INTRODUCTION TO GEOSTATISTICS

presented

26 June 2007

in

Hydrogeophysics: Theory, Methods, and Modeling
Boise State University, Boise, Idaho

Geoff Bohling
Assistant Scientist
Kansas Geological Survey
geoff@kgs.ku.edu
864-2093

These notes, tutorial, and data available at:

http://people.ku.edu/~gbohling/BoiseGeostat

Heterogeneity

Groundwater flow and transport are affected by heterogeneity in the governing properties at a range of scales. Inability to characterize these property variations completely *contributes* to uncertainty in flow and transport model predictions.

"Deterministic knowledge of a site's hydrogeology is a state we rarely, if ever, find ourselves in, although from a fundamental point of view there is no inherent element of chance in the hydrolgeological processes." -- Rubin, 2003

Latter part of statement is dominant, if not universal, philosophy: probabilistic models help us characterize/manage uncertainty due to imperfect knowledge, rather than reflecting inherent randomness in nature. (An engineering mindset?)

Geostatistics is a tool to help us characterize spatial variability and uncertainty resulting from imperfect characterization of that variability.

A major pitfall of geostatistical/stochastic modeling studies: Getting caught up in the "small" uncertainty that can be encompassed within the framework of a particular geostatistical conceptualization & modeling approach and ignoring bigger picture uncertainties, like

- Fundamentally different conceptual models of property variations
- Larger-scale problem geometry
- Consideration of different dominant processes (e.g., governing transport of a contaminant)
- Fundamentally different conceptual models of governing processes (e.g., nonlinear dynamic vs. stochastic representation of contaminant transport)

What is Geostatistics?

"In its broadest sense, geostatistics can be defined as the branch of statistical sciences that studies spatial/temporal phenomena and capitalizes on spatial relationships to model possible values of variable(s) at unobserved, unsampled locations"
(Caers, 2005)

"Geostatistics: study of phenomena that vary in space and/or time"
(Deutsch, 2002)

"Geostatistics can be regarded as a collection of numerical techniques that deal with the characterization of spatial attributes, employing primarily random models in a manner similar to the way in which time series analysis characterizes temporal data."
(Olea, 1999)

"Geostatistics offers a way of describing the spatial continuity of natural phenomena and provides adaptations of classical regression techniques to take advantage of this continuity." (Isaaks and Srivastava, 1989)

Geostatistics deals with spatially *autocorrelated* data.

Autocorrelation: correlation between elements of a series and others from the same series separated from them by a given interval. (Oxford American Dictionary)

Some spatially autocorrelated parameters of interest to hydrogeologists: facies, aquifer thickness, porosity, hydraulic conductivity

Some Geostatistics Textbooks

J. Caers, 2005, *Petroleum Geostatistics*, Society of Petroleum Engineers, 88 pages.
- o A short and fairly non-technical overview of current state-of-the-art in geostatistical techniques, focusing on methods for integrating static (property measurements at wells, etc.) and dynamic (flow) data.

C.V. Deutsch, 2002, *Geostatistical Reservoir Modeling*, Oxford University Press, 376 pages.
- o Focuses specifically on modeling of facies, porosity, and permeability for reservoir simulation.

C.V. Deutsch and A.G. Journel, 1998, *GSLIB: Geostatistical Software Library and User's Guide, Second Edition*, Oxford University Press, 369 pages.
- o Owner's manual for the GSLIB software library; serves as a standard reference for concepts and terminology.

P. Goovaerts, 1997, *Geostatistics for Natural Resources Evaluation*, Oxford University Press, 483 pages.
- o A nice introduction with examples focused on an environmental chemistry dataset; includes more advanced topics like factorial kriging.

E.H. Isaaks and R.M. Srivastava, 1989, *An Introduction to Applied Geostatistics*, Oxford University Press, 561 pages.
- o Probably the best introductory geostatistics textbook; intuitive development of concepts from first principles with clear examples at every step.

P.K. Kitanidis, 1997, *Introduction to Geostatistics: Applications in Hydrogeology*, Cambridge University Press, 249 pages.
- o A somewhat different take, with a focus on generalized covariance functions; includes discussion of geostatistical inversion of (groundwater) flow models.

R.A. Olea, 1999, *Geostatistics for Engineers and Earth Scientists*, Kluwer Academic Publishers, 303 pages.
- o Step by step mathematical development of key concepts, with clearly documented numerical examples.

Links to some software and online resources are available at

http://people.ku.edu/~gbohling/geostats

Basic Components of Geostatistics

(Semi)variogram analysis – characterization of spatial correlation

Kriging – optimal interpolation; generates best linear unbiased
estimate at each location; employs semivariogram model

Stochastic simulation – generation of multiple equiprobable
images of the variable; also employs semivariogram model

Geostatistical routines are implemented in the petroleum reservoir
modeling and aquifer modeling packages like Petrel, Roxar's Irap
RMS; and GMS (Groundwater Modeling System) -- used in the
generation of grids of facies, permeability (hydraulic conductivity),
porosity, etc. for the reservoir or aquifer.

Examples in lecture are in fact from the oil patch, but basic
problem is the same.

Exploratory Analysis of Example Data

Our example data consist of vertically averaged porosity values, in percent, in Zone A of the Big Bean Oil Field (fictitious, but based on real data). Porosity values are available from 85 wells distributed throughout the field, which is approximately 20 km in east-west extent and 16 km north-south. The porosities range from 12% to 17%. Here are the data values posted at the well locations:
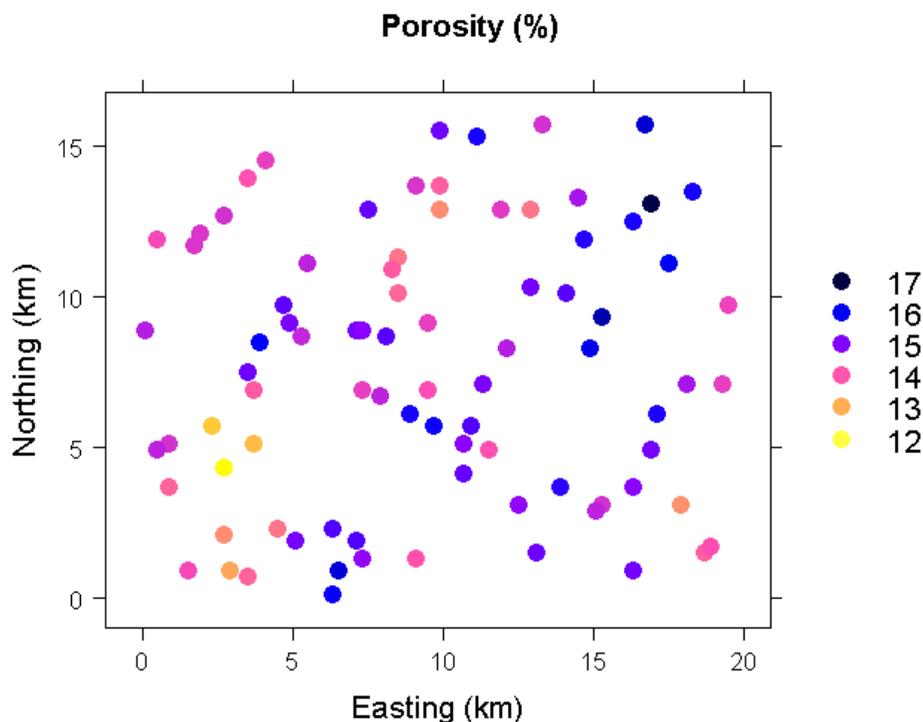

Geostatistical methods are optimal when data are
  - normally distributed and
  - stationary (mean and variance do not vary significantly in space)

Significant deviations from normality and stationarity can cause problems, so it is always best to begin by looking at a histogram or similar plot to check for normality and a posting of the data values in space to check for significant trends. The posting above shows some hint of a SW-NE trend, which we will check later.

Looking at the histogram (with a normal density superimposed) and a normal quantile-quantile plot shows that the porosity distribution does not deviate too severely from normality:
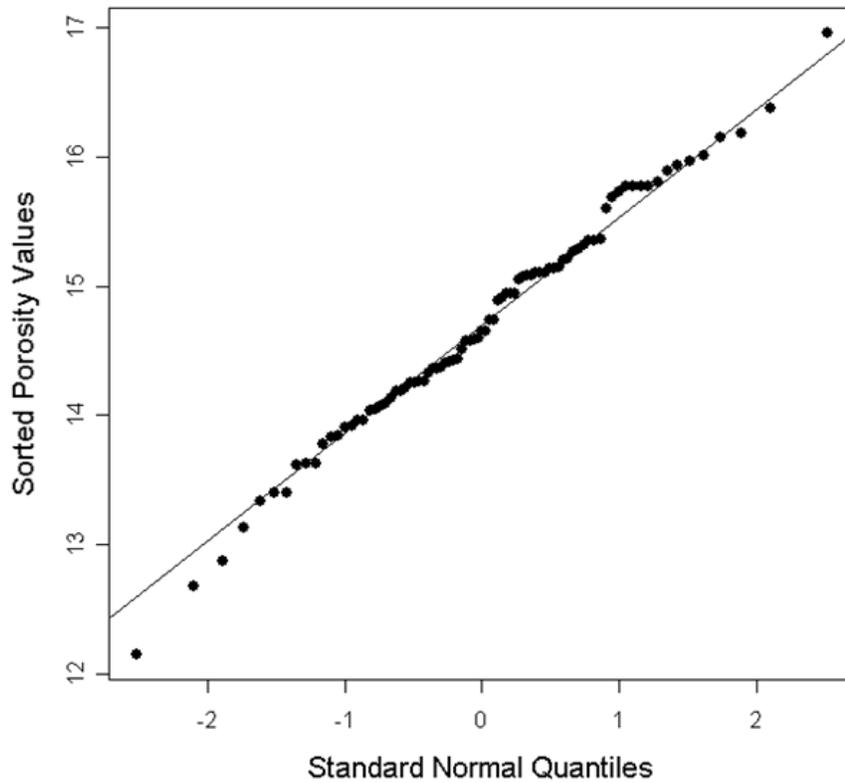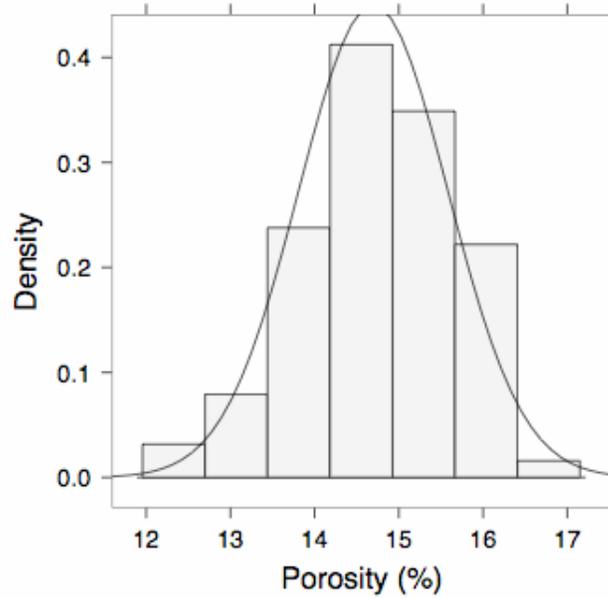
Spatial Covariance, Correlation and Semivariance

You may have already learned that covariance and correlation are measures of the similarity between two different variables:
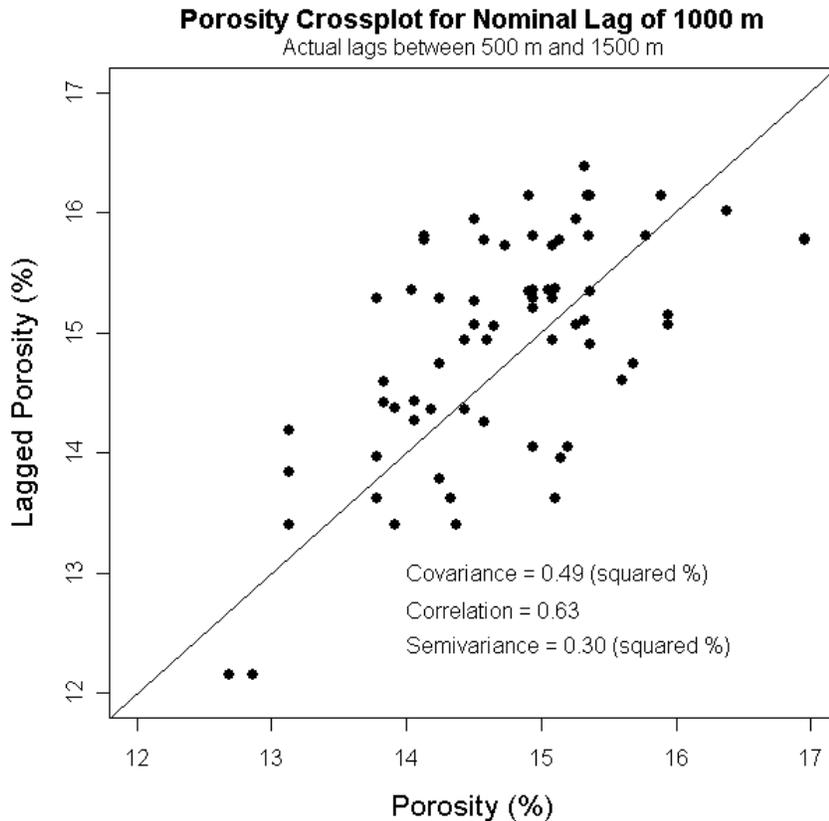
$$Cov(X,Y) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

The magnitude of the covariance increases with increasing similarity in the patterns of variation of the two variables about their respective means. The correlation coefficient, $\rho(X,Y)$, ranges from 1 for perfect positive correlation to -1 for perfect negative correlation and is in the vicinity of 0 for uncorrelated variables.

To extend these concepts to measures of spatial similarity, consider a scatterplot where the data pairs represent measurements of the *same* variable made some distance apart from each other. The separation distance is usually referred to as "lag", as used in time series analysis. We'll refer to the values plotted on the vertical axis as the lagged variable, although the decision as to which axis represents the lagged values is somewhat arbitrary.

Here is a scatterplot of Zone A porosity values at wells separated by a nominal lag of 1000 m:

Porosity Crossplot for Nominal Lag of 1000 m
Actual lags between 500 m and 1500 m

Covariance = 0.49 (squared %)
Correlation = 0.63
Semivariance = 0.30 (squared %)

Because of the irregular distribution of wells, we cannot expect to find many pairs of data values separated by exactly 1000 m, if we find any at all. Here we have introduced a "lag tolerance" of 500 m, pooling the data pairs with separation distances between 500 and 1500 m in order to get a reasonable number of pairs for computing statistics. The actual lags for the data pairs shown in the crossplot range from 566 m to 1456 m, with a mean lag of 1129 m.

9

The three statistics shown on the crossplot are the covariance, correlation, and semivariance between the porosity values on the horizontal axis and the lagged porosity values on the vertical axis. To formalize the definition of these statistics, we need to introduce some notation. Following standard geostatistical practice, we'll use:

u: vector of spatial coordinates (with components x, y or "easting" and "northing" for our 2D example)

$z(\mathbf{u})$: variable under consideration as a function of spatial location (porosity in this example)

h: lag vector representing separation between two spatial locations

$z(\mathbf{u}+\mathbf{h})$:   lagged version of variable under consideration

Sometimes $z(\mathbf{u})$ will be referred to as the "tail" variable and $z(\mathbf{u}+\mathbf{h})$ will be referred to as the "head" variable, since we can think of them as being located at the tail and head of the lag vector, **h**. The scatterplot of tail versus head values for a certain lag, **h**, is often called an **h**-scattergram.

Now, with $N(\mathbf{h})$ representing the number of pairs separated by lag $\mathbf{h}$ (plus or minus the lag tolerance), we can compute the statistics for lag $\mathbf{h}$ as

Covariance: $\displaystyle C(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} z(\mathbf{u}_\alpha) \cdot z(\mathbf{u}_\alpha + \mathbf{h}) - m_0 \cdot m_{+\mathbf{h}}$

Correlation: $\displaystyle \rho(\mathbf{h}) = \frac{C(\mathbf{h})}{\sqrt{\sigma_0^2 \cdot \sigma_{+\mathbf{h}}^2}}$

Semivariance: $\displaystyle \gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_\alpha + \mathbf{h}) - z(\mathbf{u}_\alpha)]^2$

where m_0 and $m_{+\mathbf{h}}$ are the means of the tail and head values:

$$m_0 = \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} z(\mathbf{u}_\alpha) \qquad m_{+\mathbf{h}} = \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} z(\mathbf{u}_\alpha + \mathbf{h})$$
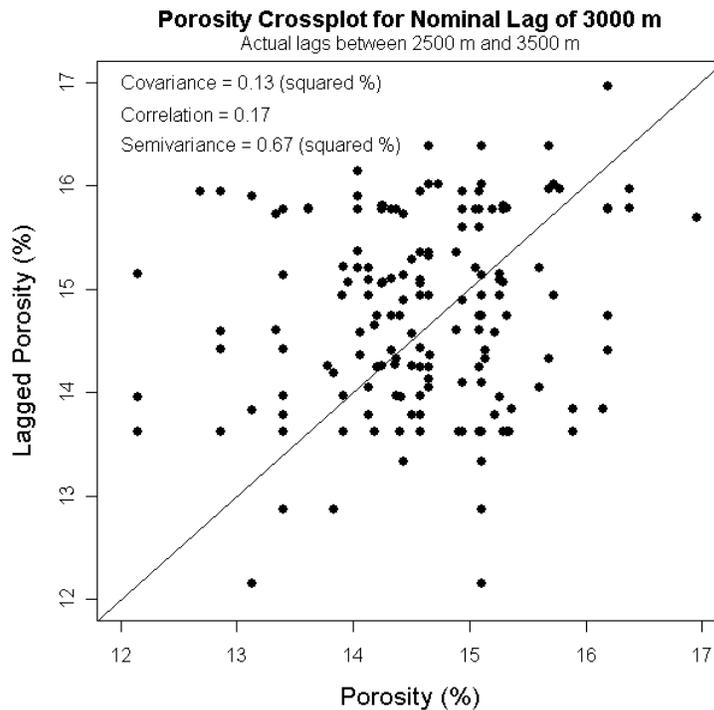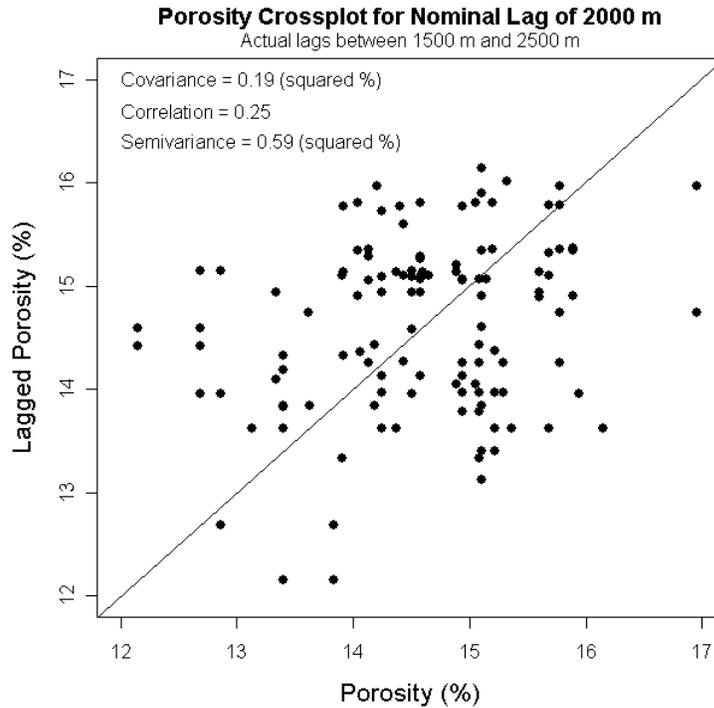
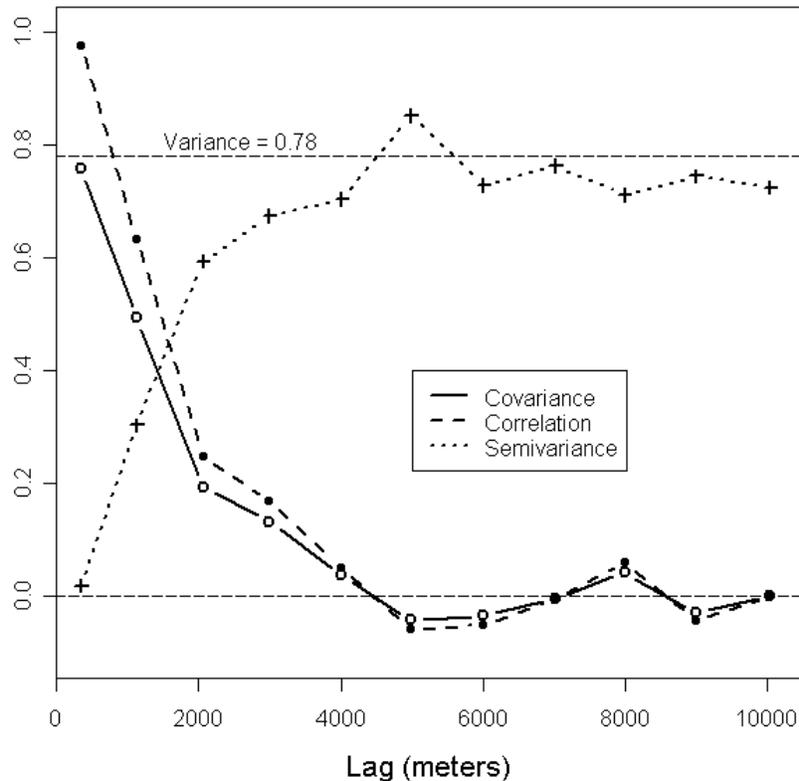and σ_0^2 and $\sigma_{+\mathbf{h}}^2$ are the corresponding variances:

$$\sigma_0^2 = \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_\alpha) - m_0]^2 \quad \sigma_{+\mathbf{h}}^2 = \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_\alpha + \mathbf{h}) - m_{+\mathbf{h}}]^2$$

Note that these definitions for $C(\mathbf{h})$ and $\gamma(\mathbf{h})$ use $N(\mathbf{h})$ in the denominator, not $N(\mathbf{h}) - 1$.

The semivariance is the moment of inertia or spread of the **h**-scattergram about the 45° (1 to 1) line shown on the plot. Covariance and correlation are both measures of the similarity of the head and tail values. Semivariance is a measure of the dissimilarity.

Here are the **h**-scatterplots for nominal lags of 2000 m and 3000 m. Note that the covariance and correlation decrease and the semivariance increases with increasing separation distance.

Porosity Crossplot for Nominal Lag of 2000 m
Actual lags between 1500 m and 2500 m

Covariance = 0.19 (squared %)
Correlation = 0.25
Semivariance = 0.59 (squared %)

Porosity Crossplot for Nominal Lag of 3000 m
Actual lags between 2500 m and 3500 m

Covariance = 0.13 (squared %)
Correlation = 0.17
Semivariance = 0.67 (squared %)

The plot above shows all three statistics versus actual mean lag for the contributing data pairs at each lag. The shortest lag shown (the nominally "zero" lag) includes six data pairs with a mean lag of 351 m. The correlation versus lag is referred to as the *correlogram* and the semivariance versus lag is the *semivariogram.* The covariance versus lag is generally just referred to as the covariance function.

The empirical functions that we have plotted – computed from the sample data – are of course just estimators of the theoretical functions $C(\mathbf{h})$, $\rho(\mathbf{h})$, and $\gamma(\mathbf{h})$, which can be thought of as population parameters. Estimating these functions based on irregularly distributed data (the usual case) can be very tricky due to the need to pool data pairs into lag bins. Larger lag spacings and tolerances allow more data pairs for estimation but reduce the amount of detail in the semivariogram (or covariance or

13

correlogram). The problem is particularly difficult for the shorter lags, which tend to have very few pairs (six in this example). This is unfortunate, since the behavior near the origin is the most important to characterize.

Under the condition of *second-order stationarity* (spatially constant mean and variance), the covariance function, correlogram, and semivariogram obey the following relationships:

$$C(\mathbf{0}) = \mathrm{Cov}(Z(\mathbf{u}), Z(\mathbf{u})) = \mathrm{Var}(Z(\mathbf{u}))$$
$$\rho(\mathbf{h}) = C(\mathbf{h})/C(\mathbf{0})$$
$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$$

In words, the lag-zero covariance should be equal to the global variance of the variable under consideration, the correlogram should look like the covariance function scaled by the variance, and the semivariogram should look like the covariance function turned upside down:

14
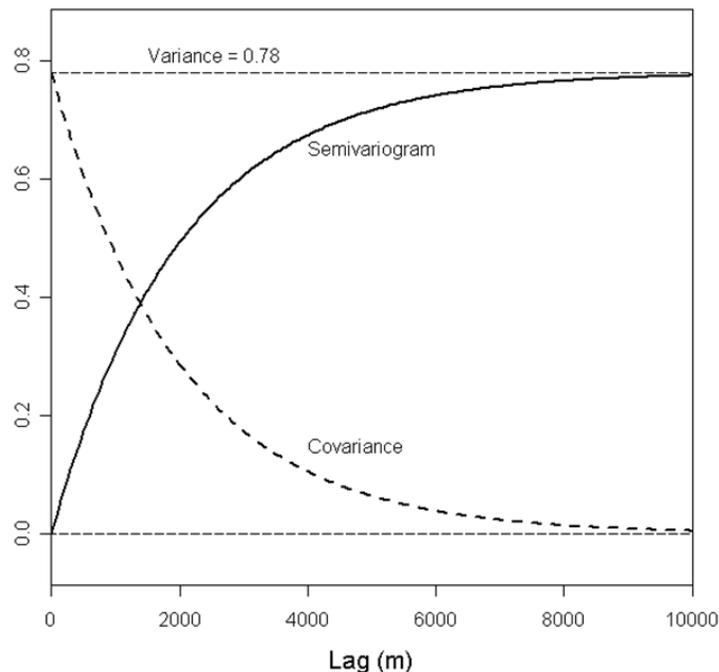
In practice, the estimated versions of the functions will violate these relationships to a greater or lesser extent due to sampling limitations and deviations from second-order stationarity.

Unlike time series analysts, who prefer to work with either the covariance function or the correlogram, geostatisticians typically work with the semivariogram. This is primarily because the semivariogram, which averages squared differences of the variable, tends to filter the influence of a spatially varying mean. Also, the semivariogram can be applied whenever the first differences of the variable, $Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})$, are second-order stationary. This form of stationarity, referred to as the *intrinsic hypothesis*, is a weaker requirement than second-order stationarity of the variable itself, meaning that the semivariogram can be defined in some cases where the covariance function cannot be defined. In particular, the semivariance may keep increasing with increasing lag, rather than leveling off, corresponding to an infinite global variance. In this case the covariance function is undefined.

Trend

If the empirical semivariogram continues climbing steadily beyond the global variance value, this is often indicative of a significant spatial trend in the variable, resulting in a negative correlation between variable values separated by large lags. Three options for dealing with lag include:

    1) Fit a trend surface and work with residuals from the trend
    2) Try to find a "trend-free" direction and use the variogram in that direction as the variogram for the "random" component of the variable (see the section on anisotropy, below)
    3) Ignore the problem and use a linear or power variogram

The semivariogram for the porosity data does not seem to indicate a significant trend.

Characteristics of the Semivariogram
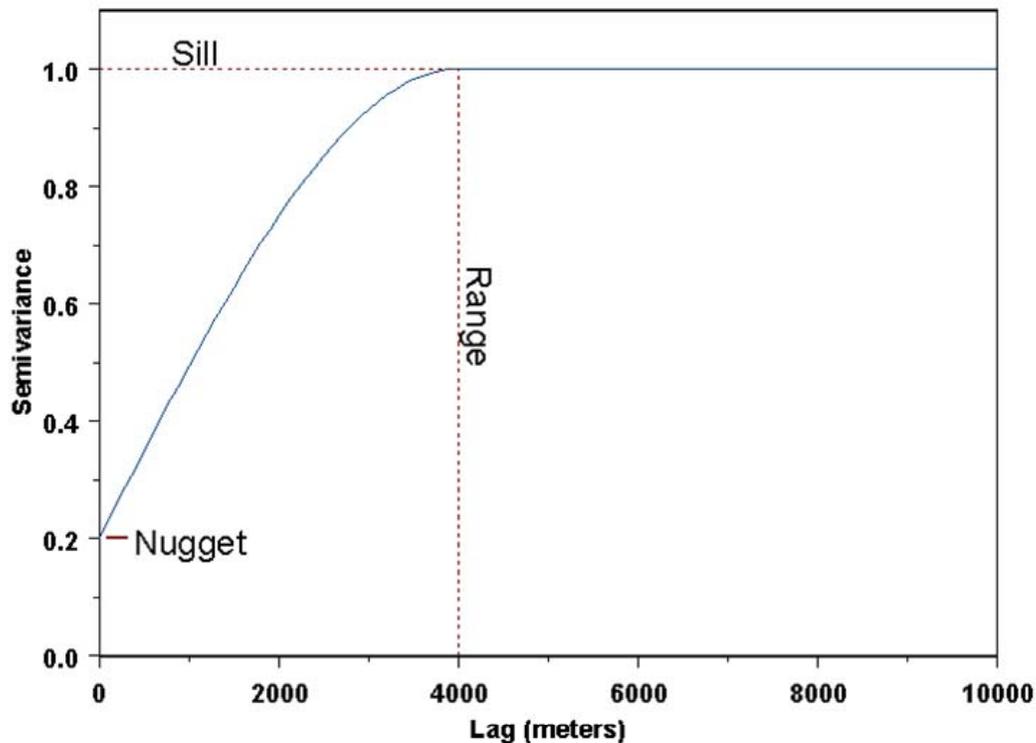
Sill: The semivariance value at which the variogram levels off. Also used to refer to the "amplitude" of a certain component of the semivariogram. For the plot above, "sill" could refer to the overall sill (1.0) *or* to the difference (0.8) between the overall sill and the nugget (0.2). Meaning depends on context.

Range: The lag distance at which the semivariogram (or semivariogram component) reaches the sill value. Presumably, autocorrelation is essentially zero beyond the range.

Nugget: In theory the semivariogram value at the origin (0 lag) should be zero. If it is significantly different from zero for lags very close to zero, then this semivariogram value is referred to as the nugget. The nugget represents variability at distances smaller than the typical sample spacing, including measurement error.

Modeling the Semivariogram

For the sake of kriging (or stochastic simulation), we need to replace the empirical semivariogram with an acceptable semivariogram model. Part of the reason for this is that the kriging algorithm will need access to semivariogram values for lag distances other than those used in the empirical semivariogram. More importantly, the semivariogram models used in the kriging process need to obey certain numerical properties in order for the kriging equations to be solvable. (Technically, the semivariogram model needs to be *non-negative definite*, in order the system of kriging equations to be *non-singular*.) Therefore, geostatisticians choose from a palette of acceptable or *licit* semivariogram models.

Using h to represent lag distance, a to represent (practical) range, and c to represent sill, the five most frequently used models are:

Nugget:
$$g(h) = \begin{cases} 0 & \text{if } h = 0 \\ c & \text{otherwise} \end{cases}$$

Spherical:
$$g(h) = \begin{cases} c \cdot \left(1.5 \left(\dfrac{h}{a} \right) - 0.5 \left(\dfrac{h}{a} \right)^3 \right) & \text{if } h \leq a \\ c & \text{otherwise} \end{cases}$$

Exponential:
$$g(h) = c \cdot \left(1 - \exp\left(\frac{-3h}{a} \right) \right)$$

Gaussian:
$$g(h) = c \cdot \left(1 - \exp\left(\frac{-3h^2}{a^2} \right) \right)$$

Power:
$$g(h) = c \cdot h^{\omega} \quad \text{with } 0 < \omega < 2$$

The nugget model represents the discontinuity at the origin due to small-scale variation. On its own it would represent a purely random variable, with no spatial correlation.

The spherical model actually reaches the specified sill value, c, at the specified range, a. The exponential and Gaussian approach the sill asymptotically, with a representing the practical range, the distance at which the semivariance reaches 95% of the sill value. These three models are shown below:
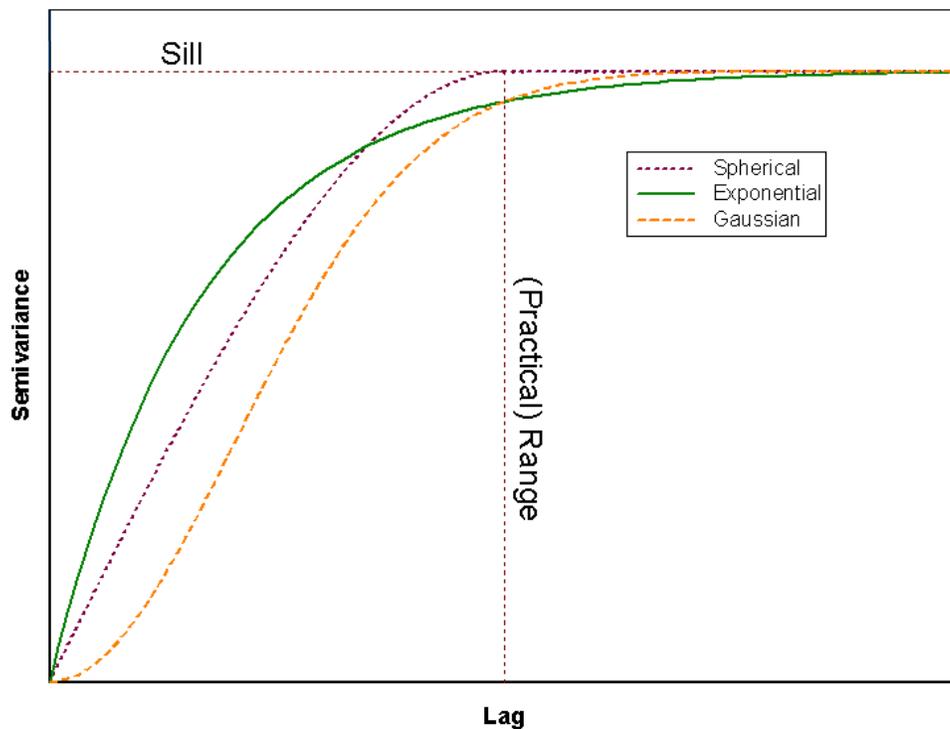

The Gaussian model, with its parabolic behavior at the origin, represents very smoothly varying properties. (However, using the Gaussian model alone without a nugget effect can lead to numerical instabilities in the kriging process.) The spherical and exponential models exhibit linear behavior the origin, appropriate for representing properties with a higher level of short-range variability.

18

Models with a finite sill, like the Gaussian, exponential, and spherical, are referred to as *transition* models and have corresponding covariance functions given by $\mathrm{cov}(h) = c - g(h)$. The power model does not reach a finite sill and does not have a corresponding covariance function. Power-law semivariogram models are appropriate for properties exhibiting fractal behavior.

Linear combinations of licit semivariogram models are also licit models, so that more complicated models may be built by adding together the basic models described above with different ranges and sills, leading to the second meaning of "sill" discussed above. For example, a variogram model might consist of a nugget of 0.1, a Gaussian component with a range of 1500 m and a sill of 0.2 and an exponential component with a range of 8000 m and sill of 0.6, giving an overall sill of 0.9:
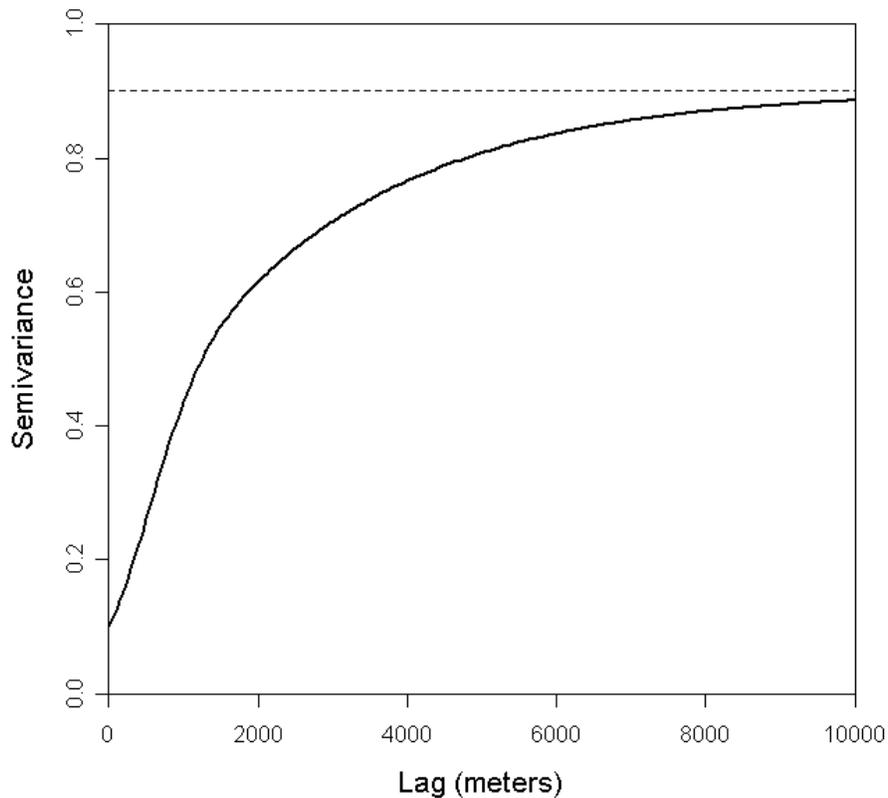
The actual process of fitting a model to an empirical semivariogram is much more of an art than a science, with different authorities suggesting different methods and protocols. Since empirical semivariograms are often quite noisy, quite a bit of subjective judgment goes into selecting a good model. Here is the semivariogram for our example porosity data, with three fitted models. In each case the sill value was fixed at the overall variance of 0.78 and the range was estimated using weighted nonlinear regression (weighting by number of data pairs for each lag):

Porosity Semivariogram with Three Models
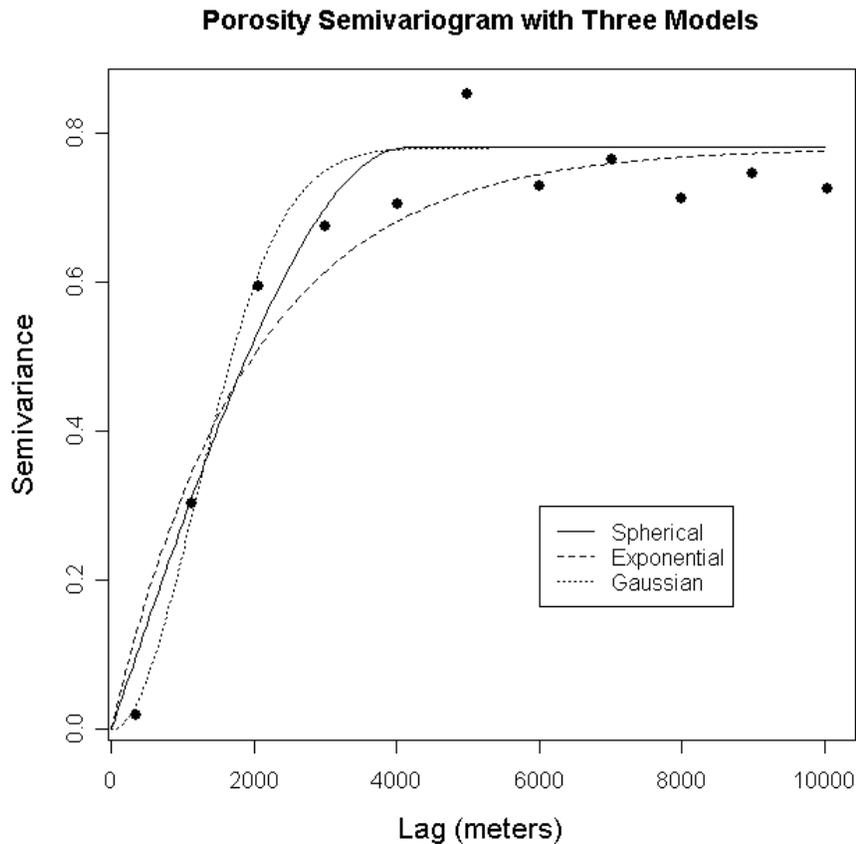
The fitted ranges for the three models are 4141 m for the spherical, 5823 m for the exponential, and 2884 m for the Gaussian. The Gaussian model gives the best fit, but the spherical is a close second.

Anisotropy

The above discussion has assumed that the spatial correlation structure is the same in all directions, or *isotropic*. In this case the covariance function, correlogram, and semivariogram depend only on the magnitude of the lag vector, $h = |\mathbf{h}|$, and not the direction, and the empirical semivariogram can be computed by pooling data pairs separated by the appropriate distances, regardless of direction. Such a semivariogram is described as *omnidirectional*.

In many cases, however, a property shows different autocorrelation structures in different directions, and an *anisotropic* semivariogram model should be developed to reflect these differences. The most commonly employed model for anisotropy is geometric anisotropy, with the semivariogram reaching the same sill in all directions, but at different ranges. In geological settings, the most prominent form of anisotropy is a strong contrast in ranges in the (stratigraphically) vertical and horizontal directions, with the vertical semivariogram reaching the sill in a much shorter distance than the horizontal semivariogram. In some settings, there may also be significant lateral anisotropy, often reflecting prominent directionality in the depositional setting (such as, along and perpendicular to channels).

The most common approach to modeling geometric anisotropy is to find ranges, a_x, a_y, and a_z, in three principal, orthogonal directions and transform a three-dimensional lag vector $\mathbf{h} = (h_x, h_y, h_z)$ into an equivalent isotropic lag using:

$$h = \sqrt{(h_x/a_x)^2 + (h_y/a_y)^2 + (h_z/a_z)^2}$$

To check for directional dependence in an empirical semivariogram, we have to compute semivariance values for data pairs falling within certain directional bands as well as falling within the prescribed lag limits. The directional bands are specified by a given azimuthal direction, angular tolerance, and bandwidth:
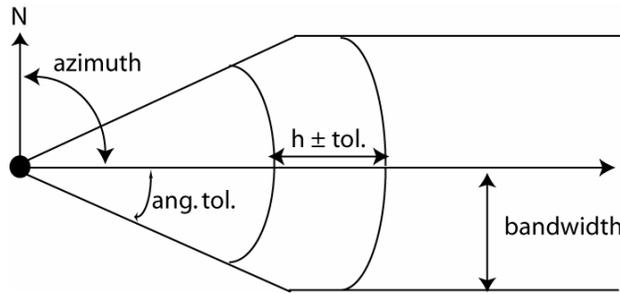
Here is the porosity semivariogram in the directions N 0° E, N 45° E, N 90° E, and N 135° E with angular tolerance of 22.5 and no bandwidth limit, together with the omnidirectional Gaussian model:
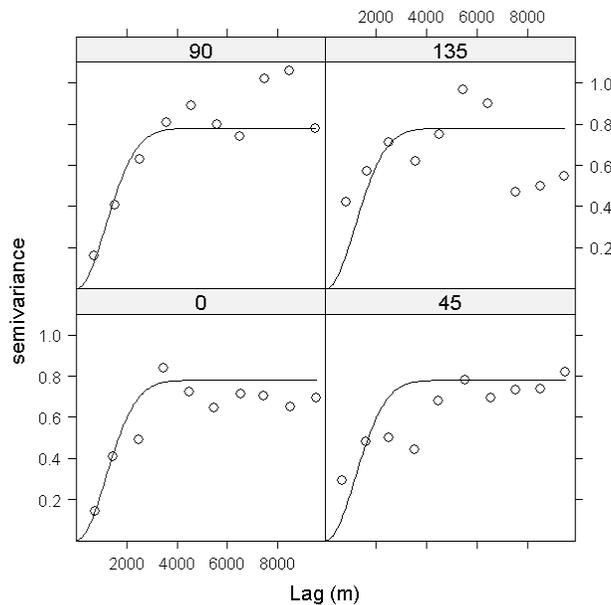

The directional semivariograms are noisier due to the reduced number of data pairs used for estimation. They do not show overwhelming evidence of anisotropy.

What is Kriging?

Optimal interpolation based on regression against observed z values of surrounding data points, weighted according to spatial covariance values.

Pronunciation: Hard "g" (as in Danie Krige) or soft "g" (á là Georges Matheron), take your pick

What is interpolation? Estimation of a variable at an unmeasured location from observed values at surrounding locations. For example, estimating porosity at $\mathbf{u}$ = (2000 m, 4700 m) based on porosity values at nearest six data points in our Zone A data:
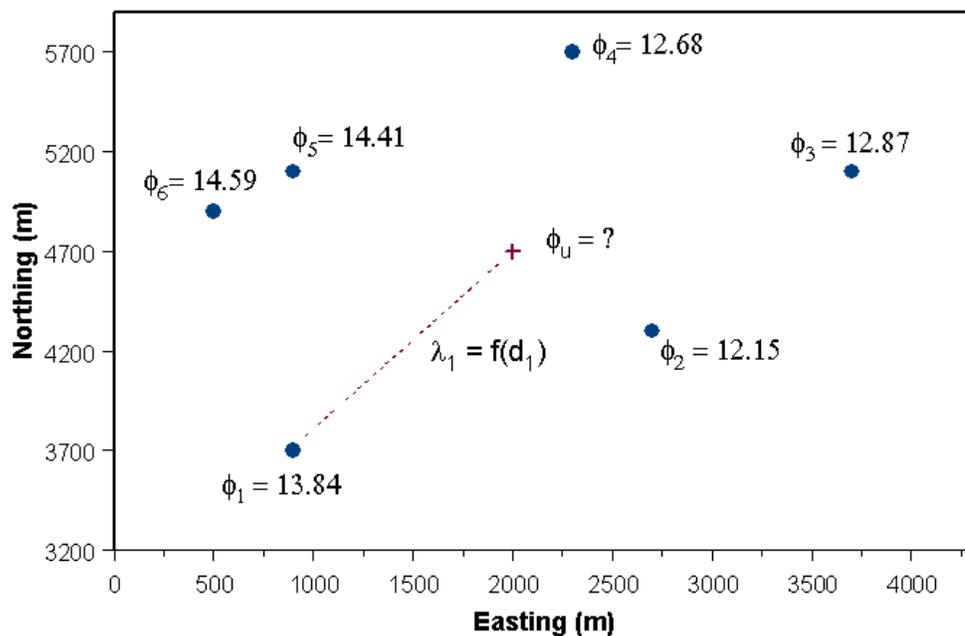
It would seem reasonable to estimate ϕ_u by a weighted average $\sum \lambda_\alpha \phi_\alpha$, with weights λ_α given by some decreasing function of the distance, d_α, from $\mathbf{u}$ to data point α.

All interpolation algorithms (inverse distance squared, splines, radial basis functions, triangulation, *etc.*) estimate the value at a given location as a weighted sum of data values at surrounding locations. Almost all assign weights according to functions that give a decreasing weight with increasing separation distance. Kriging assigns weights according to a (moderately) data-driven weighting function, rather than an arbitrary function, but it is still just an interpolation algorithm and will give very similar results to others in many cases (Isaaks and Srivastava, 1989). In particular:

- If the data locations are fairly dense and uniformly distributed throughout the study area, you will get fairly good estimates regardless of interpolation algorithm.
- If the data locations fall in a few clusters with large gaps in between, you will get unreliable estimates regardless of interpolation algorithm.
- Almost all interpolation algorithms will underestimate the highs and overestimate the lows; this is inherent to averaging and if an interpolation algorithm didn't average we wouldn't consider it reasonable

Some advantages of kriging:

- Helps to compensate for the effects of data clustering, assigning individual points within a cluster less weight than isolated data points (*or*, treating clusters more like single points)
- Gives estimate of estimation error (kriging variance), along with estimate of the variable, Z, itself (but error map is basically a scaled version of a map of distance to nearest data point, so not that unique)
- Availability of estimation error provides basis for stochastic simulation of possible realizations of $Z(\mathbf{u})$

Kriging approach and terminology

Goovaerts, 1997: "All kriging estimators are but variants of the basic linear regression estimator $Z^*(\mathbf{u})$ defined as

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha [Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)] ."$$

with

$\mathbf{u}, \mathbf{u}_\alpha$: location vectors for estimation point and one of the neighboring data points, indexed by α

$n(\mathbf{u})$: number of data points in local neighborhood used for estimation of $Z^*(\mathbf{u})$

$m(\mathbf{u}), m(\mathbf{u}_\alpha)$: expected values (means) of $Z(\mathbf{u})$ and $Z(\mathbf{u}_\alpha)$

$\lambda_\alpha(\mathbf{u})$: kriging weight assigned to datum $z(\mathbf{u}_\alpha)$ for estimation location $\mathbf{u}$; same datum will receive different weight for different estimation location

$Z(\mathbf{u})$ is treated as a random field with a trend component, $m(\mathbf{u})$, and a residual component, $R(\mathbf{u}) = Z(\mathbf{u}) - m(\mathbf{u})$. Kriging estimates residual at $\mathbf{u}$ as weighted sum of residuals at surrounding data points. Kriging weights, λ_α, are derived from covariance function or semivariogram, which should characterize residual component. Distinction between trend and residual somewhat arbitrary; varies with scale.

Development here will follow that of Pierre Goovaerts, 1997, *Geostatistics for Natural Resources Evaluation*, Oxford University Press.

We will continue working with our example porosity data, including looking in detail at results in the six-data-point region shown earlier:

Porosity (%)
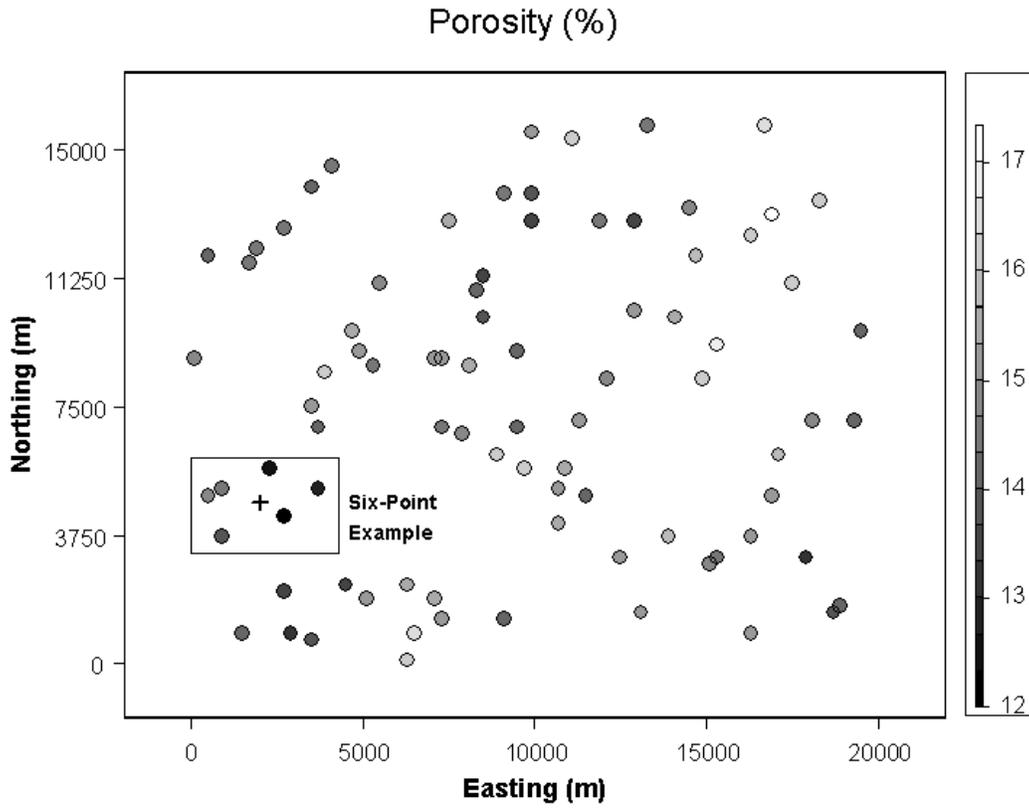
Basics of Kriging

Again, the basic form of the kriging estimator is

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha [Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)]$$

The goal is to determine weights, λ_α, that minimize the variance of the estimator

$$\sigma_E^2(\mathbf{u}) = \mathrm{Var}\{Z^*(\mathbf{u}) - Z(\mathbf{u})\}$$

under the unbiasedness constraint $\mathrm{E}\{Z^*(\mathbf{u}) - Z(\mathbf{u})\} = 0$.

The random field (RF) $Z(\mathbf{u})$ is decomposed into residual and trend components, $Z(\mathbf{u}) = R(\mathbf{u}) + m(\mathbf{u})$, with the residual component treated as an RF with a stationary mean of 0 and a stationary covariance (a function of lag, $\mathbf{h}$, but not of position, $\mathbf{u}$):

$$\mathrm{E}\{R(\mathbf{u})\} = 0$$

$$\mathrm{Cov}\{R(\mathbf{u}), R(\mathbf{u}+\mathbf{h})\} = \mathrm{E}\{R(\mathbf{u}) \cdot R(\mathbf{u}+\mathbf{h})\} = C_R(\mathbf{h})$$

The residual covariance function is generally derived from the input semivariogram model, $C_R(\mathbf{h}) = C_R(\mathbf{0}) - \gamma(\mathbf{h}) = Sill - \gamma(\mathbf{h})$. Thus the semivariogram we feed to a kriging program should represent the residual component of the variable.

The three main kriging variants, simple, ordinary, and kriging with a trend, differ in their treatments of the trend component, $m(\mathbf{u})$.

Simple Kriging

For simple kriging, we assume that the trend component is a constant and known mean, $m(\mathbf{u}) = m$, so that

$$Z^*_{SK}(u) = m + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{SK}_\alpha(\mathbf{u})\left[Z(\mathbf{u}_\alpha) - m\right].$$

This estimate is automatically unbiased, since $\mathrm{E}[Z(\mathbf{u}_\alpha) - m] = 0$, so that $\mathrm{E}\left[Z^*_{SK}(\mathbf{u})\right] = m = \mathrm{E}[Z(\mathbf{u})]$. The estimation error $Z^*_{SK}(\mathbf{u}) - Z(\mathbf{u})$ is a linear combination of random variables representing residuals at the data points, $\mathbf{u}_\alpha$, and the estimation point, $\mathbf{u}$:

$$Z^*_{SK}(\mathbf{u}) - Z(\mathbf{u}) = \left[Z^*_{SK}(\mathbf{u}) - m\right] - \left[Z(\mathbf{u}) - m\right]$$

$$= \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{SK}_\alpha(\mathbf{u}) R(\mathbf{u}_\alpha) - R(\mathbf{u}) = R^*_{SK}(\mathbf{u}) - R(\mathbf{u})$$

Using rules for the variance of a linear combination of random variables, the error variance is then given by

$$\sigma^2_E(\mathbf{u}) = \mathrm{Var}\left\{R^*_{SK}(\mathbf{u})\right\} + \mathrm{Var}\left\{R_{SK}(\mathbf{u})\right\} - 2\,\mathrm{Cov}\left\{R^*_{SK}(\mathbf{u}), R_{SK}(\mathbf{u})\right\}$$

$$= \sum_{\alpha=1}^{n(\mathbf{u})}\sum_{\beta=1}^{n(\mathbf{u})} \lambda^{SK}_\alpha(\mathbf{u})\lambda^{SK}_\beta(\mathbf{u}) C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) + C_R(0) - 2\sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{SK}_\alpha(\mathbf{u}) C_R(\mathbf{u}_\alpha - \mathbf{u})$$

To minimize the error variance, we take the derivative of the above expression with respect to each of the kriging weights and set each derivative to zero. This leads to the following system of equations:

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda^{SK}_\beta(\mathbf{u}) C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C_R(\mathbf{u}_\alpha - \mathbf{u}) \qquad \alpha = 1,\ldots,n(\mathbf{u})$$

Because of the constant mean, the covariance function for $Z(\mathbf{u})$ is the same as that for the residual component, $C(\mathbf{h}) = C_R(\mathbf{h})$, so that we can write the simple kriging system directly in terms of $C(\mathbf{h})$:

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{SK}(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C(\mathbf{u}_\alpha - \mathbf{u}) \qquad \alpha = 1, \ldots, n(\mathbf{u}).$$

This can be written in matrix form as

$$\mathbf{K} \lambda_{SK}(\mathbf{u}) = \mathbf{k}$$

where $\mathbf{K}_{SK}$ is the matrix of covariances between data points, with elements $\mathbf{K}_{i,j} = C(\mathbf{u}_i - \mathbf{u}_j)$, $\mathbf{k}$ is the vector of covariances between the data points and the estimation point, with elements given by $\mathbf{k}_i = C(\mathbf{u}_i - \mathbf{u})$, and $\lambda_{SK}(\mathbf{u})$ is the vector of simple kriging weights for the surrounding data points.

If the covariance model is licit (meaning the underlying semivariogram model is licit) and no two data points are colocated, then the data covariance matrix is positive definite and we can solve for the kriging weights using

$$\lambda_{SK} = \mathbf{K}^{-1} \mathbf{k}$$

Once we have the kriging weights, we can compute both the kriging estimate and the kriging variance, which is given by

$$\sigma_{SK}^2(\mathbf{u}) = C(0) - \lambda_{SK}^T(\mathbf{u})\mathbf{k} = C(0) - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{SK}(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u})$$

after substituting the kriging weights into the error variance expression above.

So what does all this math do? It finds a set of weights for estimating the variable value at the location u from values at a set of neighboring data points. The weight on each data point generally decreases with increasing distance to that point, in accordance with the decreasing data-to-estimation covariances specified in the right-hand vector, **k**. However, the set of weights is also designed to account for redundancy among the data points, represented in the data point-to-data point covariances in the matrix **K**. Multiplying **k** by **K**$^{-1}$ (on the left) will downweight points falling in clusters relative to isolated points at the same distance.

We will apply simple kriging to our porosity data, using the spherical semivariogram that we fit before, with zero nugget, a sill of 0.78, and a range of 4141 m:
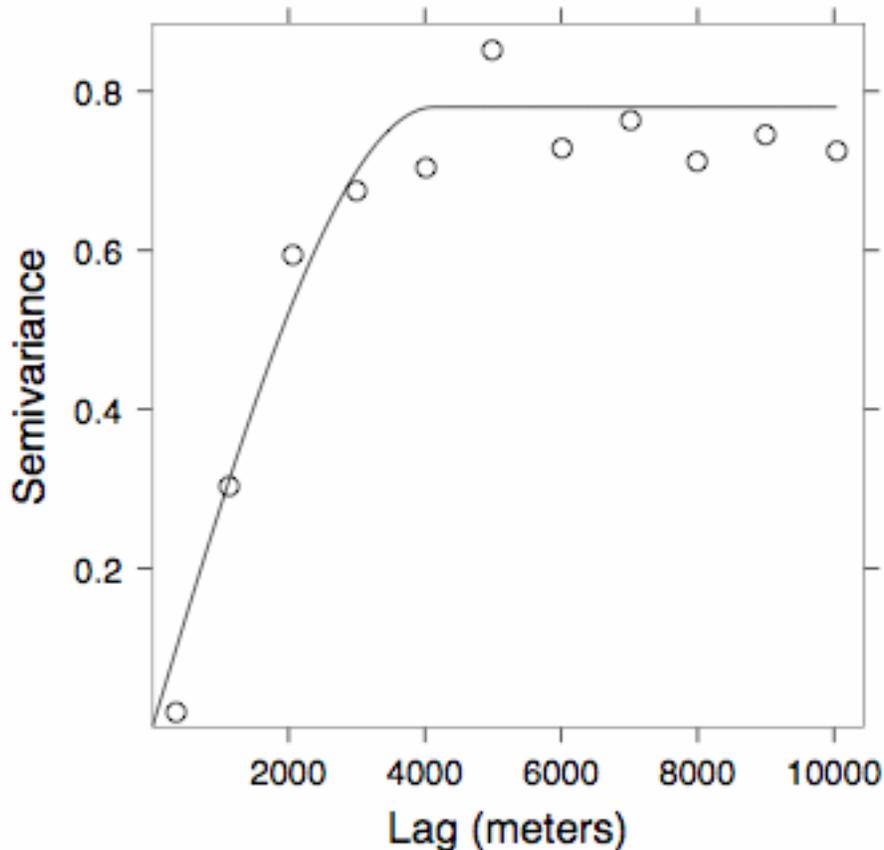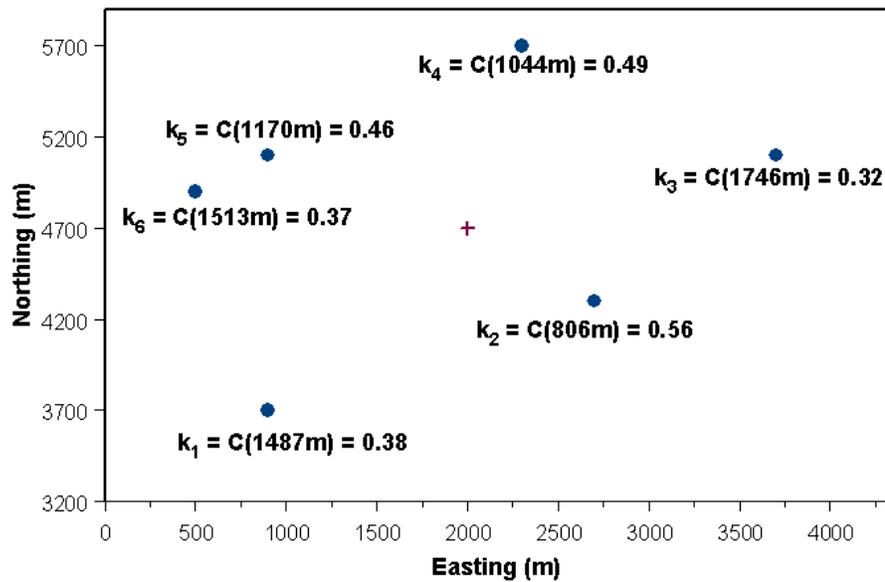
Since we are using a spherical semivariogram, the covariance function is given by

$$C(h) = C(0) - \gamma(h) = 0.78 \cdot \left(1 - 1.5 \cdot (h/4141) + 0.5 \cdot (h/4141)^3\right)$$

for separation distances, h, up to 4141 m, and 0 beyond that range.

The plot below shows the elements of the right-hand vector, $\mathbf{k} = [0.38, 0.56, 0.32, 0.49, 0.46, 0.37]^T$, obtained from plugging the data-to-estimation-point distances into this covariance function:

The matrix of distances between the pairs of data points (rounded to the nearest meter) is given by

	Point 1	Point 2	Point 3	Point 4	Point 5	Point 6
Point 1	0	1897	3130	2441	1400	1265
Point 2	1897	0	1281	1456	1970	2280
Point 3	3130	1281	0	1523	2800	3206
Point 4	2441	1456	1523	0	1523	1970
Point 5	1400	1970	2800	1523	0	447
Point 6	1265	2280	3206	1970	447	0

This translates into a data covariance matrix of

$$\mathbf{K} = \begin{bmatrix} 0.78 & 0.28 & 0.06 & 0.17 & 0.40 & 0.43 \\ 0.28 & 0.78 & 0.43 & 0.39 & 0.27 & 0.20 \\ 0.06 & 0.43 & 0.78 & 0.37 & 0.11 & 0.06 \\ 0.17 & 0.39 & 0.37 & 0.78 & 0.37 & 0.27 \\ 0.40 & 0.27 & 0.11 & 0.37 & 0.78 & 0.65 \\ 0.43 & 0.20 & 0.06 & 0.27 & 0.65 & 0.78 \end{bmatrix}$$

(rounded to two decimal places). Note in particular the relatively high correlation between points 5 and 6, separated by 447 m.

The resulting vector of kriging weights is

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \end{bmatrix} = \mathbf{K}^{-1}\mathbf{k} = \begin{bmatrix} 0.1475 \\ 0.4564 \\ -0.0205 \\ 0.2709 \\ 0.2534 \\ -0.0266 \end{bmatrix}$$

Notice that data point 6 is assigned a very small weight relative to data point 1, even though they are both about the same distance from the estimation point and have about the same data-point-to-estimation-point covariance ($k_1 = 0.38$, $k_6 = 0.37$). This is because data point 6 is effectively "screened" by the nearby data point 5. Data points 5 and 6 are fairly strongly correlated with each other and 5 has a stronger correlation with the estimation point, so data point 6 is effectively ignored. Note that the covariances and thus the kriging weights are determined entirely by the data configuration and the covariance model, not the actual data values. The porosities at points 5 and 6 could in fact be very different and this would have no influence on the kriging weights.

The mean porosity value for the 85 wells is 14.70%, and the porosity values at the six example wells are 13.84%, 12.15%, 12.87%, 12.68%, 14.41%, and 14.59%. The estimated residual from the mean at **u** is given by the dot product of the kriging weights and the vector of residuals at the data points:

$$R(\mathbf{u}) = \boldsymbol{\lambda}' \mathbf{R}_\alpha$$

$$= \begin{bmatrix} 0.15 & 0.46 & -0.02 & 0.27 & 0.25 & -0.03 \end{bmatrix} \begin{bmatrix} -0.86 \\ -2.55 \\ -1.83 \\ -2.01 \\ -0.28 \\ -0.10 \end{bmatrix} = -1.87$$

Adding the mean back into this estimated residual gives an estimated porosity of $\hat{Z}(\mathbf{u}) = R(\mathbf{u}) + m = -1.87 + 14.70 = 12.83\%$. Similarly, plugging the kriging weights and the vector k into the expression for the estimation variance gives a variance of 0.238 (squared %). Given these two pieces of information, we can represent the porosity at **u** = (2000 m, 4700 m) as a normal distribution with a mean of 12.83% and a standard deviation of 0.49%. Note that, like the kriging weights, the variance estimate depends entirely on the data configuration and the covariance function, not on the data values themselves. The estimated kriging variance would be the same regardless of whether the actual porosity values in the neighborhood were very similar or highly variable. The influence of the data values, through the fitting of the semivariogram model, is quite indirect.

Here are the simple kriging estimates and standard deviation on a 100x80 grid with 100-meter spacing using the spherical semivariogram model and estimating each grid value from the 16 nearest neighbor data points (well locations):

Estimated Porosity (%) Using Simple Kriging

Simple Kriging Standard Deviation
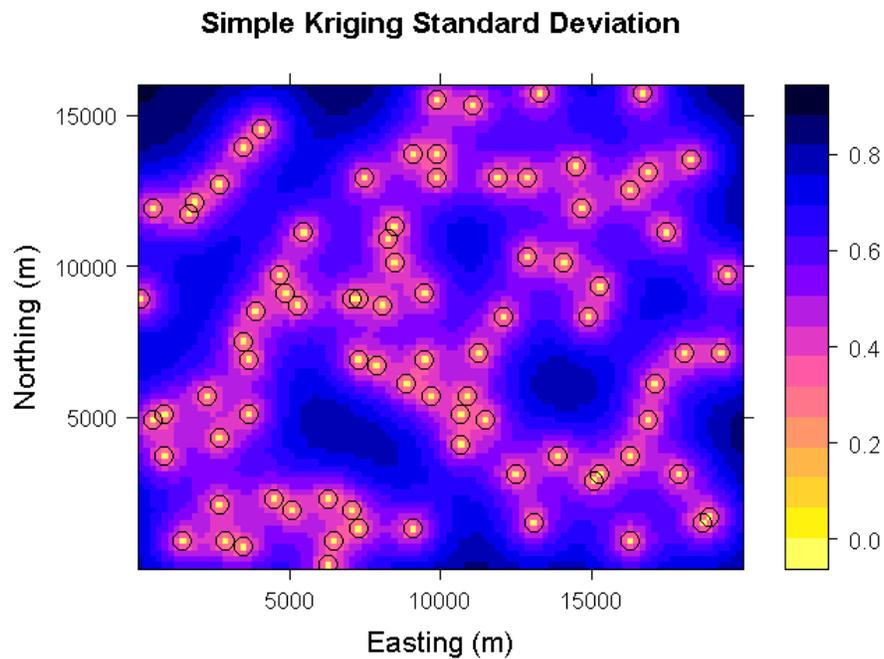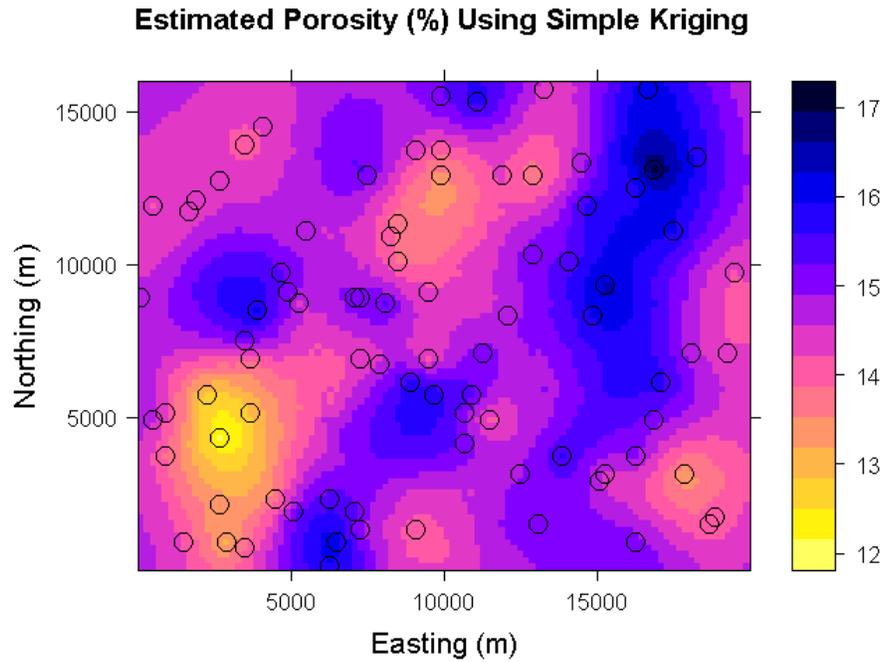
Some characteristics to note:

Smoothness: Kriged surface will basically be as smooth as possible given the constraints of the data; in many cases, probably smoother than the "true" surface.

Bullseyes: Because kriging averages between data points, local extremes will usually be at well locations; bullseyes are inevitable. This is true of almost all interpolation algorithms. Extreme form of this is artifact discontinuities at well locations when semivariogram model includes significant nugget.

Error map reflects data locations, not data values: Map of kriging standard deviation depends entirely on data configuration and covariance function; essentially a map of distance to nearest well location scaled by covariance function.

Ordinary Kriging

For ordinary kriging, rather than assuming that the mean is constant over the entire domain, we assume that it is constant in the local neighborhood of each estimation point, that is that $m(\mathbf{u}_\alpha) = m(\mathbf{u})$ for each nearby data value, $Z(\mathbf{u}_\alpha)$, that we are using to estimate $Z(\mathbf{u})$. In this case, the kriging estimator can be written

$$Z^*(\mathbf{u}) = m(\mathbf{u}) + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})[Z(\mathbf{u}_\alpha) - m(\mathbf{u})]$$

$$= \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) Z(\mathbf{u}_\alpha) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) \right] m(\mathbf{u})$$

and we filter the unknown local mean by requiring that the kriging weights sum to 1, leading to an ordinary kriging estimator of

$$Z_{OK}^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{OK}(\mathbf{u}) Z(\mathbf{u}_\alpha) \quad \text{with} \quad \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{OK}(\mathbf{u}) = 1.$$

In order to minimize the error variance subject to the unit-sum constraint on the weights, we actually set up the system minimize the error variance plus an additional term involving a Lagrange parameter, $\mu_{OK}(\mathbf{u})$:

$$L = \sigma_E^2(\mathbf{u}) + 2\mu_{OK}(\mathbf{u}) \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) \right]$$

so that minimization with respect to the Lagrange parameter forces the constraint to be obeyed:

$$\frac{1}{2} \frac{\partial L}{\partial \mu} = 1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) = 0$$

In this case, the system of equations for the kriging weights turns out to be

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{OK}(\mathbf{u}) C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mu_{OK}(\mathbf{u}) = C_R(\mathbf{u}_\alpha - \mathbf{u}) \qquad \alpha = 1,\ldots,n(\mathbf{u}) \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{OK}(\mathbf{u}) = 1 \end{cases}$$

where $C_R(\mathbf{h})$ is once again the covariance function for the residual component of the variable. In simple kriging, we could equate $C_R(\mathbf{h})$ and $C(\mathbf{h})$, the covariance function for the variable itself, due to the assumption of a constant mean. That equality does not hold here, but in practice the substitution is often made anyway, on the assumption that the semivariogram, from which $C(\mathbf{h})$ is derived, effectively filters the influence of large-scale trends in the mean.

In fact, the unit-sum constraint on the weights allows the ordinary kriging system to be stated directly in terms of the semivariogram (in place of the $C_R(\mathbf{h})$ values above). In a sense, ordinary kriging is the interpolation approach that follows naturally from a semivariogram analysis, since both tools tend to filter trends in the mean.

Once the kriging weights (and Lagrange parameter) are obtained, the ordinary kriging error variance is given by

$$\sigma_{OK}^2(\mathbf{u}) = C(0) - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{OK}(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u}) - \mu_{OK}(\mathbf{u}).$$

In matrix terms, the ordinary kriging system is an augmented version of the simple kriging system. For our six-point example it would be:

$$
\begin{bmatrix}
0.78 & 0.28 & 0.06 & 0.17 & 0.40 & 0.43 & 1.00 \\
0.28 & 0.78 & 0.43 & 0.39 & 0.27 & 0.20 & 1.00 \\
0.06 & 0.43 & 0.78 & 0.37 & 0.11 & 0.06 & 1.00 \\
0.17 & 0.39 & 0.37 & 0.78 & 0.37 & 0.27 & 1.00 \\
0.40 & 0.27 & 0.11 & 0.37 & 0.78 & 0.65 & 1.00 \\
0.43 & 0.20 & 0.06 & 0.27 & 0.65 & 0.78 & 1.00 \\
1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 0.00
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \mu
\end{bmatrix}
=
\begin{bmatrix}
0.38 \\ 0.56 \\ 0.32 \\ 0.49 \\ 0.46 \\ 0.37 \\ 1.00
\end{bmatrix}
$$

to which the solution is

$$
\begin{bmatrix}
\lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \mu
\end{bmatrix}
= \mathbf{K}^{-1}\mathbf{k} =
\begin{bmatrix}
0.1274 \\ 0.4515 \\ -0.0463 \\ 0.2595 \\ 0.2528 \\ -0.0448 \\ 0.0288
\end{bmatrix}
$$

The ordinary kriging estimate at $\mathbf{u} = (2000\ \text{m}, 4700\ \text{m})$ turns out to be 12.93% with a standard deviation of 0.490%, only slightly different from the simple kriging values of 12.83% and 0.488%.

Again using 16 nearest neighbors for each estimation point, the ordinary kriging porosity estimate and standard deviation look very much like those from simple kriging:

Estimated Porosity (%) Using Ordinary Kriging

Ordinary Kriging Standard Deviation
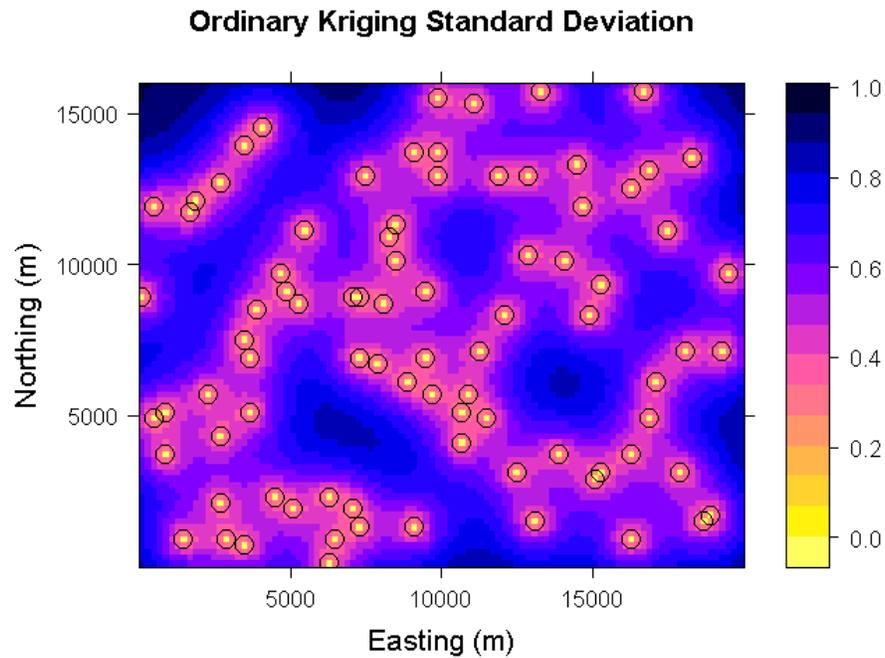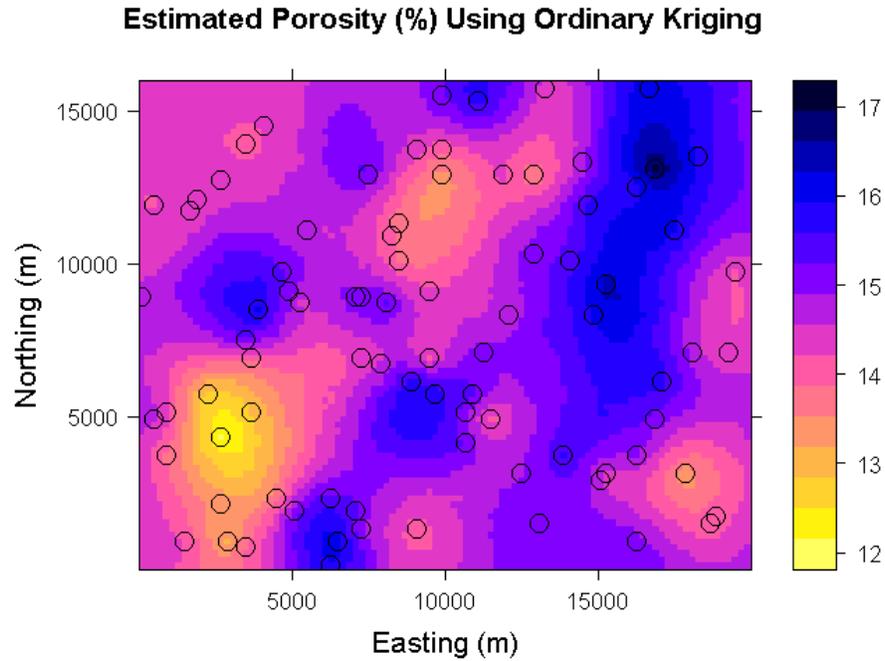
Kriging with a Trend

Kriging with a trend (the method formerly known as *universal kriging*) is much like ordinary kriging, except that instead of fitting just a local mean in the neighborhood of the estimation point, we fit a linear or higher-order trend in the (x,y) coordinates of the data points. A local linear (a.k.a., first-order) trend model would be given by

$$m(\mathbf{u}) = m(x, y) = a_0 + a_1 x + a_2 y$$

Including such a model in the kriging system involves the same kind of extension as we used for ordinary kriging, with the addition of two more Lagrange parameters and two extra columns and rows in the **K** matrix whose (non-zero) elements are the x and y coordinates of the data points. Higher-order trends (quadratic, cubic) could be handled in the same way, but in practice it is rare to use anything higher than a first-order trend. Ordinary kriging is kriging with a zeroth-order trend model.

If the variable of interest does exhibit a significant trend, a typical approach would be to attempt to estimate a "de-trended" semivariogram using one of the methods described in the semivariogram lecture and then feed this into kriging with a first-order trend. However, Goovaerts (1997) warns against this approach and instead recommends performing simple kriging of the residuals from a global trend (with a constant mean of 0) and then adding the kriged residuals back into the global trend.

A few of the many topics I have skipped or glossed over

Cokriging: Kriging using information from one or more correlated secondary variables, or multivariate kriging in general. Requires development of models for cross-covariance – covariance between two different variables as a function of lag.

Indicator Kriging: Kriging of indicator variables, which represent membership in a set of categories. Used with naturally categorical variables like facies or continuous variables that have been thresholded into categories (e.g., quartiles, deciles). Especially useful for preserving connectedness of high- and low-permeability regions. Direct application of kriging to perm will almost always wash out extreme values.

Artifact discontinuities: Kriging using a semivariogram model with a significant nugget will create discontinuities, with the interpolated surface leaping up or down to grab any data point that happens to correspond with a grid node (estimation point). Solutions: factorial kriging (filtering out the nugget component) or some other kind of smoothing (as opposed to exact) interpolation, such as smoothing splines. Or, if you really want to do exact interpolation, use a semivariogram model without a nugget.

Search algorithm: The algorithm for selecting neighboring data points can have at least as much influence on the estimate as the interpolation algorithm itself. I have used a simple nearest neighbor search. A couple of alternatives include quadrant and octant searches, which look for so many data points within a certain distance in each quadrant or octant surrounding the data point.

Stochastic Simulation

Stochastic simulation is a means for generating multiple equiprobable realizations of the property in question, rather than simply estimating the mean. Essentially, we are adding back in some noise to undo the smoothing effect of kriging. This *possibly* gives a better representation of the natural variability of the property in question and gives us a means for quantifying our uncertainty regarding what's really down there . . . or at least it gives one possible depiction of uncertainty, painted with the limited palette of whatever simulation approach we are using.

The two most commonly used forms of simulation for reservoir modeling applications are sequential Gaussian simulation for continuous variables like porosity and sequential indicator simulation for categorical variables like facies.

The basic idea of sequential Gaussian simulation (SGS) is very simple. Recall that kriging gives us an estimate of both the mean and standard deviation of the variable at each grid node, meaning we can represent the variable at each grid node as a random variable following a normal (Gaussian) distribution. Rather than choosing the mean as the estimate at each node, SGS chooses a random deviate from this normal distribution, selected according to a uniform random number representing the probability level.

For our six-point example in the porosity data, ordinary kriging gave a mean estimate of 12.93% with a standard deviation of 0.49%. In this case, if we happened to generate a uniform random number of $p = 0.665$ for this grid node, then the assigned porosity would be 13.14%, the corresponding value of the cumulative normal probability function:
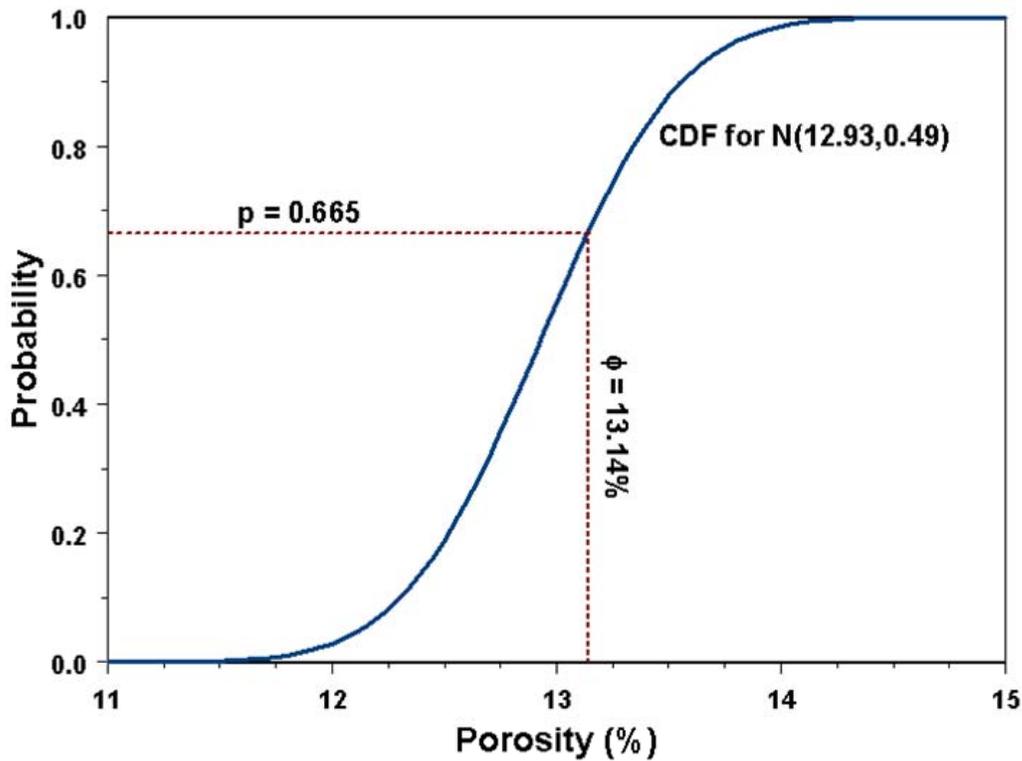
So, the basic steps in the SGS process are:

- Generate a random path through the grid nodes

- Visit the first node along the path and use kriging to estimate a mean and standard deviation for the variable at that node based on surrounding data values

- Select a value at random from the corresponding normal distribution and set the variable value at that node to that number

- Visit each successive node in the random path and repeat the process, including previously simulated nodes as data values in the kriging process

We use a random path to avoid artifacts induced by walking through the grid in a regular fashion. We include previously simulated grid nodes as "data" in order to preserve the proper covariance structure between the simulated values.

Sometimes SGS is implemented in a "multigrid" fashion, first simulating on a coarse grid (a subset of the fine grid – maybe every 10th grid node) and then on the finer grid (maybe with an intermediate step or two) in order to reproduce large-scale semivariogram structures. Without this, the "screening" effect of kriging quickly takes over as the simulation progresses and nodes get filled in, so that most nodes are conditioned only on nearby values, so that small-scale structure is reproduced better than large-scale structure.

For SGS, it is important that the data actually follow a Gaussian distribution. If they do not, we can use a normal score transform:
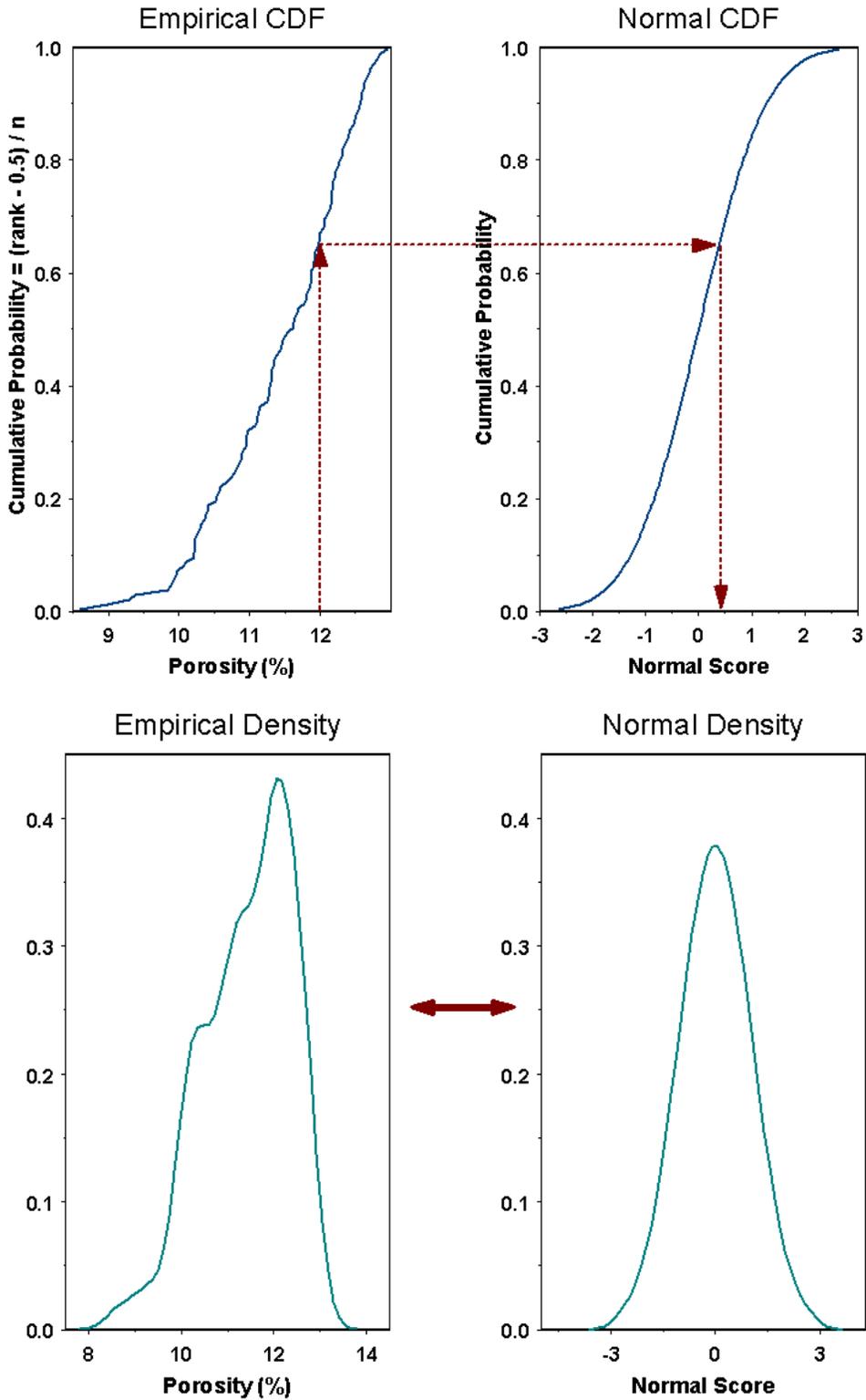
Here are six sequential Gaussian simulations of our porosity data, using the spherical semivariogram model and a 16 nearest-neighbor search:


SGS Porosity (%)
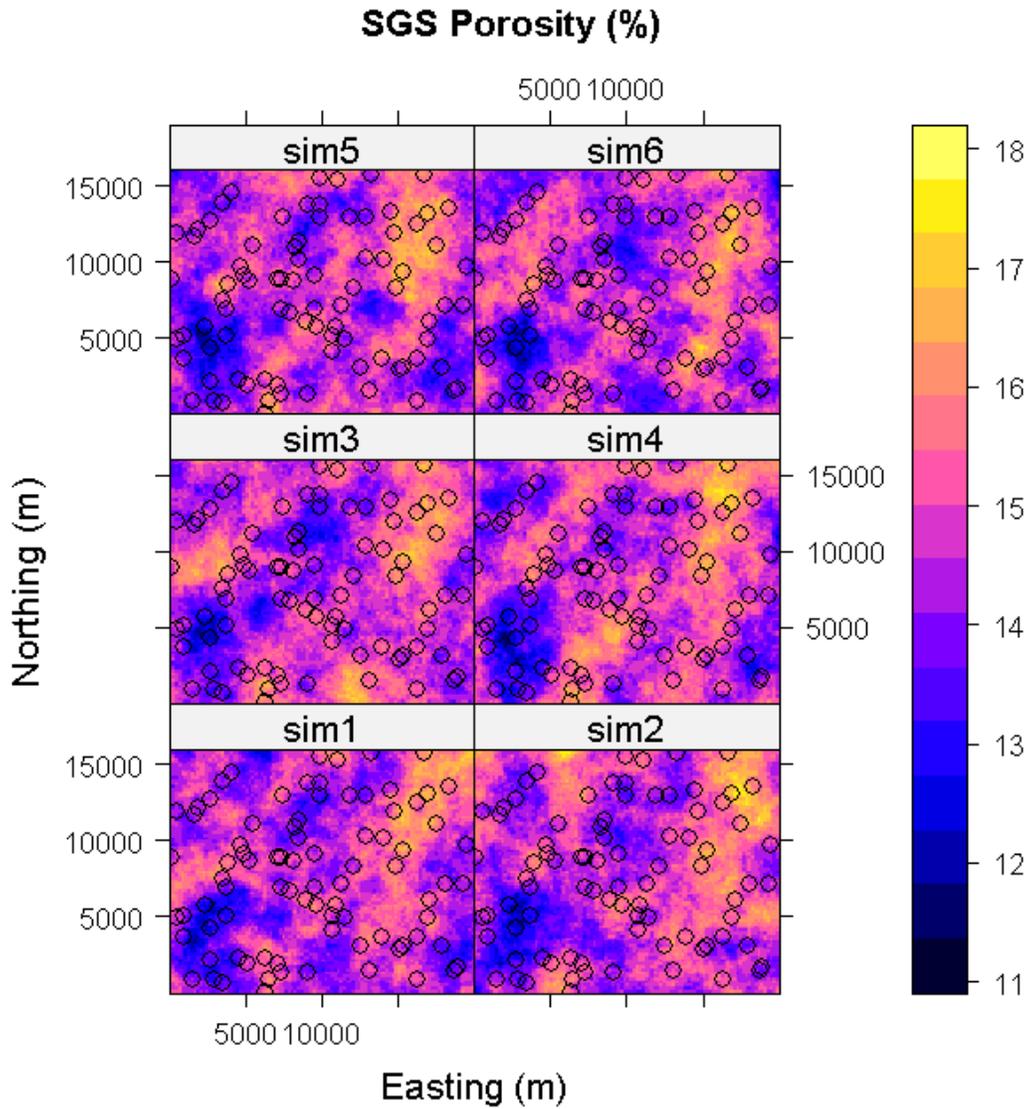
Sequential indicator simulation (SIS) is very similar to sequential Gaussian simulation, expect that indicator kriging is used to build up a discrete cumulative density function for the individual categories at each node and the node is assigned a category selected at random from this discrete CDF.

Briefly, an indicator representation for a categorical variable such as facies would be formulated as

$$i(\mathbf{u}_\alpha; k) = \begin{cases} 1 & \text{if facies } k \text{ is present at } \mathbf{u}_\alpha \\ 0 & \text{otherwise} \end{cases}$$

where you would have one indicator variable for each of the K different facies. We can then use kriging (based on indicator semivariograms) to produce a set of facies membership probabilities at each grid point, build up a CDF from the probabilities, and select a facies at random from the CDF:
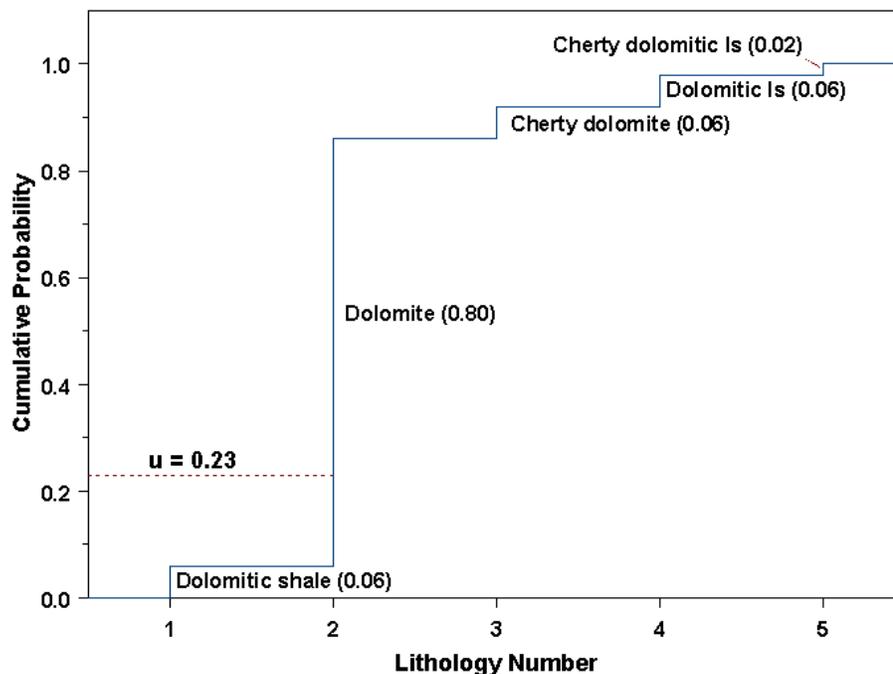
For a continuous variable such as permeability, indicator variables are built by comparing data values to a set of thresholds, z_k:

$$i(\mathbf{u}_\alpha;k) = \begin{cases} 1 & \text{if } z(\mathbf{u}_\alpha) \leq z_k \\ 0 & \text{otherwise} \end{cases}$$

We might define thresholds, for example, at the 10th, 25th, 50th, 75th, and 90th percentiles of the data distribution, for example. In this case, kriging the indicator values for the kth threshold, z_k, gives estimates of $P[Z(\mathbf{u}) \leq z_k]$ at each estimation point. Since this already is a cumulative probability, we don't need to go through the process of summing to get a CDF, although we will need to correct any violations of the expected order relationships, $P[Z(\mathbf{u}) \leq z_k] \leq P[Z(\mathbf{u}) \leq z_{k+1}]$, that happen to occur.

In this case, SIS assigns each node to a corresponding range (e.g., upper 10%) by random selection from the CDF and the resulting indicator vector gets added to the conditioning data for the remaining nodes.

Typical Reservoir or Aquifer Modeling Workflow

Basically, work from large-scale structure to small-scale structure, and generally from more deterministic methods to more stochastic methods:

- Establish large-scale geologic structure, for example, by deterministic interpolation of formation or unit tops; this creates a set of distinct zones

- Within each zone, use SIS or some other discrete simulation technique (such as object-based simulation) to generate realizations of the facies distribution – the primary control on the porosity & permeability distributions

- Within each facies, use SGS (or similar) to generate porosity distirubtion and then simulate permeability distribution conditional to porosity distribution, assuming there is some relationship between the two

Porosity and facies simulations could be conditioned to other secondary data, such as seismic. Methods also exist for conditioning to well test and production data, but these are fairly elaborate and *probably* not in very common use as yet. More typical (maybe) to run flow simulations after the fact and rank realizations by comparison to historical production & well tests.

For more on reservoir (or aquifer) modeling workflow and assessment of uncertainty from multiple realizations, see

C. V. Deutsch, 2002, *Geostatistical Reservoir Modeling*, Oxford Univeristy Press, 376 pp.