

# ACCOUNTING FOR MULTIPLE-POINT CONTINUITY IN GEOSTATISTICAL MODELING

C. V. Deutsch

University of Alberta, Edmonton, Alberta, CANADA ([cdeutsch@civil.ualberta.ca](mailto:cdeutsch@civil.ualberta.ca))

E. Gringarten

Landmark Graphics Corp., Austin, Texas USA ([egringarten@zycor.lgc.com](mailto:egringarten@zycor.lgc.com))

Most geologic phenomena show non-linear or high order connectivity, which can be appreciated from outcrops, forward geologic modeling, or dense sampling. The use of multiple-point statistics, however, has not advanced to routinely account for such information. Two-point variogram statistics remain in widespread usage because they are straightforward to infer and robust geostatistical simulation methods exist. We show two developments directed toward accounting for realistic high order geologic continuity: (1) calibration of the variogram to account for high order continuity, and (2) direct accounting for such continuity in an annealing-based simulation approach.

Variograms may be calibrated so that analytical dispersion variances match experimental dispersion variances calculated from a series of different scales. We show that, in general, the variogram range must be extended by up to 50% to account for additional high order continuity. We also show that geostatistical models created with the corrected variograms better reflect the spatial distribution and flow properties of the true underlying population. This first proposal improves the accuracy of conventional algorithms.

Direct use of multiple-points statistics may be desirable. Different methods exist for this task. The use of annealing is gradually becoming practical with increasing computer power and improved understanding of problem formulation. We show that annealing can be set up to integrate multiple-point statistics in a practical and efficient manner.

Our proposals are illustrated with two sedimentary examples. The theoretical prediction of dispersion variance is compared to observation. The proposed technique for variogram calibration is illustrated. We show more realistic geostatistical simulations and improved integration of data.

## INTRODUCTION

Geologic features rarely exhibit continuity that can be adequately described by two-point statistics. Nevertheless, two-point statistics remain in widespread usage for many reasons including (1) many software programs and accumulated expertise exist to use covariance/variogram-type statistics and (2) the direct use of multiple-point statistics is problematic.

Many proposals exist for the direct use of multiple-point statistics (direct inference of conditional probabilities, 1990, extended normal equations, Guardiano and Srivastava, 1992; annealing-based schemes, Deutsch, 1992; and neural network iterative schemes, Caers and Journel, 1998). The inference of multiple-

point statistics require a large amount of data. Historic data from previous production or analogous areas, geologic maps, or the result of forward geologic modeling provides the needed data. The relevance or *stationarity* of the multiple-point statistics depends heavily on the data source. This is exacerbated by the fact that multiple-point statistics contain all lower-order statistics including the histogram and variogram. Adopting a training-image or analogue data set amounts to assume the present study area is remarkably similar to the analogue data. Another challenge with the use of multiple-point statistics is the implementation challenge of building geostatistical models that replicate the multiple-point statistics. Developments in iterative schemes have overcome, to a large extent, this challenge.

We present two ideas in this paper (1) a procedure to calibrate the variogram to account for higher-order connectivity, which would permit remaining in the familiar world of variogram-based geostatistics and yet injecting more geological continuity, and (2) direct use of multiple-point statistics in construction of geological models.

Our first proposal consists of increasing the variogram range to account for presence of high-order and non-linear connectivity. The variogram range is increased so that the variance of larger-volumes matches that inferred from analogue data. This matching procedure amounts to transfer high-order connectivity into the variogram *without* transferring the histogram or other non-transportable features of the analogue data. It is robust and not overly dependent on the analogue data.

The second proposal is suitable for situations where all aspects of the analogue data are considered reliable; the multiple-point statistics are reproduced via an objective function. We use “histograms of different support,” introduced by R. M. Srivastava, as the multiple-point statistics to be reproduced. This alternative transports more features from the analogue data than would be possible with any variogram-based technique.

## MULTIPLE POINT DISTRIBUTIONS

There are many multiple-point statistics that could be considered. The most general one is the multiple-point histogram. Continuous variables are divided into  $K$  classes. Then, different configurations of  $N$  points are chosen. The multiple point histogram is represented by the frequency of patterns observed for the  $N$ -points simultaneously. The difficulty with such statistics is that  $K^N$  classes must be informed for just one  $N$ -point configuration. There would be 1 billion classes for  $K=10$  and  $N=9$ . This would require a large training image indeed. Moreover, there are many possible configurations of  $N$  points; aligned in a square, circle, line, and so on. Summary measures such as multiple-point covariances have been considered to reduce the number of data needed.

There are other summary measures that could be used in the context of scale-up and integration of data of different volume supports; in particular, *histograms of different support* introduced by R. M. Srivastava. Consider block average values  $z_v(\mathbf{u}) = \int_{v(\mathbf{u})} z(w)dw$  for different volumes  $v$  and all locations  $\mathbf{u}$  in the domain. A series of different block sizes  $v_i, i=1, \dots, n_v$  are considered. The block values  $z_{v_i}(\mathbf{u})$  for a specific size  $v$  are pooled into histograms. The shape and character of the  $n_v$  histograms provide information on

multiple-point connectivity. Large- $v$  average values from random or disconnected phenomena will all be close to the mean; high and low  $z$ -values average out quickly. In the case of highly correlated fields, the variance of large- $v$  average values reduces slowly.

At first approximation, we consider only the variance to summarize the information contained in such histograms of different support. We use the entire histogram later. These variances are experimental dispersion variances (see Isaaks and Srivastava, 1987 or Journel and Huijbregts, 1978):

$$\hat{D}^2(v_i, A), i = 1, \dots, n_v \quad (1)$$

Such dispersion variances may also be calculated analytically from the variogram model, that is, with average variogram values:  $D^2(v_i, A) = \sigma^2 - \bar{\gamma}(v_i, v_i)$ , where  $\sigma^2$  is the stationary variance. In presence of non-linear or high order connectivity the experimental dispersion variances  $\hat{D}^2(v_i, A)$  will be greater than those predicted by the variogram model  $D^2(v_i, A)$ . To account for the additional information in the experimental dispersion variances, in fact the multiple-point distributions of  $v$  average values, we consider two options (1) correcting the variogram to capture the non-linear connectivity, and (2) directly using the additional information in an annealing-type simulation algorithm.

## CORRECTED VARIOGRAMS

The use of variogram-based techniques is widespread and likely to remain so. Techniques that directly use multiple-point statistics have been adopted slowly due to problems of inference and implementation. For this reason, we would like to “correct” the variogram to account for the additional connectivity. In general, the variogram ranges for each nested structure must be increased to correct for greater than expected high order connectivity. Implicit multiGaussian and maximum entropy assumptions underlying geostatistics entails that the “corrected” variogram ranges will always be larger than observed from the data.

The variogram ranges are iteratively adjusted until the following objective function is minimized:

$$O = \sum_{i=1}^{n_v} \left[ \hat{D}^2(v_i, A) - (\sigma^2 - \bar{\gamma}(v_i, v_i)) \right]^2 \quad (2)$$

This objective function is a measure of mismatch between the observed dispersion variance and the theoretical prediction due to the variogram model. This problem may be solved in seconds on a PC using a simulated annealing algorithm. There are few parameters in this problem and convergence is fast. The corrected variogram ranges are used in subsequent geostatistical modeling.

## EXAMPLES

Two examples are given. Figure 1 shows an image first presented in Deutsch, 1992. A gray scale representation of the image is at the upper left corner. The values have been transformed to a standard normal distribution. The variogram is shown at the lower left: the dots are the experimental variogram in both directions, the solid line is the fitted variogram model, and the dashed line is the corrected variogram.

The correction was done by calibration with dispersion variances computed from 2x2, 5x5, and 10x10 windows. The chart to the right of Figure 1 shows the experimental dispersion variances (light colored – red – dots), the variances from the fitted variogram (x), and from the corrected model (black dots). The dots overlap, which indicates a close match with the corrected model. The variogram model consists of a nugget effect of 0.1, which remains unchanged, a first nested structure with variance contribution 0.45, where the range doubles, and a second nested structure, where the range increases by 50%.

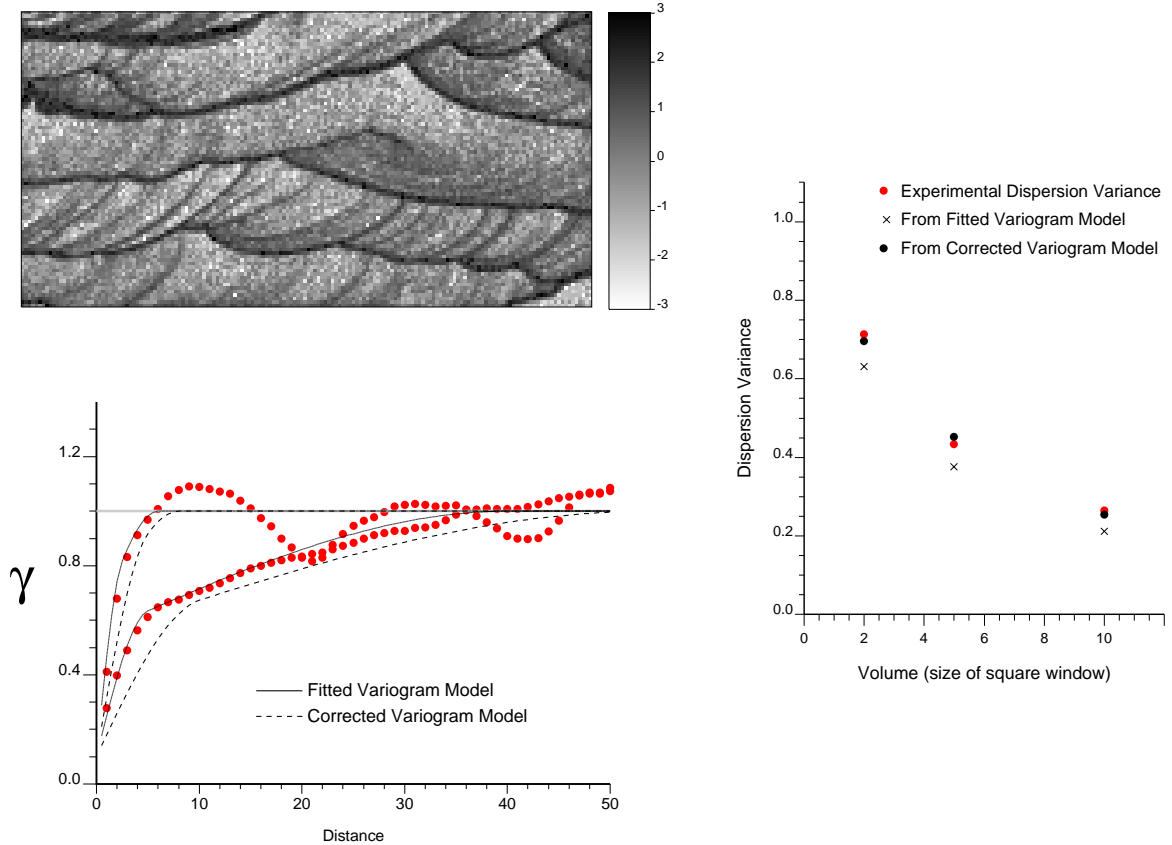


Figure 1: Aeolian sandstone with variogram and dispersion variances. The gray scale image corresponds to the content of fines transformed to a standard normal distribution

The variogram has changed. It is difficult to assess the impact of this change without processing the reference image and geostatistical simulations through some form of transfer function such as a flow simulator. The normal scores values are transformed to a lognormal permeability distribution and single-phase flow was simulated. The effective permeability in the X-direction was calculated for 68 (4 by 17) coarse grid superimposed on the image. The mean and variance of the resulting permeability values may be compared from the reference image, a simulation using the fitted variogram, and a simulation using the corrected variogram. We see that the corrected variogram leads to flow characteristics significantly closer to the reference.

	Reference	Fitted Model	Mismatch	Corrected Model	Mismatch
Mean (md)	45.3	40.9	-10%	44.0	-3%
Variance (md <sup>2</sup> )	2387	728	-70%	1936	-19%

Figure 2 shows a realization with the fitted model and the corrected model. The longer range of correlation is revealed visually on the right hand side: the corrected variogram model.

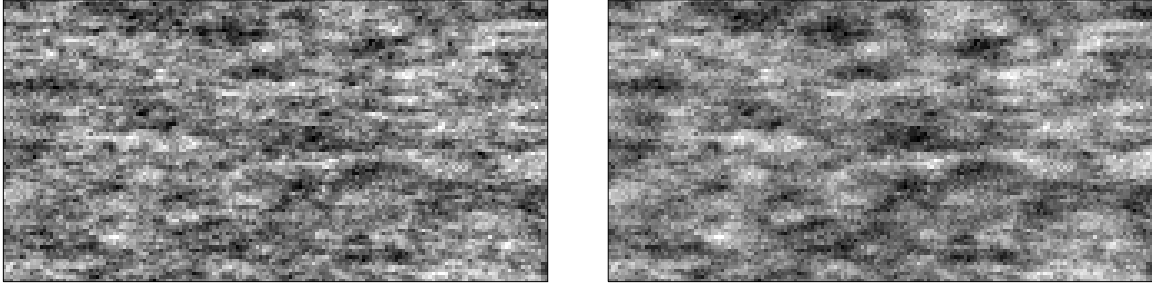


Figure 2: Gaussian simulation with fitted variogram model (left side) and corrected variogram (right side).

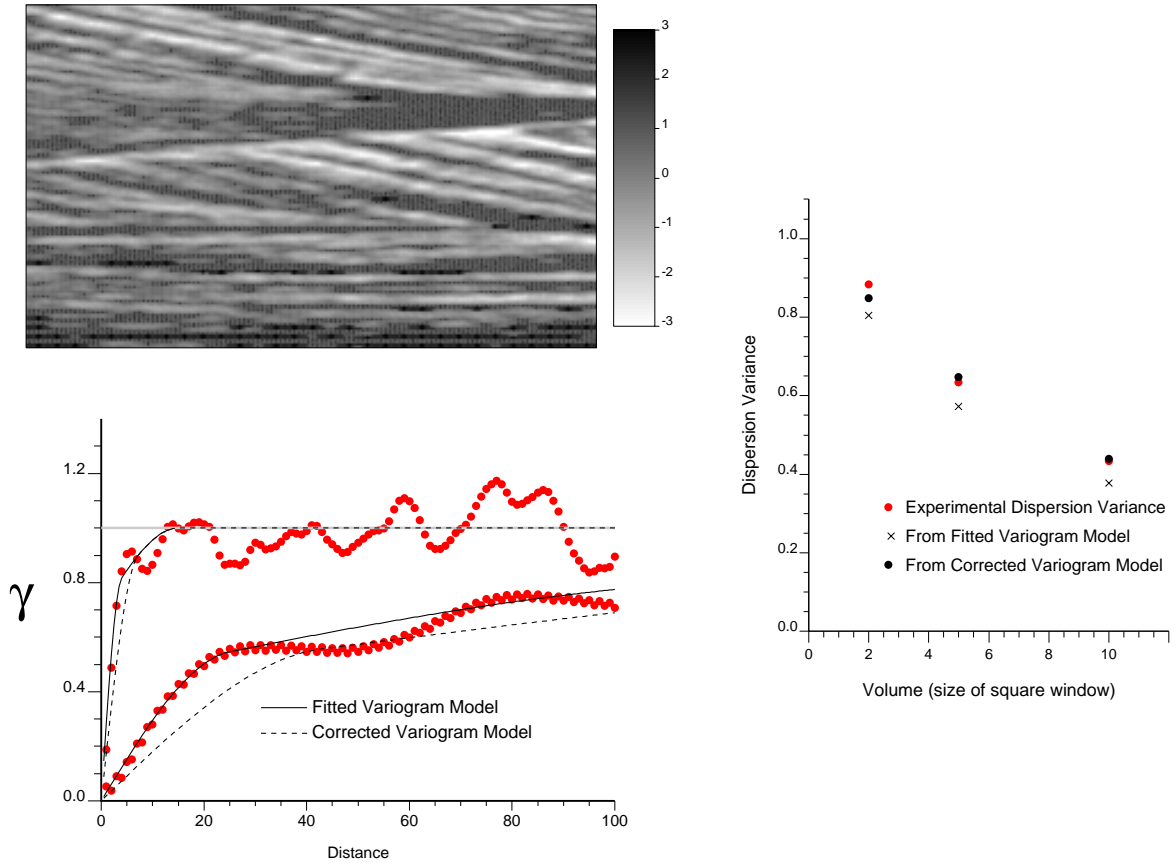


Figure 3: Sandstone cross bedding with variogram and dispersion variances. The gray scale image corresponds to the content of fines transformed to a standard normal distribution

Figure 3 shows the image corresponding to the second example. A gray scale representation of the image is at the upper left corner. The variogram is shown at the lower left: the dots are the experimental variogram in both directions, the solid line is the fitted variogram model, and the dashed line is the corrected variogram. The correction was done by calibration with dispersion variances, see chart to the right. Once again, the dots

overlap, which indicates a close match with the corrected model. The variogram model consists of three nested structures explaining 45, 25, and 30% of the variability; the range of the first structure increases by 40%, the second 35%, and no change is made to the third structure. Simulated realizations created with the fitted and corrected variogram models differ in flow character and visual appearance.

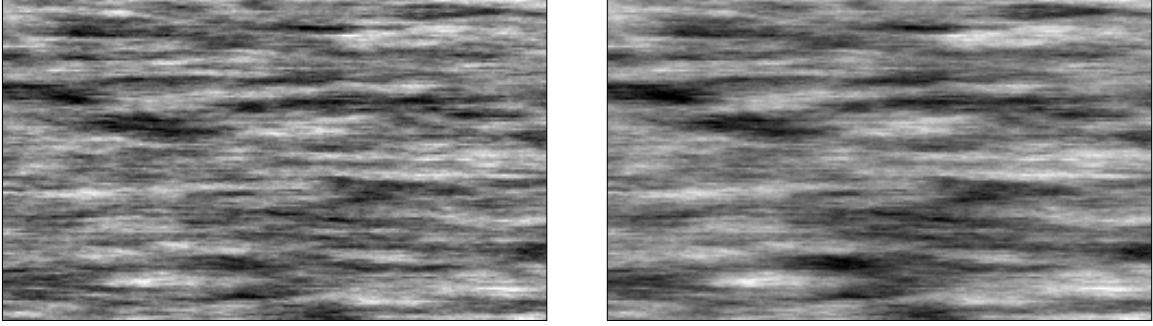


Figure 4: Gaussian simulation with fitted variogram model (left side) and corrected model (right side).

### Closer Look at Histograms of Different Support

Correcting the variogram to match experimental dispersion variances captures only one aspect of the histograms of different support: the variance. Figure 5 shows the actual histograms of different support, the theoretical prediction using the fitted variogram (Gaussian model with the lesser variance – higher valued at median), and the theoretical prediction using the corrected model (higher variance Gaussian model). The histograms on the left side correspond to the Aeolian sandstone (Figure 1) and the histograms on the right correspond to the ripples (Figure 3). The histograms at the top are for 2x2 averaging, the middle are for 5x5, and the bottom are 10x10. We see many interesting features on the histograms including skewness and modes that are not captured by the Gaussian distributions.

### DIRECT REPRODUCTION OF MULTIPLE-POINT STATISTICS

As mentioned above, there are many ways to directly reproduce multiple-point statistics in geostatistical simulation. The use of simulated annealing is powerful, but has been hampered by excessive CPU requirements the expertise required to make it work quickly and artifact-free. Recent understanding of the critical temperature (see paper by Norrena in this conference) and developments of the annealing schedule are making the method more practical. In the context of histograms of different support, a component objective function is added:

$$O_h = \omega_h \sum_{i=1}^{n_v} \sum_{j=1}^{n_z} [\hat{F}_{v_i}(z_j) - F_{v_i}^*(z_j)]^2 \quad (3)$$

where  $\omega_h$  is the weight of this component (see Deutsch 1992),  $n_v$  is the number of histograms,  $n_z$  is the number of quantiles,  $\hat{F}_{v_i}(z_j)$  is the experimental cdf, and  $F_{v_i}^*(z_j)$  is that of the realization.

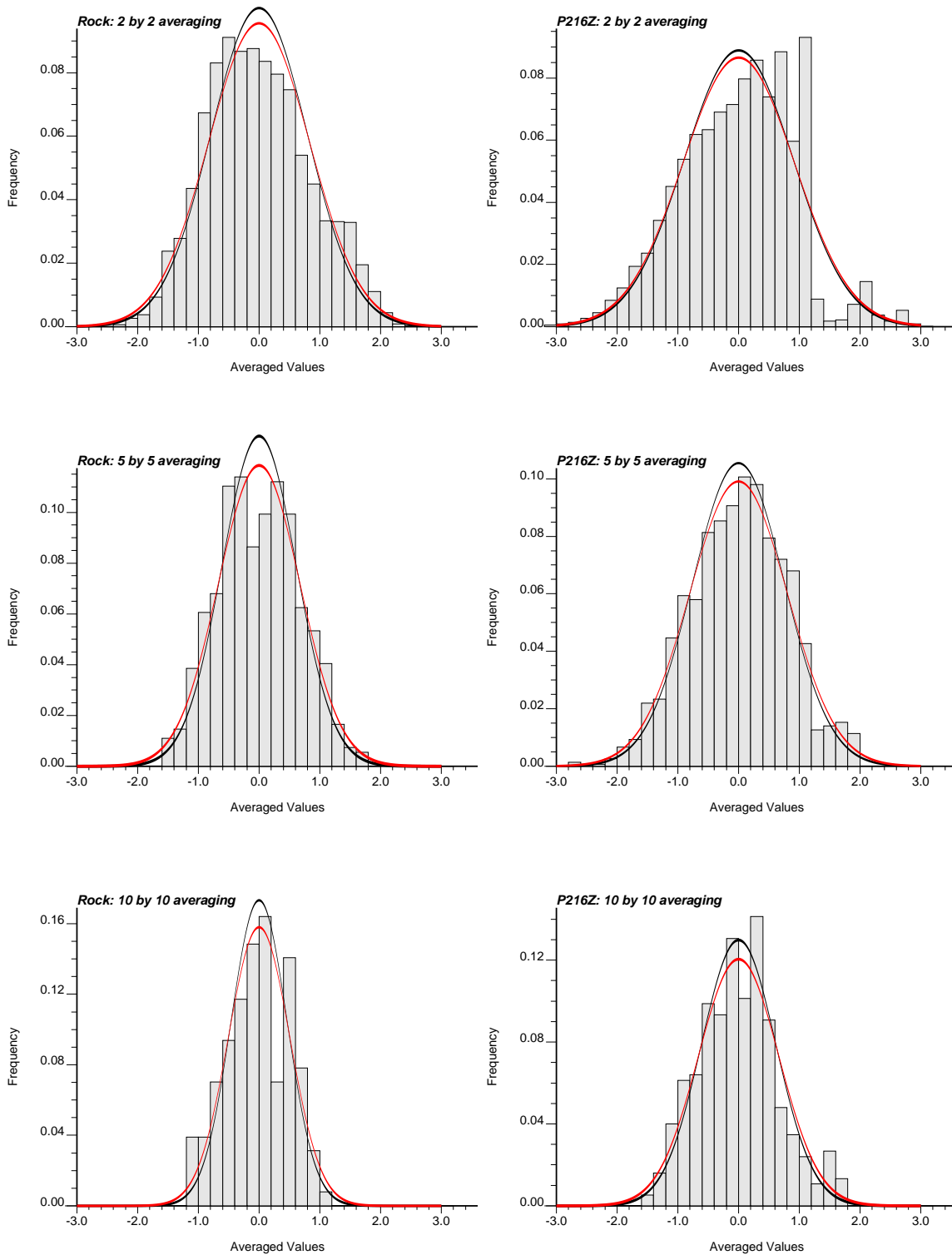
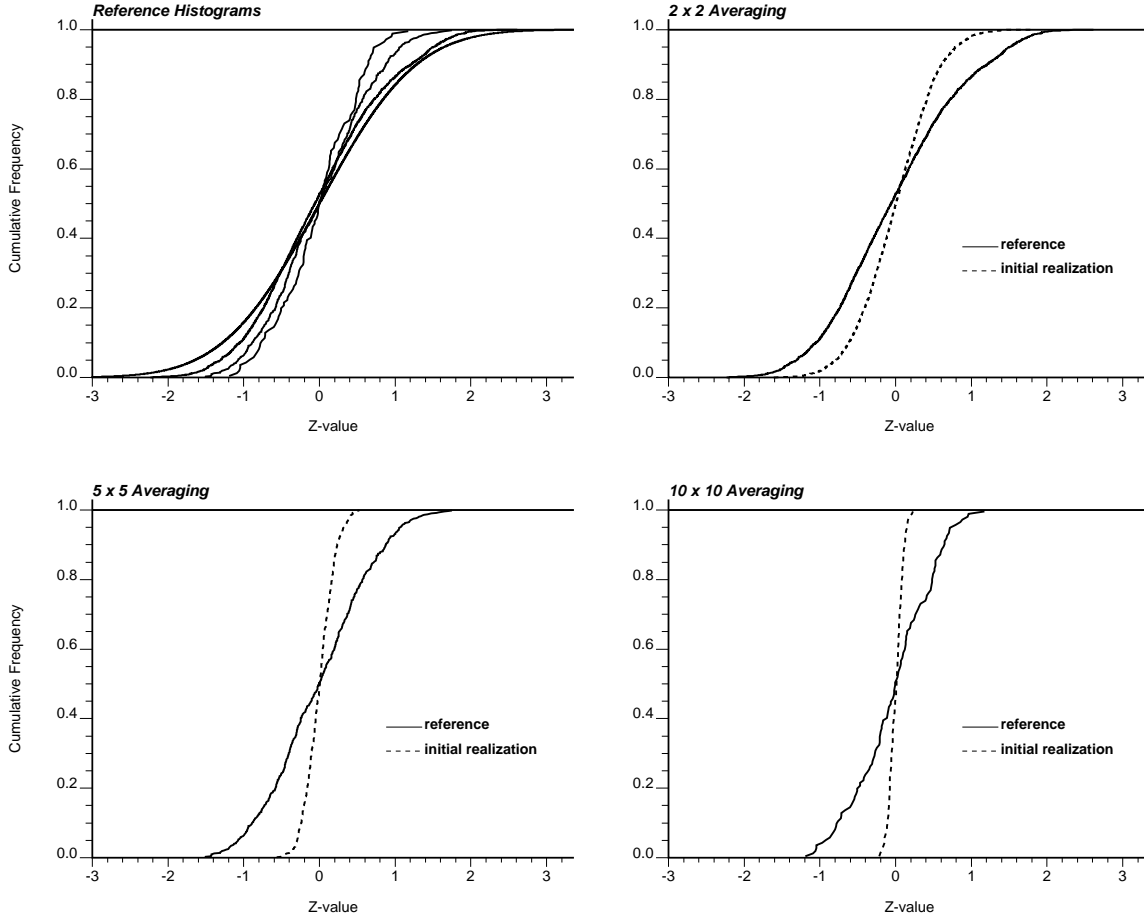


Figure 5: block histograms (histograms of different support) for the Aeolian sandstone (left) and the ripples (right) for 2 by 2, 5 by 5, and 10 by 10 averaging. The point-histogram in both cases is standard normal. Classical geostatistics would predict that block histograms would also be normal with reduced variance. Two normal distributions are shown on each histogram: one corresponds to the fitted variogram model – the lesser variance one; the second corresponds to the corrected variogram – the greater variance. The variance reduce as the volume of averaging increases, but the shapes deviate from normality. We see skewness and multiple modes (particularly on the right hand “ripple” case) that could only be reproduced in geostatistical simulation by direct means (for example simulated annealing).

The `sasim` program of GSLIB was modified to include a component objective function of this form. As with most spatial statistics, inclusion in an annealing-type program is straightforward. The histograms of different support are calculated globally at the start and locally updated through the annealing process. Some implementation details: (1) the number of quantiles  $n_z$  is chosen to capture all relevant features on the histograms of different support, for example, 50 would be adequate for the features observed on Figure 5, (2) the quantiles are chosen at equal probability intervals, that is, the first at  $\hat{F}_{v_i}(z_1) = 1/(2 n_z)$ , then incremented by  $1/n_z$ , thus, the probability values  $\hat{F}_{v_i}(z_j), j=1, K, n_z$  are equally spaced by  $1/n_z$  (3) the threshold values  $z_j$  are chosen to be the correct quantiles, (4) the realization cdf  $F_{v_i}^*(z_j)$  is computed by the proportion of block average values less than  $z_j$ , the block average values are calculated from the realization, (5) the block average values are arithmetic averages (although they could be any non-linear average), and (6) the block average values and the realization cdf values  $F_{v_i}^*(z_j)$  are updated as small scale values are perturbed. Figure 6 shows the results for the image of Figure 1 and an initial random realization.

Figure 6: block histograms for simulation. *Upper left*: small scale normal distribution and reference 2x2, 5x5, and



10x10 block histograms; *Upper right*: reference 2x2 block histogram and the result of an initial random realization. Note that the initial block histogram has significantly less variance than the reference image even though the small scale histogram is honored. Lower left and right: reference and initial histograms for 5x5 and 10x10 averaging.



The revised `sasim` program was used to generate realizations with and without the multiple-point histograms. These realizations are shown on Figure 10. The image on the right reproduces the point histogram, variogram, and multiple point histograms on Figure 7. The CPU time required to generate the two realizations were 2.03 and 2.45 minutes on an older SGI O2 workstation. The annealing was stopped when the overall objective function dropped to less than 0.001, that is,  $1/10^{\text{th}}$  of 1% of the initial objective function. All components of the objective function were able to go to “zero” simultaneously because there is no inconsistency between them.

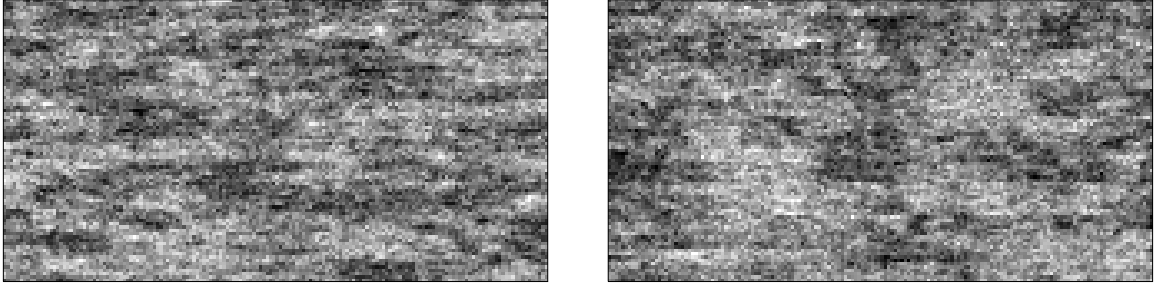


Figure 10: annealing based simulation without (left) and with (right) the three histograms of different support. Both realizations reproduce the target histogram and the fitted variogram model.

The realizations on Figure 10 could be made more similar to the reference image of Figure 1 by transporting more spatial statistics from the reference image. In fact, the reference image could be reproduced exactly by reproduction of sufficient statistics (all two-point statistics are enough); however, that is not our goal. We desire to transport only those features deemed reliable. In this case, we transport the averaging or scale-up characteristics through histograms of different support. Were the curvilinear features considered transportable, a surface-based approach would have been considered to directly replicate those features and the internal heterogeneities (see paper by Xie, Deutsch, and Cullick in these proceedings).

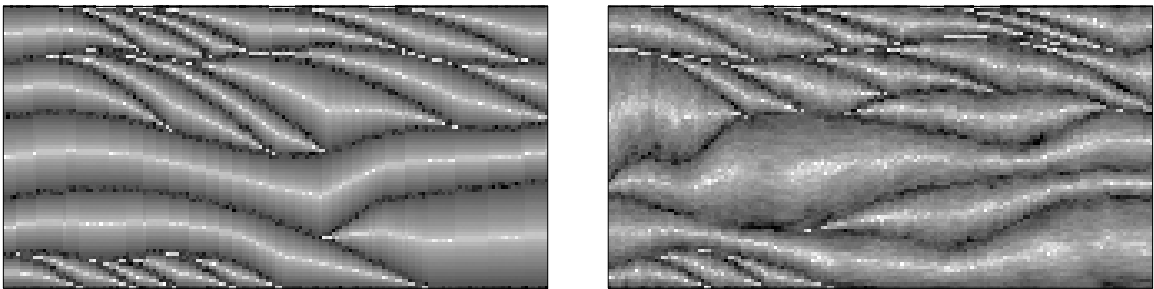


Figure 11: two different surface-based realizations using parameters taken from the reference image shown on Figure 1. This approach would be appropriate if the important heterogeneity is considered to be the high values associated to surfaces.

## CONCLUSIONS

Geostatistical simulation is being increasingly used to construct realistic heterogeneity models and to assess uncertainty. To provide reliable predictions, simulated realizations must share all known spatial features of the underlying true spatial distribution. Variogram-based techniques are in widespread use, but are limited to two-point linear connectivity; implicitly, all high-order connectivity is maximally uncorrelated. The variogram range of correlation, however, may be increased to account for such high-order connectivity. A procedure and two examples were presented. The results of this study are corroborated by flow studies where the realizations with the corrected variogram range provide flow predictions closer to reality.

Direct integration of multiple-point statistics is possible with a number of techniques. We show how simulated annealing could be used to directly reproduce histograms of different support in simulated realizations. The direct use of multiple-points statistics significantly reduces the variability between simulated realizations, which is desirable if the source of multiple-point statistics is reliable, but dangerous if the source is not directly relevant.

Future work consists of (1) gaining experience and validating variogram calibration, (2) devising means to filter univariate statistics from multiple-point statistics, that is, extract multiple-point continuity without transporting the histogram. This is straightforward with continuous variables but problematic for categorical variables, (3) selecting the appropriate multiple-point statistics, that is, the number and configuration of the  $n$ -point statistic, and (4) practical artifact-free simulation algorithms.

## REFERENCES

- Caers, J. and A.G. Journel, Stochastic Reservoir Simulation Using Neural Networks Trained on Outcrop Data, SPE 49026, *Proceedings of SPE ATCE*, New Orleans, LA, October 1998.
- Deutsch, C.V., and A. G. Journel, *GSLIB: Geostatistical Software Library*, Oxford University Press, New York, NY, Second Edition, 1997
- Deutsch, C.V., *Annealing Techniques Applied to Reservoir Modeling and the Integration of Geological and Engineering Data*, Ph.D. Thesis, Stanford University, 1992
- Guardiano, F. B., and Srivastava, R.M., Multivariate Geostatistics: Beyond Bivariate Moments, *Geostatistics Troia '92*, Volume 1, Kluwer Academic Publishers, Boston, 1993, pp. 133-144.
- Isaaks, E.I., and Srivastava, R.M., *An Introduction to Applied Geostatistics*, Oxford University Press, New York, NY, 1987
- Journel, A.G., and Huijbregts, Ch., *Mining Geostatistics*, Academic Press, New York, NY, 1978

**Dr. Clayton V. Deutsch**

Dr. Deutsch is presently an Associate Professor in the Department of Civil & Environmental Engineering at the University of Alberta. He teaches and conducts research into better ways to geostatistically predict the spatial distribution of mineral grades, petrophysical properties in petroleum reservoirs, and contaminant concentrations. Prior to joining the University of Alberta, Dr. Deutsch was an Associate Professor (Research) in the Department of Petroleum Engineering at Stanford University and Director of the Stanford Center for Reservoir Forecasting (SCRF). He has worked at Exxon Production Research Company and Placer Dome Inc.. Dr. Deutsch has a B.Sc. in Mining Engineering and M.Sc. and Ph.D. degrees in Applied Earth Sciences from Stanford University.

**Dr. Emmanuel Gringarten**

Dr. Gringarten is presently with T-Surf Corporation in Austin, Texas. He is Product Manager for geostatistics and the Jacta product, an uncertainty management software. Prior to joining T-Surf, Dr. Gringarten was a domain specialist with Landmark Graphics. He has a B.Sc. in Mathematics from Imperial College and M.Sc. and Ph.D. degrees in Petroleum Engineering from Stanford University.