = W\Lambda^{1/2}\epsilon_1 \tag{14}$$

$$z_2 = K\epsilon_1 + L\epsilon_2 \tag{15}$$

The computational costs of generating realizations of $\bar{z}$ are as follows. First there is the setup cost of forming Λ, K, and L. The FFT yields Λ in $O(\bar{n}_1 \log \bar{n}_1)$ flops. Computing $K^H = \Lambda^{-1/2} W^H \bar{R}_{12}$ requires the inverse FFT of each of the n_2 columns of $\bar{R}_{12}$, and then scaling the elements of the resultant vectors by $\Lambda^{-1/2}$. The cost of this is dominated by the $O(n_2 \bar{n}_1 \log \bar{n}_1)$ flops needed to compute the inverse FFTs and requires $O(n_2 \bar{n}_1)$ storage. Another $O(n_2^2 \bar{n}_1)$ flops are then required to form the whole product KK^H, followed by a further $O(n_2^3)$ flops to compute $R_{22} - KK^H$ and its Cholesky decomposition. As $\bar{n}_1 \approx 4n_1$ and $n_2 \ll n_1$, the total cost is $O[n_1 n_2 \max \{\log n_1; n_2\}]$ flops. Once K and L are available, each successive realization $\bar{z}$ involves calculating $\bar{z}_1$ by the FFT and z_2 by straightforward matrix vector products. Thus the cost of computing $\bar{z}$ is $O(n_1 \log n_1) + O(n_1 n_2)$ flops.

In summary, the total computational cost of calculating a succession of realizations consists of a setup cost of $O[n_1 n_2 \max \{\log n_1; n_2\}]$, followed by a cost of $O[n_1 \max \{\log n_1; n_2\}]$ for computing each realization. In-core memory storage requirements are dominated by the size of the matrix $\bar{R}_{12}$ and so are only $O(n_1 n_2)$.

The above methodology fails if the embedding matrix $\bar{R}$ in (11) is not positive definite; in this case it is not possible to find any matrix S such that $SS^H = \bar{R}$. Therefore it is important to establish conditions that will ensure $\bar{R}$ is positive definite. By the definition of a positive definite function, this in turn can be assured by finding conditions under which the periodic extension $\bar{r}$ of a positive definite covariance function r is positive definite. This issue has been explored by *Dietrich and Newsam* [1996] and *Wood* [1993], with the following results.

Given a positive definite r, so far only one set of conditions has been found that will guarantee that a positive definite periodic extension $\bar{r}$ of r can be constructed on every rectangle $[-P, P] \times [-Q, Q]$: This is that $\Delta r \geq 0$ and that r be decreasing away from the origin. Unfortunately, in two or more dimensions this condition is not satisfied by any of the standard covariance functions since it implies that r is unbounded at the origin. Nevertheless, while for a given r it is not possible to construct positive definite periodic extensions on every rectangle, it is still possible to construct such extensions on some rectangles: In particular, if r vanishes outside the rectangle $[-P, P] \times [-Q, Q]$ then the periodic extension $\bar{r}$ is guaranteed to be a covariance function. This result was apparently first noted by *Woods* [1972]; its proof is a simple special case of the argument in the next paragraph. Moreover, it indicates how we can generally ensure that $\bar{R}$ (or at worst a small perturbation of $\bar{R}$) is positive definite: Simply choose P and Q to be sufficiently large so that r essentially vanishes outside $[-P, P] \times [-Q, Q]$.

To see this, since r is positive definite then by Bochner's theorem [*Champeney*, 1987] it has a positive Fourier transform; i.e., for all u and v

$$\rho(u, v) = \int_0^\infty \int_0^\infty r(x, y) \cos [2\pi(ux + vy)] \, dx \, dy \geq 0 \tag{16}$$

Now for the periodic extension $\bar{r}$ to be a covariance function, it must also be positive definite on the cell $[0, 2P] \times [0, 2Q]$. By the analogue of Bochner's theorem for Fourier series this is only possible if the coefficients in the Fourier series expansion of $\bar{r}$ are all positive, i.e., if for all $m, n \geq 0$

$$\rho_{mn} = \int_0^P \int_0^Q r(x, y) \cos [\pi(mx/P + ny/Q)] \, dx \, dy \geq 0 \tag{17}$$

If P and Q are sufficiently large, then (16) implies that the integral in (17) is also likely to be nonnegative.

It is clearly possible, however, that P and Q may have to be very large indeed to ensure that the embedding leads to a positive definite $\bar{R}$, which would result in having to compute unacceptably large simulations. This difficulty can be avoided at the cost of introducing a small nugget effect into the simulations. Indeed, a nugget effect of size θ in the field Z ensures that $\rho(u, v) \geq \theta$ in (16). This lower bound will reduce the magnitude of the values of P and Q required to ensure that $\bar{R}$ is positive definite. In practice, given the degree of uncertainty in most geostatistical models, including a nugget effect with θ of the order of $10^{-3} \times r(0, 0)$ will give a covariance function that is consistent with the model for all practical purposes and which will have a positive definite periodic embedding for relatively small P and Q.

The extension of DN's algorithm to conditional simulations can now be summarized as follows.

1 Compute the first column $\bar{\rho}$ of $\bar{R}_{11}$ whose $\bar{n}_1$ entries are $\bar{r}(i\delta x, j\delta y)$, where $\bar{r}(x, y)$ is obtained from the covariance function $r(x, y)$ via (8). Also compute the entries $[\bar{R}_{12}]_{ij,k} = \bar{r}(i\delta x - x_k, j\delta y - y_k)$. (Here $0 \leq i \leq 2p - 1$, $0 \leq j \leq 2q - 1$, and $1 \leq k \leq n_2$.)

2. Calculate the entries of the diagonal matrix $\Lambda = (\bar{n}_1)^{1/2} W\bar{\rho}$ by the two-dimensional FFT. Check that each entry of $(\bar{n}_1)^{1/2} W\bar{\rho}$ is nonnegative. If an entry is negative, increase P and Q and/or include a small nugget effect in the simulation.

3. Compute the product $W^H \bar{R}_{12}$ by taking the inverse discrete Fourier transform of each column of $\bar{R}_{12}$ and then forming $K^H = \Lambda^{-1/2}(W^H \bar{R}_{12})$. Compute the matrix $R_{22} - KK^H$ and then form its Cholesky factorization LL^T. If the Cholesky factorization fails, indicating $R_{22} - KK^H$ is not positive definite, then again increase P and Q and/or increase the size of the nugget effect.

4. For each individual simulation, first form a vector $\epsilon = (\epsilon_1 | \epsilon_2)$ with ϵ_1 and ϵ_2 of dimensions $\bar{n}_1$ and n_2, respectively, whose entries have real and imaginary components that are zero mean independent normal variates with unit variance. Second, compute the vector $\bar{z}_1 = W\Lambda^{1/2}\epsilon_1$ via the two-dimensional FFT, and compute the vector $z_2 = K\epsilon_1 + L\epsilon_2$. Third, restrict the real and imaginary parts of the vector $\bar{z} = (\bar{z}_1, z_2)$ to the grid $\Omega_1 \cup \Omega_2$ to get two independent unconditional realizations of z. Finally, condition each unconditional simulation z on the observations z_2^* via (7).

4. Numerical Illustrations

To illustrate the method, in this section we provide experimental covariances estimated from single unconditional realizations and density plots of conditional realizations. We have not provided experimental covariances estimated from N independent realizations as, since our approach is exact (modulo computer round-off errors as are all methods), it is easy to derive from standard statistical theory a lower bound for N that ensures that the standard maximum likelihood estimate of any covariance entry has a variance below some specified a priori tolerance value.

These illustrations are for the following three covariance functions:

Exponential

$$r(x, y) = \exp[-l(x, y)] \tag{18}$$

Whittle

$$r(x, y) = l(x, y)K_1[l(x, y)] \tag{19}$$

Tensor product of exponentials

$$r(x, y) = \exp(-|x|/l_x) \exp(-|y|/l_y) \tag{20}$$

where $l(x, y) = [(x/l_x)^2 + (y/l_y)^2]^{1/2}$, l_x and l_y are the correlation length scales in the x and y directions, and K_1 denotes the second kind modified Bessel function of order 1.

For the estimation of covariance from single unconditional realizations, the setting was as follows. The sampling domain was the square lattice $\Omega_1 = \{(x_i, y_j) = (i, j) : 0 \le i \le 127, 0 \le j \le 127\}$. In other words, the length and breadth of each mesh cell were equal to one unit, with the width and breadth P and Q of Ω_1 both equal to 127.

From a single zero mean unconditional realization z with values z_{ij} at each node (i, j), the covariance estimate $\hat{r}(m, n)$ at lag (m, n) was obtained from the unbiased estimator

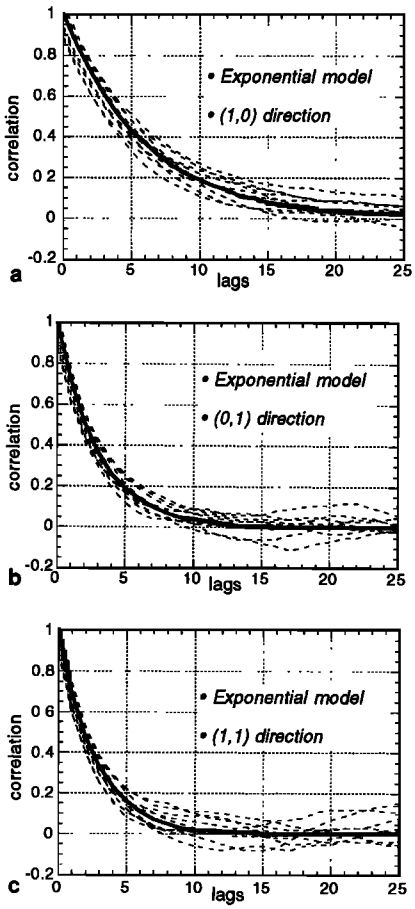

Figure 1. True covariance values (continuous line) and 10 covariance estimates obtained from 10 realizations (dashed lines) for the exponential model (equation (18)) and the first 25 lags in the (a) (1, 0), (b) (0, 1), and (c) (1, 1) directions (or similarly x, y, and $x = y$ directions).

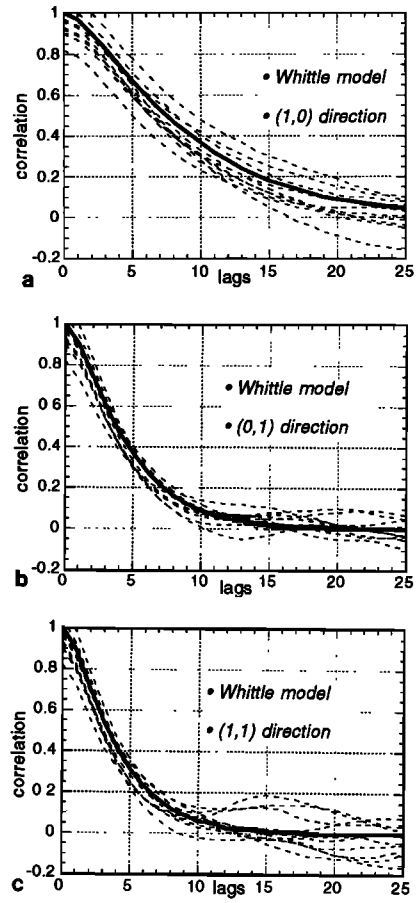

Figure 2. True covariance values (continuous line) and 10 covariance estimates obtained from 10 realizations (dashed lines) for the Whittle model (equation (19)) and the first 25 lags in the (a) (1, 0), (b) (0, 1), and (c) (1, 1) directions.

$$\hat{r}(m, n) = \frac{1}{(P - m + 1)(Q - n + 1)} \sum_{i=0}^{P-m} \sum_{j=0}^{Q-n} z_{ij} z_{i+m, j+n} \tag{21}$$

To ensure that each realization contained enough independent information to provide meaningful covariance estimates, the length scales need to be sufficiently small compared to the size of the sampling domain. To achieve this, the length scales l_x and l_y were set to 6 and 3, respectively.

For the exponential model (18) and the first 25 lags in the (1, 0), (0, 1) and (1, 1) directions (or similarly x, y and $x = y$ directions), Figure 1 plots the true covariance values (continuous line) and 10 covariance estimates obtained from 10 realizations (dashed lines). Figures 2 and 3 provide the same information for the Whittle model (19) and the tensor product of exponentials (20). The nonzero slope at the origin for the Whittle model is due to the linear interpolation used by the plotting program between discrete lags. As expected, for increasing lags less information is available to estimate correlation values, so the size of the fluctuations of the estimates around the true values increases. Nevertheless, for the first few lags, correlation estimates indicate that the approach is generating realizations with the correct distributions.

Our second set of numerical experiments was on computing conditional simulations. The steps followed here are exactly those enumerated at the end of section 3 with the setting

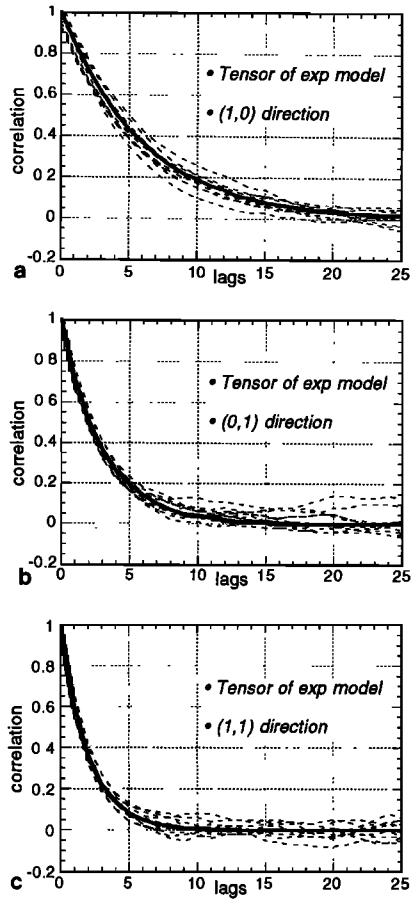

Figure 3. True covariance values (continuous line) and 10 covariance estimates obtained from 10 realizations (dashed lines) for the tensor product of exponentials (equation (20)) and the first 25 lags in the (a) (1, 0), (b) (0, 1), and (c) (1, 1) directions.

chosen as follows. The sampling domain was $\Omega_1 = \{(x_i, y_j) = (i, j) : 0 \le i \le 100,\ 0 \le j \le 80\}$ so that the length and breadth of each mesh cell were again equal to one unit and the length and breadth P and Q of Ω_1 were 100 and 80 respectively. Thus $n_1 = 101 \times 81 = 8181$. The conditioning data z_2^* consisted of $n_2 = 15$ direct measurements at the locations shown in Figure 4. The length scales l_x and l_y were set to 10 and 20 for the exponential model (18), to 10 and 6 for the Whittle model (19), and to 80 and 10 for the tensor product of exponentials (20).

Density plots of three realizations for each of the three covariance models are given, respectively, in Figures 5, 6, and 7. Again, without loss of generality, we have assumed that Z is zero mean. In each case we have also provided a contour-density plot of the associated conditional mean $\mu_{z_1|z_2^*} = R_{12}R_{22}^{-1}z_2^*$. These density plots are simply gray scale images of the conditionally simulated fields, with the field values at each node represented as a gray scale according to the legends.

The line structures in Figure 7 occur because the covariance function in (20) is separable, i.e., $r(x, y)$ can be written as the product $r(x, y) = r_1(x)r_2(y)$. This implies that the random field $Z(x, y)$ can be viewed as the product $Z(x, y) = Z_1(x)Z_2(y)$ where Z_1 and Z_2 are independent zero mean Gaussian fields with covariance $r_1(x)$ and $r_2(y)$, respectively. Such a structure may occur, for example, in horizontally stratified soils.

Finally, despite the relatively large size of the sampling domain in both sets of experiments, generating unconditional or conditional realizations took only a few minutes per realization on a Macintosh IIsi with the two-dimensional FFTs calculated by *International Mathematics and Statistics Libraries* [1991] routines.

5. Simulations Conditioned on Direct and Indirect Measurements

As indicated in the introduction, in addition to direct measurements of the field Z, indirect linear measurements of Z are also often available. That is, in addition to the vector z_2^* of values of Z on the grid Ω_2, we also have a vector $h_3^* = Az_3 + \epsilon_3$ of m_3 measurements on a grid Γ_3, where z_3 is a vector of n_3 unknown values of Z over the grid Ω_3, $\epsilon_3 \sim \mathcal{N}(0, \Sigma)$ is a vector of measurement errors that are uncorrelated with Z, and A is a $m_3 \times n_3$ matrix. As in section 2, we may assume without loss of generality that no two of Ω_1, Ω_2, and Ω_3 have nodes in common, and that Ω_2 and Ω_3 lie within the rectangular boundary of Ω_1. Note that in general n_3 will be much larger than m_3 so that A will not be invertible. A can be assumed to have full rank, however, as rank deficiency would imply that some observations in h_3^* are redundant and therefore should be removed. This could be done either by removing data in an ad hoc fashion until A has full rank, or, more systematically, via a subset selection algorithm such as that in the book by *Golub and Van Loan* [1989, p. 57].

As in generating realizations conditioned only on direct measurements, in order to generate realizations z_1 of Z on Ω_1 that are conditioned on both the direct and indirect measurements z_2^* and h_3^*, we need first to generate unconditional simulations of the zero mean Gaussian vector

$$z = \begin{pmatrix} z_1 \\ z_2 \\ h_3 \end{pmatrix}$$

with covariance matrix

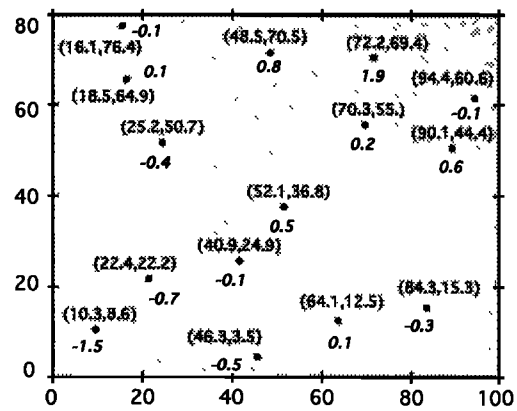

Figure 4. Location and values of the direct measurements z_2 used in the numerical illustrations. The numbers in parentheses indicate measurement locations in (x, y) coordinates while the numbers in the unshaded rectangles are the measurement values at that location.

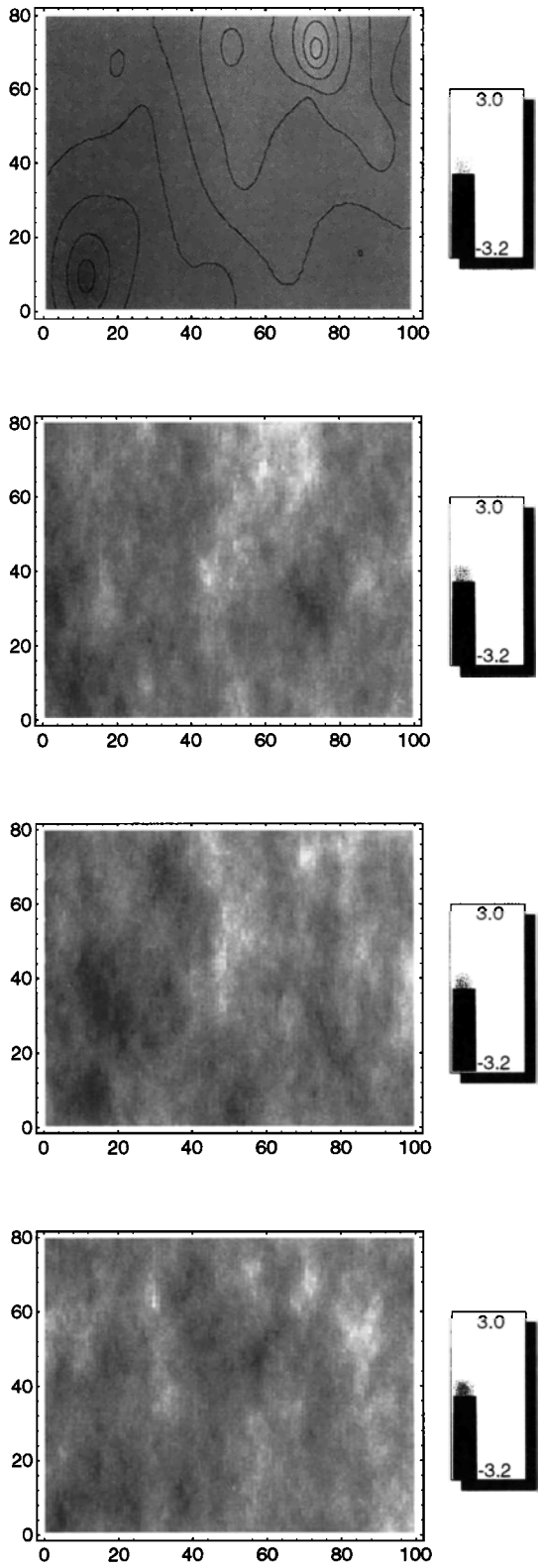

Figure 5. Density plots of three conditional realizations of $z_1|z_2$ for the exponential model (equation (18)) together with a density plot of the conditional mean $\mu_{z_1|z_2}$ with contour lines superimposed. The length scales l_x and l_y are equal to 10 and 20, respectively.

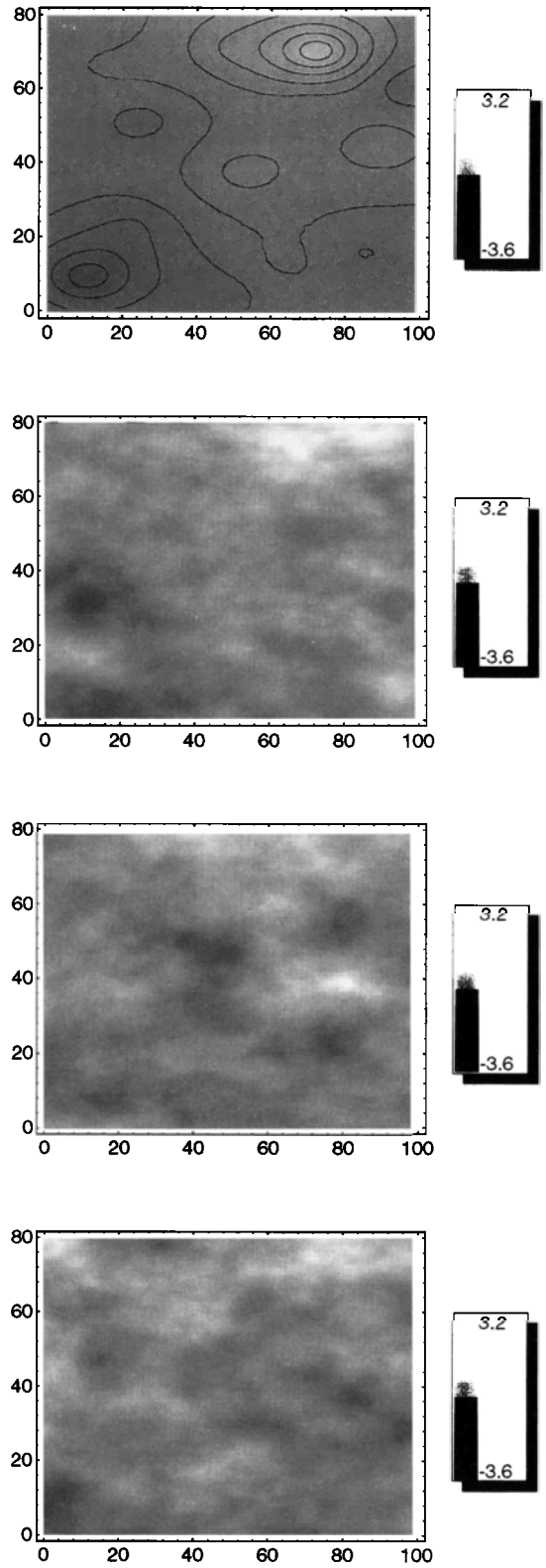

Figure 6. Density plots of three conditional realizations of $z_1|z_2$ for the Whittle model (equation (19)) together with a density plot of the conditional mean $\mu_{z_1|z_2}$ with contour lines superimposed. The length scales l_x and l_y are equal to 10 and 6, respectively.

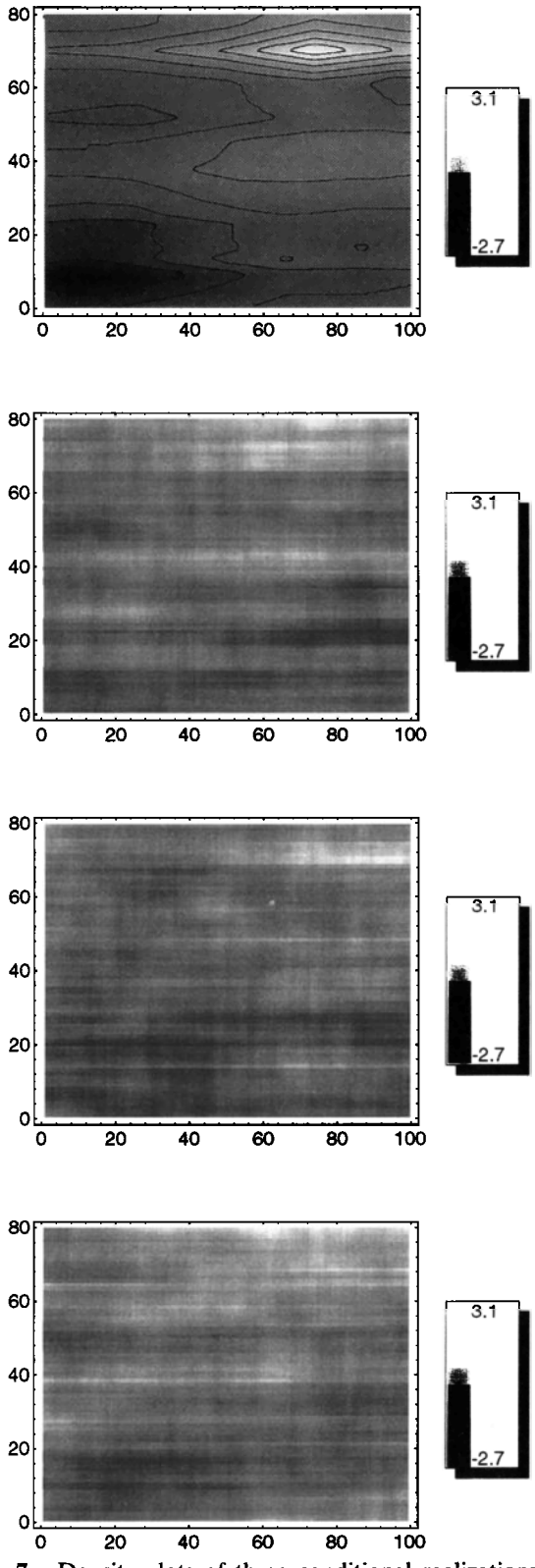

Figure 7. Density plots of three conditional realizations of $z_1|z_2$ for the tensor product of exponentials model (equation (20)) together with a density plot of the conditional mean $\mu_{z_1|z_2}$ with contour lines superimposed. The length scales l_x and l_y are equal to 80 and 10, respectively. The line structures are due to the fact that the covariance function is separable, so the random field $Y(x, y)$ can be decomposed as a product of one-dimensional fields $Y(x, y) = Y_1(x)Y_2(y)$.

$$R = E[zz^T] = \begin{pmatrix} E[z_1z_1^T] & E[z_1z_2^T] & E[z_1h_3^T] \\ E[z_2z_1^T] & E[z_2z_2^T] & E[z_2h_3^T] \\ E[h_3z_1^T] & E[h_3z_2^T] & E[h_3h_3^T] \end{pmatrix}$$

$$= \begin{pmatrix} R_{11} & R_{12} & R_{13}A^T \\ R_{12}^T & R_{22} & R_{23}A^T \\ AR_{13}^T & AR_{23}^T & AR_{33}A^T + \Sigma \end{pmatrix} \tag{22}$$

where the R_{ij} for $i, j = 1, 2, 3$ denote the correlations between field values at points in Ω_i and Ω_j. Therefore as long as the number $n_2 + m_3$ of direct and indirect field data is very much smaller than the size n_1 of the regular sampling grid Ω_1, the matrix R in (22) is only a low-rank perturbation of R_{11}. Thus the approach presented in section 3 carries through essentially unchanged. Indeed, the only changes to (5)–(7) are that R_{12} is replaced by the block $(R_{12}|R_{13}A^T)$ and R_{22} is replaced by the block

$$\begin{pmatrix} R_{22} & R_{23}A^T \\ AR_{23}^T & AR_{33}A^T + \Sigma \end{pmatrix}$$

Next, the periodic random field $\bar{Z}(x, y)$ with covariance $\bar{r}(x, y)$ defined in (8) is again invoked, and the sampling of $\bar{Z}(x, y)$ on $\bar{\Omega}_1$ as defined in (9) is denoted by $\bar{z}_1$. Then we consider unconditional simulations of the vector

$$\bar{z} = \begin{pmatrix} \bar{z}_1 \\ z_2 \\ h_3 \end{pmatrix}$$

which has covariance

$$\bar{R} = \begin{pmatrix} \bar{R}_{11} & \bar{R}_{12} & \bar{R}_{13}A^T \\ \bar{R}_{12}^T & R_{22} & R_{23}A^T \\ A\bar{R}_{13}^T & AR_{23}^T & AR_{33}A^T + \Sigma \end{pmatrix} \tag{23}$$

with $\bar{R}_{13}$ being the covariance matrix between pairs of points of $\bar{\Omega}_1$ and Ω_3 ($\bar{R}_{11}$ and $\bar{R}_{12}$ remain as defined in section 3).

As in section 3 the $\bar{R}$ of (23) is an embedding of R from (22) such that $\bar{R}_{11}$ is circulant. Therefore, provided $\bar{R}$ is positive definite, a square root $\bar{R}^{1/2}$ can be efficiently computed as before. It remains to ensure that $\bar{R}$ is positive definite. For this, we rewrite $\bar{R}$ as

$$\bar{R} = \tilde{A}\tilde{R}\tilde{A}^T + \tilde{\Sigma} \tag{24}$$

where

$$\tilde{A} = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & A \end{pmatrix} \qquad \tilde{R} = \begin{pmatrix} \bar{R}_{11} & \bar{R}_{12} & \bar{R}_{13} \\ \bar{R}_{12}^T & R_{22} & R_{23} \\ \bar{R}_{13}^T & R_{23}^T & R_{33} \end{pmatrix}$$

$$\tilde{\Sigma} = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & \Sigma \end{pmatrix} \tag{25}$$

with 0 and I denoting, respectively, the zero and identity matrices of appropriate size.

Since A has full rank and Σ is a covariance matrix, (24) shows that $\bar{R}$ will be positive definite whenever $\tilde{R}$ is positive definite. But as $\bar{R}$ is the covariance matrix of the field $\bar{Z}$ on the grid $\bar{\Omega}_1 \cup \Omega_2 \cup \Omega_3$, the arguments of section 3 show that $\bar{R}$ will be positive definite as long as $\bar{r}$ is positive definite. This will be the case if the correlation between points on opposite sides of the grid is sufficiently close to zero; otherwise an increase in the sampling size or inclusion of a small nugget effect will generally suffice. If $\bar{R}^{1/2}$ is available and given a vector $\epsilon \sim \mathcal{N}(0, I)$ of complex white noise, the vector $\bar{z} = \bar{R}^{1/2}\epsilon$ has real and imag-

inary entries with the required distribution $\mathcal{N}(0, \bar{R})$. Given $\bar{z}$, extraction of a vector z with real and imaginary entries being both $\mathcal{N}(0, R)$ is trivial. The costs of computing $\bar{z}$ are therefore a setup cost of $O[n_1(n_2 + m_3) \max \{\log n_1; n_2 + m_3\}]$ flops, followed by a cost of $O[n_1 \max \{\log n_1; n_2 + m_3\}]$ flops for computing each realization. Storage requirements are $O[n_1(n_2 + m_3)]$.

6. Conclusions

This paper has presented a new algorithm for generating simulations of a stationary Gaussian random field Z in one or more dimensions over a regular lattice of n_1 nodes conditioned on n_2 scattered direct measurements and m_3 indirect measurements that are linear functions of n_3 unknown values of Z. Fast computation requires that $m_3 \leq n_3$ and $n_2 + n_3 \ll n_1$, but these conditions are almost always met in practice.

Particular features of the approach are (1) it is exact (modulo computer round-off errors); (2) it requires only $O[n_1(n_2 + n_3)]$ storage; and (3) the total cost of calculating a succession of realizations consists of a setup cost of $O[n_1(n_2 + n_3) \max \{\log n_1; n_2 + n_3\}]$ flops, followed by a computational cost of $O[n_1 \max \{\log n_1; n_2 + n_3\}]$ flops for each realization. With $n_2 + n_3 \ll n_1$, the circulant embedding approach thus allows large conditional simulations to be speedily generated on small workstations. This was illustrated here by numerical results showing the use of the approach in two dimensions to generate simulations of anisotropic exponential and Whittle covariance functions conditioned on direct measurements.

The only significant limitations of the approach are the assumptions that the field is stationary and that the matrix $\bar{R}$ defined in (25) is positive definite. In practice, stationarity seems to be essential to get sufficient structure to make large-scale simulations computationally possible regardless of the particular method chosen. Moreover we have shown that if $\bar{R}$ is initially not positive definite, then it can be made so either by increasing the size of the sampling domain Ω_1 until the correlation between points on opposite sides of the grid is essentially zero or by introducing a small nugget effect in the simulation.

We conclude by briefly comparing the approach proposed here with some other currently used methods for conditional simulations. The spectral method [*Shinozuka and Jan*, 1972] has a very similar computational cost to the present method when used to generate unconditional simulations on rectangular grids, but the distribution of these realizations only approximates the desired distribution. It can be extended to generate simulations conditioned on measurements on certain nodes within a rectangular grid via (5) and (7), but at present it does not appear to have been extended to generate simulations conditioned on measurements at arbitrary points. Indeed any such extension is likely to be very similar to the algorithm presented here; this similarity, coupled with the fact that the present method is exact while the spectral method is only approximate, suggests that the present method has the edge.

The very popular turning bands method [e.g., *Journel and Huijbregts*, 1978] also has very low computational costs, but also has its own particular drawbacks. First, it is again an approximate method in that integrals are replaced by finite sums, and so again does not produce simulations with exactly the required covariance structure. Thus detailed error analyses or extensive experiments are needed to assess the accuracy of simulations produced by turning bands. In contrast, simulations produced by the method proposed here are Gaussian random variables with exactly the specified covariance structure. Thus the standard theory for multivariate Gaussian random variables can be used to predict the performance of the method in numerical simulations.

A second drawback is that turning bands can normally only be applied to generate simulations of nonisotropic fields. Finally, turning bands cannot be used to generate simulations conditioned on indirect measurements. Indirect data introduce a nonstationary component into the covariance matrix of the unconditional simulations, and, to the best of the authors' knowledge, the turning bands method cannot generate nonstationary fields.

The Cholesky factorization approach [*Davis*, 1987a] does give simulations with exactly the required covariance structure, but cannot be used to generate large simulations in two or more dimensions for two reasons. First, and most importantly, the computational cost scales as at least $O[(n_1 + n_2 + m_3)^2]$, and more usually as $O[(n_1 + n_2 + m_3)^3]$. This makes the method prohibitively expensive: For example, simulations over a two-dimensional rectangular grid of a few hundred nodes in both horizontal and vertical directions will only be possible with access to supercomputers having very large in-core memory storage. Yet for the same setup the circulant embedding approach can be implemented on a standard workstation. Second, in large simulations the matrix R is likely to be severely ill conditioned (see *Dietrich and Newsam* [1989] for details). In such cases, computer round-off errors may cause the Cholesky factorization of R to fail as the factorization tries to calculate the square root of negative values.

Finally, the matrix polynomial method for approximating the square root of R introduced by *Davis* [1987b] can also take advantage of the near Toeplitz structure of R to efficiently compute each step in the approximating series. Unfortunately, in practice, ill-conditioning of R will make the approximating series only slowly convergent [*Dietrich and Newsam*, 1993], so that a large amount of computation would be needed to achieve the accuracy of the embedding approach described here.

Acknowledgments. The authors would like to thank the referees for a number of helpful comments, and in particular for suggesting that the method be extended to include conditioning on indirect measurements.

References

Ahmed, S., and G. de Marsily, Comparison of geostatistical methods of estimating transmissivity using data on transmissivity and specific capacity, *Water Resour. Res.*, *23*(9), 1717–1737, 1987.

Anderson, T. W., *An Introduction to Multivariate Statistical Analysis*, 2nd ed., John Wiley, New York, 1984.

Champeney, D. C., *A Handbook of Fourier Theorems*, Cambridge University Press, New York, 1987.

Cressie, N., and D. L. Zimmerman, On the stability of the geostatistical method, *Math. Geol.*, *24*(1), 45–58, 1992.

Davis, M. W., Production of conditional simulations via the LU triangular decomposition of the covariance matrix, *Math. Geol.*, *19*(2), 91–98, 1987a.

Davis, M. W., Generating large stochastic simulations—The matrix polynomial approximation method, *Math. Geol.*, *19*(2), 99–107, 1987b.

Dietrich, C. R., and G. N. Newsam, A stability analysis of the geostatistical approach to aquifer identification, *Stochastic Hydrol. Hydraul.*, *4*(3), 293–316, 1989.

Dietrich, C. R., and G. N. Newsam, Fast and exact generation of stationary Gaussian processes via circulant embeddings of the correlation matrix, *SIAM J. Sci. Comput.*, in press, 1996.

Dietrich, C. R., and G. N. Newsam, A fast and exact method for multidimensional Gaussian stochastic simulations, *Water Resour. Res.*, *29*(8), 2861–2869, 1993.

Golub, G. H., and C. F. van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, Md., 1989.

Hoeksema, R. J., and P. K. Kitanidis, An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling, *Water Resour. Res.*, *20*(7), 1003–1020, 1984.

International Mathematics and Statistics Libraries, International mathematics and statistics libraries, version 2.0, Houston, Tex., 1991.

Journel, A. G., and C. T. Huijbregts, *Mining Geostatistics*, Academic, San Diego, Calif., 1978.

Schweppe, F. C., *Uncertain Dynamic Systems*, Prentice-Hall, Englewood Cliffs, N. J., 1973.

Shinozuka, M., and C. M. Jan, Digital simulation of random processes and its applications, *J. Sound Vibrations*, *25*(1), 111–128, 1972.

Wood, A. T. A., When is a truncated covariance function on the line a covariance function on the circle?, *Rep. CMA-SR09-93*, Cent. for Math. Anal., Aust. Natl. Univ., Canberra, 1993.

Woods, J. W., Two-dimensional discrete Markovian fields, *IEEE Trans. Inform. Theory*, *18*(2), 101–109, 1972.

C. R. Dietrich, Faculty of Environmental Sciences, School of Environmental Engineering, Griffith University, Brisbane, Queensland 4111, Australia. (e-mail: c.dietrich@cns.gu.edu.au)

G. N. Newsam, Information Technology Division, Defence Science and Technology Organisation, P. O. Box 1500, Salisbury, South Australia 5108, Australia. (e-mail: garry.newsam@dsto.defence.gov.au)

(Received January 31, 1994; revised November 3, 1994; accepted November 11, 1994.)