Direct Geostatistical Simulation on Unstructured Grids

John Manchuk, Oy Leuangthong and Clayton V. Deutsch

Department of Civil & Environmental Engineering,
University of Alberta

Abstract

Unstructured grids are commonly used in reservoir modeling and are being increasingly considered in complex mining engineering applications. Block kriging of the attributes can be easily implemented; however, this implicitly assumes linear averaging, which is not the case after Gaussian transformation or with variables that do not average linearly such as permeability. Direct simulation has been proposed as a solution; however, there are a number of important implementation considerations. This paper addresses the following considerations and develops a practical algorithm/software: (1) search for nearby relevant block and point data, (2) stabilization of the kriging equations and weights in presence of complex screening, (3) correction of the homoscedastic kriging variance to account for realistic proportional effect, (4) determination of valid conditional distribution shapes, (5) accounting for geological controls including stratigraphic surfaces and mixture of multiple facies within an unstructured grid block, and (6) accounting for directional permeability that does not average linearly.

Introduction

The use of unstructured grids, particularly in petroleum reservoir modeling, is increasing in popularity. Structured grids, commonly defined using Cartesian coordinates, lack the ability to model complex reservoir geometries such as external boundaries and internal faults [1]. Unstructured grids are being utilized to model the complex geology and geometry of reservoirs and to provide improved accuracy (a more refined grid) to more important areas. Grids such as tartan grids are used to provide a high cell density perhaps near wells and low cell density in less influential areas.

Sequential Gaussian simulation (SGS) is one of the most extensively used algorithms for continuous variable simulation; however, it is impractical when considering unstructured grids or data that come from significantly different volumes (such as seismic or production-related data). SGS relies on the multivariate Gaussian model where all conditional distributions are Gaussian and fully described by a mean and variance calculated by the well-known simple kriging equations. Unstructured grids introduce a multitude of support volumes to deal with and the variables will almost certainly not average linearly through Gaussian transform, thus SGS cannot be implemented.

There are ways to trick SGS into partially accounting for multiscale data. The covariances can be calculated after transformation. The volume scale can be ignored. The data can be pre-processed to reduce or even increase variability. The fact remains, however, that classical Gaussian techniques assume linear averaging after Gaussian transformation. A significant bias can be incurred if this assumption is adopted. The severity of the bias will be more pronounced when the univariate distribution is highly skewed.

Direct sequential simulation (DSS) [2] is becoming popular due to unstructured grids and the need to implement multiscale data. Because DSS works with data in original units, the multivariate Gaussian assumption is not explicitly required even if it is invoked through the central limit theorem. There is, however, no univariate transformation of the data and no bias in the resulting volume-averaged properties. Non-linear averaging remains an issue; all kriging-based techniques assume linear averaging in the unit system that kriging takes place.

Implementation of DSS is done with the intention of keeping data in its original units and support volumes rather than simulating on a fine (regular) grid and block averaging a posteriori to an irregular grid. Since data will be considered in its original form, larger volumes will have to be discretized to acquire a covariance value and relative to point data, this process can be very time consuming. A fast method of calculating mean covariance values between any two volumes must be developed [3]. Suggested solutions to accelerate mean covariance calculations are to limit the number of discretizations, which may hinder accuracy, or to speed up the calculation using a variogram lookup table [3].

Data used to populate unstructured grids may consist of original data at a small scale or large regularly gridded soft data. The grid volumes being populated may be of many different sizes. There is a need for an efficient search for nearby relevant data. The popular method when dealing with a fixed set of data is the super block search strategy, a variation of which could be applied to unstructured grids. A spiral search is often used when searching on a regular grid. There are also many search trees available to organize data [4, 5]. A type of search tree appealing for simulation is a quadtree for two dimensional data or an octree for three dimensions [6]. An efficient method of indexing and pointing to data prior to simulation is required such that upon simulating, searching is efficient and the data used are relevant.

A major issue when considering multiscale data are the effects of screening. If proper filtering of data prior to kriging is not used, screening can lead to anomalous weights, negative variance, and consequently inaccurate estimates. Some filtering techniques such as the octant search, iterative kriging, and the template technique have been utilized to combat screening [3]. These methods work; however their efficiency is questionable. The octant search should never exhibit screening if only one data per octant was used, however as soon as more than one data per octant is accepted; screening or the string effect is possible. Iterative kriging removes data exhibiting extreme weights; however this data may be important to the estimate. The template technique simply consumes computation time. An efficient method of stabilizing the kriging equations in the presence of complex screening is required.

A common characteristic that many data sets exhibit is a heteroscedastic relationship between the local mean and variance commonly referred to as the proportional effect [7]. Using simple kriging results in a variance that depends only on the data configuration and is independent of the data values, hence the variance is homoscedastic. When kriging with data in original units, the resulting simulated values cannot reproduce a heteroscedastic feature; a method must be developed to account for the proportional effect inherent in original data.

An important step in simulation is drawing a value from the local distributions at each location being simulated. To perform this, the shape of the local distributions must be known, preferably from the kriging mean and variance. A method of determining the local distribution shapes has been developed and will be revisited [3, 8, 9]. Basically, the local distributions are calculated and organized into a lookup table to be accessed using the kriged mean and variance.

Applying unstructured grids introduces two other issues: accounting for geological controls such as stratigraphic surfaces, and; dealing with grid blocks containing a mixture of multiple facies. An unstructured grid may not conform to the stratigraphic setting within it and this introduces problems relating to selecting relevant data for kriging and grid blocks containing multiple subsequence layers. A technique to deal with blocks that do not conform to geological controls and that contain multiple facies and subsequences is needed.

When simulating with data in original units, another problem that will be encountered is variables that do not average linearly, such as permeability. This non linear averaging property becomes a major issue when incorporating multiple support volumes into estimating a location. The location being estimated will likely have an intermediate volume different from that for the conditioning data. These non linear variables must be transformed to average linearly for implementation in DSS. Power law averaging can accomplish this transform [10].

This paper addresses these six important issues and proposes some novel approaches for resolution.

Search for Nearby Relevant Block and Point Data

A problem encountered in geostatistical modeling of unstructured grids is searching for nearby relevant block and point data. The data may consist of original data at a small scale, regularly gridded soft data, and many varying sizes of grid blocks making current methods of indexing and searching for relevant data impractical. There are many different data configurations of grid blocks along with their centroid locations (x,y,z) and geometric characteristics. There could be millions of data points and previously simulated nodes.

There are several ways a search for nearby data can be done. Using a brute force method is possible for small problems and this involves using a matrix of distances ngb by ngb in size, where ngb is the number of grid blocks. This is feasible for small data sets; however larger data sets would require so much memory it would be impractical for conventional computers RAM. Implementing the use of a super block search strategy is another option. The centroids of all grid blocks and data locations would be indexed using the conventional super block search method [15]. Another possible method of indexing large data sets is the use of search trees, which may prove to be more efficient than the other search strategies.

One type of search tree popular for use in computer graphics is a quadtree (2-dimensional) or an octree (3-dimensional) [6]. In the 2-dimensional case, a quadtree could be implemented to organize data so that operations such as point location, region location, and neighbor searches can be done quickly. For applying search trees to geostatistics, nearest neighbor searches and region queries would be important for acquiring conditioning data and previously simulated nodes. Point location operations will be needed for inserting simulated nodes into the search tree. Quadtrees work by taking the initial set of data and dividing it up into quadrants. If the number of points within each quadrant exceeds a specified maximum, they are divided into sub-quadrants. This process continues until all quadrants in the tree contain at most the specified maximum. Octrees work identically to quadtrees except there are 8 regions created when an octant is sub-divided. Quadtrees and octrees are often created such that all regions at a specific level in the tree are the same size.

Another popular type of search tree is a multidimensional binary search tree (kd-tree) that can be used to sort data of higher (k) dimensions [4, 5]. Like the octree or quadtree, data can be sorted by having one or several data points per region. A kd-tree works by splitting the data into two

regions per partition. Each cutting plane's location and orientation are chosen based on the median of the longest aligned axis for each region's data. As the process narrows to meeting a specified requirement of one or multiple data per region, a tree is built based on the median locations. Figure 2 shows a schematic of how the data are separated along with a resulting tree.

Screening occurs when near data shadow further data and also when data are configured into strings. When data are shadowed they receive negative weights in kriging. When there is an abundance of data not strongly clustered, the negative weights are very small relative to other weights and can normally be ignored [13]. They represent only up to a few percent of the total weights. Negative weights become a substantial problem when relatively few data are used to estimate a location [13]. Large negative weights can result and may lead to a negative grade estimate. When data are contiguously aligned finite strings and used in making an estimate, the string effect occurs [14]. Strings of data commonly occur in mining and petroleum applications since data are often collected along drill holes or wells. The string effect occurs due to the end point data seeming less redundant than the central data within the string and results in the weighting patters shown in Figure 3.

Stabilization of the kriging solution in presence of complex screening

Having data sorted and indexed in some manner, perhaps using a search tree, will improve the efficiency of finding data to use in kriging. A set of nearby data to be used for kriging a location can be quickly accessed from a search tree (see the next paper in this report for more details). The resulting data for that location can then be filtered further to reduce the effects of screening and the occurrence of anomalous weights.

Screening can cause extreme positive and negative weights that lead to erroneous estimates and estimation variances. One method of reducing the occurrence of extreme weights is to remove data from the kriging matrix: this iterative kriging technique [3] will remove data until the absolute value of all the weights are below a specified maximum. Iterative kriging works; however, data that may be highly influential in estimating a location could be removed from the kriging matrix resulting in a less accurate result. Another method of reducing screening is the template technique [3] which involves rejecting any data that are shadowed by a closer data. Figure 4 shows the template technique. A downfall to the template technique is its high demand on computation time.

The effects of screening can be observed using an Excel procedure for block kriging. Each volume was made adjustable such that the effects of screening when dealing with various sized volumes could be better understood. The example consists of estimating one location with six block data. A spherical variogram model was used for covariance calculations. Figure 5 shows one possible data configuration and the resulting weights.

To understand the screening effect when considering different volumes, several data configurations were put together and the kriging weights calculated (Appendix A).

An efficient method of filtering data of various support volumes for estimating a location is needed. Using the template technique should eliminate screening; however, the string effect may still occur. The template technique was applied in two dimensions and was very CPU intensive, so applying this to three dimensions would slow the process even more. A new method of filtering data used in estimating a location is the sector search method, which is somewhat similar to the template technique. The sector search method uses input dip and azimuth tolerances to create sectors in which only the nearest data is selected for kriging. A schematic of the sector

search in two dimensions is shown in Figure 6. This method reduces the occurrence of screening and depending on the size of each sector the occurrence of strings of data should be reduced as well.

The sector search subroutine works fast in two dimensions as the sectors are all pre-constructed and then translated around to points being estimated; however in three dimensions, the sectors are built starting with the nearest point with more sectors being built as un-rejected points are encountered making the process more time consuming.

Even though the sector search method removes much screened data, there may be unreasonable screening still present. A case where two points in adjacent sectors with one much closer to the location being estimated than the other will result in screening effects. Using large sectors will practically eliminate screening. To determine an optimal sector size, multiscale data should be used since the occurrence of anomalous weights and estimates is more prevalent. An attempt was made to determine an optimal size using point data; however anomalous weights did not occur regardless of the size of sectors or number of points being used.

Correction of the homoscedastic kriging variance to account for realistic proportional effect

Data in original units are often heteroscedastic; there is increased variability in high values area. This heteroscedastic behavior is commonly referred to as the proportional effect [7, 15]. An assumption in direct sequential simulation is that the kriging variance provides the variance of the local distributions of uncertainty; however the kriging variance depends only on the data configuration and is independent of their magnitude. For data following the congenial Gaussian distribution this assumption is accurate; however for data exhibiting heteroscedastic features, it is unreasonable. Because kriging will be applied directly to data in original units, the kriging variance must be adjusted such that proportional effect is reproduced.

Simple kriging is utilized in DSS because it reproduces the covariance even if the conditional probability distributions are not Gaussian [2]. Covariance reproduction using SK can be easily demonstrated (Appendix B). A problem with SK is that reproduction of the covariance only holds if the variance of the data is homoscedastic. The local distributions of data in original units could be of any shape so SK will remain useful in this context; however the variance of original data will, in most cases, be heteroscedastic. Covariance reproduction only holds if the variance of the data is homoscedastic.

To see the effects of directly simulating data that show the proportional effect, a study was done using a lognormally distributed data set. Lognormal data was chosen because there is a mathematical link between it and the more common Gaussian distribution and an equation describing the proportional effect of lognormal data exists [15]. Lognormal theory is explained in Appendix C. Knowing these relations, the kriging variance can be calibrated to honor the heteroscedasticity inherent in lognormal data.

An exhaustive lognormal data set was generated by transforming an unconditional Gaussian model (Figure 7). The mean and variance of the lognormal data were chosen to be 100 and 10000, respectively. A set of 625 samples was drawn from the model and used for simulation experimentation.

To show the proportional affect inherent in the lognormal model as well as the homoscedastic nature of Gaussian data, plots of the $y(\mathbf{u})$ and $y(\mathbf{u+h})$ values were created using a lag distance

equal to half the variogram range. By dividing the results into fifty quatiles and calculating the mean and standard deviation of each, the plots in Figure 8 were produced.

To apply direct simulation, the SGSIM program [16] was altered such that it could be used to simulate conventionally or directly with lognormal data. For the direct method, there is also the option to use a variance correction (utilizing Equation 10C in Appendix C) or to perform kriging and Monte Carlo simulation (MCS) in a naïve manner with no variance correction. This was done for comparison purposes.

Three options of simulation were explored:

Option 1   Transform a set of lognormal samples to normal space and perform kriging and MCS, then back-transform to lognormal space. This is the standard/common approach.

Option 2   Perform direct kriging with the lognormal values with an adjusted variogram and do MCS without correcting the kriging variance. This is the published approach to DSS. The limitation is that heteroscedasticity/the proportional effect is not accounted for.

Option 3   Perform direct kriging on the lognormal values with an adjusted variogram and correct the kriging variance prior to MCS. This is the new approach that we are advocating in this paper. Multiscale data can be used in direct kriging and the proportional effect is explicitly accounted for.

For each option, 100 realizations were produced and the mean and variance for every location was determined. Figure 9 shows the simulation results for all three options. To check the validity of each simulation method, reproduction of the global statistics as well as the variogram were checked. These are shown in Figure 10.

Figure 9 shows that the resulting mean and variance when a correction is applied is very similar to the SGS results; however when no correction is made as in option 2, the variance is clearly homoscedastic. To better compare the three options, plots of the mean and standard deviation were created to display reproduction of the proportional effect in a realization (Figure 11).

By performing simulation using lognormal data, it was possible to introduce a solution for dealing with the proportional effect. Lognormal data was particularly useful because the proportional effect is one of its prominent features and is analytically known. The proportional effect is a common feature of many data sets in original units and it cannot be reproduced with kriging alone as kriging only related the data configuration, not the data values. The resulting kriging variance is homoscedastic and must be corrected to reproduce the proportional effect.

By imposing a correction to the kriging variance prior to simulation it was possible to perform direct sequential simulation on a lognormal data distribution with positive results. The outcome from applying SGS and DSS with variance correction compare very well. These findings provide insight into a possible solution to DSS when dealing with the proportional effect. It is apparent that the kriging variance must be corrected such that the proportional effect is reproduced with simulation.

In the general case, a relation describing the proportional effect of a data set can be determined prior to simulation using regression (linear or curvilinear). A moving window average or the technique used to produce Figure 8 could be utilized to acquire the data for regression. Having a

relation, one can calculate the variance based on the estimate for all locations. The resulting realizations should reproduce the proportional effect observed in the initial data set.

Determination of valid conditional distribution shapes

The simple kriging system provides an estimate and an estimation variance. Once the estimation variance is corrected if needed (to account for the proportional effect), it along with the estimate are required to determine the shape of the local distribution of uncertainty at a location prior to simulation. With sequential Gaussian simulation, kriging is performed in Gaussian space and all resulting local distributions are parametric; they are Gaussian and fully described by the kriging mean and variance. Throughout simulation, all local distributions used are Gaussian and this ensures that the global distribution is reproduced. A method to determine the shape of the local distributions in original units such that the global distribution is reproduced is required. Figure 12 shows a schematic of the method.

Since the global distribution in direct space (z space) will unlikely represent any type of mathematically known distributions, a method of relating it to a Gaussian distribution is needed. Having this relation, one will be able to develop a large set of local distributions in z space from a set of local distribution in Gaussian or y space. If a specific quantile q of a y space distribution with mean m and standard deviation σ is known, the corresponding direct space quantile can be calculated using the following procedure. Equation 3 is used to acquire the inverse of the nonstandard Gaussian function $G\{m, \sigma\}$ for the quantile value q.

$$Val1 = G^{-1}{}_{\{m, \sigma\}}(q) \tag{3}$$

The corresponding probability value is then calculated from the standard Gaussian distribution, $G_{\{0,1\}}$.

$$Val2 = G_{\{0,1\}}[Val1] \tag{4}$$

$$\text{or,} \quad Val2 = G_{\{0,1\}}[G^{-1}{}_{\{m, \sigma\}}(q)] \tag{5}$$

Utilizing the link between global y space and global z space, the z space value corresponding to the probability, *Val2* can now be calculated.

$$Val3 = F^{-1}[Val2] \tag{6}$$

$$\text{or,} \quad Val3 = F^{-1}[G_{\{0,1\}}[G^{-1}{}_{\{m, \sigma\}}(q)]] \tag{7}$$

Val3 is the z quantile of a distribution of uncertainty associated with the probability value *q* [3].

By creating a series of y space distributions from a list of means and variances and repeating the above procedure for a range of quantiles, a set of z space distributions can be generated. The mean and variance of each z space distribution can be calculated and used as reference values. Upon kriging at a particular location, the resulting mean and variance can be used to reference a distribution from the previously calculated set from which a value can be simulated.

Accounting for geological controls including stratigraphic surfaces and mixture of multiple facies within an unstructured grid block

Some geological settings are characterized by a series of genetically related strata. The geology may consist of a sequence stratigraphic framework; the bounding surfaces between the layers correspond to a specific geologic time that separates two different periods of deposition or a period of erosion followed by deposition [17]. Figure 1 depicts a stratigraphic sequence.

There are issues related to dealing with unstructured grids superimposed on a system of stratigraphic surfaces that control anisotropy or facies boundaries, see Figure 13. Some of the issues shown by Figure 13 are fairly obvious; the grid does not line up with the stratigraphic surfaces; grid blocks may contain multiple facies and subsequences, and; searching for relevant data to estimate unknown locations is a problem. Another consideration is the differences in sample volume relative to the volumes being estimated. Integration of seismic data may also be a consideration. Also, the correlation of data within each subsequence and across multiple subsequences should be considered.

A possible method of dealing with data within various subsequences is to flag the data by subsequence and only use data within genetically related strata. During simulation, only data flagged the same as that being estimated will be used. When simulating blocks crossing multiple subsequences, flagging and simulating its value poses a problem. Perhaps discretizing the block, flagging the smaller components and estimating them to acquire a value or multiple values and structure within a grid block may be feasible. Because blocks may cross into multiple subsequences as well as contain multiple facies, a method of determining which portion of a grid block to use in estimating a different location is needed. What is required is an idea of the subsequence structure within grid blocks being estimated as well as those being used for conditioning data. Figure 14 shows two grid blocks and their structure to better explain why the structure within grid blocks is important in estimating unknown locations. It is possible to estimate proportions of subsequence and facies types within a grid block and retain those, however retaining the structure within each would likely be memory intensive.

Upon estimating grid blocks, the proportions of facies within each can be determined overall, but it may be better to retain the facies proportions within each subsequence within each block. Because storing information relating the subsequence structure within each block would get cumbersome, a more efficient method may be to store the overall subsequence geometry for the entire model and use an indexing method through out the blocks such that the structure for each block could be referenced when needed. Every block would then retain the facies proportions within each subsequence and a pointer telling it what the subsequence geometry is.

Since the data is multiscale, the use of direct sequential simulation will be required. The multiscale data might consist of well data, seismic data, and previously simulated nodes of varying sizes (depending on the grid).

Accounting for directional permeability that does not average linearly

To account for data that do not average linearly, such as permeability, DSS must be implemented. Conventional simulation techniques require the data to be transformed to a Gaussian distribution; however permeability does not average linearly after Gaussian transformation. Because data exist in vastly different scales such as small core based permeability and large scale production data, problems arise due to the scale difference and highly non linear averaging of permeability. By implementing a power law transform, permeability values will average linearly for use in modeling using a direct simulation approach [10]. To deal with unstructured grids, data must be linear with scale and this is accomplished using power law averaging [10].

The general formulae for power law averaging of the continuous or categorical variable K are shown by Equations 8 and 9, respectively.

$$K_{eff} = \left[\frac{1}{v} \int_v k(\mathbf{u})^\omega d\mathbf{u} \right]^{\frac{1}{\omega}} \qquad (8)$$

where v is the volume over which the average is calculated, $k(\mathbf{u})$ is the permeability at location $\mathbf{u}$ in v, and ω is an averaging exponent.

$$K_{eff} = \left[\sum_{i=1}^{n} p_i K_i^\omega \right]^{\frac{1}{\omega}} \qquad (9)$$

where n is the number of classes, p_i is the volume fraction of class i, and K_i is the permeability of class i. Equation 9 can be rewritten as Equation 10 for a system consisting of shale and sandstone.

$$K_{eff} = \left[V_{sh} K_{sh}^\omega + (1 - V_{sh}) K_{ss}^\omega \right]^{\frac{1}{\omega}} \qquad (10)$$

where K_{sh} and K_{ss} are the permeability values of shale and sandstone respectively and V_{sh} is the volume fraction of shale.

Because DSS utilizes kriging as an estimator, the variables being used must average linearly with scale. By using a power law transformation, the problems generated by multiscale data can be avoided and transformed variables will average linearly with scale. In order to transform permeability, an idea of the spatial continuity and geological setting is required. The averaging exponent depends mostly on the spatial features in the formation and not on the distribution of the data. The following procedure is used to determine a ω value: construct multiple permeability realizations based on a specific geological model; determine the effective K_X, K_Y, and K_Z values via flow simulations; calculate the directional ω values, ω_X, ω_Y, and ω_Z based on the flow simulation results; plot a histogram of the directional ω values [10].

A concern of implementing power law transformation especially when dealing with unstructured grids is that ω may not be constant over every volume support. An unstructured grid may involve many different volume support sizes to be estimated and when there is an immense scale difference, ω may change. Other concerns that affect the ω value are arbitrary boundary conditions and if the formation approaches the percolation threshold.

Power law averaging can be used to transform volume dependent variables (that scale non-linearly) to scale linearly. These variables can then be used for modeling using a direct simulation formalism, independent of their volume support. Once simulation is performed, results can be back transformed for further processing.

Conclusions

Multiscale data and unstructured grids are practically relevant for realistic reservoir modeling. Simulating in the units of the original data provides significant benefits such as unbiased accounting for multiscale data and permitting different local distributional shapes. In practice, implementation of DSS has been limited. Even something as seemingly straightforward as searching for data is complicated by the multiscale nature of the problem. In these instances, quadtrees or octrees may be particularly efficient. Grid blocks of different volumes also leads to important screening effects and destabilization of the kriging matrix, thus a preferential filtering of the data through a sector search may be appropriate. Multiscale issues are further complicated when dealing with variables that do not average linearly; a pre-processing power-law transform is a first approximation.

Accounting for small scale facies control is a challenge particularly if the blocks are large with respect to the facies variations. The grid blocks will cross multiple sequence or sub-sequence stratigraphic layers. Despite all these issues, perhaps the most important advance presented in this paper is the correction applied to the SK variance to account for the heteroscedastic nature that is often inherent to real data. The lognormal case was used to illustrate a corrective approach to effectively reproduce heteroscedasticity.

References

[1]   International Reservoir Simulation Research Institute, Brigham Young University, Provo, Utah. http://www.et.byu.edu/cheme/research/IRSRI.html

[2]   W. Xu and A.G. Journel, *DSSIM: A General Sequential Simulation Algorithm,* Stanford Center for Reservoir Forecasting, Stanford, CA, May 1994

[3]   M.J. Pyrcz and C.V. Deutsch, *Building Blocks for Direct Sequential Simulation on Unstructured Grids*, Center for Computational Geostatistics, University of Alberta, Report 4: 2001/2002

[4]   T. Kanungo et al, *An Efficient k-Means Clustering Algorithm: Analysis and Implimentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 7, July 2002

[5]   L. Devroye, J. Jabbour and C. Zamora-Cura, *Squarish k-d Trees*, Society for Industrial and Applied Mathematics, SIAM Journal of Computing, Vol. 30, No. 5, 2000, pp 1678-1700

[6]   S.F. Frisken and R.N. Perry, *Simple and Efficient Traversal Methods for Quadtrees and Octrees*, Mitsubishi Electrical Research Laboratories, 2002

[7]   B. Oz and C. V. Deutsch, *A Short Note on the Proportional Effect and Direct Sequential Simulation*, Centre for Computational Geostatistics, University of Alberta, Report 4: 2001/2002

[8]   C.V. Deutsch, T. Tran, and Y. Xie, *A preliminary report on: An approach to ensure histogram reproduction in direct sequential simulation,* Technical report, Center for Computational Geostatistics, University of Alberta, Edmonton, AB, March 2001

[9]   B. Oz, C. Deutsch, T. Tran, and Y. Xie, *A fortan 90 program for direct sequential simulation with histogram reproduction*, Computers & Geosciences, page submitted, 2001

[10]   S. Zanon, H. Nguyen, and C.V. Deutsch, *Power Law Averaging Revisited*, Center for Computational Geostatistics, University of Alberta, Report 4: 2001/2002

[11]   P. Berman, B. Dasgupta, and S. Muthukrishnan, *Exact Size of Binary Space Partitionings and Improved Rectangle Tiling Algorithms*, SIAM Journal on Discrete Mathematics v 15 no2 Feb/Apr 2002 p. 252-67

[12]   P. Kirschenhofer, C. Martinez, and H. Prodinger, *Analysis of Hoare's FIND Algorithm with Median-of-Three Partition*. April 26, 1996.

[13]   A.J. Sinclair and G.H. Blackwell, *Applied Mineral Inventory Estimation,* Cambridge University Press, 2002

[14]   C.V. Deutsch, *Kriging with Strings of Data*, Mathematical Geology, Vol. 26, No. 5, 1994

[15]   A.G. Journel and Ch.J. Huijbregts, *Mining Geostatistics*, Academic Press Limited, 1978

[16]   C.V. Deutsch and A.G. Journel, *GSLIB Geostatistical Software Library and User's Guide*, Second Edition, Oxford University Press, 1998

[17]   C. V. Deutsch, *Geostatistical Reservoir Modeling*, Oxford University Press Inc., 2002
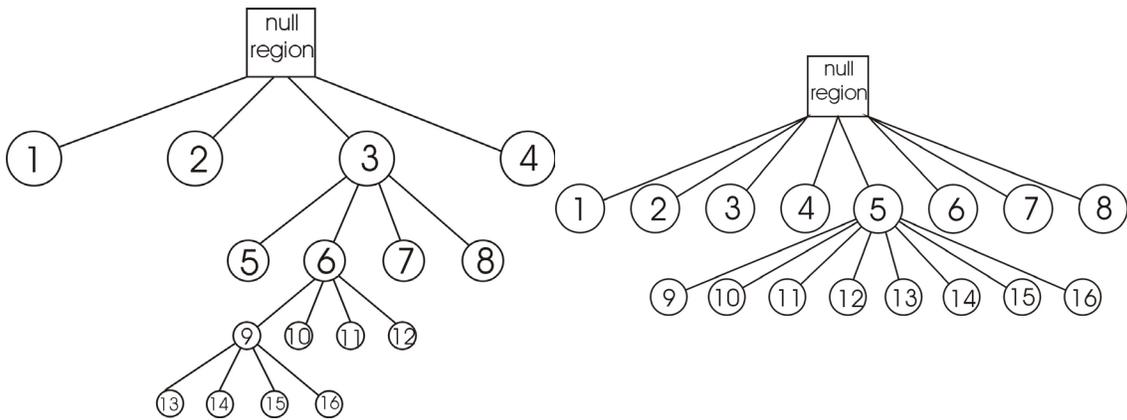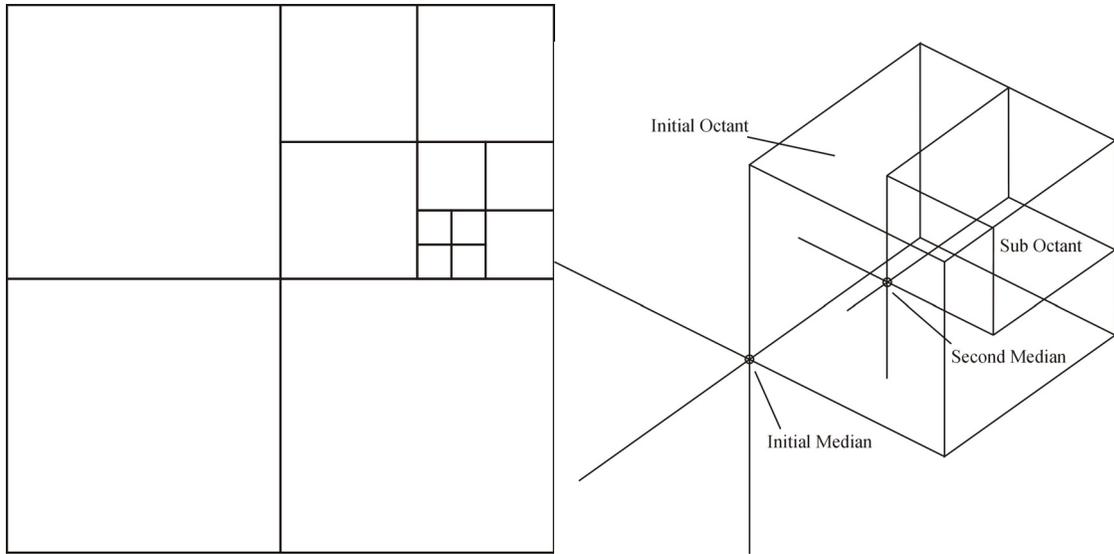
Figure 1: Quadtree (top-left) and octree (top-right) data organization. The corresponding tree representation is shown for each tree structure (bottom). The quadtree is regular whereas the octree is created based on median locations. The leaf nodes (1, 2, 4, 5, 7, 8, 10-12, and 13-16 for the quadtree and 9-16 for the octree) would contain pointers to the data within them. Cells that are split like octant 5 in the octree (bottom-right) contain pointers to the first of a set of children (cell 9).
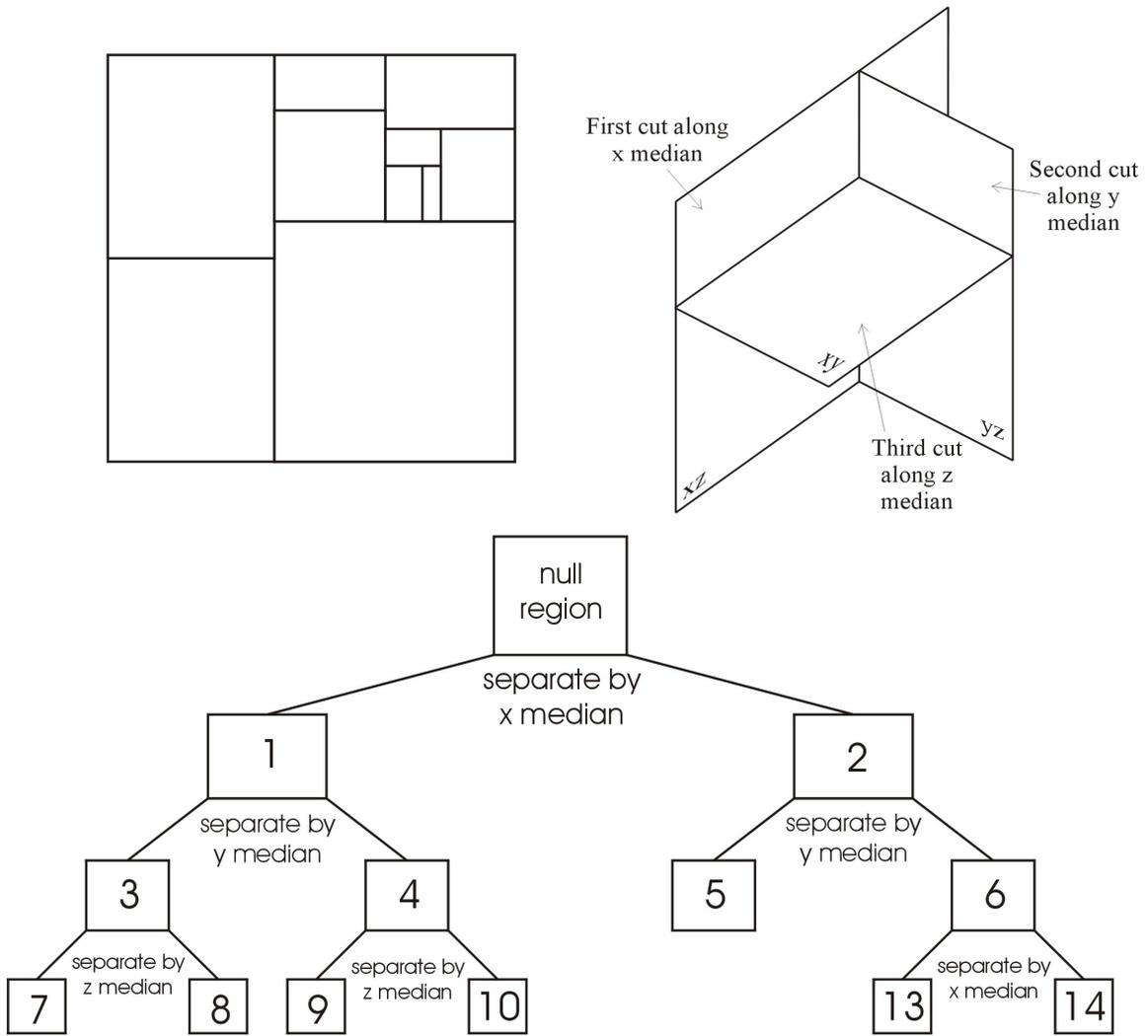
Figure 2: A kd-tree applied to two dimensions (top-left) and three dimensions (top-right) along with a representation of the tree (bottom). The data is partitioned based on median locations until there are at most a specified maximum number of points per region.
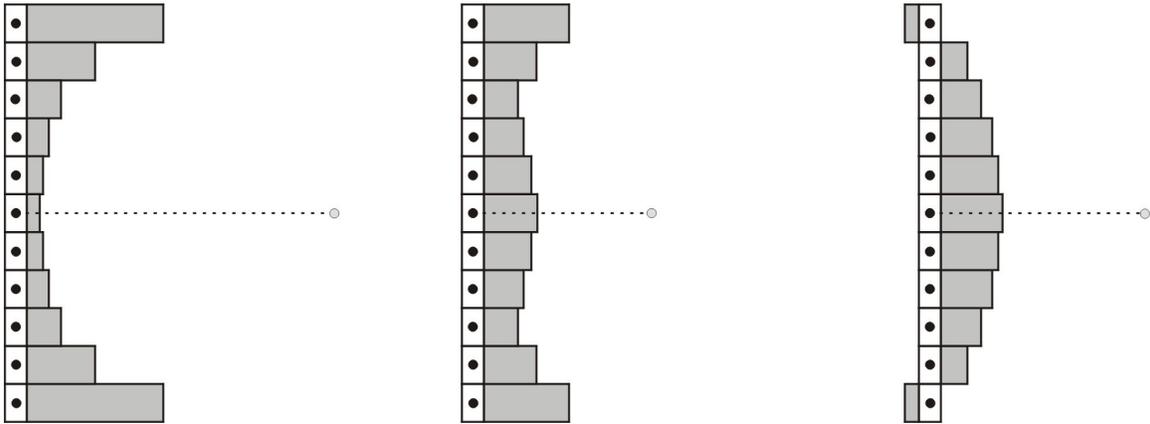
Figure 3: Examples of the string effect. Grey bars depict the kriging weight for that particular data point. Ordinary kriging with the estimation location beyond the variogram range results in the configuration on the left. When the estimate location is within the variogram range, the middle kriging weight configuration occurs. The right image shows the resulting weight configuration if simple kriging is used.
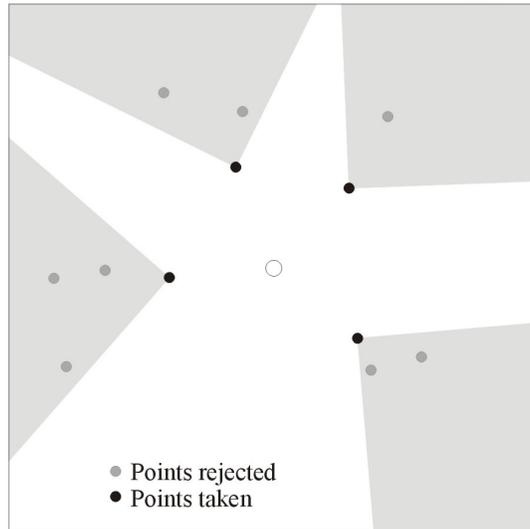
Figure 4: The template technique to reduce screening. White points are already chosen as conditioning data and black points are those rejected. The white square is the location being estimated and it acts like a light source such that the conditioning data create shadows in which all other data is rejected.
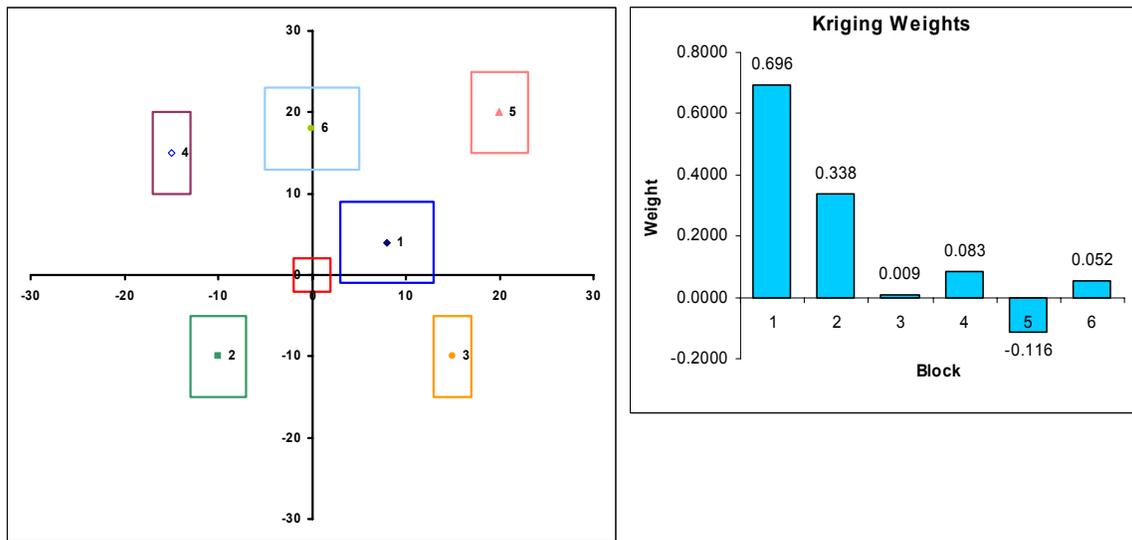
Figure 5: An example data configuration with varying block volumes. Covariance calculations were done using 4 discretizations (64 points per block). Simple kriging was used.
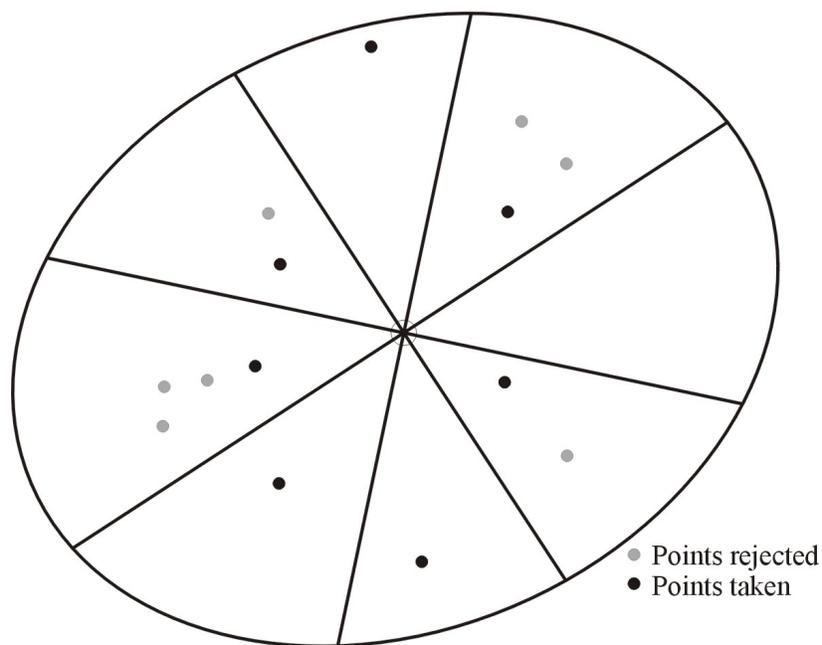
Figure 6: Results using sector search in two dimensions. The azimuth tolerance was set to 18 degrees to produce 20 sectors. Data shown with a bullet will be used in kriging and any denoted by an x were filtered out.
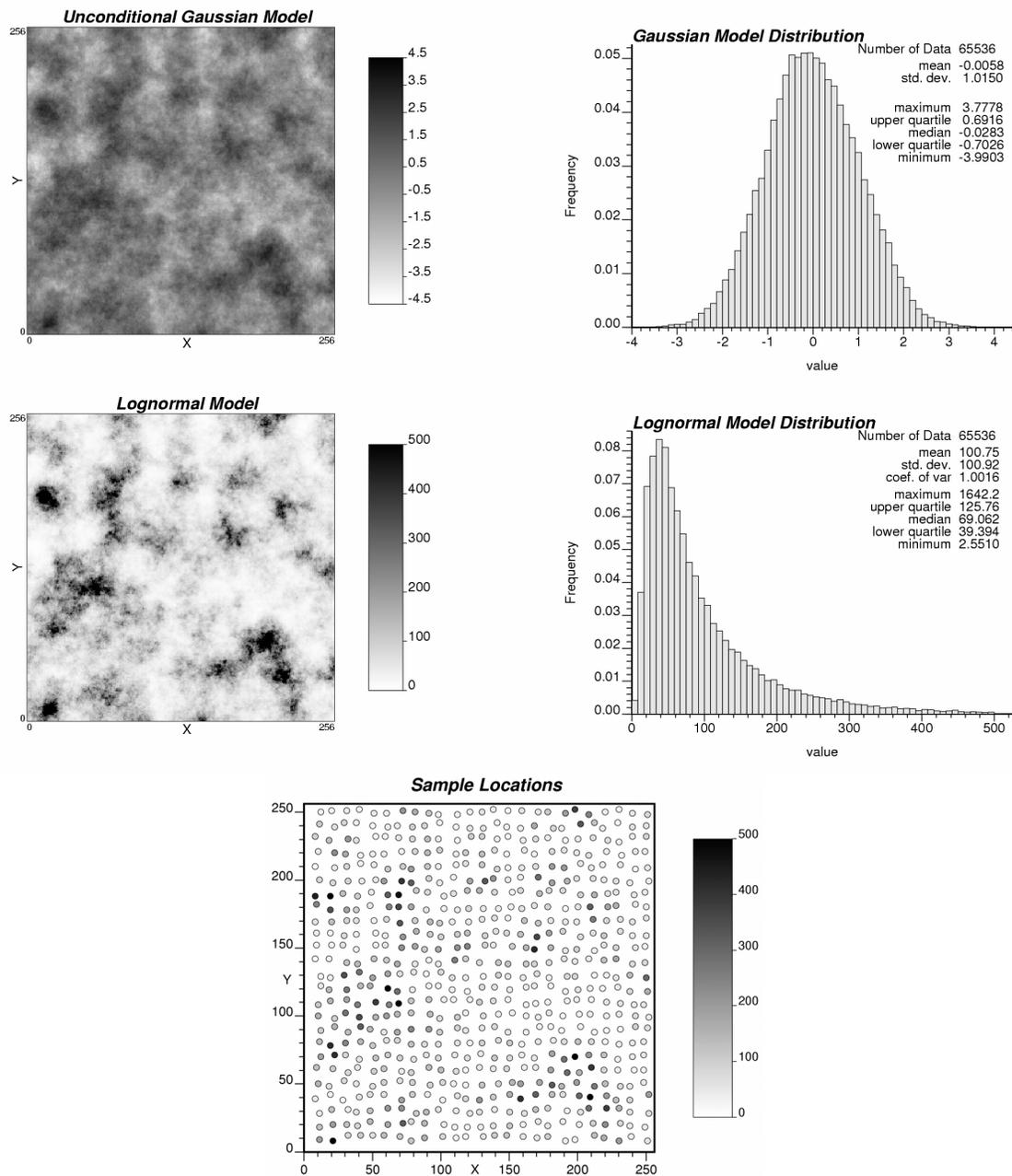
Figure 7: Unconditional Gaussian model and resulting lognormal model after transformation. The sample set used is also shown (bottom).
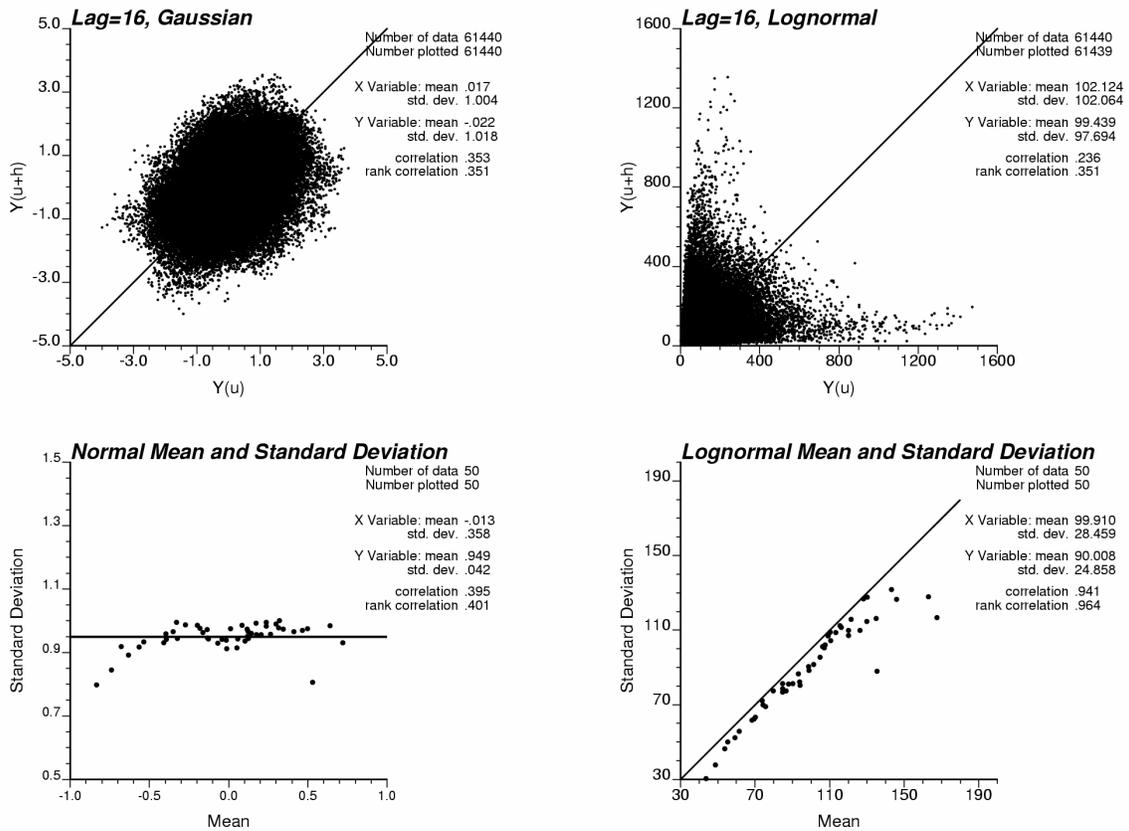
Figure 8: Scatterplots of *Y(u)* versus *Y(u+h)* for Gaussian (upper left) and lognormal (upper right) data. Gaussian data showing the variance is homoscedastic (lower left) and lognormal data displaying the proportional effect (lower right). The analytical line in the lower right plot was determined using Equation 10C in Appendix C.
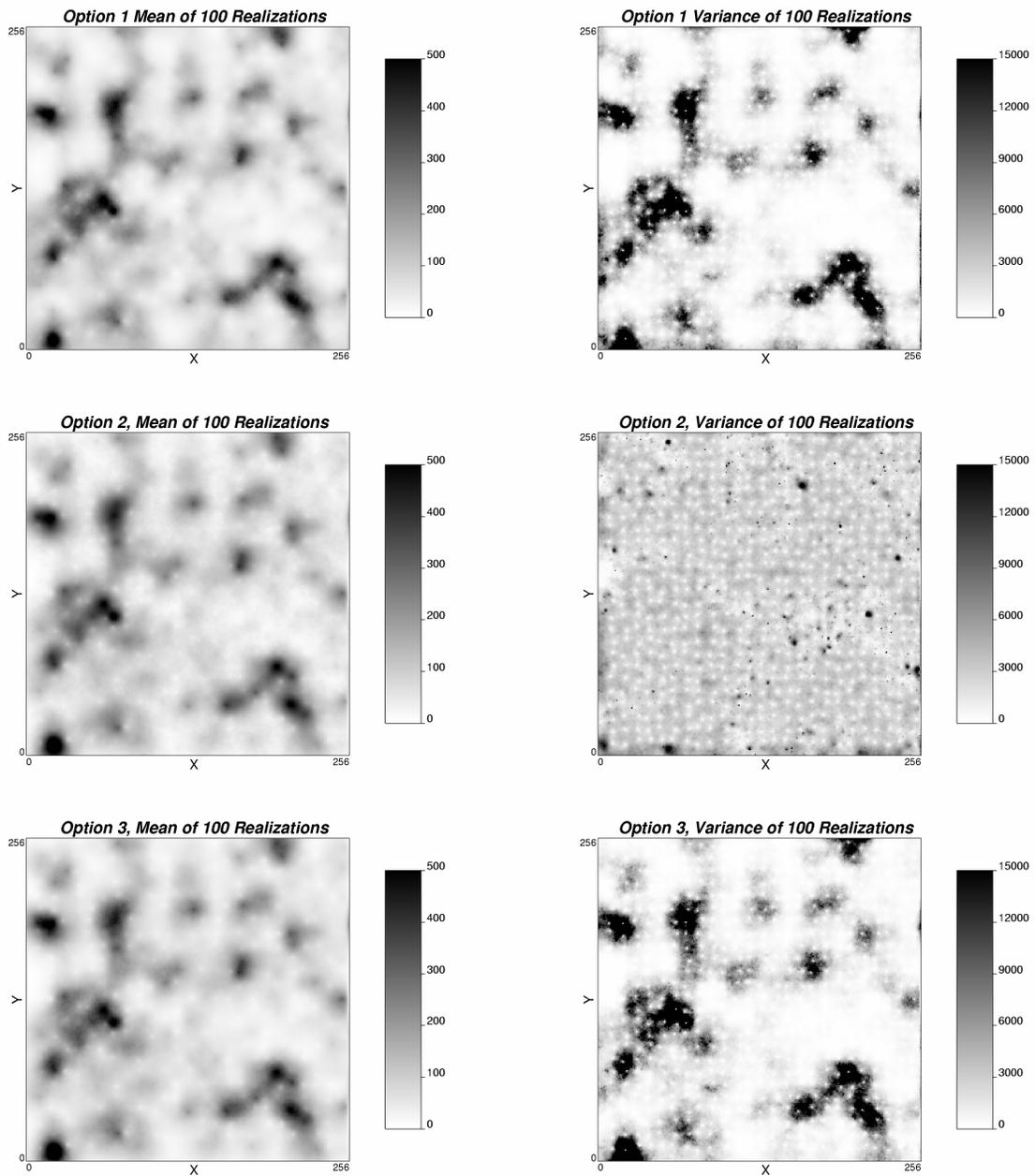
Figure 9: The mean and variance taken over 100 realizations for all three simulation approaches. Top – option 1; traditional method of data transformation prior to kriging and simulation, then back transformation to get results. Middle – Option 2; naïve direct simulation with no variance correction. Bottom – Option 3; direct simulation with a variance correction to account for the proportional effect.
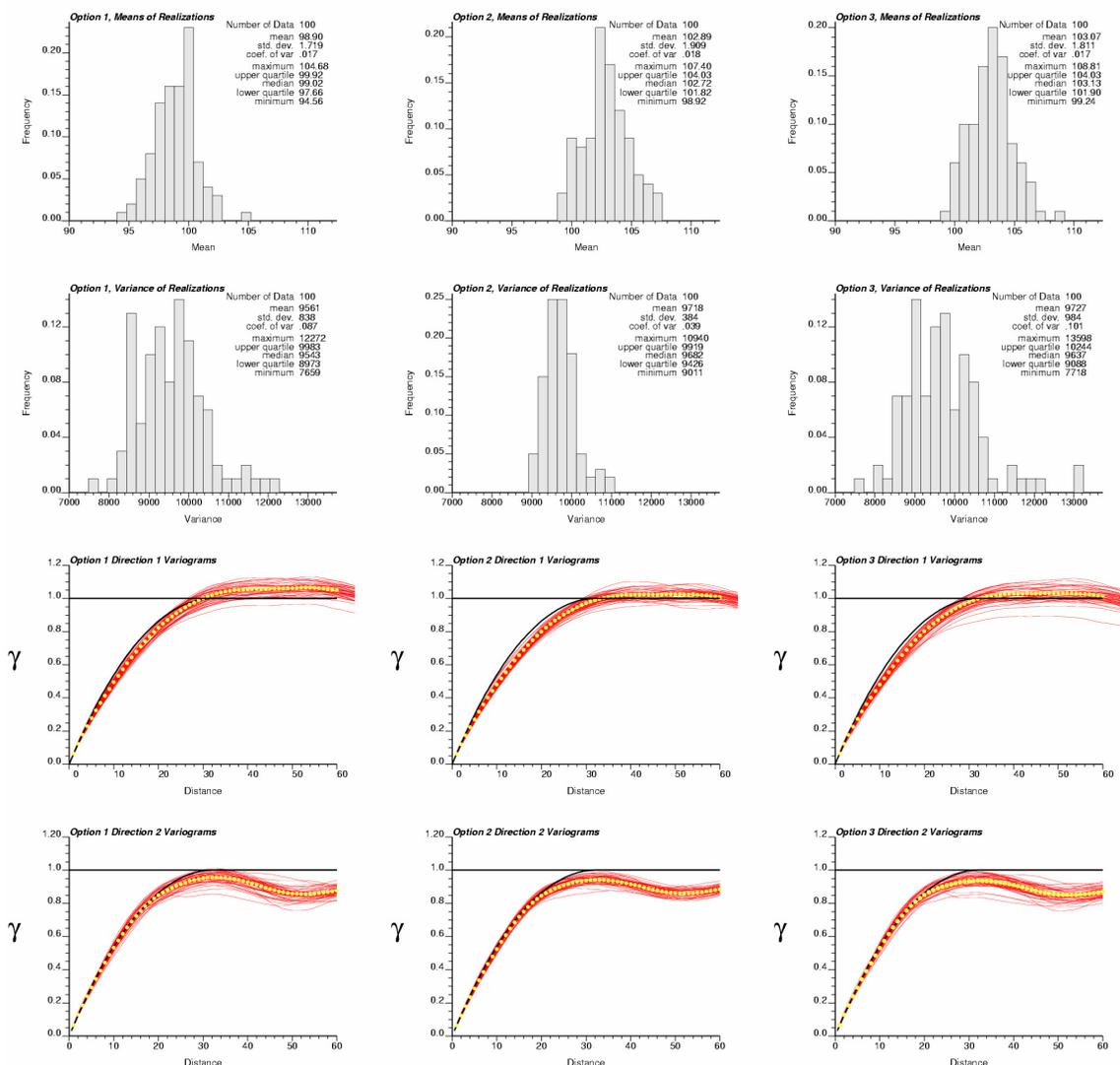
Figure 10: A check for mean, variance and variogram reproduction for the 100 realizations. Top – mean reproduction; both options 2 and 3 result in a slightly higher mean than option 1. Middle – variance reproduction; the three option compare well. Bottom – variogram reproduction; all options are close to following the analytical variogram model. Bullets represent the average of the 100 variograms.
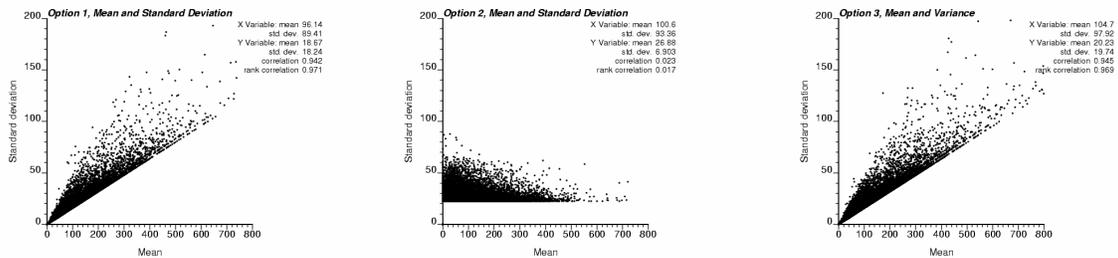
Figure 11: Local mean versus standard deviation at every estimated location for options 1 (left), 2 (middle), and 3 (right). Options 1 and 3 show the proportional effect and compare nicely. Option 2 shows a homoscedastic variance since no correction was applied.
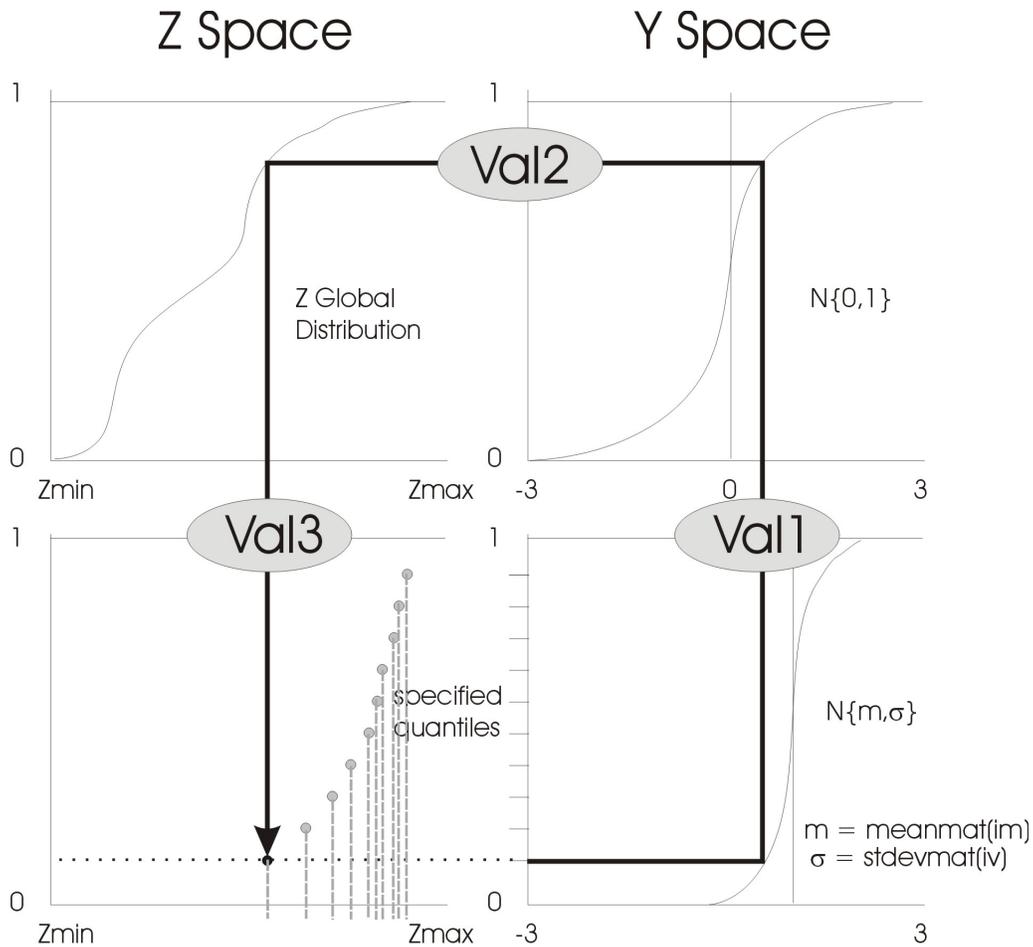
Figure 12: The graphical representation of the transformations applied to calculate the local distributions of uncertainty with a shape such that the global distribution is reproduced. The illustrated transformation is repeated for a sufficient number of quantiles to describe the local distribution.
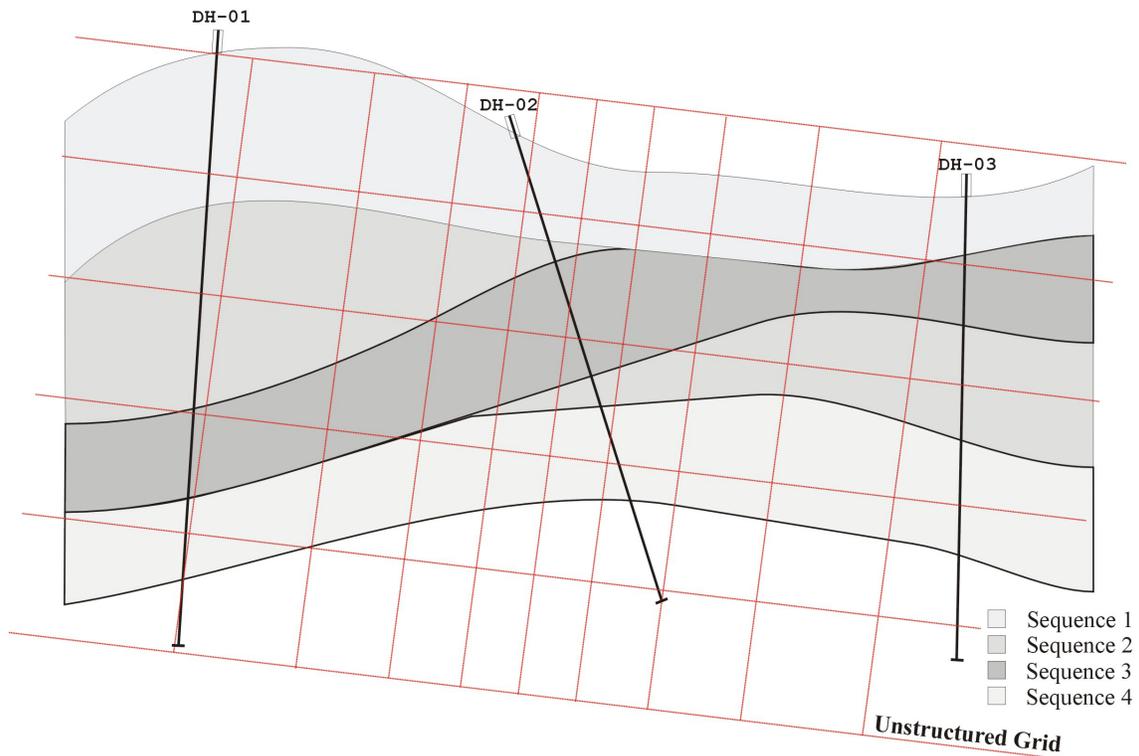
Figure 13: Stratigraphic surfaces and superimposed unstructured grid. Three hypothetical drill holes or wells are also shown.
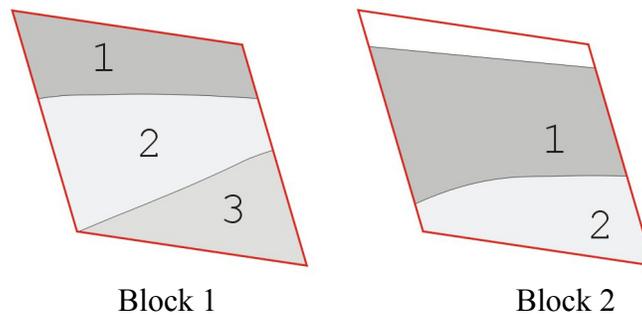
Block 1   Block 2

Figure 14: Unstructured grid block crossing multiple subsequence layers. Block 1 contains 40 % subsequence 2 and 30 % of subsequences 1 and 3 and block 2 contains 65 % subsequence 1 and 35 % subsequence 2. If the structure, facies proportions, and other characteristics such as porosity are known for block 2 and nothing is known about the block 1 except where the subsequence layers are, only data within block 2 and subsequence 1 should be used to estimate data within block 1 subsequence 1.

Appendix A

To observe the effects of screening and strings of data when dealing with volumes of various sizes, a kriging example was set up in Excel. The conditioning data consisted of six volumes with variable sizes and locations. By setting up the blocks in various configurations and performing simple and ordinary kriging, similar effects when kriging with point data can be seen.

Initially, a three block configuration was tested with blocks of equal volume. Figure A1 shows the initial configuration of blocks 1, 2, and 3. Blocks 2 and 3 remained stationary and block 1 was gradually translated along the x axis from 5 to 25 and the kriging weights calculated. The graph in Figure A1 shows the resulting weights. Initially, both blocks 2 and 3 are screened by block 1 and as the block is translated, the string effect can be observed at $x = 16$ and finally, blocks 2 and 3 screen block 1 after $x = 17$.

Using the same data configuration as in Figure 3, the block being translated was also changed in size to see the effect on the kriging weights. Three different positions were used with block one being at $x = 10$, 15, and 20 and blocks 2 and 3 being stationary. The size of block one was changed at each location to produce the graphs shown in Figure A2.
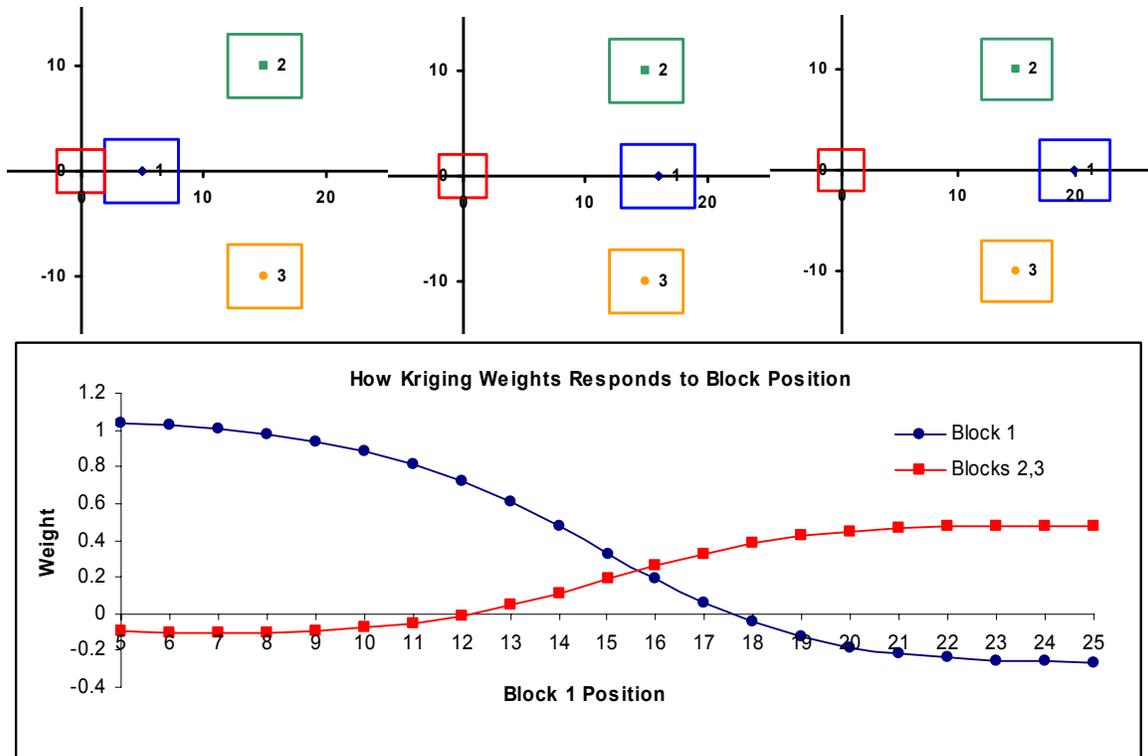
Figure A1: Top left, middle and right show various block configurations throughout the test and the graph below shows how the kriging weights varied as the configuration was changed. Both screening and the string effect can be seen.

The string effect that results from kriging with point data also occurs when kriging with volumes of equal or varying sizes. Figure A3 shows several data configurations along with graphs of the resulting kriging weights. The results are very similar to those found when using point data in the same configurations (Deutsch, 1994).
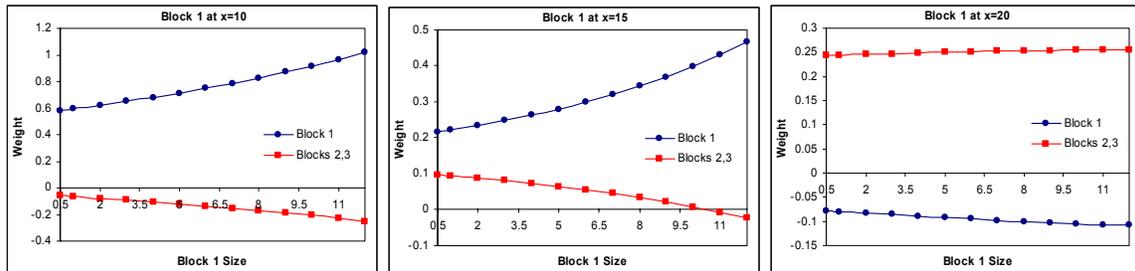
Figure A2: Plots of kriging weights versus block size. The size of block 1 was increased from 0.5 to 12 units at locations x = 10, 15, and 20 while blocks 2 and 3 where left stationary and with a side length of 6 units. The same configuration as shown by Figure 3 was used for the three blocks and simple kriging was implemented.
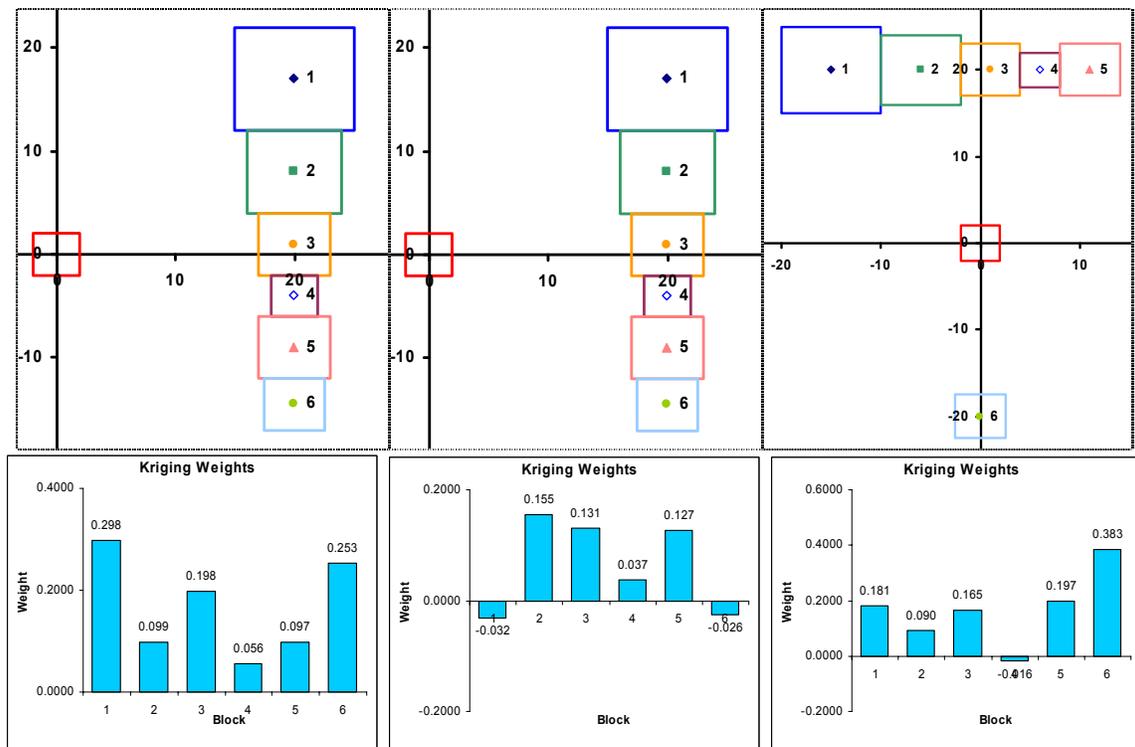
Figure A3: Data configurations and resulting kriging weights. Top left shows a string of data with varying volumes and resulting weights using ordinary kriging. The same configuration using simple kriging is centered and on the right are multiple strings of unequal size. The results are similar to those for point data.

Appendix B

The Simple Kriging Principle

Simple kriging is the key to direct sequential simulation due to the property of covariance reproduction even if the conditional probability distributions are not Gaussian. Reproducing the covariance only holds if the conditional variance is independent of the data values (homoscedastic). A proof of the covariance reproduction is provided with the following assumptions: the data stationary variable z has a mean and variance of 0 and 1, respectively. The conditional distributions are fully described by the kriging mean and variance. Equations B1, B2, and B3 describe the kriging mean and variance along with the simple kriging system considering *N* previous data:

$$z^*(\mathbf{u}) = \sum_{\alpha=1}^{N} \lambda_\alpha z(\mathbf{u}_\alpha) \tag{B1}$$

$$\sigma_{SK}^2(\mathbf{u}) = 1 - \sum_{\alpha=1}^{N} \lambda_\alpha \rho(\mathbf{u} - \mathbf{u}_\alpha) \tag{B2}$$

$$\sum_{\beta=1}^{N} \lambda_\alpha \rho(\mathbf{u}_\beta - \mathbf{u}_\alpha) = \rho(\mathbf{u} - \mathbf{u}_\alpha), \quad \alpha = 1,...,N \tag{B3}$$

Where $z^*(\mathbf{u})$ is the simple kriging mean, $\sigma_{SK}^2(\mathbf{u})$ is the simple kriging variance, and λ_α, $\alpha=1,...,N$ are the kriging weights.

A random value $R_S(\mathbf{u})$ can be drawn from a distribution described by a mean of zero and a variance equal to the kriging variance $\sigma_{SK}^2(\mathbf{u})$. The kriged mean and $R_S(\mathbf{u})$ are added together to get the simulated value for the location, $Z_S(\mathbf{u})$. An important aspect of $R_S(\mathbf{u})$ is that its value is chosen independent of the mean $Z^*(\mathbf{u})$.

$$Z_S(\mathbf{u}) = Z^*(\mathbf{u}) + R_S(\mathbf{u}) \tag{B4}$$

Now that one location has been simulated, there are *N+1* data values for simulation of the next node which will be denoted $\mathbf{u}'=\mathbf{u}_{N+1}$. The simple kriging mean and variance at $\mathbf{u}'$ are given by Equations B5 and B6 along with the kriging system shown in Equations B7 and B8:

$$Z^*(\mathbf{u}') = \sum_{\alpha=1}^{N} \lambda_\alpha z(\mathbf{u}_\alpha) + \lambda_{N+1} Z_s(\mathbf{u}) \tag{B5}$$

$$\sigma_{SK}^2(\mathbf{u}') = 1 - \sum_{\alpha=1}^{N} \lambda_\alpha \rho(\mathbf{u}' - \mathbf{u}_\alpha) - \lambda_{N+1} \rho(\mathbf{u}' - \mathbf{u}) \tag{B6}$$

$$\sum_{\beta=1}^{N} \lambda_\beta \rho(\mathbf{u}_\beta - \mathbf{u}_\alpha) + \lambda_{N+1} \rho(\mathbf{u} - \mathbf{u}_\alpha) = \rho(\mathbf{u}' - \mathbf{u}_\alpha), \quad \alpha = 1,...,N \tag{B7}$$

$$\sum_{\beta=1}^{N} \lambda_\beta \rho(\mathbf{u}_\beta - \mathbf{u}) + \lambda_{N+1} = \rho(\mathbf{u}' - \mathbf{u}) \tag{B8}$$

Where $Z^*(\mathbf{u}')$ and $\sigma_{SK}^2(\mathbf{u}')$ are the simple kriging mean and variance at location $\mathbf{u}'$ respectively. Note that the weights λ_α, $\alpha=1,...,N+1$ are *not* the same as the weights λ_α, $\alpha=1,...,N$ in Equation B1 to B3.

Once $Z^*(\mathbf{u}')$ and $\sigma_{SK}^2(\mathbf{u}')$ are known, a random value $R_S(\mathbf{u}')$ can be drawn from a distribution with a mean of zero and a variance equal to $\sigma_{SK}^2(\mathbf{u}')$. The simulated value at $\mathbf{u}'$ is calculated as follows:

$$Z_S(\mathbf{u}') = Z^*(\mathbf{u}') + R_S(\mathbf{u}') \tag{B9}$$

Let's calculate the covariance between the two simulated values:

$$
\begin{aligned}
E\{Z_S(\mathbf{u}) \cdot Z_S(\mathbf{u}')\} &= E\{Z^*(\mathbf{u}) \cdot Z^*(\mathbf{u}')\} + E\{Z^*(\mathbf{u}) \cdot R_S(\mathbf{u}')\} + \\
&\quad E\{Z^*(\mathbf{u}') \cdot R_S(\mathbf{u})\} + E\{R_S(\mathbf{u}) \cdot R_S(\mathbf{u}')\}
\end{aligned}
\tag{B10}
$$

Where $E\{Z^*(\mathbf{u}) \cdot R_S(\mathbf{u}')\}$ and $E\{R_S(\mathbf{u}) \cdot R_S(\mathbf{u}')\}$ are zero since $Z^*(\mathbf{u})$ and $R_S(\mathbf{u}')$ are independent of each other and $R_S(\mathbf{u})$ and $R_S(\mathbf{u}')$ are also independent. The remaining portions of the right hand side are non zero since the kriged means depend on one another and also because the kriged mean at the second location depends on the randomly drawn value at the first location.

$$E\{Z_S(\mathbf{u}) \cdot Z_S(\mathbf{u}')\} = E\{Z^*(\mathbf{u}) \cdot Z^*(\mathbf{u}')\} + E\{Z^*(\mathbf{u}') \cdot R_S(\mathbf{u})\} \tag{B11}$$

Expanding and simplifying the first term from the right hand side of Equation B11:

$$
\begin{aligned}
E\{Z^*(\mathbf{u}') \cdot Z^*(\mathbf{u})\} &= \sum_{\beta=1}^{N} \lambda_\beta \left[\sum_{\alpha=1}^{N} \lambda_\alpha \rho(\mathbf{u}_\beta - \mathbf{u}_\alpha) \right] \\
&\quad + \lambda_{N+1} \sum_{\alpha=1}^{N} \lambda_\alpha E\{Z_S(\mathbf{u}) Z(\mathbf{u}_\alpha)\}
\end{aligned}
\tag{B12}
$$

Now, expanding the first term of Equation B12 and recalling Equation B3:

$$\sum_{\beta=1}^{N} \lambda_\beta \left[\sum_{\alpha=1}^{N} \lambda_\alpha \rho(\mathbf{u}_\beta - \mathbf{u}_\alpha) \right] = \sum_{\beta=1}^{N} \lambda_\beta \rho(\mathbf{u}_\beta - \mathbf{u}) \tag{B13}$$

Expanding the second term of Equation B12:

$$E\{Z_S(\mathbf{u}) Z(\mathbf{u}_\alpha)\} = E\{Z^*(\mathbf{u}) Z(\mathbf{u}_\alpha)\} + E\{R_S(\mathbf{u}) Z(\mathbf{u}_\alpha)\} \tag{B14}$$

Knowing that $E\{R_S(\mathbf{u}) Z(\mathbf{u}_\alpha)\} = 0$ and $E\{Z^*(\mathbf{u}) Z(\mathbf{u}_\alpha)\} = \rho(\mathbf{u} - \mathbf{u}_\alpha)$:

$$\sum_{\alpha=1}^{N} \lambda_\alpha E\{Z_S(\mathbf{u}) Z(\mathbf{u}_\alpha)\} = \sum_{\alpha=1}^{N} \lambda_\alpha \rho(\mathbf{u} - \mathbf{u}_\alpha) = 1 - \sigma_{SK}^2(\mathbf{u}) \tag{B15}$$

Substituting B13 and B15 back into B12:

$$E\{Z^*(\mathbf{u}') \cdot Z^*(\mathbf{u})\} = \sum_{\beta=1}^{N} \lambda_\beta \rho(\mathbf{u}_\beta - \mathbf{u}) + \lambda_{N+1}\left[1 - \sigma^2_{SK}(\mathbf{u})\right] \qquad (B16)$$

Now, let's go back to Equation B11 and expand and simplify the second term of the right hand side:

$$E\{Z^*(\mathbf{u}') \cdot R_S(\mathbf{u})\} = \sum_{\alpha=1}^{N} \lambda_\alpha E\{Z(\mathbf{u}_\alpha)R_S(\mathbf{u})\} + \lambda_{N+1}E\{Z_S(\mathbf{u})R_S(\mathbf{u})\} \quad (B17)$$

The $\sum_{\alpha=1}^{N} \lambda_\alpha E\{Z(\mathbf{u}_\alpha)R_S(\mathbf{u})\}$ term is zero since $Z(\mathbf{u}_\alpha)$ and $R_S(\mathbf{u})$ are independent.

By expanding the $E\{Z_S(\mathbf{u})R_S(\mathbf{u})\}$ portion of the second term in Equation B17, it can be shown that it is equivalent to the simple kriging variance:

$$E\{Z_S(\mathbf{u})R_S(\mathbf{u})\} = E\{Z^*(\mathbf{u})R_s(\mathbf{u})\} + E\{R_S^2(\mathbf{u})\} \qquad (B18)$$

Since $E\{Z^*(\mathbf{u})R_s(\mathbf{u})\}$ is zero (the variance is homoscedastic. If the variance was heteroscedastic, $E\{Z^*(\mathbf{u})R_s(\mathbf{u})\} \neq 0$ and the covariance would not be reproduced):

$$E\{Z_S(\mathbf{u})R_S(\mathbf{u})\} = E\{R_S^2(\mathbf{u})\} = \sigma^2_{SK}(\mathbf{u}) \qquad (B19)$$

Substituting B19 back into Equation B17:

$$E\{Z^*(\mathbf{u}') \cdot R_S(\mathbf{u})\} = \lambda_{N+1}\sigma^2_{SK}(\mathbf{u}) \qquad (B20)$$

Substituting B16 and B20 into B11 and simplifying:

$$E\{Z_S(\mathbf{u}) \cdot Z_S(\mathbf{u}')\} = \sum_{\beta=1}^{N} \lambda_\beta \rho(\mathbf{u}_\beta - \mathbf{u}) + \lambda_{N+1}\left[1 - \sigma^2_{SK}(\mathbf{u})\right] + \lambda_{N+1}\sigma^2_{SK}(\mathbf{u})$$

$$E\{Z_S(\mathbf{u}) \cdot Z_S(\mathbf{u}')\} = \sum_{\beta=1}^{N} \lambda_\beta \rho(\mathbf{u}_\beta - \mathbf{u}) + \lambda_{N+1}$$

$$E\{Z_S(\mathbf{u}) \cdot Z_S(\mathbf{u}')\} = \rho(\mathbf{u}'-\mathbf{u}) \qquad (B21)$$

By working through from Equation B11 to Equation B21, the covariance is correct. The marginal covariance is reproduced.

Appendix C

Transforming normal values to lognormal

A variable, $Z \mid z(\mathbf{u}) > 0$, is lognormal with a mean m and standard deviation σ if the natural logarithm of $Z(\mathbf{u})$, $X(\mathbf{u}) = ln(Z(\mathbf{u}))$ is normally distributed with mean α and standard deviation β. Knowing the relation between $Z(\mathbf{u}) \rightarrow logN(m, \sigma)$ and $X(\mathbf{u}) \rightarrow N(\alpha, \beta)$, one can convert a Gaussian distribution, $Y(\mathbf{u}) \rightarrow N(0,1)$, into a lognormal distribution. Equations *1C* and *2C* show the relation between $X(\mathbf{u})$, $Y(\mathbf{u})$, and $Z(\mathbf{u})$, and Equations *3C* and *4C* show the relation between m and σ with α and β.

$$X(\mathbf{u}) = \alpha + \beta \cdot Y(\mathbf{u}) \qquad (1C)$$

$$Z(\mathbf{u}) = e^{X(\mathbf{u})} \qquad (1Ca)$$

Substituting 1C into 1Ca:
$$Z(\mathbf{u}) = e^{\alpha + \beta \cdot Y(\mathbf{u})} \qquad (2C)$$

$$\alpha = \ln(m) - \frac{\beta^2}{2} \qquad (3C)$$

$$\beta^2 = \ln\left(1 + \frac{\sigma^2}{m^2}\right) \qquad (4C)$$

Equations *5C* and *6C* describe the normal and lognormal probability distribution curves, which are quite similar in arrangement; however the value of the lognormal function is largely dependent on the variable value $z(\mathbf{u})$. This dependency introduces the proportional effect into the data set. Figure 1C shows the change in the distribution shapes as $Y(\mathbf{u})$ is converted into $X(\mathbf{u})$ and as $X(\mathbf{u})$ is transformed into $Z(\mathbf{u})$.

$$f(x) = \frac{\exp\left[-\frac{1}{2}\left(\frac{\ln(x) - \alpha}{\beta}\right)^2\right]}{\beta \cdot x\sqrt{2\pi}} \qquad (5C)$$

$$g(x) = \frac{\exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right]}{\sigma\sqrt{2\pi}} \qquad (6C)$$

Variogram Relationship

There is an analytical relationship between the correlograms of Gaussian and lognormal data. Equation 7C relates the correlograms and using Equations 8C and 9C, the variograms can be obtained. Knowing the variogram for lognormal data was useful for simulating in original units.

$$\rho(h) = \frac{m^2}{\sigma^2}\left[e^{\beta^2 \cdot r(h)} - 1\right] \qquad (7C)$$

$$\rho(h) = 1 - \frac{\gamma_Z}{\sigma_Z^2} \qquad (8C)$$

$$r(h) = 1 - \gamma_Y \qquad (9C)$$

where $r(h)$ is the correlation in Y-space and $\rho(h)$ is the correlation in Z-space.
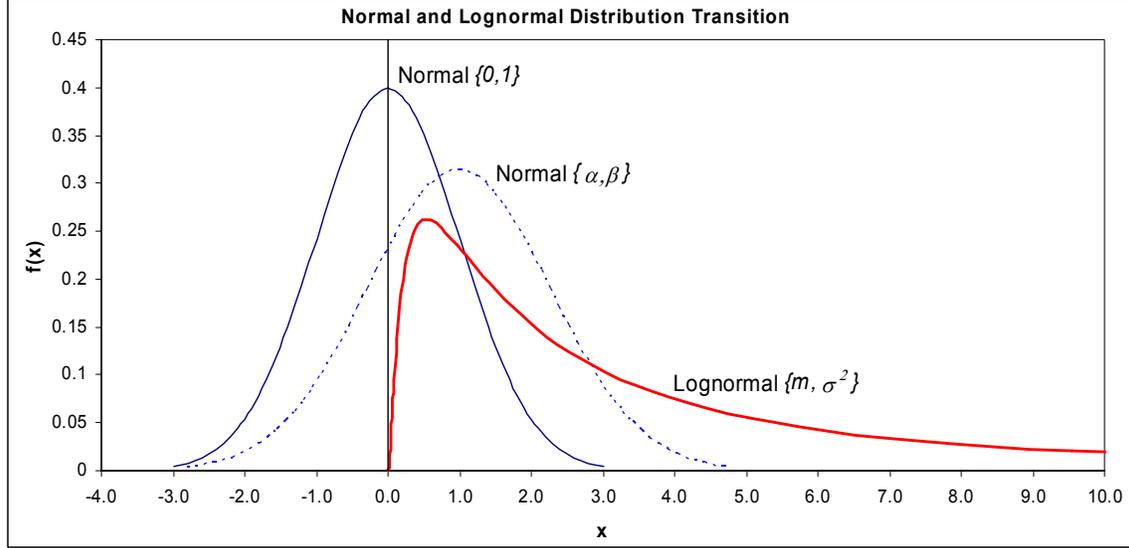
Figure 1C: Normal and corresponding lognormal distributions. The lognormal distribution was calculated from the Gaussian distribution by using the transformation equations (1 to 4) above. The lognormal distribution has a mean of 6 and a standard deviation of 3.

Homoscedastic Variance Correction

Because kriging is a linear estimator, the resulting kriging variance is homoscedastic and this is not correct for data exhibiting the proportional effect. With lognormal data, an equation exists for correcting the variance using the mean or estimate and it can be derived from Equation 4C.

$$\beta_L^2 = \ln\left(1 + \frac{\sigma_Z^2}{m_Z^2}\right) \qquad (4C)$$

$$1 + \frac{\sigma_Z^2}{m_Z^2} = e^{\beta_L^2}$$

$$\sigma_Z^2 = m_Z^2(e^{\beta_L^2} - 1)$$

Since m_Z^2 is the estimate from kriging it can be denoted by $z^*(\mathbf{u})$. From Equation A1 in appendix A it is shown that the local β_L^2 value can be determined from the kriging variance in Y-space and the global β_G^2 value.

$$\sigma_{Z,C}^2 = z^*(\mathbf{u})^2(e^{\beta_G^2 \cdot \sigma_Y^2} - 1) \qquad (10C)$$

Where $\sigma_{Z,C}^2$ is the corrected variance, σ_Y^2 is the local variance in normal space, and β_G^2 is the global variance of *ln(Z)*