

A Non-Stationary Correction of the Probability Field Covariance Bias

Chad Neufeld¹, Julián M. Ortiz², and Clayton V. Deutsch¹

¹Centre for Computational Geostatistics (CCG) – University of Alberta

²Department of Mining Engineering – University of Chile

One well known artifact of the probability field (p-field) simulation algorithm is a too large covariance near conditioning data. We derive the increased covariance as a function of the input ergodic covariance model and the locations of the conditioning data. The correct covariance in the multivariate Gaussian case is derived. A non-stationary covariance correction for p-field is proposed. Several examples are presented to illustrate the correction.

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 known 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.

Journel 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 this paper we derive the correct covariance between nodes in the presence of conditioning data in a multivariate Gaussian framework. The covariance bias in p-field simulation can be calculated. A correction factor is defined so that the output conditional p-field simulation has the cor-

rect covariance. An LU based unconditional simulation program was written for generating unconditional simulations that honor the non-stationary p-field corrected covariance field.

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 simultaneous solution the of the N simple kriging systems:

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

Substitute Equation (5) into Equation (4):

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

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

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

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

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

The conditional covariance between two simulated nodes simplifies as:

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

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:

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

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

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

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:

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

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:

$$\begin{aligned}
Cov_{pfield}\{Y_{cs(i)}, Y_{cs(j)}\} &= \sigma_i \sigma_j Cov\{Y_{s(i)}, Y_{s(j)}\} \\
Cov_{correct}(Y_{cs(i)}, Y_{cs(j)}) &= C_{ij} - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{(j)} \lambda_{\beta}^{(i)} C_{\alpha\beta}
\end{aligned}$$

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):

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

rearrange and substitute the p-field covariance:

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

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

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

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.

Unconditional Simulation Using a Non-Stationary Covariance Matrix

The LU simulation algorithm is perfectly suited to generate unconditional simulations with a non-stationary covariance function. The `lusim` program from GSLIB was modified to calculate the conditional covariance matrix for the N locations given the n data, Equation (8), perform the p-field covariance correction, Equation (13), and then generate L unconditional realizations for input to `pfsim`. Recall the correct non-stationary covariance matrix for the multivariate Gaussian case:

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

After applying the p-field covariance correction the conditional covariance matrix becomes:

$$C_{(u^{(1)}, \dots, u^{(N)} | u_1, \dots, u_n)} = \begin{bmatrix} 1 & \cdots & \frac{C_{(1)(N)} - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{(N)} \lambda_{\beta}^{(1)} C_{\alpha\beta}}{\sigma_1 \sigma_N} \\ \vdots & \ddots & \vdots \\ \frac{C_{(N)(1)} - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{(1)} \lambda_{\beta}^{(N)} C_{\alpha\beta}}{\sigma_N \sigma_1} & \cdots & 1 \end{bmatrix}$$

We then perform an LU decomposition of the corrected covariance matrix:

$$C_{(u^{(1)}, \dots, u^{(N)} | u_1, \dots, u_n)} = LU$$

Unconditional Gaussian simulations are quickly calculated by a simple matrix multiplication:

$$y = Lw$$

where w is an N by 1 vector of independent Gaussian values and y is the resulting N by 1 vector of unconditionally simulated values with the p-field corrected covariance. This is repeated L times for the multiple realizations. The covariance of the unconditional simulations is such that the resulting p-field conditional simulations will have the correct covariance.

Unconditional simulation with a non-stationary covariance matrix is not simple. There is no standard software to simulate with a non-stationary covariance function. There is a concern that the non-stationary covariance may be non-positive definite. However, we have not observed any problems as the non-stationary covariances seem to be positive semi definite by construction.

Theoretically, the correction will work for any size of grid. However, the size of the unconditional simulation is limited since an LU based approach was chosen. The practical size of the simulation grid should not exceed 5000 – 10000 nodes.

The `lusim` based program was named `pfsim_ucc` (pfsim unconditional corrected covariance). The parameters for the program are:

Line	START OF PARAMETERS:	
1	parta.dat	-file with data
2	1 2 0	- columns for X,Y,Z
3	3	-debugging level: 0,1,2,3
4	pfsim_ucc.dbg	-file for debugging output
5	pfsim_ucc.out	-file for realization(s)
6	100	-number of realizations
7	4 40.25 0.5	-nx,xmn,xsiz
8	4 28.25 0.5	-ny,ymn,ysiz
9	1 0.00 1.0	-nz,zmn,zsiz
10	112063	-random number seed
11	1 0.2	-nst, nugget effect
12	1 0.8 0.0 0.0 0.0	-it,cc,ang1,ang2,ang3
13	10.0 10.0 10.0	-a_hmax, a_hmin, a_vert

The data file on **Lines 1 and 2** specifies the locations of the conditioning data. The data values themselves are not used. The covariance bias is only dependant on the covariance function. **Lines 3 and 4** are for the debugging level and the debugging output file. The output unconditional realizations are stored in the output file on **Line 5**. **Line 6** is the number of realizations to generate. The number of realizations can be set quite high. Most of the CPU time is devoted to the LU decomposition of the C_{22} matrix. Once the decomposition is done, the realizations can be generated extremely fast. **Lines 7, 8, and 9** contain the grid definition in GSLIB format. The random number seed is in **Line 10**. **Lines 11 and greater** specify a 3-D variogram model in standard GSLIB conventions.

Example #1

The first example shows that the covariance artifact from p-field can be predicted correctly. The results from p-field are compared with sequential Gaussian simulation. The correction in Equation (13) is used and the covariance from p-field with the non-stationary unconditional simulation is compared with the multivariate Gaussian results. Consider the setting shown in Figure 1 with the following isotropic variogram:

$$\gamma(\mathbf{h}) = 0.1 + 0.9 \cdot sph_{a=10}(\mathbf{h})$$

Solving the kriging equations gives the following ccd's:

$$\begin{aligned} m_1 &= 0.445 & m_2 &= 0.326 \\ \sigma_1 &= 0.766 & \sigma_2 &= 0.810 \\ \lambda_1^{(1)} &= 0.548 & \lambda_1^{(2)} &= 0.336 \\ \lambda_2^{(1)} &= 0.228 & \lambda_2^{(2)} &= 0.406 \end{aligned}$$

The covariance between the 2 simulated locations should be identical to the predicted covariance from the multivariate Gaussian theory. The predicted conditional covariance is:

$$\begin{aligned} C_{(1,2|u_1,u_2)} &= C_{12} - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{(2)} \lambda_{\beta}^{(1)} C_{\alpha\beta} \\ &= C_{12} - \lambda_1^{(2)} \lambda_1^{(1)} C_{11} - \lambda_1^{(2)} \lambda_2^{(1)} C_{12} - \lambda_2^{(2)} \lambda_1^{(1)} C_{21} - \lambda_2^{(2)} \lambda_2^{(1)} C_{22} \\ &= 0.389 - 0.184 - 0.019 - 0.054 - 0.093 \\ &= 0.039 \end{aligned}$$

The predicted covariance from p-field is:

$$\begin{aligned} Cov\{Y_{cs1}, Y_{cs2}\} &= \sigma_1 \sigma_2 Cov\{Y_{s1}, Y_{s2}\} \\ &= \sigma_1 \sigma_2 C_{12} \\ &= (0.766)(0.810)(0.389) \\ &= 0.241 \end{aligned}$$

The covariance between the simulated locations using p-field is significantly higher than the theoretical conditional covariance. The results predicted by theory match the numerical results shown in Figures 2 and 3.

Now we apply the covariance correction for the unconditional simulation and see if the covariance from p-field is correct. The corrected covariance for the unconditional simulation is:

$$\begin{aligned} Cov\{Y_{s1}, Y_{s2}\} &= \frac{C_{12} - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{(2)} \lambda_{\beta}^{(1)} C_{\alpha\beta}}{\sigma_1 \sigma_2} \\ &= \frac{0.039}{(0.766)(0.810)} \\ &= 0.063 \end{aligned}$$

Instead of the unconditional covariance of $C_{12}=0.389$. The corrected p-field results are shown in Figure 4. The covariance, and variogram, between the simulated values is now correct.

Example #2

The second 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 non-stationary corrected covariance. Figure 5 shows the normal score data from `cluster.dat`. Figure 6 shows the isotropic variogram:

$$\gamma(\mathbf{h}) = 0.2 + 0.8 \cdot sph_{a=10}(\mathbf{h})$$

Figures 7 and 8 show the ccdf's for p-field. The ccdfs are calculated from simple kriging of the normal scores.

The sequential Gaussian simulation results are shown in Figures 9 and 10. The variogram reproduction is acceptable within normal fluctuations. Figures 11 and 12 show the p-field simulation results using an unconditional covariance field. The variograms from the simulation are too low. In other words the conditional simulation is too smooth. The p-field results improve significantly after applying the covariance correction, Figures 13 and 14.

Conclusions

Probability field simulation has gained some acceptance in modern geostatistics. However 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.

A non-stationary covariance correction was derived for p-field simulation. It corrects the unconditional simulation so the output p-field simulation will have the correct covariance.

The covariance correction is limited in size based on the LU approach used. This approach could be applied to larger problems if a non-stationary sequential simulation approach was developed.

Acknowledgments

The authors thank the industry sponsors of the Centre for Computational Geostatistics at the University of Alberta for financial support. Dr. Deutsch holds the Alberta Chamber of Resources Industry Chair in Mining Engineering; the supporting companies are gratefully acknowledged. Dr. Ortiz acknowledges the support of Codelco Chile at the Department of Mining Engineering, University of Chile. NSERC is also thanked for the industrial postgraduate scholarship that Mr. Neufeld holds; that scholarship is due to special Placer Dome support.

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] R. M. Srivastava. Reservoir characterization with probability field simulation. In *SPE Annual Conference and Exhibition*, pages 927–938, Washington, DC, October 1992.

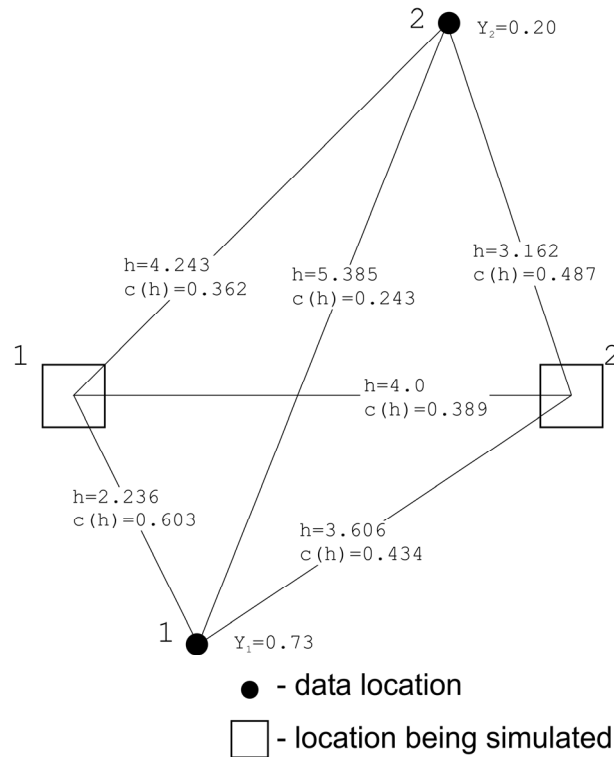


Figure 1: Configuration for example #1.

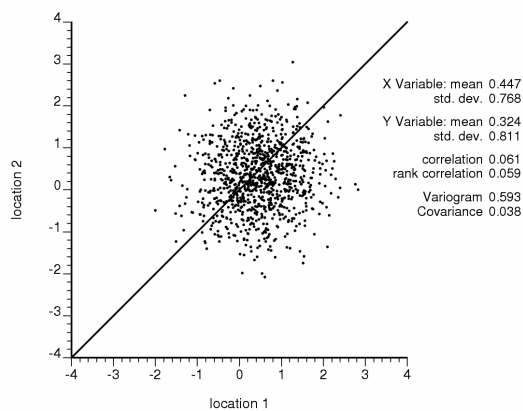


Figure 2: Sequential Gaussian simulation results.

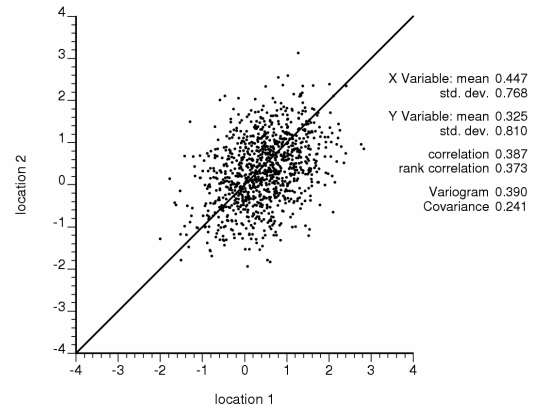


Figure 3: P-Field simulation results.

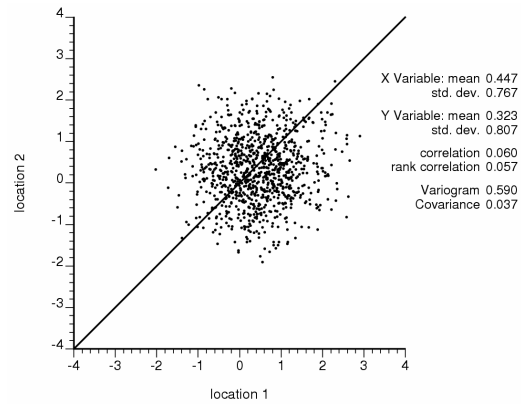


Figure 4: P-Field simulation results with the corrected unconditional covariance. The output covariance is right.

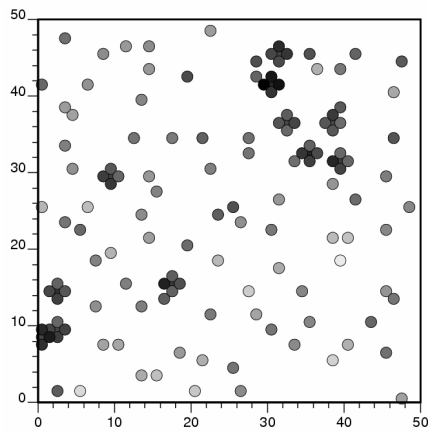


Figure 5: Data from cluster.dat.

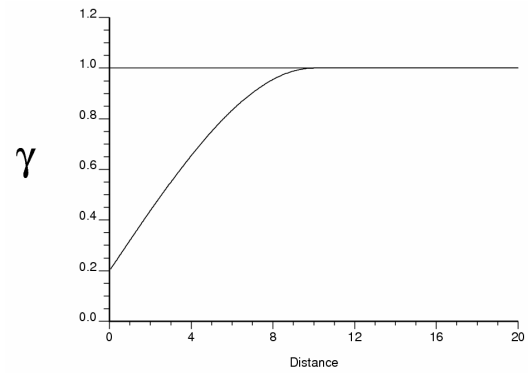


Figure 6: Isotropic variogram.

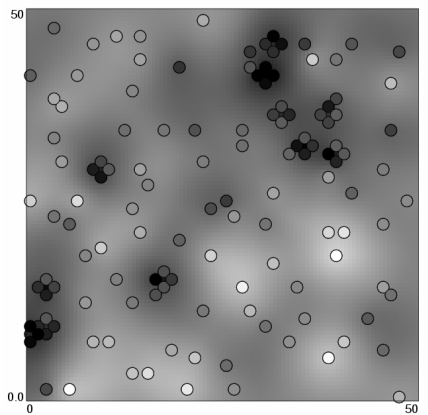


Figure 7: Simple kriging estimate.

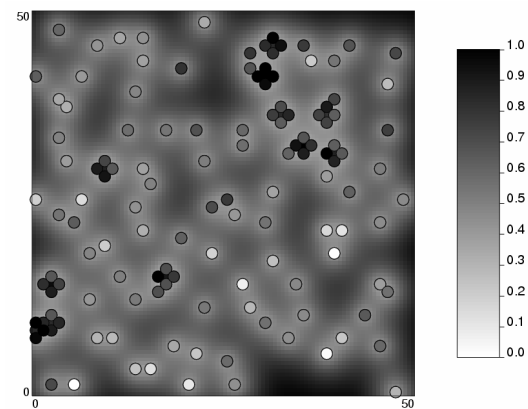


Figure 8: Simple kriging estimation variance.

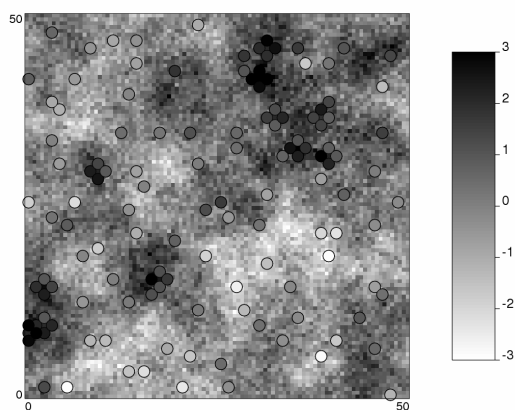


Figure 9: SGSIM Realization.

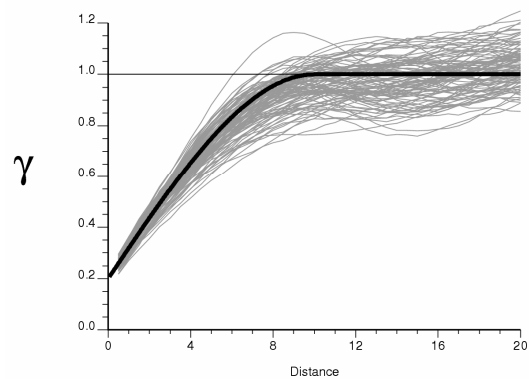


Figure 10: SGSIM variogram reproduction.

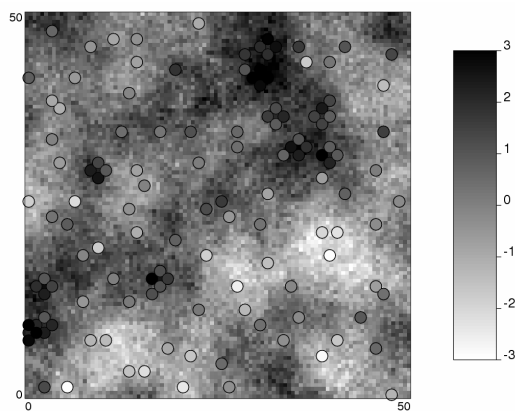


Figure 11: Uncorrected PFSIM realization.

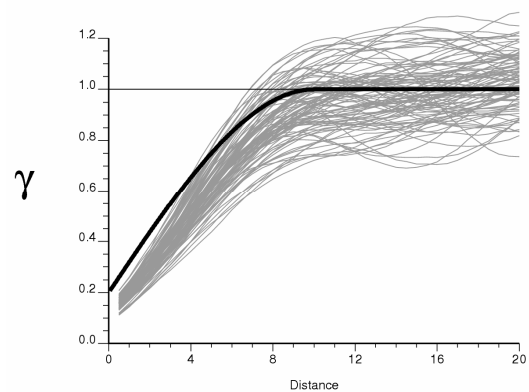


Figure 12: Uncorrected PFSIM variogram reproduction.

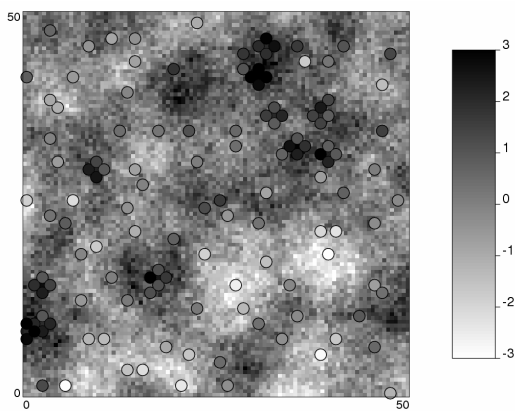


Figure 13: Corrected PFSIM realization.

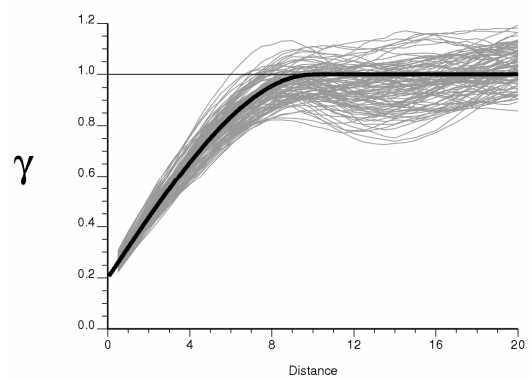


Figure 14: Corrected PFSIM variogram reproduction