

Sequential Approach to Covariance Correction for P-Field Simulation

Chad Neufeld and Clayton V. Deutsch

One well known artifact of the probability field (p-field) simulation algorithm is a too large covariance near conditioning data. Previously, we derived a theoretical correction to the covariance bias. The correction was implemented via an LU-based simulation and was limited to small grids. A mixture model local correction for the p-field covariance bias is proposed here. The mixture model correction can be used on any sized grid.

Introduction

Probability field (p-field) simulation was proposed by Srivastava [5] in 1992 and Froidevaux [2] in 1993. P-field simulation is performed in 2 steps: (1) the local distributions of uncertainty (conditional cumulative distribution functions or ccdfs) are established at each location, and (2) simulated values are drawn from the ccdfs with correlated uniform random numbers. The random numbers must be correlated, or the simulated values will be too random. An appealing aspect of p-field simulation is that the ccdfs can be based on a variety of data sources including soft data and expert judgment.

Probability field simulation is attractive because it dissociates the construction of the ccdfs and the Monte Carlo sampling from them; however, there are two well know artifacts of probability field (p-field) simulation (Pyrcz and Deutsch, 2000 [4]): (1) the local conditioning data almost always appear as local minima and maxima and (2) the covariance is not reproduced in the presence of conditioning data. The first artifact will not be covered in this paper.

Journal proved that p-field simulation correctly reproduces the input variogram in the unconditional case [3]. This is not very practical when attempting to generate conditional simulations. In the presence of conditioning data, the covariance field becomes non-stationary. The covariance values close to conditioning data are quite different compared to far away from conditioning data.

The non-stationary covariance depends on the data configuration and the variogram used for the simulation. The non-stationary covariance field is correctly reproduced in sequential Gaussian simulation because previously simulated nodes are used in the estimation of other nodes. Usually, the correlated field of random numbers used in p-field simulation is generated using a stationary covariance matrix. This gives conditional p-field simulated values the incorrect covariance.

In a previous paper [5] we derived a correction for the p-field covariance bias problem. The correction was calculated as a function of the conditioning data under a multivariate Gaussian framework. The correction algorithm was applied in an LU-based simulation program and allowed small fields to be simulated unconditionally accounting for the conditioning data. By using this corrected unconditional simulation, the output conditional p-field simulation had the correct covariance. However, the LU-based approach limited the size of the field that could be considered.

In this paper we will present an approximate local correction for p-field. Although it is not as theoretically valid compared to the previous correction, it allows larger fields to be simulated.

Covariance between Simulated Values in a Multivariate Gaussian Setting

Consider simulating N locations with n conditioning data. The covariance matrix between the data is defined using the modeled covariance function. The $n \times n$ covariance matrix between the data is:

$$C_{11} = \begin{bmatrix} C(u_1 - u_1) & \cdots & C(u_1 - u_n) \\ \vdots & \ddots & \vdots \\ C(u_n - u_1) & \cdots & C(u_n - u_n) \end{bmatrix} \quad (1)$$

where $C(u_i - u_j)$ is the covariance function for the distance between data locations i and j . Consider the covariance matrix between the n data and the N locations being simulated:

$$C_{12} = \begin{bmatrix} C(u_1 - u^{(1)}) & \cdots & C(u_1 - u^{(N)}) \\ \vdots & \ddots & \vdots \\ C(u_n - u^{(1)}) & \cdots & C(u_n - u^{(N)}) \end{bmatrix} \quad (2)$$

where $C(u_i - u^j)$ is the covariance function for the distance between the data location i and the location j being simulated. Consider the covariance matrix between the N locations being simulated:

$$C_{22} = \begin{bmatrix} C(u^{(1)} - u^{(1)}) & \cdots & C(u^{(1)} - u^{(N)}) \\ \vdots & \ddots & \vdots \\ C(u^{(N)} - u^{(1)}) & \cdots & C(u^{(N)} - u^{(N)}) \end{bmatrix} \quad (3)$$

where $C(u^{(i)} - u^{(j)})$ is the covariance function between the locations being estimated. By combining Equations (1), (2), and (3) we can obtain an expression for the conditional covariance matrix of the N points being simulated given the n data:

$$C_{(u^{(1)}, \dots, u^{(N)} | u_1, \dots, u_n)} = C_{22} - C_{12}^T C_{11}^{-1} C_{12} \quad (4)$$

It is interesting to note that the term $C_{11}^{-1} C_{12}$ is the solution the of the N simple kriging systems:

$$C_{11}^{-1} C_{12} = [\lambda] \quad (5)$$

Substitute Equation (5) into Equation (4):

$$\begin{aligned} C_{(u^{(1)}, \dots, u^{(N)} | u_1, \dots, u_n)} &= C_{22} - C_{12}^T [\lambda] \\ &= C_{22} - \begin{bmatrix} \sum_{\alpha=1}^n \lambda_{\alpha}^{(1)} C(u_{\alpha} - u^{(1)}) & \cdots & \sum_{\alpha=1}^n \lambda_{\alpha}^{(N)} C(u_{\alpha} - u^{(1)}) \\ \vdots & \ddots & \vdots \\ \sum_{\alpha=1}^n \lambda_{\alpha}^{(1)} C(u_{\alpha} - u^{(N)}) & \cdots & \sum_{\alpha=1}^n \lambda_{\alpha}^{(N)} C(u_{\alpha} - u^{(N)}) \end{bmatrix} \end{aligned} \quad (6)$$

Next, substitute the kriging equations into Equation (6). Recall the simple kriging system of equations:

$$\sum_{\beta=1}^n \lambda_{\beta} C_{\alpha\beta} = C_{\alpha 0} \quad \alpha = 1, \dots, n \quad (7)$$

After the substitution, the simplified expression for the conditional covariance between locations i and j is:

$$C_{(u^{(1)}, \dots, u^{(N)} | u_1, \dots, u_n)} = C_{22} - \begin{bmatrix} \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{(1)} \lambda_{\beta}^{(1)} C(u_{\alpha} - u_{\beta}) & \cdots & \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{(N)} \lambda_{\beta}^{(1)} C(u_{\alpha} - u_{\beta}) \\ \vdots & \ddots & \vdots \\ \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{(1)} \lambda_{\beta}^{(N)} C(u_{\alpha} - u_{\beta}) & \cdots & \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{(N)} \lambda_{\beta}^{(N)} C(u_{\alpha} - u_{\beta}) \end{bmatrix} \quad (8)$$

The conditional covariance between two simulated nodes simplifies as:

$$C_{(u^{(i)}, u^{(j)} | u_1, \dots, u_n)} = C_{ij} - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{(j)} \lambda_{\beta}^{(i)} C_{\alpha\beta} \quad (9)$$

This means that the conditional covariance between 2 points being simulated in the presence of conditioning data is exactly known. It is a function of the covariance between the data values and of the covariance between the data and locations being estimated. Note that when there are no conditioning data for the locations i and j , the correct covariance is the input covariance.

Covariance between Simulated Values using P-Field

Now that we have an expression for the conditional covariance between locations in a multivariate Gaussian setting, we need to determine the covariance between simulated locations in p-field. For the p-field proof we will only consider two locations, i and j , within a multivariate Gaussian setting.

Estimating the local cdfs is the first step for p-field simulation. Simple kriging can be used to estimate the mean and variance at each location given the n conditioning data. The mean and variance completely define the cdf in a multivariate Gaussian setting. The cdfs at the 2 locations are:

$$\begin{aligned} Y_i^* &= m_i = \sum_{\alpha=1}^n \lambda_{\alpha}^{(i)} Y_{\alpha} & Y_j^* &= m_j = \sum_{\alpha=1}^n \lambda_{\alpha}^{(j)} Y_{\alpha} \\ \sigma_i^2 &= 1 - \sum_{\alpha=1}^n \lambda_{\alpha}^{(i)} C_{\alpha(i)} & \sigma_j^2 &= 1 - \sum_{\alpha=1}^n \lambda_{\alpha}^{(j)} C_{\alpha(j)} \end{aligned} \quad (10)$$

Using the cdfs from Equation (10), the p-field simulated values at each location are:

$$\begin{aligned} Y_{cs(i)} &= m_i + \sigma_i Y_{s(i)} \\ Y_{cs(j)} &= m_j + \sigma_j Y_{s(j)} \end{aligned} \quad (11)$$

where $Y_{s(i)}$ and $Y_{s(j)}$ are unconditionally simulated values at locations i and j respectively. The unconditionally simulated values have a specific correlation structure and are normally distributed with a mean of 0 and a variance of 1. The covariance between $Y_{cs(i)}$ and $Y_{cs(j)}$ is:

$$\begin{aligned} Cov\{Y_{cs(i)}, Y_{cs(j)}\} &= E\{Y_{cs(i)} \cdot Y_{cs(j)}\} - E\{Y_{cs(i)}\} \cdot E\{Y_{cs(j)}\} \\ &= E\{(m_i + \sigma_i Y_{s(i)}) \cdot (m_j + \sigma_j Y_{s(j)})\} - E\{m_i + \sigma_i Y_{s(i)}\} \cdot E\{m_j + \sigma_j Y_{s(j)}\} \\ &= E\{m_i m_j + m_j \sigma_i Y_{s(i)} + m_i \sigma_j Y_{s(j)} + \sigma_i \sigma_j Y_{s(i)} Y_{s(j)}\} - E\{m_i\} \cdot E\{m_j\} \\ &= E\{\sigma_i \sigma_j Y_{s(i)} Y_{s(j)}\} + E\{m_i m_j\} - E\{m_i\} \cdot E\{m_j\} \\ &= \sigma_i \sigma_j Cov\{Y_{s(i)}, Y_{s(j)}\} \\ &= \sigma_i \sigma_j Cov\{\mathbf{h}\} \end{aligned} \quad (12)$$

The expected values in step 4 of Equation (12) cancel out because m_i and m_j are constants. The covariance between 2 locations in p-field simulation is a function of the standard deviation at each location and the covariance of the random field between the 2 locations. We need to correct the covariance of the random field so that the output covariance from p-field is correct.

Non-Stationary Covariance Correction

We now know the covariance between locations i and j from p-field, Equation (12), and the correct covariance between the 2 locations, Equation (9). Recall:

$$Cov_{pfield}\{Y_{cs(i)}, Y_{cs(j)}\} = \sigma_i \sigma_j Cov\{Y_{s(i)}, Y_{s(j)}\}$$

$$Cov_{correct} \left(Y_{cs(i)}, Y_{cs(j)} \right) = C_{ij} - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{(j)} \lambda_{\beta}^{(i)} C_{\alpha\beta}$$

We can define a corrected covariance for the unconditional simulation that will provide the correct covariance in the output p-field simulation. Set the p-field corrected covariance to the conditional covariance from Equation (9):

$$Cov_{pfield\ corrected} \left(Y_{cs1}, Y_{cs2} \right) = C_{ij} - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{(j)} \lambda_{\beta}^{(i)} C_{\alpha\beta}$$

rearrange and substitute the p-field covariance:

$$C_{ij} - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{(j)} \lambda_{\beta}^{(i)} C_{\alpha\beta} = \sigma_i \sigma_j Cov \left\{ Y_{s(i)}, Y_{s(j)} \right\}$$

rearrange to get a corrected covariance for the unconditional simulation that will give p-field simulation the correct non-stationary covariance field:

$$Cov \left\{ Y_{s(i)}, Y_{s(j)} \right\} = \frac{C_{ij} - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{(j)} \lambda_{\beta}^{(i)} C_{\alpha\beta}}{\sigma_i \sigma_j} \quad (13)$$

The corrected covariance for the unconditional simulation will give the p-field conditional simulations the correct spatial structure. Next, we present method to generate an unconditional simulation with a non-stationary covariance matrix.

Mixture Model Correction

The large field covariance correction is based on a mixture model. The mixture model aims to approximate the location correction from Equation (13). One component of the mixture is a standard unconditional simulation and the second component is a random field. By combining these two components, we can approximate the theoretical local p-field covariance correction.

Let Y_1 be a Gaussian unconditional simulation with a covariance function $C(\mathbf{h}) = 1 - \gamma(\mathbf{h})$. Then let Y_2 be a random Gaussian field that is $N(0,1)$ and has a covariance function $C(\mathbf{h}) = 0 \quad \forall \mathbf{h} > 0$. The combination of the Y_1 and Y_2 will be the corrected unconditional simulation for use with p-field. The combination of the Y_1 and Y_2 is done with Equation (14):

$$Y_3(\mathbf{u}) = Y_1(\mathbf{u}) \sqrt{1 - f(\mathbf{u})} + Y_2(\mathbf{u}) \sqrt{f(\mathbf{u})} \quad (14)$$

Where $f(\mathbf{u})$ is a weight applied to the random field and $1-f(\mathbf{u})$ is the weight applied to the correlated unconditional simulation. We know from the theoretical correction that in areas with little to no conditioning data, no correction is needed. In areas with dense conditioning data a large covariance correction is required. For this reason, $f(\mathbf{u})$ is calculated as a function of the kriging variance calculated from the conditioning data. Let $f(\mathbf{u})$ take the following form (Equation (15)):

$$f(\mathbf{u}) = \left(\sigma_{Y_1(\mathbf{u})} \right)^{\omega} \cdot a_0 \quad (15)$$

Where ω and a_0 are calibration parameters to control the correction. There is no method to directly calculate either ω or a_0 . Several calibration runs will be required to get correct values for ω or a_0 that result in the correct spatial structure in the resulting conditioned model.

Large Scale Unconditional Simulation

The p-field simulation program (`pfsim`) from GSLIB was modified to perform the mixture model correction. The modified program is called `pfsim_cc.for` (`pfsim` unconditional corrected covariance). The parameters for the program are:

```
Line  START OF PARAMETERS:
  1  1          -1=continuous(cdf), 0=categorical(pdf)
  2  1          -1=indicator cdfs, 0=Gaussian (mean,var)
  3  3          - number thresholds/categories
  4  0.5 1.0 2.5 - thresholds / categories
  5  cluster.dat -Within-class details: file with global dist
  6  3  0          - ivr, iwt
  7  0.0 30.0     - minimum and maximum Z value
  8  1  1.0       - lower tail: option, parameter
  9  1  1.0       - middle : option, parameter
 10  1  2.0       - upper tail: option, parameter
 11  kt3d.out    -file with input conditional distributions
 12  1  2          - columns for mean, var, or cdf values
 13  -1.0e21 1.0e21 - trimming limits
 14  sgsim.out   -file with input p-field realizations
 15  1          - column number in p-field file
 16  0          - 0=Gaussian, 1=uniform [0,1]
 17  pfsim.out   -file for output realizations
 18  1          -number of realizations
 19  50 50 1     -nx, ny, nz
 20  1.0 1.0     -omega and alpha
```

The parameters on **Lines 1-19** are the same parameters as the `pfsim` program; refer to [1] for details. The calibration parameters are specified on **Line 20**. Currently, the program only allows for Gaussian conditional distributions. The variance component of the distribution is used for the local correction.

Example

The example shows that the covariance artifact from p-field can be corrected with the mixture model approach. The effects that the calibration parameters have on the resulting spatial structure are shown and the results from the traditional p-field approach are compared to the results from the calibrated mixture model

This example uses the `cluster.dat` data file from the GSLIB book [1]. We check the variogram reproduction for sequential Gaussian simulation, p-field simulation with the stationary covariance, and p-field with the mixture model correction. In addition, we need to run a pre-processing step to get values for ω and a_0 that give good variogram reproduction. Figure 1 shows the normal score data from `cluster.dat`. Figure 2 shows the isotropic variogram:

$$\gamma(\mathbf{h}) = 0.2 + 0.8 \cdot sph_{a=1.0}(\mathbf{h})$$

Figure 3 and Figure 4 show the cdf's for p-field. The cdfs are calculated from simple kriging of the normal scores. Several calibration runs had to be made to get values for ω and a_0 . The calibration runs were for a large range of ω and a_0 values. ω ranged from 0.0 up to 10 and a_0 ranged from 0.0 to 1.8. Figure 5 to Figure 8 show 4 different cases from the calibration. The calibration parameters have a large impact on the variogram of the conditioned simulation. The combination of $\omega=4$ and $a_0=1.8$ was chosen from the results of the calibration runs.

The sequential Gaussian simulation results are shown in Figure 9 and Figure 10. The variogram reproduction is acceptable within normal fluctuations. Figure 11 and Figure 12 show the p-field simulation results using an uncorrected unconditional covariance field. The variograms from the simulation are too continuous. In other words the conditional simulation is too smooth. The mixture model corrected p-field results show a marked improvement for variogram reproduction; Figure 13 and Figure 14. The shape of the variogram is not perfect, but it is closer to the input model than the traditional p-field simulation. The corrected variogram has a Gaussian-like shape near a lag distance of zero.

Conclusions

Like all geostatistical methods, p-field algorithms have their place. There are 2 artifacts that practitioners must be aware of when using p-field: (1) data appear as local minima and maxima, and (2) the covariance is too smooth in the presence of conditioning data. Previously, a theoretical non-stationary covariance correction was derived for p-field simulation. This correction worked very well but was limited due to the LU implementation. Large grids (more than 10,000 cells) could not be simulated. The mixture model approach presented here allows a non-stationary correction to be applied to large simulation models. This allows for p-field models to be constructed without the known covariance bias. Although this correction is not exact, it should produce better results than traditional p-field with its covariance bias.

References

- [1] C. V. Deutsch and A. G. Journel. *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York, 2nd edition, 1998.
- [2] R. Froidevaux. Probability field simulation. In A. Soares, editor, *Geostatistics Troia 1992*, volume 1, pages 73–84. Kluwer, 1993.
- [3] A G. Journel. Probability fields: Another look and a proof. In *Report 8, Stanford Center for Reservoir Forecasting*, Stanford, CA, May 1995.
- [4] M. P. Pyrcz and C. V. Deutsch. Two artifacts of probability field simulation. *Math Geology*, 33(7):775–799, 2001.
- [5] C. Neufeld, J. M. Ortiz and C. V. Deutsch. A Non-Stationary Correction of the Probability Field Covariance Bias. In *Annual Report 7, Centre for Computational Geostatistics*, Edmonton, AB, 2005.
- [6] R. M. Srivastava. Reservoir characterization with probability field simulation. In *SPE Annual Conference and Exhibition*, pages 927–938, Washington, DC, October 1992.

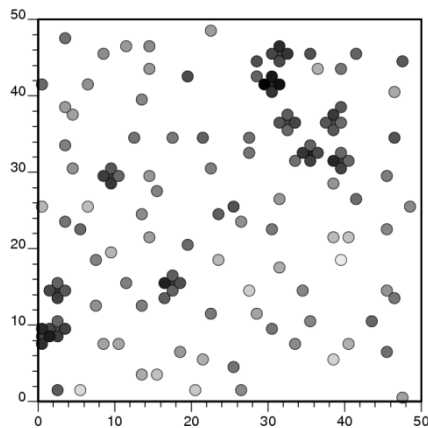


Figure 1: Data from cluster.dat.

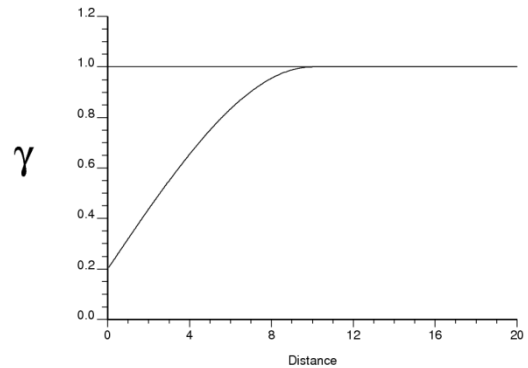


Figure 2: Isotropic variogram.

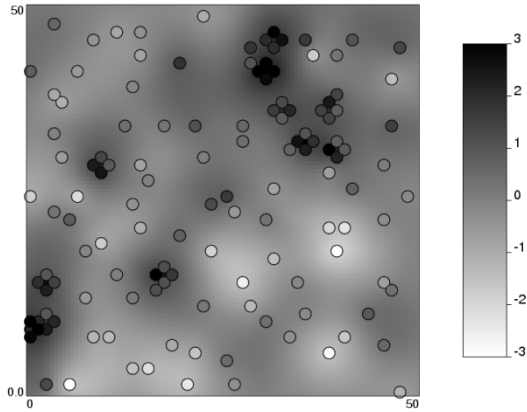


Figure 3: Simple kriging estimate.

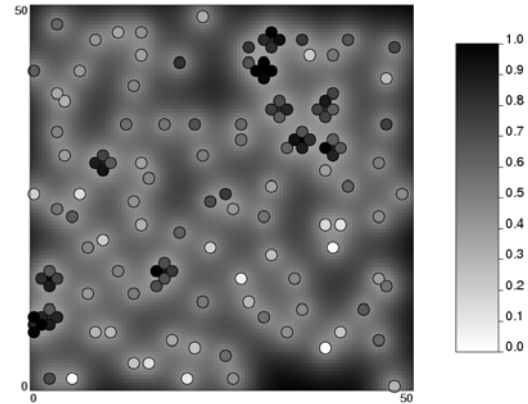


Figure 4: Simple kriging estimation variance.

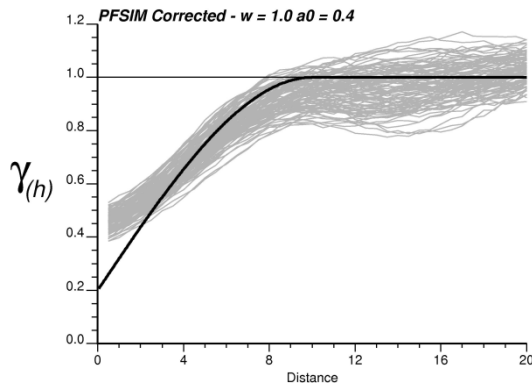


Figure 5: Variogram reproduction with $\omega=1.0$ and $a_0=0.4$.

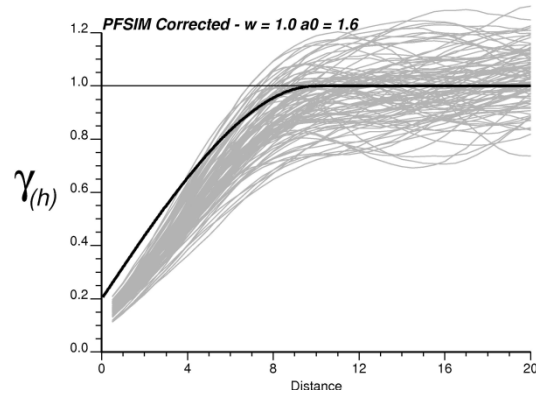


Figure 6: Variogram reproduction with $\omega=1.0$ and $a_0=1.6$.

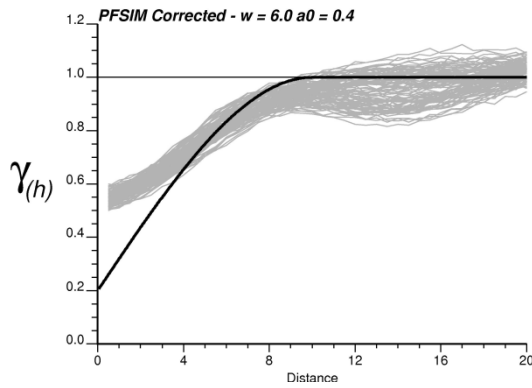


Figure 7: Variogram reproduction with $\omega=6.0$ and $a_0=0.4$.

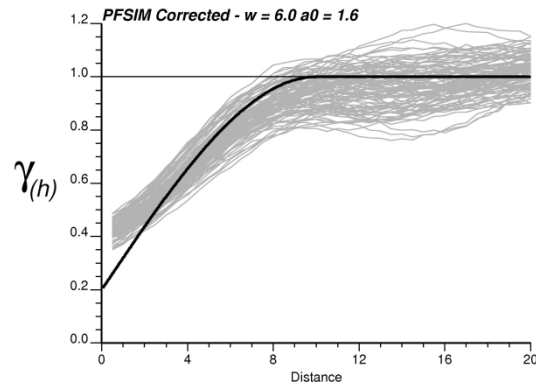


Figure 8: Variogram reproduction with $\omega=6.0$ and $a_0=1.6$.

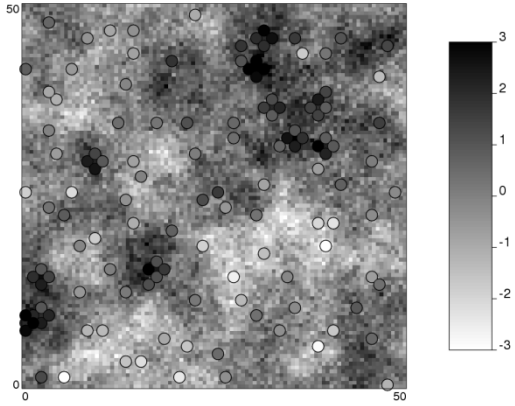


Figure 9: SGSIM Realization.

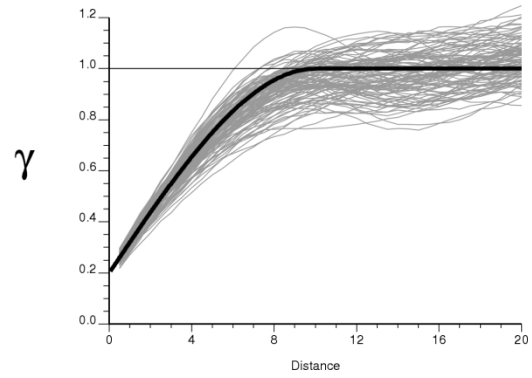


Figure 10: SGSIM variogram reproduction.

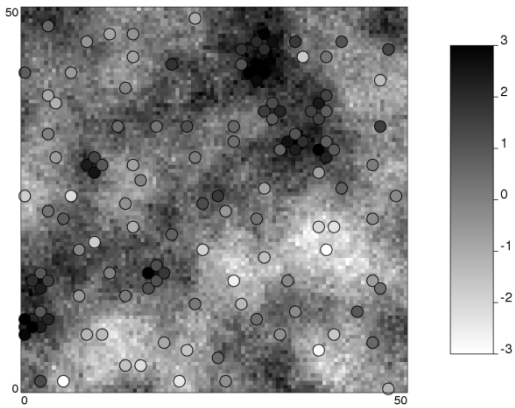


Figure 11: Uncorrected PFSIM realization.

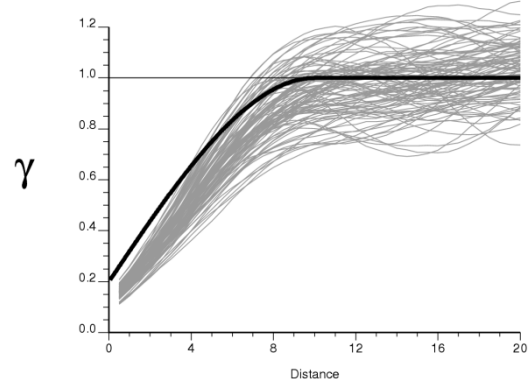


Figure 12: Uncorrected PFSIM variogram reproduction.

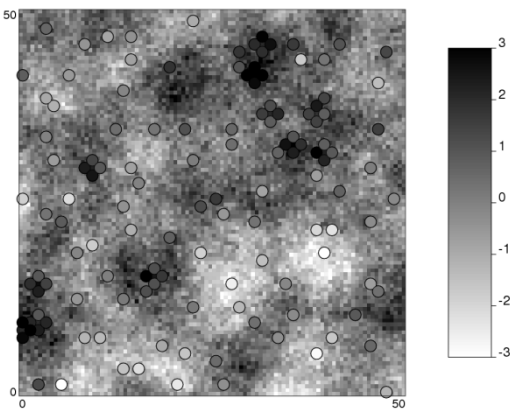


Figure 13: Corrected PFSIM realization.

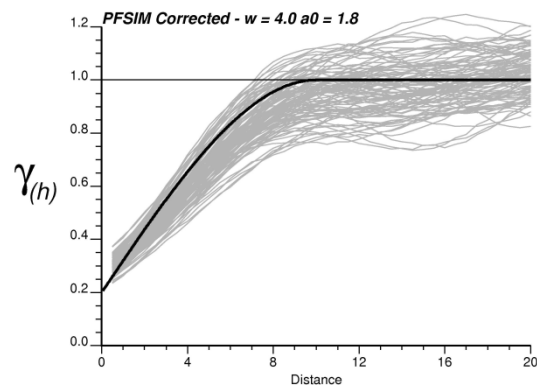


Figure 14: Mixture model corrected PFSIM variogram reproduction