Theoretical Justification for Iterative Simulation Methods such as the Gibbs Sampler

Steven Lyster

Centre for Computational Geostatistics
Department of Civil and Environmental Engineering
University of Alberta

The Gibbs Sampler algorithm has proven useful for the integration of multiple-point statistics in geostatistical realizations of discrete variables. Iterative simulation methods such as the Gibbs Sampler have advantages such as speed and ease of simulation, as local data need not be searched for and systems of equations may be solved once and the solutions retained. However, these iterative approaches are heuristic; the methods are used because they work, with less emphasis placed on why *they work. This paper explains the statistical underpinnings of iterative simulations utilizing a Gibbs sampler, including: viewing simulated realizations as random variables (or fields); the concept of exploration of the state space of a problem; and convergence of Markov chain Monte Carlo methods to the stationary joint distributions of complex variables.*

Introduction

Geostatistical simulation methods have become more sophisticated in recent years, moving from local estimation to joint simulation and accounting for multiple sources of data and increasingly complex spatial statistics. One factor holding back development of some methods is computational time required for simulating a large number of realizations. A possible solution that has been explored is the use of iterative simulation methods rather than sequential; while sequential methods require searches for nearby data and calculation of new conditional distributions at every location, iterative methods have the advantage of knowing the surrounding grid locations are fully informed and not having to re-solve sets of equations at every point. These speed advantages can often offset the fact that iterative simulation must visit each location many times.

Although the benefits of a sequential simulation framework are relatively easy to explain, the idea that an iterative simulation can be valid may be more difficult to understand. This paper will explain the theoretical underpinnings of iterative simulation in geostatistics and the relations to standard and well-known statistical resampling techniques.

Joint Uncertainty

Early use of geostatistical methods focused on estimation of unknown variables; this lead to the development of kriging and related approaches. Simulation was attractive early on because of the honouring of the variogram and removal of the smoothing effect of kriging; however, it was not until the 1980s and 90s that computers became powerful enough to make simulation a feasible tool.

While estimation allows the determination of “best” local distributions for unsampled locations, the global heterogeneity is not properly honoured. Each estimate is its own entity, unrelated to the rest. This allows calculation of such interesting properties such as local mean and variance, probability to be above or below a threshold, and probabilities of different facies at a location. However, these local estimates, when taken together, do not properly represent the global input statistics (Deutsch, 2002).

In order to properly characterize unknown joint distributions such as global tonnage and grade above a cutoff, effective permeability of a field, or connectivity of certain facies, simulation is required. Important response characteristics may then be determined by flow simulation, dig limit optimization, or other such production predicting methods. It is not possible to directly predict any of these values, as they are functions of the entire simulated field which may be comprised of several million cells. Simulation allows

an approximation of the distributions of response characteristics by sampling from the joint distribution; all realizations are deemed equally probable and representative of the underlying spatial structure. By honouring the attributes of the geologic phenomena that may be modeled, the realizations generated may then be used to determine the uncertainty in the complex responses.

State Spaces

In statistics, a *state* is a set of variables with given values. For example, a set X with two variables a and b may be expressed as:

$$X = \{a, b\} \tag{1}$$

So if a and b are fixed then X represents a single state. In this case X may be viewed as a two-dimensional random variable with components a and b. A *state space* is then the set of all possible values for a random variable; in the case of X, this is all values on a two-dimensional plane. For higher-dimensional variables the number of possible states increases geometrically. A common state space to consider in geostatistics is all possible linear estimation weights; the kriging weights are a particular state within this space. Because states may be viewed as random variables within their space, it follows that functions may be defined dependent on these variables. Following with the estimation weights example, the estimation error variance is one such function and the kriging weights are the state at which this function is minimized.

In geostatistical terms, one random variable that may be considered is the combination of the values at all locations in the model; if there are *nxyz* (unsampled) locations then this variable is of dimension *nxyz*. The state space for the model is every combination of values for all locations. If there are K values any given location may take then there are K^{nxyz} states in the space of the model. Note that the values for a single location will be discussed as in the discrete case, i.e. facies models; however, this applies in the continuous case for all practical purposes as only so many decimal places are carried and this limits the possible values. The information which is to be inferred from the models may be defined as functions on the state space of all possible values; this information could be local proportions, grade above cutoff, flow simulation results, resource in place, or any other response characteristic. Each state (possible configuration or realization) has its own set of responses.

Given the geometric nature of the space, there are a huge number of states. Even for a low number such as K=3 and a small model with *nxyz*=100,000, there are over $10^{47,000}$ states. This number may seem excessively large but the important property is that there are a finite number of states. The finite nature of the problem gives several important properties: there are some states with probability greater than zero; the sum of the probabilities of all states is equal to one; and, given some prior knowledge such as a variogram or histogram the vast majority of the states will have a probability equal to zero. The goal of geostatistical simulation is to sample from the probability distribution of this state space in an unbiased way, while honouring both the probability of each state (the likelihood function) and the available data (prior knowledge).

Most of the states which have similar simple statistics such as the variogram will be quite similar in behaviour and response characteristics. Because of this, a small number (about 100) realizations is usually enough to characterize the distribution of response characteristics. While a totally random field is a faint possibility, this is highly unlikely and therefore has an extremely low likelihood (probably equal to zero in most cases) and its response characteristics may be ignored. Most other states may be ignored in this way as well. Only those realizations which are simulated are assumed to have a high likelihood, and therefore contain a large proportion of the weight of the likelihood function. Sampling from the distribution of the state space is sufficient.

Markov Chain Monte Carlo

Markov chain Monte Carlo methods are a family of statistical algorithms which are used to sample from distributions too complex to express analytically (Robert and Casella, 2004). The basis for MCMC algorithms is the Markov chain, which is a string of variables that has the following property:

$$P\left(X_i = x_i \mid X_{i-1} = x_{i-1}, X_{i-2} = x_{i-2}, \ldots, X_0 = x_0\right) = P\left(X_i = x_i \mid X_{i-1} = x_{i-1}\right) \quad (2)$$

A Markov chain explores a state space by stepping from one state, X_{i-1}, to another, X_i, within the state space of the variable with some defined probability. In words, Equation 2 means that the next step in a Markov chain depends only on the current state; that is, how the chain arrived at X_{i-1} is irrelevant. This is called the *Markov property*. The simplest example of a Markov chain is the *random walk*. This algorithm simply adds a random value (possibly negative) to the chain at each step. An example of this is shown in Figure 1.

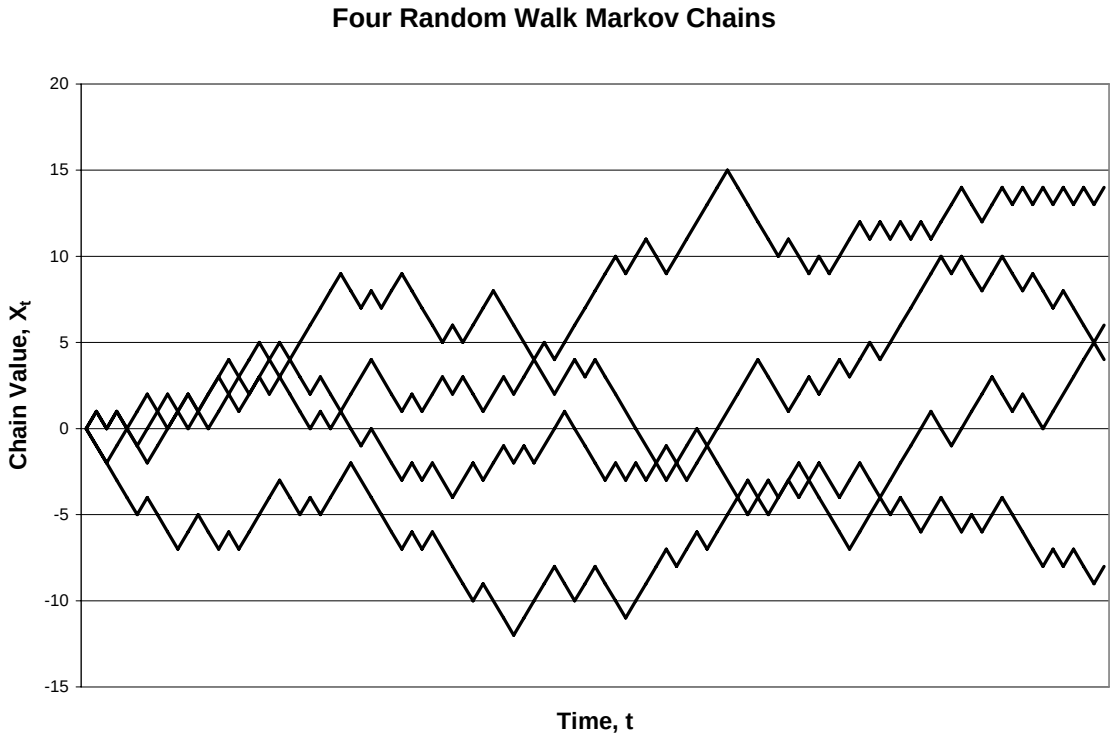

Figure 1: Example of a random walk Markov chain. At each time increment the chain steps up or down by one with equal probability.

A very useful property of Markov chains is that they will converge to stationary distributions after some time. For the random walk shown in Figure 1, the values of a walk as time increases will tend to return towards zero; however, not all of the walks will return at the same rate. This causes many walks (or realizations) taken together to have a Gaussian distribution (Lawler, 2006), even though this distribution was never used in the actual simulation.

Other, more sophisticated, MCMC algorithms are used for characterizing more complex distributions. The slice sampler (Robert and Casella, 2004) resamples components of multi-dimensional variables while keeping the others frozen, utilizing only one-dimensional sampling; the Metropolis-Hastings algorithm (Metropolis et al, 1953) uses an accept-reject criterion to cause the Markov chain to "move" towards states with higher probability; simulated annealing (Deutsch, 1992) is a variant of Metropolis-Hastings which uses an objective function to characterize "good" and "bad" steps; and the Gibbs sampler (Geman and Geman, 1984; Casella and George, 1992) uses conditional probabilities to approximate a sample from the full joint distribution of a variable.

Simulated annealing is the primary MCMC technique which has been explicitly applied in geostatistical use. The Gibbs sampler, which will be discussed further here, has been applied implicitly by numerous techniques; the Gibbs sampler is the statistical basis for what most practitioners would think of as an iterative simulation method.

The Gibbs sampler uses the conditional distributions of multiple variables to move from one state to the next. For example, the state space of a chain made up of two variables *X* and *Y* would be explored using the conditional distributions

$$x_i = f\left(x \mid y_{i-1}\right) \text{ and } y_i = f\left(y \mid x_i\right) \tag{3}$$

For two variables which are bivariate normal, *X* and *Y* would be repeatedly resampled from the conditional distributions

$$f\left(x \mid y\right) = \frac{e^{-\frac{(x-\rho y)^2}{2\left(1-\rho^2\right)}}}{\sqrt{2\pi\left(1-\rho^2\right)}}, \quad f\left(y \mid x\right) = \frac{e^{-\frac{(y-\rho x)^2}{2\left(1-\rho^2\right)}}}{\sqrt{2\pi\left(1-\rho^2\right)}} \tag{4}$$

An example of the first two steps in the Markov chain for a bivariate normal variable is shown in Figure 2. This Gibbs sampler converges very quickly to the stationary joint bivariate Gaussian distribution even though this full distribution is never explicitly used.

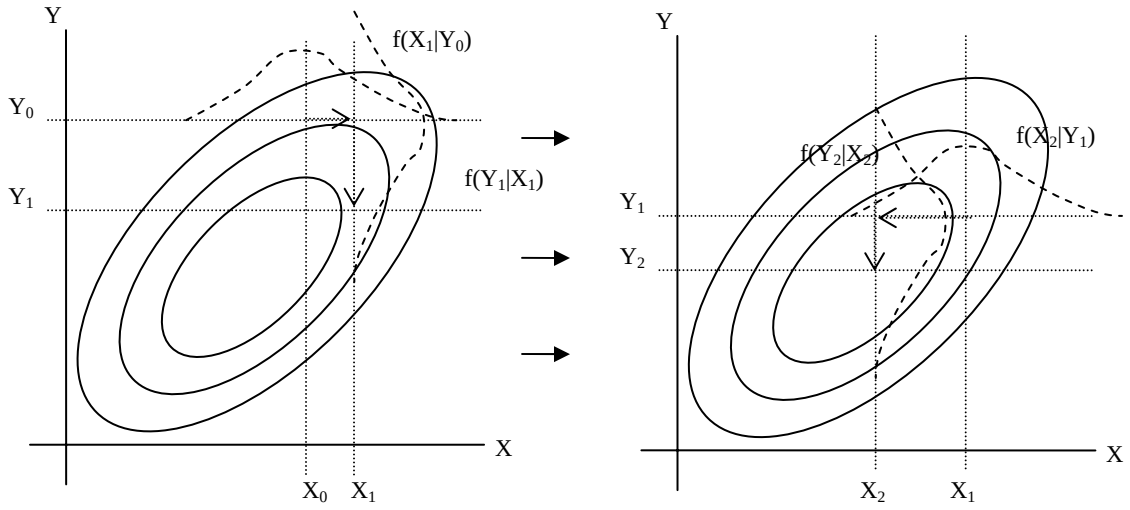

Figure 2: Example of the Gibbs sampler for a pair of bivariate normal variables.

For a higher-dimensional variable, $X^{(i)}$, i=1,…,N, sample $X_{t+1}^{(i)}$ from the distributions $f\left(X^{(i)} \mid X^{(j)}, j \neq i\right)$. This process leads the algorithm to explore the state space of the random variable, with preference to the higher-probability regions. Therefore, running this Markov chain many times is equivalent to sampling from the joint distribution of the variables which make up *X*; with enough samples the properties of the joint distribution may be determined.

In this way, a simulation of *nxyz* variables may be performed which honours the joint distribution of all variables, but does not explicitly require the joint to be known. Only the full conditional distributions are relied upon, and the Markov chain will converge to the stationary joint distribution.

In cases where there is some prior information (such as well logs or sample drillholes) the hard data may be honoured by leaving the corresponding variables frozen in place. Soft data may be honoured by modifying the conditional distributions locally to correspond to what is known. The full joint distribution will then be the likelihood function of the random variable updated with the prior information.

Markov Random Fields

The Gibbs sampler is a very powerful tool for determining unknown joint distributions while accounting for inferred spatial statistics and measured information. However, the Gibbs sampler requires the "full" conditional distributions in order to converge to the joint distribution. The conditional distributions are of dimension *nxyz – 1* and are therefore only slightly less complex than the joint, and even if the conditional distribution may be defined, such as through kriging, it may not be feasible to solve the system. In this case the theory Markov random fields may be used.

The idea behind a Markov random field is to utilize the Markov property spatially. In sequences such as that shown in Equation 2, the random variables which make up the chain have the Markov property and are therefore independent of all previous values in the chain except for the immediately preceding one; this preceding value may be viewed as the "neighbour" in the chain. Markov random fields use this same concept, except instead of being dependent on a neighbour in the dimension of time the variable being simulated is dependent on some set of neighbours in space. This interpretation of the Markov property is demonstrated below:

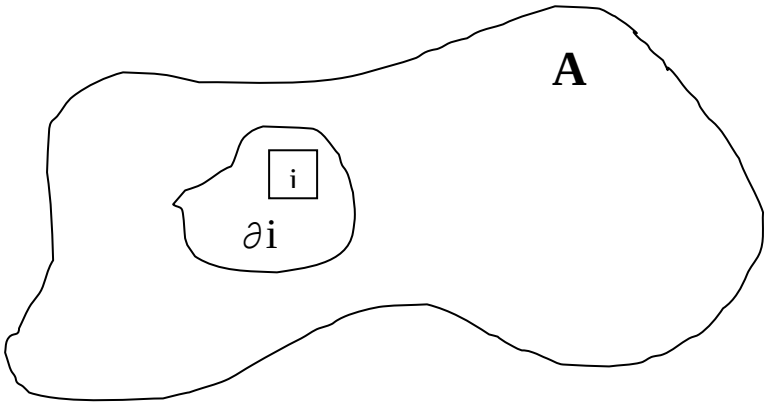

Figure 3: Example of a Markov random field.

In Figure 3, i is the variable being simulated; ∂i is the local neighbourhood of variables around i; and A is the entire set of spatial variables. The mathematical form of the Markov property is then:

$$P\left(X_i = x_i \mid X_j = x_j, j \in A, j \neq i\right) = P\left(X_i = x_i \mid X_j = x_j, j \in \partial i\right) \tag{5}$$

Similar to the Markov property in Equation 2, the distribution of X_i may be determined using only a subset of the available information. This property is used extensively in geostatistics even if it is not known explicitly. The main use of Markov random fields is in limiting search radii when looking for data for kriging or simulation; rather than using all available data, only those within some reasonable range are considered. For a Gibbs sampler of high dimension, the full conditional distributions may be simplified to conditionals considering just the information within a reasonable limit (Gelfand and Smith, 1990). This limited search will then simplify the conditional distributions to something which may be used more practically.

Iterative Geostatistical Simulation

There have been numerous implementations of iterative simulation methods in geostatistics, based on theoretical, heuristic, and image processing techniques. Simulated annealing (Deutsch, 1992) is widely used as an optimization method for maximizing or minimizing a defined objective function. Differences from target statistics (such as the variogram) may be used as the objective function for simulation of mineral deposits. If the objective function is assumed to be representative of the likelihood function of the random field, then simulated annealing explores the state space of the problem in an unbiased way and is theoretically applicable both as a statistical method as well as an optimization approach.

The Gibbs sampler has been applied in geostatistics (Srivastava, 1992), although this approach was mostly from a results-based examination of the problem rather than a theoretical exploration of the space of a high-dimensional variable. Any iterative method which perturbs an image based on local conditional probabilities may be viewed as a Gibbs sampler and the exploration of the state space as outlined above is valid, even if this is not an explicit goal of the methodology.

The relation of earth sciences problems to image processing, pattern recognition, and artificial intelligence has been noted before (Caers and Journel, 1998). Neural networks have been explored as a method for matching conceptual training images which are representative of the geology under consideration. These methods often do not have the aim of statistically exploring the space of the problem, but rather reproducing the patterns and statistics seen in the training image. Neural networks may be viewed as Gibbs samplers, with the neural network portion of the algorithms being used to calculate the conditional distributions.

Example

As a demonstration of the concept of exploration of a state space, Figure 4 shows the space for a small two-by-two grid with black and white values. There are $2^4 = 16$ states in this space. Those states which *communicate*, that is, may be reached from one another in a single step, are connected by lines. For this example, kriging is used to calculate the conditional probabilities, with adjacent nodes in the grid having correlation $\rho = 0.5$ and nodes diagonal to one another having correlation $\rho = 0.2$. The kriging system is the same for all states and is shown in Equation 6.

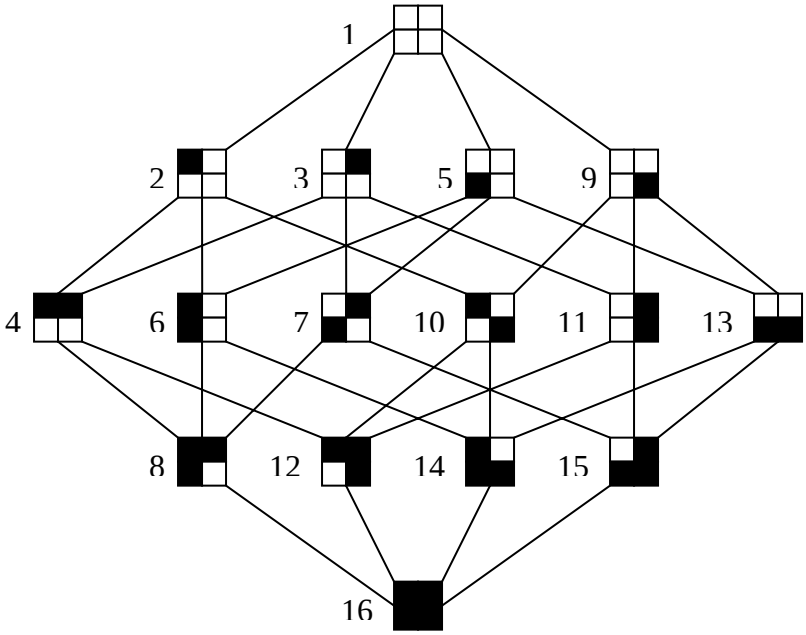

Figure 4: A state space of a simple, 2x2 grid with two possible values.

$$\begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.2 \\ 0.5 & 0.2 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.5 \\ 0.5 \end{bmatrix} \tag{6}$$

Solving this system, $\lambda_1 = -0.3714$, $\lambda_2 = 0.5714$, and $\lambda_3 = 0.5714$. The global proportions used are 0.6 for black and 0.4 for white. Because the state space in the example is such a small size the likelihood of each state may be solved analytically; alternatively, a Gibbs sampler may be used, starting from any state, and exploring the space for some number of iterations. To account for a conditioning point, the state space may

be constrained: Figure 5 shows the same state space for only those states which have white values in the upper left node.

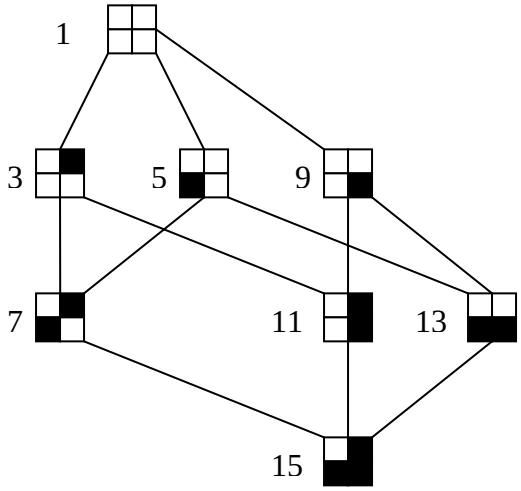

Figure 5: The same state space as in Figure 4, with a conditioning point in the upper location.

The case in Figure 5 may again be solved analytically due to the limited scope of the problem. For these examples, the Gibbs sampler produces results which exactly match the true distributions. The limiting distributions for the Gibbs samplers are shown in Table 1.

Table 1: The limiting distributions of Gibbs samplers exploring the state spaces in Figures 4 and 5.

Class	Fig 4	Fig 5
1	0.0721	0.1885
2	0.0369	0
3	0.0369	0.087
4	0.0751	0
5	0.0369	0.087
6	0.0751	0
7	0.0031	0.0083
8	0.0637	0
9	0.0369	0.1044
10	0.0031	0
11	0.0751	0.1888
12	0.0637	0
13	0.0751	0.1888
14	0.0637	0
15	0.0637	0.1471
16	0.2185	0

It is notable that those classes with more black facies have larger probabilities than those with more white, due to the global distribution; also, the sum of all of the probabilities is exactly equal to one (within rounding error). The joint distributions shown in Table 1 are not used in any way in the Gibbs samplers, but are the result of the conditional distributions as defined. For other conditional distributions, such as multiple-point statistics, the joint distributions would be different and would reflect the choice of conditionals.

When confronted by much larger problems more typical of resource estimation, analytical solutions are not feasible. In these cases, the Gibbs sampler may be used to approximate the global likelihoods, with or without conditioning data.

Practical Considerations

Markov chain Monte Carlo methods are powerful tools and may be applied in spatial simulation. However, there are practical considerations which arise. Honouring of conditioning data can sometimes be problematic, and artifacts may appear if the initial state for simulation is inappropriate or if too much weight is placed on the simulated values and not enough on the conditioning information. Integration of secondary data or multiple data types can complicate the conditional distributions and cause additional steps to be required. Also, in earth sciences data the simulations have distinct edges which can cause artifacting (Deutsch, 1992) and necessitate either modifications to the algorithm, expansion and trimming of the field, or ad hoc corrections.

Conclusions

Sampling from the state space of a random field that represents a large geological site can be difficult. Using a Markov chain in a Monte Carlo framework to explore the *nxyz*-dimensional space of the problem can be easier in some cases than sequentially simulating individual locations.

References

Caers, J. and Journel, A.G. (1998) Stochastic Reservoir Simulation Using Neural Networks Trained on Outcrop Data. *SPE Annual Technical Conference and Exhibition*, New Orleans, Oct. 1998, pp 321-336. SPE #49026.

Casella, G. and George, E.I. (1992) Explaining the Gibbs Sampler. *The American Statistician*, Vol. 46, No. 3, Aug. 1992, pp 167-174.

Deutsch, C.V. (1992) *Annealing Techniques Applied to Reservoir Modeling and the Integration of Geological and Engineering (Well Test) Data*. Ph.D. Thesis, Stanford University, 304 p.

Deutsch, C.V. (2002) *Geostatistical Reservoir Modeling*. Oxford University Press, New York, 376 p.

Gelfand, A.E. and Smith, A.F.M. (1990) Sampling-Based Approached to Calculating Marginal Densities. *Journal of the American Statistical Association*, Vol. 85, No. 410, June 1990, pp 398-409.

Geman, S. and Geman, D. (1984) Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, No. 6, Nov. 1984, pp 721-741.

Lawler, G.F. (2006) *Introduction to Stochastic Processes, 2nd Ed*. Chapman & Hall/CRC, Boca Raton, FL, 234 p.

Lyster, S., Deutsch, C.V., and Dose, T. (2006) A New MPS Simulation Algorithm Based on Gibbs Sampling. *Centre for Computational Geostatistics*, No. 8, 26 p.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953) Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, Vol. 21, No. 6, pp 1087-1091.

Robert, C.P. and Casella, G. (2004) *Monte Carlo Statistical Methods, 2nd Ed*. Springer Science and Business Media LLC, New York, 645 p.

Srivastava, M. (1992) Iterative Methods for Spatial Simulation. *Stanford Center for Reservoir Forecasting*, No. 5, 24 p.