# KRIGING

C&PE 940, 19 October 2005

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

Overheads and other resources available at:

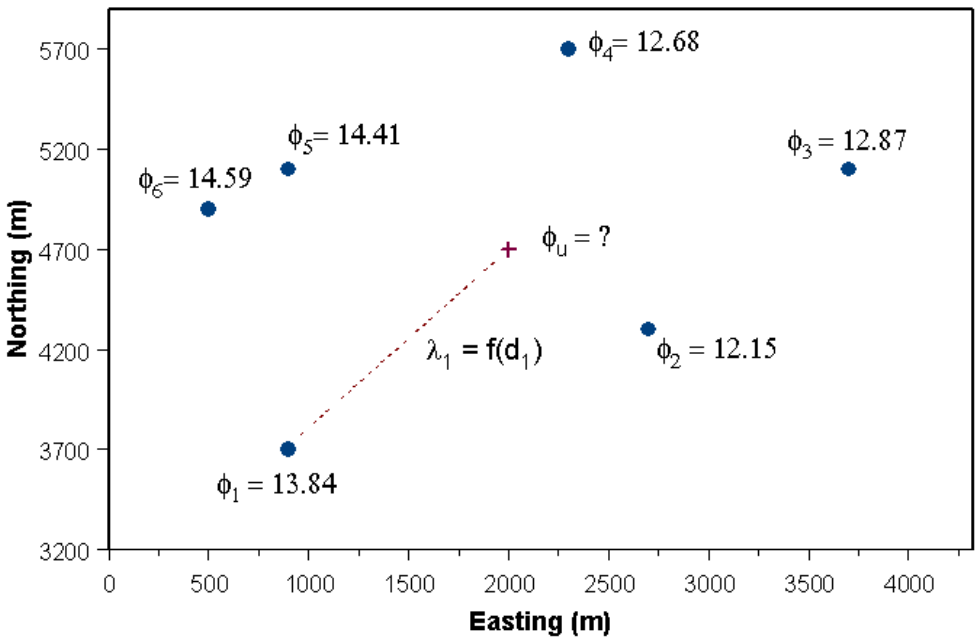
http://people.ku.edu/~gbohling/cpe940

## 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 **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 $f_u$ by a weighted average $\sum l_a f_a$, with weights $l_a$ given by some decreasing function of the distance, $d_a$, from **u** to data point $a$.

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

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{a=1}^{n(\mathbf{u})} l_a \left[ Z(\mathbf{u}_a) - m(\mathbf{u}_a) \right] .”$$

with

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

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

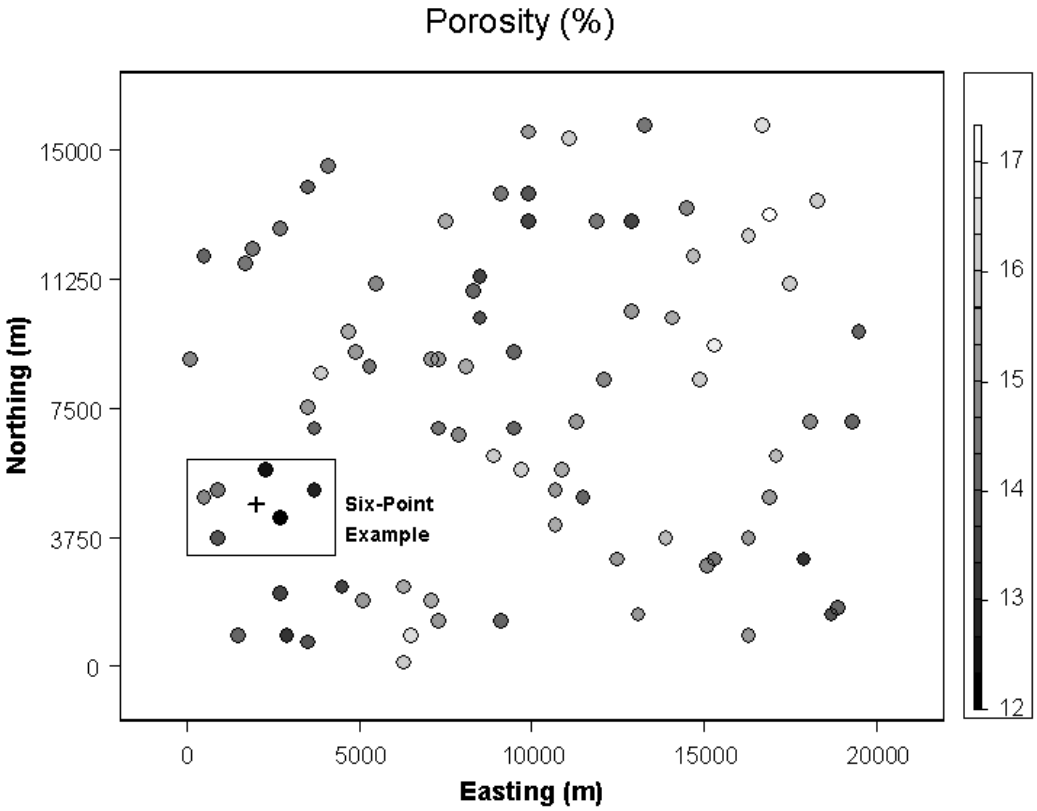
$m(\mathbf{u}), m(\mathbf{u}_a)$: expected values (means) of $Z(\mathbf{u})$ and $Z(\mathbf{u}_a)$

$l_a(\mathbf{u})$: kriging weight assigned to datum $z(\mathbf{u}_a)$ 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, $l_a$, 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:



## Basics of Kriging

Again, the basic form of the kriging estimator is

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{a=1}^{n(\mathbf{u})} l_a \left[ Z(\mathbf{u}_a) - m(\mathbf{u}_a) \right]$$

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

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

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

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, $\mathbf{u}$):

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

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

The residual covariance function is generally derived from the input semivariogram model, $C_R(\mathbf{h}) = C_R(\mathbf{0}) - g(\mathbf{h}) = Sill - g(\mathbf{h})$. 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

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

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

$$\begin{aligned} Z^*_{SK}(\mathbf{u}) - Z(\mathbf{u}) &= \left[Z^*_{SK}(\mathbf{u}) - m\right] - \left[Z(\mathbf{u}) - m\right] \\ &= \sum_{a=1}^{n(\mathbf{u})} l_a^{SK}(\mathbf{u}) R(\mathbf{u}_a) - R(\mathbf{u}) = R^*_{SK}(\mathbf{u}) - R(\mathbf{u}) \end{aligned}$$

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

$$\begin{aligned} s^2_E(\mathbf{u}) &= \mathrm{Var}\left\{R^*_{SK}(\mathbf{u})\right\} + \mathrm{Var}\left\{R_{SK}(\mathbf{u})\right\} - 2\,\mathrm{Cov}\left\{R^*_{SK}(\mathbf{u}), R_{SK}(\mathbf{u})\right\} \\ &= \sum_{a=1}^{n(\mathbf{u})} \sum_{b=1}^{n(\mathbf{u})} l_a^{SK}(\mathbf{u}) l_b^{SK}(\mathbf{u}) C_R(\mathbf{u}_a - \mathbf{u}_b) + C_R(0) - 2\sum_{a=1}^{n(\mathbf{u})} l_a^{SK}(\mathbf{u}) C_R(\mathbf{u}_a - \mathbf{u}) \end{aligned}$$

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:

$$\sum_{b=1}^{n(\mathbf{u})} l_b^{SK}(\mathbf{u}) C_R(\mathbf{u}_a - \mathbf{u}_b) = C_R(\mathbf{u}_a - \mathbf{u}) \qquad a = 1, \ldots, n(\mathbf{u})$$

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 $C(\mathbf{h})$:

$$\sum_{b=1}^{n(\mathbf{u})} l_b^{SK}(\mathbf{u})\, C(\mathbf{u}_a - \mathbf{u}_b) = C(\mathbf{u}_a - \mathbf{u}) \qquad a = 1, \ldots, n(\mathbf{u}).$$

This can be written in matrix form as

$$\mathbf{K}\, \mathrm{l}_{SK}(\mathbf{u}) = \mathbf{k}$$

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 $\mathrm{l}_{SK}(\mathbf{u})$ 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

$$\mathrm{l}_{SK} = \mathbf{K}^{-1}\mathbf{k}$$

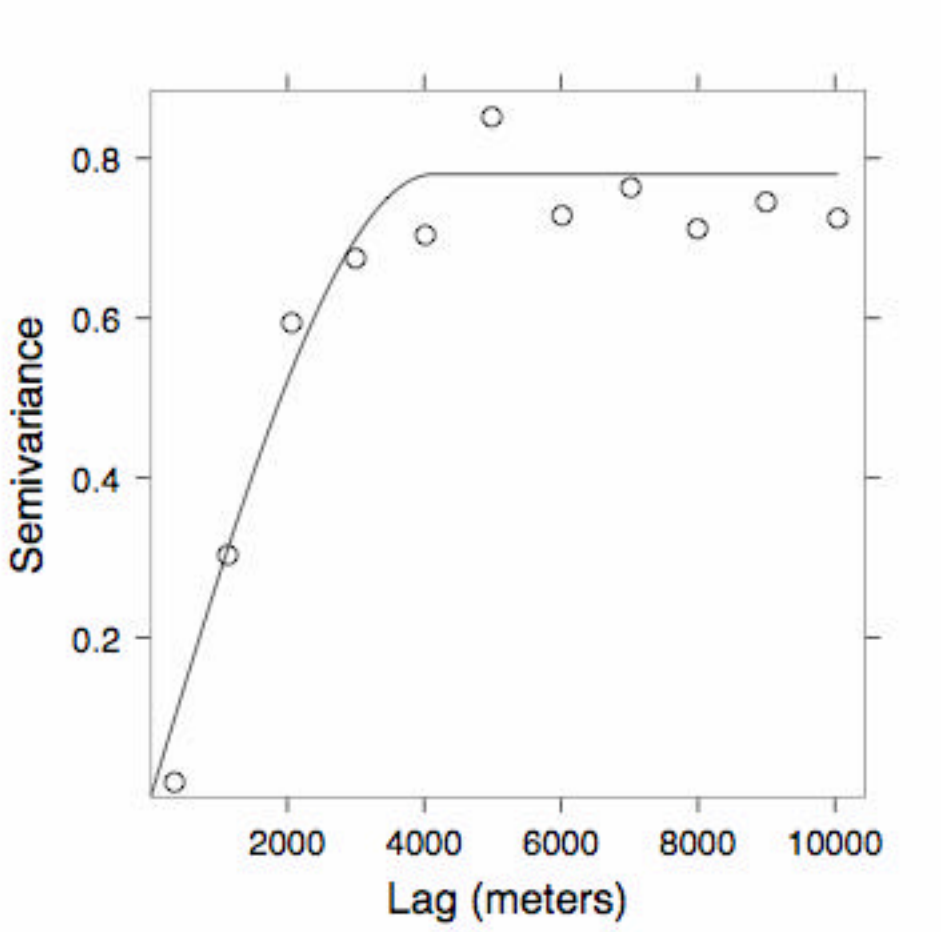
Once we have the kriging weights, we can compute both the kriging estimate and the kriging variance, which is given by

$$s_{SK}^2(\mathbf{u}) = C(0) - \mathrm{l}_{SK}^T(\mathbf{u})\,\mathbf{k} = C(0) - \sum_{a=1}^{n(\mathbf{u})} l_a^{SK}(\mathbf{u})\, C(\mathbf{u}_a - \mathbf{u})$$

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 $\mathbf{K}^{-1}$ (on the left) will downweight points falling in clusters relative to isolated points at the same distance.

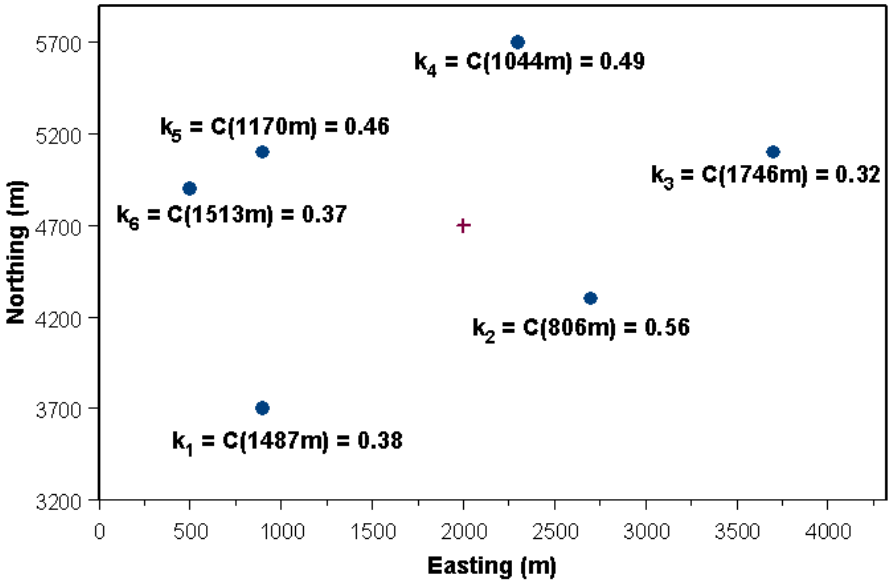
We will apply simple kriging to our porosity data, using the spherical semivariogram that we fit before, with zero nugget, a sill of 0.78, and a range of 4141 m:



Since we are using a spherical semivariogram, the covariance function is given by

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

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

$$\begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \\ l_6 \end{bmatrix} = \mathbf{K}^{-1}\mathbf{k} = \begin{bmatrix} 0.1475 \\ 0.4564 \\ -0.0205 \\ 0.2709 \\ 0.2534 \\ -0.0266 \end{bmatrix}$$

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.

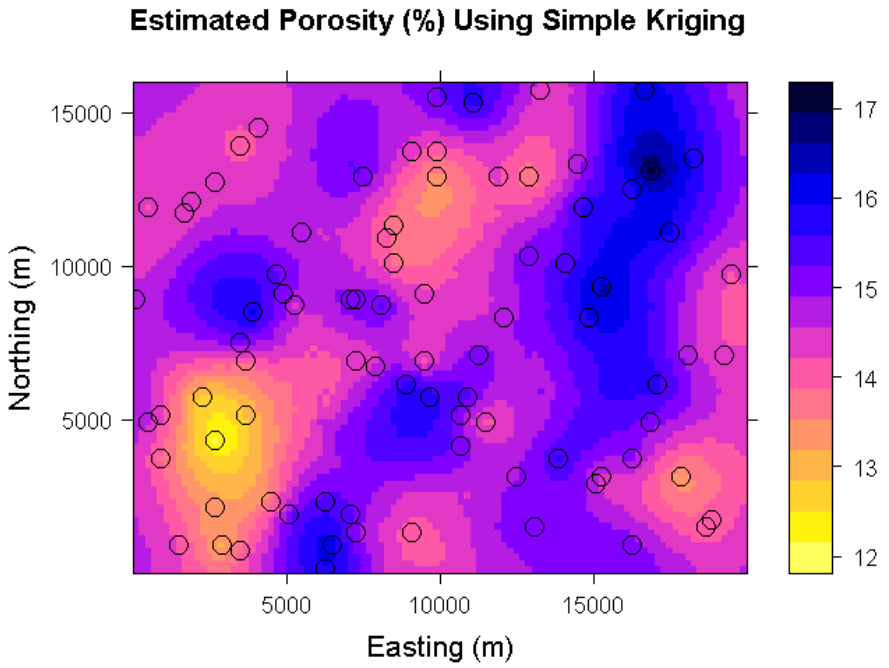
The mean porosity value for the 85 wells is 14.70%, and the porosity values at the six example wells are 13.84%, 12.15%, 12.87%, 12.68%, 14.41%, and 14.59%. The estimated residual from the mean at **u** is given by the dot product of the kriging weights and the vector of residuals at the data points:

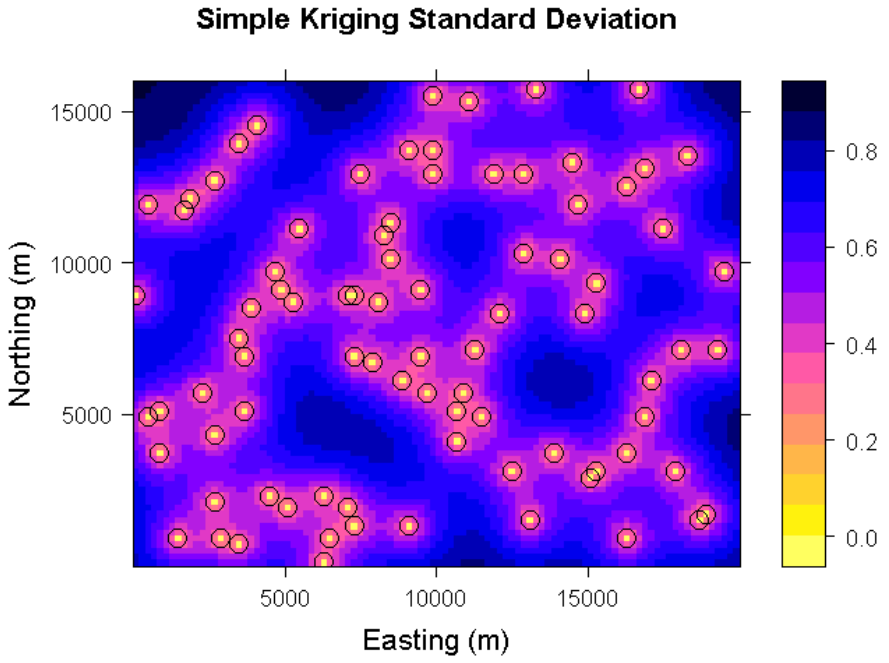
$$R(\mathbf{u}) = \mathbf{?}' \mathbf{R}_a$$

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

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):



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 $m(\mathbf{u}_a) = m(\mathbf{u})$ for each nearby data value, $Z(\mathbf{u}_a)$, that we are using to estimate $Z(\mathbf{u})$. In this case, the kriging estimator can be written

$$Z^*(\mathbf{u}) = m(\mathbf{u}) + \sum_{a=1}^{n(\mathbf{u})} l_a(\mathbf{u})[Z(\mathbf{u}_a) - m(\mathbf{u})]$$
$$= \sum_{a=1}^{n(\mathbf{u})} l_a(\mathbf{u}) Z(\mathbf{u}_a) + \left[1 - \sum_{a=1}^{n(\mathbf{u})} l_a(\mathbf{u})\right] m(\mathbf{u})$$

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

$$Z^*_{OK}(\mathbf{u}) = \sum_{a=1}^{n(\mathbf{u})} l_a^{OK}(\mathbf{u}) Z(\mathbf{u}_a) \quad \text{with} \quad \sum_{a=1}^{n(\mathbf{u})} l_a^{OK}(\mathbf{u}) = 1.$$

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, $m_{OK}(\mathbf{u})$:

$$L = s_E^2(\mathbf{u}) + 2m_{OK}(\mathbf{u})\left[1 - \sum_{a=1}^{n(\mathbf{u})} l_a(\mathbf{u})\right]$$

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

$$\frac{1}{2}\frac{\partial L}{\partial m} = 1 - \sum_{a=1}^{n(\mathbf{u})} l_a(\mathbf{u}) = 0$$

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

$$\begin{cases} \sum_{b=1}^{n(\mathbf{u})} l_b^{OK}(\mathbf{u})\, C_R(\mathbf{u}_a - \mathbf{u}_b) + m_{OK}(\mathbf{u}) = C_R(\mathbf{u}_a - \mathbf{u}) \qquad a = 1,\ldots,n(\mathbf{u}) \\ \sum_{b=1}^{n(\mathbf{u})} l_b^{OK}(\mathbf{u}) = 1 \end{cases}$$

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

$$s_{OK}^2(\mathbf{u}) = C(0) - \sum_{a=1}^{n(\mathbf{u})} l_a^{OK}(\mathbf{u})\, C(\mathbf{u}_a - \mathbf{u}) - m_{OK}(\mathbf{u}).$$

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

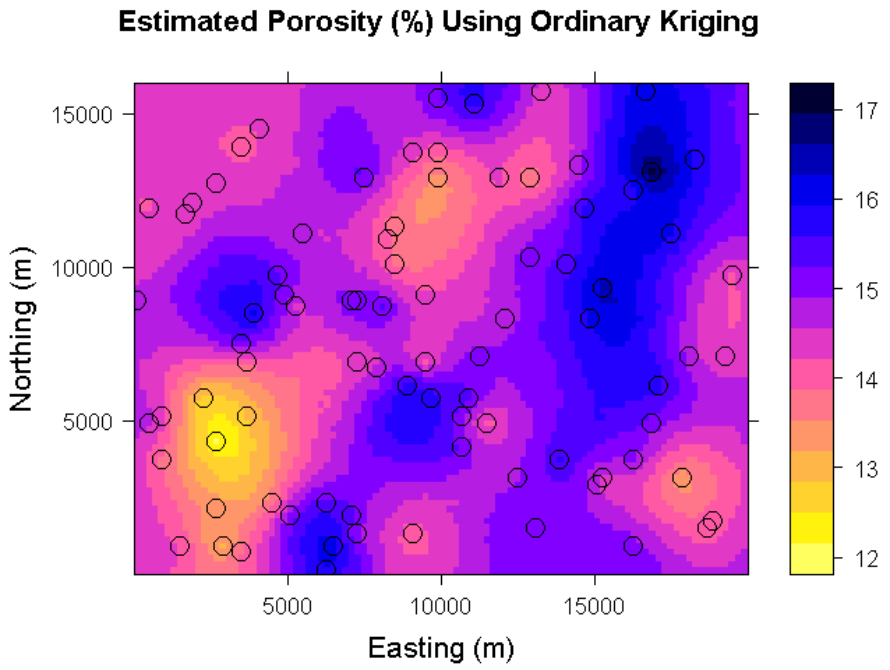
$$\begin{bmatrix} 0.78 & 0.28 & 0.06 & 0.17 & 0.40 & 0.43 & 1.00 \\ 0.28 & 0.78 & 0.43 & 0.39 & 0.27 & 0.20 & 1.00 \\ 0.06 & 0.43 & 0.78 & 0.37 & 0.11 & 0.06 & 1.00 \\ 0.17 & 0.39 & 0.37 & 0.78 & 0.37 & 0.27 & 1.00 \\ 0.40 & 0.27 & 0.11 & 0.37 & 0.78 & 0.65 & 1.00 \\ 0.43 & 0.20 & 0.06 & 0.27 & 0.65 & 0.78 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 0.00 \end{bmatrix} \begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \\ l_6 \\ m \end{bmatrix} = \begin{bmatrix} 0.38 \\ 0.56 \\ 0.32 \\ 0.49 \\ 0.46 \\ 0.37 \\ 1.00 \end{bmatrix}$$

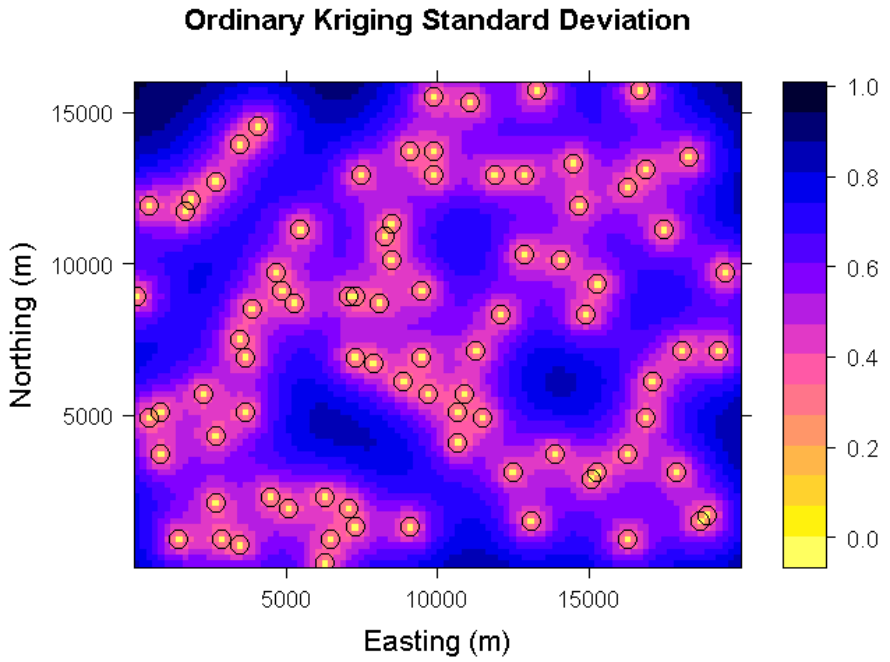
to which the solution is

$$\begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \\ l_6 \\ m \end{bmatrix} = \mathbf{K}^{-1}\mathbf{k} = \begin{bmatrix} 0.1274 \\ 0.4515 \\ -0.0463 \\ 0.2595 \\ 0.2528 \\ -0.0448 \\ 0.0288 \end{bmatrix}$$

The ordinary kriging estimate at **u** = (2000 m, 4700 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:



## 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.