Two Artifacts of Probability Field Simulation[1]

Michael J. Pyrcz[2] and Clayton V. Deutsch[3]

Probability field simulation is being used increasingly to simulate geostatistical realizations. The method can be faster than conventional simulation algorithms and it is well suited to integrate prior soft information in the form of local probability distributions. The theoretical basis of probability field simulation has been established when there are no conditioning data; however, no such basis has been established in presence of conditioning data. Realizations generated by probability field simulation show two severe artifacts near conditioning data. We document these artifacts and show theoretically why they exist. The two artifacts that have been investigated are (1) local conditioning data appear as local minima or maxima of the simulated values, and (2) the variogram model in range of conditioning data is not honored; the simulated values have significantly greater continuity than they are supposed to. These two artifacts are predicted by theory. An example flow simulation study is presented to illustrate that they affect more than the visual appearance of the simulated realizations. Notwithstanding the flexibility of the probability field simulation method, these two artifacts suggest that it be used with caution in presence of conditioning data. Future research may overcome these limitations.

INTRODUCTION

The Probability Field Simulation algorithm (PFSIM) is becoming popular due to its unique strengths. Introduced by Srivastava (1992), advantages of the method were documented further by Froidevaux (1993). Although based on a conceptually simple idea, there is no theoretical basis behind the method. Journel (1995) presented a proof that the univariate and bivariate statistics of simulated values *in the presence of no local conditioning data* are correct. The PFSIM algorithm is increasingly available in software packages such as GSLIB (Deutsch and Journel, 1997), gOcad, and other commercial software.

PFSIM separates the two key steps of geostatistical simulation, that is, data conditioning and Monte Carlo simulation. This separation of tasks leads to

---

[2]Department of Civil and Environmental Engineering, University of Alberta. e-mail: mpyrcz@ualberta.ca
[3]Department of Civil and Environmental Engineering, University of Alberta. e-mail: cdeutsch@civil.ualberta.ca

(1) conceptual simplicity through decomposition of unit operations, (2) greater CPU efficiency than conventional simulation methods such as sequential Gaussian simulation, and (3) flexibility to integrate a wide variety of secondary data through modification of the local probability distributions prior to simulation. Although PFSIM enjoys these advantages, there are two significant artifacts, which affect the visual appearance and behavior of the realizations in subsequent decision making.

The first artifact was identified in the original paper on probability field simulation by Srivastava (1992). Conditioning data are honored as local minimums or maximums. Secondly, the covariance or variogram of the simulated values deviates from the specified covariance or variogram model. In particular, the simulated values have reduced short scale variability.

This paper will (1) recall the PFSIM algorithm, (2) revisit and quantify the local extrema artifact, (3) give theoretical and experimental evidence of the covariance bias artifact, (4) present a flow simulation example to assess the impact of the artifacts, and (5) conclude with a discussion of possible corrective methods.

PROBABILITY FIELD SIMULATION

The PFSIM algorithm consists of two distinct steps: (1) establishing a set of local probability distributions, and (2) repeated Monte Carlo simulation from these distributions with correlated probability values. First, local probability distributions must be derived at each location to be simulated. These local probability distributions account for local hard conditioning data and all secondary data. The probability distribution will be a Dirac step function at the exact location of a local hard datum. Multi-Gaussian theory, indicator geostatistics, or other heurisitic calibration procedures may be used to arrive at the set of local probability distributions. The set of distributions may be denoted:

$$\{F(\mathbf{u}|(n)), \quad \forall \ \mathbf{u} \in A\} \tag{1}$$

where $F(\mathbf{u}|(n))$ are the probability distributions for all locations $\mathbf{u}$ in the area of interest A. The notation $|(n)$ implies that this set of distributions honors the n conditioning data (both hard and soft).

A wide variety of methods may be used to establish the local distributions, which is a significant advantage of PFSIM. In a multi-Gaussian setting, the local distributions are Gaussian with a mean and variance determined by simple cokriging. In an indicator setting, cumulative probability values are estimated, at fixed threshold values, by indicator kriging. For simplicity of presentation, we use Gaussian distributions; however, similar results would be obtained with probability distributions derived by other means.

The second step is to draw simulated values from the set of probability distributions (1). As explained by Srivastava, Monte Carlo simulation from each

distribution would lead to significant noise or nugget effect in the simulation results. Srivastava's key idea was to draw from the local distributions with correlated probability fields. It was conjectured that the probability fields should have the same correlation as the uniform transform of the conditioning data. By considering probability fields with a uniform [0, 1] histogram, the histogram of the simulated values has the correct level of variability. Correlation in the probability values imparts correlation to the simulated values. A particular set of correlated probabilities (l) may be denoted:

$$\{p^{(l)}(\mathbf{u}) \ \ \forall \ \ \mathbf{u} \in A\}, \quad l = 1, \ldots, L \tag{2}$$

where $p^{(l)}(\mathbf{u})$ are the probability value in the interval [0, 1] for all locations $\mathbf{u}$ in the area of interest A. The (l) superscript denotes the realization number l of the probability values; there are L realizations. These probability values have a uniform histogram and the specified uniform-score covariance structure, but need not consider any local data.

A set of simulated z values, for each l realization is created drawing from the distributions (1) with the probability values (2):

$$\{z^{(l)}(\mathbf{u}) = F^{-1}(\mathbf{u}; p^{(l)}(\mathbf{u})|(n)), \ \ \forall \ \ \mathbf{u} \in A\}, \quad l = 1, \ldots, L \tag{3}$$

The F^{-1} operation is classic Monte Carlo simulation from the specified probability distributions.

The L realizations of the z values may be used to assess joint uncertainty and make decisions. The specific nature of the uncertainty characterization and decision making depends on the context of the problem. We do not concern ourselves with these details at this time; we are concerned that the simulated values (3) are artifact free and theoretically acceptable.

Journel (1995) showed that the simulated values (3) have the correct univariate histogram and correct bivariate (spatial correlation) structure in presence of no conditioning data. Problems arise in presence of local data.

ARTIFACT ONE: LOCAL EXTREMA AT CONDITIONING DATA

Two examples were created to illustrate this artifact. In the first example, a one-dimensional (1D) space of 100 units in length was simulated. The simulation was conditioned with data with equal spacing of 10 units. The conditioning data were extracted from a nonconditional Gaussian simulation with a single spherical structure variogram with a range of fifty. Figure 1 displays three realizations of PFSIM and SGSIM, the simple kriging estimates, and the upper and lower 95th percentile values from the kriging variance. Notice how the PFSIM realizations honor the local data with a local extrema.
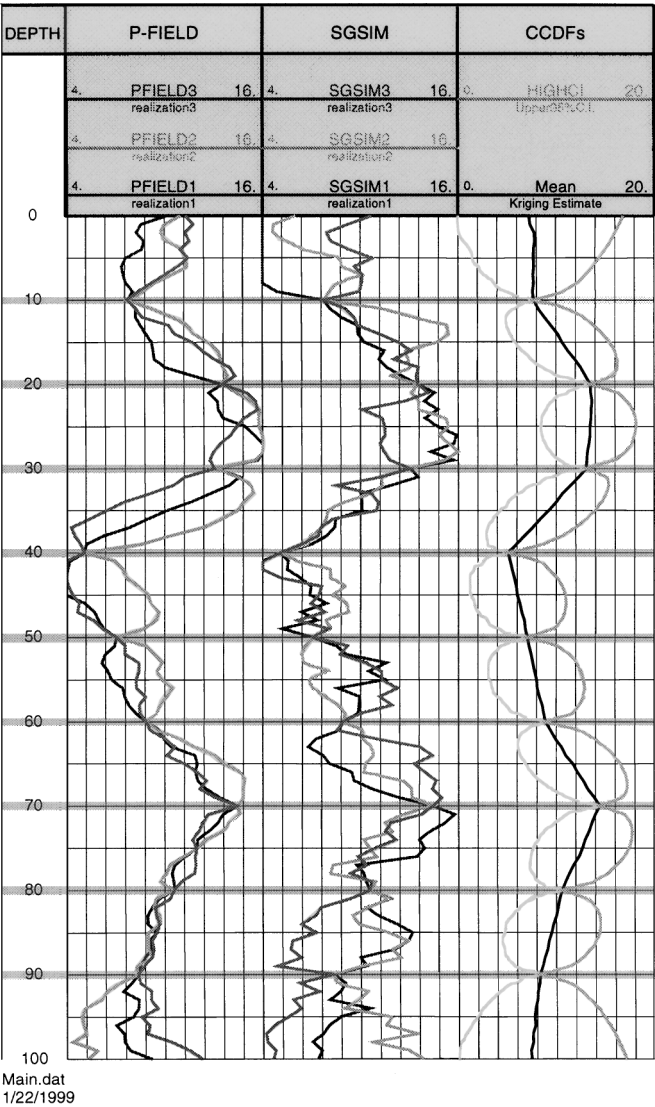
Figure 1. A 1-D example of PFSIM and SGSIM. Note the local minima and maxima artifacts of the *p*-field results.
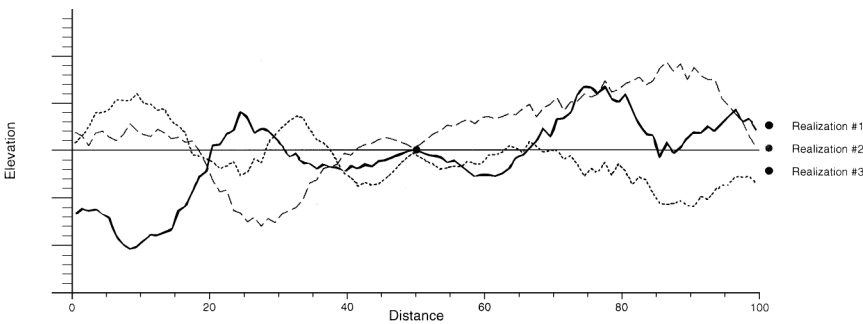
Figure 2. Three PFSIM realizations of a surface. The surfaces show the the local minima and maxima artifact.

The second example consists of simulated surfaces; see Figures 2 and 3 for cross sections through different surfaces. As seen consistently in both examples, the data are honored as local minima or maxima at the local hard data locations.

Source of Artifact

This artifact is caused by the separation of the conditioning and the Monte Carlo Simulation. As shown in Figure 4, the local probability density functions (pdfs) collapse to a Dirac step function at the local hard data locations. This reduction in pdf variance, as the estimates near the data location, causes the estimates to "pinch" toward the hard data values in the vicinity of the hard data. The artifact becomes more significant as the probabilities depart from 0.5 in the vicinity of the data location. Probability field values greater than 0.5 in the vicinity of the data cause a local minimum. Probability values less than 0.5 cause a local
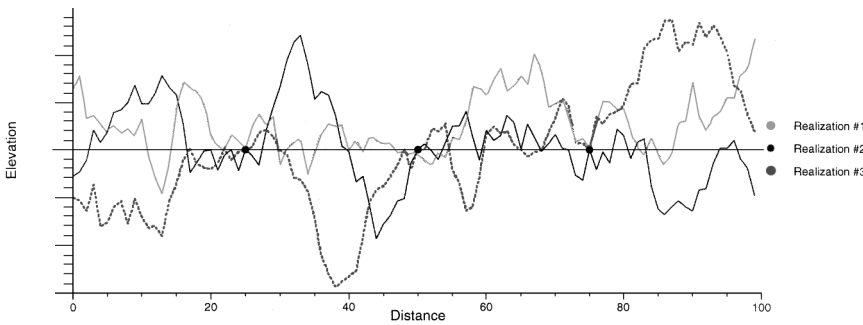

Figure 3. Three PFSIM realizations of a surface with less spatial correlation. The surfaces also show the the local minima and maxima artifact.
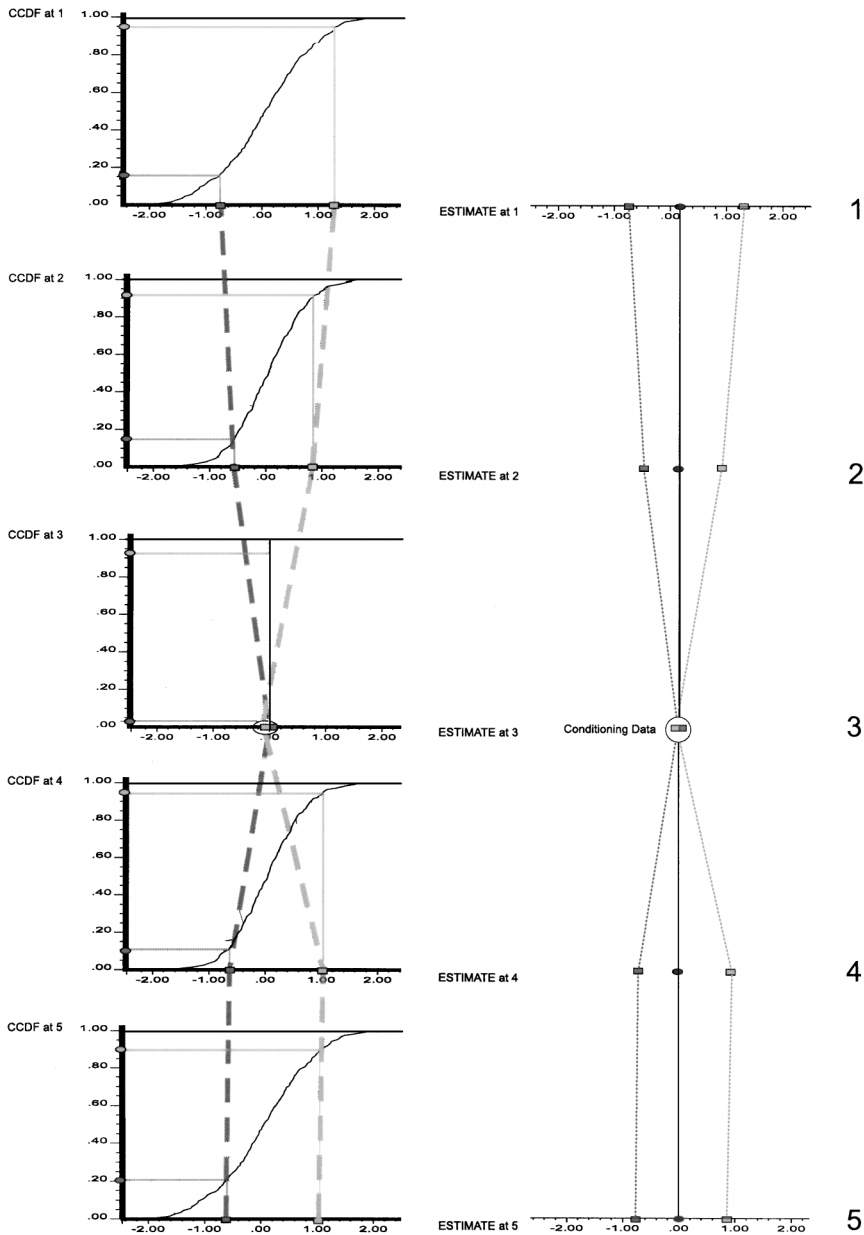
Figure 4. Two scenarios of nonconditioned correlated fields are shown. The blue line shows a local minimum caused by fields >0.5 in the vicinity of the conditioning data, while the red line shows the opposite case.

maximum. The only instance in which this artifact will not occur is when the non-conditional correlated probability value is exactly 0.5 at the data location, which is unlikely.

Multivariate Statistics

The local extrema artifact is multivariate; it can be seen visually, but not with univariate or bivariate statistics. We use multivariate statistics to quantify the artifact. A five-variate statistic is considered to measure local "texture" or variability in a one dimensional context. Five consecutive points are considered simultaneously:

$$z(\mathbf{u}+2),\, z(\mathbf{u}+1),\, z(\mathbf{u}),\, z(\mathbf{u}-1) \quad \text{and} \quad z(\mathbf{u}-2)$$

where $z(\mathbf{u})$ is the value at a data location and the values $z(\mathbf{u}+2)$, $z(\mathbf{u}+1)$, $z(\mathbf{u}-1)$ and $z(\mathbf{u}-2)$ are adjacent values.

An indicator function of two neighboring values measures whether the values are increasing or decreasing, that is,

$$i_k = \begin{cases} 1, & \text{if } z(\mathbf{u}+k) \leq z(\mathbf{u}+k-1) \\ 0, & \text{otherwise} \end{cases}$$

There are four indicator values for five adjacent z values. There are 16 different configurations of indicators. Each configuration is a particular pattern of increasing and/or decreasing values. The 16 configurations are assigned an index:

$$\text{index} = 1 + \sum_{j=1}^{4} i(j) \cdot 2^{(j-1)}$$

Figure 5 shows the 16 configurations.

For illustration, a 100-value one-dimensional example was constructed with data points at locations 10, 50, and 90. A single spherical structure with a range of 50 and no nugget effect is considered. Five thousand PFSIM and SGSIM realizations were generated and subsequently classified by the index above. Becuase all conditioning data have the same values, each index should be equally probable. SGSIM correctly demonstrates this, but the PFSIM realizations clearly show a bias toward the two indexes that represent the data point at 50 being a local minima or maxima (index 4 and 13, Figs. 6 and 7). Other examples show similar results.

The critical observation is that PFSIM creates unwarranted spatial structures in the form of discontinuities near conditioning data. A primary criteria of a
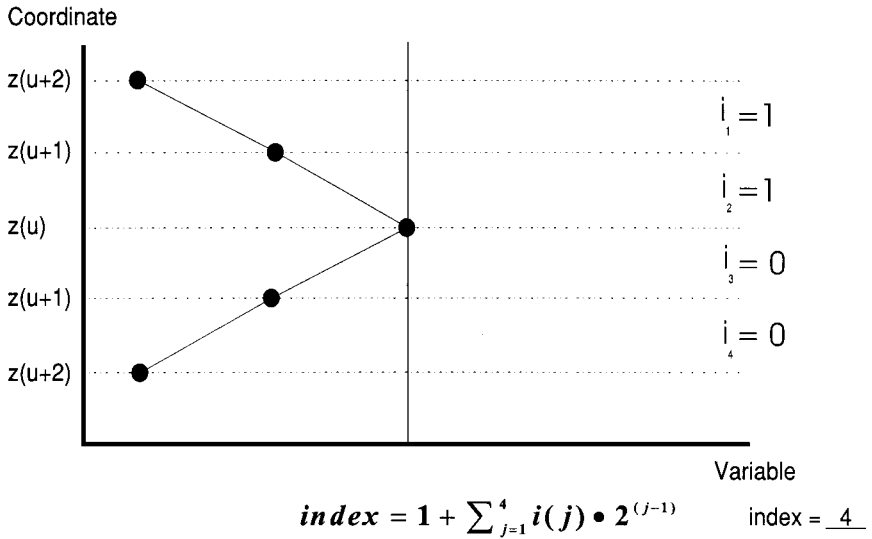
Coordinate

$$index = 1 + \sum_{j=1}^{4} i(j) \bullet 2^{(j-1)} \qquad \text{index} = \underline{\;4\;}$$

Figure 5. The multivariate index is based on the indicator transform of 1 if $z(\mathbf{u} + k)$ is greater than or equal to $z(\mathbf{u} + k - 1)$, and zero otherwise.

simulation algorithm is that realizations be free of unwarranted artifacts. No specific conclusions can be drawn since the impact of this artifact will depend on the transfer of uncertainty characterization to decision making. In general, we expect PFSIM to incorrectly produce "dimples" or "peaks" at the data locations. Practitioners must be aware of this and proceed judiciously.

ARTIFACT TWO: COVARIANCE BIAS NEAR CONDITIONING DATA

The PFSIM realizations of Figure 1 appear more continuous than the SGSIM realizations. This is not a bug. After investigation we found the variogram to be reproduced correctly with no conditioning data, but there is increased continuity in presence of conditioning data. This is expected from theory. Experimental results confirm theoretical prediction. Difficulties in the reproduction of the correct bivariate structure have been observed by Froidevaux, 1993 who stated that more work was required *to put the inference of the probability field covariance on firmer theoretical ground.*

Theoretical Background for Covariance Artifact

The theoretical background is most easily demonstrated in a multi-Gaussian setting; the distributions of uncertainty are Gaussian and defined by the simple
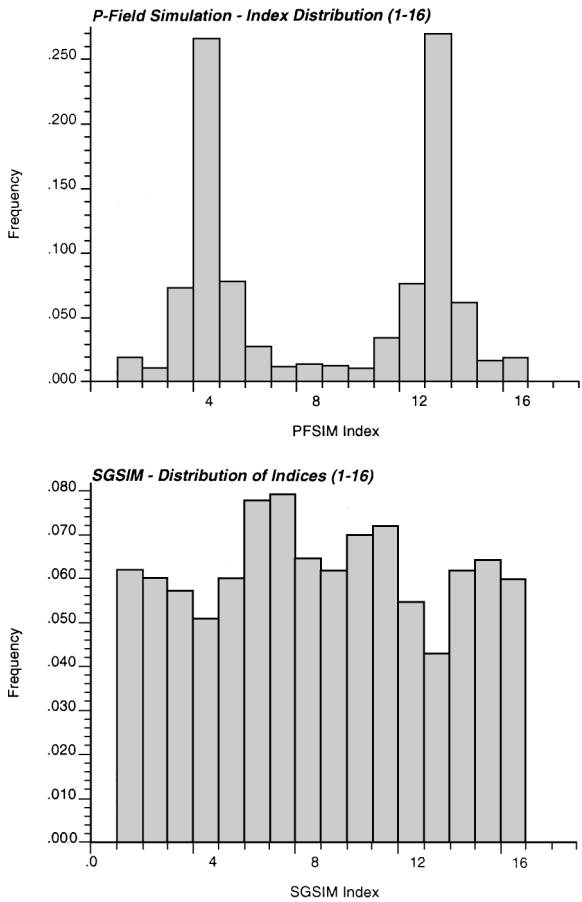
P-Field Simulation - Index Distribution (1-16)

SGSIM - Distribution of Indices (1-16)

Figure 6. The distributions of minimum/maximum indexes for PFSIM and SGSIM

cokriging means $m(\mathbf{u}), \mathbf{u} \in A$ and standard deviations $\sigma(\mathbf{u}), \mathbf{u} \in A$. As in most geostatistical theory, we work with the covariance and visualize the variogram.

First, let's consider the unconditional covariance. By definition the covariance between two values y and y' is:

$$C(\mathbf{h}) = E\{y \cdot y'\} - E\{y\} \cdot E\{y'\} \tag{4}$$

where, in p-field simulation, y and y' are drawn from Gaussian distributions with uniform, nonconditioned correlated fields, p and p'. The covariance (4) may be

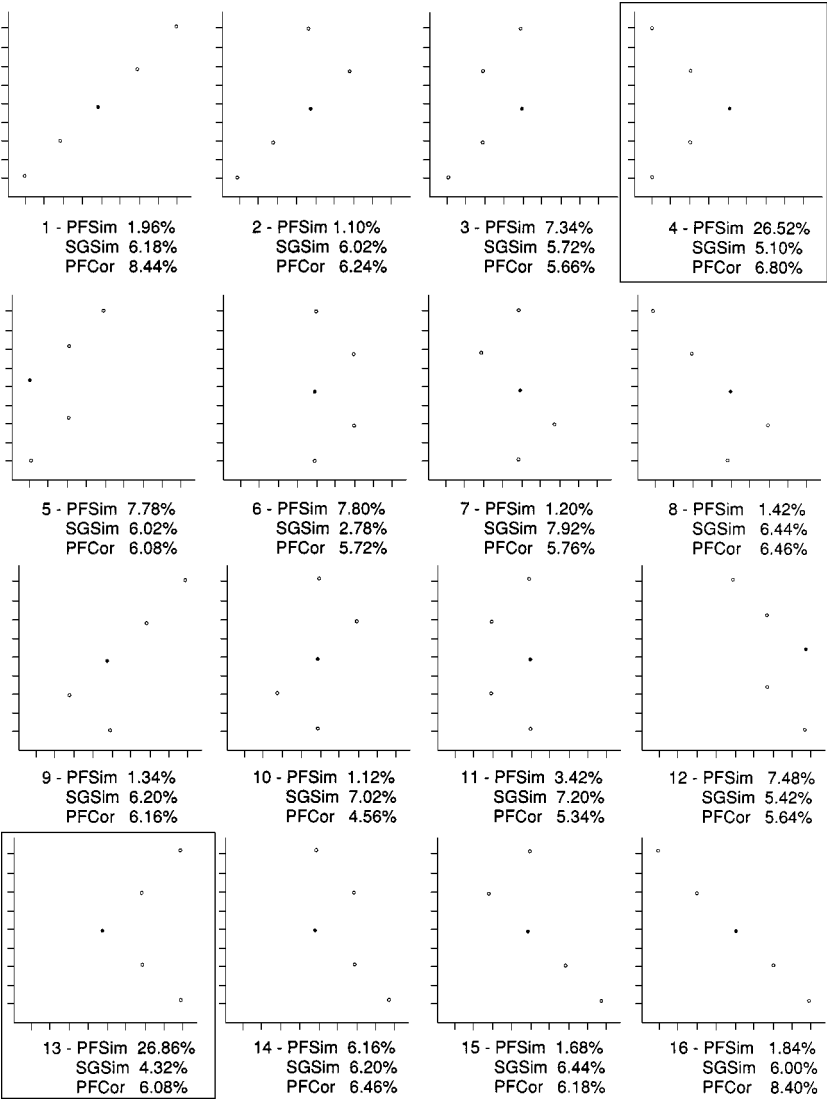| 1 - PFSim 1.96%
SGSim 6.18%
PFCor 8.44% | 2 - PFSim 1.10%
SGSim 6.02%
PFCor 6.24% | 3 - PFSim 7.34%
SGSim 5.72%
PFCor 5.66% | 4 - PFSim 26.52%
SGSim 5.10%
PFCor 6.80% |
5 - PFSim 7.78% SGSim 6.02% PFCor 6.08%	6 - PFSim 7.80% SGSim 2.78% PFCor 5.72%	7 - PFSim 1.20% SGSim 7.92% PFCor 5.76%	8 - PFSim 1.42% SGSim 6.44% PFCor 6.46%
9 - PFSim 1.34% SGSim 6.20% PFCor 6.16%	10 - PFSim 1.12% SGSim 7.02% PFCor 4.56%	11 - PFSim 3.42% SGSim 7.20% PFCor 5.34%	12 - PFSim 7.48% SGSim 5.42% PFCor 5.64%
13 - PFSim 26.86% SGSim 4.32% PFCor 6.08%	14 - PFSim 6.16% SGSim 6.20% PFCor 6.46%	15 - PFSim 1.68% SGSim 6.44% PFCor 6.18%	16 - PFSim 1.84% SGSim 6.00% PFCor 8.40%

Figure 7. Illustration of the 16 index categories and their accompanying probabilities. PFcor is PFSIM with the probabilities conditioned to 0.5 at the data, as explained in the Discussion Section.

written as

$$C(\mathbf{h}) = E\{G^{-1}(p) \cdot G^{-1}(p')\} - m(\mathbf{u}) \cdot m(\mathbf{u}')$$

In the presence of no conditioning,

$$m(\mathbf{u}) = m(\mathbf{u}') = 0$$

so, the covariance is

$$C(\mathbf{h}) = E\{G^{-1}(p) \cdot G^{-1}(p')\}$$

where the values p and p' are separated by $\mathbf{h}$ with the "correct" uniform scores covariance. The covariance $C(\mathbf{h})$ is difficult to simplify, because the inverse of the Gaussian CDF, G^{-1}, is not known analytically. Nevertheless, we know that this covariance is correct because it can be shown with appropriate conditions of ergodicity and deconditioning, that the PFSIM results are acceptable (Journel, 1995).

This ability to correctly reproduce the model covariance in the absence of conditioning data is confirmed in the following example. A two dimensional space is simulated (100×100 units) with no conditioning data. The covariance is a single spherical structure with range varying from 10 to 20. The resulting omnidirectional variograms for both PFSIM and SGSIM honor the model in expected value (see Fig. 8).

Now, let us consider the case with conditioning data. With conditioning, the distributions are no longer standard normal. We consider a pair at locations $\mathbf{u}$ and $\mathbf{u}'$, within range of conditioning data. The results will then be generalized by considering that pair translated over the field A. Near conditioning data, y and y' are defined as

$$y = m(\mathbf{u}) + \sigma(\mathbf{u}) \cdot G^{-1}(p)$$
$$y' = m(\mathbf{u}') + \sigma(\mathbf{u}') \cdot G^{-1}(p')$$

Solving for the covariance $[C^*(\mathbf{h})]$ in the presence of such conditioning,

$$
\begin{aligned}
C^*(\mathbf{h}) =\ & E\{[m(\mathbf{u}) + \sigma(\mathbf{u}) \cdot G^{-1}(p)][m(\mathbf{u}') + \sigma(\mathbf{u}') \cdot G^{-1}(p')]\} - E\{y\} \cdot E\{y'\} \\
=\ & E\{m(\mathbf{u}) \cdot m(\mathbf{u}') + m(\mathbf{u}') \cdot \sigma(\mathbf{u}) \cdot G^{-1}(p) + m(\mathbf{u}) \cdot \sigma(\mathbf{u}') \cdot G^{-1} \\
& + \sigma(\mathbf{u}) \cdot \sigma(\mathbf{u}') \cdot G^{-1}(p) \cdot G^{-1}(p')\} - E\{y\} \cdot E\{y'\} \\
=\ & E\{m(\mathbf{u}) \cdot m(\mathbf{u}')\} + m(\mathbf{u}') \cdot \sigma(\mathbf{u}) \cdot \underbrace{E\{G^{-1}(p)\}}_{0} \\
& + m(\mathbf{u}) \cdot \sigma(\mathbf{u}') \cdot \underbrace{E\{G^{-1}(p)\}}_{0} \\
& + E\{\sigma(\mathbf{u}) \cdot \sigma(\mathbf{u}') \cdot G^{-1}(p) \cdot G^{-1}(p')\} - E\{y\} \cdot E\{y'\} \\
=\ & \sigma(\mathbf{u}) \cdot \sigma(\mathbf{u}') \cdot \underbrace{E\{G^{-1}(p) \cdot G^{-1}(p')\}}_{C(\mathbf{h})} + E\{m(\mathbf{u}) \cdot m(\mathbf{u}')\} - \underbrace{E\{y\} \cdot E\{y'\}}_{0}
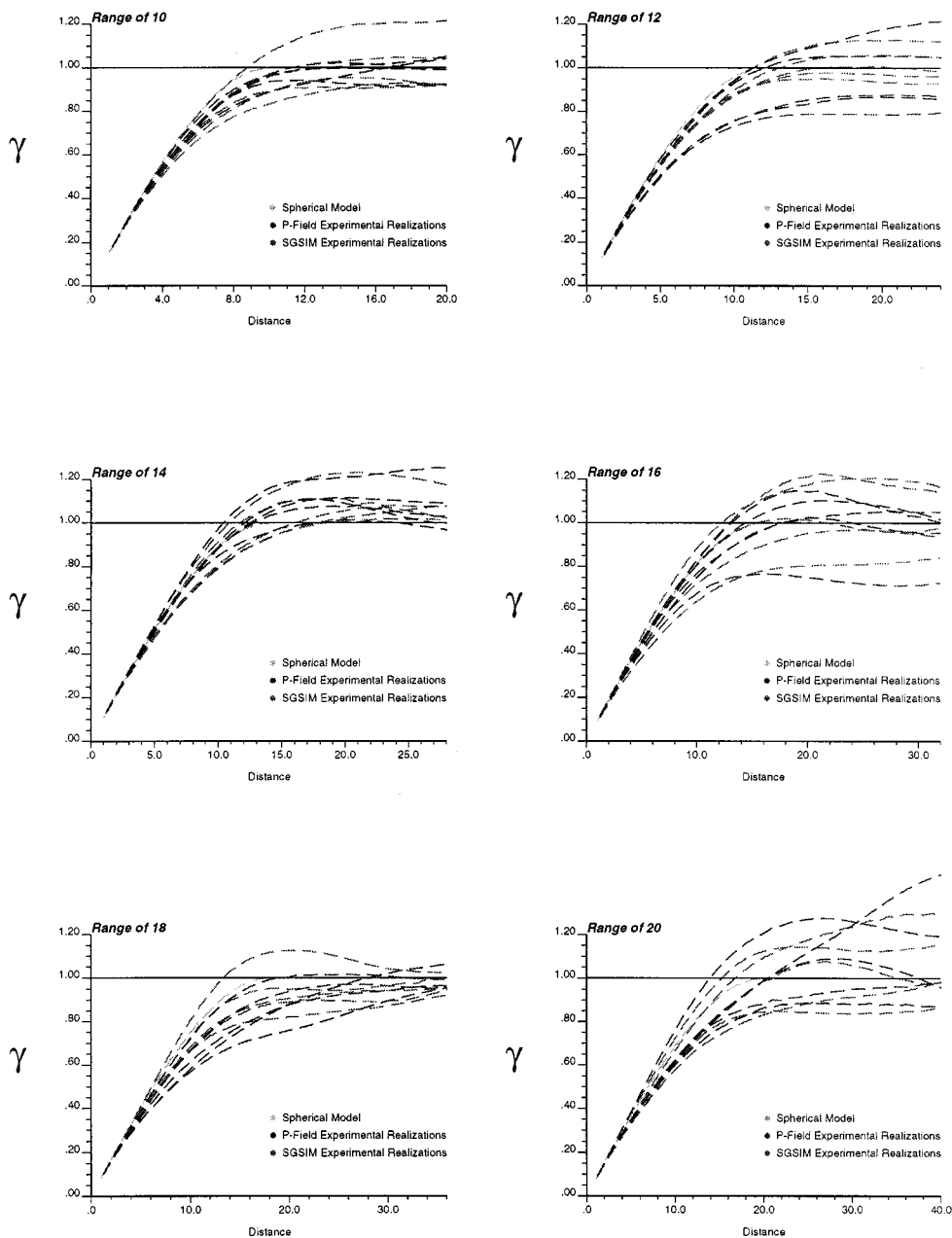\end{aligned}
$$

Figure 8. A comparison of SGSIM and PFSIM, with various ranges of correlation. All simulations are nonconditional. Note that there is no distinguishable difference in the results in the absence of conditioning data.

since $E\{G^{-1}(p)\} = E\{G^{-1}(p')\} = 0$. In addition, the global distribution of the regionalized variable is standard normal; therefore $E\{y\}$ and $E\{y'\}$, which are equivalent to $E\{m(\mathbf{u})\}$ and $E\{m(\mathbf{u}')\}$, are equal to 0, in expected value over the area A.

Therefore the biased covariance $C^*(\mathbf{h})$, between $\mathbf{u}$ and $\mathbf{u}'$, separated by vector $\mathbf{h}$, in the presence of conditioning is

$$C^*(\mathbf{h}) = \underbrace{\sigma(\mathbf{u}) \cdot \sigma(\mathbf{u}')}_{\text{Factor1}} \cdot C(\mathbf{h}) + \underbrace{E\{m(\mathbf{u}) \cdot m(\mathbf{u}')\}}_{\text{Factor2}} \tag{5}$$

The correct covariance $C(\mathbf{h})$ was derived above in presence of no conditioning. There are two factors that lead the covariance in presence of conditioning data to be different from the correct covariance (1) a term that accounts for the reduced variance of the prior distributions, and (2) a term that accounts for correlation in the means of the prior distributions.

We can further simplify this if we consider the limit case at short lag distances, $\mathbf{h} \to 0$. In such case, the conditioning data may be assumed to be the same for both locations and the following approximation can be made for Factor 2:

$$E\{m(\mathbf{u}) \cdot m(\mathbf{u}')\} = E\left\{\sum_{\alpha=1}^{n} \lambda_\alpha y_\alpha \cdot \sum_{\alpha=1}^{n} \lambda'_{\alpha=1} y_\alpha\right\}$$

$$E\left\{\sum_{\alpha=1}^{n} \lambda_\alpha y_\alpha \cdot \sum_{\alpha=1}^{n} \lambda'_\alpha y_\alpha\right\} = E\left\{\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \lambda_\alpha \cdot \lambda'_\beta \cdot y_\alpha \cdot y_\beta\right\}$$

$$= \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \lambda_\alpha \cdot \lambda'_\beta \cdot \text{Cov}(\mathbf{u}_\alpha, \mathbf{u}'_\beta)$$

$$= \sum_{\alpha=1}^{n} \lambda_\alpha \left[\sum_{\beta=1}^{n} \lambda'_\beta \cdot \text{Cov}(\mathbf{u}_\alpha, \mathbf{u}'_\beta)\right]$$

From the kriging equations above

$$\sum_{\beta=1}^{n} \lambda'_\beta \text{Cov}(\mathbf{u}_\alpha, \mathbf{u}'_\beta) = \text{Cov}(\mathbf{u}_\alpha, \mathbf{u}'_\beta)$$

$$= \sum_{\alpha=1}^{n} \lambda_\alpha \text{Cov}(\mathbf{u}', \mathbf{u}_\alpha)$$

$$= \bar{c}(\mathbf{u}, \mathbf{u}) - \sigma_k^2$$

$$= \sigma^2 - \sigma_k^2$$

which leads to the following approximation that may be used to calculate the adjusted covariance.

$$C^*(h) = \sigma_k^2 \cdot C(h) + \sigma^2 - \sigma_k^2$$
$$= \sigma^2 + \sigma_k^2[C(h) - 1] \tag{6}$$

Inspection of this approximation [Eq. (6)], shows that when the kriging variance is zero, or in face of perfect correlation, the adjusted correlation is equal to the variance, which is the sill. This is expected, since the correlation at a lag distance of 0 must be equal to the sill in order for the variogram to be equal to 0.

Also, the covariance in the presence of conditioning becomes with the covariance $C^*(h) = C(h)$, as the kriging variance approaches 1, or as the nonconditioned case is approached [Eq. (6)]. This agrees with theoretical and experimental findings, that is, the PFSIM bivariate statistics are correct in the nonconditioned case.

Returning to the general equation for adjusted covariance [Eq. (5)]:

$$C^*(\mathbf{h}) = \underbrace{\sigma(\mathbf{u}) \cdot \sigma(\mathbf{u}')}_{\text{Factor1}} \cdot C(\mathbf{h}) + \underbrace{E\{m(\mathbf{u}) \cdot m(\mathbf{u}')\}}_{\text{Factor2}}$$

The first factor, the cross variance [$\sigma(\mathbf{u}) \cdot \sigma(\mathbf{u}')$) of the sample space, is the product of the kriging standard deviation of the $\mathbf{u}$ and $\mathbf{u}'$ separated by $\mathbf{h}$. When calculating variograms, the covariance is calculated in expected value over the area of interest. Hence, the following form may be used to predict the PFSIM experimental variogram:

$$E\{C^*(\mathbf{h})\} = E\{\underbrace{\sigma(\mathbf{u}) \cdot \sigma(\mathbf{u}')}_{\text{Factor1}}\} \cdot E\{C(\mathbf{h})\} + \underbrace{E\{m(\mathbf{u}) \cdot m(\mathbf{u}')\}}_{\text{Factor2}}$$

$$= E\{\underbrace{\sigma(\mathbf{u}) \cdot \sigma(\mathbf{u}')}_{\text{Factor1}}\} \cdot C(\mathbf{h}) + \underbrace{E\{m(\mathbf{u}) \cdot m(\mathbf{u}')\}}_{\text{Factor2}} \ \forall \ \mathbf{u} \in A \tag{7}$$

As shown on the left in Figure 9, the expected cross "variance" over a space is approximately constant in a given conditioning setting and therefore independent of the lag vector.

The second factor, the expected cross mean[2] over the area, A, ($E\{(m(\mathbf{u}) \cdot m(\mathbf{u}'))\}$) is dependent on the lag vector $\mathbf{h}$. This is demonstrated on the right of Figure 9. The product of $m(\mathbf{u})$ and $m(\mathbf{u}')$ has a greater probability of being
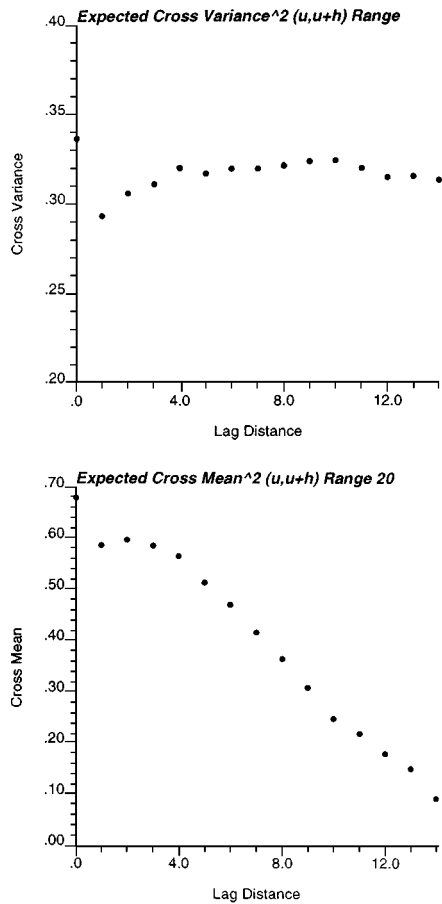
Figure 9. A plot of the cross variance and cross mean up to a lag distance of 14.

positive as the correlation increases between $\mathbf{u}$ and $\mathbf{u}'$. For example, if $m(\mathbf{u})$ is negative, and $m(\mathbf{u}')$ is well correlated, then $m(\mathbf{u}')$ will also likely be negative. This results in the cross mean^2 expressing a maximum at $\mathbf{h} = 0$, and decreasing to 0 at the range of correlation and remaining at 0 past the range of correlation.

When the lag distance is equal to 0, the formula (7) simplifies to

$$C^*(0) = \text{sill} = 1 = (\sigma(\mathbf{u}))^2 + E\{m(\mathbf{u})^2\}$$

Where $E\{m(\mathbf{u})^2\}$ is the cross mean^2 at $\mathbf{h} = 0$. This provides a method for
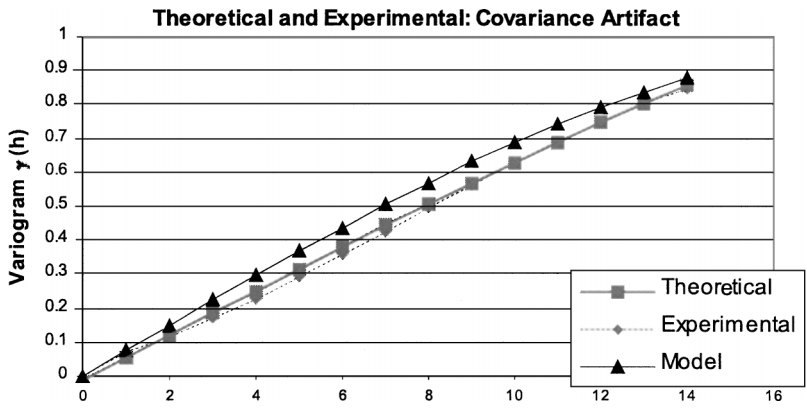
Figure 10. A comparison of the experimental PFSIM variogram and the theoretical variogram calculated with Equation (2).

approximating the constant cross variance term as

$$\overline{\sigma_k^2(\mathbf{u})} \approx (\sigma(\mathbf{u}))^2 = 1 - E\{m(\mathbf{u})^2\} \tag{8}$$

where $\overline{\sigma_k^2(\mathbf{u})}$ is the average kriging variance.

To demonstrate the use of the adjusted covariance formulas, two examples are presented. The first example uses the general equation in expected value [Eq. (7)], and considers a range of 20, from the previously mentioned 101×101 data set. The cross mean[2] and the cross variance were calculated for $\mathbf{h} = 1, \ldots, 14$ and the general formula was used to calculate the biased variogram $(1 - C^*(h))$, over the area of concern. The plot of the model, PFSIM experimental and biased or theoretical variogram are shown in Figure 10. This confirms that Equation (5) accurately predicts the experimental covariance, of PFSIM estimates.

The second example uses the short lag approximation [Equation (8)] to calculate the adjusted covariance between two close points, which have a model covariance of 0.8. Since the two points are in close proximity, the kriging variances at each location are assumed to be equal. The following table summarizes the results.

σ_k^2	$C(h)$	$C^*(h)$
1	0.8	0.8
0.8	0.8	0.84
0.6	0.8	0.88
0.4	0.8	0.92

As the amount of conditioning increases (σ_k^2 decreases) the bias in the covariance increases—that is, it departs significantly from the model covariance of 0.8. As confirmed by observation, the biased covariance is greater (variogram is lesser) than the specified model. Empirical results will now be shown that confirm the theoretical bias derived above.

Experimental Results

To illustrate the biased covariance artifact, a 101×101 model was simulated with a 10×10 evenly spaced conditioning data grid, using PFSIM and SGSIM. The simulation was based on a spherical variogram with an isotropic range. To illustrate the difference between PFSIM, SGSIM and the specified model, omni directional variograms were calculated for five realizations of each algorithm with ranges 10, 12, . . . , 20 (Fig. 11). A characteristic reduction in short scale variability is manifested in all conditioned PFSIM realizations. This divergence from the input random function (RF) model appears to increase as the scale of conditioning increases (as the range of the RF increases, or as the conditioning data density increases). The SGSIM empirical variograms reasonably approximate the RF model.

Additional experimental runs may be made to quantify this bias. As predicted by theory, the results deviate more significantly as the conditioning data increase. The bias is maximum at the short lags and decreases to 0 at the range of correlation.

Consequences

PFSIM realizations have been shown to systematically misrepresent the variogram (bivariate statistics). Once again, the impact of this problem is application dependent. The theoretical and empirical results empower the practitioner to understand and predict the magnitude of this artifact. These results reinforce the previous recommendation that caution and judgment are essential in applying PFSIM.

Some implications of this artifact can be predicted. In petroleum applications, this smoothness will lead to different flow behavior near wells, which is precisely where the flow predictions are of critical importance. In petroleum and hydrogeology applications, flow predictions are extremely sensitive to connectivity of extreme permeability or hydraulic conductivity; therefore, we anticipate significant consequences of this artifact. In mining applications, the smoothing of short-scale variability will lead to optimistic assessment of mine selectivity. The
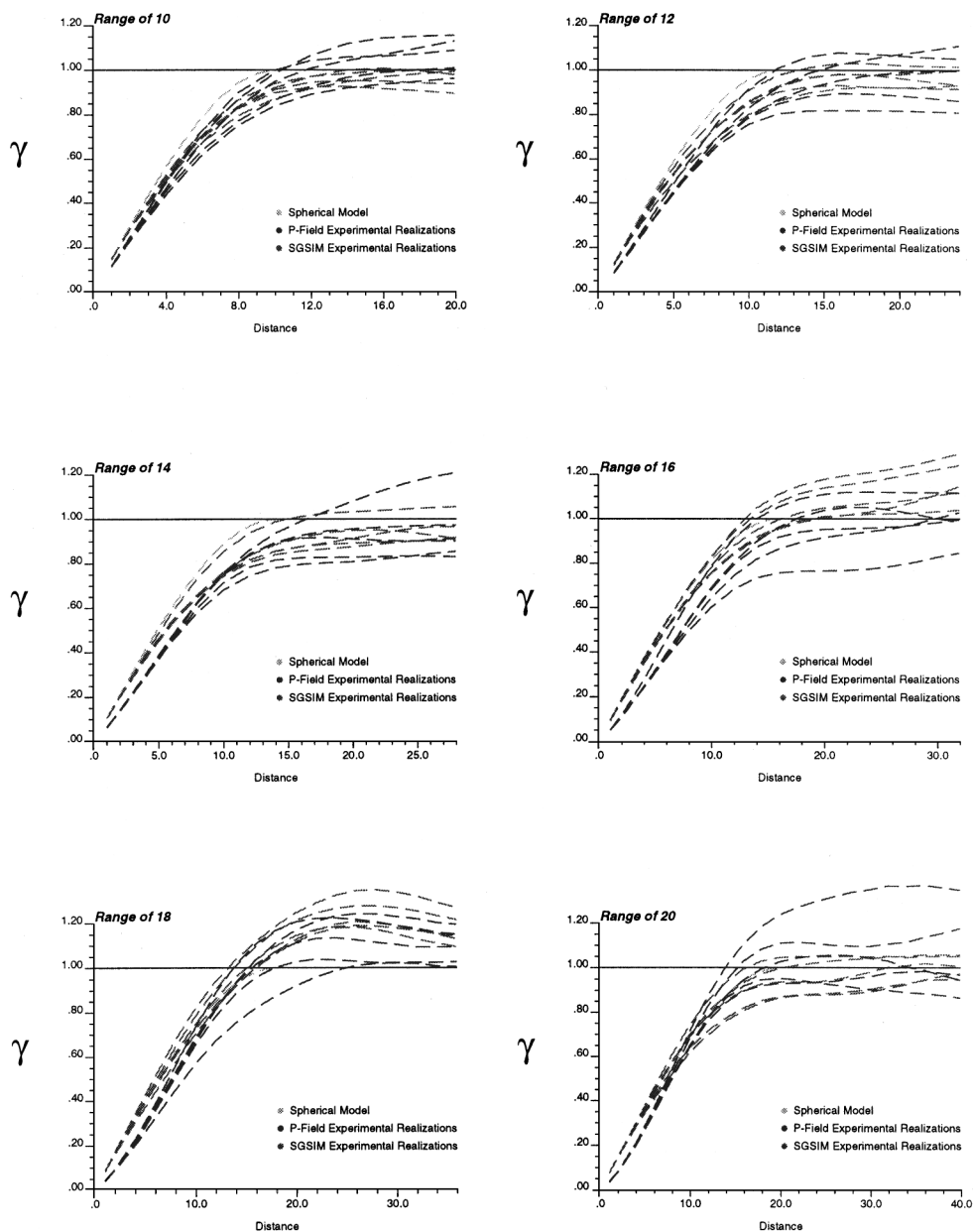
Figure 11. A comparison of experimental variograms from SGSIM and P-field simulation, with various ranges of correlation. Compare these results to the unconditional case in Figure 8. In this case there is a significant departure of the *p*-field variograms from theory.
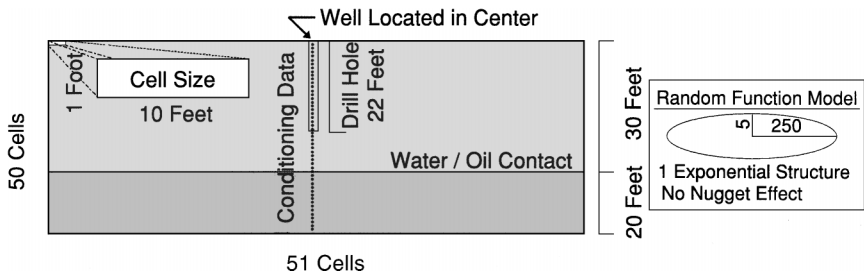
Figure 12. The layout used for the flow simulations.

drop in short-scale continuity may cause an apparent decrease in the dispersion variance of selective mining units near conditioning data, which would lead to a model.
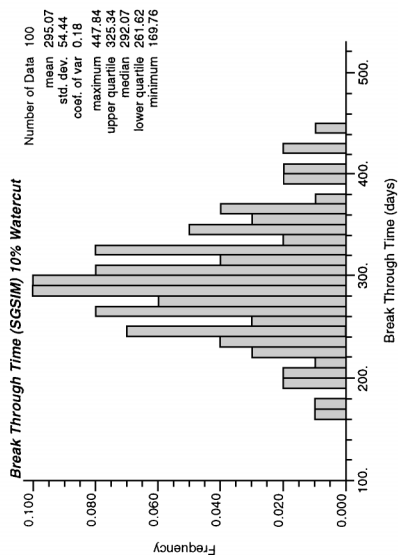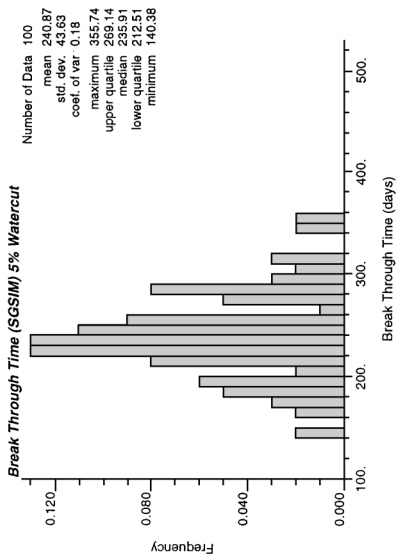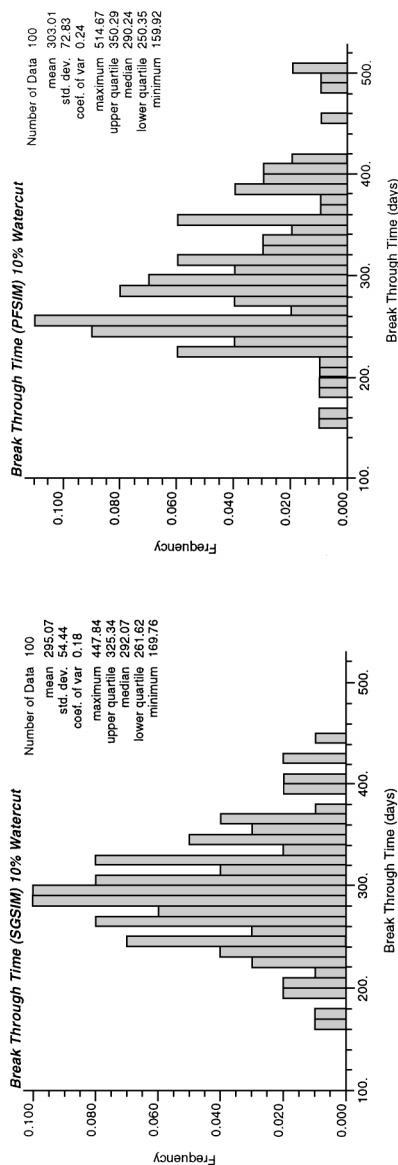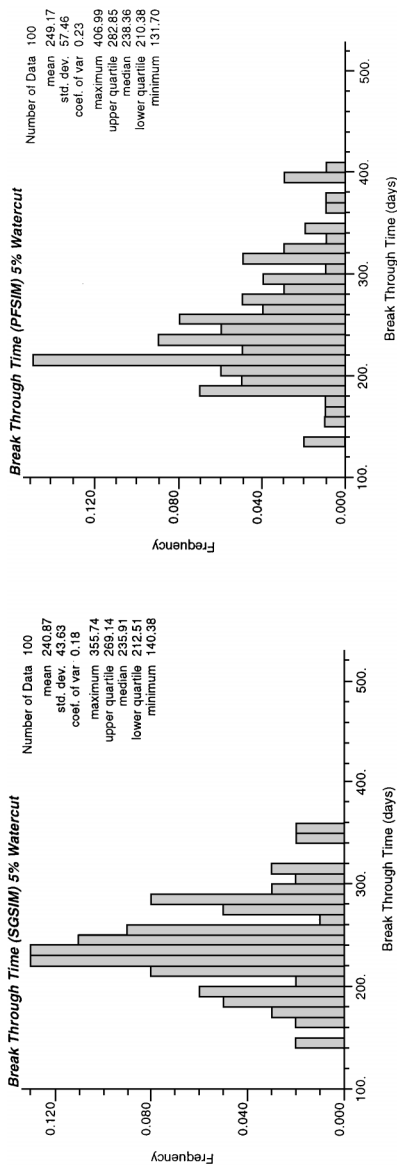
FLOW SIMULATION EXAMPLE

To investigate the effect of the PFSIM artifacts on a transfer function, multiple flow simulations were run with 100 equal probable models derived from SGSIM and PFSIM. A bottom-water drive study was performed using Eclipse100 reservoir simulator. A two dimensional cross-sectional area of 51×50 cells, with 10 by 1 ft cell dimensions was used (Fig. 12).

A porosity of 0.15 was assumed in all locations except for the bottom blocks where the porosity was increased to simulate a bottom-water drive. The formation volume factor was set at 1 res vol/standard volume, the compressibility to $1.0 \times 10^{-5}(1/psia)$, the viscosity to 0.56 cp. The well was produced with a total liquid rate of 10 barrels/day, and the total simulation time was 9.26 years (3385 days).

Two hundred flow simulations were performed. The water breakthrough time was taken as a response variable. Water cutoffs of 5, 10, and 50% were considered (Fig. 13). The watercut variances with respect to time are summarized in Figure 14. There is a divergence between the two algorithms. Consistently, PFSIM provides biased uncertainty over all times.

POSSIBLE CORRECTION SCHEMES

Since the PFSIM algorithm has been shown to fill a valuable niche among geostatistical tools, it would be well worth the effort to remediate the aforementioned artifacts while preserving PFSIM's unique strengths.
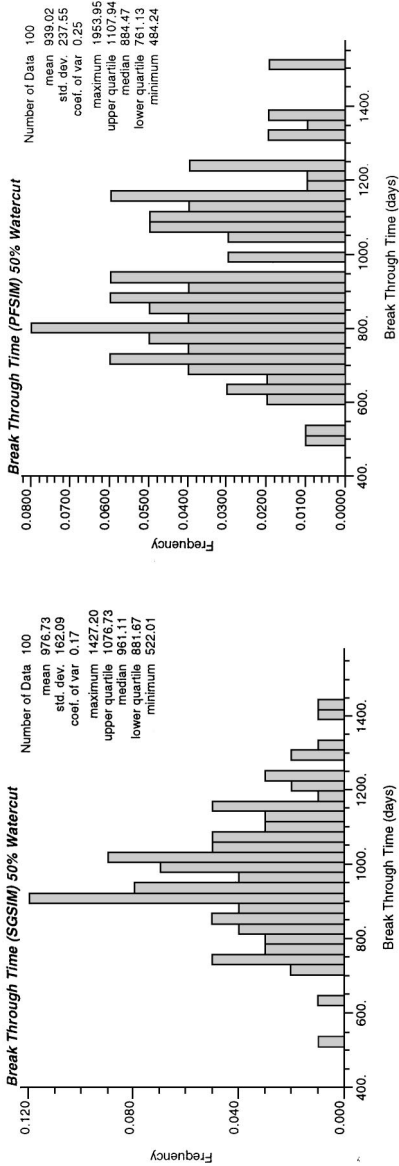
Break Through Time (PFSIM) 5% Watercut

Number of Data 100
mean 249.17
std. dev. 57.46
coef. of var 0.23
maximum 406.99
upper quartile 282.85
median 238.36
lower quartile 210.38
minimum 131.70

Break Through Time (PFSIM) 10% Watercut

Number of Data 100
mean 303.01
std. dev. 72.83
coef. of var 0.24
maximum 514.67
upper quartile 350.29
median 290.24
lower quartile 250.35
minimum 159.92

Break Through Time (SGSIM) 5% Watercut

Number of Data 100
mean 240.87
std. dev. 43.63
coef. of var 0.18
maximum 355.74
upper quartile 269.14
median 235.91
lower quartile 212.51
minimum 140.38

Break Through Time (SGSIM) 10% Watercut

Number of Data 100
mean 295.07
std. dev. 54.44
coef. of var 0.18
maximum 447.84
upper quartile 325.34
median 292.07
lower quartile 261.62
minimum 169.76

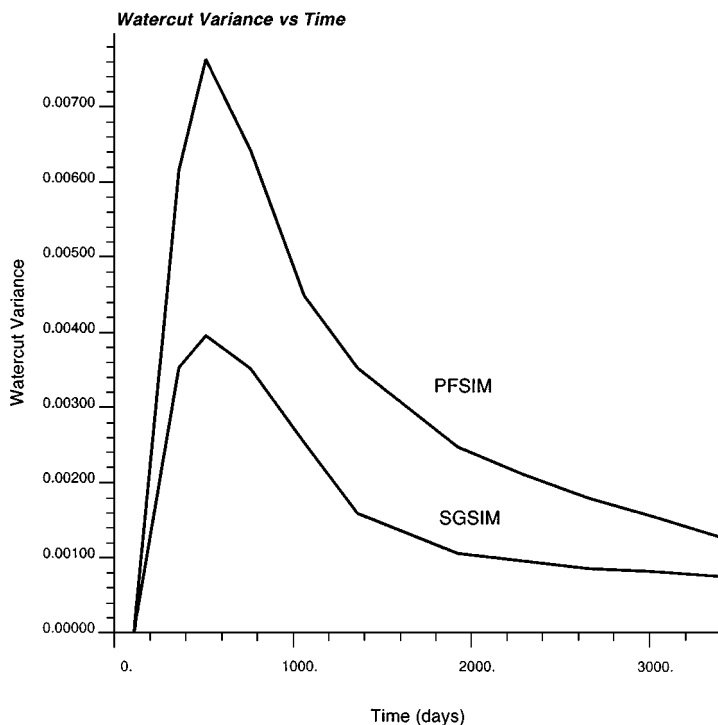**Figure 13.** The distributions of uncertainty of break through time at 5, 10, and 50% watercut.

Figure 14. The watercut variance vs. time to break through. Note the significant difference between *p*-field simulation and sequential Gaussian simulation.

An obvious solution to the local extrema artifact is to condition the correlated fields to a probability value of 0.5 at all data locations. This has been implemented. The histogram of the 16 multivariate indexes with conditioned fields (Fig. 15), is much closer to the expected uniform distribution.

The shortcoming of this correction is a significant reduction in uncertainty of the realizations. The distributions of uncertainty have less variability within range of the conditioning probabilities, that is, the distributions of probability values will not be uniform between 0 and 1. As shown in Figure 16 the pointwise and global variance are too low.

Another option for removing the extrema artifact and the covariance bias artifact would be to work with multiple point statistics. PFSIM realizations could be postprocessed with an algorithm such as simulated annealing. A multiple point statistic and a variogram (bivariate) components could be part of the objective function to be minimized. Postprocessing with iterative techniques would
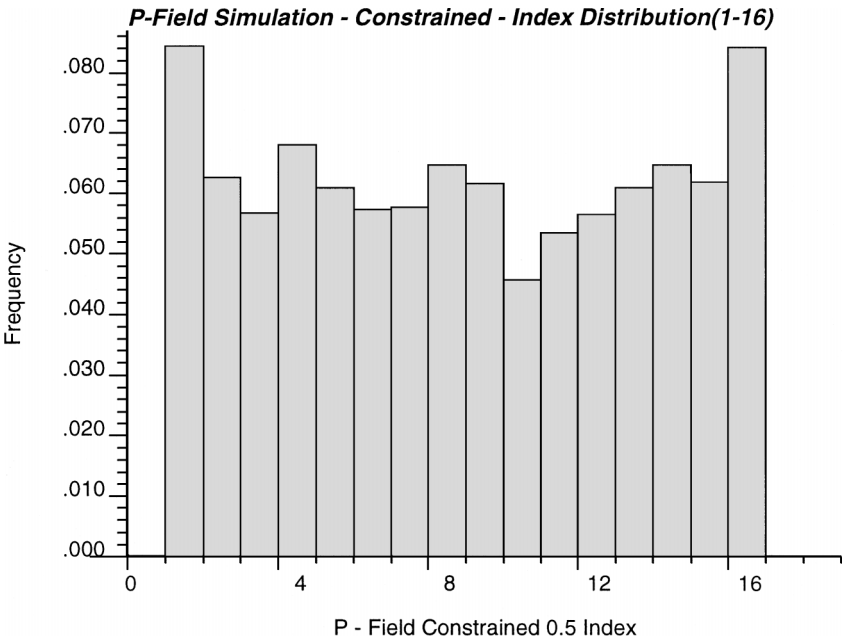
P-Field Simulation - Constrained - Index Distribution(1-16)

Figure 15. The distribution of the previously defined 16 indexes from constrained PFSIM. The distribution is uniform as it should be.

remove the advantage of speed and simplicity which are central to PFSIM's attractiveness.

CONCLUSION

The advantages of *p*-field simulation are (1) great efficiency when building multiple realizations, (2) conceptual simplicity through decomposition of steps, and (3) flexibility to integrate a wide variety of secondary data through modification of the local probability distributions prior to simulation. The disadvantages of *p*-field simulation are (1) local conditioning data commonly appear as local minima or maxima of the simulated values, and (2) the variogram model in range of the conditioning data is not honored; the simulated values have significantly greater continuity than they are supposed to.

These two flaws are inherent in the algorithm and we recommend that *p*-field simulation not be used.
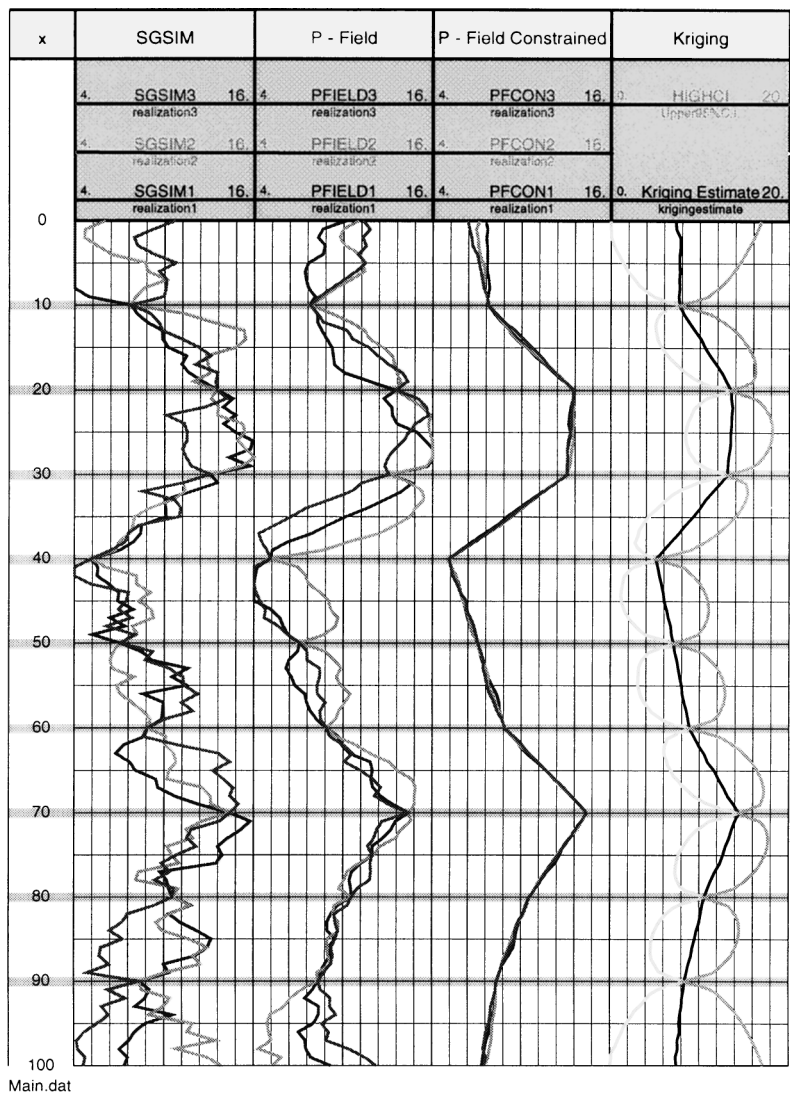
Figure 16. The previous 1D example (Fig. 1), with three realizations of constrained PFSIM added. Note the unrealistic smoothness of the constrained PFSIM results.

REFERENCES

Deutsch, C. V., Journel, A. G., 1997, GSLIB: Geostatistical software library and user's guide, Second Edition: Oxford University Press, New York, 369 p.

Froidevaux, R., 1993, Probability field simulation, *in* Soares, A., Ed., Geostatistics Troia 1992, Volume 1: Kluwer, New York, p. 73–84.

Journel, A. G., 1995, Probability fields: Another look and a proof: Unpublished report of Stanford Center for Reservoir Forecasting, Stanford, CA, 10 p.

Srivastava, R. M., 1992, Reservoir characterization with probability field simulation: SPE Formation Evaluation, v. 7, no. 4, p. 927–937.