**AAPG HEDBERG SYMPOSIUM**
**"Applied Reservoir Characterization Using Geostatistics"**
**DECEMBER 3-6, 2000, THE WOODLANDS, TEXAS**

# Determination of Reliable Histogram and Variogram Parameters for Geostatistical Modeling

*Emmanuel Gringarten (T-Surf), Peter Frykman (GEUS), and*
*Clayton V. Deutsch (Univ. of Alberta)*

*All geostatistical modeling algorithms require histogram input for the proportions of lithofacies or the distribution of petrophysical properties such as porosity and permeability. Such histogram input has a first-order control on volumetrics and flow behavior of reservoir models. Most algorithms also require variogram input, which specifies the spatial continuity of the lithofacies and petrophysical properties. In presence of sparse well data, the variogram has significant control on the appearance and flow behavior of reservoir models. Although the histogram and variogram are of unquestioned importance in reservoir modeling, they are often hastily assembled with little regard for geological relevance or statistical representivity. This paper presents guidelines and principles to assist reservoir modelers determine reliable histograms and variograms.*

## Introduction

Geostatistical techniques slavishly reproduce input lithofacies proportions and the histogram of petrophysical properties. Variations in the reproduction of the histogram are minor and attributable to ergodic statistical fluctuations and not to true uncertainty. The naïve histogram from well data must often be corrected. Wells may be preferentially located in high-pay zones and core plugs may be preferentially taken from net reservoir facies. Classical declustering techniques, such as cell declustering and polygonal declustering, may be sufficient when there are many data (tens of wells). Seismic data and/or geological trend mapping must be used in presence of few wells.

Failure to apply declustering techniques prior to geostatistical modeling can easily introduce a bias of up to 30% in the average porosity or facies proportions. Proceeding with reservoir modeling without declustering could lead to large errors in static reserves and dynamic flow predictions. The histogram is important for both cell-based geostatistical modeling *and* object-based modeling.

The variogram is a critical input to cell-based geostatistical studies: (1) it is a tool to investigate and quantify the spatial variability of the phenomenon under study, and (2) most geostatistical estimation or simulation algorithms require an analytical variogram model, which they will reproduce within statistical fluctuations. In the construction of numerical models, the variogram reflects some of our understanding of the geometry and continuity of the variable, and has a very important impact on predictions from such numerical models. Interpretation of and modeling of the variogram consistent with geological principles is critical for reliable geostatistical models.

Selection of a variogram by default or on the basis of computer visualization with extreme vertical exaggeration is questionable. Flow predictions are sensitive to the lateral continuity of critical high- and low-permeability heterogeneities.

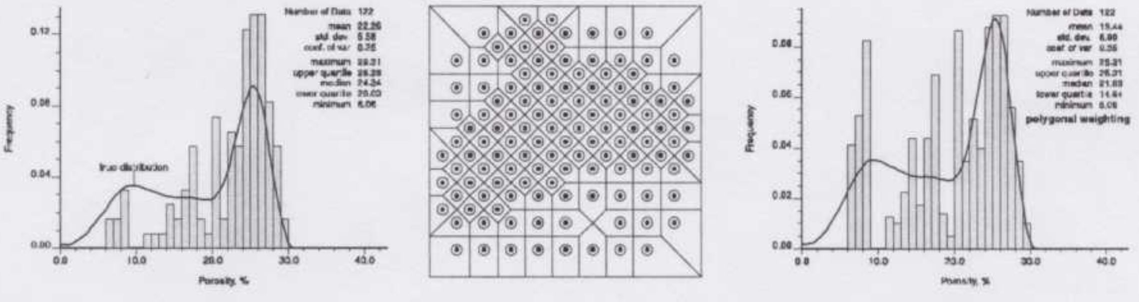
No rational geostatistical reservoir modeling workflow is complete without special attention to critical inputs such as the proportions/histograms and measures of spatial continuity (Gringarten & Deutsch 1999; Kupfersberger & Deutsch 1999).

## Classical Declustering

Many contouring or mapping algorithms automatically correct preferential clustering. Closely spaced data inform fewer grid nodes; hence, receive lesser weight. Widely spaced data inform more grid nodes; hence, receive greater weight. Even though modern stochastic simulation algorithms are built on the mapping algorithm of kriging, they do *not* correct for the impact of clustered data on the target histogram. Simulation always gives some weight to the input global distribution, particularly in areas of sparse well control. Moreover, transformation to a standard normal distribution does not remove the need for a representative distribution; back-transformation ensures that the original histogram is used.

*Declustering* techniques assign each datum a weight, $w_i$, $i=1,\ldots,n$ based on its closeness to surrounding data, and two methods have been devised to account for this effect. A datum in an area that is sampled sparsely will receive more weight than a datum in a densely sampled area. The weights are then used as frequencies of occurrence to generate the histogram or frequency distribution. Two commonly used declustering techniques are the polygonal method ((Isaaks & Srivastava 1989), p.238-239) and the cell-declustering methods (Journel 1983; Deutsch 1989).

The polygonal method first determines polygons of influence for each data location (Isaaks & Srivastava 1989). The declustering weight is taken as proportional to the area of the polygons of influence. Clustered data with small polygons of influence receive less weight than isolated locations with large polygons of influence. The weights then are used as frequencies of occurrence to generate a weighted histogram. It has been observed that polygonal declustering technique works well when the limits (boundaries) of the volume of interest are well defined and the polygons do not change in size by more than a factor of, say, 10 (largest area/smallest area).




A B C

*Figure 1: (A): Histogram of sampled values from an underlying true distribution. The sampling locations is shown in (B): The geometry of the polygons for declustering. (C): The histogram of the corrected distribution with polygonal declustering applied.*

Another widely used technique is called cell declustering. The Cell declustering algorithm first divides the volume of interest into a grid of cells. The number of data in each occupied cell is calculated. There are a number of parameters such as the location (origin and orientation) of the grid and the cell size that must be determined (Deutsch 1989). The number of data in a particular area is used to assign the weights. The main requirement of both these declustering procedures is that enough data must be present to perform the declustering to arrive at unbiased statistics.

## Soft-Data Declustering

The classical declustering methods only work when there are enough data to assign greater and lesser weights. The conventional polygonal and cell declustering methods are inadequate to determine a representative distribution unless there is adequate data coverage in both "good" and "poor" areas of the reservoir. However, if this coverage is biased, there may often be sufficient seismic or geological data to know a-priori where the high-pay good areas are; the first wells would be located in the best areas. Then, later in reservoir development, a geostatistical model may be constructed for further development planning. At that time, unbiased statistics are needed. The same secondary seismic or geological data used to collect the biased data (in a good sense) can be used to determine representative statistics for geostatistical modeling (Frykman & Deutsch 1998). The first requirement is a spatial distribution of some secondary variable like seismic impedance, a hand contoured net-to-gross map, or simply structural depth if reservoir quality is linked to depth.

In the example illustrated on Figure 2, the reservoir quality (porosity and permeability) in the reservoir are better in the highest part of the field and degrade toward the flank areas. Wells with available data are mostly situated on the crestal part of the reservoir. Since a general porosity decrease is associated with increasing depth, this imposes an under-representation of low porosity and permeability values when a model of a larger area is requested. We therefore need to derive a representative distribution for input to the geostatistical simulation. The model-based declustering technique, based on calibration with a secondary reservoir attribute (a depth model which is unbiased in this case), is presented to overcome this first challenge.

The model established for the porosity-depth relation is guided by the available well data from the vertical wells in the field, and by additional data from a well in the surrounding area, and the method relies on the merging of conditional distributions derived from a bivariate distribution model (Figure 2).

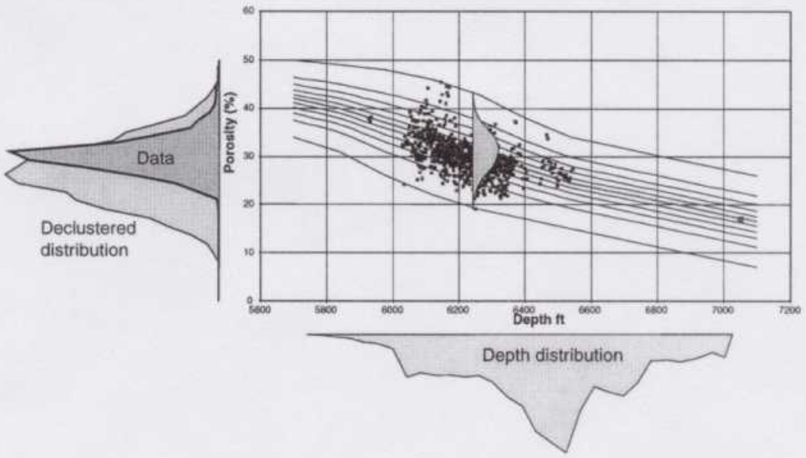



*Figure 2. The bivariate model for the relation between porosity and depth is shown as decile distribution lines. For each depth datum, a conditional distribution of porosity can be extracted, and given the depth distribution for the model volume, these conditional distributions can be merged applying the weight given by the depth distribution. This results in the declustered distribution, significantly different from that given by the initial well data.*

The method of model-based declustering has been successfully applied in studies on the Dan field, using depth and seismic impedance respectively as secondary variables (Frykman & Deutsch 1996; Vejbæk & Kristensen 2000).

## Analogue Data for Variogram Interpretation

Most wells are vertical. This makes it straightforward to infer the vertical variogram, but difficult to infer a reliable horizontal variogram. Given the overwhelming noise content in sample horizontal variograms, one evident error is to adopt a "random" variogram model, which appears to closely fit the experimental variogram. This is convenient but unrealistic. Our goal is to infer the best parameters for the underlying phenomenon; it is not to obtain a best fit to unreliable experimental statistics. Secondary information in the form of horizontal wells, seismic data, conceptual geological models, and analogue data must be considered. In all cases, however, expert judgment is needed to integrate global information from analogue data with sparse local data (Gringarten & Deutsch 1999; Kupfersberger & Deutsch 1999).

Indeed, a variogram model is not simply a mathematical convenience required by kriging-based geostatistical algorithms. The variogram is a measure of spatial correlation and informs specific aspects of geological deposition, which may have a direct impact on fluid flow. Variogram "fitting" should only come after a serious variogram "interpretation" stage. That interpretation must relate specific underlying geological phenomena to characteristic shapes observed on experimental variogram points. The interpretation methodology consists of (i) identifying the global variance of the property studied; it is the expected plateau value (sill) of the variogram model; (ii) partitioning this variance into variance regions which explain variability at different length scales; these regions are identified by specific signatures of the experimental variogram (the reader is referred to Gringarten & Deutsch 1999 for a detailed explanation of variogram interpretation and modelling).

The most important signatures of a variogram are (i) the nugget effect (unique in all directions and characterized by a discontinuity at the origin); (ii) geometric anisotropy (when different directional variograms reach a similar variance contribution at different ranges); (iii) zonal anisotropy (where the variogram only exists in one direction for a given variance region). Two other signatures are due to trends (when the variogram systematically increases above the theoretical sill) and geological cyclicity (characterised by undulation on the variogram, the so-called hole-effect).

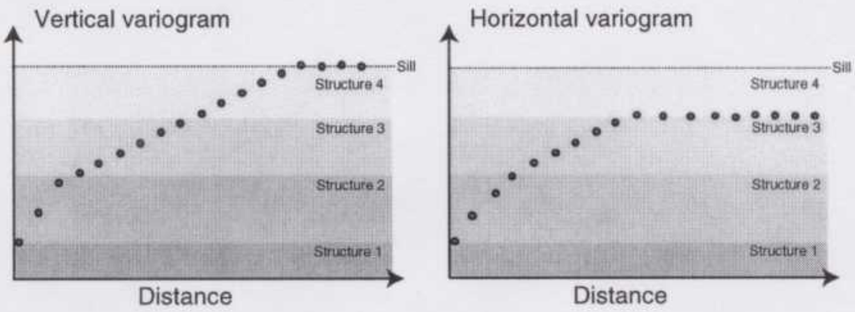



*Figure 3: Schematic of variogram interpretation principles where signatures of the experimental variograms are used to partition the total variability of the phenomena.*

Every variance region identified on the vertical variogram must exist on the horizontal variogram. The nugget effect is first, which is clear on the vertical variogram, then multiple geometric anisotropy regions, and finally there may be a zonal anisotropy for the large-scale variability region (see Figure 3). If data are too sparse to calculate ~~enough points for the~~a reliable experimental variogram, ~~T~~there are two steps to synthesize a horizontal variogram:

1. Establish or infer the horizontal-to-vertical anisotropy ratio for the geometric anisotropy structures. Conceptual geology, analogues and possibly secondary data are used for this purpose (Kupfersberger & Deutsch 1999). Typical horizontal to vertical anisotropy ratios for correlation ranges are given in this paper. The ratio varies from 10:1 in discontinuous point-bar environments to 500:1 in continuous platform carbonates.

2. If the vertical variogram flattens off at a value smaller than the theoretical sill, then the horizontal range for that variance region must be inferred as in step 1. T~~, t~~his behavior is typical of areal trends. Another type of zonal anisotropy that is caused by persistent stratification presents ~~But~~ the challenge ~~is to~~of determining~~e~~ what fraction of the variance is explained by this zonal anisotropy.~~ due to stratification that leads to persistent positive correlation in the horizontal direction.~~

The resulting synthetic horizontal variogram consists of the structures visible on the vertical variogram plus possibly a zonal anisotropy (step 2).

## Uncertainty Modeling

Multiple geostatistical realizations provide an assessment of uncertainty. Each realization is a Monte Carlo sample from the space of uncertainty defined by all decisions implicit to the modeling approach. There is no objective or correct space of uncertainty. The space of uncertainty created by multiple realizations is realistic when the conceptual geological framework and statistical parameters, such as the histogram and variogram are well known. A source of concern is that these parameters are *not* well known early in the lifecycle of a reservoir; therefore, there is more uncertainty than measured by a set of geostatistical realizations generated with the same set of underlying parameters.

A more realistic space of uncertainty is determined by a combination of the scenario-based approach and conventional geostatistical modeling. Different scenarios are defined that have alternate histograms and variograms. The probability of each scenario is determined by recursive application of Bayes Law; multiple realizations are generated for each scenario to model the uncertainty of incomplete knowledge given the scenario.

A reasonable definition of scenarios and assignment of conditional probabilities is critical. The scenarios can reflect different aspects of uncertainty, e.g., depositional style (estuarine channels versus tidal dominated), fault seal (sealing versus partial sealing), level of fracturing, and facies definition. The scenarios could also reflect uncertainty in critical statistical parameters such as the net-to-gross ratio (low, medium, and high) or the horizontal variogram range. There can be many levels to the scenario tree and a different number of levels down each branch of the tree. There could also be a different number of geostatistical realizations for each scenario.

The large-scale discrete aspects of uncertainty are quantified with the scenario-based approach. Uncertainty due to incomplete data is quantified with multiple geostatistical realizations. Reservoir uncertainty addressed with geostatistical realizations *alone* is reasonable late in the reservoir lifecycle where there is little large-scale uncertainty; small-scale models of heterogeneity are adequate to reflect uncertainty.

## Conclusions

A representative target histogram is critical for geostatistical modeling. Establishing a calibration relationship to soft data is the basis for declustering in presence of sparse data, and the maximum amount of information must be used to build this model. Both the actual data, and additional knowledge and experience are needed to construct the model. Notwithstanding our dependence on this relationship, using a poorly known calibration is better than ignoring an important sampling bias. This paper also discusses the need for a reliable variogram model and hints at ways to obtain realistic variogram parameters from ancillary information in the presence of poor experimental horizontal variogram data. The use of a systematic interpretation methodology aimed to identify variance regions at different scales on the vertical variogram and projecting them onto the horizontal variogram is the first step toward a consistent 3D variogram model. Geological understanding and analogues yielding vertical to horizontal anisotropy ratios make it a reliable model. Furthermore, due to the inference nature of input parameters, it is good practice to perform uncertainty modelling through a scenario-based approach of histogram and variogram parameters - and in general of key yet uncertain modelling parameters.

## References

Deutsch, C.V. 1989: DECLUS: A Fortran 77 Program for Determining Optimum Spatial Declustering Weights. Computers & Geosciences 15(3), 325-332.

Frykman, P. & Deutsch, C.V. 1998: Model-based Declustering for Deriving Representative Distributions Prior to Conditional Simulations. Second IMA Conference on Modelling Permeable Rocks, Churchhill College, Cambridge, March 23-25, 1998, 1 pp.

Gringarten, E. & Deutsch, C.V. 1999: Methodology for Variogram Interpretation and Modeling for Improved Reservoir Characterization. SPE Annual Technical Conference and Exhibition, Houston, Texas, 3-6 October 1999, (SPE 56654).

Isaaks, E.H. & Srivastava, R.M. 1989: An Introduction to Applied Geostatistics, 561 pp. New York: Oxford University Press.

Journel, A.G. 1983: Non-parametric estimation of spatial distributions. Mathematical Geology 15(3), 445-468.

Kupfersberger, H. & Deutsch, C.V. 1999: Methodology for integrating analog geologic data in 3-D variogram modeling. American Association of Petroleum Geologists Bulletin 83(8), 1262-1278.