A Conditional Finite Domain (CFD) Approach to Calculate Variogram Uncertainty

Behrang Koushavand and Clayton V. Deutsch

The variogram is central to many geostatistical studies. Due to limited data available in the early stages of modeling, there is considerable uncertainty in all statistical parameters including the variogram. The uncertainty in the experimental variogram should be quantified and transferred into subsequent modeling. All previously studied methods to assess variogram uncertainty are unconditional to data values and just depend on data locations. Therefore, the uncertainty is unrealistically high. A conditional finite domain based variogram uncertainty algorithm is presented that is more robust and realistic. The configuration of sampled data is kept the same new configurations are generated by translation and rotation of original data. A number of realizations are generated for each of new configuration conditional to the original data, then the experimental variogram is calculated for each configuration and realization. The variance of each lag is calculated. The calculated variogram uncertainty is reasonable.

Introduction

Variogram estimates from a limited data set are uncertain. That uncertainty can be estimated, incorporated into the results, and used in applications. Variogram uncertainty has been considered by different authors. Webster and Oliver (1992), Müller and Zimmerman (1999) and Bogaert and Russo (1999) measured the variogram uncertainty in sampling schemes and suggested different methods to minimize this uncertainty. Cressie (1985), Ortiz and Deutsch (2002), and Pardo-Igúzquiza and Dowd (2001) suggested similar expressions for the covariance matrix of experimental variogram estimates to the ergodic variogram (Marchant and Lark 2004).

The semivariogram (often called the variogram for brevity) is defined under a second order stationarity assumption as (Matheron, 1971):

$$\gamma(\mathbf{h}) = \frac{1}{2} E\left\{ \left[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u}) \right]^2 \right\} \tag{1}$$

It is estimated by the method of moments (Journel and Huijbregts, 1978):

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

where $N(\mathbf{h})$ is the number of pairs of values for the separation vector $\mathbf{h}$.

For a fixed set of K lag vectors, the variance–covariance matrix of the experimental variogram is a K×K matrix Σ. The pq element is (Marchant and Lark, 2004):

$$\left[\Sigma \right]_{pq} = \mathrm{Cov}\left[\hat{\gamma}(\mathbf{h}_p), \hat{\gamma}(\mathbf{h}_q) \right] \tag{3}$$

The diagonal elements give the variance of the variogram estimates:

$$\left[\Sigma \right]_{pp} = \mathrm{Cov}\left[\hat{\gamma}(\mathbf{h}_p), \hat{\gamma}(\mathbf{h}_p) \right] = \mathrm{Var}\left(\hat{\gamma}(\mathbf{h}_p) \right) \tag{4}$$

Several authors (Cressie, 1985; Ortiz and Deutsch, 2002; Pardo-Igúzquiza and Dowd, 2001) assumed the distribution of each lag is Gaussian, and defined by a mean corresponding to the ergodic variogram $\gamma(\mathbf{h})$, and the variance of each lag, given by equation (4). Koushavand (2008) et.al used a two parameter (shape and scale) Gamma distribution that is more flexible, since a gamma distribution is a general form of a Gaussian distribution. This distribution has only positive values that fits better the squared differences of the variogram. It is also asymmetric with flexibility to fit the actual distribution shape. The variance covariance matrix of the variogram has been derived analytically and the variogram joint uncertainty has been generated by correlated gamma distributions. They also present a spatial bootstrap method that converges to the analytical results but is faster with no data limitations.

Methodology

The Conditional Finite Domain (CFD) method is used to generate conditional realizations inside the domain. The sample data configuration is fixed and randomly rotated and translated to generate new configurations inside the domain. A maximum radius (RMax) parameter controls the maximum translation distance. Rotation is allowed within 360 degrees if there is no strong anisotropy in the variogram. New data are simulated conditional to the original data values. Experimental variograms are calculated using simulated data location. This procedure is repeated to a reasonable set of data configurations. The variogram uncertainty will be more realistic than conventional method.

The `Randrot` program has been designed to generate randomly translated and rotated a given data locations to generate a new set of configurations. Figure 1 (top) shows a sample parameter file. Data file is the original data locations. Rmax parameter is the maximum radius of translation, irot is a flag for rotation, if it set to 0 new configurations will not rotated and if it set to 1 new data configurations will generate by translation and rotation. The number of configurations is defined with nconfig.

The `Pickreal` (CCG catalog 2007) program has been modified to read output file of randrot, the set of configurations, and extract the specified configuration.

`Sgsim` (Deutsch and Journel, 1998) program was modified to read location being simulated form an external file and generate conditional realizations. Figure 1 (bottom) shows a sample parameter file.

Case Study

A sub-vertical tabular deposit is drilled to provide data on vein thickness and grades. The thickness values are used here. The sample locations are presented in Figure 2. Figure 3 left shows the histogram of thickness in original units. The following isotropic variogram model is fitted to the omni-directional experimental variogram of the normal score transformation of thickness:

$$\gamma(\mathbf{h}) = 0.15 + 0.85.\mathrm{Exp}\left(\frac{\mathbf{h}}{250}\right) \tag{6}$$

The experimental variogram and fitted model are shown in Figure 3 right. Figure 4 shows experimental variograms of 100 unconditional realizations. Figure 5 shows location map, histogram and variogram reproduction of conditional simulation using original data for 2 data configurations: translated data configuration at left and rotated and translated data configuration at right. To assess the effect of Rmax at uncertainty and find the optimum RMax, various number of RMax were chosen:

0.1,0.5,1,5,10,20,30,40,50,60,70,80,90,100,150,250,250,300,350,400.

50 configurations were generated for each RMax value and for each configuration 50 conditional simulation is generate. For each RMax, 2500 experimental variograms were calculated. The values for each lag were extracted and gamma distribution was fitted to each lags. The results are as expected, that is, as the RMax increases, there is greater uncertainty in the variogram. Also, uncertainty increases as the rotation increases.

The red data set is quite uncertain and has a trend: high values are in the north and low values at south. The mean of most fitted gamma distributions are greater than expected variogram value which is because of trend on the data.

Conclusions

The data are not directly used in the conventional method for variogram uncertainty calculation. The CFD approach is more realistic and data values have a significant impact on variogram uncertainty. Considering rotation for an anisotropic data set will artificially increase uncertainty. The Rmax value should be chosen very carefully based on data spacing.

References

Bogaert, P., and Russo, D., 1999, Optimal spatial sampling design for the estimation of the variogram based on a least squares approach: Water Resour. Res., v. 35, no. 4, p. 1275–1289.

Cressie, N., 1985, Fitting variogram models by weighted least squares: Math. Geology, v. 17, no. 5, p. 563–586.

Davis, B. M., and Borgman L.E., 1982, A note on the asymptotic distribution of the sample variogram: Math. Geology, v.14, no. 2, p. 189–193

Deutsch, C. V., and Journel, A. G., 1998, GSLIB: Geostatistical software library and users guide, 2nd ed.: Oxford University Press, New York, 369 p.

Journel, A. G., and Huijbregts, C. J., 1978, Mining geostatistics: Academic Press, London, 600 p.

Koushavand, B., Ortiz, C. J., and Deutsch, C. V., 2008, A Methodology to Quantify and Transfer Variogram Uncertainty through Kriging and Simulation, In Center for Computational Geostatistics, paper 301, Annual Report 10

Marchant, B. P., and Lark, R. M., 2004, Estimation Variogram Uncertainty: Math Geology, v. 36, no, 8, p. 865-898

Matheron, G., 1971, The theory of regionalized variables and their applications: Ecole des Mines, Fontainbleau, France, 211 p.

Müller, W. G., and Zimmerman, D. L., 1999, Optimal designs for variogram estimation: Environmetrics, v. 10, no. 1, p. 23–37.

Ortiz, C. J., and Deutsch, C. V., 2002, Calculation of uncertainty in the variogram: Math. Geology: v. 34, no. 2, p. 169–183.

Pardo- Igúzquiza, E., and Dowd, P. A., 2001, Variance–covariance matrix of the experimental variogram: Assessing variogram uncertainty: Math. Geology, v. 33, no. 4, p. 397–419.

Webster, R., and Oliver, M. A., 1992, Sample adequately to estimate variograms of soil properties: J. Soil Sci., v. 43, no. 1, p. 177–192.

```
             Parameters for ROTCOORD
             ***********************
START OF PARAMETERS:
data.dat                    -file with data
1   2                       -  columns for original X and Y coordinates
locs.dat                    -file for output with new coordinates
50   1                      -Rmax and irot: Flag of Rotation clockwise (0=No, 1=yes)
100  69069                  -nconfig(number of configurations) and random number seed
```

```
             Parameters for PICKREAL
             ***********************

START OF PARAMETERS:
locs.dat                    -file with simulation
67                          -   number of data
5                           -realization to write out
loc.dat                     -file for output realization
```

```
             Parameters for SGSIM
             ********************
START OF PARAMETERS:
data.dat                    -file with data
1  2  0  4  0  0            -  columns for X,Y,Z,vr,wt,sec.var.
-100.0      1.0e21          -  trimming limits
0    1                      -transform the data (0=no,1=yes) jackknife. (0=no,1=yes)
sgsim.trn                   -  file for output trans table
0                           -  consider ref. dist (0=no, 1=yes)
histsmth.out                -  file with ref. dist distribution
1  2                        -  columns for vr and wt
0.0   15.0                  -  zmin,zmax(tail extrapolation)
1      0.0                  -  lower tail option, parameter
1     15.0                  -  upper tail option, parameter
0                           -debugging level: 0,1,2,3
sgsim.dbg                   -file for debugging output
sgsim.out                   -file for simulation output
SSS                         -number of realizations to generate
loc.dat                     -file with simulate locations
5  6  0                     -  columns for X,Y,Z
69069                       -random number seed
0    20                     -min and max original data for sim
40                          -number of simulated nodes to use
0                           -N/A
1     3                     -multiple grid search (0=no, 1=yes),num
0                           -maximum data per octant (0=not used)
1000  1000 10.0             -maximum search radii (hmax,hmin,vert)
 0.0   0.0   0.0            -angles for search ellipsoid
51    51    11              -size of covariance lookup table
0     0.60   1.0            -ktype: 0=SK,1=OK
none                        -  N/A
0                           -  N/A
1    0.15                   -nst, nugget effect
2    0.85  0.0 0.0 0.0      -ivt, cc, ang1, ang2, ang3
250  250  250               -a_hmax, a_hmin, a_vert
```

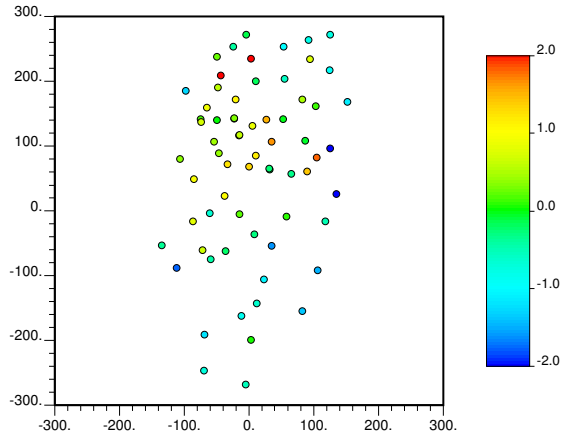**Figure 1.** Parameter files of randrot (top) and sgs_b (bottom**)**

Figure 2. Location map of samples and thickness content taken from the database red.dat.

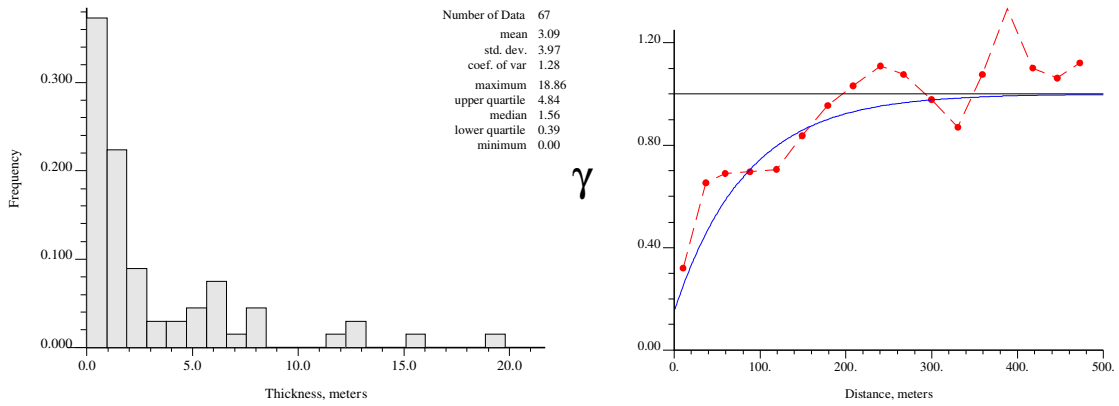|	
Number of Data	67
mean	3.09
std. dev.	3.97
coef. of var	1.28
maximum	18.86
upper quartile	4.84
median	1.56
lower quartile	0.39
minimum	0.00

Figure 3. Histogram of thickness in meters and Experimental variogram (dashed line) and fitted model (solid line) of normal scores of thickness
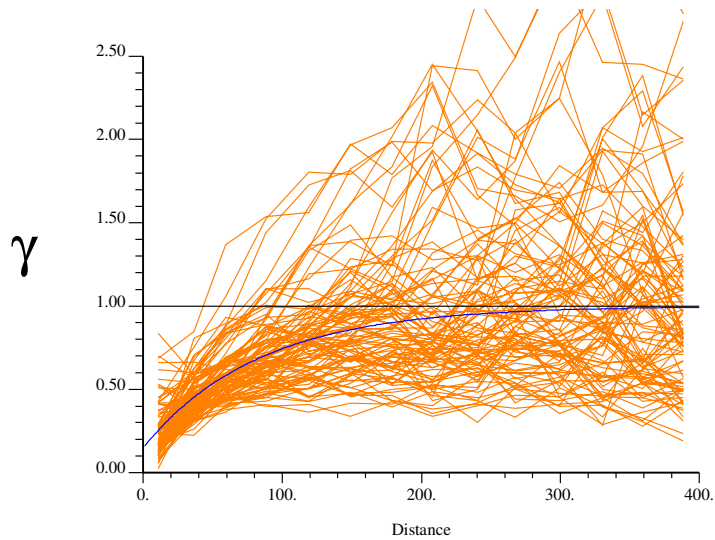

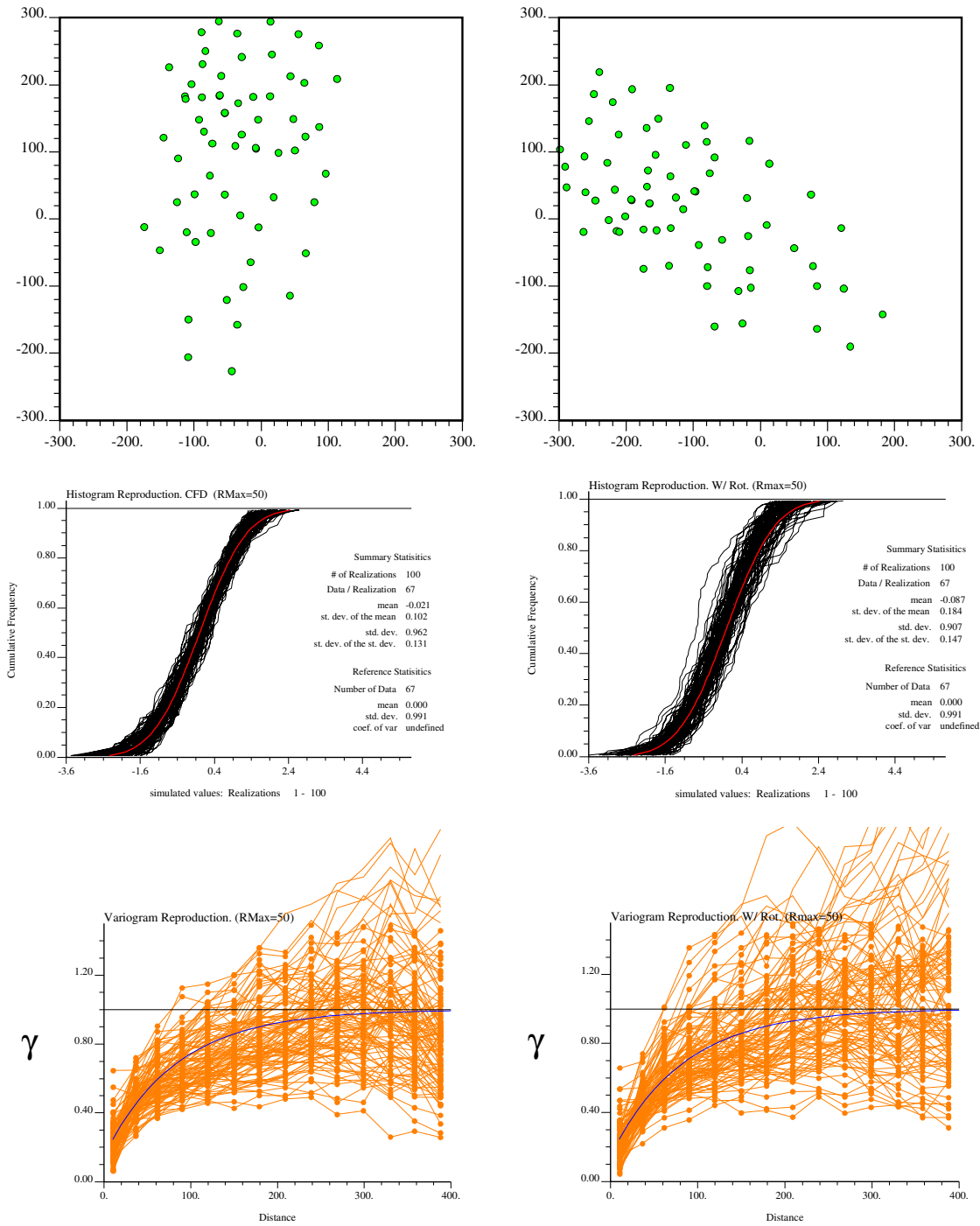**Figure 4.** Unconditional based variogram Uncertainty

Figure 5. Location map, Histogram reproduction and variogram reproduction of sconditional simulated for translated (left) and rotated and translated (right) data set.