Mathematical Geology, Vol. 28, No. 6, 1996

SHORT NOTE

On Conditional Simulation to Inaccurate Data[1]

Dean S. Oliver[2]

INTRODUCTION

Of the many methods available for conditional simulation to exact data, two are particularly interesting because of the contrast in approaches. One approach is to generate a "rough" field (an unconditional simulation) with the same covariance as the true field, then to subtract a smooth correction that forces the simulated field to pass through the data (Journel and Huijbregts, 1978). A second approach is to calculate a "smooth" estimate that passes through the data, then use the LU decomposition of the estimation error covariance to add a stochastic component to the estimate. It is known that these methods are equivalent (Krzanowski, 1988; Dietrich and Newsam, 1995).

The first method is used with the method of turning bands and with spectral methods because these methods are efficient at generating unconditional simulations. The second method is less commonly used because it requires a square-root decomposition of the estimation error covariance matrix (or operator). If the problem of interest is large, then taking this square root can be difficult and, because the problem is not stationary after measurements have been introduced, it seems unlikely that this method could be used for large problems, although Dietrich and Newsam (1995) have pointed out that taking the square root of conditional covariance matrices may be done efficiently using Chebychev polynomial approximations.

The introduction of errors into the measurements requires a modification of even the first method. Clearly, it is not desirable to force the conditional simulation to honor inaccurate measurements exactly, but the ideas behind the two approaches generally are valid and can be shown to give equivalent results

[1]Received 4 December 1995; revised 7 March 1996.

[2]Chevron Petroleum Technology Company, P.O. Box 446, La Habra, California, 90633-0446; e-mail: dsol@chevron.com

0882-8121/96/0800-0811$09.50/1

if small modifications are made. I use this equivalence to develop an efficient method for calculating a square root decomposition of the estimation error covariance for use in simulation.

Although the problem of errors in observations has received little attention in the geostatistical literature, Marcotte (1995) recently proposed several approaches based on filtering of data measurement error.

PRELIMINARIES

Consider the spatial random field, $Z(x)$, which is the sum of a known trend, $m(x)$, and a fluctuation, $\epsilon(x)$, that is

$$Z(x) = m(x) + \epsilon(x) \tag{1}$$

For simplicity, assume that the covariance, C_E, between $\epsilon(x)$ and $\epsilon(x')$ for two points in the domain of Z, is known and defined by

$$C_E(x, x') = E[\epsilon(x)\epsilon(x')] \tag{2}$$

where the expected value of the fluctuation, $E[\epsilon(x)]$, is zero everywhere.

Suppose, also, that N observations of phenomena that are linearly related to Z have been made and that these observations have errors that jointly are distributed normally with zero mean and $(N \times N)$ covariance matrix, C_D, that is, the N-dimensional data vector, d^{obs}, is given by

$$d^{\text{obs}} = FZ + e, \qquad \text{where} \quad e \sim N_N(0, C_D) \tag{3}$$

If, for example, the ith observation is a measurement of the value of Z at a location x_i, then

$$F_i Z = \int \delta(x - x_i) Z(x)\, dx = Z(x_i)$$

where the Dirac delta function, $\delta(x)$, is the seen to be the kernel of the data operator for point sampling of the field. In this simple example, $d_i^{\text{obs}} = Z(x_i) + e_i$, but, clearly, other types of measurements, such as travel time in crosswell seismic tomography, would have different, more complex, kernels.

Note that F is a linear operator mapping the model space containing Z (infinite dimensional) into the data space containing d^{obs} (N-dimensional). The transpose of F maps elements of the dual of the data space (N-dimensional) into elements of the dual of the model space (infinite dimensional). It also is possible to think of Z as a vector field defined on M grid points, in which instant, Z, m, and ϵ are M-dimensional vectors, C_E is an $(M \times M)$ matrix, F is an $(N \times M)$ matrix, and F^T is a standard matrix transpose of F. See Järvinen (1976) for a lucid description of the connection between finite and infinite dimensional ma-

trices, or Chapter 4 of Lanczos (1961) for a description of the connection between differential operators acting on functions, and matrices acting on vectors.

In much of the following, the explicit dependence of the random variable on the spatial coordinate will be suppressed to simplify the presentation.

SMOOTH PLUS ROUGH

Using the notation of Equations (1) and (3), the best estimate of $Z(x)$ based on the measurements can be written as follows (see Rodgers, 1976; Hachich and Vanmarcke, 1983):

$$\hat{Z} = m + C_E F^T (F C_E F^T + C_D)^{-1} (d^{\text{obs}} - Fm) \tag{4}$$

or, more explicitly (see Tarantola, 1987; McLaughlin and Townley, 1996),

$$\hat{Z}(x) = m(x) + \left| \int_D C_E(x, \xi) F^T(\xi)\, d\xi \right| \cdot \left| \int_D \int_D F(\eta) C_E(\eta, \zeta) F^T(\zeta)\, d\eta d\zeta + C_D \right|^{-1} \cdot \left| d^{\text{obs}} - \int_D F(\xi) m(\xi)\, d\xi \right| \tag{5}$$

Similarly, the estimation error covariance is

$$C_{E'} = C_E - C_E F^T (F C_E F^T + C_D)^{-1} F C_E \tag{6}$$

A conditional simulation can then be generated using the square root of the estimation covariance as follows:

$$Z^{CS} = \hat{Z} + C_{E'}^{1/2} S \tag{7}$$

When Z^{CS} is a continuous field, S in Equation (7) is a random field with a Dirac covariance measure. If, however, Z^{CS} is a defined only at M discrete locations, then $S \sim N_P(0, I_P)$ and $C_{E'}^{1/2}$ is a $(M \times P)$ matrix of rank P (see Rao, 1965, p. 440). A square root, $C_{E'}^{1/2}$, of the estimation covariance must satisfy the following relationship:

$$C_E = C_E^{1/2} [C_E^{1/2}]^T \tag{8}$$

which, for continuous fields, can be written as

$$C_E(x, x') = \int_D C_E^{1/2}(x, \xi) C_E^{1/2}(\xi, x')\, d\xi \tag{9}$$

ROUGH PLUS SMOOTH

Another way to approach the problem is to subtract a relatively smooth correction from the unconditional simulation, the advantage over the previous method being the relative simplicity of generating unconditional simulations. Journel and Huijbregts (1978) discussed this method in detail for the situation in which the observations of a regionalized variable are exact, that is without error. Because the observations were assumed to be exact, the observed values and the simulated values had to be the same at the observation locations.

When the observations are not exact, we can only require that the conditional realizations have the same expectation as Z and that the covariance of the conditional realizations be equal to the estimation covariance of Z. Conditional simulations with these properties can be generated by replacing $[Z - \hat{Z}]$ in the following identity,

$$Z = \hat{Z} + [Z - \hat{Z}] \tag{10}$$

with another random field isomorphic to $[Z - \hat{Z}]$. The unconditional simulation,

$$Z^{\text{ucs}} = m + C_E^{1/2} S \tag{11}$$

clearly has the same mean and covariance as Z. Measurements or observations, d^{ucs}, of the unconditional realization are generated by adding random errors, E, to the exact observations, FZ^{ucs}, that is,

$$d^{\text{ucs}} = FZ^{\text{ucs}} + E \tag{12}$$

where the errors are drawn from the same distribution as the true measurement errors,

$$E = C_D^{1/2} S_D \tag{13}$$

and S_D is a vector of N independent random normal deviates. The best estimate of Z^{ucs} based on inaccurate observations of the unconditional simulation is

$$\hat{Z}^{\text{ucs}} = m + C_E F^T (F C_E F^T + C_D)^{-1} (d^{\text{ucs}} - Fm) \tag{14}$$

Note that Equation (14) is identical to Equation (5) except that simulated data, d^{ucs}, have been used instead of the actual observations. A formula for the conditional simulation of Z is obtained by substitution of Equations (11) and (14) into Equation (10).

$$\begin{aligned} Z^{cs} &= \hat{Z} + [Z^{\text{ucs}} - \hat{Z}^{\text{ucs}}] \\ &= \hat{Z} + [C_E^{1/2} S - C_E F^T (F C_E F^T + C_D)^{-1} ((F Z^{\text{ucs}} + E) - Fm)] \\ &= \hat{Z} + ([C_E^{1/2} \quad 0] - C_E F^T (F C_E F^T + C_D)^{-1} [F C_E^{1/2} \quad C_D^{1/2}]) \begin{bmatrix} S \\ S_D \end{bmatrix} \end{aligned} \tag{15}$$

Z^{cs} is again seen to be the sum of a smooth, deterministic term and a rough, stochastic term.

EQUIVALENCE

Equations (7) and (15) represent two different ways of generating simulations of random fields that are conditional to inaccurate data. Clearly, the expectations of Z^{cs} from Equations (7) and (15) are identical. To show that the two methods are equivalent, we need to also establish that the covariance of realizations generated by the second method is equal to the estimation covariance used in the first method. From Equation (15), the covariance of the conditional simulation of Z is

$$\begin{aligned} E\{(Z^{cs} - \hat{Z})(Z^{cs} - \hat{Z})^T\} = &([C_E^{1/2} \quad 0] - C_E F^T (FC_E F^T + C_D)^{-1} \\ &\cdot [FC_E^{1/2} \quad C_D^{1/2}]) \\ &\cdot E\left\{ \begin{bmatrix} S \\ S_D \end{bmatrix} [S^T \quad S_D^T] \right\} ([C_E^{1/2} \quad 0] \\ &- C_E F^T (FC_E F^T + C_D)^{-1} [FC_E^{1/2} \quad C_D^{1/2}])^T \end{aligned}$$

Because S and S_D are independent,

$$E\left\{ \begin{bmatrix} S \\ S_D \end{bmatrix} [S^T \quad S_D^T] \right\} = I$$

and the covariance of the conditional simulation is

$$\begin{aligned} C_E^{cs} &= ([C_E^{1/2} \quad 0] - C_E F^T (FC_E F^T + C_D)^{-1} [FC_E^{1/2} \quad C_D^{1/2}]) \\ &\quad \cdot ([C_E^{1/2} \quad 0] - C_E F^T (FC_E F^T + C_D)^{-1} [FC_E^{1/2} \quad C_D^{1/2}])^T \\ &= ([C_E^{1/2} \quad 0] - C_E F^T (FC_E F^T + C_D)^{-1} [FC_E^{1/2} \quad C_D^{1/2}]) \\ &\quad \cdot \left(\begin{bmatrix} C_E^{T/2} \\ 0 \end{bmatrix} - \begin{bmatrix} C_E^{T/2} F^T \\ C_D^{T/2} \end{bmatrix} (FC_E F^T + C_D)^{-1} FC_E \right) \\ &= C_E^{1/2} C_E^{T/2} - C_E F^T (FC_E F^T + C_D)^{-1} FC_E^{1/2} C_E^{T/2} \\ &\quad - C_E^{1/2} C_E^{T/2} F^T (FC_E F^T + C_D)^{-1} FC_E \\ &\quad + C_E F^T (FC_E F^T + C_D)^{-1} (FC_E^{1/2} C_E^{T/2} F^T + C_D^{1/2} C_D^{T/2}) \\ &\quad \cdot (FC_E F^T + C_D)^{-1} FC_E \\ &= C_E - C_E F^T (FC_E F^T + C_D)^{-1} FC_E \end{aligned}$$

which is simply the a posteriori covariance that was calculated in the first method. Thus, the two methods generate equivalent random fields.

AN EXAMPLE

Consider the simple situation in which the data are point measurements of the regionalized variable, and the prior covariance of the value of ϵ at two points is a function only of the separation. A simple expression for a "square root" of the *a posteriori* covariance can be obtained from Equation (15).

$$\begin{aligned} L(x, x') &= [C_E^{1/2} \quad 0] - C_E F^T (F C_E F^T + C_D)^{-1} [F C_E^{1/2} \quad C_D^{1/2}] \\ &= \left| C_E^{1/2} - \sum_{i=1}^{N} C_E(x - x_i) \sum_{j=1}^{N} (K^{-1})_{ij} C_E^{1/2}(x' - x_j) \right. \\ &\quad \left. - \sum_{i=1}^{N} C_E(x - x_i) \sum_{j=1}^{N} (K^{-1})_{ij} [C_D^{1/2}]_{jk} \right| \end{aligned} \tag{17}$$

where $K_{ij} = C_E(x_i - x_j) + C_{Dij}$. For some covariance models, $C_E^{1/2}$ is easy to calculate. For example, Oliver (1995) showed that if $C_E(r) = \sigma^2 \exp(-r/a)$ in 3-D, then $C_E^{1/2}(r) = \sigma/\sqrt{2\pi a r^2} \exp(-r/a)$.

A conditional realization can be generated fairly easily using Equation (15), which can be rewritten as follows:

$$Z^{cs} = (\hat{Z} - C_E F^T K^{-1} C_D^{1/2} S_D) + (C_E^{1/2} - C_E F^T K^{-1} F C_E^{1/2}) S \tag{18}$$

The first term on the right side of Equation (18) is the "smooth" part of the conditional field. The stochastic component of the first term can be attributed to the addition of random errors to the "data" calculated from the unconditional simulation. The second term is a nonstationary, weighted moving average process. It can be calculated in the same way as any weighted moving average, with the additional complication that the weights are functions of location.

DISCUSSION

If S is a vector of independent random normal deviates distributed as $N(0, I_N)$ and $Z = LS$, then Z is distributed as $N(0, LL^T)$. For any desired covariance matrix, $C = LL^T$, a square matrix L can be determined numerically using, for example, the Cholesky algorithm. Almost no one, however, uses the Cholesky method for conditional simulation because the cost of factoring a nonstationary covariance matrix is prohibitive for large problems.

Equations (7) and (15) are both "square root" approaches to conditional simulation. The advantage of Equation (15) is that it uses the square root of the stationary *a priori* covariance to generate the square root of the *a posteriori*

covariance. Because it is possible to factor a stationary covariance operator analytically for most covariance models, it is not necessary to perform Cholesky decomposition to use Equation (15) for conditional simulation.

The square root, L, given by Equation (17) is not square; it, however, does satisfy $C = LL^T$. The disadvantage of this is that the L given by Equation (17) is slightly larger than necessary, indicating that we must generate and use a few more random deviates than if we used Cholesky decomposition to calculate L for simulation. Of course, if C_E is conditioned poorly, as it will be, for example, if a Gaussian covariance model is used, the optimal decomposition could be calculated from a singular value decomposition algorithm, in which instance L would be smaller than the Cholesky factors.

The method described in this paper has two advantages in certain situations. The first is that the calculation of the square root of the conditional covariance is nearly trivial, and depends only on the number of data, not on the size of the domain or the fineness of the discretization. The second advantage is that only the coefficients of the inverse of the kriging matrix need be stored, unlike methods that calculate the square root directly.

REFERENCES

Dietrich, C. R., and Newsam, G. N., 1995, Efficient generation of conditional simulations by Chebyshev matrix polynomial approximations to the symmetric square root of the covariance matrix: Math. Geology, v. 27, no. 2, p. 207–228.

Hachich, W., and Vanmarcke, E. H., 1983, Probabilistic updating of pore pressure fields: Jour. Geotech. Eng., v. 109, no. 3, p. 373–387.

Järvinen, R. D., 1976, Finite and infinite dimensional linear spaces: Marcel Dekker, New York, 168 p.

Journel, A., and Huijbregts, C. J., 1978, Mining geostatistics: Academic Press, New York, 600 p.

Krzanowski, W. J., 1988, Principles of multivariate analysis: a user's perspective: Clarendon Press, Oxford, 563 p.

Lanczos, C., 1961, Linear differential equations: Van Nostrand, London, 564 p.

Marcotte, D., 1995, Conditional simulation with data subject to measurement error: Postsimulation filtering with modified factorial kriging: Math. Geology, v. 27, no. 6, p. 749–762.

McLaughlin, D., and Townley, L. R., 1996, A reassessment of the groundwater inverse problem: Water Resources Res., v. 32, no. 5, p. 1131–1161.

Oliver, D. S., 1995, Moving averages for Gaussian simulation in two and three dimensions: Math. Geology, v. 27, no. 8, p. 939–960.

Rao, C. R., 1965, Linear statistical inference and its applications: John Wiley & Sons, New York, 522 p.

Rodgers, C. D., 1976, Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation: Jour. Geophys. Res., v. 14, no. 4, p. 609–624.

Tarantola, A., 1987, Inverse problem theory: methods for data fitting and model parameter estimation: Elsevier, Amsterdam, The Netherlands, 613 p.