ROSSI, M.E. Practical aspects of large-scale conditional simulations. *Application of Computers and Operations Research in the Minerals Industries,* South African Institute of Mining and Metallurgy, 2003.

Practical aspects of large-scale conditional simulations

M.E. ROSSI
GeoSystems International, Milpitas, California, USA

All large-scale applications of conditional simulations intend to benefit from the 'correct' (or 'accurate') characterization of uncertainty, adequately describing the variability observed from the data and its impact on the final objective. This paper discusses a number of relevant issues that highlight the extent to which, in reality, a model of uncertainty is dependent on the random function model chosen. Often in the presence of large number of conditioning data most available random function models will provide similar models of uncertainty. When conditioning is scarce, the resulting models of uncertainty will vary more widely, since the final output will depend mostly on the underlying assumptions of the model chosen.

The application described in this paper is taken from a large porphyry Cu operation in northern Chile, and it demonstrates the impact of several variables on the resulting models of uncertainty. In particular, some of the decisions involved include the number of conditioning data and the underlying Random Function model used (Gaussian or indicator-based). Also, implementation-specific parameters must be decided upon, such as grid size used, random number generator, simple or ordinary kriging, multiple-grid simulation, and simulation of irregular three-dimensional bodies (as opposed to regular grids). Some practical guidelines are proposed to deal with these decisions.

Introduction

Geostatistical conditional simulations are becoming increasingly popular as tools that provide models of uncertainty at different stages of a mining project. They have been used as grade control tools in daily operations (Rossi[1]), to analyse risk related to resource classifications (Rossi and Camacho[2]), to assess the uncertainty of minable reserves at the project's feasibility stage (Van Brunt and Rossi[3]), and to assess mineralization potential in certain settings. Other applications include assessment of recoverable reserves and drill hole spacing optimization studies.

For a general background of the theory of geostatistical conditional simulations the reader is referred to Goovaerts[4] or Journel[5]. Geostatistical conditional simulations are used to build models that reproduce the full histogram and modelled measures of spatial continuity of the original, conditioning data. Therefore, they honour the spatial characteristics of the spatial variable of interest as represented by the three-dimensional sample (conditioning) data. A 'Large-Scale' conditional simulation is arbitrarily defined here as a deposit-wide or industrial-scale simulation (see for example Rossi[6]), generally consisting of several million nodes (simulating 20 to 70 million nodes is relatively common). These types of simulations are typically used in a well-scoped risk assessment study, or other similar application.

The simulation model should correctly represent the proportion of high and low values, the mean, the variance, and other univariate statistical characteristics of the data, as represented by the histogram. It should also correctly reproduce the spatial continuity of the variable, including the connectivity of low and high grade zones, anisotropies, relative nugget effect, and other characteristics of the variogram model.

Conditional simulations are built on fine grids, as fine as possible given the hardware available, so that they correspond to approximately the support size of the original samples. The vertical resolution of the grid should be a function of the support data, for example the size of the mining bench, if modelling a variable mined by open pit. Larger grid sizes may still be used sometimes because of the amount of computer time and hard disk space involved.

In building a conditional simulation model, many of the conditions and requirements of linear and non-linear estimations apply, most importantly regarding stationarity decisions. Shifts in geologic settings require the separation of the data into different populations. Detailed knowledge of the behaviour of extreme and outlier values in the sampled population is required. Issues such as limiting the maximum simulated grade should be carefully considered.

The simulation method itself should be decided based on the type of deposit, the Random Function model chosen, the quantity and quality of available samples, the possibility of using 'soft' or fuzzy information, and the desired output. All these are, to a large extent, subjective decisions. These and other implementation parameters, along with the chosen algorithm and simulated domain, have a bearing on the output simulations, and the resulting uncertainty model.

A required decision is whether to use a parametric or non-parametric approach. Examples of each are the Sequential Gaussian (Isaaks[7]) and Sequential Indicator (Alabert[8]) simulations. The latter is more complicated, based on multiple indicator kriging techniques (Journel[9]), and requires definition of several indicator cutoffs. The former is simpler and quicker, although more restrictive in its basic assumptions. As with any estimation exercise, spatial

continuity measures, such as correlograms, should be obtained and modelled.

When a number of conditional simulations have been run and checked, for each node defined in the grid, there are the same number of node values available. These set of grades for each node, all equi-probable by construction, are interpreted to describe the model of uncertainty for each block, i.e., a *cumulative conditional distribution function* (ccdf) curve. Preferably, a large number of simulations are needed to describe this curve better. However, and due to practical limitations, a much smaller number is generally used as an initial approximation. The model of uncertainty developed for each point within the area of interest can be described as (Journel[9]):

$$F(z;\underline{x}\,|\,(n)) = \Pr ob\left\{Z(\underline{x}) \le z\,|\,(n),\ \alpha = 1,...,n\right\} \qquad [1]$$

Here $F(z;\underline{x}|(n))$ represents the cumulative conditional distribution frequency curve for each node x of the simulated grid, obtained using the (n), $\alpha=1,...,n$ conditioning samples. It provides the probability of that point in the grid (represented in the model by the Random Variable $Z(x)$) of being above (or below) any given value z, given those (n) conditioning values. Each simulated value for the node will represent one point in the ccdf described in [1]. One important question in practice is: how many simulations are needed in order to adequately describe the ccdf [1]?

For the case study described here, sequential Gaussian simulations (SGS) were used. SGS is based on a multi-Gaussian Random Function model assumption for the spatial process being simulated. The original data is first transformed into a Gaussian distribution using a Normal Scores transformation, also known as anamorphosis. This process transforms any sample distribution into a univariate Gaussian distribution. Then, tests are performed to validate the Gaussian assumption made, such as the relationship between indicator semi-variograms obtained from the transformed Gaussian samples and the theoretical Gaussian semi-variogram. The test only validates the adequacy of the bivariate Gaussian distribution assumption (see, for example, Goovaerts[1]), representing a necessary but not a sufficient condition for multi-Gaussianity. In practice, however, if the bivariate assumption is accepted, then a multi-Gaussian assumption is applied without any further considerations.

The underlying concept in this paper is that uncertainty is not a property of the physical attribute being modelled, but rather of the Random Function (RF) model developed. This RF is somewhat subjective, and only appropriate or 'correct' if it adequately describes the more important aspects of the variable of interest. The latter are in turn a function of the problem being analysed. Therefore:

- The uncertainty model that can be derived from conditional simulations is subjective, and only relevant to the underlying RF model; and
- Applications that can be derived from the uncertainty model (such as risk assessments) are only useful and 'realistic' if relevant to the specific problems or questions posed by the same, specific RF model.

A common example is the assessment of uncertainty of a block model, used to define resources and reserves of a deposit. The uncertainty model should be built using the same underlying RF model, in order to describe the uncertainty related to the block model built. If an indicator-based approach has been used to build the block model, then the series of conditional simulations have to be also indicator-based. It makes little sense to study the risk of a mine plan based on an ordinary kriged block model using Sequential Indicator Simulations. Similarly, the same geologic model used to constrain the block model has to be used to constrain the simulation model.

This paper will describe the impact on the uncertainty model of some of the parameters used in obtaining conditional simulations per the GSLIB implementation of the Sequential Gaussian Simulation (SGS) algorithm (Deutsch and Journel[10]).

Case study

The example developed below is based on two databases: a blast hole database and a drill hole database, corresponding to a porphyry copper deposit in northern Chile. The examples and results shown below correspond to uncertainty models of Total Copper (TCu) within a single bench, although the original simulation model was indeed deposit-wide, with over 55 million nodes simulated. Here, this single bench is taken as an example of the minimum volume of interest, and thus all the results discussed correspond to the uncertainty model for the entire bench. In practice, other volumes, such as bi-weekly, monthly or yearly planned production volume, geologic units, or other relevant units are used as the basis for risk analysis.

Figures 1 through 3 show the histograms and basic statistics for the composites dataset, the blast hole dataset, and the combined dataset, respectively. Note how the composites are smaller in number, and also a smaller average grade than the blast holes. This is a common phenomenon in porphyry copper deposits in northern Chile, and is generally related to inappropriate drilling techniques used for sampling, and the mineralogy of higher-grade samples. Also note that the results of any exercise involving blast holes will be dominated by their statistical characteristics. The combined dataset clearly reflects the general statistics of the blast holes dataset.

The correlogram models used were the same in all cases, obtained from the combined datasets, have a low nugget effect, and two spherical sub-horizontal structures with almost 600 m maximum range in the N-S direction and a 10-to-1 horizontal-to-vertical anisotropy. There are also flat-lying NW higher-grade cross structures that influence the first structure of the correlogram model. The kriging plans used applied no horizontal anisotropy, and a 7:1 and a 2:1 horizontal-to-vertical anisotropy for the blast holes and

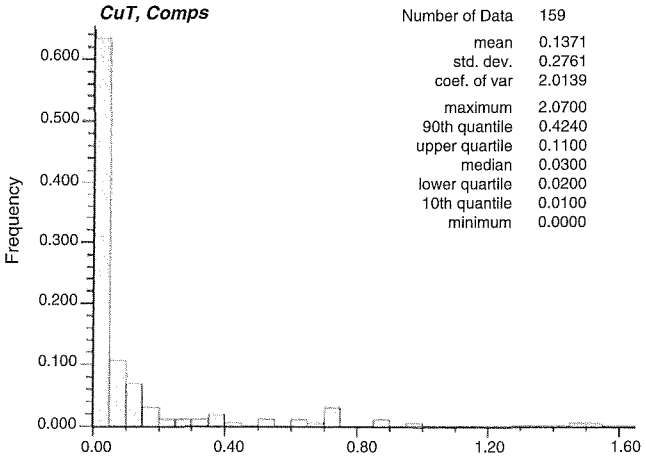

Figure 1. Histogram and statistics of the composites

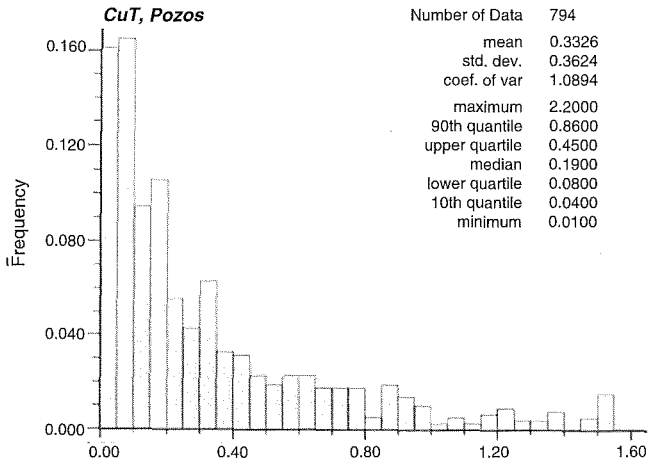

Figure 2. Histogram and statistics of the blast holes

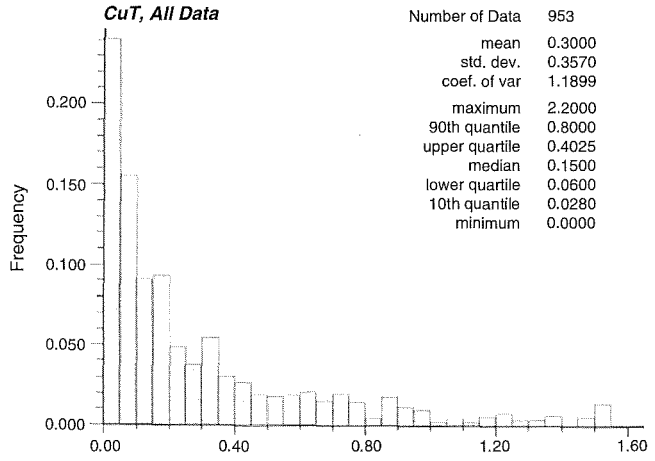

Figure 3. Histogram and statistics of the composites and blast holes combined

composites datasets, respectively. Also, the maximum search radius when using composites only was 120 m, while when using the combined dataset was 35 m. In all cases, a minimum of 1 sample and a maximum of 16 samples were used to obtain a simulated value. Octant search was used in all cases, allowing a maximum of 4 samples per octant.

Only a few statistics were used in this paper to represent and compare the different models of uncertainty obtained. The choice of these statistical parameters is somewhat subjective, but appropriate to illustrate the impact of the different RF models implemented. The main statistics presented here are:

80th Inter-percentile range:

$$IP_{80} = P_{90} - P_{10} \qquad [2]$$

Where IP_{80} is the 80th *IP* Range, P_{90} is the 90th Percentile of the conditional cumulative distribution function [1], and P_{10} is the 10th percentile of the same distribution.

The arithmetic average of the ccdf [1] for each simulated node and the complete bench:

$$m_j = \frac{1}{nsim(j)} \sum_{i=1}^{nsim(j)} s(i,j) \; (\forall j \in bench) \qquad [3]$$

Where *nsim* is the number of simulated values obtained for each node, j represents the node, and $s(i)$ represents each simulated value for the node. The means for each node are then again averaged over the whole bench.

The coefficient of variation of ccdf [1] for the complete bench:

$$CV = \frac{\sigma}{m} \qquad [4]$$

Where σ is the standard deviation of the ccdf [1], and m is defined in [3].

Summary of uncertainty models compared

Several models of uncertainty for Total Copper (TCu) were obtained for a bench (Bench 4330) of this large porphyry copper mine. These uncertainty models differ in some of the basic parameters used to obtain the set of conditionally simulated values. All other main variables, including variography, geologic domains used (stationary zones), etc. remained the same. As mentioned above, all parameter references relate to the GSLIB implementation of the Sequential Gaussian Algorithm, the only simulation algorithm used in this paper.

- *Uncertainty model used as 'Base Case Model'*—The Conditional Simulation Model used as 'Base Case Model' is based on the Sequential Gaussian Algorithm (as are all others considered in this paper), obtained on a 1 x 1 x 15 m grid of simulated nodes, and 10 simulated values for each. Also, in addition to the 15 m composites from the exploration drill hole data, the model used the more dense blast hole data (corresponding to the 15 m bench height used at the mine), ordinary kriging and multiple grid search to obtain the simulated values, and a single initial seed for the random number generator (reset for each simulation by the software). There is no particular significance to using this uncertainty model as the Base Case against which all others are compared.
- *Changes in the dataset used*—Alternative conditional simulation models were obtained using the exploration drill hole data set only. Blast holes are mostly drilled on a 9 x 11 m and a 10 x 10 m grid. The exploration drill hole data is drilled to approximately a 70 x 70 m grid. This is a fairly common in-fill grid used for 'Feasibility Level' resource definition in Cu porphyry-type deposits such as this, and provides one of the closest drill hole spacing usually available in this type of deposit. These two data sets 15 m composites from the exploration drill holes and exploration drill holes plus 15 m bench blast hole data) were used to obtain models for all the changes mentioned next.
- *Changes in simulated grid size*—Conditional Simulation Models were obtained for a 1 x 1 m grid, a 2 x 2 m grid, a 3 x 3 m grid, a 4 x 4 m grid, and a 5 x 5 m grid. The simulated grid was always based on the Bench Height (15 m), the vertical support of both the blast holes and drill hole composites. Different simulation models were obtained using the two datasets mentioned above.
- *Changes in kriging method*—Both ordinary kriging (the "Base Case") and simple kriging with a global mean were used in obtaining the conditional simulations. As before, both datasets were used as well, such that 4 different comparative models are obtained.
- *Changes in multiple grid search*—A multiple grid search algorithm, particular to the implementation in GSLIB, is usually a preferred option, as it results in faster processing speeds. The option analysed here was

the no use of a multiple search grid, to analyse its impact, if any, on the uncertainty model.

- *Changes in the seed of the random number generator*—Alternatively to using the standard option in GSLIB (one seed for the random number generator for multiple simulations, and then allowing the software to reset the seed for each new simulation), 10 simulations were obtained independently of each other by changing the input random number seed. Again, this was done using both datasets described above.
- *Changes in the number of simulations used to define the ccdf* [1]—The Base Case was repeated using 20, 30, 40, and 50 simulated values (instead of only 10) to obtain ccdf [1] for each node. Also, the same series of uncertainty models were obtained using the exploration composites only. Only the case with 50 simulations is discussed in this paper.

Results

Only the main results of this exercise are presented here. Figure 4 shows the Base Case described above, including the simulations obtained using both blast holes and exploration composites, and the simulations obtained using only exploration composites. The following main conclusions can be reached:

- The average simulated grade for the simulations obtained using blast holes and composites is higher than the average grade of simulations using composites alone. This is explained by the differences in TCu averages between the original datasets (Figures 1 through 3).
- The CV (second axis) for the simulations based on composites only is higher than the corresponding CV for the blast hole plus composites simulations. This also corresponds to a reasonable reproduction of the CV of the original datasets (Figures 1 through 3).
- The IP80 statistic is again smaller for the simulations obtained from blast holes plus composites, corresponding to what is observed in the original datasets. The IP80 is the more robust statistic of the three chosen here, and thus has little variation from one simulation to the next. For clarity, in the following Figures only the average grade of the simulations and its CV are shown.
- All three statistics for the simulations based on blast holes plus composites vary more from one simulation to the next, compared to the simulations obtained using exploration composites only. This result is not as counter-intuitive as it may appear initially, since more data does not always imply less variability in the ccdf [1] resulting from the simulations.

Figure 5 shows the comparisons of average grades and CVs of the uncertainty models for different grid sizes, for the simulations obtained using the combination of blast holes and exploration composites. Generally, the smaller the grid, the higher the average simulated grade, and the lower the simulated CV. This is particularly true for the 2 x 2 m grid, and not as evident when comparing the 4 x 4 m and the 5 x 5 m grids. The same is true for the simulations obtained from exploration composites only (not shown here).

Figure 6 shows the same comparison for the uncertainty models obtained with simple kriging (SK), compared to the base case (OK). Note that the SK variant gives lower average grades if using the exploration composites, but very similar averages are obtained when adding the blast holes. This is likely a result of the influence of the global mean of the data used, which is more reproduced less well if composites only are used. For the CV, the SK variant results in less variability for the combined dataset, while the opposite is true if using the exploration composites. Again, this is attributed to the characteristics of the SK method used.

Figure 7 shows the comparison for the uncertainty models obtained changing the random number generator seed for each simulation. Note that there is a significant difference in average simulated grades (from the blast holes and exploration composites combined) for the two variants in random number seeds assignment. This is unexpected, as it is generally accepted that the method of assigning random number seeds should be inconsequential to the resulting uncertainty model.

Figure 8 shows the comparison for the uncertainty models

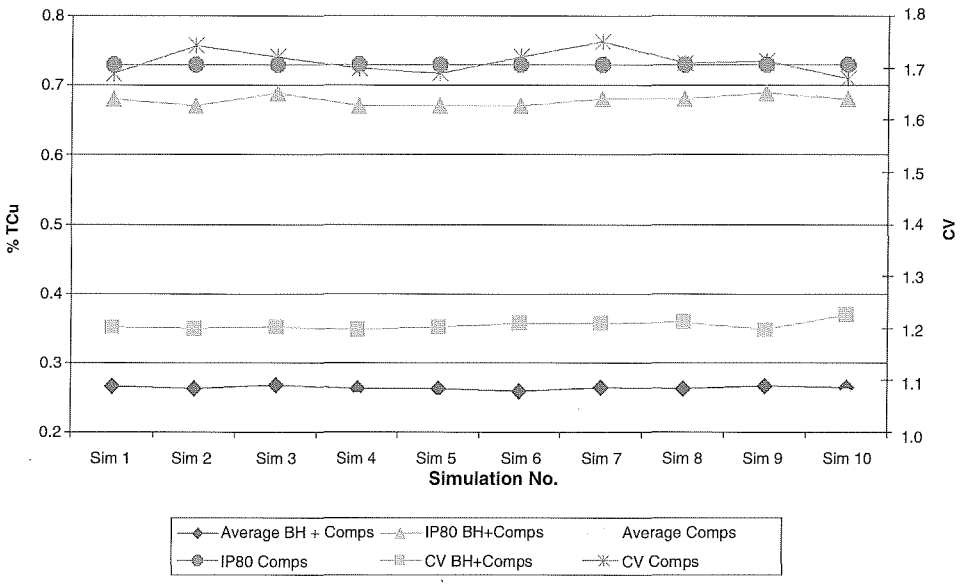

Figure 4. Simulations base case models, both datasets (1 x 1 x 15 m grid)

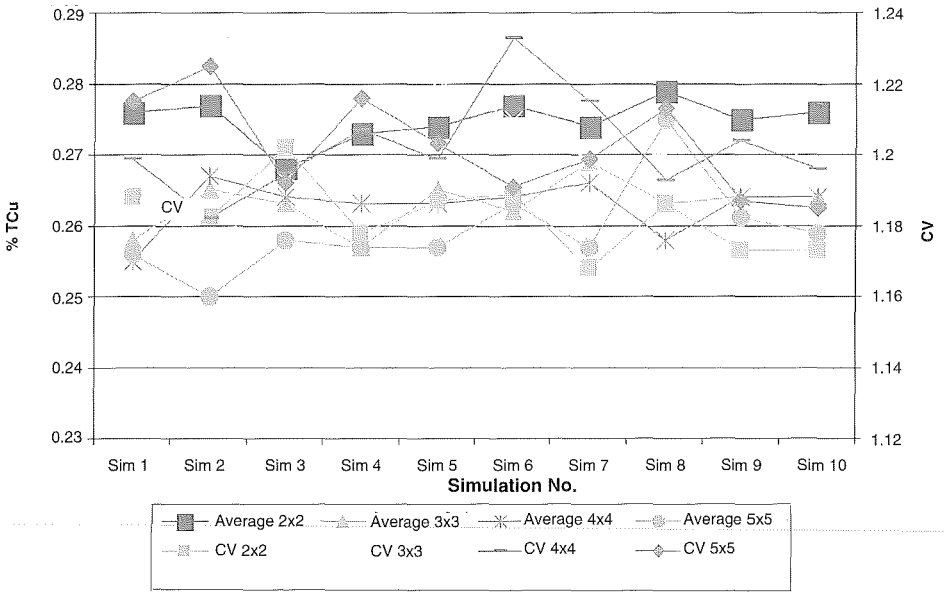

Figure 5. Simulations on different grid sizes, blast holes plus composites

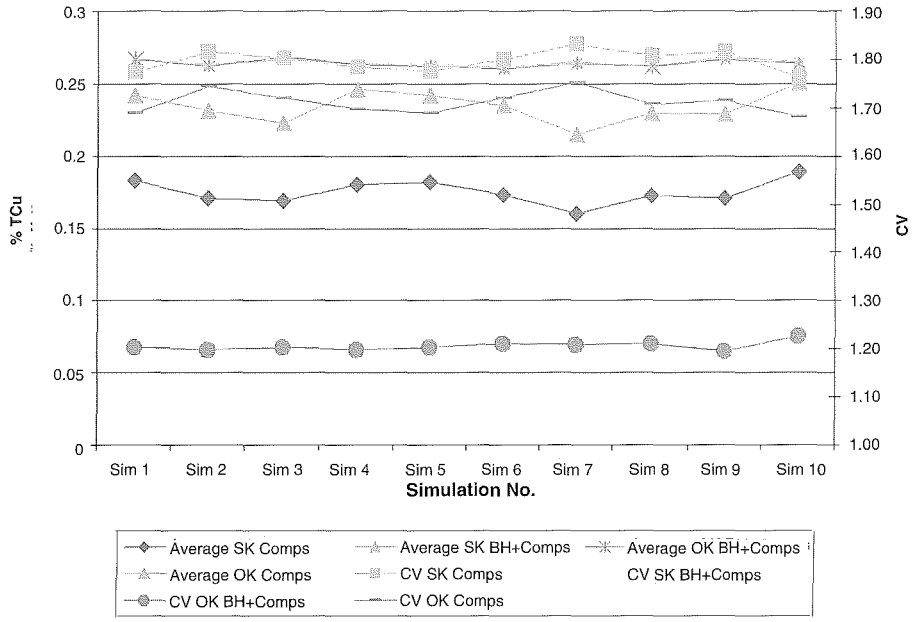

Figure 6. Simulations based on simple and ordinary kriging, both datasets

corresponding to the use or no use of multiple grid searches, another option characteristic to the GSLIB implementation used here. Note that the use of either option does not impact significantly the resulting model of uncertainty.

Figure 9 shows the comparison of the Base Case for both datasets (blast holes plus exploration composites, and exploration composites alone) when running 50 simulations instead of 10 to determine ccdf [1]. Note how the main conclusions derived from Figure 4 have not changed significantly. The simulated averages derived from exploration composites are smaller and vary more than the averages from the blast holes plus composites simulations. Note that Figure 4 shows the IP80 statistic, while Figure 9 does not. Also note that in both cases after simulation 10 the approximate minimum and maximum values of the ccdf [1] are found. Strictly, the minimum and maximum are found within 30 and 40 simulations for the exploration composites and blast hole plus composites-based simulations, respectively. This suggests that, for this TCu distribution at least, 20–30 simulations are more than sufficient to characterize its ccdf [1]. Also, after ten simulations the minimum and maximum are within 1–2% of the absolute minimum and maximum found in later simulations, which is a reasonable approximation for most risk analyses and other studies to be performed later.

Conclusions

The purpose of this work was to demonstrate how sensitive the models of uncertainty are to some of the implementation parameters chosen, in this case using the

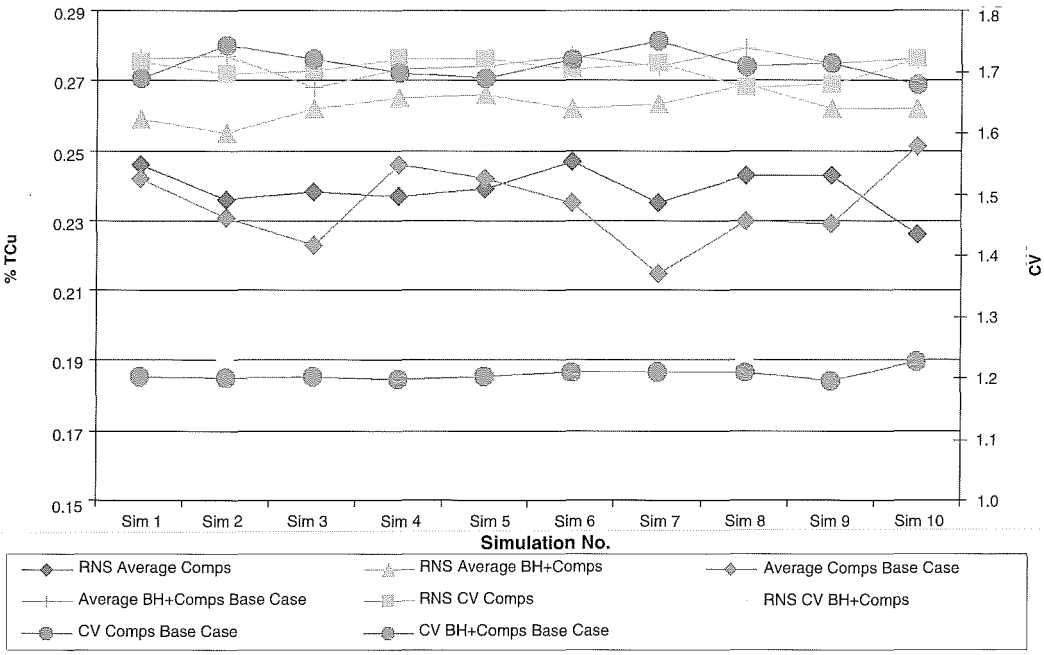

Figure 7. Simulations based on assigning random number seeds by two different methods, both datasets

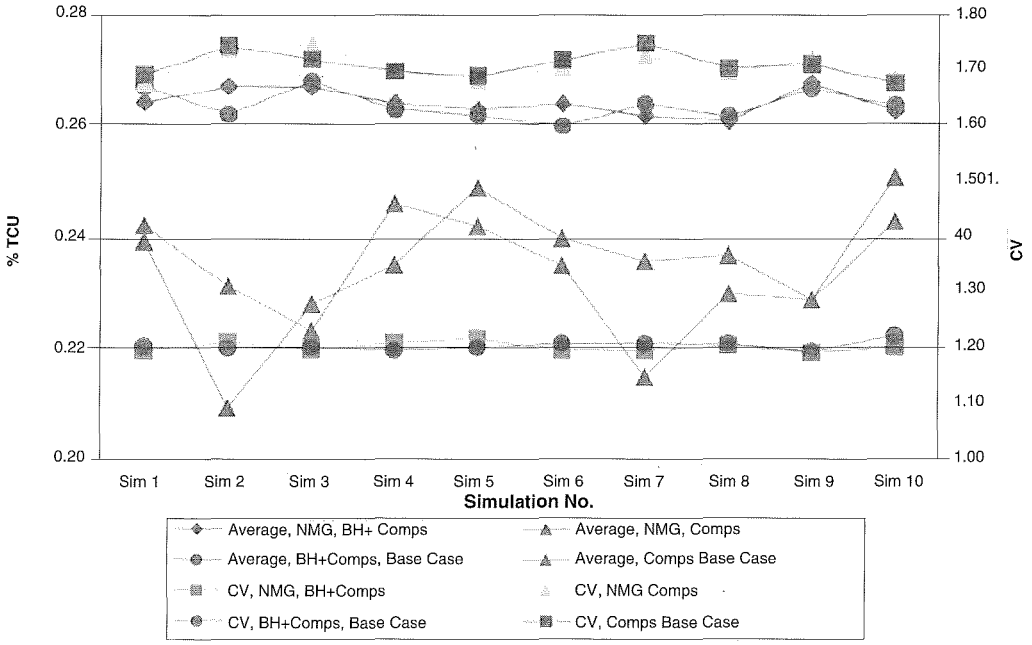

Figure 8. Simulation based on using simple searches, as opposed to using multiple grid searches (base case), both datasets

GSLIB software. It is emphasized that the parameters chosen here (number of simulations, grid size, datasets to be used, kriging method to be used, method to determine random number seed, and use or not of multiple grid search) are a few examples of the decisions that usually need to be made before implementing a large-scale simulation study, including among others the Random Function Model chosen, simulation domains, variogram models, etc.

The slight and few changes implemented, and how they impact the models of uncertainty obtained, lead to the following main conclusions:

- All the parameters analysed, with the exception of the multiple grid option, have a significant impact in obtaining a model of uncertainty for the TCu distribution described.
- More data input into the simulation model does not necessarily result in less variability. In this case, using composites (less and more sparse data, with a much larger coefficient of variation) result in simulations that have higher variability. But if the variability of the original blast hole dataset were higher than the composites dataset, then the simulations would try to reproduce it, as they do here, regardless of the amount of data available. This comment relates to the use of distinct datasets in the simulations.
- The sensitivity to implementation parameters shown here suggests that the objectives of the simulation study

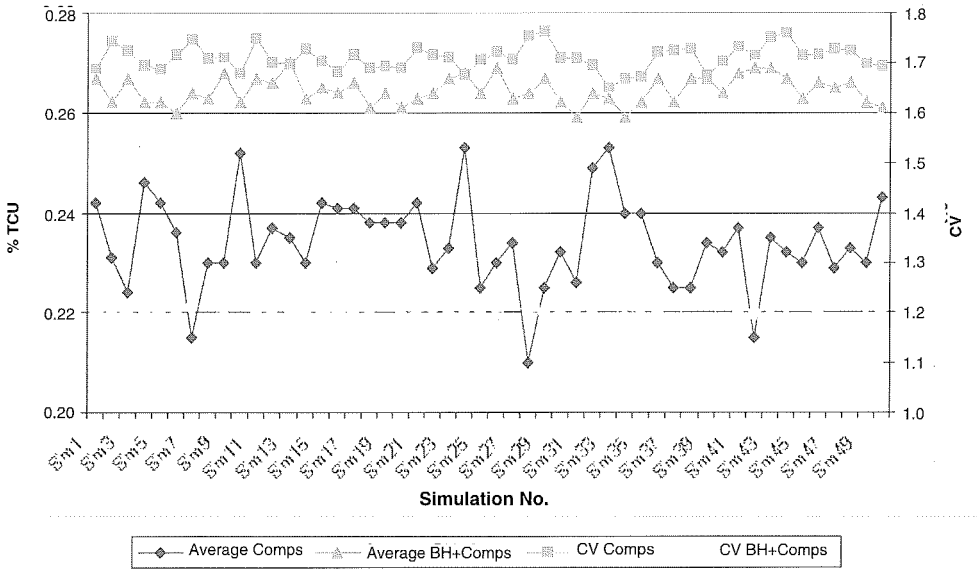

Figure 9. 50 Simulations, base case for both datasets

must be very well-defined and constrained in order for its results to be meaningful.

- The simulation models obtained can and will vary significantly depending on the choices made for each parameter implemented. The model of uncertainty is thus a subjective concept that results from choices made at the time of implementation, i.e., from the overall Random Function model chosen. Therefore, it is argued that no 'objective', or 'true' uncertainty model exists.

References

1. ROSSI, M.E. 'Optimising Grade Control: A Detailed Case Study', *Proceedings of the 101st Annual Meeting of the Canadian Institute of Mining, Metallurgy, and Petroleum* (CIM), Calgary, Canada, May, 2–5. 1999.
2. ROSSI, M.E. and CAMACHO V.J. 'Applications of Geostatistical Conditional Simulations to Assess Resource Classification Schemes', *Proceedings of the 102nd Annual Meeting of the Canadian Institute of Mining, Metallurgy, and Petroleum* (CIM), Quebec City, Canada, April 29-May 2, 2001.
3. VAN BRUNT, B.H. and ROSSI, M.E. 'Mine Planning Under Uncertainty Constraints', *Proc. of the Optimising With Whittle '99 Conference,* March 22–25, 1999, Perth, Australia.
4. GOOVAERTS, P. *Geostatistics for Natural Resources Evaluation*, Oxford University Press, 1997. 483 pp.
5. JOURNEL, A.G., and HUIJBREGTS, CH.J. *Mining Geostatistics*, Academic Press, 1978. 600 pp.
6. ROSSI, M.E. Uncertainty and Risk-Models for Decision-Making Processes, *Proceedings of the 28th International Symposium on Computer Applications in the Mining Industry (APCOM)*, Golden, Colorado, USA, October, 1999.
7. ISAAKS, E.H. The Application of Monte Carlo Methods to the Analysis of Spatially Correlated Data, PhD thesis, Stanford university, Stanford, CA, 1990. 213 pp.
8. ALABERT, F. Stochastic Imaging of Spatial Distributions Using Hard and Soft Information, MSc. thesis, Stanford University, Stanford, CA, 1987. 197 pp.
9. JOURNEL, A.G. Fundamentals of Geostatistics in Five Lessons, Stanford Centre for Reservoir Forecasting, Stanford University, Stanford, CA., 1988.
10. DEUTSCH, C.V. and JOURNEL, A.G. *GSLIB: Geostatistical Software Library and User's Guide,* Oxford University Press, New York, 340 pp. plus diskettes, 1998.