

Minimum Acceptance Criteria for Geostatistical Realizations

Oy Leuangthong,^{1,2} Jason A. McLennan,¹ and Clayton V. Deutsch¹

Received 21 March 2004; revised 17 May 2004

Geostatistical simulation is being used increasingly for numerical modeling of natural phenomena. The development of simulation as an alternative to kriging is the result of improved characterization of heterogeneity and a model of joint uncertainty. The popularity of simulation has increased in both mining and petroleum industries. Simulation is widely available in commercial software. Many of these software packages, however, do not necessarily provide the tools for careful checking of the geostatistical realizations prior to their use in decision-making. Moreover, practitioners may not understand all that should be checked. There are some basic checks that should be performed on all geostatistical models. This paper identifies (1) the minimum criteria that should be met by all geostatistical simulation models, and (2) the checks required to verify that these minimum criteria are satisfied. All realizations should honor the input information including the geological interpretation, the data values at their locations, the data distribution, and the correlation structure, within “acceptable” statistical fluctuations. Moreover, the uncertainty measured by the differences between simulated realizations should be a reasonable measure of uncertainty. A number of different applications are shown to illustrate the various checks. These checks should *be an integral part of any simulation modeling work flow*.

KEY WORDS: Model validation, verification, simulation.

INTRODUCTION

Geostatistics has grown to a comprehensive methodology and suite of tools for both estimation and simulation. Conventional kriging algorithms are looked upon favorably for their ability to estimate accounting for the spatial variability of the data. The only problem is the overly smooth distribution of estimates that is not representative of the true variability. Recently, simulation is gaining in popularity to estimation techniques because of its ability to improve heterogeneity characterization and assessment of joint uncertainty.

Geostatistical simulation is built on the foundations of kriging. In its development, simulation retains

most of the positive attributes of kriging, that is, exact data reproduction and use of the spatial correlations between data. The smoothing effect in estimation is corrected by accounting for the variability between the simulated locations. Multiple realizations of the deposit allow for the assessment of joint uncertainty.

In 1984, Parker claimed that “production of the reserve model is now on the critical path of the project . . . [it] has forced geostatistics to be practiced in a ‘production mode.’ That many successful geostatistical studies of the past were the result of careful methodical research and checking is often forgotten.” (Parker, 1984, p. 930) Almost 20 years later, this sentiment rings true yet in modern practice. The only difference is that in today’s technologically advanced environment, this “production mode” is at a bigger scale.

The advancement of technology and the availability of simulation algorithms in many commercial software has popularized the use of geostatistics in both the mining and the petroleum industry. The

¹Department of Civil & Environmental Engineering, 220 Civil/Electrical Engineering Building, University of Alberta, Edmonton, Alberta, Canada T6G 2G7.

²To whom correspondence should be addressed; e-mail: oy@ualberta.ca.

algorithms and engines are faster than ever before, and they are designed to facilitate easy setup of a geostatistical study. The ease of application may turn geostatistical modeling into a “black box” practice, allowing for more inferior models (Srivastava, 1996, p. 56). Furthermore, the tools required for careful checking of these models are rarely provided as part of the software; the only recourse is to perform model checking tasks outside of the commercial software, but this may be impractical because of the large size of the files and the data manipulation required.

This paper reinforces the need for careful checks to ensure success in a modern geostatistical application. Specifically, the minimum criteria for confirming consistency in a model are addressed, and the required checks that must be performed are outlined. An application to a gold deposit illustrates the model results and shows the basic tools to verify consistency of the model with input information.

Geostatistical simulation produces a model of uncertainty that is represented by multiple sets of possible values distributed in space; one set of possible outcomes is referred to as a *realization*. Some checks described in this paper are performed on a single realization, whereas others are performed on the set of multiple realizations. The term *suite of realizations* or *ensemble* are used to refer to this latter set of realizations.

MINIMUM CRITERIA

The term model validation usually refers to some measure of the *goodness* of a model. This goodness may be measured by reconciling model results to some additional hard data. In mining, this hard data can be blast hole data; whereas in petroleum, reconciliation may be more complex because only dynamic production data are available. Whatever the additional information, model validation has a specific reference to the predictive ability of the model, suggesting that the model is representative of the physical reality (Oreskes, Shrader-Frechette, and Belitz, 1994, p. 642).

Confirming that a model reproduces the input is the scope of what we infer by minimum criteria. It does not necessarily imply that the model is geologically realistic or good for production forecasting. The key deposit-specific inputs to a geostatistical simulation consist of the actual data and the variogram measure of spatial correlation. At the very least, the information contained in these two inputs must be re-

produced by a numerical model. A simulated model *must* reproduce:

- (1) data values at their location,
- (2) distribution of the attribute of interest, and
- (3) the spatial continuity characterized by the variogram model.

The latter two statistics are reproduced in *expected* value, thus statistical fluctuations about these values are anticipated. In the situation of simulation models of multiple variables, the multivariate distributions and corresponding summary statistics also should be honored.

Specific implementation details may cause problems. In such situations, a careful examination should be conducted to confirm the cause of violation, and whether this is acceptable. Some of these checks can be performed on individual realizations while others require consideration of the full suite of multiple realizations.

The credibility of any model is not only dependent on satisfying the above checks, but also on ensuring that the model parameters are appropriate. Reliable inference of model parameters is critical to model credibility and deserves some attention; however, that is not the subject of this note.

The practitioner is faced with many decisions in the process of model construction, use of declustering tools, variogram modeling, size of the model to generate, number of realizations, type of kriging to apply, and a general multitude of implementation details that differ with the software. Careful documentation and justification of these decisions are important for repeatability of the models. Although this documentation may help to improve the construction of future models, it does not act as an error-checking tool for the current model under construction. For this, several validation tools exist that should be integrated into the modeling work flow.

As simple as this sounds, the first check should be a visualization of the realizations (in 3D if possible). This visualization should highlight low and high valued areas. The project geologist should be satisfied with the variability of the high and low values and their overall distribution. The variability or uncertainty should be reasonable and plausible, for example, there should be no high values in clearly low areas and vice versa. Comparisons against simple geologic contours of trends, generated by methods such as hand contouring, inverse distance and other estimation techniques, also would provide a level of comfort and confidence in the simulation models. The geologist

should be neither intimidated by the geostatistical procedures nor swayed into accepting any strange results.

Once the realizations are deemed geologically plausible, validation tools, such as cross validation and the jackknife, could be used. The basic idea is to estimate an attribute at a location where the true value is known. In cross validation, a data value is removed and the location is estimated using all other neighboring data. Conversely, the jackknife refers to resampling without replacement. As a result, cross validation is known as the “leave one out” approach, and the jackknife approach is known as the “keep some back” approach (Deutsch, 2002). These methods provide an indication of the goodness of modeling parameters. Cross validation should yield the following results for a model with “good” parameters:

- Cross plot of the estimate vs. the true value should show a high correlation coefficient.
- Distribution of errors should be symmetric, with a mean of zero and a low variance.
- Cross plot of the error vs. the estimate should be centered about zero error, satisfying a property termed “conditional unbiasedness” (Isaaks and Srivastava, 1989; Krige, 1999; Krige and Assibey-Bonsu, 2000).

Although these techniques may be used to fine tune the variogram model, cross validation results usually are insensitive to minor changes in the variogram model. The main use of cross validation is to identify mistakes or problem data. It does not prove which simulation or estimation technique is optimal.

Data Reproduction

From the exactitude property of kriging, the estimate at a data location is exactly the data value and the error variance is zero (Goovaerts, 1997; Journel and Huijbregts, 1978), that is,

$$z_K^*(\mathbf{u}) = z(\mathbf{u}_\alpha), \quad \forall \mathbf{u} = \mathbf{u}_\alpha, \quad \alpha = 1, \dots, n$$

where $z_K^*(\mathbf{u})$ is the kriged estimate of the random variable, Z , at location \mathbf{u} in the domain, and $z(\mathbf{u}_\alpha)$ is the data values at location \mathbf{u}_α , $\alpha = 1, \dots, n$. Because simulation relies on the kriged estimate and the kriging variance to define the conditional cumulative distribution function (ccdf), simulation also will reproduce the data exactly at their locations.

To verify that all data that *should* be reproduced are reproduced, a crossplot of the data and the simulated values at the data locations should be generated. First, all data should be accounted for via a detailed inventory for data reproduction. In some instances, there is a valid reason why some data are not reproduced. For instance, some implementations of simulation allow for data assignment to grid nodes as a way to speed up distance calculations. As a result, the total number of assigned data may be less than the total number of data available for conditioning. There are a number of reasons for this (Fig. 1):

- The sample coordinates may lie outside of the 3D grid as specified by the model limits.
- The data lie inside the 3D grid, but their value is trimmed. Although these samples are located inside the 3D grid to be simulated, their data

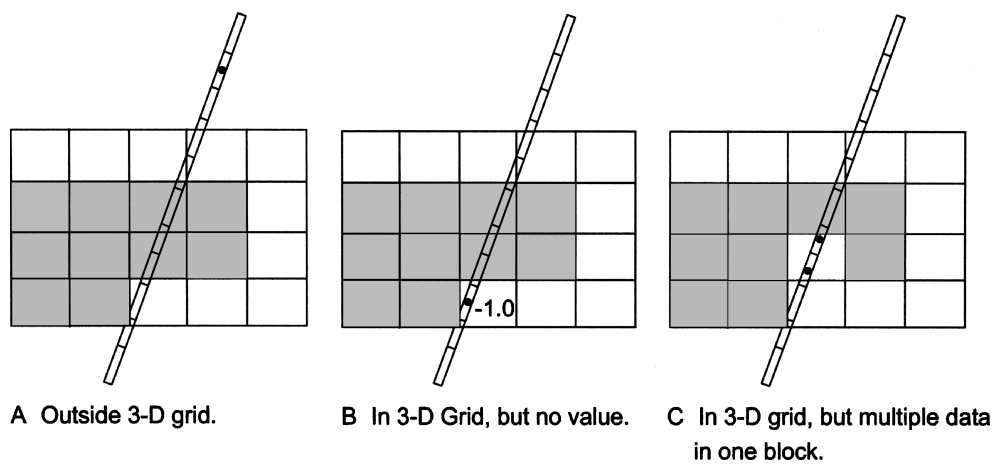


Figure 1. Schematic illustrations of possible reasons why composite is not assigned to grid block: A, sample is available but lies outside of simulation grid; B, sample lies within simulation grid, but is unformed; and C, two samples lie within same grid block, but only closest sample is assigned.

values indicate that either no sample is available or it is specified as missing or an outlier. Hence it would not be assigned to a grid block.

- The data is inside the 3D grid but there are multiple data close to one grid block; another sample is closer to the same block centre and is assigned to the grid block. The first sample then is not considered in the simulation.

The exact breakdown of samples that are not assigned as a result of the given reasons should be tabulated to ensure that all nonreproduced data are accounted for. In this type of implementation, all data are used to define the global distribution but only the assigned data are used to condition the simulation.

Once the exact number of data that *should* be reproduced is determined, a crossplot of this subset data against the model values at the same locations should show a 1:1 correspondence between these two sets of values (see Fig. 2). Slight deviations may be the result of numerical precision of the transformation look-up table for back transformation of simulated values to original units, which is the situation for the low values in Figure 2. Significant deviations from the 45° line should be investigated further to determine the cause.

Of course, there also is the option to *not* assign data to grid nodes. In this example, unless a sample has the same coordinates as a grid block center, it will not be reproduced exactly. In this implementation, the corresponding crossplot to Figure 2 should show

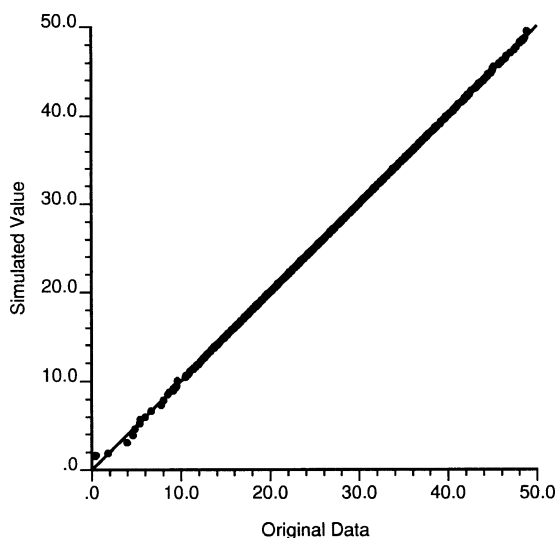


Figure 2. Crossplot of original data and simulated values at data locations.

slight deviations from the 45° line with a high correlation near 1.0. The potential for large deviations from this 45° line depends on the proportion of the nugget effect.

Histogram Reproduction

The second check that should be performed is to verify that the histogram is reproduced. This requires that the target histogram is clearly identified. If the representative histogram is different from the equally weighted data histogram, then this target distribution should be clearly specified in the simulation input parameters.

Moreover, the practitioner also should be clear in the declustering approach employed to establish the representativity of the distribution. In this matter, appropriate considerations regarding the different techniques should be addressed. For instance, if a geology model is available that clearly defines the geological boundaries, then polygonal declustering may be preferable to cell declustering.

In the multivariate context, if the simulation approach will consider secondary data, then the target distributions also may require accounting for both the joint and marginal distributions of the two variables in the weighting scheme. For instance, use of the stepwise conditional transformation requires transforming one variable conditional to another (Leuangthong and Deutsch, 2003); hence, the target distribution of the conditionally transformed variable will require use of the bivariate distribution to calibrate the marginal distribution to obtain the target histogram.

To verify global reproduction of the histogram, the histogram of the model should be examined. It may be impractical to examine the histograms from all realizations; however, a few randomly selected realizations should be checked. In this type of visual checking, the key features to note include reproduction of (1) the histogram shape, (2) the range of the simulated values, and (3) the summary statistics, such as the mean, median, and the variance (see Fig. 3).

Alternatively, a quantile–quantile (Q–Q) plot may provide a better indication of histogram reproduction, as binning may hide some features in the histogram (see Fig. 4). This type of check permits multiple realizations to be visualized at once. This amounts to plotting multiple distributions onto the same Q–Q plot and assessing whether the suite of distributions honors the input histogram with some statistical fluctuations.

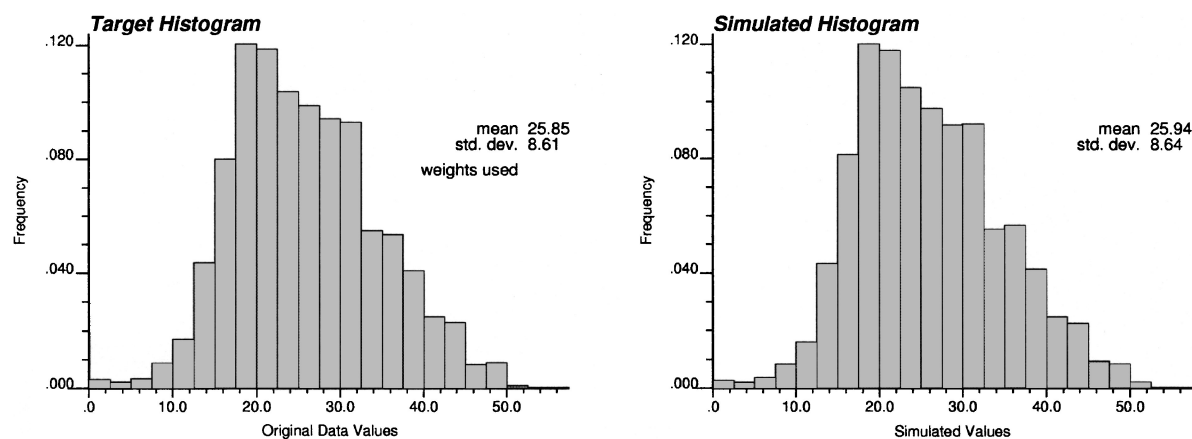


Figure 3. Histogram reproduction for attribute of interest: representative histogram (left) compared to simulated histogram from one realization (right).

Reproduction of Summary Statistics

The previous check was an examination of the reproduction of the histogram shapes and some statistics on an individual realization basis. This particular check differs in that the reproduction of the summary statistics will be examined over the entire suite of multiple realizations (which may be referred to as the ensemble). For each realization, there is a global mean and a global variance. For the ensemble, the summary statistics can be checked. An easy way to check this reproduction is to plot the histograms of the mean and variance from all the realizations. We would expect

that for many realizations, the mean of the means should reproduce the target mean, and the mean of the variance distribution should reproduce the target variance. A box plot on both distributions with a reference value will show the position of the target statistic relative to the distribution of simulated statistics. Figure 5 shows an example of this type of check performed over 40 realizations. The dot plotted on the box plot shows the target statistic relative to the distribution of the simulated statistics; this can be compared with the reported mean value of the summary statistic. Departures from these expectations may be cause for concern. Although this check (for the ensemble) should be performed, usual practice is to check only a few realizations and any departures from the target histogram and variogram are typically attributed to ergodic fluctuations (Srivastava, 1996). There has been little work to establish *acceptable* ergodic fluctuations. Deutsch and Journel (1998) discussed the effect of domain size and the variogram model on ergodic fluctuations; Goovaerts explored the magnitude of ergodic fluctuations and the space of uncertainty from four different simulation algorithms (Goovaerts, 1999); Srivastava touched on the ability of simulation to fairly sample from the space of uncertainty (Srivastava, 1996, p. 60); Chilés and Delfiner mentioned the use of the integral range as a measure of practical ergodicity (Chilés and Delfiner, 1999). Another way to “predict” the amount of ergodic fluctuations one should expect is to calculate the dispersion variance of the domain relative to the assumption of an infinite domain, $D^2(A, \infty)$; this can be calculated numerically. If ergodic fluctuations

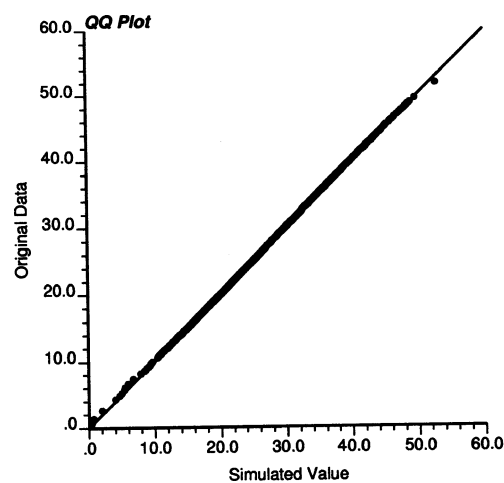


Figure 4. Q-Q plot of the original data distribution and simulated values distribution to check histogram reproduction.

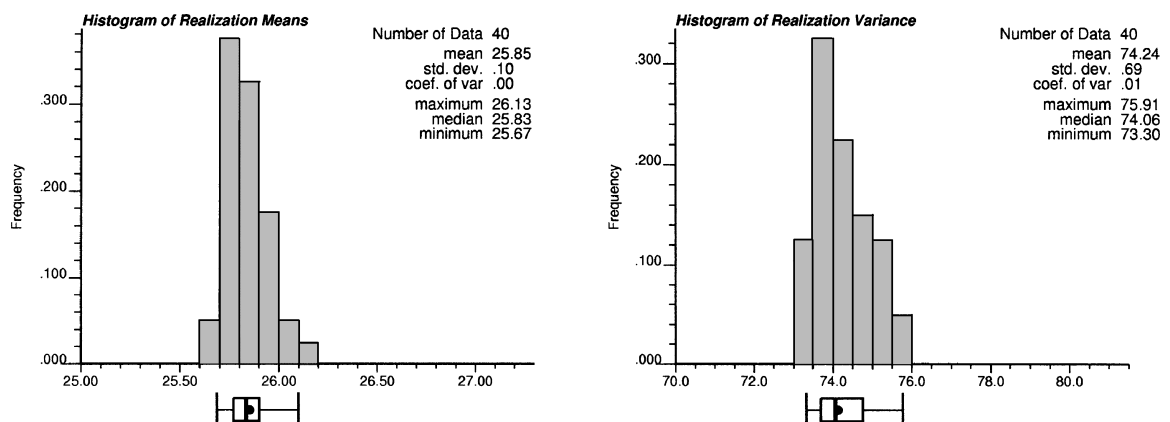


Figure 5. Reproduction of summary statistics for model variable: histogram of means (left) and variances (right) from multiple simulated realizations. Box plots on x-axis shows 95% probability interval (outside lines), 50% probability interval (box) and median (vertical bold line inside box). Dot indicates mean value of summary statistic from target (declustered) distributions.

exceed $D^2(A, \infty)$, then model parameters should be checked. More work is required to explore this issue.

Variogram Reproduction

Now that the data and the histograms have been checked for reproduction, we can proceed by checking reproduction of the second order statistics, specifically the variogram. For Gaussian simulation, it is important to note that this check must be performed in normal or transformed space (prior to back transformation), since only the normal scores variogram is imposed directly.

The variogram should be calculated for multiple realizations, and compared to the input variogram model in the same direction (Fig. 6). The model variogram should be reproduced within *acceptable* ergodic fluctuations (see previous discussion).

OTHER CHECKS FOR CREDIBILITY

Unlike cross validation where the true value is known, the probabilistic models developed using geostatistical tools are built with some degree of uncertainty.

One basic check is to verify that the probability intervals of the local distributions are consistent with the underlying model of uncertainty. For a specific probability interval (PI), p , we should expect to learn that for multiple realizations, the proportion of times the true value falls within the PI is approximately equal to p for all p in $[0,1]$ (Deutsch, 1996). For instance, a symmetric PI of 80% ($p = 0.80$) indicates

that the lower and upper probability values in the interval is 0.10 and 0.90, respectively. Ideally, the proportion of times the true value falls within the 80% PI should be close to 0.80. If this fraction is greater than 0.80, then the probability interval is too wide, and the local uncertainty may be too high. Conversely, if the fraction is smaller than 0.80, then the probability interval is too narrow and the distribution has too low a variance.

This can be checked for a series of PI from $[0,1]$, and the probability intervals plotted against the fraction of true values that fall within these intervals. Figure 7 shows this crossplot for a simulation model that was reconciled against blast hole samples. For the 70% probability interval, the fraction of the true values

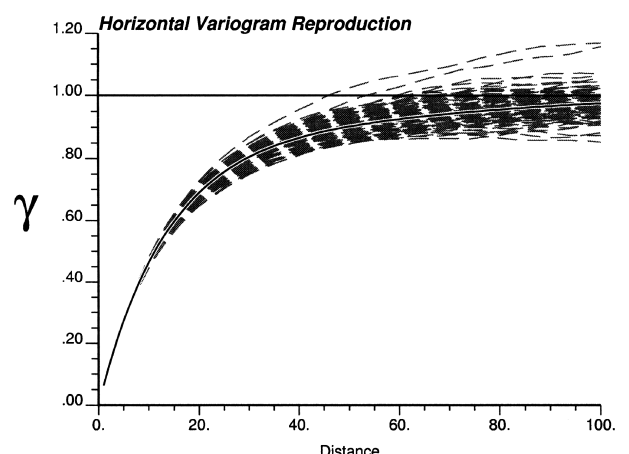


Figure 6. Reproduction of horizontal variogram: input variogram model (outlined by a white line), and resulting variograms in multiple realizations are shown in dashed line.

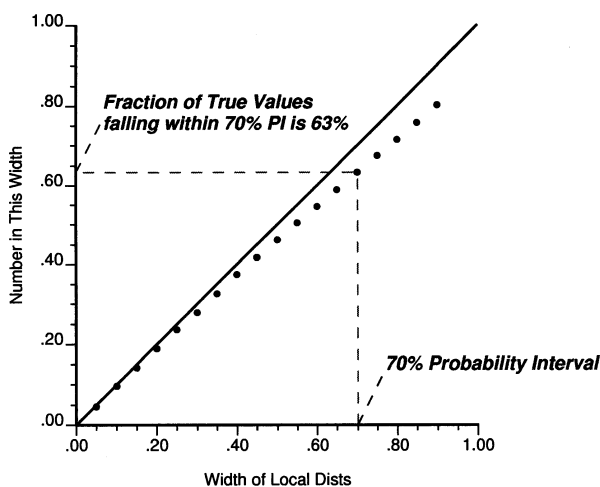


Figure 7. Crossplot of fraction of true values falling within a specific probability interval (PI) and probability interval. Note that these paired data fall close to 45° line.

falling within this interval is 63%. For this and all other intervals shown in the plot, the match between these pairs of numbers is sufficiently close to the ideal situation of falling on the 45° line.

APPLICATION

The grades in this example are from a 2D gold deposit. There are 52 available gold drillhole sample data. Figure 8 shows the drillhole sample locations and the histogram of gold grades. Samples are clustered in the high grade area close to the ground surface. The gold grades (g/t) are skewed positively with a mean and variance of 1.563 and 1.641, respectively. Preferential sampling in the high-grade area requires declustering to obtain the target (or reference) distribution. This target distribution also is shown in Figure 8, with a noticeably reduced mean and variance

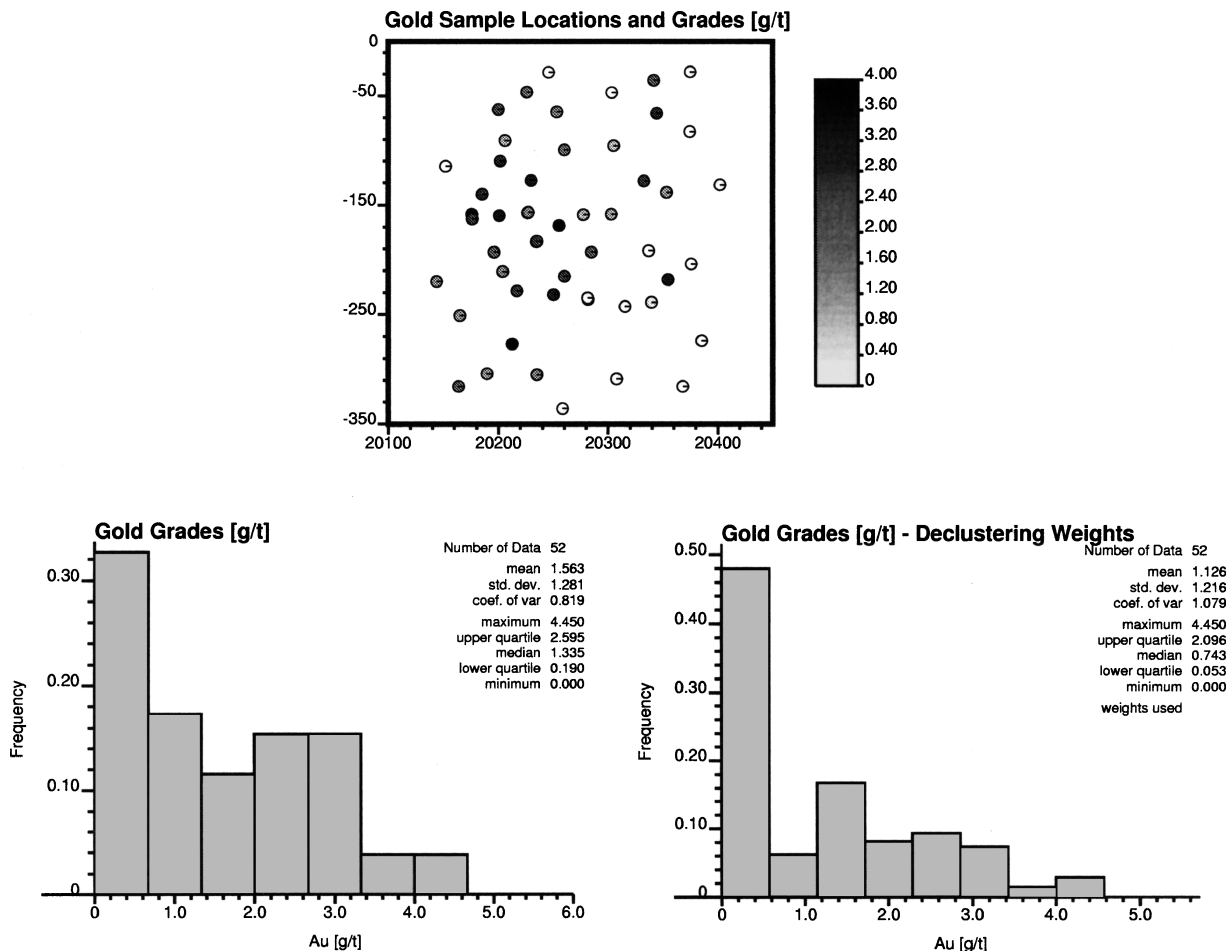


Figure 8. Location map of drillhole samples (top). Histogram of gold grades (bottom left) and declustered histogram of grades (bottom right).

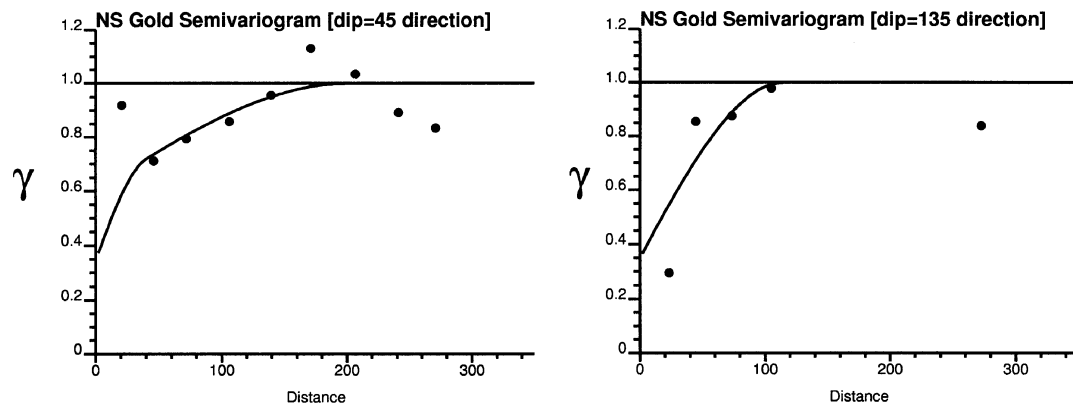


Figure 9. Normal scores variogram for gold in dip direction of 45° (left) and 135° (right).

of 1.126 and 1.479, respectively. This is the distribution that must be reproduced in expected value by geostatistical simulation.

Variography is performed using the normal score values of the drillhole data. Concluding a directional

variogram investigation, the best correlation is shown in the dip direction of 45° and is selected as the principal or major direction; the direction of minor continuity is the perpendicular direction of 135° dip. Figure 9 shows the final experimental variogram points scaled

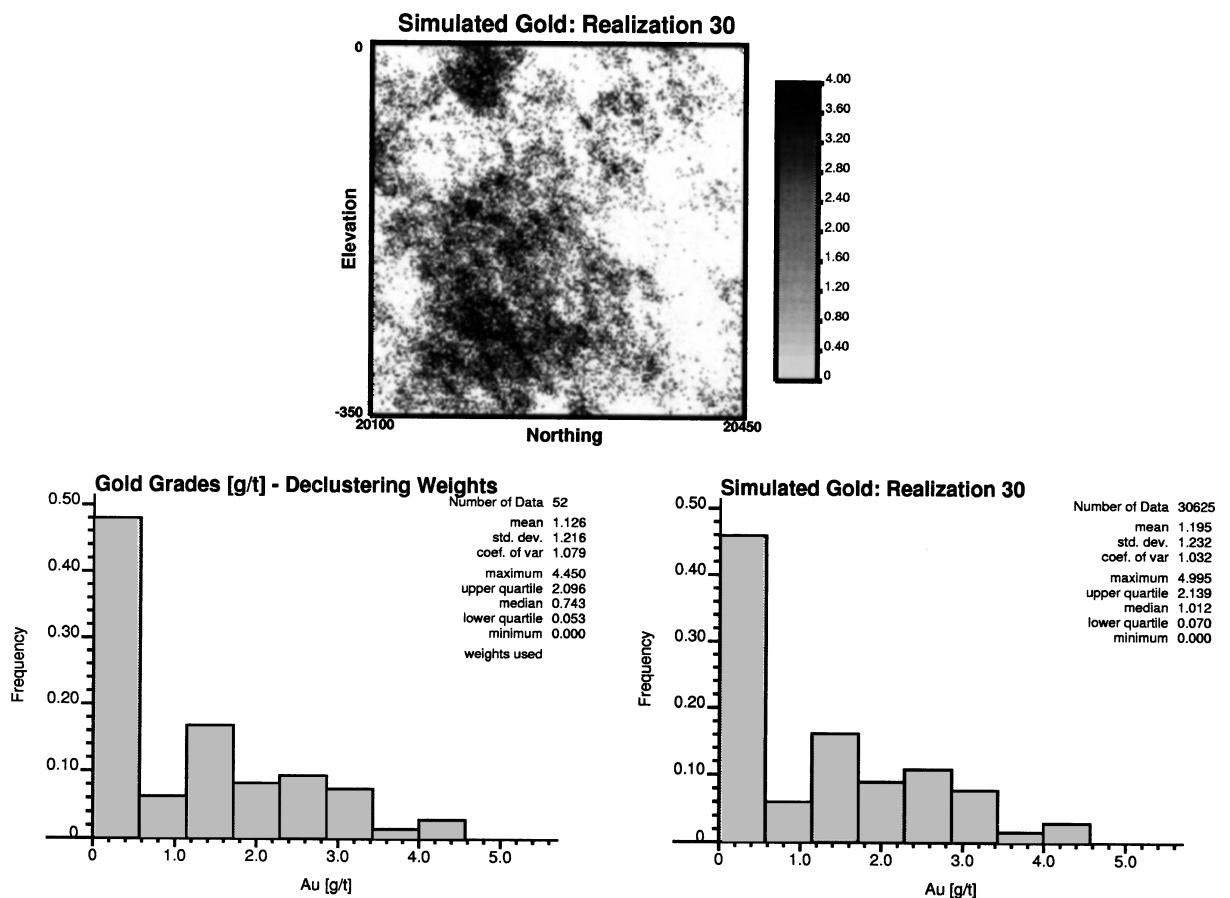


Figure 10. Map of simulated gold grades for realization 30 (top), target declustered gold distribution (left) and distribution of simulated gold grades from realization 30 (right).

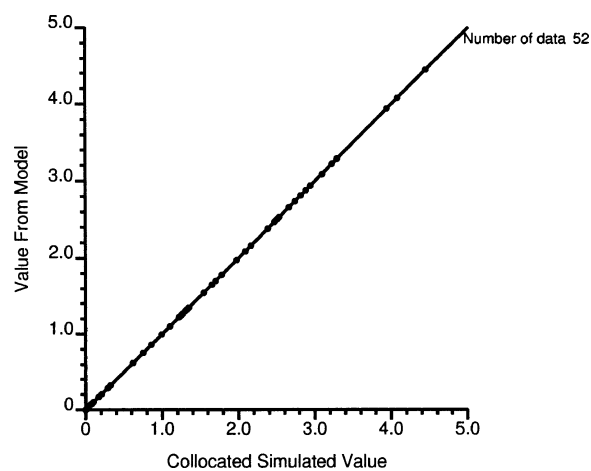


Figure 11. Crossplot of true data value and collocated simulated value to check for data reproduction.

according to the number of pairs, which is shown above each point, used in the calculation and the variogram model lines for both directions. The final variogram model fitted to both directions simultaneously is:

$$\gamma(\mathbf{h}) = 0.35 + 0.25 \cdot \text{Sph}(\mathbf{h}) + 0.40 \cdot \text{Sph}(\mathbf{h})$$

$a_{\max} = 40$

$a_{\min} = 115$

$a_{\max} = 200$

$a_{\min} = 115$

where a_{\max} is the direction of maximum continuity with a dip of 45°, and a_{\min} is the direction of minimum continuity with a dip of 135°.

Sequential Gaussian simulation is used to create 100 realizations of gold grades at a square 2 m support. The Northing–Elevation (Y–Z) view and

histogram of the gold grades simulated for the 30th realization are shown in Figure 10. For easy comparison, the declustered histogram also is shown. Notice the agreement between the realization’s distribution of simulated gold grades and the input declustered histogram.

The check for data reproduction after simulation is a cross plot of the original gold samples against the collocated simulated values. This crossplot is shown for the 30th realization in Figure 11. All of the pairs fall on the 45° line. The input data are honored at the original sample locations.

The reproduction of the summary statistics is checked by considering all 100 realizations of the gold distribution. The mean and variance for each realization are summarized as a histogram of the means and a histogram of the variances. These histograms are shown in Figure 12. The mean of the means is 1.107 and the mean of the variances is 1.449 compared to the input declustered distribution’s mean and variance of 1.126 and 1.479, respectively. The input distribution is satisfactorily reproduced.

To check that the variogram model is reproduced satisfactorily, the variogram is calculated in both the major and minor direction for each of the 100 realizations. All 100 calculated variograms are shown as dashed lines, and the input model variogram is shown as a solid line in Figure 13. The variogram model is satisfactorily reproduced in both directions.

DISCUSSION

Although this paper is focused mainly on the *minimum* acceptance criteria, there are a number of

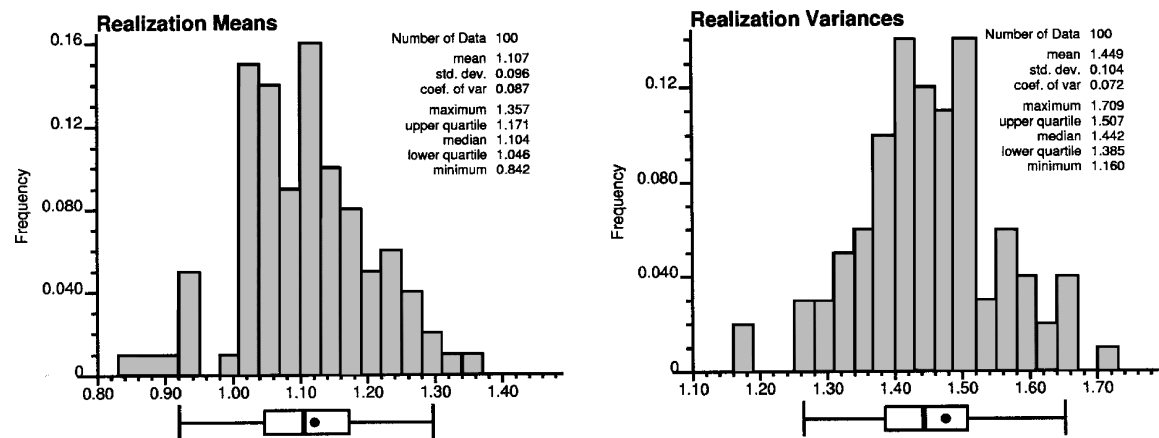


Figure 12. Histograms of mean (left) and variance (right) for 100 realizations to check for reproduction of summary statistics.

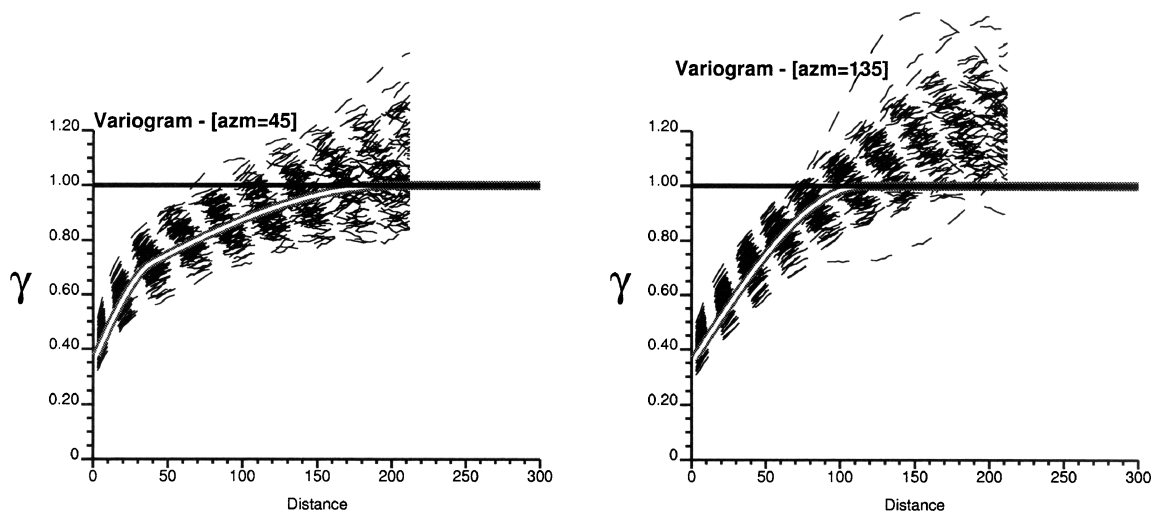


Figure 13. Variogram in direction of major (left) and minor (right) continuity directions. Input model is shown as dark solid line, whereas variograms corresponding to 100 realizations are shown as lighter dashed lines.

considerations that should be addressed prior to using these models for decision making.

The four basic criteria identified as the minimum criteria in this paper are not new. Geostatisticians have known about these checks, but the general practice of model checking is lacking.

Any model of spatial uncertainty should include a prior assessment of data uncertainty and possible errors. This involves a clear and complete documentation of the errors and uncertainty inherent in the data. In light of this assessment, the practitioner then can gauge when the data or statistics are honored *too* well. Essentially, this practice is aimed at documenting data that must be honored and recognizing the level of reproduction that would be acceptable.

Once the models are constructed, one idea is to methodically check *each* realization and “sign off” on it, after verifying that the realization honours all the input information and conforms to the geological interpretation. The set of signed-off realizations then can be passed through to the next stage of decision making.

Another consideration is to reconcile the model results with existing “output” or production data. For instance, in a petroleum context, the flow response of the reservoir is of primary importance. A set of realizations could be processed through a flow simulator and the model response could be checked against real production data. Similarly in a mining context, the simulated values could be reconciled against blasthole data or actual mill production to assess the predictive ability of the model. The historical performance of

simulation models for other similar sites can also be used as a measure for comparison.

CONCLUSIONS

Today, much emphasis is placed on the construction of numerical models for risk qualified decision-making. Many different stochastic simulation algorithms exist to build these models, and the application of conventional approaches is made popular through commercial software. Faster, more efficient computers facilitates this modeling process. Model checking is as important to the decision making process as the actual model construction, but unfortunately, this area may be overlooked in practice.

For geostatistical realizations, there are essentially four minimum criteria that *must* be satisfied. These are reproduction of (1) data values at data locations, (2) the target histogram, (3) the target summary statistics, and (4) the input covariance model. In the multivariate context, this list also should include reproduction of the multivariate distribution and the corresponding summary statistics. An application to a 2D gold deposit shows the relevant checks required to ensure the realizations are consistent with the simulation approach.

ACKNOWLEDGMENTS

We would like to acknowledge the industry sponsors of the Centre for Computational Geostatistics at

the University of Alberta for supporting this research. We are also thankful to R. Mohan Srivastava for sharing his many ideas and suggestions on this topic.

REFERENCES

Chilés, J., and Delfiner, P., 1999, Geostatistics: modeling spatial uncertainty: John Wiley & Sons Inc., New York, 695 p.

Deutsch, C. V., 1996, Direct assessment of local accuracy and precision, *in* Baafi, E. Y., and Schofield, N. A., eds., Geostatistics Wollongong '96, vol. 1, Kluwer Acad. Publ., Dordrecht, Holland, p. 115–125.

Deutsch, C. V., 2002, Geostatistical reservoir modeling: Oxford Univ. Press, New York, 376 p.

Deutsch, C. V., and Journel, A. G., 1998, GSLIB: Geostatistical Software Library and users guide (2nd edn.): Oxford Univ. Press, New York, 369 p.

Goovaerts, P., 1997, Geostatistics for natural resources evaluation: Oxford Univ. Press, New York, 483 p.

Goovaerts, P., 1999, Impact of simulation algorithm, magnitude of ergodic fluctuations and number of realizations on the spaces of uncertainty of flow properties: Stochastic Environmental Research and Risk Assessment, v. 13, p. 161–182.

Isaaks, E., and Srivastava, R. M., 1989, An introduction to applied geostatistics: Oxford Univ. Press, New York, 561 p.

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

Krige, D., 1999, Conditional bias and uncertainty of estimation in geostatistics: 1999 APCOM Intern. Symp., Soc. of Mining Engineers, Golden, Colorado, p. 3–14.

Krige, D., and Assibey-Bonsu, W., 2000, Limitations in accepting repeated simulations to measure the uncertainties in recoverable resource estimates particularly for sections of an ore deposit: GEOSTATS 2000, Cape Town, South Africa, p. 830–836.

Leuangthong, O., and Deutsch, C. V., 2003, Stepwise conditional transformation for simulation of multiple variables: Math. Geology, v. 35, no. 2, p. 155–173.

Oreskes, N., Shrader-Frechette, K., and Belitz, K., 1994, Verification, validation and confirmation of numerical models in the earth sciences: Science, v. 263, no. 5147, p. 641–646.

Parker, H. M., 1984, Trends in geostatistics in the mining industry, *in* Verly, G., David, M., Journel, A., and Marechal, A., eds., Geostatistics for natural resources characterization: v. 2, Kluwer Acad. Publ., Dordrecht, Holland, p. 915–934.

Srivastava, R. M., 1996, Matheronian geostatistics: where is it going? *in* Baafi, E. Y., and Schofield, N. A., eds., Geostatistics Wollongong '96, v. 1, Kluwer Acad. Publ., Dordrecht, Holland, p. 53–66.