

The Effect of the Simulation Path in Sequential Gaussian Simulation

J. A. McLennan (jam12@ualberta.ca)
University of Alberta, Edmonton, Alberta, CANADA

Abstract

Sequential simulation requires that subsequent point/block locations be conditioned to previously simulated values. A simulation path or order of simulation must be chosen. Conventional wisdom is to use a random path. There are two alternatives that are occasionally brought up: (1) a row-by-row regular path that would require less computer resources because of a regular kriging configuration and (2) a path that spirals away from original sample data that may more correctly honor sample data. The problems with these two alternatives are documented in this short note. Conventional wisdom is justified in the recommendation of a random path.

Introduction

In the sequential simulation approach, subsequent point/block locations are simulated according to a random path. A random path can be created with the use of any valid random number generator. First, a random number is assigned to each simulation location; the simulation locations are then sorted such that the assigned random numbers are in an ascending/descending order. The path defined by visiting the simulation points/blocks in this order is random. Figure 1 illustrates this random path implementation for a 3x3 block setting. Sorting large arrays of numbers can be time consuming and the limited period length of linear congruential random number generators could also be used (Srivastava, 1988).

At each point/block location, an attribute value is drawn from a local conditional cumulative distribution function (ccdf). The ccdfs are modeled using some *flavor* of kriging that considers original sample data and previously simulated grid nodes. In practical implementation, the number of conditioning data retained is restricted due to limited computer resources and our desire for fast results. Doubling the number of conditioning data retained leads to four-fold and eight-fold increases in computer memory and CPU demand, respectively. Nearby conditioning data are selected according to closest variogram distance until either a pre-defined maximum number of conditioning data are retained or the boundary of a pre-defined search neighborhood (often the variogram range) is reached.

Searching for nearby conditioning data is a recurrent process that needs to be optimized. If the data are configured irregularly within the simulation grid, we can simply compute and sort the variogram distance between close conditioning data and the location to be simulated; however, if the conditioning data occupy a regular grid, a more efficient *spiral search* can be implemented. Optimum search parameters will often involve using a combination of both search routines; the first to search for irregularly spaced sample data, and the second to search for regularly spaced previously simulated grid nodes.

As sequential simulation progresses, the distances between conditioning data and the simulation locations decrease relative to the variogram range. As a consequence, reproduction of long-range

variogram structure can be poor. To attenuate this, a multiple grid search is implemented. A multiple grid search simulates in several stages. In each subsequent stage, simulation is performed on a finer grid covering the same area (or volume) and conditioned to values simulated at the previous stage, Tran (1994). This continues until the final grid density is simulated. Long-range variogram structure is honored in the early stages while short-range variogram structure is honored in the later stages. Improved variogram reproduction using a multiple grid search routine comes at no cost to CPU speed or computer memory.

Variogram reproduction depends on the simulation parameters used, some of which include the simulation grid density, the overall field size, the span of the search neighborhood, the number of conditioning data retained and whether or not a multiple grid search is used. A small experiment was performed using various combinations of these parameters to better understand their relationship. Using the “sgsim” program from GSLIB, 100 unconditional Gaussian simulations were constructed for each combination. All input variogram models had a single spherical structure with a 10% relative nugget effect. For each combination, the 100 exhaustive variograms were calculated and, to nullify ergodicity, averaged. The results are shown in Figure 2. Variogram reproduction improves as the simulation grid density becomes coarser, the field size expands, the search neighborhood increases (up to the variogram range), the number of conditioning data retained escalates and a multiple grid search is implemented.

Practitioners understand the effect that different simulation parameters have on variogram reproduction. Due attention to practical advice results in consistent and successful variogram reproduction; however, there is an increasing anxiety among geostatisticians to accept the high computer resources and unrealistic *simulated features* manifested from current simulation algorithms. Regular and spiral simulation paths are considered to be possible solutions to the aforementioned problems, but they have yet to be well researched and documented.

A *regular* simulation path is described as simulating subsequent point/block locations according to a row-by-row sequence. The first location simulated will populate a corner of the 2-D/3-D simulation grid. A location perpendicular to and directly beside the first would be simulated next. Simulation continues along this direction until the last location in the row is simulated. Subsequent rows parallel and directly beside previously simulated rows are then visited similarly. If the final simulation grid is 3-D, the row-by-row sequence will only simulate the first of multiple layers and will need to be repeated for all subsequent layers parallel and directly beside previously simulated layers.

A regular simulation path ensures that the majority of subsequent locations are modeled using the same arrangement of conditioning data. As a result, the CPU time required to solve the necessary kriging matrices and the memory allocation needed for previously simulated nodes significantly decreases.

A spiral simulation path considers subsequent point/block locations according to their closeness to the original sample data. Based on calculated variogram distance, a predefined maximum number of conditioning data and a predefined search neighborhood, the sample data closest to each simulation location are retained; the simulation locations are sorted so that these distances are in an *ascending* order. The path defined by following the simulation locations in such order will simultaneously spiral away from all the original sample data.

It is a common perception that if simulation were to begin at locations close to the sample data and continue for locations increasingly farther away, the original data would be more faithfully

honored and the resulting simulated maps would be more realistic, that is, absent of any simulated features.

Although a regular and spiral simulation path would relieve computer demand and might generate more realistic grade distributions, respectively, the validity of each must be assessed before they can be implemented in practice. Both paths should, above all, sufficiently reproduce the input/model variogram, be absent of any artifacts and accurately reproduce the true local cdfs. Both the regular and spiral simulation path will be shown to fail to meet any of these criteria. The regular and spiral paths will be described in more detail and implemented in various simulation settings.

Regular Path

The Path. Consider the $n \times l \times q$ block simulation setting in Figure 3 where n , l and q are the block indices. Reference X - Y - Z directions are orientated so that n cycles from 1 to N in the positive X direction, l cycles from 1 to L in the positive Y direction and q cycles from 1 to Q in the positive Z direction ($Q = 1$ in 2-D). Each possible n - l - q combination refers to a block location, e.g., block 1-2-3 corresponds to the location where $n = 1$, $l = 2$ and $q = 3$. The order in which a regular path could visit the $N \times L \times 1$ layer of blocks (shaded in 3-D and shown in plan view in 2-D) is shown. The first and second blocks visited are located at 1-1-1 and 2-1-1 and labeled as “1” and “2”, respectively. The path continues in the positive X direction until the $l = 1$ rows’ last block location (N -1-1, labeled as “ N ”) is visited. The $l = 2$ row (1-2-1, 2-2-1, ..., N -2-1, labeled as “ $N+1$, $N+2$, ... $2N$ ”) parallel and directly beside the $l = 1$ row is simulated next. This process of cycling through the n index with l constant and $q = 1$ is then repeated for $l = 3, \dots, L$ until the layers’ final block location (N - L -1, labeled as “ $L \times N$ ”) is visited. In 3-D, we would need to repeat this process of cycling the n index within the slower cycling l index for $q = 1, \dots, Q$ until the final block location (N - L - Q) in the model is visited. There are other possible regular paths. We specify a regular path by writing the indices in a particular order, e.g., an nlq regular path implies visiting the block locations according to the l index cycling faster than the q index and slower than the n index. In total, there are two possible regular paths in 2-D and six possible regular paths in 3-D.

Advantage. Consider unconditionally simulating 10,000 block locations with $N = L = 100$ according to a regular nl path (n cycles faster than l) within the square 2-D ($Q = 1$) field shown in Figure 4. The regular nl path (versus a random path) simplifies the matrix computations and reduces the memory allocation needed to simulate the field. For a maximum of 25 previously simulated blocks retained to condition the cdfs, the shaded region (drawn to scale) would contain 8,832 of the 10,000 possible block locations to which the exact same arrangement of conditioning data is retained. This corresponds to solving a single matrix for a single set of kriging weights applicable to 88.3% of the block locations (see bottom of Figure 4).

If a grid of simulation locations is visited according to a random path, the conditioning data configuration and corresponding set of kriging weights could be different for each simulation location; however, visiting the simulation locations according to a regular path allows a template of kriging weights to be applied to the majority of the simulation field. Since CPU speed increases as the number of matrices to be solved decreases, the same number of simulation locations could be simulated significantly faster using a regular path in place of a random path. Also, random simulation paths require the allocation of enough memory to hold a simulated value at each simulation location. This is necessary because it is possible for any previously simulated location to be recalled for conditioning subsequent randomly visited locations. The field of simulated values is written to the output file only after the last location is simulated. Using a

regular path, however, the number of previously simulated values that have no chance to be used as conditioning data would continually increase, that is, because a regular path retains the same conditioning data configuration for the majority of locations, it becomes impossible for an increasing number of previously simulated locations to be recalled within the dominating conditioning data configuration template. Such values could be continually written to the output file during the simulation, sparing significant initial memory allocation.

The Problem. Despite the advantages of a regular simulation path, it has yet to be implemented in practice – it is ill-famed due to suspicions that artificially created spatial structures, such as a spread of artifacts along rows, would result in poor variogram reproductions. All suspicion aside, if the same conditioning data configuration is used to simulate the majority of subsequent simulation locations, we cannot expect the input variogram to be reproduced for distances outside the span of this dominating data configuration. There is a lack of enough implementation or documentation, however, to prove/disprove these expectations and help support using a regular path in practice.

Implementation. We now implement a regular simulation path in 2-D and 3-D (see figure 5). There are 2,500 blocks in 2-D with $N = L = 50$ ($Q = 1$) and 421,875 blocks in 3-D with $N = L = Q = 75$ to be simulated. The X - Y - Z directions are orientated so that the n - l - q indices cycle in the Easting-Elevation-Northing directions, respectively. The blocks measure 2m a side. The model variogram has a 10% relative nugget effect with isotopic ranges of 10 and 15m in 2-D and 3-D, respectively. All possible regular paths in 2-D and 3-D were implemented without conditioning data. To attenuate the effect of ergodic fluctuations, 100 realizations were generated for each possible path and the 100 corresponding calculated variograms were averaged. This process was repeated for a random simulation path. The results are summarized in Figure 5. All possible regular simulation paths unsatisfactorily reproduce the model variograms, but, as expected, the random simulation path successfully reproduces the variogram models.

Discussion. A characteristic feature of all the calculated regular path variograms is a substantial amount of added correlation in the direction where the slowest cycling index resides. Neither the input variogram model nor the average random path variogram predicts this imparted anisotropy – it is an *artifact* of the regular simulation path. A representative illustration of this artifact is presented in Figure 6. The average Northing direction variograms, calculated from the 100 random path realizations and the 100 nlq regular path (q is the slowest cycling index in the Northing direction) realizations, are shown. Also, a central Northing-Easting cross-section through one of the 100 random path realizations and one of the 100 regular path realizations are shown. The artifact is clearly visible. There is added correlation in the Northing direction nlq regular path variogram and cross-section, inconsistent with both the model.

The six regular 3-D simulation paths (refer to Figure 5) were initially implemented without conditioning data, by retaining a maximum of 25 previously simulated nodes and using a grid density of $75 \times 75 \times 75$, an overall field size of $150 \times 150 \times 150$ m and an isotopic range and search radius of 15m. Changing these parameters will mitigate, but not remove, the regular path artifact. Various combinations of these parameters were considered and the six regular simulation paths were re-implemented for each combination. Although conditioning data mollified the artifact, using fewer previously simulated values, coarser grid densities, smaller field sizes and smaller search neighborhoods amplified the artifact. A similar re-implementation and result was obtained in 2-D.

For the same reason a regular path simulation is less demanding of CPU time and computer memory, it is not suitable for practical implementation. The same arrangement of conditioning

data is used to simulate the majority of locations; therefore, variogram reproduction is only satisfactory for distances up to the span of the dominating conditioning data configuration.

The exact manner in which the regular path variogram deviates from the model variogram for distances past the span of the dominating data configuration is not easily explained. It is reasonable to assume, however, that the artifact originates from the relative influence given to the dominating conditioning data configuration, that is, the template of kriging weights (refer to Figure 5). The artifact could be predicted by somehow propagating the template of kriging weights to varying distances and directions. Further analysis of possible propagation techniques and results would be needed to be able to predict the artifact for all distances and directions.

Spiral Path

The Path. Consider the 2-D point simulation setting in the top of Figure 7. The dark boxes are sample data and the points to be simulated (shown as dots) populate a regular grid. A spiral path through the simulation locations is to be established. Centered on a particular sample datum, the shaded region is enlarged to show its surrounding simulation locations (shaded circles). The distances between the simulation locations and the central sample datum are calculated as $d-1$, $d-2$, $d-3$, ..., $d-n$, where n is the number of different distances. These calculated distances could then be summarized as being the radii of a family of circles centered on that particular sample datum. Each circle intersects a set of point locations to be simulated. The same procedure is repeated for all sample data, i.e. there are as many sets of simulation locations intersected by the circle of radius $d-1$ (or any radius), as there are sample data. The order of simulation is then defined by simulating groups of locations according to an ascending radii order: the first group of locations simulated intersect the circles of radius $d-1$, the next group intersect the circles of radius $d-2$ and so on until the group of locations intersected by the circles of radius $d-n$ are simulated. Simulation locations equidistant from sample data are visited randomly. In 3-D, a spiral path is established with the same procedure, except the distances would be summarized as spheres instead of circles.

Possible Advantage. There is a common suspicion among some practicing geostatisticians that if hard (sample) data were somehow more influential than previously simulated values, the resulting simulated maps would be more realistic and absent of any “simulated features”. The theme here is that sample data, compared to simulated data, should be more faithfully honored in conditioning the local cdfs. A spiral simulation path can put this theme into practice via simulating locations as close and as long as possible to the original sample data.

The Problem. The problem with a spiral simulation path is similar to the problem of a regular simulation path. The search routine would systematically retain conditioning data configurations that span small distances relative to the variogram range and variogram reproduction can only be expected to be satisfactory for distances up to the span of the conditioning data configuration. Further, proper implementation and documentation of the spiral path has yet to be performed to warrant its use or rejection in practice.

Implementation. The spiral simulation path is implemented in 2-D for various sample data configurations. There are 50x50 (2,500) block locations measuring 1m a side to be simulated. The model variogram has a 10% relative nugget effect with an isotopic range of 10m. The averages of 100 variograms from 100 spiral path realizations and 100 random path realizations are calculated and compared relative to each other and the input variogram model. Figure 8 shows the results for (1) a single centrally located sample (2) four samples roughly equidistant from each other and the

50x50m field border and (3) 29 exploration samples (“data.dat” from GSLIB). The spiral path variograms all deviate from the model variogram. The random path variograms all satisfactorily reproduce the model.

Discussion. The results are not surprising. An arrangement of conditioning data spanning distances less than the variogram range is constantly retained for the local ccdf models, thus, satisfactory variogram reproduction for distances exceeding this span is not expected.

Just as we did for the regular simulation path, we re-implemented the spiral simulation path with different combinations of simulation parameters. As expected, using more previously simulated values, coarser grid densities, larger field sizes and a large search neighborhood improves variogram reproduction. We also tried different amounts and arrangements of original sample data configurations, from which it seems variogram reproduction improvement and number of original sample data are positively correlated (this can also be seen in Figure 8 by the relative size of the broken-line circles).

The spiral simulation path approach seems to work decently, that is, with enough conditioning data, e.g. “data.dat” (Figure 8), there are minor, yet noticeable, deviations from the model. However, this path offers no computer resource advantages and does not appear to “value” or “weight” the original sample data any more than a random simulation path does.

The regular and spiral simulation paths are similar in that they fail to reproduce long-range variogram structure, yet different by the manner in which they do so. Both regular and spiral simulation paths constantly retain conditioning data configurations spanning smaller distances than the model variogram range; however, a regular simulation path fails to retain conditioning data from the particular direction in which the slowest cycling index resides (refer to Figure 4), whereas a spiral simulation path retains conditioning data from all previously simulated directions. So, although the mutual result is unsatisfactory variogram reproduction from both paths, a regular path variogram exhibits added correlation in the direction for which the slowest cycling index exists whereas a spiral path variogram exhibits added correlation unbiased in direction (compare Figures 6 and 8).

Local Conditional Cumulative Distribution Functions (ccdfs)

Other than what has already been presented, the validity of a regular and spiral simulation path could be assessed via a local conditional cumulative distribution function (ccdf) analysis: the local ccdf means and variances created from either of these paths should agree with the true ccdf means and variances (created from simple kriging). Consider the example in Figure 9. There are 121 (11x11) 1x1m blocks to be simulated. There is a central sample with a normal score deviate of +2.0. The variogram model is a single spherical structure with an isotopic range of 10m; notice that even the four corner blocks, located ~ 7.8m away, are within the 10m range of the model variogram. The true mean and variance of the 121 local ccdfs are calculated using simple kriging, then derived from the average of 1000 random path realizations, the average of 1000 regular path realizations and the average of 1000 spiral path realizations. The result of comparing the averaged ccdf means and variances to the true mean and variance at each of the 121 locations is summarized in Figure 9. As expected, the average random path ccdfs are perfectly correlated to the true ccdfs; however, the average regular path ccdfs are not as well correlated (0.84 versus 1.00 and 0.90 versus 0.99 for the means and variances, respectively). The average spiral path ccdfs seem to be well correlated to the true ccdfs, but deviate from the true ccdfs in a specific manner: for locations far away (near the corners), the spiral path ccdf means and variances

underestimate and overestimate the true cdf means and variances, respectively – a pattern not predicted by the random path. The example and results re-emphasize the notion that regular and spiral paths are unsuitable for practical implementation.

Conclusion

Although regular and spiral simulation paths *could* significantly improve the efficiency and quality of simulation algorithms, both force only close conditioning data to be retained resulting in poor variogram reproduction. The conventional random path using the multiple grid option is recommended in practice.

References

- Deutsch, C.V. and Journel A.G. (1998). *GSLIB: Geostatistical Software Library: and User's Guide*, Oxford University Press, New York, 2nd Ed.
- Tran, T. T. (1994). *Improving Variogram Reproduction on Dense Simulation Grids*, Computers and Geosciences, 20(7/8):1161-1168.

A 3x3 block setting:

Block 1	Block 2	Block 3
Block 4	Block 5	Block 6
Block 7	Block 8	Block 9

Random number assignment:

<i>rnd #:</i> 0.3866	<i>rnd #:</i> 0.8635	<i>rnd #:</i> 0.7176
<i>rnd #:</i> 0.0244	<i>rnd #:</i> 0.4352	<i>rnd #:</i> 0.0969
<i>rnd #:</i> 0.5376	<i>rnd #:</i> 0.3380	<i>rnd #:</i> 0.7205

The list of sorted locations:

block 4	→	0.0244	<i>increasing rnd #</i> ↓
block 6	→	0.0969	
block 8	→	0.3380	
block 1	→	0.3866	
block 5	→	0.4352	
block 7	→	0.5376	
block 3	→	0.7176	
block 9	→	0.7205	
block 2	→	0.8635	

The order the blocks are visited:

<i>visited 4th</i>	<i>visited 9th</i>	<i>visited 7th</i>
<i>visited 1st</i>	<i>visited 5th</i>	<i>visited 2nd</i>
<i>visited 6th</i>	<i>visited 3rd</i>	<i>visited 8th</i>

Figure 1 – Random Path. An illustration of the procedure to establish a random path for a 3x3 block simulation setting. The 9 block locations (not according to GSLIB convention), the random number assigned to each block, the sorted list of block locations (in terms of an *ascending* random number order) and the resulting random simulation path is shown.

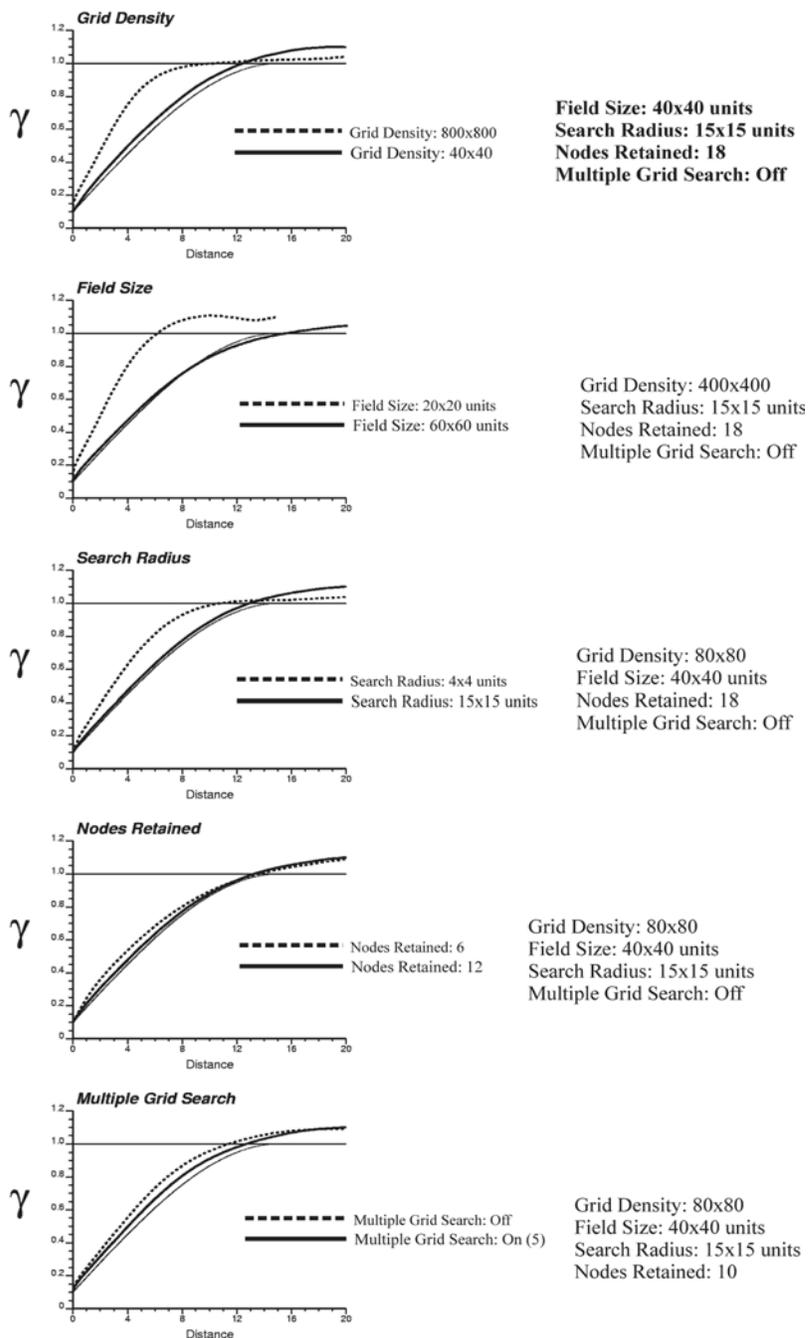
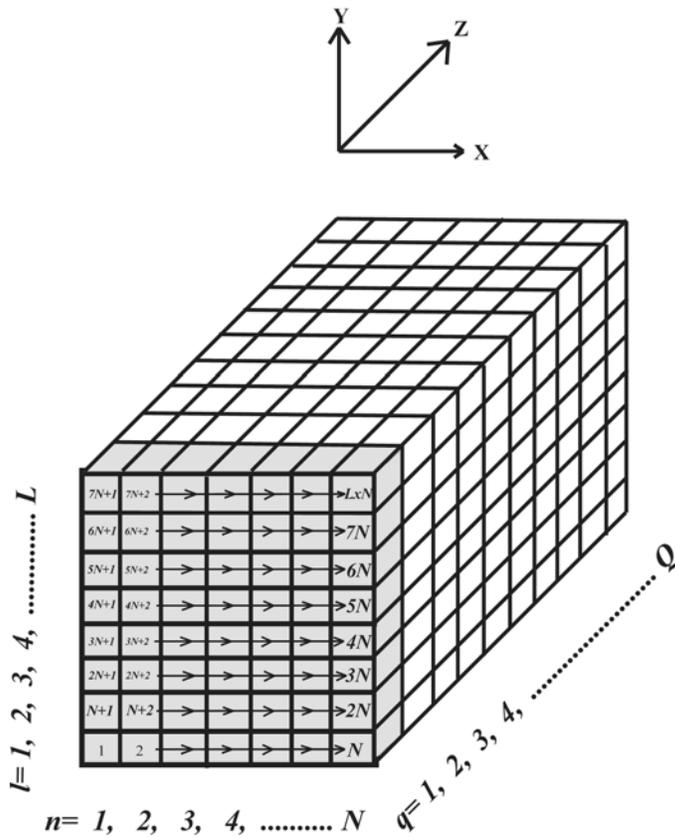


Figure 2 – Simulation Parameters and Variogram Reproduction. An illustration of the effect different simulation parameters have on variogram reproduction. An average variogram is calculated for various combinations of simulation parameters and shown as either a heavy broken line or heavy solid line. In each of the five plots, the varied parameter is indicated by the title and the inscribed variogram legend, the constant parameters are listed to the right and the model variogram is shown as a thin line. Variogram reproduction improves as the simulation grid density becomes coarser, the field size expands, the search neighborhood increases, the number of conditioning data retained escalates and a multiple grid search is implemented.

3 Dimensions



2 Dimensions

shaded portion in 3-D model ($q = 1$)

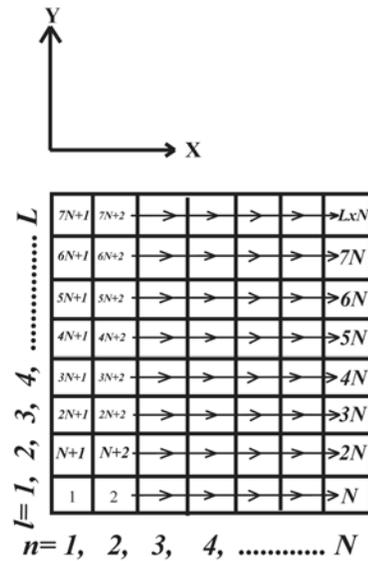
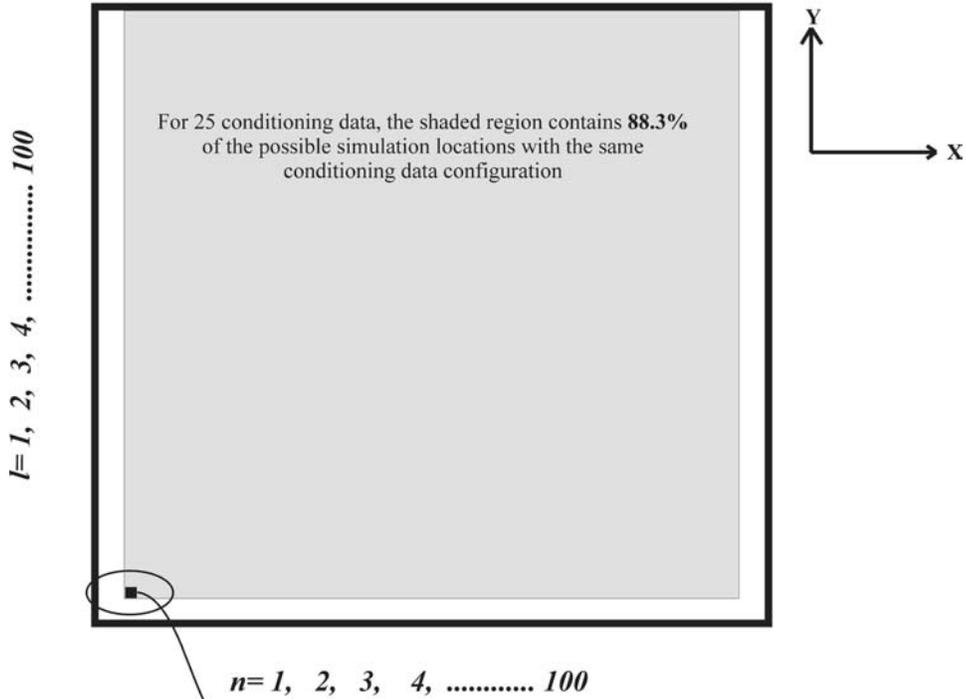


Figure 3 – Regular Path. Reference X-Y-Z directions are orientated so that the n index cycles from 1 to N in the positive X direction, the l index cycles from 1 to L in the positive Y direction and the q index cycles from 1 to Q in the positive Z direction. The order in which a regular path could visit the $N \times L \times 1$ ($q = 1$) layer of blocks, shaded in 3-D and shown in plan view in 2-D, is established by cycling the n index slower than the l index with $q = 1$. The corresponding regular path in 3-D would repeat this procedure, as in 2-D, within the slowest cycling q index.

$$N = L = 100$$



Kriging Weight Template (25 Conditioning Data)

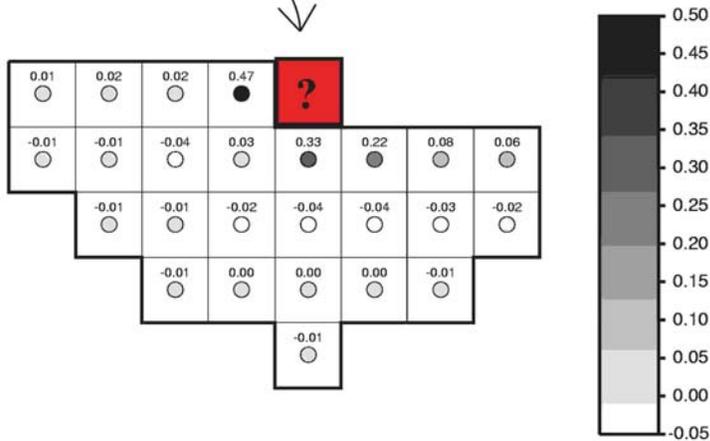
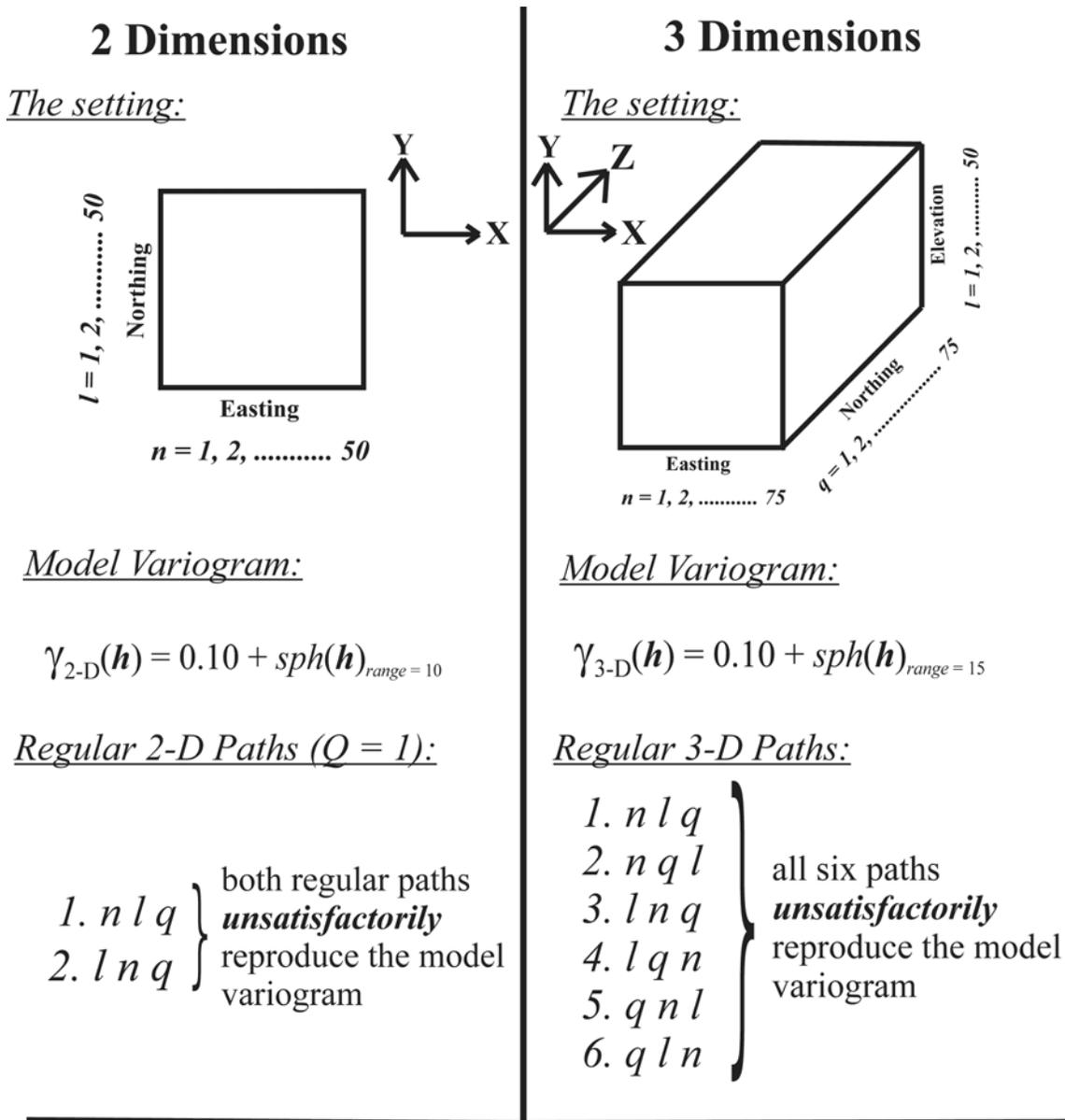


Figure 4 – Regular Path Data Configuration. An illustration of the mechanism that makes a regular simulation path more efficient than a random path. The square field (top) bounds 10,000 possible simulation locations. If a maximum of 25 conditioning data were to be retained per ccdf model, the shaded region would contain 88.3% (8,832) of the block locations where the exact same 25 conditioning data configuration would be used. This 25 data configuration and corresponding template of kriging weights is shown (bottom).



- The random simulation path **successfully** reproduces both the 2-D and 3-D model variograms.

Figure 5 – Regular Path Simulation Results. A summary illustration of the regular path implementation and results. Reference X-Y-Z directions are orientated so that the $n-l-q$ indices cycle in the Easting-Elevation-Northing directions, respectively. $N = L = 50$ ($Q = 1$) in 2-D and $N = L = Q = 75$ in 3-D. The variogram has a 10% relative nugget effect and isotropic ranges of 10 and 15m in 2-D and 3-D, respectively. All possible regular paths yield unsatisfactory variogram reproduction; however, the random simulation path yields satisfactory variogram reproduction.

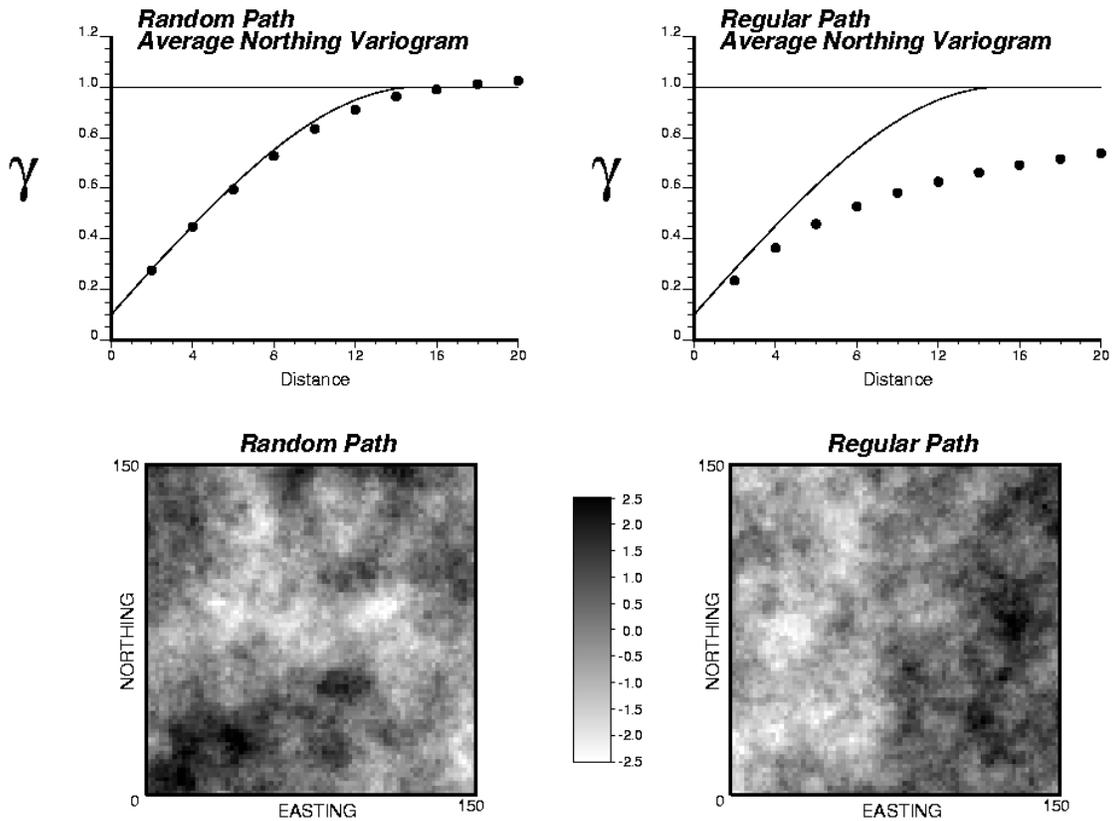


Figure 6 – Regular Path Artifact. An illustration of the regular path artifact via the previously implemented *nlq* regular path. The average random and regular path Northing variograms are shown as solid dots; the dark solid line is the model. A central Northing-Easting slice through one of the 100 random and regular path realizations are also shown. There is a substantial amount of correlation added to the Northing direction for which the input variogram does not predict.

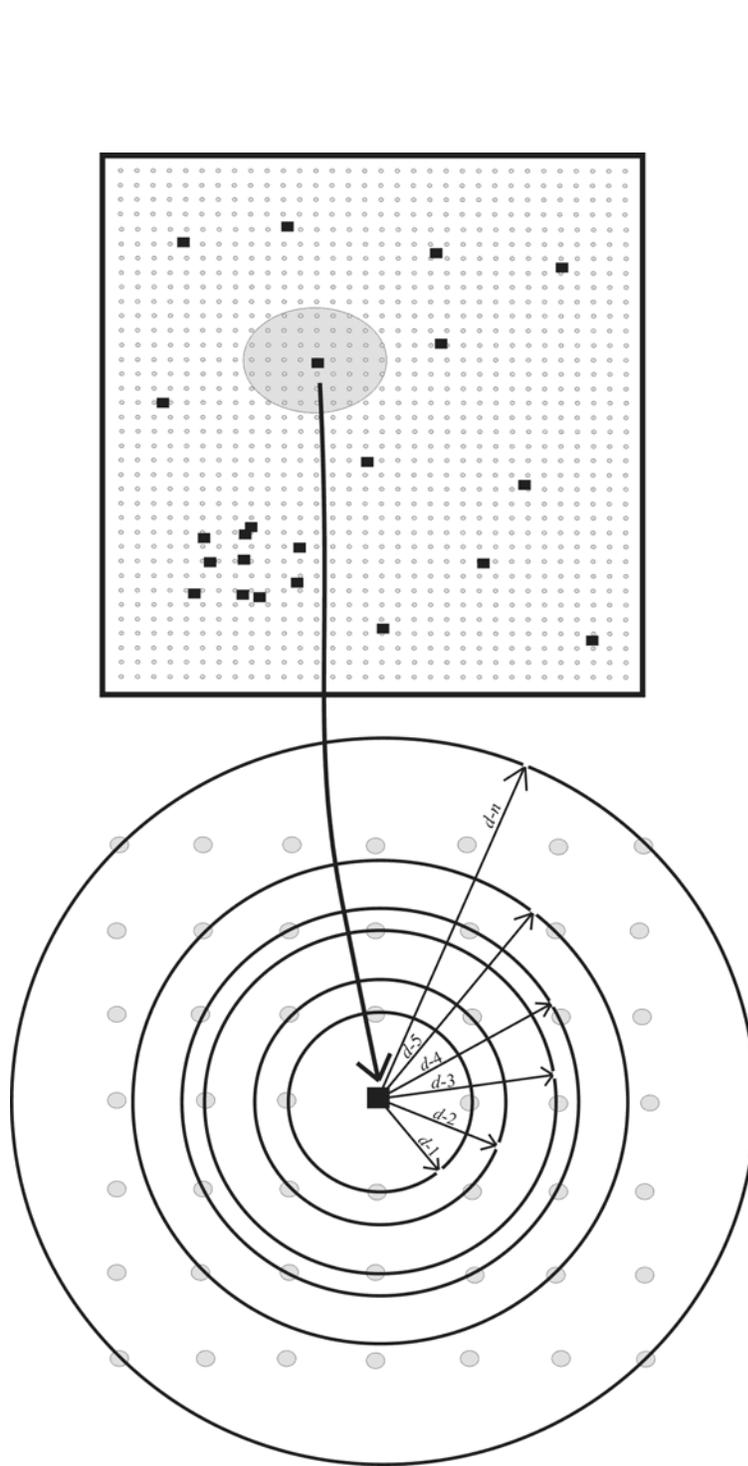
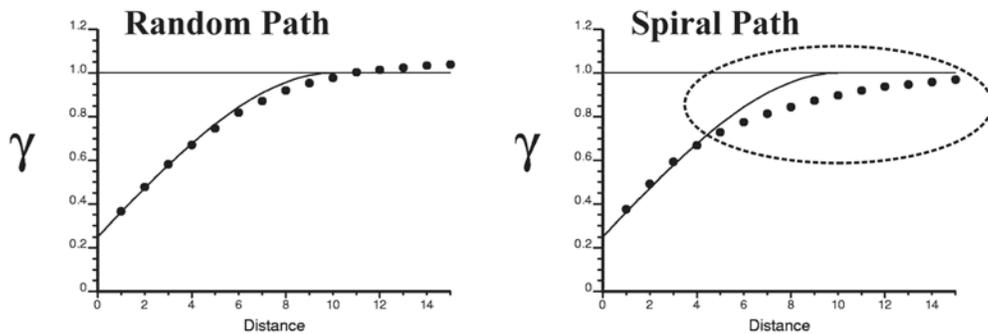
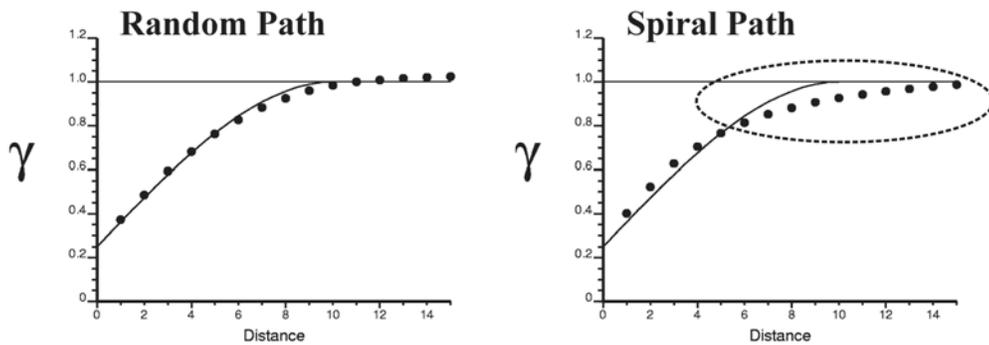


Figure 7 – Spiral Path. An illustration of the procedure to establish a spiral simulation path. A blown up version of the shaded area shows a central sample datum surrounded by some of its locations to be simulated. The distances between the sample datum and the point locations are calculated and summarized by the family of circles of radii $d-1$, $d-2$, $d-3$, ..., $d-n$ shown, where n is the number of different distances. The order of a spiral simulation path is defined by simulating groups of point locations according to an ascending radii order. A spiral path in 3-D would be constructed similarly, except the distances would be summarized as spheres.

Data Configuration (1) - A Single Central Sample



Data Configuration (2) - Four Samples



Data Configuration (3) - 29 Samples (data.dat)

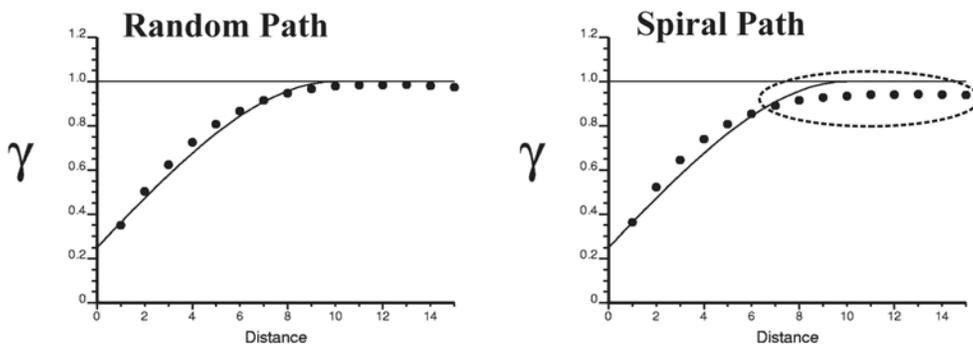


Figure 8 – Spiral Path Simulation Results. The results of implementing the spiral path for three different sample data configurations. The average of 100 variograms from 100 random path and 100 spiral path realizations are calculated (shown as solid dots) and compared relative to each other and the model (solid line). In all cases, the spiral path variogram departs from the model variogram (see the points enclosed by the broken-lined circles) while the random path variograms all satisfactorily reproduce the model.

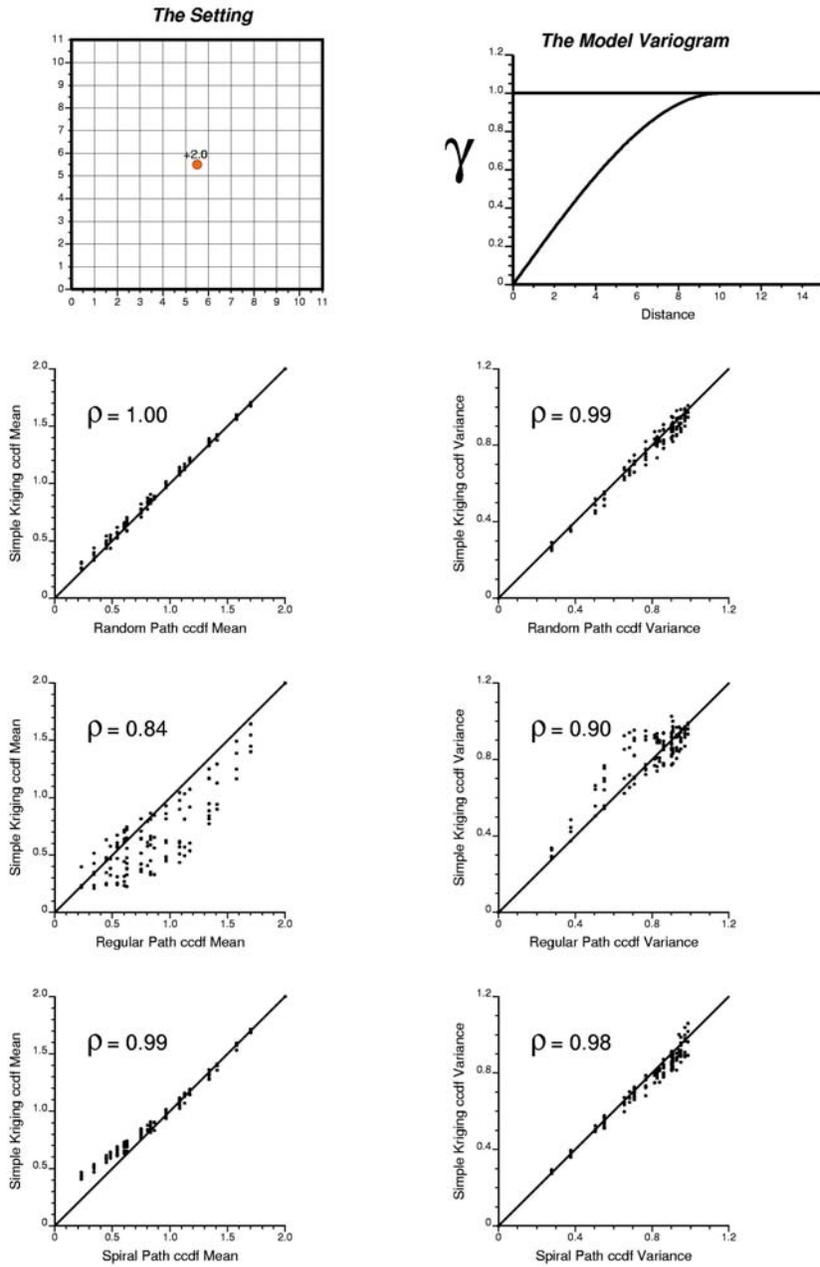


Figure 9 – Local Conditional Cumulative Distribution Functions (ccdfs). An illustration of a local ccdf analysis for regular and spiral simulation path implementations. The 121 (11x11) 1x1m block locations are to be simulated with an isotropic spherical variogram of range 10m (top right) and a centrally located sample with a normal score deviate of +2.0. The ccdf means and variances created from the average of 1000 random, regular and spiral path realizations are paired with the true means and variances created by simple kriging. Compared to a random path, the regular path ccdfs are less correlated to the true ccdf means and variances. The spiral path ccdfs shift lower and become wider (the means and variances are underestimated and overestimated, respectively) for locations further away from the central sample. The results re-emphasize the notion of not using a regular or spiral simulation path for practical implementation.