Mathematical Geology, Vol. 29, No. 1, 1997

Markov Chain Monte Carlo Methods for Conditioning a Permeability Field to Pressure Data[1]

Dean S. Oliver,[2] Luciane B. Cunha,[3,4] and Albert C. Reynolds[5]

Generating one realization of a random permeability field that is consistent with observed pressure data and a known variogram model is not a difficult problem. If, however, one wants to investigate the uncertainty of reservior behavior, one must generate a large number of realizations and ensure that the distribution of realizations properly reflects the uncertainty in reservoir properties. The most widely used method for conditioning permeability fields to production data has been the method of simulated annealing, in which practitioners attempt to minimize the difference between the "true" and simulated production data, and "true" and simulated variograms. Unfortunately, the meaning of the resulting realization is not clear and the method can be extremely slow. In this paper, we present an alternative approach to generating realizations that are conditional to pressure data, focusing on the distribution of realizations and on the efficiency of the method. Under certain conditions that can be verified easily, the Markov chain Monte Carlo method is known to produce states whose frequencies of appearance correspond to a given probability distribution, so we use this method to generate the realizations. To make the method more efficient, we perturb the states in such a way that the variogram is satisfied automatically and the pressure data are approximately matched at every step. These perturbations make use of sensitivity coefficients calculated from the reservoir simulator.

KEY WORDS: conditional simulation, Markov chain, Monte Carlo, sampling, pressure data, sensitivity, well test.

INTRODUCTION

Because the process of predicting flow and transport in petroleum reservoirs is nonlinear, it is generally impossible to calculate directly the probability distribution for future reservoir performance. Instead, we are forced to estimate the

[1]Received 11 September 1995; revised 26 March 1996.

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

[3]The University of Tulsa, Department of Petroleum Engineering, Tulsa, Oklahoma 74104.

[4]Present address: Petróleo Brasileiro Research Center, Ilha de Fundao, Q7, Rio de Janiero, Brasil 21949--900; e-mail: lubo@cenpes, petrobras.gov.br

[5]The University of Tulsa, Department of Petroleum Engineering, Tulsa, Oklahoma 74104; e-mail: PE_ACR@centum.utulsa.edu

0882-8121/97/0100-0061$12.50/1

probability distribution from the outcomes of flow predictions for a large number of realizations of the reservoir. For this method to work, it is essential that the permeability and porosity realizations used in the flow simulation adequately reflect the uncertainty in reservoir properties. Thus, the realizations must be drawn from the probability distribution for the reservoir models. If we generate a large number of realizations of the reservoir permeability field, we should expect to generate a few models from regions of the model space that have low probability density. Methods that do not generate models from the "tails" of the distribution are not sampling correctly the distribution of possible reservoirs.

There are many stochastic methods that work well for multinormal distributions, or for independent stochastic variables, but stochastic simulation of general multivariate distributions can be difficult, especially when the number of variables is large. The type of problem that we envision is one in which the permeability in a large number of simulator gridcells must be determined for reservoir flow and transport prediction. We assume that there are at least some prior expectations regarding the distribution of permeability and that there are also some production data available. For our examples, it will be the conditioning to production data that makes the generation of realizations difficult.

The paper is organized as follows. We begin by presenting background material on the use of Markov chain Monte Carlo methods for sampling. This is followed by a critique of the current practice of using simulated annealing to minimize objective functions as well as a discussion of the conditional probability distribution for Gaussian random fields. We then apply these ideas to the simulation of Gaussian random fields with known mean and variogram. The method that we use has close connections to the method of moving averages so we discuss that relation, and the shape of the smoothing functions for one and two dimensions. A brief description of the application to conditional simulation with point data leads us to the main topic which is the problem of conditioning Gaussian random fields to nonlinear data such as transient pressure measurements.

General background on Markov chain Monte Carlo (MCMC) methods for conditional simulation can be obtained from Hammersley and Handscomb (1964), Hastings (1970), Geman and Geman (1984), and Neal (1993). Useful references on the application of Markov chain Monte Carlo methods to the earth sciences include publications by Tjelmeland, Omre, and Hegstad (1994) and Hegstad and others (1993).

MARKOV CHAINS FOR SAMPLING

In the context of reservoir characterization, we can think of a sequence of possible states or *realizations* of the reservoir permeability distribution as a Markov chain if the probability of generating some particular new realization

depends only on the preceding realization in the sequence. We will denote a particular stochastic realization or map of reservoir properties by the symbol, m^i, where the superscript i refers to the ith possible realization of a denumerable set of possible realizations, not to the ith realization in the Markov chain. Each one of these possible realizations, m^i, will have a probability, π_i, associated with it. In our context π_i will be the probability of the ith possible realization being the correct map of reservoir properties. Of course each of these probabilities is extremely small but some realizations will be far more likely than others. If we were to generate many realizations for prediction of reservoir performance, we would want the realizations to be sampled with the correct probability. Unfortunately, in most real applications we cannot characterize easily the probability of the ith realization so we must instead use a method that relies only on the relative probabilities to generate realizations. This can be done using Markov chains if we are careful in the specification of the conditional probability, p_{jk}, of transition to realization m^k from realization m^j.

If we can determine transition probabilities p_{ij} such that it is possible to get from any one state to another in a finite number of transitions and such that the probability of state m^j is the sum of the probabilities of being in a state m^i times the probability of transition from m^i to m^j, that is

$$\pi_j = \sum_i \pi_i p_{ij} \tag{1}$$

then the Markov chain will be stationary and ergodic (independent of initial conditions) and π_j will be the probability distribution for the realizations (Feller, 1968; Hammersley and Handscomb, 1964). In practice, it may take a large number of transitions to reach the stationary distribution.

Generation of a Markov chain with the desired distribution will require calculation of transition probabilities p_{ij} that satisfy Equation (1). We will discuss later the components of π_j and why they are so difficult to calculate. For now, we will state simply that although it may be impossible to calculate π_j, it may be relatively easy to calculate the ratio, π_j/π_i, of the probability of being in state m^j to the probability of being in state m^i.

Metropolis and others (1953) observed that the problem of calculating a permissible transition matrix could be simplified if the transition matrix satisfied a reversibility condition

$$\pi_i p_{ij} = \pi_j p_{ji} \tag{2}$$

They also proposed splitting the matrix into two components:

$$p_{ij} = \alpha_{ij} q_{ij} \tag{3}$$

where q_{ij} is the probability of proposing a transition from state m^i to state m^j and α_{ij} is the probability of accepting the proposed transition. The q_{ij} can be

selected somewhat arbitrarily, subject to the restriction that, because they are probabilities, they take values between zero and one. It should be clear, however, that some selections for q_{ij} will result in high probabilities of acceptance of transition and thus substantially reduce the computation time. The α_{ij} are not determined completely by Equations (2) and (3). Hastings (1970), therefore, proposed using

$$\alpha_{ij} = s_{ij}\left[1 + \frac{\pi_i q_{ij}}{\pi_j q_{ji}}\right]^{-1} \tag{4}$$

where s_{ij} is a symmetric matrix selected to ensure that $0 \leq \alpha_{ij} \leq 1$. From Equation (4) and the constraints on α_{ij} we obtain

$$0 < s_{ij} \leq 1 + \frac{\pi_j q_{ij}}{\pi_j q_{ji}} \tag{5}$$

Because $[s_{ij}]$ is symmetric, then from Equation (5) we also obtain

$$0 < s_{ij} \leq 1 + \frac{\pi_j q_{ji}}{\pi_i q_{ij}}$$

Based on these constraints, Hastings' (1970) choice for s_{ij} is

$$s_{ij} = \min\left\{1 + \frac{\pi_i q_{ij}}{\pi_j q_{ji}},\ 1 + \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right\}$$

which gives the following acceptance probability for state j with respect to state i:

$$\alpha_{ij} = \min\left\{1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right\} \tag{6}$$

Note that if the proposed transitions *are* symmetric, that is if $q_{ij} = q_{ji}$, then the calculation of whether to accept a transition is based only on the ratio of the probability of being in the two states. If the proposed transition is rejected, the old state is repeated in the chain. Swapping permeability values in two randomly selected grid blocks is an example of a symmetric transition that has been used in reservoir characterization (Farmer, 1992; Deutsch and Journel, 1992; Sagar, Kelkar, and Thompson, 1993). We will restrict ourselves to the α_{ij} given by Equation (6) because this selection is known to be relatively efficient at sampling the distribution compared to other possible selections (Ripley, 1987, p. 114).

The focus of our efforts will be on improving the efficiency of the overall method by increasing the acceptance probability for proposed transitions q_{ij}. Because π_i is a fixed, but unknown, function of the permeability distribution,

we can only improve the efficiency by improving the process for proposing transitions.

"OBJECTIVE FUNCTIONS"

Most published attempts to condition permeability fields to production data rely on either simulated annealing or the genetic algorithm to reduce the error misfit between production data calculated from a theoretical model of the reservoir and the actual production data (Sen and others, 1992; Sagar, Kelkar, and Thompson, 1993; Deutsch, 1992; Ouenes and others, 1993). Studies that are based on simulation of Gaussian permeability fields typically also attempt to

gram (Deutsch and Journel, 1992). The most widely used measure of misfit is the sum of the squared differences. To fit this into the optimization framework on "objective function" is formed from the weighted sum of the misfit terms

priori knowledge of what is reasonable, on closely spaced measurements from an outcrop, on limited samples obtained throughout the field, or on some combination of all of these. The important point is that the variogram model used for kriging and simulation is an estimate of the theoretical variogram for the ensemble of possible reservoirs or, equivalently, for a reservoir of infinite extent.

The local variogram, $\gamma_v(h)$, is the variogram that one would calculate from complete knowledge of a single realization of the random field within a domain v of limited extent. The expected value of the local variogram (for lags that can be calculated) is the theoretical variogram, but the local variogram can fluctuate substantially around the expected value when the dimensions of the domain v are only a few times larger than the lags of interest for the comparison of local variogram. Unfortunately, the expected fluctuations in the calculated covariance are somewhat difficult to compute because they involve the calculation of moments of order 4 of the random variable (Journel and Huijbregts, 1978, p. 193). It does seem that, for random variables whose spatial correlation is governed by a linear variogram, the relative fluctuation variance, that is, $E\{[\gamma(h) - \gamma_v(h)]^2\}/[\gamma(h)]^2$ approaches zero as h/L for lags h that are small compared to dimensions L of the local domain (Journel and Huijbregts, 1978, p. 193). We would expect this same behavior to hold for other variogram models, including exponential and spherical.

Instead of calculating the fluctuation variance for large lags, we can estimate the variability of the dispersion variance $\bar{C}(v, v)$ of a random function for the domain v. The dispersion variance will be a measure of the variability of the variance of point values within a limited domain. In other words, if we calculate the variance of sample values about the sample mean for a domain whose dimensions are large compared to the range of the variogram (assuming that it has one), we should expect to obtain the sill value for the variogram. If, however, our sample domain is small then all of the samples will be correlated highly and we expect to calculate a small sample variance. Journel and Huijbregts (1978, p. 61–68) show that the dispersion variance in a domain v is given by

$$D^2(0/v) = C(0) - \bar{C}(v, v) \tag{7}$$

where

$$\bar{C}(v, v) = \frac{1}{v^2} \int_v dr' \int_v dr\, C(r - r') \tag{8}$$

Using this result, we calculate that, for a one-dimensional exponential

the range of the variogram can we expect that the variance of point values within the domain will be close to the variance given by the sill of the variogram. Unless our domain is large compared to the range of the variogram we should not expect to generate realizations whose variance matches the sill of the variogram, and we should not attempt to achieve conditional simulations by matching the variogram exactly.

Thus, although it might be feasible to estimate the probability distribution for the variation of the experimental variogram about the theoretical variogram, the terms in the objective function for a simulated annealing approach would have to be weighted such that the realizations of the variogram are drawn from this distribution at the final temperature. This is not a new idea. In a discussion of ergodicity, Deutsch and Journel (1992, p. 126) state that "exact reproduction of the model statistics by each simulated realization may not be desirable, or possible." As usually practiced, however, the "objective of (simulation by simulated annealing) is for the semivariogram $\gamma^*(h)$ of the simulated realization to match the prespecified semivariogram model $\gamma^*(h)$" (Deutsch and Journel, 1992, p. 154).

THE CONDITIONAL PROBABILITY DISTRIBUTION

Although it is difficult to estimate the probability distribution for the local variogram when the theoretical variogram is known, it is straightforward to calculate the probability distribution for the random variables under the same assumption. Specifically, if the theoretical covariance, C_M, and the mean, m_0 are both known *a priori* for a Gaussian random field then the probability distribution for the random variables is multivariate normal:

$$\Pi \propto \exp\{-\tfrac{1}{2}[m - m_0]^T C_M^{-1}[m - m_0]\} \tag{9}$$

The mean of the random variable in Equation (9) can be a function of position, as can the covariance. Because this is a simple distribution, it is possible to use any number of standard geostatistical methods of simulation including sequential Gaussian simulation and the *LU* decomposition method (Scheuer and Stoller, 1962). We will discuss the application of Markov chain Monte Carlo methods to the sampling of this distribution in the following section.

Recall that the goal of simulation is not to determine the maximum of the probability distribution but to sample it. Although simulated annealing is an MCMC method, when it is used without careful consideration of the stopping "temperature," the results will not be random realizations from the correct probability density function. If, for example, simulated annealing were used to determine the maximum of Equation (9) or, equivalently, the minimum of $[m - m_0]^T C_M^{-1}[m - m_0]$, one would obtain $m = m_0$ independently of the vario-

gram. So, if we are not careful, we obtain the kriging solution instead of a stochastic simulation. Incidentally, simulated annealing has been used occasionally to sample the model space in such a way that the distribution of samples approximates the *a posteriori* distribution in the model space (Koren and others, 1991; Hegstad and others, 1993).

Conditioning the permeability field to production data requires that we be able to characterize the likelihood of obtaining the measured data from a simulator when the true values of the random permeability field are known. Two sources of error enter the problem: measurement or observation errors result in the inaccuracy of the measuring instrument or sloppy technique, and modeling errors, which are the result of oversimplification or inaccuracy of the flow simlator. It generally is difficult to characterize the distribution of errors accurately, but if we assume that the errors of both types are distributed normally, then we can combine them into one data error term and write the conditional probability distribution for the permeability field as

$$\Pi \propto \exp\{-\tfrac{1}{2}[d_{\text{obs}} - g(m)]^T C_D^{-1}[d_{\text{obs}} - g(m)] - \tfrac{1}{2}[m - m_0]^T C_M^{-1}[m - m_0]\} \tag{10}$$

where C_D is the sum of the data error covariance matrix and the modeling error covariance matrix (Tarantola and Valette, 1982). Throughout, $g(m)$ is the predicted value of the production data that one obtains from using the permeability field m in the flow simulator. Although we have made the assumption that the errors in the data are distributed normally, the probability distribution given by Equation (10) would be multivariate normal only if $g(\cdot)$ were a linear function of the random variables m. The function $g(\cdot)$ is not linear for transient pressure data and evaluation of Equation (10) requires a flow simulation for each selection of the random variables. Because, in the simplest situation, there will be thousands to tens of thousands of gridcells, each with a value of permeability assigned to it, and because a complete evaluation of the probability distribution would require a flow simulation for every possible combination of permeabilities, clearly it is impossible to sample exhaustively the parameter space.

CONDITIONING TO A VARIOGRAM AND A MEAN VALUE

Although there are many standard methods for generating realizations that are conditioned to a mean and a variogram (or covariance), we will describe a Markov chain Monte Carlo method of sampling because it will provide much of the background for the more complicated examples that follow. Consider a state j, which is an array of log permeability values, one for each of the M

$$m^i = (m_1, m_2, m_3, \ldots, m_M)^T$$

It is well known that the distribution of states is multivariate normal, that is, $m^i \sim N_M(m_0, C_M)$, if and only if

$$m^i = m_0 + LZ^i \tag{11}$$

where L is a "matrix square root" of C_M satisfying

$$C_M = LL^T \tag{12}$$

and the components of Z^i are independent normal deviates with mean value zero and variance one (Rao, 1965, p. 440). This is the basis of the what may be termed the LU or Cholesky decomposition method of conditional simulation (Ripley, 1981; Davis, 1987b; Alabert, 1987).

Because every array Z^i of independent normal deviates is related uniquely to an array m^i of log-permeability values by Equation (11), we can define the transitions along the Markov chain in terms of perturbations of Z instead of m. A particular value of Z^i, or equivalently, m^i, is referred to as the ith state of the model or Markov chain and is obtained by defining a component of the random vector Z for each node of the reservoir model,

$$Z^i = (Z_1, Z_2, Z_3, \ldots, Z_M)^T$$

To implement the Markov chain Monte Carlo method, we need to calculate the probability of drawing a new state and the probability of accepting a transition to the new state once it has been drawn.

INDEPENDENT REALIZATIONS OF THE PERMEABILITY FIELD

Consider the transition to a state j which is obtained from state i by drawing new values of Z_k at every grid node. If we denote the new values with primes to differentiate from the old values, then the proposed values of the normal deviates are

$$Z^j = (Z_1', Z_2', Z_3', \ldots, Z_M')^T \tag{13}$$

Because these are drawn independently from the normal distribution $N(0, 1)$, we can write the probability of drawing the state j when the current state is i as

$$\begin{aligned} q_{ij} &= c \exp\left(-\tfrac{1}{2} Z^j \cdot Z^j\right) \\ &= c \exp\left[-\tfrac{1}{2} (m^j - m_0)^T C_M^{-1} (m^j - m_0)\right] \end{aligned} \tag{14}$$

which is independent of the current state.

The probability of accepting the proposed transition is given by Equation

(6) which contains π_j, π_i, and q_{ji} in addition to the term we have just calculated. The probability, q_{ji}, of proposing the reverse transition to state i from the state j is

$$q_{ji} = c \exp\left(-\tfrac{1}{2} Z^i \cdot Z^i\right)$$

$$= c \exp\left[-\tfrac{1}{2} (m^i - m_0)^T C_M^{-1} (m^i - m_0)\right] \tag{15}$$

and the probabilities of being in state i or j are

$$\pi_i = c \exp\left[-\tfrac{1}{2} (m^i - m_0)^T C_M^{-1} (m^i - m_0)\right] \tag{16}$$

$$\pi_j = c \exp\left[-\tfrac{1}{2} (m^j - m_0)^T C_M^{-1} (m^j - m_0)\right] \tag{17}$$

From Equations (14) through (17), it is easy to show that

$$\frac{\pi_j q_{ji}}{\pi_i q_{ij}} = 1$$

so that *every* proposed transition q_{ij} of the type given by Equation (13) will be accepted with probability $\alpha_{ij} = 1$.

It may seem strange to term this a Markov chain Monte Carlo method because the Markov chain that is generated is simply a sequence of independent realizations generated using the *LU* decomposition method. The advantage is that, unlike many other methods, there is no wasted effort in proposing and rejecting large numbers of perturbations or transitions that do not meet the acceptance criteria. Also, the fact that each element of the chain is an independent realization makes the chain a more efficient sampler of the probability distribution than a chain of correlated realizations.

CORRELATED REALIZATIONS

Most proponents of simulated annealing for reservoir characterization propose transitions that are local perturbations to the permeability field. Two particularly widespread methods of perturbing the field are to "swap" the values at two locations, or to draw a new value at one location from the cumulative distribution function. Many of the transitions proposed using the swapping approach are rejected because they violate the variogram constraint. We will outline a transition that is local, that is, it only alters the permeabilities in a limited region, but that also satisfies the variogram.

Suppose, for example, that we randomly select (from a uniform distribution) the kth node, and draw a new value of Z_k so that the array Z^j corresponding to state j is given by

$$Z^j = (Z_1, Z_2, \ldots, Z_k', \ldots, Z_M)^T$$

Because the Z_k are independent and only one value is being altered, the probability of selecting state j when the current state is i is

$$q_{ij} = \frac{1}{M\sqrt{2\pi}} \exp\left(-\frac{1}{2} Z_k'^2\right)$$

The probability of proposing the reverse transition is

$$q_{ji} = \frac{1}{M\sqrt{2\pi}} \exp\left(-\frac{1}{2} Z_k^2\right)$$

The probability of being in state i is

$$\begin{aligned}\pi_i &= c \exp\left[-\tfrac{1}{2}(m^i - m_0)^T C_M^{-1}(m^i - m_0)\right] \\ &= c \exp\left[-\tfrac{1}{2} Z^i \cdot Z^i\right] \\ &= c \exp\left[-\tfrac{1}{2}(Z_1^2 + Z_2^2 + \cdots + Z_k^2 + \cdots + Z_M^2)\right]\end{aligned}$$

and the probability of being in state j is nearly the same:

$$\begin{aligned}\pi_j &= c \exp\left[-\tfrac{1}{2}(m^j - m_0)^T C_M^{-1}(m^j - m_0)\right] \\ &= c \exp\left[-\tfrac{1}{2} Z^j \cdot Z^j\right] \\ &= c \exp\left[-\tfrac{1}{2}(Z_1^2 + Z_2^2 + \cdots + Z_k'^2 + \cdots + Z_M^2)\right]\end{aligned}$$

Again, we determine that

$$\frac{\pi_j q_{ji}}{\pi_i q_{ij}} = 1$$

so that all proposed transitions are accepted.

The difference between this approach and the approach proposed in the previous section is in the size of the region that is modified during a transition from one state to the next. In the previous section, we proposed to draw new values of all the normal deviates at each transition, whereas in this section, we proposed drawing a new value of Z_k at only one location per transition. This may sound similar to the standard practice of drawing a new value of permeability from the cumulative distribution function but our modification of the permeability field comes through the relationship between Z^i and m^i. The change in m resulting from the change in Z is given in Equation (11),

$$\delta m^{ij} = L\delta Z^{ij} \tag{18}$$

where

$$\begin{aligned}\delta Z^{ij} &= Z^j - Z^i \\ &= (0, 0, \ldots, Z'_k - Z_k, \ldots, 0)^T \\ &= (0, 0, \ldots, \delta Z_k, \ldots, 0)^T\end{aligned}$$

so the perturbation to the permeability field that results from a change, δZ_k, in the value of the kth normal deviate is given by the product of δZ_k with the kth column of L.

Although our notation and our frequent reference to the LU decomposition might seem to imply that L is a lower triangular matrix, we only require that L satisfy Equation (12). Because the columns of L determine the shape of the perturbation of the permeability, we digress from the problem of generating the Markov chains to examine the columns of L for various decompositions.

SHAPE OF THE PERTURBATIONS

One-Dimensional Domain

If C_M is the $M \times M$ matrix of covariances of the M random variables on the grid nodes, then the decomposition of C_M ino the product of a lower triangular matrix, L, and its transpose L^T, provides a set of weights that can be applied to the array of normal deviates for the simulation of correlated random variables on the nodes. Similarly, we can construct a symmetric square-root matrix R such that $C_M = RR$ (see Appendix). The columns of R can be used like the columns of L in Equation (18) to generate a perturbation of the permeability field that is consistent with the variogram. Figures 1 and 2, respectively, show selected columns of L and R for an exponential covariance with a practical range of 9.

Two-Dimensional Perturbations

The shapes of the perturbations in two and three dimensions also are determined by the columns of the square root of the covariance matrix. To visualize the "shape," we arrange the elements of one column of L or R in the two-dimensional positions corresponding to the arrangement of permeability values. In Figure 3, we have plotted one particular column (corresponding to the perturbation of the 66th component of Z^i) for an LU decomposition of a stationary exponential covariance, and for the symmetric square-root decomposition. The "symmetric" decomposition R is symmetric only approximately because of some numerical inaccuracy in the calculation of the eigenvectors of the covariance, whereas L is localized but not symmetric. Oliver (1995) has described a

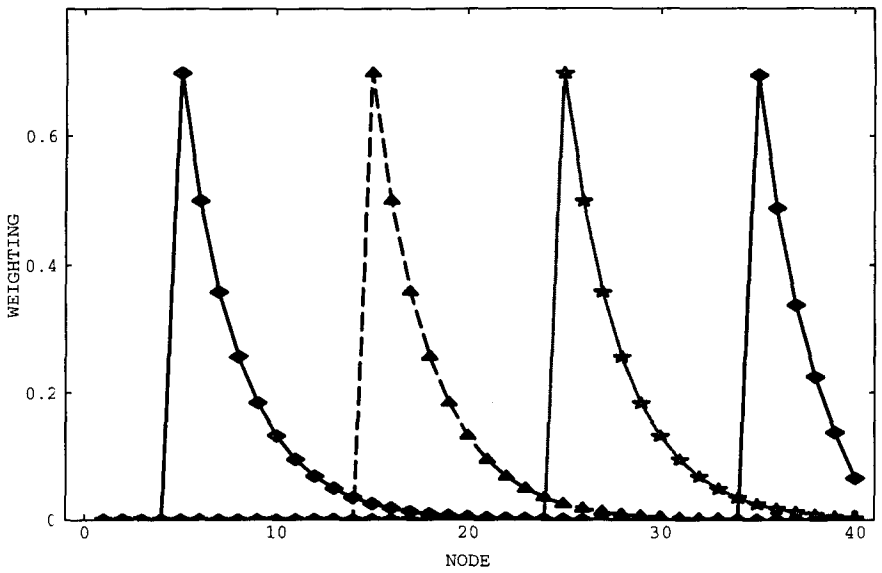


Figure 1. 5th, 15th, 25th, and 35th columns of lower triangular matrix L for exponential covariance function with practical range of 9.

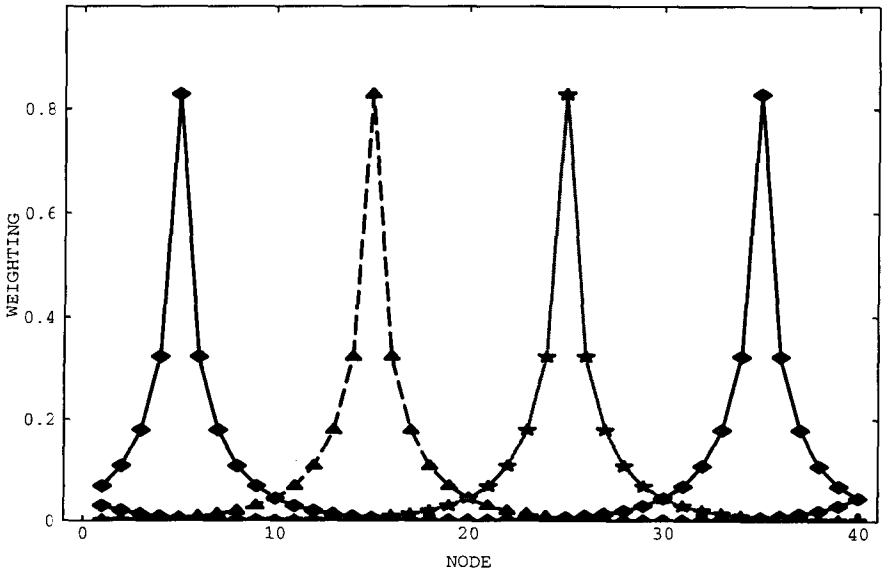


Figure 2. 5th, 15th, 25th, and 35th columns of symmetric square-root matrix R for exponential covariance function with practical range of 9.

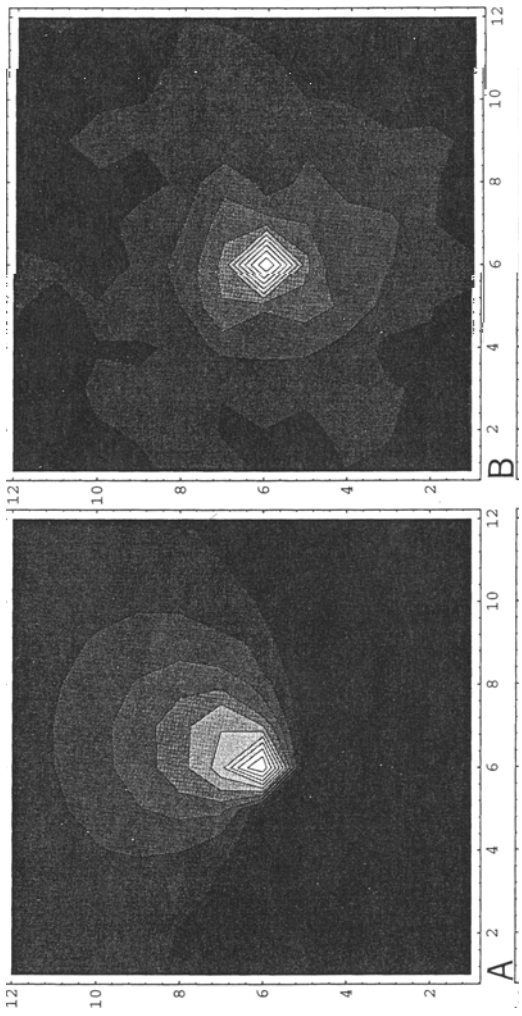


Figure 3. 66th columns of: A lower triangular matrix L; B, the symmetric square-root matrix R for a two-dimensional exponential covariance function with practical range of 9.

method for calculating the shapes of the symmetric perturbations, for several variogram models, which does not require calculation of the square root of the matrix.

CONDITIONING TO TRANSIENT PRESSURE DATA

Suppose that, in addition to *a priori* knowledge of the theoretrical variogram and mean, we have access to measured production data that can be used for conditioning the simulations. We know that the realizations must be sampled from Equation (10). We can use the same perturbations in our Markov chain procedure that were used for conditioning to the variogram, or we can improve convergence further by incorporating the "linear part" of the production data. We will start with the simpler method and then explain the improvements that can be made by using the sensitivity functions in the perturbations.

Local Perturbations Based on the Variogram

Consider the transition from one state to the next, consisting of a change in only one of the components of the vector of normal deviates. The ith possible state of the model is specified by the values of the array of normal deviates:

$$Z^i = (Z_1, Z_2, \ldots, Z_k, \ldots, Z_M)^T$$

and the jth state by

$$Z^j = (Z_1, Z_2, \ldots, Z'_k, \ldots, Z_M)^T$$

Because we draw the values of Z_k from the normal distribution, the probability of proposing a transition to state j from state i is

$$q_{ij} = \frac{1}{M\sqrt{2\pi}} \exp\left(-\frac{1}{2} Z_k'^2\right)$$

and the probability of proposing the reverse transition is obvious.

We calculate the probability of being in the state i from Equation (10), that is

$$\pi_i = c \exp\{-\tfrac{1}{2}\, \Delta d^{iT} C_D^{-1} \Delta d^i - \tfrac{1}{2}\, [m^i - m_0]^T C_M^{-1} [m^i - m_0]\}$$

$$= c \exp\{-\tfrac{1}{2}\, \Delta d^{iT} C_D^{-1} \Delta d^i - \tfrac{1}{2}\, [LZ^i]^T (L^T)^{-1} L^{-1} [LZ^i]\}$$

$$= c \exp\{-\tfrac{1}{2}\, \Delta d^{iT} C_D^{-1} \Delta d^i - \tfrac{1}{2}\, Z^i \cdot Z^i\}$$

where we have introduced the notation, $\Delta d^i \equiv d_{\text{obs}} - g(m^i)$ for the misfit of the production data.

Combining these four results, and canceling the common terms, we obtain

$$\frac{\pi_j q_{ij}}{\pi_i q_{ij}} = \frac{c \exp\left(-\frac{1}{2}\,\Delta d^{j^T} C_D^{-1} \Delta d^j - \frac{1}{2}\, Z^j \cdot Z^j\right) \exp\left(-\frac{1}{2}\, Z_k^2\right)}{c \exp\left(-\frac{1}{2}\,\Delta d^{i^T} C_D^{-1} \Delta d^i - \frac{1}{2}\, Z^i \cdot Z^i\right) \exp\left(-\frac{1}{2}\, Z_k'^2\right)}$$

$$= \frac{\exp\left(-\frac{1}{2}\,\Delta d^{j^T} C_D^{-1} \Delta d^j\right) \exp\left[-\frac{1}{2}\left(Z_1^2 + \cdots + Z_k'^2 + \cdots + Z_M^2\right) - \frac{1}{2} Z_k^2\right]}{\exp\left(-\frac{1}{2}\,\Delta d^{i^T} C_D^{-1} \Delta d^i\right) \exp\left[-\frac{1}{2}\left(Z_1^2 + \cdots + Z_k^2 + \cdots + Z_M^2\right) - \frac{1}{2} Z_k'^2\right]}$$

$$= \frac{\exp\left(-\frac{1}{2}\,\Delta d^{j^T} C_D^{-1} \Delta d^j\right)}{\exp\left(-\frac{1}{2}\,\Delta d^{i^T} C_D^{-1} \Delta d^i\right)} \tag{19}$$

Thus, when a square root matrix, R or L, is used to generate the perturbations to the permeability field, the acceptance criterion, Equation (6), depends only on the misfits to the production data. If the proposed permeability field generates production data that fit the observed data more closely than the previous permeability field, then we accept the perturbation and update the field. If the fit is worse, we accept the transition with probability given by Equation (19). It is unnecessary to calculate the misfit to the variogram, and convergence is accelerated because *all* proposed transitions are consistent with the probability distribution for the variogram.

Perturbations that Use Sensitivity Functions

We can increase further the probability of generating acceptable transitions by selecting transitions that are approximately consistent with the production data. To see how this might be done, let us examine the likelihood part of the total probability density function:

$$\Pi_{D|M}(d_{\text{obs}}|m) \propto \exp\left\{-\tfrac{1}{2}\,[d_{\text{obs}} - g(m)]^T C_D^{-1} [d_{\text{obs}} - g(m)]\right\}$$

where, in general, $g(m)$ is a nonlinear functional of m.

Select a reference state, say m_r, and expand $g(m)$ in a Taylor series around m_r.

$$g(m) = g(m_r) + G_r(m - m_r) + \epsilon(m)$$

where

$$\epsilon(m) \equiv g(m) - g(m_r) - G_r(m - m_r)$$

and $G_r m$ is a *linear* functional of m. Also, define a reference data vector, d, such that

$$d = d_{\text{obs}} - g(m_r) + G_r m_r$$

so

$$\begin{aligned} d_{\text{obs}} - g(m) &= d_{\text{obs}} - [g(m_r) + G_r m - G_r m_r + \epsilon(m)] \\ &= [d_{\text{obs}} - g(m_r) + G_r m_r] - G_r m - \epsilon(m) \\ &= d - G_r m - \epsilon(m) \end{aligned}$$

The [redacted] likelihood of obtaining the observed data [redacted] a particular set of

[redacted]

random variables, m, can be factored into the product of Gaussian and non-Gaussian distributions:

$$\begin{aligned} \Pi_{D|M}(d_{\text{obs}}|m) &\propto \exp\{-\tfrac{1}{2}[d - G_r m - \epsilon(m)]^T C_D^{-1}[d - G_r m - \epsilon(m)]\} \\ &= \exp\{-\tfrac{1}{2}[d - G_r m]^T C_D^{-1}[d - G_r m] \\ &\quad + [d - G_r m]^T C_D^{-1}\epsilon(m) - \tfrac{1}{2}\epsilon(m)^T C_D^{-1}\epsilon(m)\} \end{aligned}$$

The same factorization then can be used in the probability distribution for the random variables.

$$\begin{aligned} \Pi_M(m) &\propto \exp\{-\tfrac{1}{2}[m - m_0]^T C_M^{-1}[m - m_0] - \tfrac{1}{2}[d - G_r m]^T C_D^{-1}[d - G_r m]\} \\ &\quad \times \exp\{[d - G_r m]^T C_D^{-1}\epsilon(m) - \tfrac{1}{2}\epsilon(m)^T C_D^{-1}\epsilon(m)\} \\ &\propto \exp\{-\tfrac{1}{2}[m - \mu]^T C_{M'}^{-1}[m - \mu]\} \\ &\quad \times \exp\{[d - G_r m]^T C_D^{-1}\epsilon(m) - \tfrac{1}{2}\epsilon(m)^T C_D^{-1}\epsilon(m)\} \end{aligned} \tag{20}$$

where the reference mean, μ, is calculated from

$$\mu = m_0 + C_M G_r^T [G_r C_M G_r^T + C_D]^{-1}[d - G_r m_0] \tag{21}$$

Chain of Independent Realizations

We again specify the state i of the model by the values of an array of independent normal deviates,

$$Z^i = (Z_1, Z_2, Z_3, \ldots, Z_M)^T$$

that are related to the random permeability field as

$$m^i = \mu + LZ^i \quad \text{where} \quad C_{M'} = LL^T$$

We will propose transitions to a new state, j, for which every element of the array of normal deviates has been resampled:

$$Z^j = (Z_1', Z_2', Z_3', \ldots, Z_M')^T$$

Because the components of Z^j are independent, the probability of proposing the transition from state i to state j is

$$q_{ij} = (2\pi)^{-M/2} \exp(-\tfrac{1}{2} Z^j \cdot Z^j)$$

From Equations (20) and (22), this is equivalent to

$$\begin{aligned} q_{ij} &\propto \exp\{-\tfrac{1}{2}[m^j - \mu]^T C_{M'}^{-1}[m^j - \mu]\} \\ &\propto \exp\{-\tfrac{1}{2}[m^j - m_0]^T C_M^{-1}[m^j - m_0] \\ &\quad - \tfrac{1}{2}[d - G_r m^j]^T C_D^{-1}[d - G_r m^j]\} \end{aligned} \tag{23}$$

The probability of proposing the reverse transition, from state j to i, is determined by replacing m^j by m^i in Equation (23).

Recall that we can calculate the probability of being in state i from Equation (20):

$$\begin{aligned} \pi_i &\propto \exp\{-\tfrac{1}{2}[m^i - \mu]^T C_{M'}^{-1}[m^i - \mu]\} \\ &\quad \times \exp\{[d - G_r m^i]^T C_D^{-1} \epsilon(m^i) - \tfrac{1}{2}\epsilon(m^i)^T C_D^{-1}\epsilon(m^i)\} \end{aligned}$$

Combining these four results, and canceling the common terms, we obtain

$$\begin{aligned} \frac{\pi_j q_{ji}}{\pi_i q_{ij}} &= \frac{\exp\{[d - G_r m^j]^T C_D^{-1}\epsilon(m^j) - \tfrac{1}{2}\epsilon(m^j)^T C_D^{-1}\epsilon(m^j)\}}{\exp\{[d - G_r m^i]^T C_D^{-1}\epsilon(m^i) - \tfrac{1}{2}\epsilon(m^i)^T C_D^{-1}\epsilon(m^i)\}} \\ &= \frac{\exp\{[d_{\text{obs}} - g(m^j)]^T C_D^{-1}\epsilon(m^j) + \tfrac{1}{2}\epsilon(m^j)^T C_D^{-1}\epsilon(m^j)\}}{\exp\{[d_{\text{obs}} - g(m^i)]^T C_D^{-1}\epsilon(m^i) + \tfrac{1}{2}\epsilon(m^i)^T C_D^{-1}\epsilon(m^i)\}} \end{aligned} \tag{24}$$

which we use in Hasting's criterion, Equation (6), to decide whether to accept the transition to state j. Note that if $\epsilon(m) = 0$ for all m then the ratio in Equation

(24) is equal to one, and all proposed transitions are accepted. If $\epsilon(m)$ is small, we expect high acceptance ratios.

Incidentally, this is approximately the approach used previously by the authors (Oliver, 1994, 1996; Chu, Reynolds, and Oliver, 1995) to generate realizations that were conditional to well test data. In those situations, however, the proposed transitions were not checked against the acceptance criterion because the nonlinearity was presumed to be small. We will examine the consequences of rigorously applying the Markov chain Monte Carlo method to the numerical problem when we consider computational examples.

Local Perturbations

One problem with modifying all of the random variables at each transition is that the number of accepted transitions can be low if the non-Gaussian part of the distribution is large and the observed production data are accurate (because if the data are accurate, then only perturbations that closely match the observed data are accepted). The probability of proposing a large transition that results in a good match to accurate production data can be small. An alternative is to make a large number of small transitions as discussed previously. In this approach, the transition from state i,

$$Z^i = (Z_1, Z_2, \ldots, Z_k, \ldots, Z_M)^T$$

to state j,

$$Z^j = (Z_1, Z_2, \ldots, Z'_k, \ldots, Z_M)^T$$

involves changing only one component of the array of normal deviates. The probability of proposing this transition is

$$q_{ij} = \frac{1}{M\sqrt{2\pi}} \exp\left(-\frac{1}{2} Z_k'^2\right)$$

whereas the probability of proposing the reverse transition is

$$q_{ji} = \frac{1}{M\sqrt{2\pi}} \exp\left(-\frac{1}{2} Z_k^2\right)$$

The probabilities of being in states i or j were reported in the last section but the results are more convenient when expressed in terms of the normal deviates:

$$\pi_i \propto \exp\left\{-\tfrac{1}{2}(Z_1^2 + \cdots + Z_k^2 + \cdots + Z_M^2)\right\}$$

$$\times \exp\left\{[d - G_r m^i]^T C_D^{-1} \epsilon(m^i) - \tfrac{1}{2}\epsilon(m^i)^T C_D^{-1} \epsilon(m^i)\right\}$$

and

$$\pi_j \propto \exp\{-\tfrac{1}{2}(Z_1^2 + \cdots + Z_k'^2 + \cdots + Z_M^2)\}$$

$$\times \exp\{[d - G_r m^j]^T C_D^{-1} \epsilon(m^j) - \tfrac{1}{2}\epsilon(m^j)^T C_D^{-1} \epsilon(m^j)\}$$

Combining these four results, and canceling the common terms, we obtain

$$\frac{\pi_j q_{ji}}{\pi_i q_{ij}} = \frac{\exp\{[d - G_r m^j]^T C_D^{-1} \epsilon(m^j) - \tfrac{1}{2}\epsilon(m^j)^T C_D^{-1} \epsilon(m^j)\}}{\exp\{[d - G_r m^i]^T C_D^{-1} \epsilon(m^i) - \tfrac{1}{2}\epsilon(m^i)^T C_D^{-1} \epsilon(m^i)\}}$$

$$= \frac{\exp\{[d_{\text{obs}} - g(m^j)]^T C_D^{-1} \epsilon(m^j) + \tfrac{1}{2}\epsilon(m^j)^T C_D^{-1} \epsilon(m^j)\}}{\exp\{[d_{\text{obs}} - g(m^i)]^T C_D^{-1} \epsilon(m^i) + \tfrac{1}{2}\epsilon(m^i)^T C_D^{-1} \epsilon(m^i)\}} \tag{25}$$

which is the same as Equation (24). Thus, the acceptance criteria is the same whether we do local perturbations or global perturbations. The acceptance *probability*, however, will be different for the two different types of transition because in the first approach m^j is "close" to m^i whereas in the second approach the two states can be different. The disadvantage is that many of the proposed transitions will turn out to have no effect on the misfit of the production data because, for example, the perturbed cells are located outside the region of influence of the data. If this were the situation, we would obtain $g(m^j) = g(m^i)$ and $\epsilon(m^j) = \epsilon(m^i)$ so that the acceptance probability, by Equation (25), would be equal to one. Conceptually, we can determine whether this is a good approximation without running the flow simulator by calculating the product of the sensitivity coefficients with the proposed transition. If the product $G_r(m^j - m^i)$ is approximately zero, then the production data are insensitive to the proposed change, and it is unnecessary to run the flow simulator.

COMPUTATIONAL EXAMPLES

Here, we use Markov chain Monte Carlo methods to generate realizations of the log-permeability field that are consistent with transient pressure data and correctly sample the *a posteriori* probability density function. For the first example the permeability is log-normally distributed with log variance of 0.25 and log mean equal to 3.4. In the second example the mean is 3.4 but the variance is increased to 1.0. When we solve these example problems, we assume that the permeability of each gridblock is a scalar and that the 2-D spatial continuity can be described by an isotropic spherical variogram model with a range of 600 ft. In both synthetic examples, the reservoir is square, 1500 × 1500 ft, with no flow boundaries along the edges. Reservoir performance was simulated using a uniform spatial grid with 100 × 100 ft gridblocks, so there are 225 unknown permeability values in each model.

Ten observations of pressure were obtained at each of five wells whose locations are shown in Figure 4. The central, active well produced 750 barrels per day of fluid; the other four are observation wells, used only to acquire pressure data. Measurements of the pressure began at 0.05 days and ended at 5.7 days. The error variance for all pressure measurements is assumed to be 0.15 psi^2. All computations were done on a Pentium-90.

VARIANCE EQUAL TO 0.25

The "true" permeability distribution, shown in Figure 4, is an unconditional simulation generated using the Cholesky decomposition method for a spherical variogram. Values of $\ln(k)$ range from 2.0 to 5.0, or, in terms of k, from 7 to 148 md. "True" pressure data for the five wells were generated using the true gridblock permeabilities in a finite difference flow simulator. We used these data in three of the MCMC methods described previously to generate Markov chains of possible permeability fields. In the first method, the proposed perturbations are local and are based on the square root of the prior covariance, so they honor only the variogram. In the second and third methods, the sensitivity coefficients from the reservoir simulator are used to modify the direction and the magnitude of the proposed perturbations. The perturbations in the second method are local, in the third method they are global, meaning that the permeability in every gridblock is modified in every transition.

One way to compare the three methods is to look at the rate of generation of acceptable independent realizations for each. Figure 5 shows the values of the "objective function" [i.e., the argument of the exponential in Eq. (10)] vs. the perturbation number. The perturbation number in this figure refers to the number of transitions proposed, but only the values of accepted transitions are plotted.

Acceptance of a perturbation based on Hasting's acceptance criterion, Equation (6), does not necessarily indicate that the resulting state is drawn from the correct probability distribution. There usually is a transient period at the beginning of a Markov chain that must be discarded on the grounds that equilibrium has not been reached. For the MCMC methods that use the *a posteriori* covariance to generate perturbations, the transient period is so short that it is difficult to see any evidence of it in Figure 5. Close examination of the data shows that the transient period ends after about 30 perturbations. It might seem from Figure 5 that the transient period for the first MCMC method, which uses local perturbation based only on the variogram, also is short. A careful inspection, however, indicates that the values of the objective function probably are decreasing after 50,000 perturbations so, for this method, all of the first 50,000 realizations are in the transient period.

When the stationary distribution of the Markov chain is reached, the values

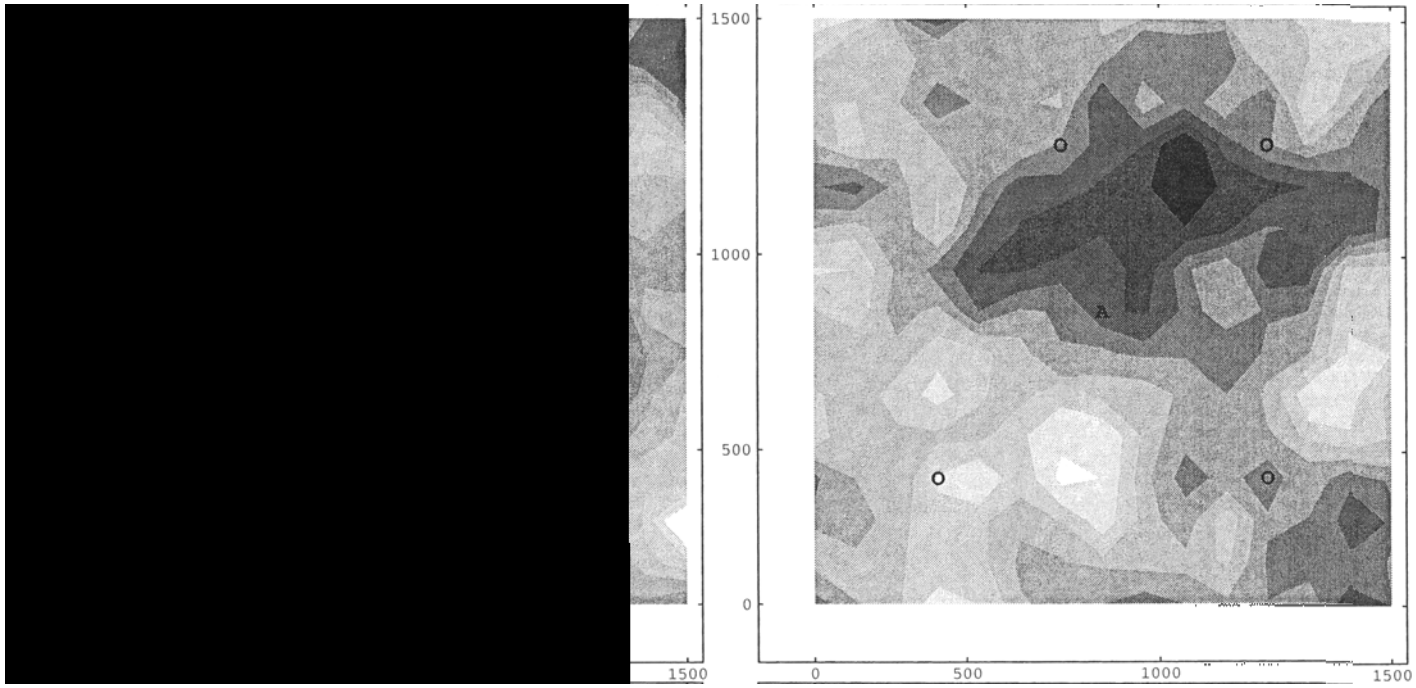


Figure 4. Log-permeability distribution ranges from less than 2.25 in black regions to greater than 4.75 in white region. Well locations are denoted with "A" for active well and "O" for observation well. "True" log-permeability is on left and conditional realization is on right.

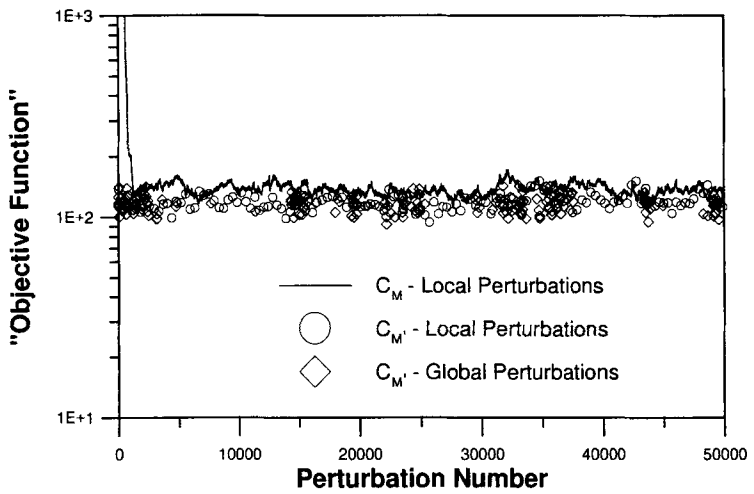


Figure 5. Values of objective function for three Markov chain Monte Carlo methods.

of the objective function cluster around a value greater than or equal to $M/2$ where M is the number of model parameters. Because the pressure data used in these examples to calculate the maximum *a posteriori* model were exact, and the prior model was close to the *a posteriori* model, the average value of the objective function in the stationary phase should be approximately equal to $M/2 = 112.5$. The observed mean of the objective function, computed for the set of accepted realizations for the method that uses local perturbations based on the *a posteriori* covariance is 119 with a corresponding estimate of the standard deviation given by 10. The method that uses global perturbations based on the *a posteriori* covariance resulted in a mean value of 111 and a standard deviation of 10.

Each of the three methods required approximately 13 hours to attempt 50,000 perturbations. The Markov chain Monte Carlo method that used local perturbations based on the variogram did not generate any legitimate images of the permeability field during this period. The MCMC methods that used global perturbations based on the *a posteriori* covariance matrix generated more than 200 legitimate, independent realizations. Using local perturbations based on the *a posteriori* covariance matrix, approximately 37,500 highly correlated images were accepted. As expected, using local perturbations greatly increases the number of states accepted but successive states, or realizations, differ only slightly.

Recall that the ultimate goal usually is not to generate simply possible reservoir models, but to draw inferences based on predictions from the realizations. These predictions may be expensive computationally so it usually is better to have 10–100 independent realizations from the *a posteriori* distribution

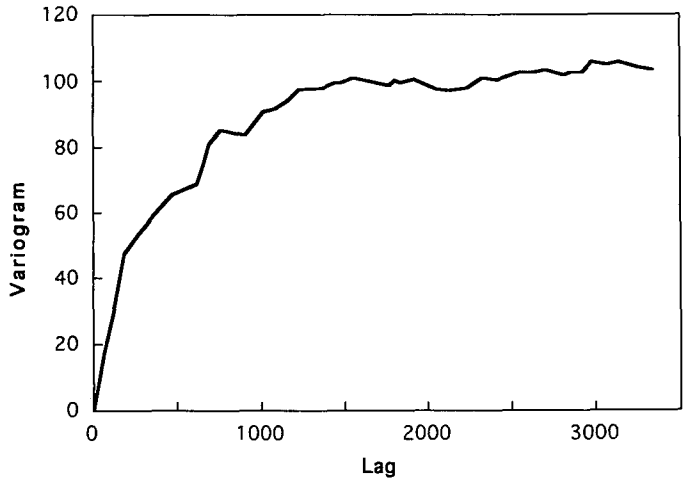


Figure 6. Experimental variogram for objective function of chain of new realizations generated using local perturbations based on *a posteriori* covariance.

than to have 10,000 dependent realizations. Because the dependence decreases with "lag," an approximately independent set of realizations can be obtained by retaining every rth state from the chain, where r is the value of the lag for which the autocorrelation is approximately zero. In Figures 6 and 7 we have plotted the "experimental variogram" of the chain of values of the objective

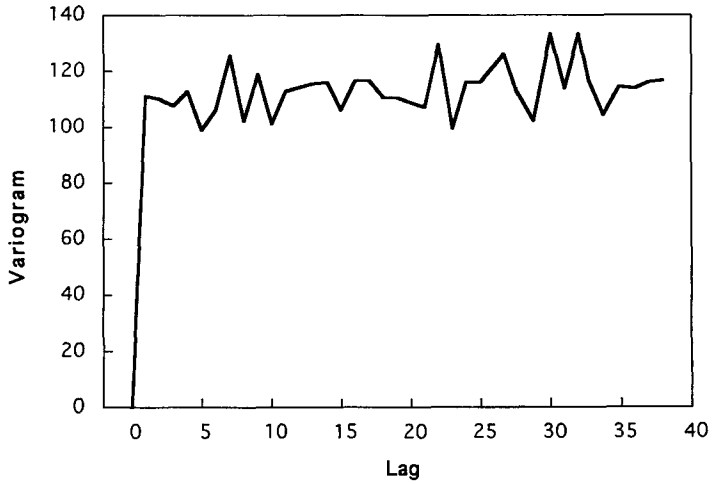


Figure 7. Experimental variogram for objective function of chain of new realizations generated using global perturbations based on *a posteriori* covariance.

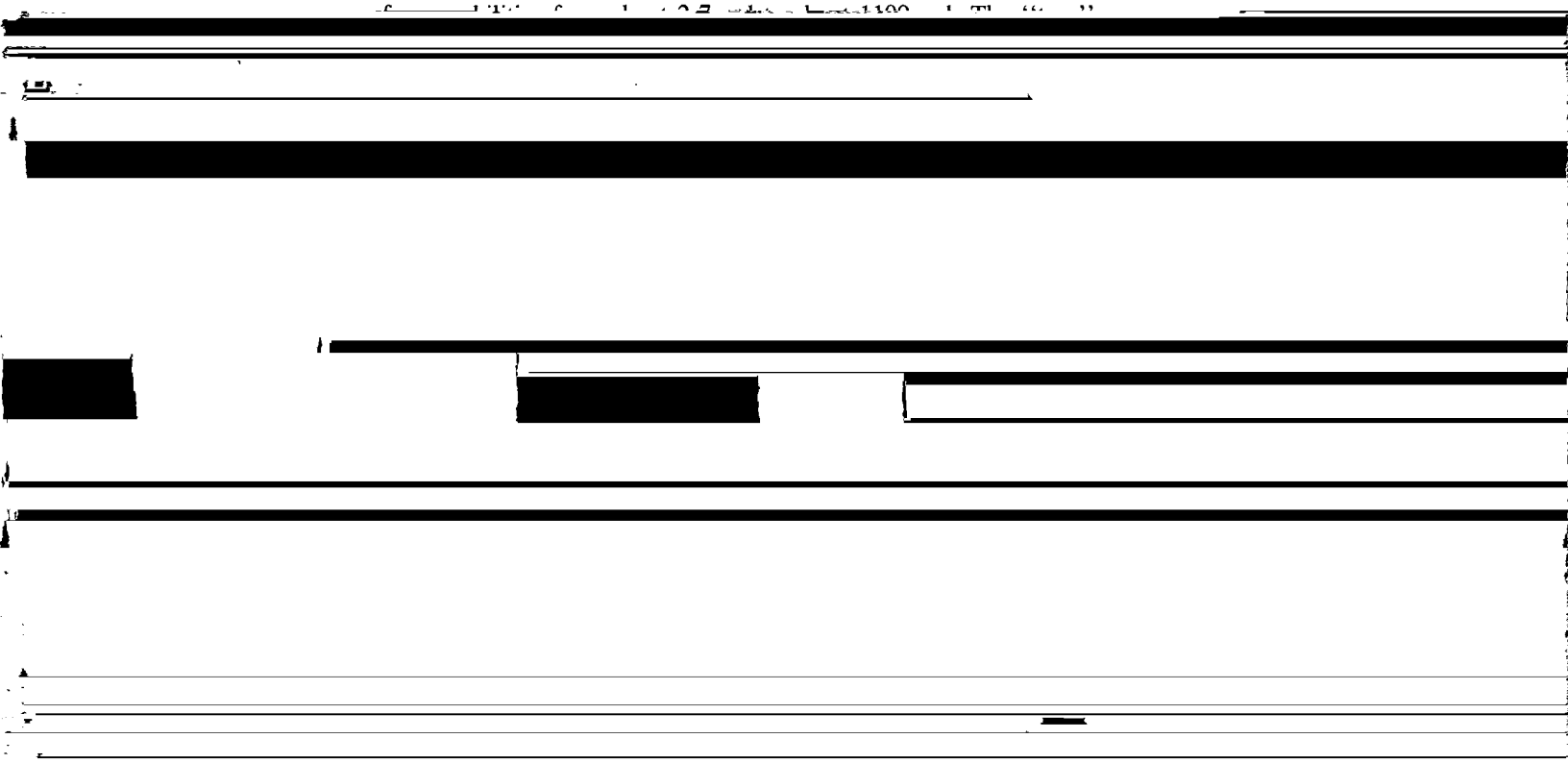
function for the second and third MCMC methods. The experimental variogram is calculated from

$$\gamma(s) = \frac{1}{2N_s} \sum_{t=0}^{N_s-1} (S(m_t) - S(m_{t+s}))^2 \tag{26}$$

where $S(m_t)$ is the value of the argument of the exponential in Equation (10) for the tth accepted transition in the chain. It is clear from Figure 7 that successive new states generated using global perturbations are uncorrelated whereas new states generated from local perturbations are correlated slightly to lags of approximately 1200. If, following the suggestion of Tierney (1994), we retain only every 1200th new state we would be left with 31 independent realizations from local perturbations compared to 200 from global perturbations. For probabilistic inference, each realization would be weighted by the frequency of occurrence of the state within the chain. After thinning to the set of independent realizations, the weighting is equal to the number of *attempted* transitions between the occurrence of the independent states.

VARIANCE EQUAL TO 1.0

In this example, the variance of the log-permeability field is increased to 1.0, all other parameters of the distribution are the same as in the first example. The "true" log-permeability field, generated using the Cholesky decomposition method with the same random seed as the previous example, is shown in Figure 8. Values of $\ln(k)$ range from about 1.0 to about 7.0, corresponding to a range



data again are obtained from a finite difference simulation of flow in the synthetic reservoir.

Figure 9 shows values of the "objective function" of accepted states vs.

Figure 8. Log-permeability distribution ranges from less than 1.5 in black regions to greater than 6.5 in white region. Well locations are denoted with "A" for active well and "O" for observation well. "True" log-permeability is on left and conditional realization is on right.

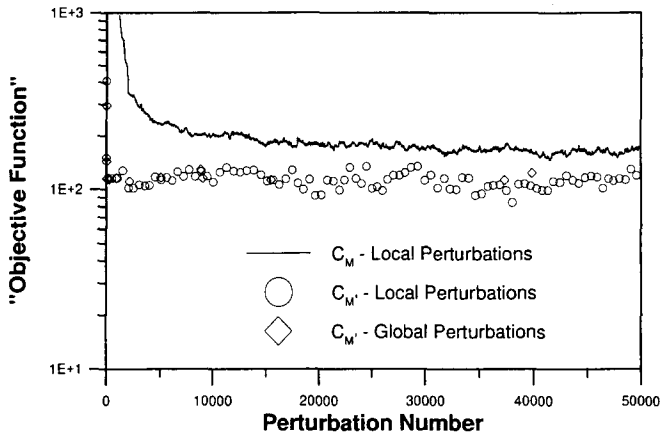


Figure 9. Values of objective function for three Markov chain Monte Carlo methods.

result in a correct sampling of the conditional probability distribution for the random variables, given a theoretical variogram model and observations that are related to the random variables through a possibly nonlinear functional. By carefully selecting the shape and amplitude of the proposed perturbations, high acceptance ratios can be obtained in the MCMC method of generating realizations.

For linear problems, such as conditioning a permeability field to "hard data" and a variogram model, we outlined procedures for generating perturbations that will be accepted. These methods are more useful, however, when conditioning the permeability field to observations that are *not* linearly related to the random variables. In the nonlinear situation, we showed that by incorporating the sensitivity information from a flow simulator into the sampling distribution for proposed transitions, the probability of acceptance of the transitions can be increased significantly compared to methods that do no use this information.

Although a high acceptance ratio is desirable, it generally is more important for the perturbation schemes to be efficient at producing *independent* realizations from a MCMC procedure. In Figures 6 and 7, we determined that, for the moderately nonlinear example ($\sigma^2 = 0.25$), only 1 of 1200 accepted transitions from local perturbation results in independent realizations whereas all accepted transitions from global perturbations result in independent realizations. In the more strongly nonlinear example ($\sigma^2 = 1.0$), a similar analysis shows that approximately 1800 accepted transitions are required to generate an independent realization from local perturbations. Finally, for the linear problem, an analysis of the probability of visiting 90% of the sites indicates that approximately 1 in

sity Petroleum Reservoir Exploitation Projects, and the U.S. Department of Energy under BDM-Oklahoma, Inc. Subcontract No. 93-0001. Financial support for the graduate studies of Luciane B. Cunha was provided by the preceding sources, Conselho Nacional de Desenvolvimento Científico e Technológico (CNPq) and Petróleo Brasileiro S.A. (PETROBRÁS).

REFERENCES

Alabert, F., 1987, The practice of fast conditional simulations through the LU decomposition of the covariance matrix: Math. Geology, v. 19, no. 5, p. 369–386.

Chu, L., Reynolds, A. C., and Oliver, D. S., 1995, Computation of sensitivity coefficients for conditioning the permeability field to well-test pressure data: In Situ, v. 19, no. 2, p. 179–223.

Davis, M., 1987a, Generating large stochastic simulations—the matrix polynomial approximation method: Math. Geology, v. 19, no. 2, p. 99–107.

Davis, M., 1987b, Production of conditional simulations via the LU decomposition of the covariance matrix: Math. Geology, v. 19, no. 2, p. 91–98.

Deutsch, C. V., 1992, Annealing techniques applied to reservoir modeling and the integration of geological and engineering (well test) data: unpubl. doctoral dissertation, thesis, Stanford University, 325 p.

Deutsch, C. V., and Journel, A. G., 1992, GSLIB: Geostatistical Software Library and user's guide: Oxford Univ. Press, New York, 340 p.

Dietrich, C. R., 1993, Computationally efficient generation of Gaussian conditional simulations over regular sample grids: Math. Geology, v. 25, no. 4, p. 439–451.

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.

Farmer, C. L., 1992, Numerical rocks, *in* King, P. R., ed., Mathematics of oil recovery: Clarendon Press, Oxford, p. 437–447.

Feller, W., 1968, An introduction to probability theory and its applications, v. I (3rd ed.): John Wiley & Sons, New York, 509 p.

Geman, S., and Geman, D., 1984, Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images: IEEE Trans. Pattern Analysis and Machine Intelligence, v. PAMI-6, no. 6, p. 721–741.

Gupta, A. D., 1992, Stochastic heterogeneity, dispersion, and field tracer response: unpubl. doctoral dissertation, Univ. Texas, Austin, 248 p.

Hammersley, J. M., and Handscomb, D. C., 1964, Monte Carlo methods: John Wiley & Sons, New York, 178 p.

Hastings, W. K., 1970, Monte Carlo sampling methods using Markov chains and their applications: Biometrika, v. 57, no. 1, p. 97–109.

Hegstad, B. K., Omre, H., Tjelmeland, H., and Tyler, K., 1993, Stochastic simulation and conditioning by annealing in reservoir description, *in* Armstrong, M., and Dowd, P. A., eds., Geostatistical simulation: Kluwer Acad. Publ., Dordrecht, The Netherlands, p. 43–55.

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

Koren, Z., Mosegaard, K., Landa, E., Thore, P., and Tarantola, A., 1991, Monte Carlo estimation and resolution analysis of seismic background velocities: Jour. Geophys. Res., v. 96, no. B12, p. 20,289–20,299.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., 1953, Equations of state calculations by fast computing machines: Jour. Chem. Physics, v. 21, no. 6, p. 1087-1092.

Neal, R. M., 1993, Probabilistic inference using Markov chain Monte Carlo methods: Tech. Rept. CRG-TR-93-1, Dept. Computer Science, Univ. Toronto, 140 p.

Oliver, D. S., 1994, The incorporation of transient pressure data into reservoir characterization: In Situ, v. 18, no. 3, p. 243-275.

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

Oliver, D. S., 1996, Multiple realizations of the permeability field from well test data: Soc. Petrol. Eng. Jour., v. 1, no. 2, p. 145-154.

Ouenes, A., 1992, Application of simulated annealing to reservoir characterization and petrophysics inverse problems: unpubl. doctoral dissertation, New Mexico Tech., 205 p.

Ouenes, A., Bréfort, B., Muenier, G., and Dupéré, S., 1993, A new algorithm for automatic history matching: Application of simulated annealing method (SAM) to reservoir inverse modeling: unsolicited manuscript SPE 26297, 30 p.

Pérez, G., 1991, Stochastic conditional simulation for description of reservoir properties: unpubl. doctoral dissertation, Univ. Tulsa, 245 p.

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

Ripley, B. D. 1981, Spatial statistics: John Wiley & Sons, New York, 252 p.

Ripley, B. D., 1987, Stochastic simulation: John Wiley & Sons, New York, 237 p.

Sagar, R. K., Kelkar, B. G., and Thompson, L. G., 1993, Reservoir description by integration of well test data and spatial statistics, *in* Proc. 68th Ann. Tech. Conf. Exhibition: Soc. Petrol. Eng., p. 475-489.

Scheuer, E. M., and Stoller, D. S., 1962, On the generation of normal random vectors: Technometrics, v. 4, no. 2, p. 278-281.

Sen, M. K., Gupta, A. D., Stoffa, P. L., Lake, L. W., and Pope, G. A., 1992, Stochastic reservoir modeling using simulated annealing and genetic algorithm, *in* Proc. 67th Ann. Tech. Conf. Exhibition: Soc. Petrol. Eng., p. 939-950.

Srivastava, R. M., 1994, The interactive visualization of spatial uncertainty, *in* Proc. Univ. Tulsa Centennial Petroleum Engineering Symposium: Soc. Petrol. Eng., p. 87-95.

Tarantola, A., and Valette, B., 1982, Inverse problems = quest for information: Jour. Geophysics, v. 50, no. 3, p. 159-170.

Tierney, L., 1994, Markov chains for exploiting posterior distributions: Tech. Rept. 560 (revised), School of Statistics, Univ. Minnesota, 25 p.

Tjelmeland, H., Omre, H., and Hegstad, B. K., 1994, Sampling from Bayesian models in reservoir characterization: Tech. Rept. Statistics No. 2/1994, Univ. Trondheim, 15 p.

APPENDIX

Square Root of the Covariance

The covariance matrix is symmetric and positive definite so it can be decomposed into the product $C_M = W\Lambda W^T$ where the columns of W are orthonormal eigenvectors of C_M and Λ is the diagonal matrix of eigenvalues of C_M Also, $W^TW = I$ so

$$C_M = (W\Lambda^{1/2}W^T)(W\Lambda^{1/2}W^T)$$
$$= R^2$$

and R is clearly symmetric.

We then could use $m = \mu + RZ$ instead of $m = \mu + LZ$ to generate perturbations. When C_M comes from a variogram, this method has the nice property of making all perturbations to the permeability distribution have the same shape when we use the normal deviates to generate the perturbation.

The use of the square-root decomposition of the covariance matrix has been previously proposed in the literature. Davis (1987a) suggested its use in simulation because the calculation of the matrix square root could be accomplished efficiently by a polynomial approximation. More recently, Dietrich (1993) proposed a fast method for multidimensional simulation of Gaussian random fields that also uses the square root of the covariance matrix instead of the LU decomposition. Dietrich and Newsam (1995) took the polynomial method of calculating the square root of the covariance a step further by suggesting its use for conditional simulation.