# Accounting for Different Sampling Errors in Exploratory Drilling Campaigns in Resource/Reserves Modelling

Miguel A. Cuba, Julian M. Ortiz and Oy Leuangthong

*Exploratory drilling is in most cases the only source of information for ore deposit evaluation and for long term mine planning. This information is subject to sampling and measurements errors; this is an unavoidable fact despite the emergence of improved technologies to minimize these errors. Often, information from previous drilling campaigns is also available and must be considered; these tend to show greater measurement errors than their more recent counterparts. Errors are introduced in every stage of measurement: sampling, logging, chemical analysis, and so on. These can be attributed to a variety of factors including instrumentation, sampling protocols and/or procedures. Borehole data typically show 3-5% error when good sampling protocols are carried out, and blast hole data can show as much as 10-30% error given the nature of the data collection practice. While these sampling errors are unavoidable, we can minimize and account for these errors in our resource models. Geostatistical simulation aims to account for spatial variability in the attribute and obtain a statement of uncertainty of the resource/reserve as a result of this variability. Rarely is sampling error considered in this process. For long term mine planning, it is typical to assume that the borehole data are accurate at the data locations. Short term mine models rely primarily on blast hole data and once again, the model takes no account of the inherent inaccuracies and imprecision of the data. Mine plan decisions are made centrally on these resource models without integration of these errors. This paper considers a novel approach to integrate sampling errors from different types and sources of data into the simulation process. Simulation using a Gibbs Sampler is performed at the sample locations, in order to impose the spatial correlation of the attribute of interest and ensure the sampling error is considered. This iterative approach yields multiple realizations of possible values at sample locations. At each location, the distribution of simulated values over a large number of realizations is distributed around the sampled value with the correct relative variance, given by the estimated sampling error. Each realization can then be used to condition a conventional estimation/simulation over a regular grid, allowing the easy introduction of the uncertainty in the data due to their imprecise nature.*

## Introduction

Sampling errors are inevitably attached to most measures of the grades of elements of interest in the mining industry (Sinclair and Blackwell, 2000; Dominy and others, 2002; Stoker and Gilfillan, 2001). Good practice can minimize the impact of sampling errors through proper procedures of quality assurance and quality control (Gy, 1982; Pitard, 2000; Magri and Ortiz, 2000). However, in many mining projects, information from older campaigns is available and should be used even with its associated sampling error. This error can reach relatively high values, but the sample set still provides relevant information about the spatial continuity and local trends of the data, so it must be included in the resource/reserve modelling (Emery and others, 2005).

Including sampling errors in estimation and simulation models has been addressed by several authors. Accounting for qualitative interval information through soft kriging (Journel, 1986) or fuzzy kriging (Diamond, 1988, 1989) is possible to incorporate soft information. Multigaussian kriging and indicator techniques also provide amenable frameworks to impose inequality constraints (Freulon, 1993, 1994; Goovaerts, 1997). Nonetheless, integrating information from several campaigns with different sampling errors is still a challenge. In this paper, we discuss a methodology to account for the sampling error locally, preserving the spatial correlation of the deposit, in a simulation context.

Stochastic simulation approaches honours the conditioning data reproduction at their respective locations; this translates to a local variance equal to zero and a mean of the simulated values equal to the conditioning

data value at all the data locations. Thus, there is no uncertainty at the conditioning data locations and the data values are considered as "true" data.

On the other hand if the conditioning data is assumed to have a certain degree of uncertainty with some known global relative error (GRE) this characteristic of the data should be reproduced in the resulting simulated model as well as the spatial correlation. The uncertainty of the conditioning data can then be transferred to the simulated model.

The proposed approach is to simulate as many datasets as the number of realizations that are required of the model. All the datasets should reproduce (1) the spatial correlation of the initial data in expected value and (2) the expected variability at the sample locations due to sampling error. Only one variogram model is required to simulate all the realizations with the different simulated datasets.

The proposed algorithm draws samples at each data location such that the provided measure of sample error is reproduced. One possible approach to satisfy this objective is to randomly draw samples from a normal distribution centered about the data value with a variance determined from the sampling error. This random drawing at each data location should yield acceptable results when the sample error is relatively small; however in the case of integrating data from different sampling campaigns where the sample error increases with vintage of data, an uncorrelated sampling will degrade the spatial continuity characterized by the variogram. We propose an alternative approach that samples from each location to reproduce the sample error such that the correlation structure of the attribute is preserved. For this task, we draw samples from the cumulative distribution function (cdf) at the locations of the conditioning data following a random path and conditioning the distributions to all current simulated values at sample locations. This is repeated over many iterations, and only samples that approach the target global sampling error (GRE in relative terms) are accepted. The algorithm stops when this target is reached.

We show a comparison between the spatial uncorrelated random sampling and the spatial correlated approach. In both cases the local errors are honoured, but the uncorrelated case leads to a degradation of spatial correlation at the short range.

**Global Sampling Error**

A number of different variability measures could be used to characterize the sampling error. We could consider different sampling distributions at every location; however, it is common to consider a global sampling error that affects the entire set of samples. This is often reported as a relative error, where the variance is proportional to the squared mean at each sample location. For this approach, we use the Global Relative Error (GRE) to characterize the sampling variability. The GRE is defined as:

$$GRE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Z_s(\mathbf{u}_i) - Z_o(\mathbf{u}_i)}{Z_o(\mathbf{u}_i)}\right)^2} \qquad \forall i = 1,\ldots,n \tag{1}$$

Where $Z_o$ are the original data values, $Z_s$ are the sampled values of the alternative simulated datasets, and $n$ is the number of sample locations. It should be noticed that this definition implies that simulated values will be drawn around the actual measured value, which we know has an error attached. One could implement other approaches to account for sampling error variability, but these are not explored in this study.

**Proposed Methodology**

The spatially correlated approach simulates the dataset sample values by means of a Gibbs Sampler. The idea is to compute simulated values from the conditional cumulative distribution functions (ccdf) at each sample location of the initial dataset, following a random path. Two main objectives are considered for this approach: (1) to reproduce the experimental variogram in expected value (Eq. (2)), and (2) to reproduce the relative error.

$$E\left\{\hat{\gamma}_s(\mathbf{h})\right\} = \hat{\gamma}_0(\mathbf{h}) \tag{2}$$

Where $\hat{\gamma}(\mathbf{h})$ is the experimental variogram, and the subscripts s and o denote simulated and original data sets, respectively.

The second objective of reproducing the relative error has two components. Firstly, the global relative error is targeted (see Eq. (1)); this calculates the relative error within the entire dataset. Secondly, a local relative error (LRE) is also targeted such that the relative error at any one data location is also reproduced in expected value, that is:

$$LRE = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\left(\frac{Z_s^l(\mathbf{u}) - Z_o(\mathbf{u})}{Z_o(\mathbf{u})}\right)^2} \tag{3}$$

Where $L$ is the number of realizations of datasets and this is calculated over all realizations for each sample location $\mathbf{u}$.

The spatial correlation reproduction is imposed by conditioning each distribution function with the current simulated sample values at other sample locations and constrained to satisfy both the GRE and LRE. Each sampled dataset approaches the target GRE and LRE through iterations, so the resulting datasets match the variogram, the GRE and the LRE.

Since Gaussian simulation is used, the correlated sampling is performed in normal score units (Deutsch and Journel, 1998). Therefore, each ccdf can be defined by two parameters (mean and variance) and the GRE and LRE comparison is performed after back transformation to original units (where GRE and LRE are measured).

Specifically, the procedure can be summarized as follows:

1. Define the GRE and variogram of normal scores to be matched
2. For each realization:
    a. Initialize the algorithm by generating spatially correlated values with unconditional simulation. These values are not related to the actual samples taken at data locations. Compute the GRE between the current sample values (simulated) and the original sample values.
    b. Select a random path to visit each sample location. At each location do the following:
        i. Compute the ccdf of the normal scores via simple kriging. The kriging estimate and variance are associated with the conditional mean and variance of the local distribution.
        ii. Draw a simulated value from this distribution by Monte-Carlo simulation
        iii. Back transform the simulated value to original units
    c. Update the value of the GRE (Eq. (1)). If close to the target GRE, keep the simulated value and move to the next location; otherwise, reject and draw a new value by going back to Step 2b(ii), or move on to the next node.
    d. Repeat Step 2b and 2c until the GRE and the variogram matches the target within some reasonable tolerance.
3. Over all realizations, visit each location and check that the target LRE is locally reproduced (Eq. (3)). If this is reproduced within some reasonable tolerance, then accept the set of sample data realizations and proceed to geostatistical modelling.

Since the GRE is computed with reference to the actual sampled values, this approach ensures simulated values at data locations will be reasonably close to the actual sampled value, and globally the sampling error will be honoured. Furthermore, the simulated values are spatially correlated due to their construction with the Gibbs Sampler. Many realizations can be computed in this fashion. Each one represents one possible scenario of sample values, consistent with the available information, the spatial continuity and the sampling error of the drilling campaign.

The sampling algorithm is able to simulate many datasets that on average reproduce the spatial correlation of the initial dataset. For different targets of GRE, the dispersion of the experimental variograms of the simulated datasets increases proportionally to the GRE target but remains unbiased.

The sampling algorithm considers the following assumptions:

- Relative errors are constant values into campaigns or sub datasets.
- Sampling errors are interpreted as spatial uncorrelated random processes that lead to the increment that lead to the increment of an independent component in the dataset that consequently increment the nugget effect in the semi – variogram.
- The local distributions of the local sample realizations in the dataset are not Gaussian distributed.

**Application**

A synthetic example is prepared to illustrate the implementation of this methodology. Consider an initial dataset that consists of 400 data samples that are regularly spaced on an interval of 250m x 250m; this regular spacing avoids any clustering effects. The dataset was drawn of one realization of an unconditional simulated map with an isotropic variogram.

The distribution of the initial data is lognormal in original units with a minimum value of 13.73 to a maximum value of 4176.34. In order to work with the sampling algorithm, the normal scores equivalent of the data is needed. The fitted variogram model of the normal score values is isotropic with the range of 1200m, which is consistent with the reference variogram of the synthetic data set.

$$\gamma_{ref}\left(\mathbf{h}\right) = Sph_{1200}\left(\mathbf{h}\right)$$

A GRE of 30% is arbitrarily chosen and imposed on the initial dataset.

For comparison purposes, we consider the case of uncorrelated sampling of a local distribution of uncertainty that is defined based on the GRE. Locally, the distributions are assumed to be normally distributed (which is consistent with a distribution of errors), with a mean given by the original value and a variance that is a function of the target GRE value:

$$\sigma^2(\mathbf{u}_i) = Z_o^2(\mathbf{u}_i) \times GRE^2 \tag{4}$$

Multiple realizations of the dataset can be simulated in this uncorrelated fashion while still honouring the target GRE. For comparison, we generate 50 realizations. Consistency requires that we also impose reproduction of the local relative error (LRE), and for this example, the LRE is constrained to be between a minimum and a maximum of 0.29 and 0.31 respectively (±0.01). Only samples that satisfy this condition are accepted.

Both the uncorrelated sampling and the proposed correlated sampling procedure are compared based on the resulting GRE and the experimental variograms calculated using the resulting simulated datasets. Figure 2 shows the comparison of the GRE reproduction from both approaches, as well as the reproduction of the mean of the distribution over multiple realizations of the sample dataset. In both cases, the GRE is well reproduced; this is expected by construction.

The reproduction of the variogram is verified by checking the normal scores variogram since this is the variogram used in the correlated sampling approach. This requires that each simulated dataset is normal score transformed. Figure 3 shows the results of this check for both the uncorrelated and the correlated sampling approach. In both cases, the overall structure of the experimental variogram is honoured on average by the 50 simulated datasets; however, at short range we can clearly see departures from the original experimental variogram values in the uncorrelated sampling case. This artificial inflation of the nugget effect is significant in cases where the GRE is large (as is the case here). The proposed correlated sampling approach shows good reproduction of the experimental variogram over all ranges.

Figure 4 shows the local distributions for 50 possible samples at three locations selected for being at the low, median and high data ranges of the initial data set. They all have the distribution of sampled values that satisfy the local mean, GRE and LRE sampling conditions as well as the spatial correlation feature. Unlike the uncorrelated sampling approach, these distributions clearly show that the shape of the distribution is irregular and need not be Gaussian.

**Transfer of Uncertainty**

Having characterized the variability of the sampling values at each data location, the next natural step is to transfer this uncertainty into the numerical models of the corresponding geological unit or domain. A straightforward approach to transfer uncertainty in the data values is to directly infer each simulated model using one of the simulated dataset using correlated sampling. For example, for 100 simulated models the input data should consist of 100 different datasets that honour the spatial correlation and the target GRE and LRE values. In this way, the uncertainty due to the sampling error of the input data is transferred to the rest of the simulated locations in the model.

Figure 5 shows the comparison between the conventional approach of conditioning all realizations to the original dataset and the proposed approach of transferring sample error through to the simulated model. The distributions in the global mean and standard deviation are more uncertain than the conventional approach. This is a consistent and expected result since sample error implies uncertainty in the sample data, thus the conditioning data are more uncertain and this is inevitably transferred to the simulation model.

**Conclusions**

Sampling variability is a property of the datasets that has to be considered and reproduced in the models unless it is assumed negligible. It has the same importance as the sample values of interest in the dataset. The proposed approach of constructing multiple realizations of the dataset reproduces the global relative error, as the measure of sample error, and the spatial correlation of the data. An uncorrelated sampling scheme artificially inflates the nugget effect and degrades the spatial structure at short ranges; this is avoided using the correlated sampling scheme proposed in this paper.

The transfer of uncertainty in the sample dataset through to simulation is straightforward. Greater uncertainty in the conditioning data necessarily translates to greater uncertainty in the simulation results. This was shown in terms of the global mean, standard deviation and also in variogram reproduction. In the latter check, greater fluctuation is apparent when we consider sample error in the conditioning information.

The extension of this methodology to include data from different drilling campaigns is simple. The GRE and LRE in the methodology can be imposed for different data sets. Depending on the vintage of information and associated sample errors, we may wish to limit the variogram calculation to be based on the most reliable or least-sample-error dataset. In the instance where multiple drilling campaigns yield datasets that show consistent spatial structure, then the merging of these datasets can yield a relatively reliable estimate of the variogram.

Furthermore, while the proposed methodology used the GRE as the metric for sample error, an alternate expression to capture sample error could easily be considered.

**References**


Deutsch, C.V. and Journel, A.G. (1998) *GSLIB: Geostatistical Software Library and User's Guide, 2nd Edition*, Oxford University Press, 370p.

Diamond, P. (1988) *Interval-valued random functions and kriging intervals*, Mathematical Geology, Vol. 20, No. 3, p. 145-165.

Diamond, P. (1989) *Fuzzy kriging. Fuzzy Sets and Systems*, Mathematical Geology, Vol. 33, No. 3, p. 315-332.

Dominy, S.C., Noppé, M.A., and Annels, A.E. (2002) *Errors and Uncertainty in Mineral Resource and Ore Reserve Estimation: The Importance of Getting it Right*. Exploration and Mining Geology, v. 11, no. 1-4, p. 77-98.

Emery, X., Bertini, J. P., and Ortiz, J. M. (2005) *Resource and Reserve Evaluation in the Presence of Imprecise Data*, CIM Bulletin, Vol. 98, n° 1089, 12 p., Paper 37, September/October 2005.

Freulon, X., (1993) *Conditioning a Gaussian model with* inequalities in Proceedings, Fourth International Geostatistics Congress—Geostatistics Tróia'92 Edited by A. Soares. Kluwer Academic, Dordrecht, 1, p. 201-212.

Freulon, X. (1994) *Conditional Simulation of a Gaussian Random Vector with Nonlinear and/or Noisy Observations in Geostatistical Simulations* Edited by M. Armstrong and P.A. Dowd Kluwer Academic, Dordrecht, p. 57-71

Goovaerts, P. (1997) *Geostatistics for natural resources evaluation*, New York: Oxford University Press, 480 p.

Gy, P. M., (1982) *Sampling of particulate materials - theory and practice*. Elsevier, Amsterdam, 431p.

Journel, A.G. (1986) *Constrained interpolation and qualitative information: The soft kriging approach*, Mathematical Geology, Vol. 18, No. 3, p. 269-286.

Magri, E. and Ortiz J. (2000) *Estimation of Economic Losses Due to Poor Blast Hole Sampling in Open Pits*, in Geostatistics 2000, Cape Town: Proceedings of the Sixth International Geostatistics Congress, W. J. Kleingeld and D. G. Krige, editors, Geostatistical Association of Southern Africa , vol. 2, p. 732-741.

Pitard, F.F. (2000) *Pierre Gy's Sampling Theory and Sampling Practice 2nd Edition*, CRC Press, 488p.

Sinclair, A.J., and Blackwell, G.H. (2002) *Applied mineral inventory estimation.* Cambridge University Press, Cambridge, 381 p.

Stoker P.T. and Gilfillan J.F. (2001) *The Resource Database, in Mineral Resource and Ore Reserve Estimation – The AusIMM Guide to Good Practice (Ed: A C Edwards)*, p 31–36 The Australasian Institute of Mining and Metallurgy: Melbourne

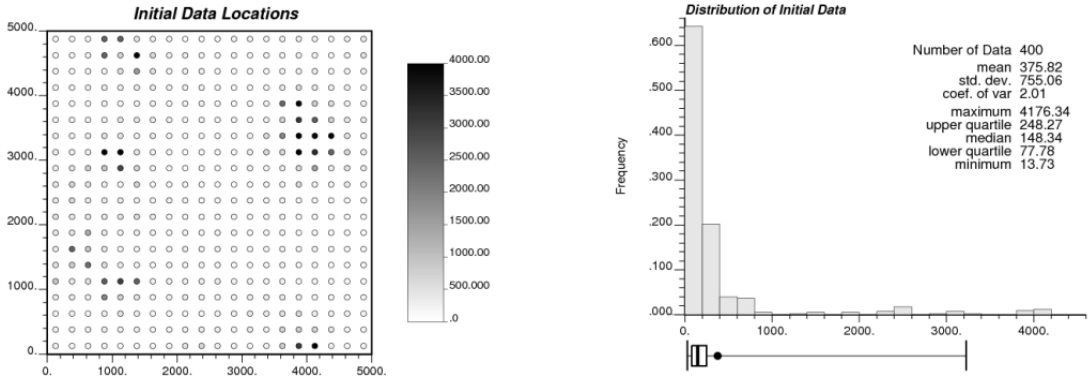


**Figure 1**, Location map of the initial data (left), histogram of initial values (right)

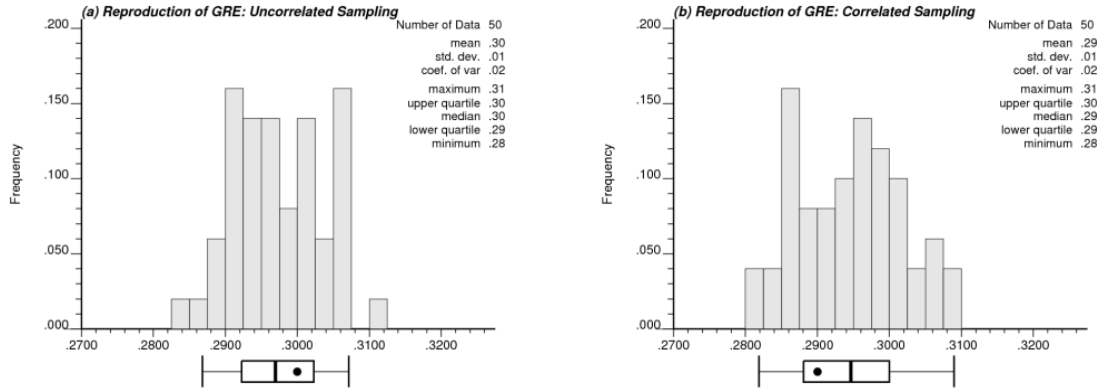


**Figure 2**, Reproduction of target GRE of 30%: (a) using uncorrelated sampling approach and (b) using correlated sampling approach

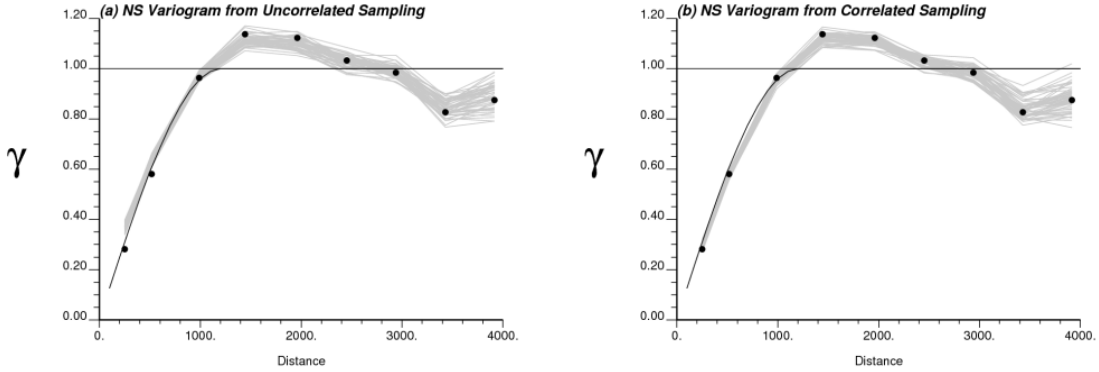


**Figure 3**, Experimental variogram (gray lines) using 50 simulated datasets with 30% GRE: (a) from uncorrelated sampling approach, and (b) from proposed correlated sampling approach. Initial experimental variogram values are shown as black dots, and the fitted variogram model, shown as black solid line, is based on original experimental points.

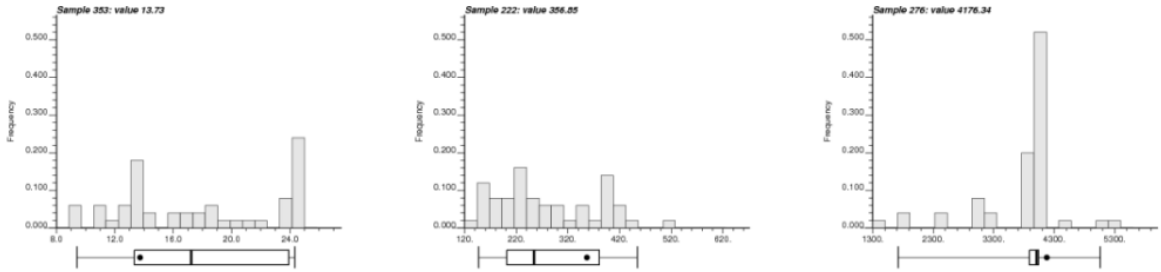


**Figure 4**, Local distributions for the low valued location (left), close to the mean valued location (center), and the maximum value location (right) from correlated sampling approach

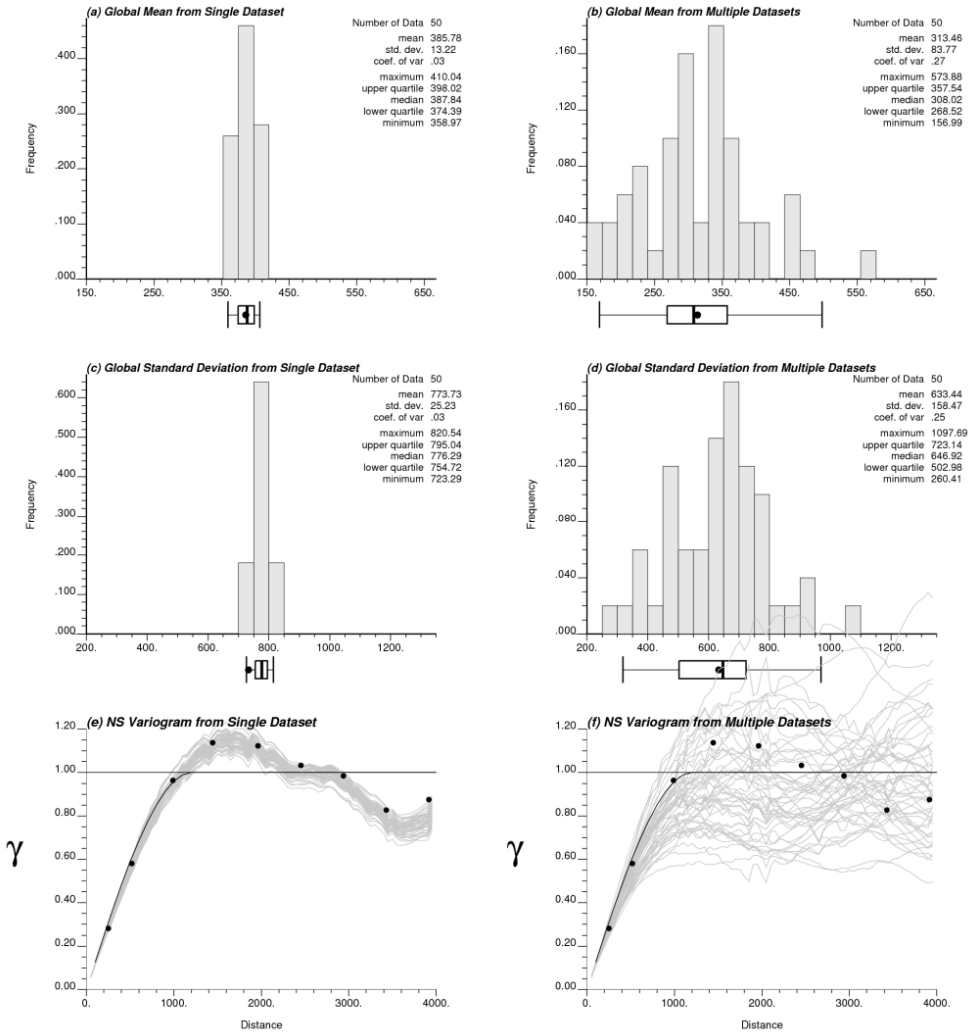


**Figure 5**, Comparison of simulation results using a single sample dataset (left column) and multiple simple datasets (right column): (a) and (b) distribution of global means over 50 realizations (top row), (c) and (d) distribution of standard deviation over multiple realizations (middle row), and (e) and (f) normal scores variogram reproduction (bottom row).