Uncertainty Upscaling

Julián M. Ortiz (jmo1@ualberta.ca) and Clayton V. Deutsch (cdeutsch@ualberta.ca)
Department of Civil & Environmental Engineering
University of Alberta

Abstract

Quantification of uncertainty of a spatially distributed variable at any scale can be handled through geostatistical simulation. Large computation time, storage, and post-processing of the realizations are required to obtain a final assessment of block uncertainty.

Multi-Gaussian kriging is a flexible alternative to Gaussian simulation. It computes the kriging estimate and variance after normal score transformation of the original samples. Under the multi-Gaussian assumption, all marginal and conditional distributions are Gaussian, hence fully defined by their mean and variance. These Gaussian conditional distributions are obtained by simple kriging and can be back-transformed to the original units of the variable of interest. An estimate and any quantile can be easily retrieved.

The main disadvantage of performing multi-Gaussian kriging is that change of support is not straightforward. We propose a methodology to overcome this limitation of multi-Gaussian kriging by considering a matrix simulation to generate multiple probability fields. Each probability field is used to draw spatially correlated point values from the point-support conditional distributions, and multiple realizations of the average can be obtained. This permits the calculation of the average over the block and its uncertainty. These blocks may correspond to selective mining units or to volumes from longer production periods relevant for engineering decisions. They can even be disjoint blocks, such as when several faces are mined at the same time. POSTMG, a Fortran program to perform these calculations, is described and a case study is provided.

Introduction

Geostatistical simulation is used to generate optimum estimates under different loss functions and to assess uncertainty at different supports relevant for engineering decisions. A typical application of simulation is to generate grade models in a mine and then evaluate the uncertainty in grade for monthly, quarterly, and yearly production. However, when the models are large, computing time and storage may prohibit the use of simulation. Furthermore, it is rarely possible to simulate more than a few tens of millions of nodes with current computers and software.

The use of estimation methods that allow assessment of local uncertainty can be seen as advantageous. Indicator kriging and multi-Gaussian kriging are two typical approaches to define the point distributions. Unlike simulation methods, there is no requirement to keep
all the previously simulated (or estimated) points in memory, since only the sample data are used to calculate the estimate and its variance. Only one pass is required to fully assess the local uncertainties. With simulation methods, the entire grid must be visited as many times as realizations are required. Finally, postprocessing the results is required in both cases. In the case of simulated models, block averages are calculated for every realization and these are used to build the distribution of uncertainty for the block volume. If estimation methods are used to define the point distributions of uncertainty, a fast simulation method can be used to generate unconditional probability fields that allow the assessment of block uncertainty.

We briefly review the methods to calculate point uncertainty. Then, we discuss the approaches to evaluate the block uncertainty. The methodology and implementation details for the program POSTMG are discussed along with a case study to illustrate its application.

**Point-Scale Uncertainty**

### Indicator Kriging

Indicator kriging permits the direct estimation of the conditional distribution at an unsampled location, that is, its point distribution of uncertainty [6, 7, 10]. It allows the random variable to have different spatial continuity for high and low values, giving more flexibility than Gaussian methods that lock the connectivity of extremes with the single variogram of normal scores.

The indicator formalism requires the data to be coded as probabilities [1]:

\[
i(u_\alpha; z_k) = \begin{cases} 
1, & \text{if } z(u_\alpha) \leq z_k \\
0, & \text{otherwise}
\end{cases} \quad k = 1, \ldots, K
\]

(1)

where \(z(u_\alpha)\) is the value at the data location \(u_\alpha\) and \(z_k, k = 1, \ldots, K\) is the threshold value. Additionally, constraint intervals, soft categorical data, and soft continuous data can be added to improve the estimation of the conditional probability.

At every data location, the \(z\) value is transformed into \(K\) indicator values. The point distribution of uncertainty can be inferred by kriging the indicator function at every threshold. The original \(n\) data are converted into \(K\) sets of \(n\) indicator variables. Each one of those sets of data is used to estimate the value of the indicator at all unsampled locations in the grid, that is, the probability of having \(z(u) \leq z_k, k = 1, \ldots, K\).

To perform simple indicator kriging [14], the stationary mean of the indicator random function is required. This mean is given by the cumulative distribution function of the random function \(Z(u)\):

\[
E[I(u; z_k)] = \text{Prob}[Z(u) \leq z_k] = F(z_k)
\]

The stationary simple kriging estimate of the indicator at that threshold is written:

\[
[i(u; z_k)]_{SK} = [\text{Prob}[Z(u) \leq z_k|(n)]]_{SK}^* = \sum_{\alpha=1}^{n} \lambda_{\alpha}^{SK}(u; z_k) \cdot i(u_\alpha; z_k) + [1 - \sum_{\alpha=1}^{n} \lambda_{\alpha}^{SK}(u; z)]F(z_k)
\]

(2)

where the weights \(\lambda_{\alpha}^{SK}(u; z_k)\) are the unique solution of the simple kriging system:
\[ \sum_{\beta=1}^{n} \lambda_{\beta}^{SK}(u; z_{k}) \cdot C_{I}(u_{\beta} - u_{\alpha}; z_{k}) = C_{I}(u - u_{\alpha}; z_{k}) \quad \alpha = 1, ..., n \] (3)

Notice that a stationary covariance indicator function \( C_{I}(h; z_{k}) \) has to be inferred for each threshold \( k = 1, ..., K \) in Equation 3.

Alternatively, ordinary kriging could be used, where the local mean is estimated from the conditioning data in a neighborhood of the location being estimated, rather than considering the marginal cumulative probability. Median indicator kriging can be used to reduce the requirement of inferring and modelling \( K \) indicator covariances. It can be applied if the \( K \) indicator random functions \( I(u; z_{k}) \) have their indicator variograms and cross variograms proportional to a common variogram model [8].

Since the estimations are made independently for each threshold, there is no guarantee that the set of estimated indicators will give a valid cumulative distribution, hence corrections for order relation deviations are routinely performed [2, 4].

Indicator techniques allow the integration of secondary information, as well as soft data. They provide a flexible and powerful framework to estimate conditional distributions.

Multi-Gaussian Kriging

Multi-Gaussian kriging works by means of a transformation of the sample data into normal scores [16, 17]. Under the assumption of multivariate spatial Gaussianity, all marginal and conditional distributions are Gaussian in shape and only their means and variances are required to fully define them.

The spatially distributed data \( z(u_{\alpha}), \alpha = 1, ..., n \) are declustered and a normal score transformation is performed, generating the normal values \( y(u_{\alpha}), \alpha = 1, ..., n \) (Figure 1).

Variogram calculation and modelling of the normal values is then required. Notice that the values required for variogram inference are the normal score transforms without considering the declustering weights [2].

Kriging is performed to estimate the normal score value at every unsampled location in a regular grid. Kriging provides a mean and a variance. Under the multi-Gaussian assumption, the shape of the conditional distribution is known to be Gaussian, hence the full conditional distribution in the original units of the variable can be retrieved by back-calculating the \( z \) values for given percentiles. The mean can be back-calculated by numerical integration, as illustrated in Figure 2. The full distribution of uncertainty in original units can be retrieved in the same manner. In theory, the available information allows us to calculate the estimate with minimum estimation variance by simple kriging. Ordinary kriging could be used in case there are local variations in the mean, so long as the variable can be considered locally stationary. The estimation variance in this case will be higher than the one obtained by simple kriging. Ordinary kriging could be too conservative when estimating the local uncertainty.

One of the advantages of transforming the variable to normal scores is that there is implicit control over any heteroscedasticity of the variable, that is, change in the local variance as a function of the local mean. In mining applications, this phenomenon is known as the proportional effect. Practice has shown that in general for positively skewed distributions,
Figure 1: The normal score transformation is illustrated for a data $z_i$. The cumulative frequency is read in the original distribution and the value $y_i$ of a standard normal distribution, that is, a Gaussian distribution with mean 0 and variance 1, corresponding to that cumulative frequency is assigned to the data location. The anamorphosis function is shown at the bottom.
Figure 2: Calculation of the mean by numerical integration. The local uncertainty distribution is given by the kriging estimate and variance and the assumption that the shape is normal (bottom right). Several quantiles are calculated in the illustration. The nine deciles of the distribution, $y_1, \ldots, y_9$, are back-transformed (top) and the corresponding values, $z_1, \ldots, z_9$, are used to calculate the mean (bottom left).
the local standard deviation is directly proportional to the local mean. Furthermore, the relationship is in general very close to linear [5, 11].

The process of transforming the value to a standard Gaussian distribution and then back-transforming the conditional distributions “filters” the proportional effect, because of the changing slope in the anamorphosis function (see bottom of Figure 1).

Multi-Gaussian kriging provides an easy approach to calculate the conditional distributions, with the same requirements, in terms of inference, as Gaussian simulation: a representative histogram is transformed and the variogram of the normal scores is calculated and modelled.

**Block Uncertainty**

Assessing block uncertainty can be done using geostatistical simulation methods. Sequential methods work by visiting randomly all the uninformed nodes in the simulation grid. The point distribution of uncertainty is computed at each location given the sample data and previously simulated nodes, and a value is drawn from this distribution. This value is used for all subsequent nodes as conditioning data; this ensures reproduction of the spatial correlation [9]. Due to the requirement to include previously simulated points to condition the subsequent nodes in the random path, storage may become an issue if very large grids are considered (over ten million nodes). Furthermore, the time required to compute multiple realizations may also be a problem. If \( N \) nodes are being simulated for \( L \) realizations, then \( N \cdot L \) kriging systems must be solved. Since the simulated points are also used to condition the estimation in subsequent nodes, the kriging matrices tend to be large in size, increasing the computer time required to invert them. In estimation, only \( N \) kriging systems must be solved and these are in general not as large as the ones considered in the case of simulation, and storage is required only for the sample data as conditioning.

The typical procedure to assess uncertainty over block values using simulation is:

1. \( L \) dense grid simulated realizations are generated.
2. Block values are calculated for each one of the \( L \) realizations.
3. The \( L \) block values are used to construct the distribution of uncertainty (histogram) for the block variable.
4. Mean and any quantile of the block distribution can be retrieved from the set of \( L \) block values.

An alternative to simulation is to use the local distributions of uncertainty to infer the block support uncertainty. In order to combine the point support uncertainty distributions into a single block distribution, the spatial correlation must be taken into account.

**Methodology**

Probability field (p-field) simulation allows combining these point conditional distributions into a block distribution by generating spatially correlated values from the point distributions [3, 15]. These values are averaged to obtain a simulated value that is used to build the
distribution of block values. The idea is to generate spatially correlated probability values that are used to draw from the point distributions. For example, if a strong spatial correlation exists, the probability value used to draw from a point distribution will be similar to the probability value used to draw from the point distribution at a nearby location. This permits generating correlated values from these point distributions.

Because the p-field is used to draw from the conditional distributions, it can be generated unconditional, hence a quick algorithm such as spectral methods or matrix decomposition can be used.

The following methodology is proposed to assess block uncertainty:

- Calculate a representative distribution by declustering the \( z \) data.
- Transform the data to normal scores considering the declustered distribution.
- Infer and model a three-dimensional variogram of the normal scores.
- Calculate the multi-Gaussian kriging estimate and variance at point support by kriging the normal score data using the corresponding variogram of normal scores.
- Define the blocks of interest for uncertainty calculation, based on the objective of the study:
  - Regular blocks can be defined for the calculation of recoverable reserves or for resource or reserve classification.
  - A production volume (irregular shape and probably disconnected) can be considered for assessing uncertainty in average grade for a planned production period.
- For every block of interest, retrieve the points located within the block and their locations.
- Generate multiple unconditional probability fields using a matrix method, with the normal score variogram.
- For every probability field:
  - Generate the simulated values by drawing from the conditional distributions using the probability field.
  - Back-transform every simulated value to the original units, using the transformation table, and average the values in original units to obtain a single realization of the average.
- Pool together the average values for a particular block, and retrieve its variance or any probability interval as a measure of uncertainty.

This methodology has been implemented in a Fortran program called \texttt{POSTMG}, which is described in the \textbf{Appendix}. 


Case Study: Porphyry Copper Deposit

The objective of this study is to show the implementation of the method proposed for calculating the uncertainty at block support from point grade uncertainties.

Drillhole Data

A drillhole database with copper grades in 12m composites from a porphyry copper mine is available for this study. The data base contains East, North and elevation coordinates, and the grade in percent by weight. Figure 3 shows the declustered histogram of composites. 1281 samples are available. The data range from 0 to approximately 7 % Cu and the distribution is positively skewed. The coefficient of variation is approximately 0.5, which can be considered relatively low. It is a typical value for deposits of this type. The median is very close to the mean value.

A plan view of the drillholes at a particular bench is shown in Figure 4. The copper grades are shown at the bench level with a tolerance of 12 m. The average spacing between drillholes is around 50 m. In many zones drillholes are spaced even closer.

Variogram of Normal Scores

The variogram of the normal scores is calculated and modelled. The final model is presented in Table 1 and the experimental and fitted variograms in the three main directions of anisotropy are shown in Figure 5.

Multi-Gaussian Kriging

Multi-Gaussian kriging is performed, that is, the normal scores of the data are used as conditioning information to obtain kriging estimates and variances at point support within the domain of interest.

The grid simulated is defined in Table 2.
Figure 4: Plan view showing the drillhole information for bench 3922.

<table>
<thead>
<tr>
<th>Nugget Effect</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure 1</td>
<td>Spherical</td>
</tr>
<tr>
<td>Sill Contribution</td>
<td>0.15</td>
</tr>
<tr>
<td>Range N30°W</td>
<td>20.0</td>
</tr>
<tr>
<td>Range N60°E</td>
<td>60.0</td>
</tr>
<tr>
<td>Range Vertical</td>
<td>45.0</td>
</tr>
<tr>
<td>Structure 2</td>
<td>Exponential</td>
</tr>
<tr>
<td>Sill Contribution</td>
<td>0.70</td>
</tr>
<tr>
<td>Range N30°W</td>
<td>160.0</td>
</tr>
<tr>
<td>Range N60°E</td>
<td>105.0</td>
</tr>
<tr>
<td>Range Vertical</td>
<td>220.0</td>
</tr>
</tbody>
</table>

Table 1: Normal scores variogram model parameters.

<table>
<thead>
<tr>
<th>Direction</th>
<th>Number of Nodes</th>
<th>Coordinate Centre of First Block (m)</th>
<th>Size (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easting</td>
<td>100</td>
<td>24551.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Northing</td>
<td>150</td>
<td>25201.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Elevation</td>
<td>12</td>
<td>3881.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Table 2: Grid definition.
The search strategy was defined considering 4 and 16 samples as minimum and optimum numbers for estimating a location. The search ellipsoid was defined with its main axis rotated to N30°E. The maximum, minimum and vertical radii that define the ellipsoid were 160, 110, and 220 m.

Figure 6 shows the estimates and variances in transformed units.

Block Distributions

The multi-Gaussian output was post-processed with the program PostMG. Averaging of the point distribution was done to blocks of 10 by 10 by 12 m. The numerical discretization of the Gaussian distributions to obtain the mean, variance and quantiles in original units was set to 200 quantiles. Finally, 100 realizations of the probability fields were considered, from which the simulated point values were drawn, back-transformed and averaged to obtain the distribution of block grades. Figure 7 shows the estimates and variances in original units at point support (top) and at block support (bottom). Notice that the variance does not depend exclusively on the spatial configuration of the data, but also on its local mean.

Possible Applications

These block grade distributions of uncertainty could be used to determine recoverable reserves. One of the output files of PostMG contains quantiles for each block. The number of quantiles reported is set by the user. These distributions could be used to calculate the proportion of each block with a grade higher than a given cutoff. Adding these proportions and calculating the grade of the material above the cutoff grade provides an estimate of the recoverable reserves.

A second possible application is the use of PostMG for classification. The program could be modified to calculate locally an estimated distribution of uncertainty for blocks of different sizes. These could represent production for different periods. Blocks could be
Figure 6: Maps of estimates and variances of normal scores.

Figure 7: Maps of estimates and variance in original units at point support (top) and at block support (bottom) from PostMG.
classified based on their spread around the mean, at a given confidence level.

Comments

Although it is known that p-field simulation generates some artifacts [13], it is still used in practice. The argument to justify its use is that when considering block averages, the bias in the spatial correlation and the local extrema are no longer relevant problems.

It is the opinion of these authors that the bias in the variogram is of importance and the realizations generated with this method do not reproduce the target measure of continuity. It is therefore advisable to correct the variogram used to generate the p-fields to obtain realizations that correctly reproduce the variogram. This has not been implemented in the previous example, but should be considered for future work [12].

Conclusions

A methodology to post-process the point distributions of uncertainty obtained with multi-Gaussian kriging, to obtain block grades estimates with an associated measure of uncertainty is illustrated. The methodology has been implemented for an application to a copper deposit. The program PostMG is provided along with the documentation for its use.

The block grade distributions of uncertainty account for the spatial correlation between the points inside the block, via a p-field simulation of the point grades. The correlated probabilities contained in every realization of the probability field are used to draw from the conditional distributions from multi-Gaussian kriging. The simulated point values are back-transformed with the corresponding global transformation table and averages in original units are calculated. These averages can be pooled together into a histogram that represents the block grade uncertainty. Any summary measure of uncertainty can be retrieved from these histograms, and they can be used for classification, production uncertainty assessment and calculation of recoverable reserves.

References


Appendix: Program POSTMG

A Fortran program has been prepared to post-process multi-Gaussian output. The input information required is: the multi-Gaussian kriging grid with means and variance of the local point distributions, the transformation table to back-transform the Gaussian values simulated, and the variogram of normal scores, which is required to generate the p-fields.

The program requires the following parameters (see also the parameter file in Figure 8):
Parameters for POSTMG
**************************

START OF PARAMETERS:
1 - i=grid, 0=arbitrary prod. volume
kt3d.out - input file with Gaussian mean and var.
1 2 - columns: if 1:m,var; if 0:X,Y,Z,m,var
50 0.5 1.0 - if 1: point support: nx,xmn,xsiz
50 0.5 1.0 - ny,ymn,ysiz
50 0.5 1.0 - nz,zmn,zsiz
5 5.0 10.0 - block support: nbx,bxmin,bxsiz
5 5.0 10.0 - nby,bymn,bysiz
5 5.0 10.0 - nbz,bzmn,bzsiz
nscore.trn - file with transformation table
postMG.out - output file with point m and var in original units
quantiles.out - output file with selected quantiles of point ccdfs
postMGbl.out - output file with block m and var in original units
quantilesbl.out - output file with selected quantiles of block ccdfs
200 - number of quantiles for numerical integration
0.10 - quantile interval
0.0 15.0 - zmin,zmax(tail extrapolation)
1 0.0 - lower tail option, parameter
1 15.0 - upper tail option, parameter
100 - number of realizations for change of support
1 0.1 - nscores variogram: nst, nugget effect
1 0.9 0.0 0.0 0.0 - it,cc,ang1,ang2,ang3
10.0 10.0 10.0 - a_hmax, a_hmin, a_vert

Figure 8: Parameter file for program POSTMG.

- **IGRID**: Flag that indicates if a regular grid is being input (IGRID=1) or scattered points are used (IGRID=0)

- **DATAFL**: Name of file with multi-Gaussian kriging output. If IGRID=1, then only the columns for the mean and the variance of the Gaussian conditional distribution at point support is required for all the points in the grid; if IGRID=0, then the coordinates are also required.

- **COLX,COLY,COLZ,COLM,COLV**: Column number for X, Y, and Z (if required), and for means and variances of local distributions of uncertainty.

- **NX,XMN,XSIZ, NY, YMN, YSIZ, NZ, ZMN, ZSIZ**: Grid definition of the multi-Gaussian kriging input at point support (only required if IGRID=1). The number of nodes NX, NY, and NZ, coordinates of the first point in the grid XMN, YMN, and ZMN,
and spacing of the three-dimensional grid XSIZ, YSIZ, and ZSIZ are required. These values follow the convention of all the GSLIB programs.

- **NBX,BXMN,BXSIZ,NBY,BYMN,BYSIZ,NBZ,BZMN,BZSIZ**: Grid definition for the required block output. The same parameters as for the grid input are required.

- **TRANSFL**: Name of the file with the transformation table. This table can be generated with the program NSCORE in GSLIB. It contains the original values and normal score transforms in two columns without a header.

- **OUTFL,OUTFL2,OUTFL3,OUTFL4**: Names of the four output files generated. The file OUTFL contains the mean and variance for the point support in original units. The file OUTFL2 contains the values of the variable for a set of quantiles specified by the user (see next). These two files also include the coordinates of the points if IGRID=0. The file OUTFL3 has the mean and variance in original units at block support. Finally, the file OUTFL4 writes out the quantiles for the block distributions. The files OUTFL3 and OUTFL4 output values for a single block if IGRID=0, since all the points are considered to belong to the same production volume.

- **NDISC**: Discretization of the point distributions for numerical integration. This allows improvement in the accuracy of the back-transformation of the full point support distribution of uncertainty.

- **NQUANT**: Quantile interval for reporting. This number must be greater than 0 but smaller than 1. A typical value would be 0.10, in which case the nine deciles are output to OUTFL2 and OUTFL4 for the point and block distributions.

- **ZMIN,ZMAX**: Minimum and maximum values for tail extrapolation during the back-transformation of the quantiles.

- **LTAI,LTPAR,UTAIL,UPAR**: Parameters for the extrapolation of the tails. LTAIL defines the model to extrapolate the lower tail. LTAIL=1 entails a linear extrapolation to the value LTPAR; LTAIL=2 implies a power model with parameter LTPAR. Similarly, UTAIL indicates the type of extrapolation for the upper tail. UTAIL=1 indicates a linear model up to the value UPAR; UTAIL=2 implies a power model with power UPAR; UTAIL=4 considers a hyperbolic extrapolation with parameter UPAR. In the case of the hyperbolic extrapolation, the values are truncated at ZMAX.

- **NSIM**: Number of realizations of the p-fields to compute the block support statistics.

- **NST(1),C0(1),IT(i),CC(i),ANG1(i),ANG2(i),ANG3(i),AA(i),AA1,AA2**: The normal scores variogram model parameters. As with all programs in GSLIB, NST(1) corresponds to the number of structures, C0(1) is the nugget effect, IT(i) is the variogram type for the structure i, CC(i) is its sill contribution, ANG1(i), ANG2(i), and ANG3(i) are the rotation angles for the principal directions of anisotropy, AA(1), AA1, and AA2 are the ranges in the directions of maximum continuity (h_{max}), minimum continuity (h_{min}), and perpendicular to both (\textit{vert}).