Geostatistics for Reservoir Characterization

A.G. Journel, Stanford U.

SPE Member

Abstract

Geostatistics and, more specifically, stochastic modeling of reservoir heterogeneities are being increasingly considered by reservoir analysts and engineers for their potential in generating more accurate reservoir models together with usable measures of spatial uncertainty. Geostatistics provides a probabilistic framework and a toolbox for data analysis with early integration of information. The uncertainty about the spatial distribution of critical reservoir parameters is modeled and transferred all the way to a risk-conscious reservoir management. The stochastic imaging (modeling) algorithms allow the generation of multiple, equiprobable, unsmoothed reservoir models yet all honoring the data available.

Introduction

Numerical models of reservoirs often fail to capture the heterogeneities that are critically important for reservoir performance. With production always being the primary function of a well, the available data are typically biased toward the more productive regions in a reservoir and are regretfully sparse. The interpolation and gridding algorithms commonly used by industry further exacerbate the problem since they are low-pass filters that tend to smooth out the little spatial variability that the sparse data reveal. While core plugs and well logs are not the only sources of information, other data, such as geophysical information, are often difficult to integrate since they have different levels of reliability and are representative of very different volumes of rock. History matching on historical production information and well test data does not guarantee reliable forecasts of a reservoir's future performance.

All of this is not news; these problems have come to the forefront as industry focuses on enhancing recovery from known reservoirs. With performance prediction for EOR processes calling for better numerical models of a reservoir rock and fluid properties, geostatistics is receiving renewed attention.

Until recently, geostatistics was often associated with only one of its important contributions and was used as a synonym for kriging, a multiple regression technique that has been most commonly used as a gridding procedure. By reducing geostatistics to another canned gridding software, early users failed to realize its full potential as a set of spatial data analysis tools, as a probabilistic language to be shared by geologists, geophysicists, and reservoir engineers, and as a vehicle for integrating various sources of uncertain information. Over time, more reasonable expectations and has broadened the scope of applications.

Spatial Data Analysis

Geostatistics begins with an emphasis on describing and modelling the spatial variability of reservoir properties and the spatial correlation between related properties such as porosity and seismic velocity. These models can then be used in the construction of numerical models for a variety of purposes—interpolations for a property whose average is critically important, stochastic simulations for a property whose extremes are critically important.

Whether one needs to transfer information from one reservoir to another, or between different units within the same reservoir, or whether one needs to transfer information from one discipline to another, a quantitative vehicle is necessary. Geostatistical models of spatial variability and dependence provide a quantitative summary of geological observations, and can therefore serve as a vehicle. Geostatistical models make it easier to compare data from different sedimentary basins, from different formations, and from different horizons, and are therefore valuable aids to any attempt at building a

References and Illustrations at end of paper
classification system. They also enable geologists, for example, to put their valuable information in a format that can be used by reservoir engineers.

Geological interpretation of the direction of maximum continuity can be corroborated through the use of directional correlograms and the plot of correlation ranges, see figure 1. In the absence of sufficient data in a particular reservoir, correlograms and other statistical characteristics can be borrowed from other relevant data sets such as originating from more mature fields in similar geological environments, outcrop studies or geological cross-sections and maps. By capturing the critical heterogeneities in a quantitative form, directional correlograms enable the use of this important information that often goes ignored since it commonly has only a qualitative expression.

Geostatisticians commonly use variograms to describe spatial continuity. While a correlogram, such as that shown in figure 1, shows the decay in the correlation between sample values as a function of increasing separation distance, the variogram shows the increase in dissimilarity between sample values versus increasing separation distance.

Though variograms and correlograms are commonly used to study spatial continuity of a particular variable, the same tools can be applied to the study of the cross-continuity of different variables at different locations. One could, for example, compare porosity at a particular location to travel time at a location nearby. Once modelled, this spatial cross-correlation can be used in a multivariate regression procedure known as cokriging for building a porosity map not only from the available porosity data but also from the more abundant seismic information.

Non-linear transformations of the data values can also be very useful in spatial data analysis. The spatial continuity of very skewed data, such as permeability values, is usually better understood through analysis of the logarithms of these data. Another example is the non-linear binary indicator transform, which can be defined as 1 if the permeability sample value is greater than a given threshold and 0 otherwise. The variogram of such an indicator transformation provides a measure of spatial connectivity of the high permeability values that are responsible for flow paths and the low permeability values that serve as barriers to flow.

In addition to variograms and correlograms, a geostatistician's toolkit contains a wide variety of plots and graphs for detecting, corroborating and modelling patterns of spatial variability and interdependence.

**Generalized Regression (Kriging - Cokriging)**

Any unsampled value, a porosity for example, can be estimated by generalized regression from surrounding measurements of the same value once the statistical relationship between the unknown being estimated and the available sample values has been defined. This is exactly what the correlogram provides: a prior model of the statistical similarity between data values. As was pointed out in the previous section, this generalized regression can also include nearby measurements of some different variable—seismic travel time, for example, or a facies code. When using such secondary information, one also needs a prior model of the cross-correlogram, which provides information on the statistical similarity between different variables at different locations. These generalized regression algorithms are collectively known as "kriging" and "cokriging". They generalize traditional regression as applied in well log analysis in two senses:

(i) The data used (the "independent variables" in traditional statistical jargon) need not be independent one from another. This allows redundancy to be taken into account, an important factor when using several related well logs simultaneously.

(ii) The sample correlation is modelled before being applied to the regression analysis. This allows for filtering certain aspects (frequencies) of the sample correlation. For example, white noise and more generally the high frequency components of the data can be filtered out, leaving an interpolated map that reflects only the large scale trends of these data. Conversely, a particular trend known from indirect information can be built into the map in such a way that it honors as closely as possible the available data.

**Integrating Data of Different Types**

Multivariate regression, or "cokriging" as a geostatistician would call it, is usually not a convenient framework for the integration of too widely different types of data such as quantitative geological information, which is usually only indicative in nature, and direct laboratory measurements.

At a particular location where one does not yet have a porosity measurement, a consideration of the lithofacies information might provide a reasonable interval within which the unknown value should fall. If we are certain that we are in a particular type of sandstone, for example, we might know that the porosity must be somewhere in the interval from 10% to 30%. If, in addition, we also have enough core plug measurements within that type of sandstone to build a histogram we could go further than simply stating the previous constrained interval. We could use that histogram to provide a probability distribution that might, for example, tell us that the unknown porosity is more likely to be on the low end of our 10% to 30% range than on the high end.

The indicator framework of geostatistics and the soft kriging algorithm allow an updating of such prior distribution using nearby data that may be either "soft" or "hard". To continue with our previous example, we could locally update the previous probability distribution obtained from lithofacies considerations with more specific local information. This could be soft information, such as the fact that the location in question is in the upper half of a fining-upwards sequence, or hard information, such as the fact that a full core measurement taken in a well only 50 feet away had a porosity of 15.6%.

The result of this updating is a "posterior" probability distribution that provides the probability for the unsampled value to belong within any given class of values, say, between 10% and 15% porosity. From such a distribution, any "optimal" estimate for the unsampled value can be derived once an optimality criterion has been specified. The optimality criterion preferred by most statisticians is the least square error criterion; for this definition of optimality, the mean of the posterior
probability distribution provides the best estimate. One could choose instead to use the absolute error, rather than the squared error, as the yardstick by which to measure the goodness of an estimate; for this optimality criterion, the median of the posterior probability distribution turns out to be the best estimate.

Rather than choosing some convenient, but arbitrary, definition of optimality, one could also analyze the intended use of any estimate and come up with a specific "loss function" that describes the penalty associated to any particular level of error. Such a loss function would likely be asymmetric since the impact of an underestimation of a particular magnitude is rarely the same as the impact of an overestimation of the same magnitude. This project-specific loss function can also be minimized, yielding a "best" estimate that aims specifically at minimizing the impact of error on each specific project.

Indicator geostatistics encodes all amenable information into a series of indicator (binary) data. Hard data result in a coding that consists entirely of 0's and 1's; soft data result in a coding that consists of intermediate values between 0 and 1. These indicator data are then interpreted as prior probability for any unsampled variable to take upon any particular value. No matter what their origin, be it well logs, seismic data or geological interpretation, all indicator data are pooled together. Experimental indicator variograms are then calculated to reveal the patterns of spatial continuity in these indicator data. Once the pattern of spatial continuity is well understood, it can be modelled and used in a multiple indicator regression to yield the posterior probability distribution of any unsampled value.

Traditionally, data of different types have been processed separately, leading to several different models—a geological model, a geophysical model, a production model, —which are difficult, if not impossible, to merge. Indicator geostatistics takes a quite different route by first merging all of the relevant information through a common coding of that information and, then, producing reservoir models consistent with that information.

Stochastic Imaging of Reservoir Heterogeneities

The thrust of modern geostatistics is not least-squares spatial regression but the building of probability distributions that characterize the present uncertainty about a reservoir rock and fluid properties. These probability distributions should account for all relevant information through models of the spatial dependence between each piece of information and the unknown variable of interest. As more information is collected, the uncertainty about the unknown variable of interest lessens and the spread of the posterior probability distribution decreases.

The "unknown" can be a single particular unsampled value, say porosity, at a single location, or it can be the unsampled values of a particular variable at many locations (the nodes of a regular grid, for example). It can even be the unsampled values of all relevant variables—porosity, permeability, saturations, pressures,...—at many locations. With a single variable at a single location, one has a 1-variate problem; with a single variable at N locations, the problem becomes an N-variate problem; with K interdependent variables at N locations, the problem becomes a (K · N)-variate problem.

Just as one can draw a series of outcomes from a univariate probability distribution, one can also draw a series of outcomes from a multivariate probability distribution. With a (K · N)-variate probability distribution representing the uncertainty in K rock and fluid properties at N locations, each drawing from such a multivariate probability distribution represents a stochastic image of the reservoir. If N is very large, several million nodes on a 3-dimensional grid, each outcome provides a high resolution image of the reservoir rock and fluid properties. Such stochastic images honor all of the available information, both hard and soft, at all locations and all of the structural information entered through the variograms and cross-correlograms, yet they are different one from each other, see figure 2.

The differences between these stochastic images provide a directly usable visualization of the uncertainty about the reservoir rock and fluid properties. Where all the different outcomes agree, there is little or no uncertainty; where they differ most there is maximum uncertainty. For example, a streak of high permeability values that persists on 90% of the images may be considered as reliable, whereas a streak that is present on only 50% of the images is less likely to exist. If the uncertain streaks happen to be in areas that are highly consequential to reservoir performance, the need for additional data has been established together with the locations at which these data are needed. Stochastic imaging might, for example, reveal that there is a lot of uncertainty about whether or not a particular EOR injector/producer pattern contains a high permeability streak near the top of the producing unit. If the nature of the EOR process makes gravity override a concern, the exercise of stochastic imaging will have identified the need for more accurate information on the upper portion of the pattern in question.

Stochastic images represent alternative, equiprobable numerical models of the reservoir rock and fluid properties and, as such, can be used as input to flow simulators for sensitivity analysis. Since a complex flow simulator could be very expensive to run, it is often necessary to consider only a limited number of reservoir models representative of the range of uncertainty: some unfavorable ones, some intermediate ones and some favorable ones. This selection can be done by running a stripped-down simulator that is fast and efficient yet also captures the relevant feature of the flow problem under consideration. A simple particle tracking algorithm, for example, might be useful in identifying the main flow paths, thereby allowing a trained engineer quickly to assess whether a particular reservoir model will produce favorable or unfavorable results.

References


Correlograms $\rho(|h|, \alpha)$ is a plot versus $|h|$ of the linear correlation coefficient between all pairs of porosity data $\phi(x)$, $\phi(x+\Delta)$ separated by approximately the same distance $|h|$ in approximately the same direction $\alpha$. The correlogram $\rho(|h|, \alpha)$, measures the loss of correlation as the separation distance $|h|$ between two sample values increases. The distance at which the correlation vanishes is called the range $a_\alpha$. The polar plot, $a_\alpha$ vs. $\alpha_1$, of these range values provides a measure of anisotropy, with the largest range corresponding to the direction of best spatial continuity (correlation).
Figure 2: Four equiprobable realizations of the sand-shale sequence over a vertical section. (The vertical exaggeration is 10:1)