

GEOSTATISTIC IN RESERVOIR CHARACTERIZATION: FROM ESTIMATION TO SIMULATION METHODS

H. Mata Lima*

ABSTRACT

In this article objective have been made to reviews different geostatistical methods available to estimate and simulate petrophysical properties (porosity and permeability) of the reservoir.

Different geostatistical techniques that allow the combination of hard and soft data are taken into account and one refers the main reason to use the geostatistical simulation rather than estimation. Uncertainty in reservoir characterization due to variogram assumption, which is a strict mathematical equation and can leads to serious simplification on description of the natural processes or phenomena under consideration, is treated here. Multiple-point geostatistics methods based on the concept of training images, suggested by Strebelle (2000) and Caers (2003) owing to variogram limitation to capture complex heterogeneity, is another subject presented.

This article intends to provide a review of geostatistical methods to serve the interest of students and researchers.

Key words: *Petrophysical properties, Geostatistic, Estimation, Simulation, Uncertainty.*

RESUMEN

Este artículo presenta una revisión de diversos métodos geoestadísticos disponibles para estimar y para simular características petrofísicas (porosidad y permeabilidad) de la formación geológica (roca depósito del petróleo).

Se presentan diversas técnicas geoestadísticas que permiten la combinación de datos *hard* y *soft* y se explica la razón principal para utilizar la simulación geoestadística en vez de estimación. También se explica la incertidumbre en la caracterización del depósito debido a la asunción del variogram. El hecho de que el variogram sea una simple ecuación matemática conduce a la simplificación seria en la descripción de los procesos o de los fenómenos naturales bajo consideración. Los «métodos geoestadísticos del Multiple-point» (*Multiple-point geostatistics methods*) basados en el concepto de *training images*, sugerido por Strebelle (2000) y Caers (2003), debido a la limitación del variogram para capturar heterogeneidad compleja es otro tema presentado.

Este artículo se propone proporcionar una revisión de métodos geoestadísticos que sean de interés para estudiantes e investigadores.

Palabras claves: *Características petrofísicas, geoestadística, estimación, simulación, incertidumbre.*

Introduction

The spatial distribution of physical and hydraulic properties in natural materials is difficult to predict deterministically. Scarcity related to sampling programs is another factor that complicates the prediction of the subsurface properties. Hence, prediction of the spatial occurrence of rock properties via stochastic process ought to be done for effective numerical

modelling of physical and hydraulic processes that act on heterogeneous properties.

Geostatistics is a branch of applied statistics developed by George Matheron of the Centre de Morphologie Mathématique in Fontainebleau, France (Dorsel and Breche, 1996) and its original purpose was centered on estimating changes in ore grade within a mine. Geostatistics uses the *regionalized* variables; a *regionalized* variable varies in a

* Technical University of Lisbon. CMRP. Av. Rovisco Pais, 1. 1049-001 Lisboa. Portugal. E-mail: helima@ist.utl.pt

continuous manner spatially so that data values from points nearer each other are more correlated. *Regionalized variables* describe phenomena with geographical distribution (e.g. distance downwards, elevation of ground surface, etc.). The phenomenon exhibit spatial continuity; however, it is not always possible to sample every location. Therefore, unknown values must be estimated from data taken at specific locations that can be sampled. The size, shape, orientation, and spatial arrangement of the sample locations are termed the support and influence the capability to predict the unknown samples. If any of these characteristics change, then the unknown values will change. The sampling and estimating of *regionalized variables* are done so that a pattern of variation in the phenomenon under investigation can be mapped such as a contour map for a geographical region. Therefore, contrary to traditional statistics where independence between samples is generally assumed, geostatistics takes advantage of the fact that samples located in proximity to one another are often more similar than those obtained at large separation distances, i.e., it provides a means of quantifying this spatial correlation in material properties and then exploiting that information for use in both interpolation and stochastic simulation techniques.

Samples data are interpreted as from a random process, in geostatistical models (Soares, 2000). According to the same author, a stochastic process is only, in model context, a way to analyses data from experimental samples. Hence, the first step of geostatistical modeling is to concept a random process that matches the experimental data or observations.

The geostatistic allows the quantification of natural resources characteristics or spatial phenomena (e.g. mining resource). Therefore, one must have models that allow deduce the magnitude of phenomena in the not sampled space from available information (Soares, 2000).

In order to perform any type of geostatistical estimation or simulation one requires an assumption of stationarity. Geostatistic is completely imbibed by the principle of stationarity. To determine the histogram of the petrophysical property (porosity, permeability) of a reservoir one relies on stationarity in the sense that information from various wells are pooled together into one single histogram. Implicitly, it is assumed that, in terms of a histogram, all the property values in that histogram originate from a single «population». Both histogram and variogram (based on the assumptions of stationarity) are used for estimation and simulation purpose throughout the reservoir, frequently in interwell region where a few a-priori information is present (Caers and Zhang, 2002).

Different reservoir rocks shows differences in pores/fracture dimension and connection, (...) which can change with depth and areal location. Hence, one must understand as well as possible the distribution of reservoir properties, i.e. reservoir characterization, in order to enhance oil recovery.

In the last years geostatistical methods have been used for stochastic characterization of the reservoir rocks by generating multiple reservoir models constrained to geologic, seismic and production data (Caers and Zhang, 2002). According to the same authors geostatistic intends to achieve the following three objectives: i) provide reservoir models that depict a certain believed or interpreted geological heterogeneity; ii) provide a quantification of uncertainty through multiple reservoir models, all honoring that same geological heterogeneity; and iii) integrate various types of data, each type bringing information on possibly different scales and with different precision.

This section presents several geostatistical techniques used to estimate and simulate the porosity and permeability of the reservoir rocks.

It is explained how log and core porosity/permeability measurements and soft information on porosity/permeability can be used in the simulation of spatial variable realizations.

Reasons to use geostatistics

Geostatistical techniques is considered suitable to study spatially distributed phenomena due to their capability to quantify the uncertainty in the derived estimates (Gandin, 1970; Matheron, 1970; Delhomme, 1978; de Marsily, 1986).

Geostatistics allows a good estimate of property values between data points – interwell properties. The conjugation of these property values (data and estimated) is a property grid. Geostatistics methods, such as, *kriging and conditional simulation*, make accurate grids and can combine, by *cokriging and cosimulation*, many different types of *hard* and *soft data*, and is able to quantify the uncertainty in a reservoir description. Geostatistical should be viewed as a tool to integrate data (e.g., geological interpretation, core and log data, seismic data and production data). Table 1 presents some examples of *hard* and *soft data* (*soft data* must be calibrated to the *hard data*).

Table 1.—Examples of *Hard* and *Soft data*

<i>Hard Data</i>	<i>Soft Data</i>
Lithofacies assignments from core	Seismic data
Core porosity	Well log porosity
Core permeability	Well log permeability

One use grid to characterize a reservoir because it is the form of input data required for reservoir flow simulator and others software. Usually grid corresponds to rock properties but, for instance, reservoir fluid can be used too. This situation avoids the problem due to different data structures (data inconsistency) that usually require a great number of programs and procedures to do simply data conversion.

Geostatistical methods provide an appreciable accuracy of forecasting based on the reservoir model because they produce grids which represents reservoir heterogeneity; can quantitatively combine many different types of hard and soft data; and, can quantify the uncertainty in a reservoir description. Although, geostatistical quantification and formalization of physical concepts relies on strict mathematical rules and equations which leads to serious simplification on description of the natural processes or phenomena under investigation (Caers and Zhang, 2002).

Geostatistical estimation and simulation

In this section one presents a brief description of basic geostatistical theory and the methods of geostatistical interpolation and simulation that are commonly used for reservoir modelling purpose.

Geostatistical estimation, kriging, is used to obtain the unbiased estimates with minimum variance. The kriging methods are in general not so interesting for reservoir study because the use of kriged fields as input parameters to the reservoir flow equation yields biased estimates of fluid flow velocities. More interesting are geostatistical simulation methods that give multiple equally likely realizations of the attribute at any location in the domain and an «unbiased» (it is not really unbiased but less biased or more close to field condition) estimate of the reservoir flow velocities may be obtained.

All geostatistical methods are based on stationarity principles, since different measurement data within the study domain are pooled together in order to estimate the statistics. The mean is estimated by calculating the arithmetical average in case no preferential sampling was done. However, it is very common the use of a preferential sampling scheme. This may be due to physical constraints or because a zone of study domain is of special interest. In case of a preferential sampling scheme the cell declustering technique (Deutsch and Journel, 1998) can be used to estimate the mean.

The covariance function is calculated by plotting the separation distance of the all pairs of measurement data versus the variance for those pairs of mea-

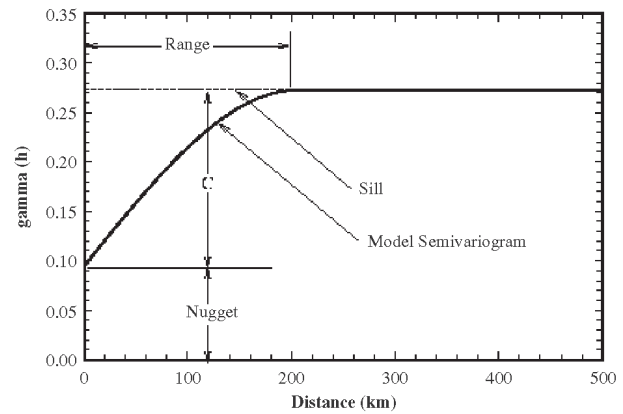


Fig. 1.—Features of a semivariogram and parameters defining the search area (after Englund and Sparks, 1988).

surement data. The result is a data set with for each of the comparison pairs a covariance and the separation distance. However, in practice, the variogram is normally used to model the spatial correlation.

The semivariogram (see figure 1) is estimated by grouping pairs of measurement data in distance classes. The pairs in one distance class have a similar, but not necessarily equal, separation distance. The experimental semivariogram $[\gamma^*(h)]$ is defined as:

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (1)$$

where: h - is a directional vector; $\gamma^*(h)$ - is the experimental semivariogram for a vector h ; $N(h)$ - is the number of pairs of data locations a vector h ; $z(x_i)$ - is the value for attribute z at location x_i , and $z(x_i + h)$ - is the value for attribute z at a vector h from location x_i .

By the way, it is important to refer that the definition of the variogram does not require the existence of a constant mean and finite variance for the random function. In this case a sufficient condition is that the random function increments $Z(x) - Z(x+h)$ (h being the separation distance) are stationary of order two (Goovaerts, 1997). This condition is also called the intrinsic hypothesis (Journel and Huijbregts, 1992).

Normally the semivariogram is estimated by allowing a tolerance of Δh around the separation distance h , in order to get enough comparison pairs in each class. After estimating the experimental semivariances for different distance classes a model variogram has to be obtained because semivariances for any possible separation vector h will be needed. Not any function can be used as a variogram model: the positive definite condition should hold. Some

models are commonly used because they are known to be positive definite.

The *Spherical* model:

$$\gamma(h) = \begin{cases} c_o + c \left[1.5 \frac{h}{a} - 0.5 \left(\frac{h}{a} \right)^3 \right] & h \leq a \\ c_o + c & h > a \end{cases} \quad (2)$$

where: c_o is the nugget effect, c the sill minus the nugget and a the range. Others parameters have the same meaning referenced before.

The *Exponential* model:

$$\gamma(h) = c_o + c \left[1 - \exp\left(\frac{-3h}{a}\right) \right] \quad (3)$$

The *Gaussian* model:

$$\gamma(h) = c_o + c \left[1 - \exp\left(\frac{-3h^2}{a^2}\right) \right] \quad (4)$$

The *Power* model:

$$\gamma(h) = c_o + c(h^\omega) \quad 0 < \omega < 2 \quad (5)$$

Since the positive definite condition holds for any linear combination of the models above, a variogram model can be a combination of two or more of the specified models, each model having different values for its parameters.

The models presented above (equation 2 to 5) are for isotropic variogram models. But, one knows that in reality we are facing with the anisotropy, i.e. the change of the pattern of spatial variability with orientation. Anisotropy can be distinguished between *geometric* and *zonal* anisotropy (see, for instance, Hendricks Franssen, 2000; Soares, 2000; Pebesma, 2001).

Geometric anisotropy: directional semivariogram have same shapes and sill values, but different range values.

Zonal anisotropy: means that the sill value is a function of the direction. There is situation in which both sill and range values are a function of direction.

This category of *anisotropy* is typical in stratified phenomenon whose the spatial continuity through the layer is widely different from the variability between layers (Hendricks Franssen, 2000).

The definition of the variogram is quite important since the result of stochastic simulation is based on it. According to Wingle (1997) and Ortiz and Deutsch (2002) by evaluating the uncertainty in the variogram, the greater range of uncertainty associated with the simulated results becomes apparent. Wingle (1997) refers that the uncertainty in the simulation process can be more completely evaluated by using methods such as jackknifing, latin-hypercube sam-

pling, and expert opinion in defining the semivariogram models to be used for stochastic simulation.

Jackknifing: in this procedure the experimental semivariogram is calculated with one (or more) data point(s) removed from the data set. By repeating this procedure for every point in the data set, a series of n (n = number of samples) experimental semivariograms is calculated. For each lag distance there are now n $\gamma^*(h)$ values. Using these values, confidence limits can be approximately determined, for the mean $\gamma^*(h)$ at a particular distance. When these are plotted, the error bars define the possible range of the modeled variogram (given a specific confidence level – 95% is usually used). This technique intend to help the modeler in optimizing further data collection or identifying a likely range of reasonable model variograms.

Latin-Hypercube: once statistical distribution of the experimental semivariograms has been calculated, variograms can be fit through the zone defined by the error-bars. The objective is not to make a single best estimate of the character of the subsurface (i.e. a single variogram), rather the objective is to select model semivariograms representative of the range of possible conditions at the site. This range of variograms should be used with the original data to conduct kriging and stochastic simulation to generate multiple interpretations of the subsurface.

One is convinced that *expert opinion* is more useful to contribute to the definition of an accurate variogram model. Indeed, another methodologies follow mathematical rules and equations which obviously simplify the reality. According to Soares (2000), the process of fitting experimental semivariogram by a theoretical model is widely constraint to the expert opinion of the modeler about the spatial phenomenon under studying.

Cautionary notes about semivariogram definition

After the calculation of the semivariogram values for different h , one needs to fit the variogram which represents the total area. This is one of the most important (or most important) geostatistical steps because it summarizes the structural characteristics (dispersion or continuity, anisotropy, nested structures) of the spatial phenomenon in a single variogram model (Soares, 2000).

«Variogram modelling is a critical step in any geostatistical study; however, a reliable variogram is difficult to infer in presence of sparse data».

Ortiz and Deutsch, 2002

The process of defining the variogram is decisive and requires that the modeler be conscious about the following:

Wingle (1997)

— There is not a method to directly measure the uncertainty, error, or confidence limits associated with an experimental semivariogram;

— $\gamma^*(h)$ is calculated as the mean of squared differences for a given distance. Therefore, it is a variance of the data for that distance and not really a mean;

— the variance about $\gamma^*(h)$ increases with separation distance;

— the expert of the modeler is determinant to reduce uncertainty in the variogram definition;

— the variogram cannot reproduce with complete accuracy the field heterogeneity since it is a statistical tool (guided by strict mathematical rules and equations) which unavoidably cause serious simplification on description of the natural processes or phenomenon (Caers and Zhang, 2002);

— outliers can perturb the interpretation of data because they distort the variogram shape. So it should be removed if the calculated variogram without outlier match, in a better way, others values of $\gamma^*(h)$ (Soares, 2000). But it should not be removed without further justification (for more about outliers see Iglewicz and Hoaglin, 1993 and Belsley *et al.*, 2004).

Geostatistical estimation

The first geostatistical step, described above, is to obtain the parameters of a given random function. The next step corresponds to the use of the random function model to estimate (or interpolate) the attribute (e.g. porosity, permeability) of interest at locations where there are not measurements.

The variograms explained above is used to obtain model covariance: i) for any pair of locations; ii) between measurement points; and, iii) between a measurement point and a location where one want to estimate the value of the attribute.

There is several form of kriging developed for estimation purpose. The estimation algorithms most commonly used are the following:

Continuous variables	Categorical variables
Simple kriging (SK)*.	Indicator kriging (with or not soft data).
Ordinary kriging (OK)*.	Maximum probability classification.
Universal kriging (UK)*.	Rescaled maximum probability classification.
Ordinary Kriging with external drift (KED)*.	Dynamic classification.
Cokriging and collocated cokriging*.	
Indicator kriging*.	
Indicator collocated cokriging.	

* These methods are described, below, in this study.

Simple kriging (SK)

The SK is the most vulgar form of kriging. One assumes that the mean of a set of random variable related to sampled values and non-sampled points are known. This algorithm is applied whether one must know the mean of random function or good information about the trend of phenomenon is available (Soares, 2000).

The simple kriging unbiased estimator with minimum variance for the value of an attribute at a certain location is given by:

$$[Z_{SK}(x)]^* = \sum_{i=1}^{N(x)} \lambda_i^{SK}(x)[Z(x_i) - m] + m \quad (6)$$

where: $N(x)$ - is the number of measurement data; $\lambda_i^{SK}(x)$ - the assigned kriging weight to each of the observation data. The following factors could be considered in assigning the weights: closeness of the data to the location being estimated, clustering of the data locations, anisotropic continuity (preferential direction) and magnitude of continuity/variability; $Z(x_i)$ - the data value at the i -th observation location; and m - the stationary constant mean. The term $\lambda_i^{SK}(x)$ is obtained by solving the *simple kriging* system.

The variance of the SK is expressed as follows:

$$\sigma_{SK}^2(x) = C(0) - \sum_{i=1}^{N(x)} \lambda_i^{SK}(x)C(x_i - x) \quad (7)$$

where: C - is the model spatial covariances; and $C(0)$ - is the spatial covariance for separation distance zero.

Ordinary kriging (OK)

Ordinary kriging (also called *punctual kriging*) «allows to accounts for local variations in the mean by limiting the domain of stationarity to the local neighbourhood centred on the location where an estimation of the attribute value is required». Due to this particularity, OK is more robust to non-stationarity (Hendricks Franssen, 2000). The ordinary kriging estimate is:

$$[Z_{OK}(x)]^* = \sum_{i=1}^{N(x)} \lambda_i^{OK}(x)Z(x_i) \quad (8)$$

The unknown local mean does not appear in the expression because it is filtered out by forcing the kriging weights to sum to one:

$$\sum_{i=1}^{N(x)} \lambda_i^{OK}(x) = 1 \quad (9)$$

From the following system of equations the kriging weights are calculated:

$$\begin{bmatrix} C(x_1 - x_1) & \dots & C(x_1 - x_{N(x)}) & 1 \\ \vdots & \ddots & \vdots & 1 \\ C(x_{N(x)} - x_1) & \dots & C(x_{N(x)} - x_{N(x)}) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1^{OK}(x) \\ \vdots \\ \lambda_{N(x)}^{OK}(x) \\ \mu_{OK}(x) \end{bmatrix} = \begin{bmatrix} C(x_1 - x_1) \\ \vdots \\ C(x_{N(x)} - x_{N(x)}) \\ 1 \end{bmatrix} \quad (10)$$

where $\mu(x)$ is the Lagrange multiplier which is introduced to guarantee that the kriging weights sum to one under the condition that an unbiased estimate with minimum variance is obtained.

The variance is calculated as follows:

$$\sigma_{OK}^2(x) = C(0) - \sum_{i=1}^{N(x)} \lambda_i^{OK}(x) C(x_i - x) - \mu_{OK}(x) \quad (11)$$

Universal kriging (UK)

In some cases it may be inappropriate to consider the mean even constant in a local neighbourhood. In such a case the local mean in a neighbourhood is modelled as a function of the coordinates. The *universal kriging* is similar to that of *ordinary kriging*, but used when a trend, or slow change in average values, in the samples exists.

The UK is a kriging with trend estimator and is given by:

$$[Z_{UK}(x)]^* = \sum_{i=1}^{N(x)} \lambda_i^{UK}(x) Z(x_i) \quad (12a)$$

with the following constraints:

$$\sum_{i=1}^{N(x)} \lambda_i^{UK}(x) f_k(x_i) = f_k(x), \quad k = 0, \dots, K \quad (12b)$$

where: (x) - is the location where an estimation is required; (x_i) - is one of the N measurement locations; f_k - are known trend function; $f_k(x_i)$ - is the function value at a measurement location; and $f_k(x)$ - is the function value at a location (grid cell) where an estimation is required.

In order to make uncomplicated the demonstration of the UK, let the variable $Z(x)$ be equal to the sum between a mean $m(x)$ and a residual $R(x)$ with a null mean and covariance C_R :

$$Z(x) = m(x) + R(x) \quad (13)$$

For the UK the variance is given as follows:

$$\sigma_{UK}^2(x) = C_R(0) - \sum_{i=1}^{N(x)} \lambda_i^{UK}(x) C_R(x_i - x) \quad (14)$$

In some cases one can consider the trend $m(x)$ as a linear function of an external auxiliary variable $Y(x)$ (Marechal, 1984; in Soares, 2000). This process is called *Kriging with external drift*, which can be seen as a branch of *universal kriging*, and accounts for more complex trends.

Cokriging

When, beyond the principal variable Z_1 , one has a secondary variable Z_2 (covering the field) strongly

correlated with the variable Z_1 , estimation process by kriging ought to take account of secondary information wherever possible (Goovaerts, 1997; Soares, 2000).

To estimate the variable $Z_1(x)$ in an unsampled location, the values of both variable [$Z_1(x_i)$ and $Z_2(x_j)$] are linearly combined as follows:

$$[Z_{CK}(x)]^* = \sum_{i=1}^{N_1(x)} a_i Z_1(x_i) + \sum_{j=1}^{N_2(x)} b_j Z_2(x_j) \quad (15a)$$

where: $N_1(x)$ - is the number of data for the principal variable Z_1 ; $N_2(x)$ - is the number of data for the secondary variable Z_2 ; $Z_1(x_i)$ and $Z_2(x_j)$ - are respectively the value of the principal variable at the i -th measurement location and the value of the secondary variable at the j -th measurement location; and a_i and b_j - are the cokriging weights.

In order to guarantee an unbiased estimator, the following constraints are imposed to the weights:

$$\sum_{i=1}^{N_1(x)} a_i = 1 \quad \text{and} \quad \sum_{j=1}^{N_2(x)} b_j = 0 \quad (15b)$$

The cokriging weights are obtained by solving the following system:

$$\begin{bmatrix} M_{ii} & M_{ij} & 1_i & 0_i \\ M_{ji} & M_{jj} & 0_j & 1_j \\ 1_i^T & 0_j^T & 0 & 0 \\ 0_i^T & 1_j^T & 0 & 0 \end{bmatrix} \begin{bmatrix} a_i \\ a_j \\ b_i \\ b_j \end{bmatrix} = \begin{bmatrix} [C_{ii}(x_i - x)]^T \\ [C_{ji}(x_j - x)]^T \\ 1 \\ 0 \end{bmatrix} \quad (15c)$$

where: M_{ii} , M_{ji} , M_{ij} and M_{jj} are the matrices with the covariances and cross-covariances for the principal and secondary variable; $C_{ii}(x_i - x)$ and $C_{ji}(x_j - x)$ are covariances and cross-covariances between the data and the location where the estimate is required; 1_i , 1_j , 0_i and 0_j are unit and null vectors and the superscript T means transpose.

The variance is expressed as:

$$\sigma_{CK}^2(x) = C_{ii}(0) - b_i - \sum_{i=1}^{N_1(x)} a_i C_{ii}(x_i - x) - \sum_{j=1}^{N_2(x)} a_j C_{ji}(x_i - x) \quad (16)$$

Indicator kriging

Indicator kriging (Journel, 1983) differs from simple or ordinary kriging in that a range of parameter values are reduced to discrete indicators (integer values) by defining threshold values. Indicator description makes it possible to «krige» qualitative parameters such as lithology which, for instance, could be defined as indicator 1 for silt, indicator 2 for silty-sand, and indicator 3 for fine sand.

In this technique the measurement data re transformed into indicator data by the following transform:

$$i_z(x) = \begin{cases} 1 & Z(x) < z \\ 0 & Z(x) \geq z \end{cases} \quad (17)$$

where: i_z - is the indicator transform and z the threshold value (which represent the values of especial interest like an established critical porosity/permeability).

The ordinary indicator kriging estimate is:

$$I(x; z) = \sum_{j=1}^{N(x)} \lambda_j^K(x; z) I_j(x; z) \quad \text{with} \quad \sum_{j=1}^{N(x)} \lambda_j^K(x; z) = 1 \quad (18)$$

where I_z is the indicator data and z the threshold.

The indicator kriging system that gives the indicator kriging weights is expressed by:

$$\begin{bmatrix} C_1(x_1 - x_1; z) & \dots & C_1(x_1 - x_{N(x)}; z) & 1 \\ \vdots & \ddots & \vdots & 1 \\ C_1(x_{N(x)} - x_1; z) & \dots & C_1(x_{N(x)} - x_{N(x)}; z) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1^K(x; z) \\ \vdots \\ \lambda_{N(x)}^K(x; z) \\ \mu^K(x; z) \end{bmatrix} = \begin{bmatrix} C_1(x_1 - x; z) \\ \vdots \\ C_1(x_{N(x)} - x; z) \\ 1 \end{bmatrix} \quad (19)$$

In the next section one presents the simulation techniques that ought to be used instead the interpolation algorithms since its produce more accurate results for spatial variable.

For a full explanation of the methods, briefly described above, see Olea (1974); Journel and Huijbregts (1992); Davis (1986, 1987); Goovaerts (1997); Wingle (1997); Deutsch and Journel (1998); and Soares (2000).

Geostatistical simulation

In the first place it is important to stress that interpolations process explained above do not give accurate results as the simulation of multiple equally likely realizations. The reason is because the interpolation methods do not represent the «realistic» maps due to smoothing. Estimation is locally accurate and smooth, appropriate for visualizing trends, inappropriate for flow simulation where extreme values are important, and does not assess of global uncertainty; while simulation reproduces histogram, honors spatial variability (variogram), allows an assessment of uncertainty with alternative realizations possible, and is appropriate for flow simulation.

The main goal of simulation in geostatistics is to produce «realistic» maps of a phenomenon rather than minimizing the prediction error, which usually leads to smooth maps, not really representative of the real field. As referred by Soares (2002), simulation model for spatial phenomenon intend to generate images which characterize the resource by reprodu-

cing the proportion and high or low spatial continuity of different structures, heterogeneities and extreme classes of related histograms. Reproduces the histogram of an original variable is an obligation of any simulation approach. The simulation is performed to reproduces the variability of the phenomenon through the distribution function of $Z(x) - F_z(z) = \text{prob}\{Z(x) < z\}$ and the variogram $\gamma(h)$.

Continuous variables	Categorical variables
Sequential Gaussian Simulation (SGS)*.	Sequential Indicator Simulation.
Direct Sequential Simulation and Cosimulation (Co-DSS)*.	Simulated Annealing.
Probability Field Gaussian Simulation .	Post Processing of Simulation using Annealing.
Sequential Indicator Simulation (SIS).	Truncated Gaussian with or without a linear trend.
Simulate Annealing.	Categorical to continuous (facies to physical property) using SGS.
Cloud transform.	Categorical to continuous using cloud transform.

Examples: i) continuous variables: lithotypes, soil types, forest species, etc.; ii) Categorical variables: grades, porosity, etc.

Sequential simulation is a stochastic algorithm widely used in geostatistical to reproduce spatial distribution and uncertainty of variables of different natural resources, in part, owing to its simple implementation (Soares, 2001).

Sequential Gaussian Simulation (SGS)

The *Sequential Gaussian Simulation* (SGS) is widely used in geostatistics (Deutsch and Journel, 1998). To produce «realistic» maps when data are normally distributed, one estimates the parameters of the local probability density function at each location of the test set, and randomly generates a value from this distribution.

The first part of a SGS is to check if the known data are normally distributed, and if not, to apply normal-score transformation on them (Deutsch and Journel, 1998). Then, one calculates the variogram of the transformed data according to equation 1. The next part is the simulation process itself.

1. One chooses at random a point to estimate inside the unknown data set.

2. Apply the kriging procedure, using the known data set and the modelised variogram, to estimate the mean μ_{krig} and variance σ_{krig}^2 of this value. These two parameters are then considered respectively as the mean and variance of the local Gaussian probability density function of the point.

3. One randomly chooses a value for the unknown point, following the law $N(\mu_{krig}, \sigma_{krig}^2)$.

4. The simulated point is then considered as a known point and will be used to simulate the next randomly chose unknown point.

5. Repeat from step 1, until there are no more unknown points.

Gilardi *et al.* (2001), refer that «the result of one simulation is thus a «noisy» version of a kriging procedure, which reproduces the statistical histogram and the variogram of the known data, giving a more realistic aspect to the output but a lower prediction performance». However, when performing multiple sequences of simulation, one is able to draw reliable probability maps.

Confidence interval calculation

As the simulated data are normally distributed, it is also easy to estimate the confidence interval of a prediction for each data point. The procedure used by Gilardi *et al.* (2001), is the following:

- Compute the mean μ_{sim} and standard deviation σ_{sim} of the simulation of each data point;
- Define the 95% confidence interval bounds as the 2.5% and 97% quantils of the cumulative density function of the normal law defined by $N(\mu_{sim}, \sigma_{krig})$. When these bounds have been calculated, one is applying the N-Score back-transformation both on them and on the mean, in order to get back the real output space.

Procedure of SGS experiments

One has to simulate a test set using the information given by a train set. The SGS 95% confidence interval calculation may be done like this:

1. Normal score transformation of the train outputs if necessary;
2. Variogram modelling using the transformed outputs of the test set;
3. 100 or more sequential Gaussian simulation procedures of the set;
4. Evaluation of the confidence intervals as described above (confidence interval calculation);
5. Back-transformation of the confidence intervals bounds if necessary.

For the SGS the Gaussian hypothesis imposes a normal-score transformation and back-transformation of data if they are not yet normal. Such transformation seems to cause artifacts in solutions as observed by Gilardi *et al.* (2001). The consequence

of mathematical simplicity is the characteristic of maximum spatial entropy, i.e., low and high values are disconnected. The SGS is recommended for porosity and not appropriate for permeability.

For permeability simulation the simulate annealing is becoming very popular.

Direct Sequential Simulation and Cosimulation (DSS)

The *Direct Sequential Simulation and Cosimulation* (Co-DSS) was proposed by Soares (2001) and has been used in several study developed by Center for Petroleum Modelling of Technical University of Lisbon (CMRP/IST), concerning to simulation of petrophysical properties (e.g. porosity and permeability), with satisfactory results.

Sounds like the DSS intend to substitute the *Sequential Indicator Simulation* (SIS) and *Sequential Gaussian Simulation* (SGS) since it does not require the transformation of continuous variable into a binary or a Gaussian variable. As refers Soares (2001), the goal «is to use the local simple kriging estimates of the mean and variance, not to define the local cumulative distribution function (*cdf*) but to sample from the global *cdf*. Simulated values of the original variable are drawn from intervals of the global *cdf*, which are calculated with the local estimates of the mean and variance». The DSS entails the co-simulation procedure in the absence of any transformation (e.g. into a set of indicator variables or a standard Gaussian variable) of the original variables.

This method is fully described in Soares (2000, 2001), hence it is not extensively treated in this work.

The method (Co-DSS) proposed by Soares (2001) is based in the following principle:

If the local *cdf*'s are centered at the simple kriging estimate

$$Z(x_u)^* = m + \sum_{\alpha} \lambda_{\alpha}(x_u)[z(x_{\alpha}) - m] \quad (20)$$

x_{α} being the conditioning data (original and previously simulated values), with a conditional variance identified with the simple kriging variance $\sigma_{sk}^2(x_u)$; no matter what probability distribution we choose, the spatial covariance model or variograms are reproduced in the final simulated maps.

One of the main advantages of this algorithm (DSS) over traditional SIS and SGS is that it allows a joint simulation dealing directly with the original variables (Soares, 2001). The DSS method is recommend both for porosity and permeability.

The results of any sequential simulation must be checks according to the following questions: i) honor

data?; ii) honor global proportions?; iii) honor variogram?; and, iv) honor concept of geology?

Final Remarks

Stochastic modeling of reservoir properties typically includes two steps: the geometry of facies is simulated first and then the spatial distribution of petrophysical variables (e.g., porosity and permeability) is simulated within each facies (Haldorsen and Damsleth, 1990). This two-step approach allows reproduce large scale heterogeneities as generate by facies boundaries and smaller scale variability within facies. In his article Azpirixaga *et al.* (2000) refers a more detailed description about the steps concerned to reservoir study: i) the first approach is to perform a calibration between seismic and well stratigraphy; ii) attribute calculation is done for each horizon; iii) geostatistical techniques are applied to build a quantitative relationship between rock quality and attributes. This relation is then incorporated into a *stochastic model* in order to characterize the reservoir and a fluid flow simulation is performed.

As referred before, there are several algorithms available to simulate *categorical variables*: Boolean algorithms (Ripley, 1987; Haldorsen *et al.*, 1988; Suro-Pérez, 1991), multiple truncations of a Gaussian field (Journel and Isaaks, 1984; Matheron *et al.*, 1987; Xu and Journel, 1993), sequential Gaussian simulation and indicator-based algorithms (Deutsch and Journel, 1998), simulate annealing (Farmer, 1991; Deutsch and Journel, 1994), and direct sequential co-simulation (Soares, 2001). Each algorithm generates a set of alternative, equally probable, categorical maps conditional to available information such as facies-type at sample locations, spatial continuity and transition probabilities among categories as deduced from a training image or a structural analysis of the data.

Although multiple realizations are necessary for the user to appreciate the uncertainty in the spatial distribution of facies, there is a tendency in retaining a single realization. This unique categorical map may be the realization that pleases best the geologist or, more often, it is the first realization generated (Goovaerts, 1994). The author proposes that, in such case, an alternative would consist of using a single estimated (rather than simulated) categorical map deduced from a classification of the grid nodes. Soares (1992) developed a classification algorithm that preferentially allocates nodes to the category with largest local probability of occurrence under the constraint of reproduction of global proportions. Because local probabilities are

established using kriging algorithm, Soares' classification is typically too smooth and does not reproduce (cross) variograms models (Goovaerts, 1994). To correct for this smoothing effect and reproduce better transition probabilities between categories Goovaerts (1994) proposes to post-process Soares' or any other estimated categorical map by a MAP (maximum a posteriori or steepest descent-type) algorithm with a three-components objective function.

Annealing allows integration of disparate sources of information (multiple-points statistics, connectivity functions) which cannot be readily taken into account by other simulation algorithms (simulation annealing is more often used to describe reservoir permeability). Indicator-based algorithms can provide local probabilities for any specific facies to prevail at any location; such local probabilities are convenient vehicles to integrate prior information such as trends in facies proportions or the absence of a facies in a particular area.

Goovaerts (1994), proposes two criteria for the incorporation of local probabilities into a simulated annealing objective function: i) maximization of the average local probability, and ii) reproduction of local class proportions. But this criterion was not described here because one believes that the direct sequential co-simulation is a very suitable method to simulate porosity and permeability with sufficient accuracy.

Well and seismic information give vertical and lateral characterization of the reservoir, respectively. Geostatistical techniques could be seen as a way to combine well and seismic data to produce best description of the reservoir. Boechmann (2000), refers that 3D seismic data is possibly the only data source for detecting reservoir heterogeneities in the interwell regions.

Todini (2001) refers that no account of the effects of uncertainty in parameters estimates is commonly made, wich may lead to three major inconsistencies: i) a bias in the point estimation of the multi-dimensional variable; ii) a different spatial distribution of the measure of uncertainty; and iii) an inappropriate model choice for the variogram.

Although geostatistic is able to evaluate property values between data points and produces grids which represent reservoir heterogeneity, it cannot reproduce with complete accuracy the field heterogeneity. Geostatistics techniques might allows a good reduction of the magnitude of error because they (kriging and conditional simulation) make accurate grids and can combine, by *cokriging and cosimulation*, many different types of *hard* and *soft* data, and is able to quantify the uncertainty in a reservoir description.

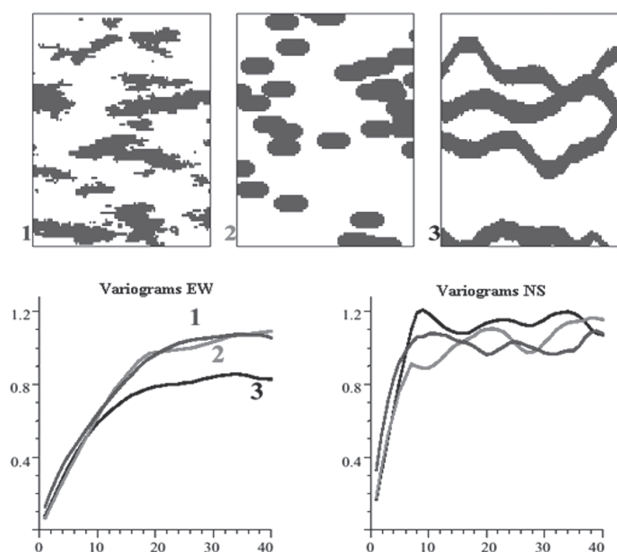


Fig. 2.—The variogram as a poor descriptor of geological heterogeneity. Three different geological heterogeneities result in three similar variograms (after Caers and Zhang, 2002).

Caers and Zhang (2002) brought up that the variogram is a statistical tool (guided by strict mathematical rules and equations) which unavoidably cause serious simplification on description of the natural processes or phenomenon and obviously might not reproduce the reality, since it is too limiting in capturing geological heterogeneity from outcrops. Hence, different types of reservoir heterogeneities may produce a similar experimental variogram (see figure 2).

Due to the limitation of the variogram to capture complex heterogeneity several authors (Strebelle, 2000, 2002, Strebelle and Journel, 2001, Caers and Zhang, 2002) have suggested a multiple-point geostatistics methods based on the concept of training images. However, the geostatistical assumption is preserved, i.e. when the information is not repetitive (stationary) it cannot be captured and reproduced by a model that relies on stationarity and because that some outcrop models cannot be directly used in multiple-point geostatistics as a reservoir analog or training image (Caers and Zhang, 2002).

In the other hand, Ortiz and Deutsch (2002) present an approach to calculate the uncertainty in the variogram and a methodology to transfer this uncertainty through geostatistical simulation and decision making.

ACKNOWLEDGMENTS

The author wants to thanks the FCT (Fundação para Ciência e Tecnologia) and GLOBALVIA – Consultores de Engenharia, SA for all help that has been provided to students and beginning researchers through the last years.

References

- Almeida, J., Renaldo, C., Luis, J., Schuhler, E. y Diet, J.P. (2000). Integration of seismic impedance in stochastic modelling of porosity and quality using Probabilistic Neural Networks and Sequential Direct Co-simulation. In: *Applied Reservoir Characterization Using Geostatistics – The Value of Spatial Modeling*. AAPG Hedberg Research Conference '2000, The Woodlands, Texas, 4 pp.
- Azpirixaga, I., Galli, A. y Jimenez, Z. (2000). Seismic Constraints in a Stochastic Reservoir Model and Flow Fluid Simulation, In: *Applied Reservoir Characterization Using Geostatistics – The Value of Spatial Modeling*. AAPG Hedberg Research Conference '2000, The Woodlands, Texas, 8 pp.
- Beier, R.A., Hardy, H.H., y Lewis, J.J.M. (2000). Geostatistical modeling of permeability in Aeolian Transverse Dunes. In: *Applied Reservoir Characterization Using Geostatistics – The Value of Spatial Modeling*. AAPG Hedberg Research Conference '2000, The Woodlands, Texas, 10 pp.
- Belsley, D. A., Kuh, E. y Welsch, R. E. (2004). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons, New York, 292 pp.
- Boechmann, M. (2000). Geostatistics vs. Neural Networks – Benefits and Pitfalls in 3D Property Modeling, In: *Applied Reservoir Characterization Using Geostatistics – The Value of Spatial Modeling*. AAPG Hedberg Research Conference '2000, The Woodlands, Texas, 4 pp.
- Caers, J. y Zhang, T. (2002). Multiple-point geostatistics: a quantitative vehicle for integrating geologic analogs into multiple reservoir models, In: *Integration of Outcrop and Modern Analog Data in Reservoir Models: AAPG Memoir* (G.M. Grammer et al., edit.). American Association of Petroleum Geologist (AAPG), Tulsa, 24 pp.
- Caers, J. (2003). History Matching Under Training-Image-Based Geological Model Constraints. *SPE Journal*, 8, 218-226.
- Davis, J.C. (1986). *Statistics and Data Analysis in Geology*. 2nd edition. John Wiley and Sons. New York, 646 pp.
- Davis, B.M. (1987). Uses and Abuses of Cross-Validation in Geostatistics. *Mathematical Geol.*, 19, 241-248.
- Delhomme, J.P. (1978). Kriging in the Hydrosciences, *Advanced Water Resources*, 1, 251-266.
- De Marsily, G. (1986). *Quantitative Hydrogeology*. Academic Press, London, 464 pp.
- Deutsch, C.V. y Journel, A.G. (1994). The Application of Simulated Annealing to Stochastic Reservoir Modeling. *SPE ATIS*, vol. 3, SPE paper 23565.
- Deutsch, C.V. y Journel, A.G. (1998). *GSLIB: Geostatistical Software Library and User's Guide*. 2nd edition, Oxford University Press, New York, 368 pp.
- Dorsel, D. y Breche, T. (1996). Environmental Sampling and Monitoring Primer: kriging, Online information: <http://www.cee.vt.edu> (acceded in 16 November, 2002), Virginia Polytechnic Institute and State University, Virginia, 9 pp.
- Englund, E. y Sparks, A. (1988). *Geostatistical Environmental Assessment Software (GEOEAS)*. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, EPA/600/4-88/033, Las Vegas, 174 pp.

- Farmer, C. (1991). Numerical Roacks. In: *The Mathematical Generation of Reservoir Geology* (J. Fayers y P. King, edit.). Oxford University Press, New York.
- Gandin, L.S. (1970). *The Planning of Meteorological Station Network*. WMO Tech. Note 111, World Meteorological Organization, Geneva, 120 pp.
- Gilardi, N., Melliush, T. y Maignan, M. (2001). Confidence Evaluation for Risk Prediction, Institut Dalle Molle d' Intelligence Artificielle Perceptive. In: *2001 Annual Conference of the International Association for Mathematical Geology*, Cancún, Mexico, 15 pp.
- Goovaerts, P. (1994). *Prediction and Stochastic Modelling of Facies Types Using Classification Algorithms and Simulated Annealing*. Annual report, Stanford Center for Reservoir Forecasting, Stanford University, 70 pp.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York. 483 pp.
- Haldorsen, H., Brand, P. y Macdonald, C. (1988). Review of the Stochastic Nature of Reservoirs. In: *Mathematics in Oil Production* (S. Edwards, y P. King, edit.). Clarendon Press, Oxford, 109-209.
- Haldorsen, H. y Damsleth, E. (1990). Stochastic Modeling, *J. Petroleum Technology*, April. 404-412.
- Hendricks Franssen, H.J. (2000). *Inverse Stochastic Modelling of Groundwater Flow and Mass Transporty*. Doctoral thesis, Universidad Politécnica de Valencia, Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos, Valencia, 363 pp.
- Iglewicz, B. y Hoaglin, D.C. (1993). *How to Detect and Handle Outliers*. Quality Press, American Society for Quality Control, Milwaukee, Wisconsin.
- Journel, A.G. (1983). Non-parametric distributions of spatial distributions. *Mathematical Geol.*, 15, 445-468.
- Journel, A.G. y Isaaks, E. (1984). Conditional Indicator Simulation: Application to a Saskatchewan Uranium Deposit. *Mathematical Geol.*, 16, 685-718.
- Journel, A.G. y Huijbregts, C.J. (1992). *Mining Geostatistics*. Academic Press, New York, 600 pp.
- Matheron, G. (1970). La Théorie des Variables Regionalisées et ses Applications. Fascicule 5, Les Cahiers du Centre de Morphologie Mathématique, Ecole des Mines de Paris, Fontainebleau.
- Matheron, G., Beucher, H., De Fouquet, C., Galli, A., Guerillot, D. y Ravenne, C. (1987). Conditional Simulation of the Geometry of Fluvio-deltaic Reservoirs, SPE paper # 16753, Society of Petroleum Engineering.
- Olea, R.A. (1974). Optimal Contour Mapping Using Universal Kriging. *J. Geophys. Res.*, 79: 695-702.
- Ortiz, J. y Deutsch, C. (2002). Calculation of Uncertainty in the Variogram. *Mathematical Geol.*, 34: 169-183.
- Pebesma, E.J. (2001). *Gstat User's Manual*. Dept. of Physical Geography, Utrecht University, Utrecht, 100 pp.
- Ripley, B. (1987). *Stochastic Simulation*. John Wiley & Sons, New York, 237 pp.
- Soares, A.O. (1992). Geostatistical Estimation of Multiphase Structures. *Mathematical Geol.*, 24, 149-160.
- Soares, A.O. (2000). *Geoestatística para as Ciências da Terra e do Ambiente*. Coleção Ensino da Ciência e da Tecnologia, IST Press, Lisboa, 206 pp.
- Soares, A.O. (2001). Direct Sequential Simulation and Cosimulation. *Mathematical Geol.*, 33, 911-926.
- Strebelle, S. (2000). *Sequential Simulation Drawing Structures from Training Images*. Doctoral thesis, Department of Geological and Environmental Sciences, Stanford University, Stanford, 187 pp.
- Strebelle, S. (2002). Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics. *Mathematical Geol.*, 34, 1-21.
- Suro-Pérez, V. (1991). *Generation of a Turbiditic Reservoir: the Boolean alternative: Stanford Center for Reservoir Forecasting*. Stanford University, Stanford.
- Todini, E. (2001). Influence of parameter estimation in Kriging: Part 1. Theoretical Development. *Hydrology & Earth System Sciences*, 5, 215-223.
- U.S. Army Corps of Engineers, USACE (1993). *Hydrologic Frequency Analysis*, Engineering Manual EM 1110-2-1415. Hydrologic Engineer Center, Davis, CA. 149 pp.
- Wingle, W. (1997). *Evaluating Subsurface Uncertainty Using Modified Geostatistical Techniques*. Doctoral thesis, Colorado School of Mines, Colorado, 202 pp.
- Xu, W. y Journel, A.G. (1993). *GTSIM: Gaussian Truncated Simulation of Lithofacies*. Stanford Center for Reservoir Forecasting, Stanford University, Stanford.

Recibido el 24 de noviembre de 2004
Aceptado el 5 de mayo de 2005