A Practical Approach to Validate the Variogram Reproduction from Geostatistical Simulation

Julián M. Ortiz and Oy Leuangthong

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

The variogram model is one of the most relevant parameters in geostatistical estimation and simulation methods. The sample variogram is inferred from available data, which may be subject to spatial bias and proportional effect. Over this sample variogram, a licit variogram model is fit and is carried through the process of estimation and/or simulation of the random function usually without regard to its uncertainty. The simulated realizations are required to adequately reproduce this input variogram model. We propose a methodology to test the validity of the output variograms from a suite of realizations computed using a reference variogram model. The test is based on a multivariate Gaussian hypothesis for the resulting variogram values at different lags. Hotelling's T^2 statistic is used to verify the hypothesis that the mean sample variogram vector is equal to the vector of input variogram values for a set of lags. The T^2 statistic is distributed as a random variable with F-distribution with p and n-p degrees of freedom for a given confidence level α. A simple methodology is presented that requires the computation of simple statistics of the output realizations and can be easily implemented. The test can be used to tune the search parameters used for simulation, such as maximum number of samples and previously simulated nodes used for computing the conditional distribution at every node. Two examples show the proposed test. The results are discussed with emphasis on limitations and future research.

Introduction

Geostatistical methods for estimation and simulation call for a variogram model that characterizes the spatial variability of the random function considered. This variogram model is obtained from fitting a licit function to the experimental variogram values calculated from the available sample data.

Estimation methods such as kriging use the variogram model to measure the closeness and redundancy of the data with each location being estimated. In the case of simulation, the variogram model is an input to compute the conditional distributions from which simulated values can be drawn. Furthermore, the simulated random fields should reproduce the spatial character of the phenomenon, that is, they should be characterized by the same variogram.

The variogram model called for in estimation and simulation is inferred based on an estimated variogram determined from the data. Matheron (1965) first proposed a method-of-moments approach to estimate the sample variogram:

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{N(\mathbf{h})}\sum_{i=1}^{N(\mathbf{h})}\left[z_i(\mathbf{u}) - z_i(\mathbf{u}+\mathbf{h})\right]^2$$

where $N(\mathbf{h})$ is the number of pairs of data separated by a vector **h,** and **u** denotes a location within a domain $\mathcal{A}$. This numerical approximation laid the foundation for most theoretical and practical development in the area of variogram modeling and uncertainty.

Since the data are not exhaustive and often show clusters, the estimation of the sample variogram is not completely certain. Given its importance in geostatistical methods such as change of support, kriging and simulation, it is not surprising that the issue of variogram uncertainty and fitting has been extensively covered in the literature. However, the variogram uncertainty is rarely accounted for when applying estimation and simulation.

The objective of this study is to develop a formal approach to validate geostatistical simulations by comparing the resulting variograms over multiple realizations with the reference variogram model, in order to suggest the acceptance or rejection of the computed realizations as correct numerical representations of the phenomenon under study.

In the next section, we review the main advances on variogram estimation, fitting and uncertainty determination. Next, we discuss a methodology to perform hypothesis testing on the variograms calculated from simulated realizations as compared with a reference model. The methodology is then illustrated with some examples, to show its implementation. Finally, the usefulness of these results for validating simulated realizations is discussed and further avenues of research are suggested.

Variogram Estimation and Its Uncertainty

Variogram Fitting

Davis and Borgman (1979) developed the characteristic function of the variogram estimator, $\hat{\gamma}(\mathbf{h})$, for an equally-spaced, one-dimensional, stationary Gaussian random function (RF) model. They tabulated the sample distribution of the variogram estimator, using a Finite Fourier Transform inversion. In 1982, Davis and Borgman further proved that the distribution of sample of variogram is indeed asymptotic:

$$L\left\{\frac{\hat{\gamma}(\mathbf{h})-\gamma(\mathbf{h})}{\sigma[\gamma(\mathbf{h})]}\right\} \rightarrow N(0,1) \text{ as } N(\mathbf{h}) \rightarrow \infty$$

Many authors have focused on the derivation of the variance/covariance matrix of the experimental variogram, mainly with the purpose to determine an optimum fit for the variogram. David (1977) proposed the use of an ordinary least squares approach to minimize $\sum_{i=1}^{nh}\left[\hat{\gamma}(\mathbf{h}_i)-\gamma(\mathbf{h}_i)\right]^2$, where *nh* is the number of lags. Cressie (1985) later approximated the variance of the variogram estimates for a Gaussian variable

$$\text{Var}[\hat{\gamma}(\mathbf{h})] \approx \frac{2[2\gamma(\mathbf{h})]^2}{N(\mathbf{h})}$$

These were then used for variogram fitting using a weighted least squares (WLS) approach, where the weights account for the numbers of pairs within each class. It can be shown that the variogram estimator, for a Gaussian variable, is a linear combination of independent χ-square random variables, each with one degree of freedom (Cressie, 1991):

$$2\hat{\gamma}(\mathbf{h}) = \sum_{i=1}^{n}\lambda_i(\mathbf{h})\chi_{1,i}^2$$

Cressie goes on to show the use of this result for the robust estimation of the variogram.

Several authors have since proposed similar least squares approximation approaches to fitting the variogram: Genton (1998) proposed a generalized least squares error (GLSE), Bogaert and Russo (1999) devised a least squares parametric estimator to estimate the variance-covariance matrix of the variogram estimator, for the purpose of optimizing a sample design to improve estimation of variogram parameters. Other approaches to variogram fitting include maximum likelihood, Bayesian method and fuzzy modeling (Chilès and Delfiner, 1999).

Switzer (1984) proposed to linearly transform the data to give uncorrelated quantities with constant variance, confidence limits were then calculated to estimate certain variogram parameters including the scale, nugget effect, range and general shape.

Variogram Uncertainty

Variogram uncertainty stems from the fact that the fitting is performed considering a sample variogram that has a finite number of lag vectors computed, for a limited number of directions, and from sample data that may not be completely representative of the entire domain of interest, due to clusters, preferential sampling or biased sampling.

Computing the uncertainty of the variogram may be useful when quantifying uncertainty through geostatistical modeling, as it provides an additional source of variability to the numerical characterization of the phenomenon. This uncertainty calculation is generally done under a strong multivariate Gaussian assumption.

Ortiz and Deutsch (2002) developed an analytical expression for the pointwise variogram uncertainty, that is, for the uncertainty of the variogram for a particular lag value. They calculated the uncertainty in the variogram by assuming a known variogram model. They showed that the uncertainty in the variogram is the average covariance between pairs of pairs used to calculate the variogram for the particular lag:

$$\sigma^2_{2\hat{\gamma}(\mathbf{h})} = \frac{1}{N^2(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \sum_{j=1}^{N(\mathbf{h})} C_{ij}(\mathbf{h})$$

where

$$C_{ij}(\mathbf{h}) = Cov\left\{\left[Z(\mathbf{u}_i) - Z(\mathbf{u}_i + \mathbf{h})\right]^2, \left[Z(\mathbf{u}_j) - Z(\mathbf{u}_j + \mathbf{h})\right]^2\right\}$$

$C_{ij}(\mathbf{h})$ is a four-point covariance that can be computed from two-point covariances if the random variables are considered multivariate Gaussian. This approach can therefore be valid for the variogram of normally transformed values, as used when performing Gaussian simulation.

Pardo-Igúzquiza and Dowd (2001) showed a very thorough review of the approaches used for assessing the variograms uncertainty. They derived the variance-covariance matrix of the experimental variograms, $[Cov\{\hat{\gamma}(\mathbf{h}), \hat{\gamma}(\mathbf{h}')\}]$, and compared it with approximate solutions proposed by Cressie (1985). They also built upon the limit distribution of the experimental variograms derived by Davis and Borgman (1982) to propose a simultaneous confidence interval for the experimental points of the variograms. Similar to Ortiz and Deutsch's (2002) approach, this required examination of a fourth-order statistic; however, the expression developed by Pardo-Igúzquiza and Dowd accounts for the joint uncertainty of the variogram between different lags.

Marchant and Lark (2004) conducted simulation experiments to estimate variogram uncertainty for two simulation field sizes and three different sampling schemes. Integral to their approach was the use of a generalized least squares (GLS) approach to variogram fitting. Based on these experiments, they concluded that Pardo-Igúzquiza and Dowd's approach (2001) provided a good estimate of variogram uncertainty due to ergodic errors.

Proposed Methodology

After simulation, variogram reproduction is often checked by calculating directional variograms from multiple realizations and comparing against the input model variogram for the same directions. Gaussian simulation is the most common simulation technique used in practice; as such, variogram reproduction in this context is performed in Gaussian units, that is, prior to back transformation to original units. This check is relatively straightforward and often visually verified. We propose to apply a multivariate hypothesis test, based on Hotelling's T^2-statistic to quantify whether the fit is acceptable within a 95% confidence level (that is, α=0.05).

A number of lags p are chosen for verification, and the resultant n realization variograms form the sample variogram values for each lag. Considering all relevant lags taken together requires a joint test of how well the variogram model is being reproduced. We construct a hypothesis test based on the sample mean variogram values and compare against the model variogram value for that lag distance. The null hypothesis is constructed as:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$$

where $\boldsymbol{\mu}$ is the *p x 1* mean vector obtained from the realization variograms for *p* different lags, and $\boldsymbol{\mu_0}$ is the *p x 1* mean vector based on the model variogram for the same *p* lag distances. The test is constructed to measure the squared distance of the sample mean from the reference mean, standardized by the sample covariance. In the univariate context, the familiar Student t-distribution is common to perform the hypothesis test. The hypothesis is rejected if the observed t-statistic value exceeds a specific percentage threshold of the t-distribution with the same degrees of freedom (Johnson and Wichern, 1998). For this multivariate context, the Hotelling's T^2 statistic is distributed as (Johnson and Wichern, 1998):

$$\frac{(n-1)p}{n-p}F_{p,n-p}$$

where $F_{p,n-p}$ represents an *F*-distribution with *p* and *n-p* degrees of freedom; *p* is the number of lag distances to check, and *n* is the number of sample variogram values available at each lag and corresponds to the number of realizations generated. Under this multivariate context, the null hypothesis is rejected if

$$T^2 = n(\overline{\mathbf{x}} - \boldsymbol{\mu_0})'\mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu_0}) > \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$$

where **S** is the sample covariance matrix and calculated as

$$\frac{\sum_{i=1}^{n}(\mathbf{x_i} - \overline{\mathbf{x}})(\mathbf{x_i} - \overline{\mathbf{x}})'}{n-1}$$

In summary, the methodology proposed for variogram validation in simulation is:

- Define a variogram model (reference)
- Perform Gaussian simulation to obtain *n* realizations
- Compute the experimental variograms from each realization for a number of lags *p*
- Calculate the sample *p x 1* mean vector of variogram values for the realizations $\overline{\mathbf{x}}$
- Calculate the sample *p x p* covariance matrix of variogram values
- Compute the Hotelling's T^2 statistic
- Compute the value for the *F*-distribution with *p* and *n-p* degrees of freedom
- Reject the hypothesis that the mean vector of variogram values from the realizations is equal to the input variogram if $T^2 > \frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)$

Examples

As a first example, consider a stationary Gaussian random function with the following input variogram model for simulation (see Figure 1):

$$\gamma(h) = 0.10 + 0.90Sph_{\substack{ah\max=100 \\ ah\min=50}}(h)$$

Suppose that we now want to check variogram reproduction say for only the maximum continuity direction, by considering only four lag distances (p = 4): 20, 40, 60 and 80m. We calculate the realization variograms and determine the experimental values for each of these four lags for each of ten realization (n = 10) (see Table 1).

From Table 1, we can calculate the average variogram value for each lag, taken over the 10 realizations. We can also determine the reference variogram values from the input model corresponding to these same lag distances:

$$\bar{X} = \begin{bmatrix} 0.358456 \\ 0.584801 \\ 0.769379 \\ 0.895639 \end{bmatrix} \qquad \mu = \begin{bmatrix} 0.3664 \\ 0.6112 \\ 0.8128 \\ 0.9496 \end{bmatrix}$$

$$\bar{X} - \mu = \begin{bmatrix} -0.007944 \\ -0.026399 \\ -0.043421 \\ -0.053961 \end{bmatrix}$$

The information in Table 1 also permits us to calculate the sample covariance matrix *S* and its inverse:

$$S = \begin{bmatrix} 0.000893221 & 0.00294439 & 0.00483074 & 0.00631575 \\ 0.00294439 & 0.0104177 & 0.0178715 & 0.023472 \\ 0.00483074 & 0.0178715 & 0.0331223 & 0.0460079 \\ 0.00631575 & 0.023472 & 0.0460079 & 0.0677182 \end{bmatrix}$$

$$S^{-1} = \begin{bmatrix} 66286.953 & -44449.673 & 26691.548 & -8909.735 \\ -44449.673 & 33192.171 & -21393.422 & 7175.513 \\ 26691.548 & -21393.422 & 14891.916 & -5191.758 \\ -8909.735 & 7175.513 & -5191.758 & 1885.902 \end{bmatrix}$$

Given the information above, we can now calculate Hotelling's T^2 statistic as:

$$T^2 = 10 \begin{bmatrix} -0.007944 \\ -0.026399 \\ -0.043421 \\ -0.053961 \end{bmatrix}' \begin{bmatrix} 66286.953 & -44449.673 & 26691.548 & -8909.735 \\ -44449.673 & 33192.171 & -21393.422 & 7175.513 \\ 26691.548 & -21393.422 & 14891.916 & -5191.758 \\ -8909.735 & 7175.513 & -5191.758 & 1885.902 \end{bmatrix} \begin{bmatrix} -0.007944 \\ -0.026399 \\ -0.043421 \\ -0.053961 \end{bmatrix}$$
$$= 0.841$$

We now compare this to the corresponding *F*-distribution value for α=0.05:

$$\begin{aligned} \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha) &= \frac{(10-1)*4}{10-4} F_{4,6}(0.05) \\ &= \frac{9*4}{6}(4.53) \\ &= 27.18 \end{aligned}$$

Since T^2=0.841 < 27.18 above, therefore we cannot reject the null hypothesis and this variogram reproduction remains possibly valid. Notice that not rejecting the hypothesis does not ensure that variogram reproduction is adequate.

Now suppose we extend this same example by considering 100 realizations and checking the hypothesis for 10 different lags from 10 to 100m in increments of 10m. The resulting T^2 statistic is 27.762 while the *F*-distribution value for *p*=10, *n*=100 and α=0.05 gives a value of

$$\begin{aligned} \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha) &= \frac{(100-1)*10}{100-10} F_{10,90}(0.05) \\ &= \frac{99*10}{90}(1.94) \\ &= 21.34 \end{aligned}$$

Since T^2=27.762 > 21.23, therefore we reject the null hypothesis and the variogram reproduction is deemed poor. We can conclude from this example that if we do not consider a sufficient number of lags, the hypothesis may yield false positive results.

As a second example, we consider the case where the simulation is modified to permit a larger number of simulated values to be used in inference, considering the same variogram model as before. For this case, we use 48 previously simulated values for the simulation (compared to the 12 used in the previous example). The resulting variogram reproduction plots are shown in Figure 3. Once again, 100 realizations are constructed, and the T^2 statistic is calculated for 20 lags ranging from 6 to 120m in increments of 6m. The resulting T^2 statistic for this case is 11.57. Compare this now with the F-distribution value:

$$\begin{aligned}\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha) &= \frac{(100-1)*20}{100-20}F_{20,80}(0.05)\\ &= \frac{99*20}{80}(1.70)\\ &= 42.08\end{aligned}$$

Since T^2=11.57 < 42.08, therefore we cannot reject the null hypothesis and these simulation results cannot be deemed to be poor reproductions.

Discussion

The proposed methodology proves useful for validating the output realizations from a Gaussian simulation method, as compared to an input variogram model computed from the data. Defining the appropriate number of samples and previously simulated nodes in the search parameters for constructing the conditional distributions at every node, can be guided by this test.

Currently, the T^2 statistic is calculated for one direction at a time. This implies that the practitioner would have to run this test for the three principal directions. Of course, if any one direction yields a rejection result, this is an indicator to the practitioner that the simulation parameters should be revisited to improve results.

Checking too few lags may yield misleading results and lull the practitioner into believing that the variogram reproduction is adequate. Checking more lags should mitigate this possibility.

Checking too few realizations will inevitably yield hypothesis test results that may be unreliable; this is a consequence of the unreliability of the average variogram values and sample variogram covariance based on the simulated realizations. In practice, a minimum of 20 realizations should yield stable values in the average variogram and the sample covariance matrix of the simulation variograms.

Note that the hypothesis test can only determine cases where we reject the null hypothesis, that is, we can only conclude those cases where the reproduction is considered to be too poor to satisfy a certain level of confidence, 1-α. In instances where the null hypothesis is not rejected, this does *not* imply that the reproduction is good. It merely states that it is *possibly* good.

Future Work

This research opens several new avenues. Undoubtedly designing a test that permits considering two or more directions simultaneously is of interest. A similar test for the fitted variogram model as compared with the sample variogram would be of interest, prior to inference by kriging or simulation. This would be used as a means to test a goodness-of-fit of the variogram model to the sample variogram values that accounts for the inherent variogram uncertainty. This would call for the analytical determination of the variance-covariance matrix of variogram values for several lags. This test would allow verifying the validity of extreme scenarios where the fitting of the variogram is very continuous or very erratic, provided these models pass the test.

A plot similar to the accuracy plot could be considered, but applied to the variogram instead of the uncertainty distribution of simulated values. This would consider multiple lags wherein the L realization values at each lag would determine the distribution of uncertainty in the variogram. We would test the 'accuracy' of the reproduction by considering different probability intervals and simply count the frequency that the input model falls within the probability interval (Deutsch, 1997). Precision and accuracy of the variogram reproduction would be interpreted in a similar fashion and this plot would allow adjusting the fitting to provide a reliable variogram model for subsequent modeling steps.

References

Bogaert, P. and Russo, D., “Optimal Spatial Sampling Design for the Estimation of the Variogram based on a Least Squares Approach”, *Water Resources Research*, 35 (4), 1999, p. 1275-1289.

Chilès, J.P. and Delfiner, P., *Geostatistics: Modeling Spatial Uncertainty*, John Wiley & Sons Inc., New York, 1999, 695 pp.

Cressie, N., “Fitting Variogram Models by Weighted Least Squares”, *Mathematical Geology*, 17 (5), 1985, p. 563-586.

Cressie, N., *Statistics for Spatial Data*, John Wiley & Sons Inc., New York, 1991, 900pp.

David, M., *Geostatistical Ore Reserve Estimation*, Elsevier, Amsterdam, 1977.

Davis, B.M. and Borgman, L.E., “Some Exact Sampling Distributions for Variogram Estimators”, *Mathematical Geology*, 11 (6), 1979, p. 643-653.

Davis, B.M. and Borgman, L.E., “A Note on the Asymptotic Distribution of the Sample Variogram”, *Mathematical Geology*, 14 (2), 1982, p. 189-193.

Deutsch, C.V., “Direct assessment of local accuracy and precision”. In “*Geostatistics Wollongong '96*”. Baafi, E.Y., Schofield, N.A. (Eds.), Kluwer Academic Publishers, 1997, pp. 115-125.

Genton, M.G., “Variogram Fitting by Generalized Least Squares Using an Explicit Formula for the Covariance Structure”, *Mathematical Geology*, 30 (4), 1998, p. 323-345.

HyperStat Online, http://davidmlane.com/hyperstat/F_table.html, Accessed on October 11, 2006.

Johnson, R.A. and Wichern, D.W., 1998: Applied Multivariate Statistical Analysis, Prentice Hall, New Jersey, 4th Edition.

Marchant, B. P. and Lark, R. M., “Estimating Variogram Uncertainty” ”, *Mathematical Geology*, 36 (8), 2004, p. 867-898.

Matheron, G. *La Théorie des Variables Régionalisée et ses Applications*, Masson, Paris, 1965.

Ortiz, J.M. and Deutsch, C.V., “Calculation of Uncertainty in the Variogram”, *Mathematical Geology*, 34 (2), 2002, p. 169-183.

Pardo-Igúzquiza, E. and Dowd, P., “Variance-Covariance Matrix of the Experimental Variogram: Assessing Variogram Uncertainty”, *Mathematical Geology*, 33 (4), 2001, p. 397-419.

Switzer, P., “Inference for Spatial Autocorrelation Functions”, in Verly, G., David, M., Journel, A.G. and Marechal, A., eds., *Geostatistics for Natural Resources Characterization*, D. Reidel Publishing Company, Dordrecht, 1984, pp. 127-140.

Table 1: Realization variograms for 10 realizations at four different lag distances in the maximum continuity direction.

Realization	Lag			
	20	40	60	80
1	0.362980	0.591920	0.788360	0.941170
2	0.351010	0.564000	0.722170	0.811390
3	0.350950	0.558470	0.701530	0.770120
4	0.362190	0.577510	0.736480	0.844610
5	0.368190	0.624720	0.845510	1.009400
6	0.345340	0.543540	0.713100	0.836570
7	0.374620	0.638110	0.841650	0.971500
8	0.368120	0.628210	0.863890	1.025590
9	0.349860	0.563060	0.746390	0.866170
10	0.351300	0.558470	0.734710	0.879870

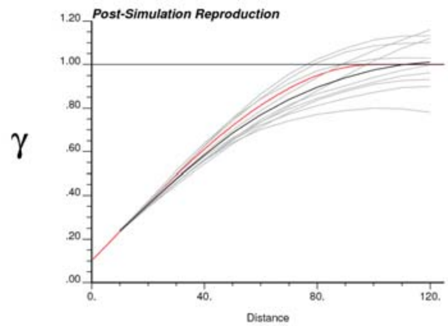

Figure 1: Post-simulation check for variogram reproduction based on 10 realizations. The red variogram shows the input model, the black line shows the average variogram of the simulated realizations, and the gray lines correspond to the variogram from each of the 10 realizations.

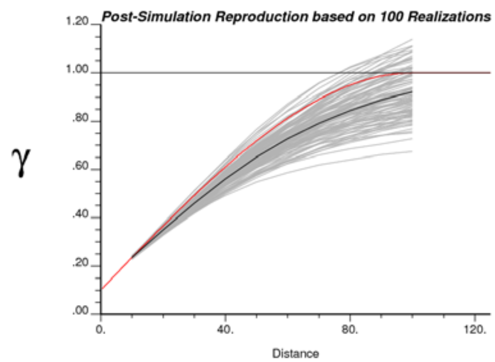

Figure 2: Post-simulation check for variogram reproduction based on 100 realizations. The red variogram shows the input model, the black line shows the average variogram of the simulated realizations, and the gray lines correspond to the variogram from each of the 100 realizations.

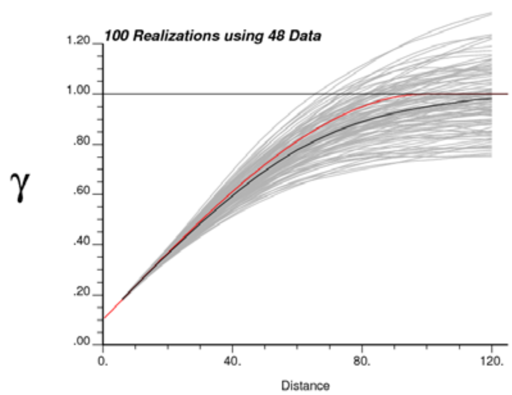

Figure 3: Post-simulation check for variogram reproduction based on 100 realizations using 48 data. The red variogram shows the input model, the black line shows the average variogram of the simulated realizations, and the gray lines correspond to the variogram from each of the 100 realizations.