Variogram Reproduction in Sequential Simulation: Interaction Between Screening and Search Strategy

Olena Babak

Centre for Computational Geostatistics (CCG)
Department of Civil and Environmental Engineering
University of Alberta

Sequential simulation techniques such as sequential Gaussian simulation and sequential indicator simulation, work with an input variogram model to create realizations that reproduce local data, an input histogram and a variogram model. A limited search is often used to manage computer resources. This causes the variogram to be poorly reproduced. Reasons for such unsatisfactory reproduction are considered in this paper. A link between the perfect screening property of the exponential variogram and inadequate reproduction of input variograms is established. A method for finding the exponential variogram reproduced in such simulation is provided. Several small examples illustrating the methodology are considered.

Introduction

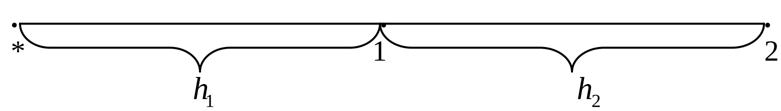
Stochastic simulation is widely used to quantify uncertainty in regionalized variables. Multiple equiprobable realizations of the spatial distribution of either continuous or categorical attributes are used for resource and reserve assessment (Leuangthong, Ortiz and Deutsch, 2004-2005).

The variogram is an extremely important input to geostatistical simulation. It is a measure of spatial correlation, geometry and continuity of the variable of interest. Appearance, behavior and predictions from the created geostatistical models as well as spatial continuity of the variables we model depend heavily on modeling and reproduction of the variogram (Deutsch, 2002).

In this paper we investigate the commonly observed problem of spherical variogram reproduction by sequential simulation with limited search. We explain the exponential shape of the variograms reproduced by sequential simulation with limited search by the perfect screening property of the exponential variogram. We also propose an efficient method for finding the exponential variogram that will be reproduced by sequential simulation with limited search strategy with an input spherical variogram prior even to conducting any simulation. We illustrate our methodology with several small examples.

Perfect Screening of the Exponential Variogran

Let us consider the problem of estimation of an unknown value of the variable of interest at location u^* using the data at locations u_1 and u_2 based on Simple Kriging, when all three locations lie on one line. The data configuration is shown below.



Assume that the variogram of the data is simple (single structure) isotropic exponential,

$$\gamma(h) = 1 - e^{-\frac{h}{a}},$$

where a denotes the range of correlation. Then, the estimate at the location u^* is given by

$$Y^* = \lambda_1 Y_1 + \lambda_2 Y_2 + [1 - \lambda_1 - \lambda_2]m,$$

where Y_1, Y_2 denote data values of the variable of interest at the locations u_1 and u_2, m is the known stationary mean and λ_1, λ_2 are weights given to data at locations u_1 and u_2 calculated from the well known system of normal equations,

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} C_{1*} \\ C_{2*} \end{bmatrix}. \qquad (1)$$

In the case of exponential variogram, for the data configuration shown above the elements in the right-hand side covariance vector and the left-hand side covariance matrix of equation (1) are given by

$$C_{11} = C_{22} = 1,$$

$$C_{12} = C_{21} = e^{-\frac{h_2}{a}},$$

$$C_{1*} = e^{-\frac{h_1}{a}},$$

$$C_{2*} = e^{-\frac{(h_1+h_2)}{a}}.$$

Then, it is obvious that the weight corresponding to the second data Y_2 is equal to 0, since

$$\lambda_2 = \frac{\begin{vmatrix} C_{11} & C_{1*} \\ C_{21} & C_{2*} \end{vmatrix}}{\begin{vmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{vmatrix}} = \frac{\begin{vmatrix} 1 & e^{-\frac{h_1}{a}} \\ e^{-\frac{h_2}{a}} & e^{-\frac{(h_1+h_2)}{a}} \end{vmatrix}}{\begin{vmatrix} 1 & e^{-\frac{h_2}{a}} \\ e^{-\frac{h_2}{a}} & 1 \end{vmatrix}} = \frac{e^{-\frac{(h_1+h_2)}{a}} - e^{-\frac{(h_1+h_2)}{a}}}{1 - e^{-2\frac{h_2}{a}}} = 0$$

So, the estimate at the location u^* does not depend on the data value at the location u_2, and is determined entirely by the data at the location u_1 and the stationary mean, that is,

$$Y^* = \lambda_1 Y_1 + [1 - \lambda_1]m,$$

We will refer to this property of the exponential variogram as *perfect screening*. The perfect screening property is a property that holds only for exponential variogram; it does not hold for any other variogram type. Moreover, perfect screening cannot be observed when using ordinary kriging for estimation of the value of the variable of interest. This follows directly from the system for the ordinary kriging weights.

Variogram Reproduction in Sequential Simulation with Limited Search Strategy

The input variogram is not always reproduced in sequential simulation such as sequential Gaussian simulation (SGS). The lack of reproduction is worse when a limited search is used. Figure 1 shows an example. Specifically, Figure 1 shows the reproduction of the input spherical variogram:

$$\gamma(h) = 0.8\,Sph_{a_1=20}(h) + 0.2\,Sph_{a_2=60}(h)$$

based on an unconditional SGS with number of simulated data used for subsequent sequential generation of the unknown values equal to 8. Clearly, the variograms of the simulated values do not have the shape of the input spherical variogram. The variograms have a shape similar to the exponential variogram. This may seem quite surprising at first. There is, however, a very simple explanation for this phenomenon. Using the limited search strategy is equivalent to using a large number of data, but setting the weight corresponding to most of them (to data beyond a certain search radii) equal to zero. This is equivalent to using an exponential variogram, which has a property of perfect screening.

To verify this observation and conclusion, a method for finding an exponential variogram reproduced by sequential simulation with limited search strategy is proposed in the next section. Further, we restrict our attention to considering only input spherical variograms.

Method for Finding Exponential Variogram reproduced by Sequential Simulation with Limited Search Strategy

It can be noted from Figure 1 that the reproduction of the input spherical variogram to the stochastic sequential simulation at small lag distances is good. Thus, the exponential variogram that is truly reproduced by sequential simulation should confirm the same behavior. Mathematically, this means that the gradients of the input spherical variogram and reproduced exponential variogram are equal for some small distance, that is

$$\frac{\partial \gamma_{Sph}(h)}{\partial h}\Big|_{h=h_0} = \frac{\partial \gamma_{Exp}(h)}{\partial h}\Big|_{h=h_0}, \tag{2}$$

where $\gamma_{Sph}(h)$ denotes the input spherical variogram consisting of k structures each with contribution c_i and range of correlation a_i, that is,

$$\gamma_{Sph}(h) = \sum_{i=1}^{k} c_i\, Sph_{a_i}(h) = \sum_{i=1}^{k} c_i \left[1.5\frac{h}{a_i} - 0.5\frac{h^3}{a_i^{\,3}}\right]; \tag{3}$$

$\gamma_{Exp}(h)$ denotes the reproduced exponential variogram with an unknown range of correlation b,

$$\gamma_{Exp}(h) = Exp_b(h) = 1 - e^{-\frac{h}{b}}; \tag{4}$$

and $h_0 = 2s$, with s equal to the size of the block partitioning used in sequential simulation (this relation was determined experimentally).

Note that using definitions (3), (4) directly, we can rewrite equality (2) as

$$1.5\sum_{i=1}^{k} c_i \left[\frac{1}{a_i} - \frac{h_0^2}{a_i^{\,3}}\right] = \frac{1}{b} e^{-\frac{h_0}{b}}. \tag{5}$$

The range of correlation of the reproduced exponential variogram can be found by solving Equation (5).

It must be noted that when there is a significant difference between ranges $a_i,\ i = 1,\ldots,k,$ of the input spherical variogram (for instance, the range of the first structure is 5 or more times smaller than the range of the second structure), a single structured exponential variogram (4) may not yiled the behavior of the reproduced variograms by the sequential simulation. Therefore, in that case, a condition that the exponential variogram is a single structure needs to be relaxed. Instead, it needs to be assumed that the exponential variogram reproduced by sequential simulation has the same number of nested structures with the same contributions $c_i,\ i = 1,\ldots,k$, as the input spherical variogram and only the ranges $b_i,\ i = 1,\ldots,k,$ of the exponential nested structures are unknown,

$$\gamma_{Exp}(h) = \sum_{i=1}^{k} c_i \, Exp_{b_i}(h).$$

Then in addition to equation (2), we need to impose *k-1* additional constraints to find all *k* unknown ranges of correlation in order to find the exponential variogram reproduced by sequential simulation. In the case of two structured input spherical variogram, the second condition is the equality of spherical variogram and reproduced exponential variogram at lag distance h_1 corresponding to $c_1 + \frac{c_2}{2},$ that is,

$$\gamma_{Sph}(h_1) = \gamma_{Exp}(h_1),$$

or

$$\sum_{i=1}^{2} c_i \left[1.5\frac{h_1}{a_i} - 0.5\frac{h_1^{\,3}}{a_i^{\,3}}\right] = \sum_{i=1}^{2} c_i \left[1 - e^{-\frac{h_1}{b_i}}\right].$$

To illustrate this approach to find the exponential variogram that is reproduced when applying sequential simulation with limited search strategy and input spherical variogram, several small examples will be considered in the next section.

Small examples

Several simulation studies are documented to investigate the proposed method for finding the exponential variogram reproduced by sequential simulation with a limited search strategy. In all cases unconditional SGS was applied with the number of simulated data used for subsequent sequential generation of the unknown property values equal to 8. In order to reduce the impact of ergodic fluctuations, the size of the domain for simulation was chosen several (at least 6) times larger then the largest range of correlation.

Specifically, data was simulated on the grid of 512 x 512 (262144 in total) cells of size 1 by 1 in the X and Y directions, respectively. Twelve different input spherical variograms were considered:

$$\gamma_{Sph,1}(h) = 0.3\,Sph_{a_1=20}(h) + 0.7\,Sph_{a_2=60}(h);$$
$$\gamma_{Sph,2}(h) = 0.8\,Sph_{a_1=20}(h) + 0.2\,Sph_{a_2=60}(h);$$
$$\gamma_{Sph,3}(h) = 0.7\,Sph_{a_1=16}(h) + 0.3\,Sph_{a_2=64}(h);$$
$$\gamma_{Sph,4}(h) = 0.5\,Sph_{a_1=36}(h) + 0.5\,Sph_{a_2=84}(h);$$
$$\gamma_{Sph,5}(h) = 0.6\,Sph_{a_1=30}(h) + 0.4\,Sph_{a_2=84}(h);$$
$$\gamma_{Sph,6}(h) = 0.5\,Sph_{a_1=26}(h) + 0.5\,Sph_{a_2=64}(h);$$
$$\gamma_{Sph,7}(h) = 0.4\,Sph_{a_1=10}(h) + 0.6\,Sph_{a_2=80}(h);$$
$$\gamma_{Sph,8}(h) = 0.3\,Sph_{a_1=10}(h) + 0.7\,Sph_{a_2=60}(h);$$
$$\gamma_{Sph,9}(h) = 0.8\,Sph_{a_1=18}(h) + 0.2\,Sph_{a_2=64}(h);$$
$$\gamma_{Sph,10}(h) = 0.3\,Sph_{a_1=50}(h) + 0.7\,Sph_{a_2=84}(h);$$
$$\gamma_{Sph,11}(h) = 0.6\,Sph_{a_1=30}(h) + 0.4\,Sph_{a_2=64}(h);$$
$$\gamma_{Sph,12}(h) = 0.3\,Sph_{a_1=5}(h) + 0.7\,Sph_{a_2=64}(h).$$

The reproduction of all six considered variograms is shown in Figure 2. This figure also shows the respective exponential fits obtained by method outlined above. These exponential variograms are also listed below

$$\gamma_{Exp,1}(h) = Exp_{a_1=69.2}(h);$$
$$\gamma_{Exp,2}(h) = Exp_{a_1=40.1}(h);$$
$$\gamma_{Exp,3}(h) = Exp_{a_1=35.3}(h);$$
$$\gamma_{Exp,4}(h) = Exp_{a_1=94.8}(h);$$
$$\gamma_{Exp,5}(h) = Exp_{a_1=74.8}(h);$$
$$\gamma_{Exp,6}(h) = Exp_{a_1=68}(h);$$
$$\gamma_{Exp,7}(h) = 0.4\,Exp_{a_1=14.6}(h) + 0.6\,Exp_{a_1=119.5}(h);$$
$$\gamma_{Exp,8}(h) = 0.3\,Exp_{a_1=16.2}(h) + 0.7\,Exp_{a_1=87.9}(h);$$
$$\gamma_{Exp,9}(h) = Exp_{a_1=36}(h);$$
$$\gamma_{Exp,10}(h) = Exp_{a_1=133.5}(h);$$
$$\gamma_{Exp,5}(h) = Exp_{a_1=70.2}(h);$$
$$\gamma_{Exp,6}(h) = 0.3\,Exp_{a_1=6}(h) + 0.7\,Exp_{a_1=96.2}(h).$$

It is clear from Figure 2 that the reproduction of the input spherical variogram in the sequential Gaussian simulation with limited search strategy is not great. The shape of the reproduced variograms is exponential. Moreover, it is seen that the method for finding exponential variogram reproduced by sequential simulation with limited search strategy performs quite well. The exponential variograms closely match the reproduced variogram. This, of course, confirms the above made conclusion about the correspondence between perfect screening of the exponential variogram and the limited search strategy in sequential simulation with input spherical variogram.

It was also noted in this study that with increase in the number of simulated values in subsequent simulation of other values in the domain of interest, spherical variogram reproduction by sequential simulation improves largely (the shape of the reproduced variograms no longer resembles an exponential). Moreover, with very large number of data, e.g., 128 data, the reproduction of the input variogram is correct (see Figures 3 and 4 for the reproduction of spherical variograms 1 and 12, respectively). So, in general, we can conclude that in order to obtain good reproduction of the input variogram to sequential simulation, large search strategy (number of previously simulated values) should be used as a rule.

Conclusions

The problem of unsatisfactory variogram reproduction by sequential simulation techniques was considered. The limited search strategy causes the problem. A perfect screening property of the exponential variogram was used for the explanation for this commonly observed phenomenon. A method for finding the exponential variogram reproduced by sequential simulation with limited strategy was also proposed. This method was shown to perform very well in predicting exponential variograms in several small simulation studies.

This helps us understand variogram reproduction in sequential simulation. The solution was already known – use a large number of data, say 48, when using SGS or SIS.

References

Deutsch, C.V., 2002: *Geostatistical Reservoir Modeling*, Oxford University Press.

Leuangthong, O., Ortiz J.M., and Deutsch, C.V: 2004-2005, *On the Scaling and Use of Multivariate Distributions in Geostatistical Simulation*, Centre for Computation Geostatistics, Report7.

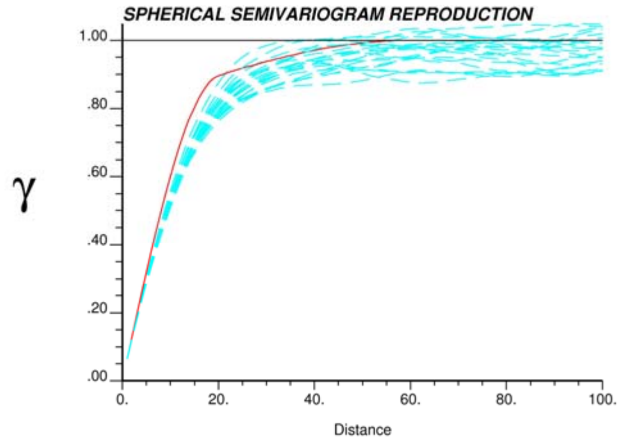


Figure 1: Commonly observed problem in reproduction of the input spherical variogram to sequential simulation with limited search strategy. Solid red line denotes the input spherical variogram to sequential simulation. Blue dashed lines denote the variograms reproduced by sequential simulation with limited search strategy.

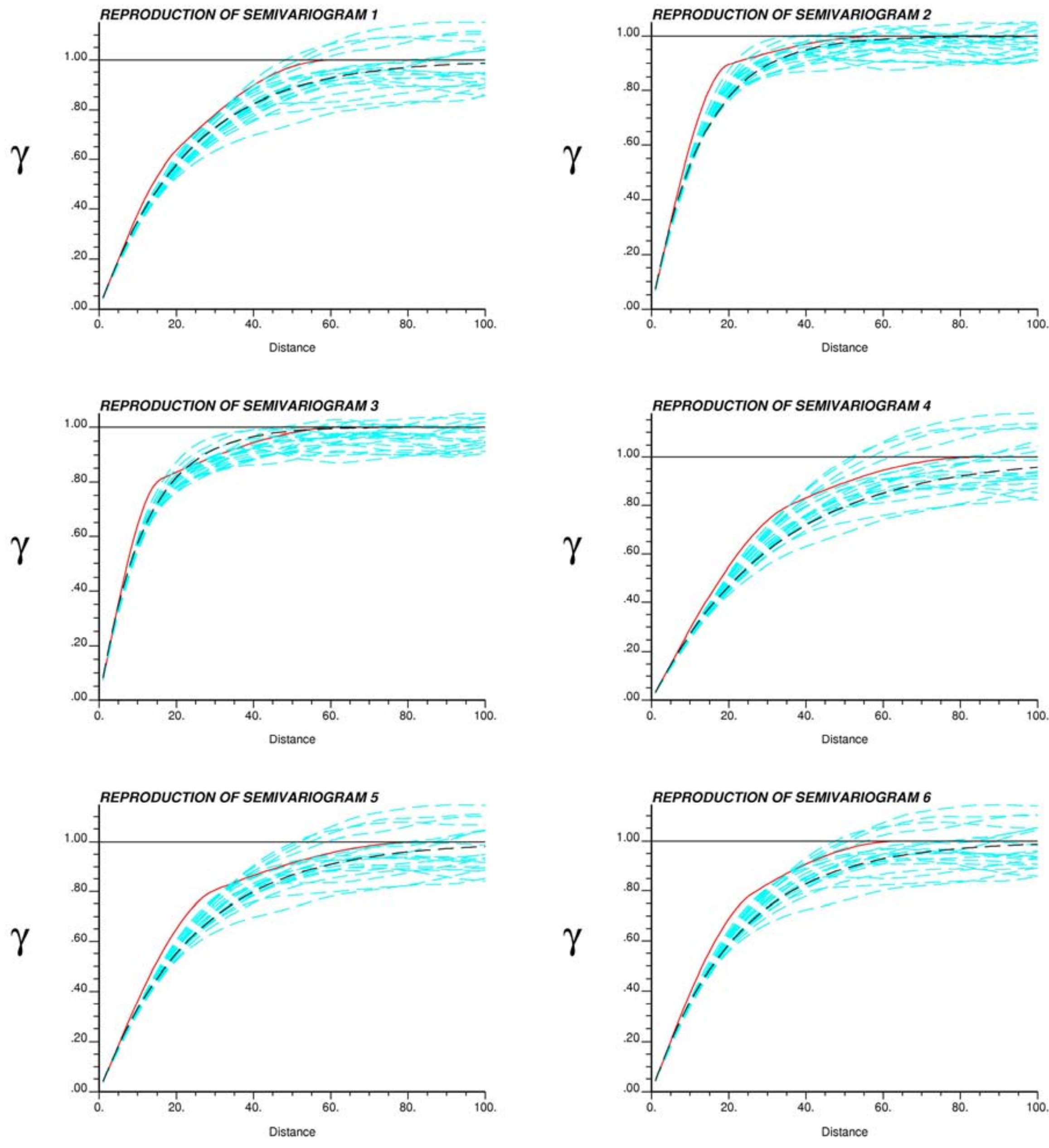


Figure 2: Reproduction of the input spherical variogram by the sequential simulation with limited search strategy for the six cases considered in small examples section. Solid red lines denote the input spherical variograms. Blue dashed lines denote the respective variograms reproduced by sequential simulation with limited search strategy. Black dashed lines denote the proposed exponential variograms. (1/2)

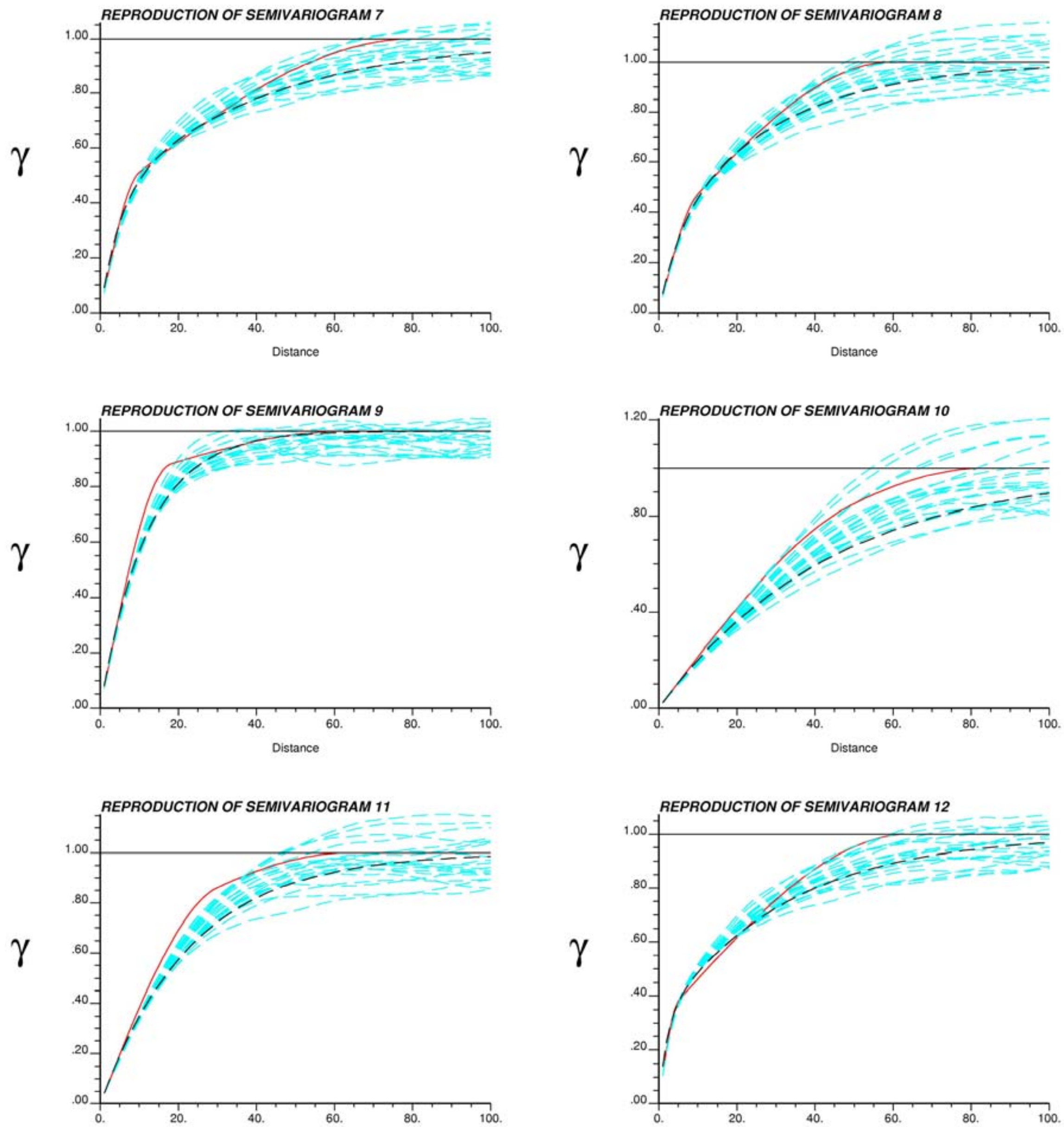


Figure 3: Reproduction of the input spherical variogram by the sequential simulation with limited search strategy for the six cases considered in small examples section. Solid red lines denote the input spherical variograms. Blue dashed lines denote the respective variograms reproduced by sequential simulation with limited search strategy. Black dashed lines denote the proposed exponential variograms. (2/2)

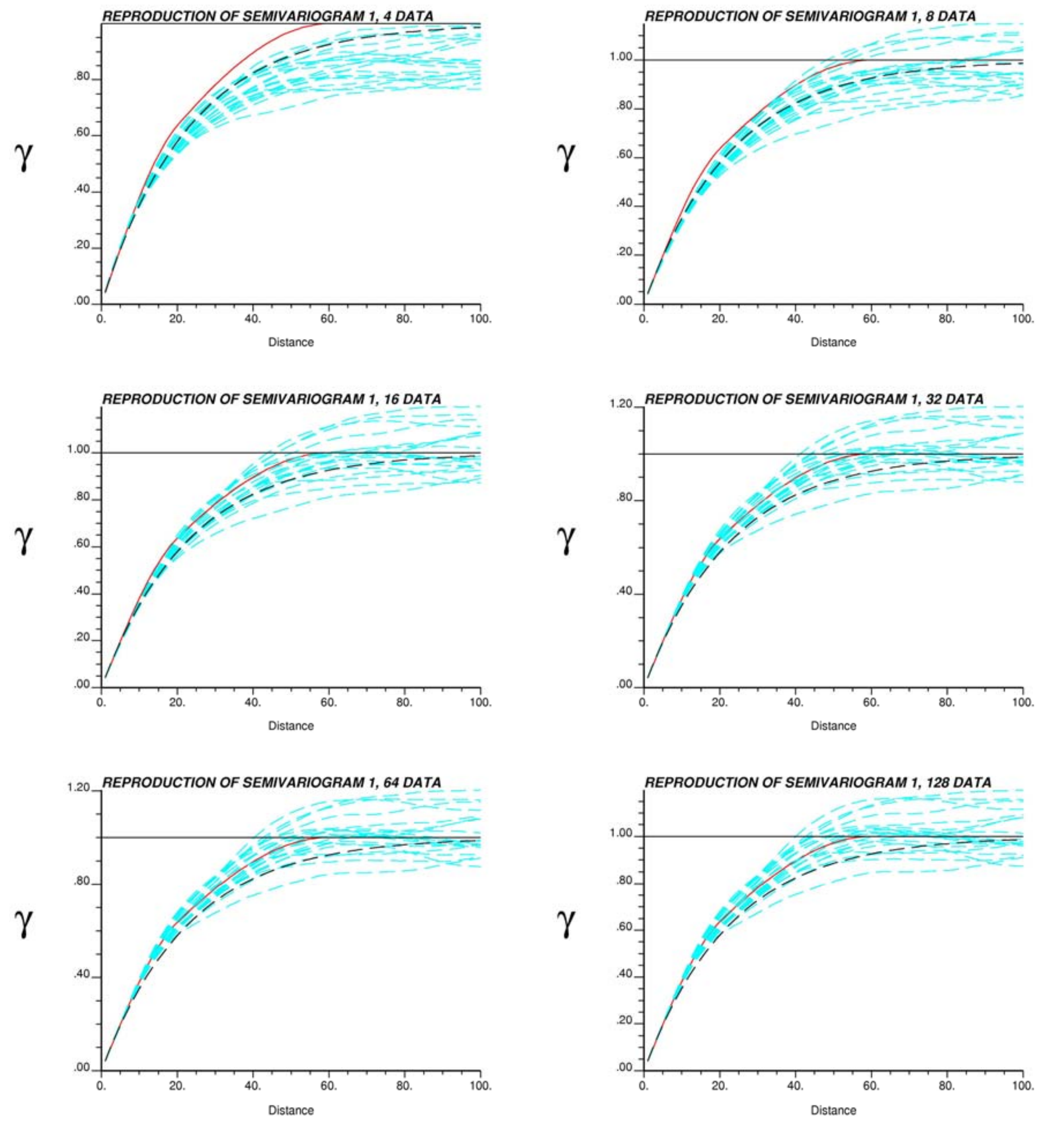


Figure 4: Reproduction of the input spherical variogram 1 (see text) by the sequential simulation with increase of the number of data used in simulation. Solid red lines denote the input spherical variogram. Blue dashed lines denote the respective variograms reproduced by sequential simulation. Black dashed lines denote the exponential variogram reproduced by sequential simulation with limited search strategy.

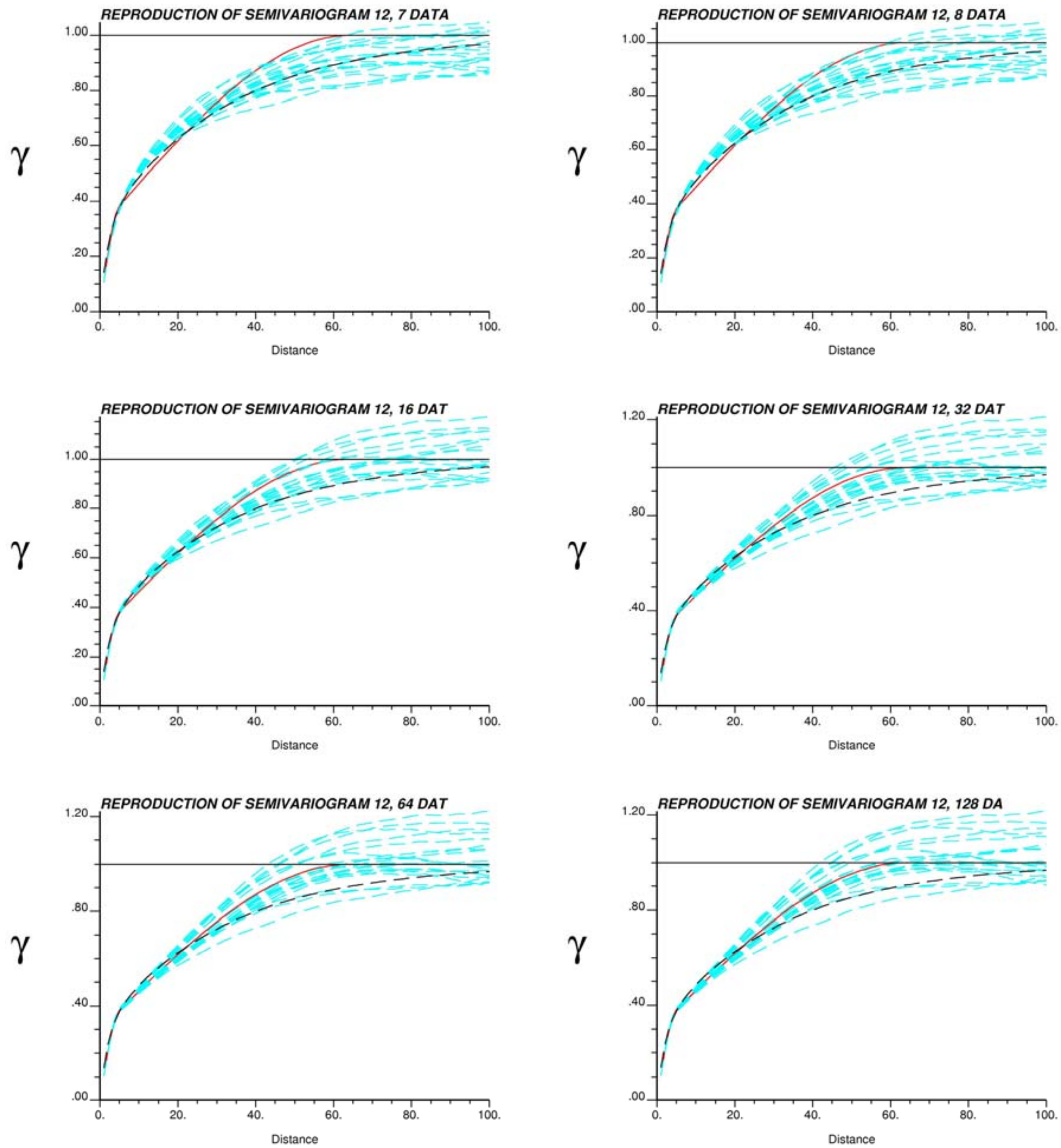


Figure 5: Reproduction of the input spherical variogram 12 (see text) by the sequential simulation with increase of the number of data used in simulation. Solid red lines denote the input spherical variogram. Blue dashed lines denote the respective variograms reproduced by sequential simulation. Black dashed lines denote the exponential variogram reproduced by sequential simulation with limited search strategy.