

# Automatic Report Generation for Geostatistical Modeling Software

John G. Manchuk and Clayton V. Deutsch

*Many parameters are involved in sequential Gaussian simulation and they require tuning to ensure the set of realizations reproduce the required statistics. For large problems, that is, ones with grids having many cells, generating realizations to check that the histogram and variogram are reproduced can be very time consuming. A goal of this work is to provide a tool that allows more efficient checking of realizations for tuning parameters. It is possible to evaluate realizations without building them. The error between the expected covariance of realizations and the known covariance matrix defined by the input variogram can be measured. It is also possible to predict the variance of the means of a set of realizations, the expected variance of a set of realizations, and the expected variogram reproduction at a variety of lags for a set of realizations, all without the need to generate them. A program that assess these measures and makes suggestions to adjust parameters is developed.*

## 1. Introduction

Geostatistics often involves building complex models of geology and geological properties for a variety of applications. The modeling process can be quite time consuming and performing several runs to calibrate input parameters may be impractical. It would be beneficial to know the quality and accuracy of models that will result from a set of parameters without having to actually build a full set of realizations. Consider having to check variogram reproduction for a large set of realizations: 1 – a substantial amount of time is spent building the realizations, sometime up to eight hours (usually done overnight); 2 – the resulting variogram reproduction may be unacceptable and parameters require adjustment, after which another overnight run of model building is done. Unacceptable reproduction of the variogram may be in the form of a systematic bias where the experimental average over all realizations consistently falls below the input model, or variograms of each realization indicate a significant amount of variance at each evaluated lag. Other checks may include histogram reproduction, variation in the mean and variance, recovery of trends, and variation in statistics through some post-processing step.

Some recent research from the Centre for Computational Geostatistics has concentrated on selecting parameters for geostatistical modeling. A paper by Guo and Deutsch (2008) looked at the relationships between kriging weights to determine an adequate number of data to use in kriging. The average and variance of the kriging variance were observed while increasing the number of conditioning data in one, two and three dimensions. It was concluded that 30, 40, and 60 data were adequate in 1D, 2D and 3D problems respectively. Another work by Villalba and Deutsch (2010) looked at predicting the expected fluctuations in the mean of a set of realizations due to ergodicity. It is possible to predict the expected variation in the mean of set of realizations given a particular variogram function and domain size. In the case of a non-ergodic domain, that is, a domain with extents less than the correlation length of the random variable, the variation in the mean will be greater than zero and one should anticipate a certain amount of variation in histogram reproduction in final models.

It is possible to predict several other outcomes for a set of realizations using SGS without generating them. A recent paper by Emery and Peláez (2011) focused on reconstructing the covariance matrix of a realization under a variety of different conditions such as unconditional, conditional, with multiple variables, and using different random paths and number of data in kriging. The ability to reconstruct the covariance matrix permits the overall accuracy of the realizations to be measured in comparison to the input variogram model. It is also possible to assess the expected variation in the variogram for a set of realizations without generating them.

This paper utilizes the methods discussed by Emery (2011) and identifies two other measures that can be recovered from the covariance matrix: the expected variation in the mean of realizations as discussed by Villalba and Deutsch (2010), and the expected variance for each realization. Measures to assess the goodness of realizations from sequential Gaussian simulation (SGS) are combined into a program that generates a report for users. The results can be used to decide if parameters should be adjusted to improve models without the potentially time consuming process of building them.

## 2. Background

Assessing the quality of a realization from SGS is ultimately dependent on how well the algorithm can reproduce the covariance structure of the underlying random variable. SGS is a widely used algorithm for modeling geological properties (Alabert, 1987; Deutsch and Journel, 1998). It involves estimating conditional distributions for a set of points and is based on a factorization of the joint probability of points and conditioning data defined by Equation 1, where  $P$  is probability,  $n$  is the number of conditioning data and  $N$  is the number of points. Points are typically defined as a set of grid cell centers, with  $N$  being the number of grid cells. For Gaussian random functions (RF), conditional distributions are given by simple kriging and used to generate Gaussian random values,  $y_n, \dots, y_{n+N}$ .

$$\begin{aligned}
 P\{y_{n+1} \leq Y_{n+1}, \dots, y_{n+N} \leq Y_{n+N} \mid y_1, \dots, y_n\} = \\
 P\{y_{n+1} \leq Y_{n+1} \mid y_1, \dots, y_n\} \times \\
 P\{y_{n+2} \leq Y_{n+2} \mid y_1, \dots, y_n, y_{n+1}\} \times \\
 \vdots \\
 P\{y_{n+N} \leq Y_{n+N} \mid y_1, \dots, y_n, y_{n+1}, \dots, y_{n+N-1}\}
 \end{aligned} \tag{1}$$

SGS is typically used to generate a set of realizations that reproduce input distribution functions, heterogeneity, and the variogram; however, determining the complete factorization is not always possible.

For very small problems ( $n + N \leq 10,000$ ) when all conditioning data and previously simulated values can be included in kriging, the path taken through the grid points in SGS has no effect on results; however,  $N$  is typically large ( $N \geq 10^6$ ) for realistic problems and the conditional probabilities are approximated using a set of nearest neighbors. In this case, several parameters including the path have an effect on results. One approach to avoid path-related artifacts in realizations is based on random or quasi-random paths. For regular grids, the multigrid approach is common (Chiles and Delfiner, 1999; Gomez-Hernandez and Cassiraga, 1994). It is a stratified sampling method (Asmussen, 2007) that provides improved reproduction of the variogram over pure random sampling, especially when anisotropy and more complex nested variogram models are involved.

Other parameters that impact the quality of realizations from SGS include: geometry of the domain being modeled; search window geometry; the presence of conditioning data; the number of nearest neighbors to use for kriging, and; the type of kriging. Effects of domain geometry relate to the theory of ergodicity: as the dimensions of a domain tend to infinity, fluctuations in statistical properties of random fields generated in the domain tend to zero (Chiles and Delifiner, 1999). The size of a domain is measured relative to the variogram involved and is direction dependent. A good overall measure is the average variogram in Equation 2, where  $D$  is the domain,  $|D|$  is the volume of  $D$ ,  $\gamma$  is the variogram, and  $x$  and  $y$  traverse  $D$ , which is typically a three dimensional regular grid. In 3D, the integrals over  $D$  would be triple integrals, while  $x$  and  $y$  would have three components.

$$\bar{\gamma}_D = \frac{1}{|D|^2} \int \int_D \gamma(x-y) dx dy \tag{2}$$

Consider a 1D problem and a variogram with a sill of one and range equal to  $A$ . When  $D < A$ ,  $\gamma(x-y) < 1, \forall xy \in D$  and  $\bar{\gamma}_D < 1$ , indicating the domain is non-ergodic and may result in significant fluctuation in statistics such as the mean and variance of the realizations. When  $D \gg A$ ,  $\gamma(x-y) = 1$  for a very large fraction of  $xy \in D$  and  $\bar{\gamma}_D \cong 1$ , indicating the domain is ergodic. For a Gaussian RF, the mean and variance of every realization would approximately equal zero and one. A program called GAMMABAR is available for doing these calculations.

Statistical fluctuations due to a non-ergodic domain are not an indication of realization quality; however, when implementing algorithms such as SGSIM from GSLIB (Deutsch and Journel, 1998), one may be encouraged to try to obtain Gaussian realizations with zero mean and unit variance. One of the diagnostic outputs from SGSIM is the following:

---

Realization 4: number	=	900	
mean	=	0.0440	(close to 0.0?)
variance	=	0.2217	(close to gammabar(V,V)? approx. 1.0)

---

This output was for a square 2D domain 30 units across and an isotropic zero-nugget spherical variogram with range 50 units. One might be encouraged to start manipulating parameters to obtain a mean and variance of zero, which would be a futile exercise since the domain is non-ergodic. Such diagnostic information should be used as an indication of ergodicity, not to tune parameters unless the domain is ergodic. The SGS assessment algorithm presented in this work provides measures of the expected variation in the mean of realizations and the expected variance of realizations as measures of ergodicity, which are discussed in the next section.

When approximating the factorization in Equation 1, search window geometry affects realization quality since it defines the data that are found for kriging, and how they are found. The latter is controlled by the anisotropy of the search ellipsoid: by altering the orientation and anisotropy of the search, different sets of nearest neighbors can be found. Ideally, the search should be setup to find the set of data with highest correlation and is achieved using a search ellipse that is defined by the anisotropy and orientation of the variogram. Other numerical issues may still have an adverse impact on realization quality, for example, the string effect and screening. Screening effects are especially noticeable with the Gaussian variogram model. Notwithstanding the numerical issues, search effects can be related back to the average variogram Equation, which is re-written as Equation 3, where  $S_i$  is the search domain positioned at each  $x_i \in D$  and the outer integral over  $D$  has been replaced by a sum over the cells in a grid.

$$\bar{\gamma}_D = \frac{1}{N|D||S|} \sum_{i=1}^N \int_{S_i} \gamma(x_i - y) dy, \quad x_i \in D, y \in S_i \quad 3$$

When the search ellipse is very small, for example the same size as a grid cell, the average variogram is approximately zero because  $\gamma(x_i - y) \cong 0 \forall y \in S_i$ , assuming the nugget effect is zero. This indicates realizations should have maximum variation in the mean and a variance of zero; however, the opposite effect occurs with SGS for numerical reasons. When the search is very small, for all points in  $D$  no data for kriging is found and the program assumes the global mean and variance apply. The result will be an uncorrelated Gaussian random field with zero mean and unit variance. Effects are observed when making the search slightly larger than 1 grid cell, in which case at least 1 data is found in the search and the variogram is actually sampled for Equation 3.

An example using a square domain of 50 units and isotropic zero-nugget spherical variogram with range 10 is used to demonstrate too small a search. The average variogram in this case is 0.97, indicating the problem is ergodic, that is, realizations should have zero mean and unit variance. Using a search radius of 1 grid cell, diagnostic output from SGSIM is as follows:

---

Realization 9: number	=	2500
mean	=	-0.0572 (close to 0.0?)
variance	=	0.7028 (close to gammabar(V,V)? approx. 1.0)

---

In this case, expanding the search is the obvious solution to obtain better realizations.

The presence of conditioning data typically has an adverse effect on the quality of realizations, especially when coupled with a small search. Variogram models derived from conditioning data are in almost all cases not an exact match to the experimental variogram of the data. Such noise causes a mismatch between the experimental variogram of realizations and the input variogram model. Positioning of the samples can also cause issues especially when the data are oriented in long strings of samples, which is almost always the case for mining and petroleum applications. Search options such as the octant search are available to help mitigate data positioning effects.

Using a small search for conditional simulation can lead to more variation in variogram reproduction because the variogram of conditioning data may not be stationary. The variogram in local domains defined by the search ellipse deviate from the input model and lead to noise. Measuring the effect of data without generating realizations is difficult because covariance is not dependent on data values, while computing experimental variograms is. Only the effects due to the spatial arrangement of conditioning can be measured. In cases where the assumption of second order stationarity may be suspect, one should consider splitting the domain into stationary sub-domains, or adopting a locally varying model (Boisvert, Manchuk and Deustch, 2009).

The number of nearest neighbors to use for kriging has a similar effect as search ellipsoid geometry. Increasing the number of neighbors is similar to increasing the search radii; however, there are subtle differences in the performance of each during construction of a realization. Consider using  $m$  nearest neighbors, an isotropic

variogram with range  $A$ , a fixed isotropic search radius of  $R$  units,  $R < A$ , and a multigrid random path. If the coarsest level of the random path involves grid cells that are separated by  $d$ ,  $R < d < A$ , then no data are found for kriging at any of the locations in the coarsest level and the kriging variance is overestimated. Eventually,  $m$  points are found and the variance is approximated more accurately. Such a scenario can lead to poor variogram reproduction beyond  $R$ . If a much larger search is specified such as  $R = 2A$  while keeping  $m$  the same, more data are involved earlier on in the realization and variogram reproduction up to and beyond  $A$  is improved. The approximation of the average variogram over the domain is also improved.

Lastly, the type of kriging used influences the quality of realizations. Simple kriging is theoretically the correct method since it is proven to reproduce the covariance between two grid cells in a realization. In practice, other methods are employed such as ordinary kriging and collocated cokriging. With ordinary kriging, the covariance is not reproduced exactly, but can be too high or low. Constraining the kriging weight to a unit sum can lead to data close to a location being estimated receiving too low a weight, or data far away receiving too high a weight. Consider an extreme case with one conditioning data and the first grid cell visited while generating a realization. Also let the search radius equal the variogram range,  $R = A$ , and the data is  $A$  units away from the first cell. Using ordinary kriging, the data receives a weight of one while using simple kriging the weight is zero. The ordinary kriging variance in this case is given by Equation 4, where  $\sigma^2 = 1$  is the global variance,  $\lambda = 1$  is the kriging weight, and  $\eta = -1$  is the Lagrange parameter ensuring the sum constraint.

$$\sigma_{OK}^2 = \sigma^2 - \lambda - \eta = 2 \quad 4$$

The simple kriging variance  $\sigma_{SK}^2 = 1$ .

Collocated cokriging is another case where realization quality is degraded. Under a Markov screening assumption, only the secondary variable at a location being estimated is retained, leading to variance inflation. The cause is generally due to accumulation of covariance errors between the primary and secondary data and is described by Babak and Deutsch, 2009. The solution is to use intrinsic collocated cokriging that does not suffer from variance inflation. When the Markov approach is used, a variance reduction parameter can be tuned using the SGS assessment algorithm discussed later.

### 3. Methodology

The premise of measuring the quality of realizations from SGS is based on reconstructing the covariance matrix,  $\hat{C}$ , of the realization. An approach to do so is discussed by Emery and Peláez (2011). Once  $\hat{C}$  is available, it is possible to assess a variety of quality indicators including: the overall covariance accuracy; the expected variation in the mean of a set of realizations; the expected variance of each realization; and the expected variogram reproduction.

Reconstructing the covariance matrix is discussed in terms of unconditional simulation using simple kriging, but the approach is similar for other implementations. The approach described by Emery and Peláez (2011) is essentially LU simulation (Davis, 1987), only backwards: instead of populating the covariance matrix and computing the Cholesky factorization to generate realizations, the factorization is generated from a realization to compute the covariance matrix. For sequential simulation, when all previously simulation values are used for each subsequent step the two approaches are equal; however, in realistic cases where the search neighborhood and nearest neighbors are limited, the covariance matrix from the latter is a degraded version of the former.

It is required to generate one realization to rebuild the covariance matrix; however, the result is not dependent on simulated values, but only on the path taken and other input parameters. Multiple realizations do not lead to different results, which can save a significant amount of time for parameter tuning. As a realization progresses, a lower triangular matrix,  $L^{-1}$ , is populated as follows, where  $\sigma_i$  is the kriging standard deviation of the  $i^{th}$  grid node visited in the sequence,  $\lambda_{ji}$ ,  $j = 1, \dots, i - 1$  is the kriging weight assigned to previously simulated node  $j$ , and  $N$  is the number of grid nodes.

$$L^{-1} = \begin{bmatrix} 1/\sigma_1 & 0 & 0 & \dots & 0 \\ -\lambda_{12}/\sigma_2 & 1/\sigma_2 & 0 & \dots & 0 \\ -\lambda_{13}/\sigma_3 & -\lambda_{23}/\sigma_3 & 1/\sigma_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\lambda_{1N}/\sigma_N & -\lambda_{2N}/\sigma_N & -\lambda_{3N}/\sigma_N & \dots & 1/\sigma_N \end{bmatrix} \quad 5$$

A potential issue is that the matrix is  $N \times N$  and is impractical for realistic problems; however, the quality of SGS realizations is not significantly tied to the grid, but is more dependent on the size of the domain covered by the grid. It is possible to obtain a measure of quality on a much coarser grid and this is related to numerical approximation of the average variogram. For example, a cubic domain of 50 units and a spherical variogram of 10 units converges with increasing  $N$  according to Table 1.

Once  $L^{-1}$  is populated, the covariance matrix is reconstructed using Equation 6, which is the Cholesky decomposition of  $\hat{C}$ . Computing the inverse of  $L^{-1}$  is straightforward and not overly computationally demanding; however, the matrix multiplication requires a significant amount of computations and was parallelized to minimize execution time.

$$\hat{C} = (L^{-1})^{-1} ((L^{-1})^{-1})^T = LL^T \quad 6$$

**Table 1: Convergence of average variogram with increasing  $N$ .**

$N$	1	8	27	64	125	216
$\bar{\gamma}$	0.001299	0.875081	0.962979	0.98438	0.992002	0.994465
$N$	343	512	729	1000	1331	
$\bar{\gamma}$	0.995438	0.995836	0.995981	0.996035	0.996159	

Having an estimate of the covariance matrix of the realizations, it is possible to measure the error with the actual covariance based on the input variogram model. Emery and Peláez (2011) use the Frobenius norm defined by Equation 7, where  $C$  is the true covariance matrix and  $\|C\|$  is the square root of the sum of the squared entries.

$$\varepsilon = \frac{\|\hat{C} - C\|}{\|C\|} \quad 7$$

An acceptable amount of error is not clear and may be problem dependent. Ideally, the target error should be zero; however, to be practical one should target values between 0 and 5 %.

It is also possible to assess the expected variogram and error at each lag from the covariance matrix estimate. Each entry in  $\hat{C}$  represent a pair of grid cells that are separated by a lag vector,  $\mathbf{u}$ . All entries in  $\hat{C}$  corresponding to a specified lag can be summarized by a mean, variance and range indicating the expected variation in the variogram from a set of realizations.

The variance of the mean of a set of realizations can also be computed based on the expansion of Equation 8, where  $\mu$  is the mean of a realization defined by Equation 9 and  $\bar{\mu}$  is the mean of the realization means defined by Equation 10.  $y_i$  is the value of grid cell  $i$  and  $M$  is the number of realizations.

$$Var\{\bar{\mu}\} = E\{\bar{\mu}^2\} - E\{\bar{\mu}\}^2 \quad 8$$

$$\mu = \frac{1}{N} \sum_{i=1}^N Y_i \quad 9$$

$$\bar{\mu} = \frac{1}{M} \sum_{m=1}^M \mu_m = \frac{1}{MN} \sum_{m=1}^M \sum_{i=1}^N y_i \quad 10$$

For Gaussian random fields,  $E\{\bar{\mu}\} = 0$ . Expanding the first term in Equation 8 and simplifying yields Equation 11, which is the average estimated covariance over the domain.

$$\text{Var}\{\bar{\mu}\} = E\{\bar{\mu}^2\} = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \hat{C}_{ij} \quad 11$$

In a similar manner, an equation for the expected variance of a set of realizations is obtained starting from Equation 12, where  $\bar{\sigma}^2$  is the mean of the realization variances,  $\sigma^2$ .

$$E\{\bar{\sigma}^2\} = \frac{1}{M} \sum_{m=1}^M \sigma_m^2 = \frac{1}{MN} \sum_{m=1}^M \sum_{i=1}^N (y_i - \mu_m)^2 \quad 12$$

Expanding and simplifying, the expression is given by Equation 13, where  $\bar{C}_i$  is the average covariance between grid cell  $i$  and all other grid cells, that is, the average of row  $i$  of  $\hat{C}$ , and  $\bar{C}$  is the average covariance equal to Equation 11

$$E\{\bar{\sigma}^2\} = \frac{1}{N} \sum_{i=1}^N (C_{ii} - 2\bar{C}_i + \bar{C}) \quad 13$$

It is not possible to assess the variance of the variance of a set of realizations based solely on the covariance matrix as there is fourth order moments involved. Covariance is a second order moment. The four features that are derived for a set of realizations generated from SGS are used together to measure the expected quality of the realizations.

#### 4. Program and Parameters

A program called SGSACCURACY was written to perform the assessments suggested in the previous section. The primary purpose is to determine the statistics of a set of realizations without generating them and use the results to tune parameters and input statistics. Parameters are identical to those for SGSIM (Deutsch and Journel, 1998) with one additional parameter that is the name of an output file for a report that is generated (Table 2). A convenience to using the same parameter file as SGSIM is that both programs can be executed using the same parameter file. Having the additional line for the report file will not affect SGSIM. After tweaking parameters using SGSACCURACY, the actual set of realizations can be generated using the same parameter file and SGSIM.

**Table 2: Parameters for SGSACCURACY.**

Line	Parameter	
1	data.dat	-file with data
2	1 2 0 3 0 0	- columns for X,Y,Z,vr,wt,sec.var.
3	-10.0 1.0e21	- trimming limits
4	0	-transform the data (0=no, 1=yes)
5	sgsim.trn	- file for output trans table
6	0	- consider ref. dist (0=no, 1=yes)
7	histsmth.out	- file with ref. dist distribution
8	1 2	- columns for vr and wt
9	0.0 15.0	- zmin,zmax(tail extrapolation)
10	1 0.0	- lower tail option, parameter
11	1 15.0	- upper tail option, parameter
12	1	-debugging level: 0,1,2,3
13	sgsim.dbg	-file for debugging output
14	sgsim.out	-file for simulation output
15	1	-number of realizations to generate
16	30 1.08 2.16	-nx,xmn,xsiz

17	30	1.08	2.16		-ny, ymn, ysiz	
18	1	0.5	1.0		-nz, zmn, zsiz	
19	69069				-random number seed	
20	0	36			-min and max original data for sim	
21	20				-number of simulated nodes to use	
22	1				-assign data to nodes (0=no, 1=yes)	
23	1	3			-multiple grid search (0=no, 1=yes), num	
24	0				-maximum data per octant (0=not used)	
25	65.0	65.0	10.0		-maximum search radii (hmax, hmin, vert)	
26	0.0	0.0	0.0		-angles for search ellipsoid	
27	51	51	11		-size of covariance lookup table	
28	1	0.60	1.0		-ktype: 0=SK, 1=OK, 2=LVM, 3=EXDR, 4=COLC	
29	../data/ydata.dat				- file with LVM, EXDR, or COLC variable	
30	4				- column for secondary variable	
31	1	0.0			-nst, nugget effect	
32	1	1.0	0.0	0.0	0.0	-it, cc, ang1, ang2, ang3
33			10.0	10.0	10.0	-a_hmax, a_hmin, a_vert
34	report.out				-file for report output	

Most parameters will not have an effect on the realization quality measures including normal score transform (Line 4); number of realizations (Line 15); and the size of the covariance lookup table (Line 27), although larger tables generally lead to faster program execution. Parameters that will effect results include the presence of conditioning data (Line 1); the grid specification (Lines 16 – 18); the number of data to use for kriging (Line 20 and 21); the random path (Line 23) and octant search (Line 24); the search ellipsoid (Line 25 and 26); the kriging type (Line 28) and the variogram (Line 31 – 33).

The random number seed has an influence on the random path; however, it will only lead to minor numerical differences and does not influence the results. Regardless of the number of realizations entered, the accuracy program will only generate one to reconstruct the covariance matrix. In this initial version, execution time limits the problem size, but as mentioned, results are influenced primarily by the domain size, not the discretization. A small test was setup to confirm this statement in 2D and 3D. The number of grid cells was varied for the same domain size of 50 units square and 50 units cubed. An isotropic spherical variogram with zero nugget-effect and range 10 units was used. Nearest neighbors was set to 50 and the search was as large as the domain. A graph of covariance error as a function of  $N^{1/D}$  shows the problem converges to a stable value for a discretization of 25 in 2D (625 nodes) and roughly 12 in 3D (1,728 nodes), see Figure 1. Execution times are shown in Figure 2 with the matrix multiplication done in parallel using 4 processors. For the number of nodes giving convergence, the execution times are reasonable (< 1 minute); however, these results are likely dependent on the parameters going into SGSIM. In cases where more nodes are needed, the execution time quickly becomes unreasonable. Note that the errors are higher in the 3D case. To bring the error down to the same level as the 2D case, excessive previously simulated nodes are needed in kriging (200 to 300 nodes); however, the variogram reproduction using less (50 nodes) was good. This is due to the challenge in reproducing the variogram in a single realization versus reproducing the variogram in expected value.

Contrary to many algorithms where error decreases as  $N$  increases, the opposite is observed for assessing error in the estimated covariance matrix. When  $N < k$ , where  $k$  is the number of nearest neighbors used, all previously simulated nodes are used and the error is zero. As  $N$  increases beyond  $k$ , an estimate of the actual error is approached. In any case, the matrix multiplication requires considerable computational time and it is not recommended to exceed  $N = 10000$ . This is somewhat of a limitation, although one may consider performing checks in 2D slices at various orientations for large 3D problems, or checking smaller 3D sub-domains for any local variation. Running the accuracy check with a coarser grid permits the variogram to be checked at coarser lags as well and if higher resolution is needed, a smaller domain must be considered. This is a point of future work.

Output from the program is a report file that contains a review of the critical parameters, an overall assessment including covariance error, expected variation in the mean, the expected variance, and expected error in variogram reproduction in  $x$ ,  $y$  and  $z$ . If an error seems large, some recommendations are made, and lastly data for plotting expected variograms with error bars is output. A sample report file is shown for a 2D domain of 50 units square, an exponential variogram with range 30, and using 20 nearest neighbors for kriging (Table 3). In this

case, the non-ergodic domain is identified by a high variance of realization means and low expected variance. The accuracy is good even with a low number of neighbors, which is often the case with the exponential variogram due to its excellent screening properties.

**Table 3: Sample SGSACCURACY report.**

---

SEQUENTIAL GAUSSIAN SIMULATION REPORT

Review of critical parameters:

Simulation: UNCONDITIONAL  
 Kriging type: SIMPLE  
 Number of previously simulated nodes: 20  
 Multigrid: YES. Number of refinements: 3  
 Octant search: NO

Review of overall assessment

Covariance error: 0.019409  
 Expected variance of realization means = 0.146720  
 Expected variance of realizations = 0.852964  
 Variogram reproduction SSE in x = 0.000139  
 Variogram reproduction SSE in y = 0.000175

Comments and Recommendations

The overall accuracy appears to be good  
 The expected variance of the mean is high or  
 the expected variance of the realizations is low:  
 - This is indicative of a domain that is small (non-ergodic)  
 relative to the range of correlation of the variogram.  
 This is not incorrect, but SGSIM will not yield zero mean  
 unit variance realizations

Variogram Reproduction Assessment

X Direction

Lag	Distance	True_Variogram	Expected_Variogram	Variance	Minimum	Maximum
0	0.000	0.000000	0.000315	0.161E-06	-.00012	0.00220
1	2.500	0.221199	0.221502	0.158E-06	0.22104	0.22331
2	5.000	0.393469	0.393381	0.639E-06	0.38749	0.39592
3	7.500	0.527633	0.525882	0.631E-05	0.51468	0.53051
4	10.000	0.632121	0.628610	0.139E-04	0.61589	0.63575
5	12.500	0.713495	0.708782	0.176E-04	0.69444	0.71702

---

An example of an expected variogram with error bars is shown in Figure 3. In this example, a spherical variogram with range 30 was used in a domain 50 units square. The number of neighbors was set to 10 to cause noticeable errors, while the search was set to the domain so long range variogram reproduction should be maintained. Errors are low near the origin, but grow as the range extends beyond the 10 nearest neighbors. The expected variogram falls below the input model and this is caused by a non-ergodic domain. If the search radius is constrained to a small value, for example 10 units, variogram reproduction is poor even with 50 neighbors (Figure 4).

## 6. Conclusions and Future Work

A utility was developed to permit the assessment of realizations from SGS without having to generate them. It is based on reconstructing the covariance matrix of a grid based on parameters such as nearest neighbors and random path. It allows users to tune parameters for SGS prior to building realizations so that their quality is known ahead of time. Unfortunately, the algorithm is very computationally demanding and this limits the grid dimensions that can be evaluated. It is possible to obtain an understanding of how SGS will perform for a given set of parameters using a coarser grid. An area for future research is to develop approximations to the measures of realization quality that do not require computing the  $N \times N$  covariance matrix. Another point of future work is to make it possible for the program to parse other parameter files such as for kriging or ultimate SGSIM.



**References**

Alabert, F.G., 1987, Stochastic imaging of spatial distributions using hard and soft information. M.Sc. Thesis, Stanford University, Stanford, CA, 169

Asmussen, S., Glynn, P.W., 2007 Stochastic simulation: algorithms and analysis. Springer Science+Business Media, 482

Babak, O., Deutsch, C.V., 2009. Improved spatial modeling by merging multiple secondary data for intrinsic collocated cokriging. Journal of Petroleum Science and Engineering 69, 93-99.

Boisvert, J.B., Manchuk, J.G., Deutsch, C.V., 2009, Kriging in the presence of locally varying anisotropy using non-Euclidean distance. Mathematical Geosciences, 41(5), 585-601

Chiles, J.P., Delfiner, P. (1999), Geostatistics: modeling spatial uncertainty. John Wiley and Sons, 720

Davis, M.W., 1987, Production of conditional simulations via the LU triangular decomposition of the covariance matrix. Mathematical Geology 19(2): 91-98

Deutsch, C.V., Journel, A.G., 1998, GSLIB: Geostatistical software library and user's guide. Oxford University Press, 384

Emery, X., Peláez, M., 2011, Assessing the accuracy of sequential Gaussian simulation and cosimulation. Computational Geoscience, DOI 10.1007/s10596-011-9235-5, Springer Science+Business Media, 17 pages

Gomez-Hernandez, J.J., Cassiraga, E.F. (1994) Theory and practice of sequential simulation. Geostatistical Simulations, M. Armstrong and P.A. Dowd (eds), Kluwer, Dordrecht, Netherlands, 111-124

Guo H., Deutsch, C.V., 2008, Choosing an adequate number of conditioning data for kriging. Centre for Computational Geostatistics, Report 10, Paper 122, 4 pages

Villalba, M.E., Deutsch, C.V., 2010, Expected ergodic fluctuations in the mean. Centre for Computations Geostatistics, Report 12, Paper 110, 8 pages

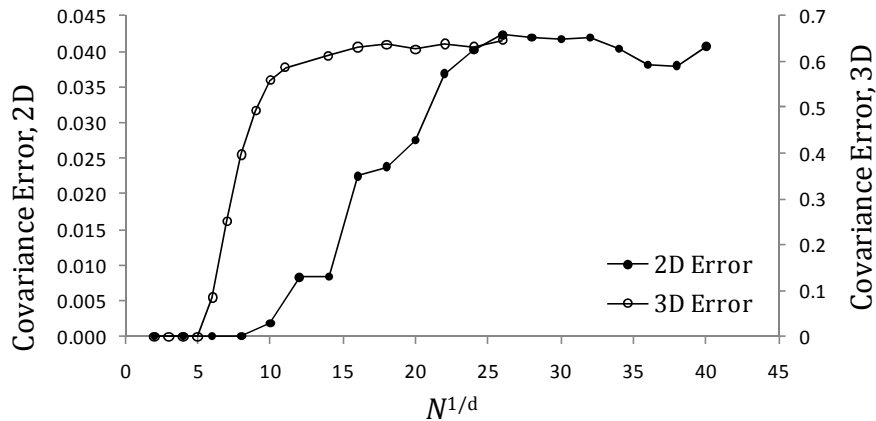


Figure 1: Covariance error as a function of  $N^{1/d}$ , where  $d$  is the dimensionality.

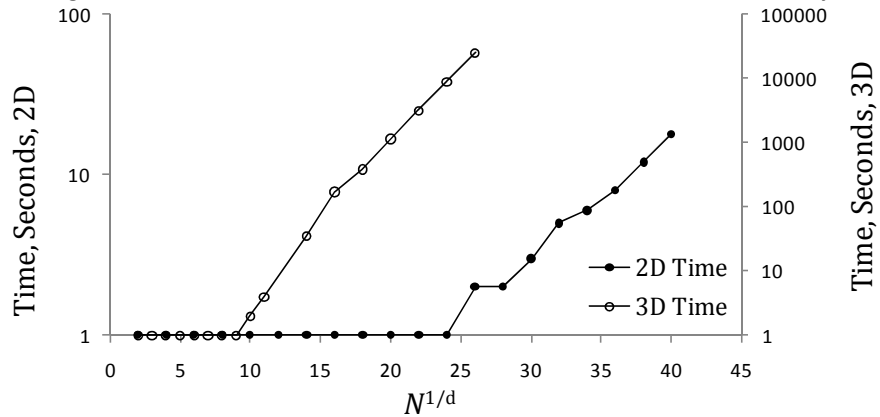


Figure 2: Execution time using 4 processors for computing the covariance error as a function of  $N^{1/d}$ , where  $d$  is the dimensionality.

Figures

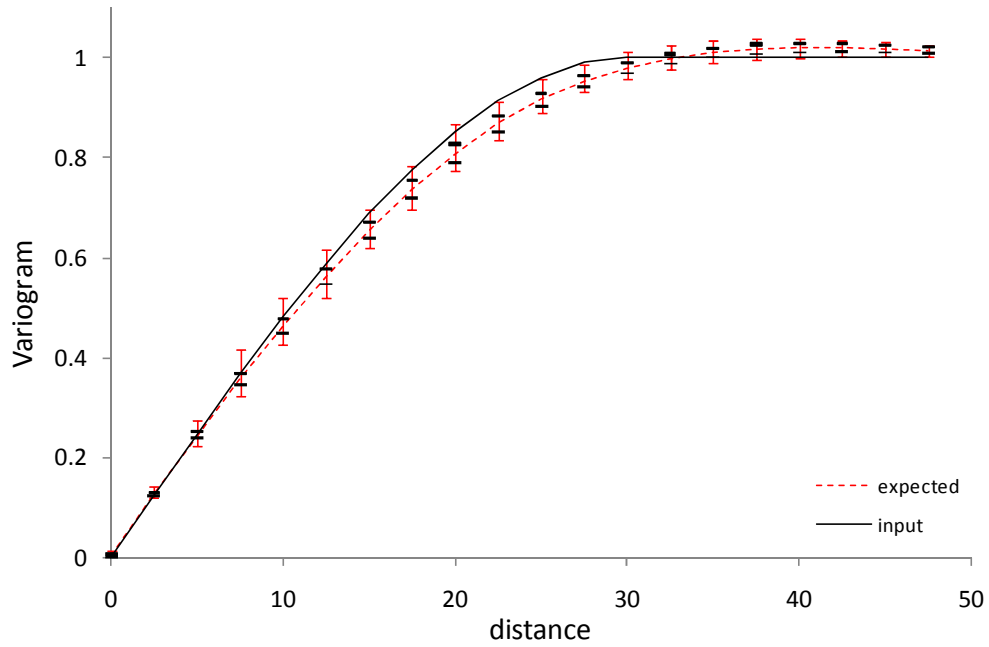


Figure 3: Example plot of expected variogram from a set of realizations compared to the input model. Error bars and +/- 1 standard deviation are shown for the expected variogram.

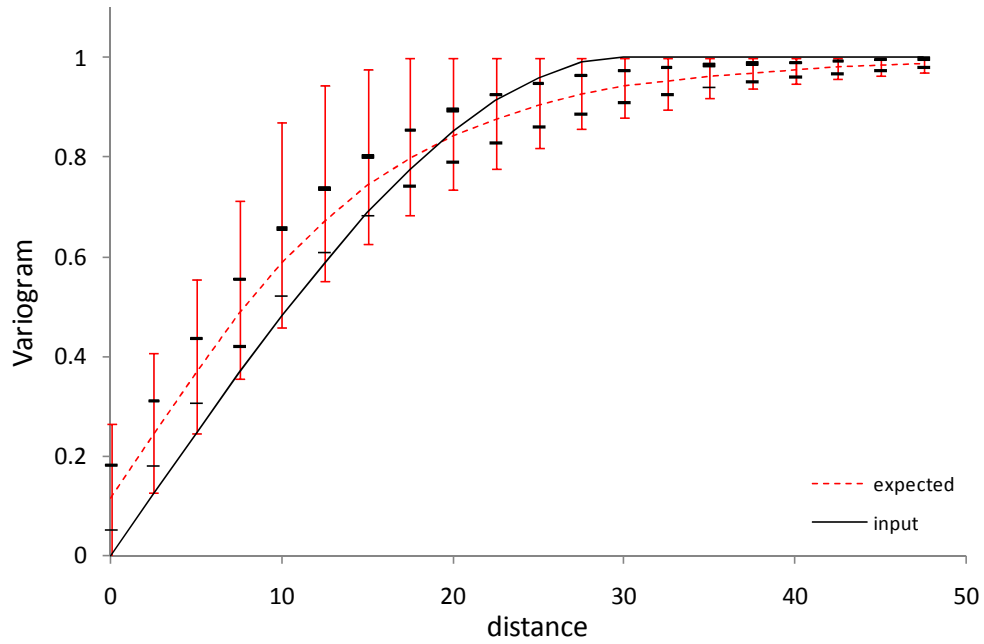


Figure 4: Example plot of expected variogram from a set of realizations compared to the input model. Error bars and +/- 1 standard deviation are shown for the expected variogram.