

UltimateSGSIM: Non-Stationary Sequential Gaussian Cosimulation by Rock Type

Clayton V. Deutsch (cdeutsch@ualberta.ca) and Stefan Zanon (szanon@ualberta.ca)
University of Alberta, Edmonton, Alberta, CANADA

The sequential Gaussian simulation (SGS) algorithm is popular because of its remarkable robustness; however, the flexibility offered by SGS is rarely used as much as it could be. The UltimateSGSIM program was generalized from the latest FORTRAN 90 SGSIM program to provide practitioners with access to some additional flexibility. Some of the special features of UltimateSGSIM include: (1) simulation by rock type accounting for hard and soft boundaries between rock types; one integer coded rock type can be set to “missing” and that rock type “keyed out” and not simulated, (2) multiple variables are simulated simultaneously by either (i) full cokriging, (ii) collocated cokriging with a Markov model, or (iii) after stepwise conditional transformation, (3) the mean values of any or all of the different variables can be made locally variable; the self-healing algorithm can be applied with this locally variable mean option (or with the collocated cokriging option), (4) the three angles that define anisotropy can be made locally variable to improve reproduction of large-scale curvilinear features, and (5) the correlation coefficients in collocated cokriging can be made locally variable.

Sequential Gaussian Simulation

Alabert, 1987, devised sequential simulation under Journel’s supervision. Isaaks, 1990, developed the Gaussian alternative also under Journel’s supervision. The first edition of GSLIB (Deutsch and Journel, 1992) provided the public-domain code `sgsim` that was widely adopted as a stand-alone program and within commercial software. The well established “recipe” of sequential Gaussian simulation:

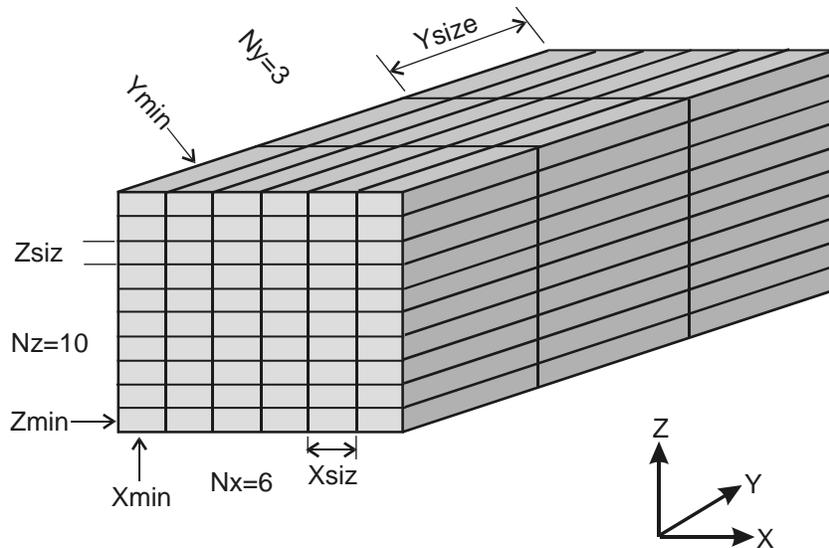
1. Transform the data to Gaussian (normal) deviates: $y(\mathbf{u}_i) = G^{-1}(F_Z(z(\mathbf{u}_i)))$. Where $y(\mathbf{u}_i)$ is the Gaussian deviate at location \mathbf{u}_i , following a standard normal distribution G , and $z(\mathbf{u}_i)$ is the original data value following distribution F_Z .
2. Loop over all grid nodes in a random order:
 - i. Search for nearby data and all previously simulated values.
 - ii. Calculate the conditional mean y^* and variance σ_{SK}^2 by simple kriging.
 - iii. Draw a value from the non-standard normal conditional distribution $N(y^*, \sigma_{SK}^2)$ by Monte Carlo simulation.
3. Back transform all simulated values: $z(\mathbf{u}) = F_Z^{-1}(G(y(\mathbf{u})))$.
4. Repeat the entire procedure with a different random number seed (that changes both the random path and the Monte Carlo simulation) for another realization.

The sequential aspect of this simulation algorithm amounts to a decomposition of the multivariate distribution into a series of univariate conditional distributions. It is exact. In practice, not all values within the range of correlation are used for kriging. This leads to a Markov-type approximation that is not a limitation provided that sufficient conditioning data are retained.

Some common extensions to this core algorithm are cosimulation with multiple variables. The collocated alternative is commonly used (Xu and others, 1992; Verly, 1993; Journel, 1999; Shmaryan and Journel, 1999). A locally varying mean is also used to account for a secondary variable (Deutsch and Journel, 1998). Locally varying directions of continuity (Deutsch and Lewis, 1992; Xu, 1994) and locally varying correlation coefficients (Lyll and Deutsch, 2000) have also been implemented. These features and others will be combined into a version of SGSIM dubbed `UltimateSGSIM`. The programming language is FORTRAN90 with dynamic memory allocation. Source code is provided to facilitate integration into other software and to modify the code even further. The old version of SGSIM will no longer be maintained.

Basics: Grid and Searching

`UltimateSGSIM` works with the standard grid node centered Cartesian grid that all other GSLIB programs use; however, any or all of the coordinates could be the result of some geometric transformation that accounts for stratigraphic/geological correlation.



The Z-axis increases upwards. “Depth” could be considered the Z coordinate with all units increasing downwards; however, the angles must be corrected and the output ordering must be understood (effectively written from the top downwards).

There are three angles that define anisotropy. The original X/Y/Z axes are associated to Horizontal Minor (X), Horizontal Major (Y), and Vertical directions (Z). The search and variogram radii are specified for *hmax*, *hmin*, and *vert*, which are essentially the Y/X/Z directions, unless the angles are rotated. There are three angles that rotate the search ellipsoid or variogram distance calculation to any arbitrary orientation. The first angle (*ang1*) rotates the X/Y coordinates around the Z-axis; clockwise is positive. The second angle (*ang2*) rotates the Y/Z coordinates around the rotated X-axis; clockwise is negative. The third angle (*ang3*) rotates the Z/X coordinates around the rotated Y-axis; clockwise is positive. Any or all of these three angles can be made locally variable in `UltimateSGSIM`.

The grid for `UltimateSGSIM` should be relatively fine compared to the ultimate block grid required for calculations; an appropriate scale-up is required. For this reason, data are assigned to grid nodes. All data outside the grid volume are not used for conditioning, but could be used to construct the data distribution for normal scores transform. The search for previously simulated grid nodes and original data assigned to grid nodes is the same: a spiral search using the anisotropy of the search ellipsoid.

A multiple grid scheme is recommended where the grid nodes are simulated on successively finer grids; searching and data are assigned at the finest grid regardless of the multiple grid search level. The multiple grid search helps reproduce the large range variogram (Tran, 1996).

An octant search could be considered if the data fall along strings from drill holes or wells; otherwise, the string effect or screening of kriging can sometimes be seen in the simulated models showing too much variability away from the strings.

The data for simulation must come from the same rock type or an allowed rock type. A “logic matrix” specifies the boundaries between data and rock types. A “hard” boundary is appropriate in certain geological settings where there is no correlation across rock type boundaries. “Soft” boundaries may be appropriate when there are gradational boundaries across rock types. Soft boundaries may be one-way or two-way, that is, a sample in r can be used for a grid node in rock type r' but a sample in r' does not have to be used in the simulation of a grid node in rock type r . All allowable data of the same type (regardless of what rock type) are treated the same in kriging. Note that the same data value may have different normal score transform values depending on the rock type being estimated. `UltimateSGSIM` keeps track of this.

In cokriging, multiple data types can be used. Each type is searched separately leading to some number of each different data type. `UltimateSGSIM` allows one pre-existing secondary variable coming from, say, seismic data. The user has the responsibility to reconcile multiple pre-existing secondary variables into one with some type of regression or neural network.

The 0th variable is the pre-existing variable (if available). In presence of K variables there are $n_k, k=0, \dots, K$ data available at each grid node: $y_k(\mathbf{u}_{jk}), j_k=1, \dots, n_k$. The next step in `UltimateSGSIM` is to determine the mean and variance values of the conditional normal distributions.

Basics: Transformation

The variables (secondary variable and all K variables) often require transformation to Gaussian or normal score values. Conventional normal scores transformation requires each variable (in each rock type) to be transformed to a Gaussian distribution. Consider K variables and R rock types:

$$\begin{aligned} Y_{1,r} &= G^{-1}[F_{1,r}(z_{1,r})] \\ Y_{2,r} &= G^{-1}[F_{2,r}(z_{2,r})] \\ &\vdots \\ Y_{K,r} &= G^{-1}[F_{K,r}(z_{K,r})] \end{aligned} \quad r=1, \dots, R \quad (1)$$

A polynomial approximation is used for the Gaussian distribution: G/G^{-1} . The original variable distributions $\{F_{k,r}, k=1, \dots, K, r=1, \dots, R\}$ are specified by either the conditioning data or as reference histograms from some other source. The values making up these distributions can be weighted. Linear interpolation/extrapolation are used in the back transformation. The more esoteric power law and hyperbolic options available in `GSLIB` are for the dangerous situation of few data (<20), which is not recommended for `UltimateSGSIM`.

A stepwise conditional transform can be applied to simplify cosimulation and to account for non-linear, heteroscedastic, and constrained relationships (Leuangthong and Deutsch, 2001). This transformation requires sufficient data (approximately 20^K data per rock type) to accomplish the transformation with a series of conditional transformations.

$$\begin{aligned}
Y_{1,r}' &= G^{-1}[F_{1,r}(z_{1,r})] \\
Y_{2,r}' &= G^{-1}[F_{2,r|1,r}(z_{2,r} | z_{1,r})] \\
&\vdots \\
Y_{K,r}' &= G^{-1}[F_{K,r|1,\dots,K-1,r}(z_{K,r} | z_{1,r}, \dots, z_{K-1,r})]
\end{aligned} \quad r = 1, \dots, R \quad (2)$$

Note that the K variograms required in case of stepwise transformation are of the transformed variables $\{Y'_{k,r}, k=1, \dots, K, r=1, \dots, R\}$ and not the original variables.

Basics: Kriging

Once the nearby data have been identified, simple kriging is the default type of kriging:

$$y_k^*(\mathbf{u}) = \sum_{i=1}^{n_k} \lambda_{i,k} \cdot y_k(\mathbf{u}_{i,k}) \quad (3)$$

The weights and kriging variance are calculated by the well-known kriging equations. Ordinary kriging is an option under trend modeling. The sum of the weights is constrained to one to rely less on the global mean of 0. The kriging variance is larger because of the requirement to estimate the mean from the local data. The larger kriging variance also causes the global variance of the simulated values to increase, which in turn causes the global histogram not to be reproduced. The ad-hoc self-healing algorithm could be used to keep the variance under control (Zanon, et. al., 2002). Self-healing can also be used with the local mean option.

$$y_{LVM,k}^*(\mathbf{u}) - m_k(\mathbf{u}) = \sum_{i=1}^{n_k} \lambda_{i,k}^{LVM} \cdot (y_k(\mathbf{u}_{i,k}) - m_k(\mathbf{u}_i)) \quad (4)$$

The non-stationary mean could be used for any subset of variables and rock types. The input mean values are transformed in the same way as the input data for that variable and rock type.

In presence of multiple variables, there are four approaches to account for the secondary variable:

1. *Independent*: each variable is simulated independently, that is, equation 3 or 4 is applied K times successively.
2. *Collocated*: each variable is simulated in turn using all previous collocated variables as conditioning.
3. *Stepwise*: the variables are transformed in a stepwise conditional manner and then back transformed in reverse order to preserve the correct correlation. The transformed variables are essentially simulated independently.
4. *Full Cokriging*: a linear model of coregionalization is used to simulate the variables with the specified correlation structure.

The collocated cokriging alternative may also require self-healing to correct for a too-large variance. The variance values generated by kriging may be modified by the self healing algorithm (see Zanon et. al. 2002 for details), which may result in the variogram not being reproduced.

Full cokriging is probably theoretically the best approach, but it requires all of the variograms to be fit with a licit positive definite linear model of coregionalization. Computer-aided fitting may make this more practical than conventional wisdom tells us. The full cokriging estimator of variable k' is written:

$$y_{k'}^*(\mathbf{u}) = \sum_{k=0}^K \sum_{i=1}^{n_k} \lambda_{i,k}^{k'} \cdot y_k(\mathbf{u}_{i,k}) \quad (5)$$

The cokriging weights $\lambda_{i,k}^{k'}$ are different for every variable k' ; however, the left-hand side kriging system is the same. `UltimateSGSIM` takes advantage of this and solves for all of the kriging weights simultaneously using multiple weight and right hand side covariance vectors.

Collocated cokriging and full cokriging generate a K -dimensional mean vector \mathbf{m} and a K -dimensional standard deviation vector \mathbf{s} . A vector of simulated values \mathbf{y}_{CS} must be drawn from these non-standard Gaussian distributions that reproduces the correlation at lag $\mathbf{h}=0$. This joint simulation is done in two steps:

1. Simulate a vector of standard Gaussian values \mathbf{y} with the correct correlation by LU simulation:

$$\mathbf{y} = \mathbf{L} \bullet \mathbf{w} \quad (6)$$

where \mathbf{L} is the result of LU decomposition using the $\mathbf{h}=0$ correlation matrix and \mathbf{w} is a vector of independent Gaussian deviates.

2. Convert the standard Gaussian values to non-standard deviates:

$$\mathbf{y}_{CS} = \mathbf{y} \bullet \mathbf{s} + \mathbf{m} \quad (7)$$

There are different ways to accomplish the joint drawing of non-standard Gaussian deviates; however, this method is efficient many drawings from different non-standard distributions can be achieved with a single LU decomposition. A proof that this is correct is contained in the Appendix.

Parameters

The parameter file could be very large, which would make it difficult to use the program for simple exercises. To make `UltimateSGSIM` easier to operate, the parameter file has been divided into a number of required and optional blocks. The required input parameters include the following blocks:

- **MAIN**: number of realizations, random number seed, grid size, and output options.
- **SRCH**: search anisotropy and radii, number of data, and other search parameters.
- **VARG**: variogram definition for at least one variable and rock type.

The optional blocks:

- **DATA**: specification of conditioning data file.
- **TRAN**: details of transformation including limits and reference distributions.
- **TRND**: specification of non-stationary parameters such as the mean and anisotropy.
- **SECV**: secondary variable data file.
- **ROCK**: valid rock types, rock type model, and logic matrix.
- **MULT**: specification of how to treat multiple variables.

`UltimateSGSIM` reads through the parameter file three times to get the **MAIN**, **SRCH**, and **VARG** parameters. Defaults are assigned for all optional blocks; then, it reads through for the parameters in the optional blocks. Finally, the program may have to read the variograms again depending on the number of rock types and variables. Very little is written to the screen, which could optionally be a file. Most parameters are echoed to the debugging file.

The **MAIN** block of parameters:

```

1                -number of realizations to generate
1                -number of variables being simulated
69069           -random number seed
50  0.5  1.0    -nx,xmn,xsiz
50  0.5  1.0    -ny,ymn,ysiz
1   0.0  1.0    -nz,zmn,zsiz
sgsim.out      -file for simulation output
0              - output format: (0=reg, 1=coord, 2=bin)
0              -debugging level: 0,1,2,3
sgsim.dbg      -file for debugging output

```

The number of variables to simulate must be specified up front. The other new parameter from the original **SGSIM** is the output format: option 0 is as before with multiple variables as multiple columns, option 1 will also put the X/Y/Z coordinates in the output file, and option 2 will write the values in binary (cycling fastest over the number of variables, then X, then Y, then Z, and finally the realization number).

The **SRCH** block of parameters:

```

12              -number of simulated nodes to use
1              -multiple grid search (0=no, 1=yes)
0  3            -octant search (0=no, 1=yes), number
10.0  10.0  10.0 -maximum search radii (hmax,hmin,vert)
0.0  0.0  0.0  -angles for search ellipsoid

```

There is no two-part search. The data are assigned to the closest grid node unless they fall outside the grid in which case they are not used. The size of the covariance lookup table is automatically calculated and allocated.

The **VARG** block of parameters:

```

1                -number of variograms
0  1  1          -rock type, variable 1, variable 2
2  0.4          - number of structures, nugget effect
1  0.25  0.0  0.0  0.0 - type,variance,ang1,ang2,ang3
                    50.0  50.0  50.0 - a_hmax, a_hmin, a_vert
1  0.35  0.0  0.0  0.0 - type,variance,ang1,ang2,ang3
                    1000  1000  1000 - a_hmax, a_hmin, a_vert

```

The specification of a variogram is as before. There may be many variograms specified for multiple variables and multiple rock types. The rock type, variable 1, and variable 2 specify the specific variogram. The first variogram encountered is the default. The order of the variograms does not matter.

These three blocks (**MAIN**/**SRCH**/**VARG**) are required. Of course, with only these three blocks the simulated realizations must be unconditional and standard normal. The program searches for these blocks first and then attempts to read the remaining

The **DATA** block of parameters is required for conditional simulation:

```

cluster.dat     -file with data
1  2  0  0  0   - columns for X,Y,Z,wt,rock type
3  4            - columns for variables
-1.0           1.0e21 - trimming limits

```

Each data point can have coordinates, declustering weight, rock type, and the variables being simulated. The trimming limits are not for removing or cutting outliers; they are global and are used to remove missing values. The declustering weight applies to all variables. If the data are not equally sampled and a different declustering weight should apply; then, the user has to enter

that through the reference distribution option in the TRAN block. If required, the rock type will come from the 3-D rock type model when the rock type is not contained in the data file.

The **TRAN** block of parameters specifies the transformation and back transformation:

```

1 -transform the data (0=no, 1=yes)
200 -number of quantiles to keep
3 -number of min/max values
1 1 0.0 15.0 - variable, rock type, min, max
1 2 0.0 15.0 - variable, rock type, min, max
1 3 0.0 15.0 - variable, rock type, min, max
1 -number of reference distributions
1 1 - variable number, rock type number
data01.dat - file with ref. distribution
1 0 - columns for value and weight
1 2 - variable number, rock type number
data02.dat - file with ref. distribution
1 0 - columns for value and weight

```

No parameters are needed from this block if the option is to not transform the data. Of course, additional parameters can be left in the parameter file to recall the format in future runs. The tail interpolation/extrapolation is always linear (the `nscore/backtr` programs can be used as pre- and post-processing for `UltimateSGSIM` if a more elaborate option is required. The minimum and maximum values are taken from the input data files unless specified. The user gives the number of minimum and maximum values and then one line per entry with the variable number, rock type number, minimum, and maximum. The distribution will be taken from the data unless a reference distribution is specified. There must be data for every variable in every rock type if the transformation option is turned on.

The **TRND** block of parameters specifies the possibility of trends or non-stationarities:

```

0 -consider ordinary kriging (0=no, 1=yes)
0 -consider self healing (0=no, 1=yes)
0 -consider LVM (0=no, 1=yes)
2 -number of locally varying means
1 1 - rock type number, variable number
lvmf101.dat - file with local mean
1 - column for value
1 2 - rock type number, variable number
lvmf102.dat - file with local mean
1 - column for value
0 -locally varying angles (0=no, 1=yes)
locang.dat - file with local angles
1 0 0 - columns for three angles

```

The default is simple kriging and no non-stationary parameters. If this block is present in the parameter file, then it will be read for the possibility of locally varying (non-stationary) parameters. Ordinary kriging should only be used when the data spacing is dense relative to the variogram range; otherwise, the kriging variances are too high and the global histogram will not be reproduced. Locally varying means can be used for some variables and some rock types. No local mean values will be used if the option is turned off. There is an option for self-healing with each local mean and rock type; this will prevent a too-large variance and lack of histogram reproduction. Any or all of the three angles can be made locally variable. These angles will override the variogram angles specified in the **VARG** block.

The **SECV** block of parameters specifies the existing secondary variables at the same grid resolution as being simulated.

```

0 -consider secondary data (0=no, 1=yes)
sgsim01.dat -file with secondary variable

```

```

1          - column for value
1          -transform secondary data (0=no, 1=yes)

```

The default is no secondary data. If this block is present in the parameter file, then it will be read for the possibility of secondary data coming from remote sensing or some previous simulation.

The **ROCK** block of parameters specifies the rock type model as an optional input. The parameters for this option are:

```

0          -use rock type model (0=no, 1=yes)
3          -number of rock types to model
1 2 3      -the rock types to model
sisim.out  -file with rock types
1          - column number for rock type
1          -consider logic matrix (0=no, 1=yes)
1 1 0      - sample in rock type 1 (0=hard, 1=soft)
1 1 0      - sample in rock type 2 (0=hard, 1=soft)
0 0 1      - sample in rock type 3 (0=hard, 1=soft)

```

The number of rock types to model and their integer numbers are required if a choice is made to use rock types. The file with the rock types must also be present and at the same grid resolution as that being simulated. All rock types not modeled will be left as missing values in the SGS output; multiple rock types and missing values can be “keyed out.” There are no errors if rock types are chosen that are not in the input file; however, there must be corresponding histogram and variogram input for each variable and each rock type specified if the data are being transformed.

The “logic matrix” is used to specify whether or not data from one rock type should intervene in modeling another rock type. The default is for no logic matrix to be considered (option 0) and all rock types are modeled separately. Alternatively, the user can specify the rock types that can be affected by a sample in each rock type. With R rock types, there must be R lines for the logic matrix; one for a sample in each rock type. Each line refers to a sample in one rock type. A 0 means that the sample cannot be used to estimate that particular rock type; a 1 means that it can be used. In the example above, there is a soft boundary between rock types 1 and 2 and rock type 3 is modeled by itself. Note that the boundaries can be soft in one direction. Consider:

```

1          - consider logic matrix (0=no, 1=yes)
1 1 0      - sample in rock type 1 (0=hard, 1=soft)
0 1 0      - sample in rock type 2 (0=hard, 1=soft)
0 0 1      - sample in rock type 3 (0=hard, 1=soft)

```

In this case, a sample in rock type 1 can be used for rock types 1 and 2, but a sample in rock type 2 can only be used in rock type 2.

The **MULT** block of parameters specifies the details of joint simulation of multiple variables. The following input parameters are considered:

```

2          -1=independent, 2=collocated, 3=stepwise, 4=full cokriging
10         -number of classes for stepwise
multivariate.dat -file with transformation data
4 5 6      - columns for variables
0          - column for weight
0.6 0.7 0.5 -correlation coefficients: 0 with 1,2,3,..
          0.4 0.3 -correlation coefficients: 1 with 2,3,..
          0.5      -correlation coefficients: 2 with 3,..

```

Multiple variables may be simulated independently (no correlation), with collocated cokriging, after stepwise conditional transformation, or with full cokriging. The number of secondary classes for stepwise conditional transformation applies to all variables. The order of the variables

is from 1 to K as specified in the input conditioning data file. The number here is so that the correlation

Testing Protocol for Ultimategsim

A recursive testing scheme has been setup for UltimateSGSIM program to ensure that it is properly performing the different types of sequential Gaussian simulation. The GSLIB data is being used to test the different algorithms. In some cases these files will need to be altered to accommodate new simulating methods and/or new data will be used.

The test protocol has been designed to check the statistics for individual and multiple realizations. The results from UltimateSGSIM will be compared to existing programs when available. For a single realization a histogram and pixel plot are created to visually inspect the results and to look at histogram reproduction. Then 51 realizations are created and the mean and variance are calculated at every location over all the realizations. Pixel plots and histograms are created for the mean and variance. The mean map will help to remove the effects of ergodic fluctuations in the process and at each location this should be close to the same value independent of the new and old program for the same kriging type. The variance map shows the degree of ergodic fluctuation at each location providing a sense of uncertainty. At locations of hard data the variance should be zero if the program is reproducing the input data.

Variogram reproduction was calculated for ten of the multiple realizations to show how two point statistics were being reproduced. If required a variogram could be created of the averaged mean map that will be much smoother but this has not been used so far. A scatter plot of the average mean and variance was also created to show the spread of variability for each mean value but this does not provide any new information, it only presents it differently. The final test, so far, is a scatter plot of the primary and secondary data for the single realization to look at correlation reproduction.

The parts of the UltimateSGSIM tested so far are SK, OK, LVM, CCK, and full CK has been started. The main parts left are stepwise, self-healing and kriging using different rock types.

Conclusion

This program fills a need for a flexible version of SGS that permits simultaneous simulation of multiple variables by rock type with non-stationary means and directions. The design is more modular than the original version. Dynamic memory allocation makes the program efficient for a variety of grid sizes.

There are still features that could be included. The piecewise simulation of large grids that was implemented by Julian Ortiz is one example.

References

- Alabert, F.G., (1987) *Stochastic Imaging of Spatial Distributions Using Hard and Soft Information*, Masters Thesis, Stanford University, Stanford, CA.
- Deutsch C.V., Journel A.G. (1998) *GSLIB: Geostatistical Software Library: and User's Guide* Oxford University Press, New York, 2nd Ed.
- Goovaerts, P. (1997) *Geostatistics for natural resources evaluation*, Oxford University Press, New York.
- Isaaks, E. H. (1990) *The Application of Monte Carlo Methods to the Analysis of Spatially Correlated Data*, Ph.D. Thesis, Stanford University.
- Journel, A. G., (1999) Markov Models for Cross-Covariances, *Math. Geology*, 31(8), pp 955-964.
- Shmaryan, L. E. and A. G. Journel, (1999) Two Markov Models and Their Application, *Math. Geology*, 31(8), pp 965-988.
- Xu, W, T. T. Tran, R. M. Srivastava and A. G. Journel. 1992. Integrating Seismic Data in Reservoir Modeling: the Collocated Cokriging Alternative, 1992 SPE Annual Technical Conference and Exhibition, Washington, DC, SPE Paper Number 24742, pp 833-842.
- Verly, G., (1993) Sequential Gaussian Cosimulation: A Simulation Method Integrating Several Types of Information, *Geostatistics Troia '92*, vol 1., ed. A. Soares, Kluwer Academic Publishers, pp. 543-554.
- Zanon, S., T. Faechner, and C.V. Deutsch, (2002) "Improved Integration of Secondary Data using Self-Healing Sequential Gaussian Simulation," 30th International Symposium on Computer Applications in the Mineral Industries (APCOM), February 25-27, 2002, Phoenix, Arizona

Appendix: Drawing from Multivariate Gaussian (From Oy Leuangthong)

Consider n variables, for which simultaneous cokriging yields the parameters of each of the conditional univariate distributions.

$$Y_{c_i} = Y_i \cdot s_i + m_i \quad (1)$$

with

$$Y_i = \sum_{k=1}^n L_{ik} \cdot w_k \quad (2)$$

where w_k , $k=1, \dots, n$ is independent, identically distributed (iid) as $N(0, 1)$, and L_{ik} is the element in the i^{th} row and k^{th} of the lower triangular matrix from the Cholesky decomposition of the correlation matrix at $\mathbf{h} = 0$, denoted simply as $\boldsymbol{\rho}$.

$$\boldsymbol{\rho} = \begin{bmatrix} 1.0 & \cdots & \rho_{1n} \\ \vdots & \ddots & \vdots \\ \rho_{n1} & \cdots & 1.0 \end{bmatrix}$$

So Equation (1) becomes

$$Y_{c_i} = \left(\sum_{k=1}^n L_{ik} \cdot w_k \right) \cdot s_i + m_i \quad (3)$$

The mean of Y_{c_i} is

$$\begin{aligned} E\{Y_{c_i}\} &= E\left\{ \left(\sum_{k=1}^n L_{ik} \cdot w_k \right) \cdot s_i + m_i \right\} \\ &= s_i \cdot \left(\sum_{k=1}^n L_{ik} \cdot E\{w_k\} \right) + m_i \\ &= 0 + m_i \\ &= m_i \end{aligned}$$

The covariance between Y_{c_i} and Y_{c_j} is

$$Cov\{Y_{c_i} Y_{c_j}\} = E\{Y_{c_i} \cdot Y_{c_j}\} - E\{Y_{c_i}\} \cdot E\{Y_{c_j}\}$$

with

$$\begin{aligned} E\{Y_{c_i} \cdot Y_{c_j}\} &= E\left\{ \left[\left(\sum_{k=1}^n L_{ik} \cdot w_k \right) \cdot s_i + m_i \right] \cdot \left[\left(\sum_{l=1}^n L_{jl} \cdot w_l \right) \cdot s_j + m_j \right] \right\} \\ &= s_i s_j \sum_{k=1}^n \sum_{l=1}^n L_{ik} L_{jl} \cdot E\{w_k w_l\} + m_j s_i \sum_{k=1}^n L_{ik} \cdot E\{w_k\} \\ &\quad + m_i s_j \sum_{l=1}^n L_{jl} \cdot E\{w_l\} + m_i m_j \end{aligned} \quad (4)$$

Since $w_k, k=1, \dots, n$ is iid $N(0, 1)$, then $E\{w_k\} = 0, \forall k$ and

$$\begin{aligned} E\{w_k w_l\} &= 1.0, \text{ if } k = l \\ &= 0, \text{ otherwise} \end{aligned}$$

Equation (4) becomes

$$E\{Y_{c_i} \cdot Y_{c_j}\} = s_i s_j \sum_{k=1}^n L_{ik} L_{jk} + m_i m_j$$

The covariance is then

$$\begin{aligned} Cov\{Y_{c_i} Y_{c_j}\} &= s_i s_j \sum_{k=1}^n L_{ik} L_{jk} + m_i m_j - m_i m_j \\ &= s_i s_j \sum_{k=1}^n L_{ik} L_{jk} \end{aligned}$$

Recall from Cholesky Decomposition of the correlation matrix that the correlation between the i^{th} and j^{th} element is given by

$$\rho_{ij} = \sum_{k=1}^n L_{ik} L_{jk}$$

Therefore the covariance between Y_{c_i} and Y_{c_j} is

$$Cov\{Y_{c_i} Y_{c_j}\} = s_i s_j \rho_{ij}$$

which reproduces the covariance matrix at $\mathbf{h}=0$, with s_i and s_j as the standard deviation of the i^{th} and j^{th} variable provided by cokriging and ρ_{ij} is the correlation at $\mathbf{h}=0$.

Recall of Cholesky Decomposition

Consider the following correlation matrix at $\mathbf{h}=0$, denoted simply as $\boldsymbol{\rho}$.

$$\boldsymbol{\rho} = \begin{bmatrix} 1.0 & \cdots & \rho_{1n} \\ \vdots & \ddots & \vdots \\ \rho_{n1} & \cdots & 1.0 \end{bmatrix}$$

Cholesky decomposition yields the following lower triangular matrix, \mathbf{L} (the upper triangular matrix is simply the transpose of the lower triangular matrix, $\mathbf{U}=\mathbf{L}'$):

$$\mathbf{L} = \begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ l_{(n-1)1} & & \ddots & 0 \\ l_{n1} & \cdots & \cdots & l_{nn} \end{bmatrix}$$

that is, for $n=3$, \mathbf{L} becomes:

$$\mathbf{L} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix}$$

To determine the initial correlation matrix based on the lower matrix, all that is required is to multiply the lower and the upper triangular matrices together:

$$\rho = \mathbf{LU}$$

$$\rho = \mathbf{LL}'$$

Correspondingly, the i^{th} and j^{th} element of this correlation matrix is obtained by

$$\rho_{ij} = \sum_{k=1}^n l_{ik}l_{jk}$$

Parameters for UltimateSGSIM

```

START OF MAIN:
1          -number of realizations to generate
1          -number of variables being simulated
69069     -random number seed
50  0.5   1.0   -nx,xmn,xsiz
50  0.5   1.0   -ny,ymn,ysiz
1   0.0   1.0   -nz,zmn,zsiz
sgsim.out -file for simulation output
0          - output format: (0=reg, 1=coord, 2=bin)
4          -debugging level: 0,1,2,3
sgsim.dbg -file for debugging output

START OF SRCH:
12        -number of data to use per variable
1         -multiple grid search (0=no, 1=yes)
0         3   -octant search (0=no, 1=yes), number
10.0  10.0  10.0 -maximum search radii (hmax,hmin,vert)
0.0   0.0   0.0 -angles for search ellipsoid

START OF VARG:
1         -number of variograms
1  1  1   -rock type, variable 1, variable 2
2   0.10 - number of structures, nugget effect
1   0.25 0.0  0.0  0.0 - type,variance,ang1,ang2,ang3
           5.0  5.0  5.0 - a_hmax, a_hmin, a_vert
1   0.65 0.0  0.0  0.0 - type,variance,ang1,ang2,ang3
           25.0 25.0 25.0 - a_hmax, a_hmin, a_vert
2  1  1   -rock type, variable 1, variable 2
2   0.90 - number of structures, nugget effect
1   0.05 0.0  0.0  0.0 - type,variance,ang1,ang2,ang3
           5.0  5.0  5.0 - a_hmax, a_hmin, a_vert
1   0.05 0.0  0.0  0.0 - type,variance,ang1,ang2,ang3
           25.0 25.0 25.0 - a_hmax, a_hmin, a_vert

START OF DATA:
cluster.dat -file with data
1  2  0  0  6 - columns for X,Y,Z,wt,rock type
3  4  5   - columns for variables
-1.0      1.0e21 - trimming limits

START OF TRAN:
1         -transform the data (0=no, 1=yes)
200      -number of quantiles to keep (default=200)
3        -number of min/max values
1  1  0.0  15.0 - rock type, variable, min, max
2  1  0.0  300.0 - rock type, variable, min, max
1  3  0.0  15.0 - rock type, variable, min, max
0        -number of reference distributions
1  1     - rock type number, variable number
cluster.dat - file with ref. distribution
3  0     - columns for value and weight
1  2     - rock type number, variable number
data02.dat - file with ref. distribution
1  0     - columns for value and weight

```

continued on next page...

```

START OF TRND:
0 -consider ordinary kriging (0=no, 1=yes)
0 -consider self healing (0=no, 1=yes)
0 -consider LVM (0=no, 1=yes)
2 -number of locally varying means
1 1 - rock type number, variable number
lvmfl01.dat - file with local mean
1 - column for value
1 2 - rock type number, variable number
lvmfl02.dat - file with local mean
1 - column for value
0 -locally varying angles (0=no, 1=yes)
locang.dat - file with local angles
1 0 0 - columns for three angles

START OF SECV:
0 -consider secondary data (0=no, 1=yes)
sgsim01.dat -file with secondary variable
1 - column for value
1 -transform secondary data (0=no, 1=yes)

START OF ROCK:
0 -use rock type model (0=no, 1=yes)
2 -number of rock types to model
21 22 -the rock types to model
Working/rock.out -file with rock types
1 - column number with rock type
0 -consider logic matrix (0=no, 1=yes)
1 1 0 - sample in rock type 1 (0=hard, 1=soft)
1 1 0 - sample in rock type 2 (0=hard, 1=soft)
0 0 1 - sample in rock type 3 (0=hard, 1=soft)

START OF MULT:
2 -1=independent, 2=collocated, 3=stepwise, 4=full cokriging
10 -number of classes for stepwise
multivariate.dat -file with transformation data
4 5 6 - columns for variables
0 - column for weight
0.6 0.7 0.5 -correlation coefficients: 0 with 1,2,3,..
0.4 0.3 -correlation coefficients: 1 with 2,3,..
0.5 -correlation coefficients: 2 with 3,..

```

Figure 1: Example parameter file for UltimateSGSIM with all parameter blocks