

## Semivariograms and Kriging

Dr. Zhi Wang  
Department of Earth and Environmental Sciences  
California State University, Fresno

### Descriptive

- Similar to autocorrelations, but here we **focus on the difference between pairs of observations separated by a lag  $h$** .
- Although measurements taken very close to each other tend to be similar, variations are generally highly irregular.

### Relevant questions

- Should samples be taken randomly or at uniform, irregular intervals?
- How close together should samples be taken?
- Do variations differ with direction of sampling?
- Scale variant or invariant?
- How many samples are needed for a reliable semivariogram?

### Semivariance

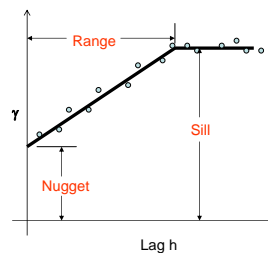
Semivariance:

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

$N(h)$  is the total number of pairs of observations separated by  $h$ .

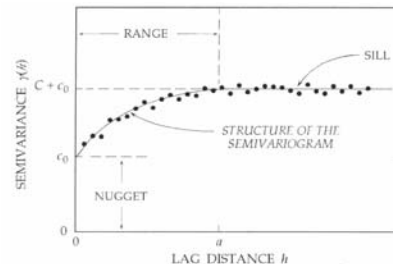
### Semivariograms

- **Nugget variance:** a non-zero value for  $\gamma$  when  $h = 0$ . Produced by various sources of unexplained error, or spatial variations occurring over smaller distances than the smallest sampling interval.
- **Sill:** for large values of  $h$  the variogram levels out, indicating that there **no longer is any correlation between data points**. The sill should be equal to the variance of the data set.
- **Range:** is the value of  $h$  where the sill occurs (or 95% of the value of the sill). Mirror image of correlation length.

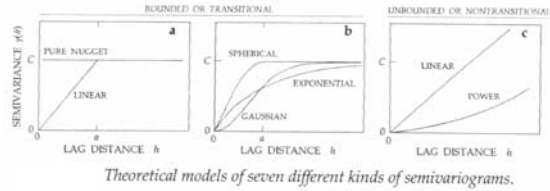


### Sample size

- In general, **30 or more points are needed** to generate a reasonable sample variogram.
- **The most important part of a variogram is its shape near the origin** as closest points are given more weight in the interpolation process.



## Variogram models



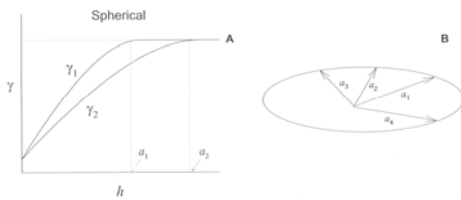
## Variogram models

Equations for the seven semivariogram models

Model	Equation	Consideration
<i>Bounded or Transitional</i>		
Pure nugget	$\gamma(h) = \begin{cases} 0 & h = 0 \\ C & h > 0 \end{cases}$	$h = 0$ $h > 0$
Linear	$\gamma(h) = \begin{cases} Ch/a & 0 \leq h \leq a \\ C & h > a \end{cases}$	$0 \leq h \leq a$ $h > a$
Spherical	$\gamma(h) = \begin{cases} C \left[ \frac{3h}{2a} - \frac{1}{2} \left( \frac{h}{a} \right)^3 \right] & 0 \leq h \leq a \\ C & h > a \end{cases}$	$0 \leq h \leq a$ $h > a$
Exponential	$\gamma(h) = C[1 - \exp(-h/a)]$	$h \geq 0$
Gaussian	$\gamma(h) = C[1 - \exp[-(h/a)^2]]$	$h \geq 0$
<i>Unbounded or Nontransitional</i>		
Linear	$\gamma(h) = mh$	$h \geq 0$
Power	$\gamma(h) = mh^\beta$	$h \geq 0; 1 < \beta < 2$

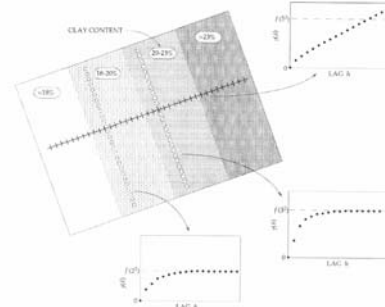
## Directional dependence

- There may be higher spatial autocorrelation in one direction than in others, which is called anisotropy:



- Anisotropy is incorporated in the variogram model by means of a linear transformation.

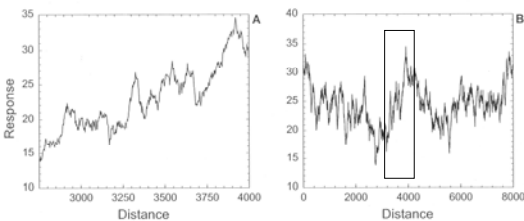
## Directional dependence



Semivariograms derived from soil samples taken along three transects in a field having nonuniform distributions of clay content.

## Drift (local trend)

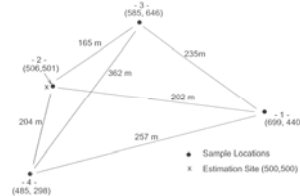
- Some attributes change in a systematic manner across the landscape:



In A there is an apparent trend in the values (changing mean). In B more data was collected and drift is no longer apparent (scale dependency, knowledge of the data).

## Kriging

- Kriging is based on the idea that you can make inferences about a random function  $Z(x)$  given data points  $Z(x_1), Z(x_2), \dots, Z(x_n)$ .



$$Z(x) = m(x) + \gamma(h) + \varepsilon''$$

3 components: structural (constant mean), random spatially correlated component and residual error.

## Inference

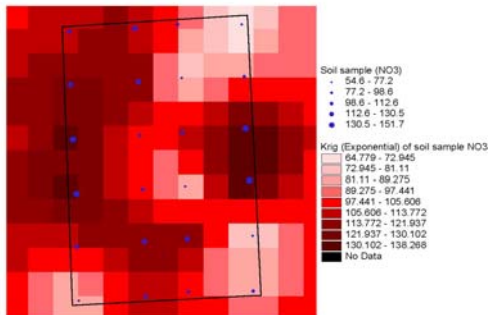
- Kriging produces **the best linear unbiased estimate** of an attribute at an unmeasured site, **knowing only the variogram**.
- **Ordinary kriging**: Most robust and most used
- **Lognormal kriging**: Ordinary kriging of the logarithms of the measured values (strongly positively skewed data that approximate a lognormal distribution)
- **Universal kriging** accounts for drift (local trend in variograms).
- **Punctual kriging**: produces values for non-sampled points.
- **Block kriging**: gives values for inferred areas instead of inferred points. Estimates for blocks have lower variance because of the averaging of the small scale fluctuations of the function  $Z(x)$  over the area of the block.
- **Co-kriging**: uses 2 or more variables that are correlated between themselves in the estimation of values for one of them (e.g: soil bulk density and soil water content).

## Examples from ArcView

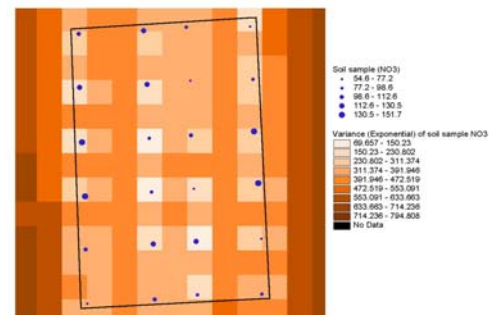


ESRI web site: Kriging interpolator 3.2 for ArcView Spatial Analyst

## Example: kriging results



## Example: variance map



## Computer exercise

- Use GEOSTAT software to produce semivariograms of soil attributes (data provided in the Excel data-file).
- Determine a best fit model of the semivariogram.
- Get data and redraw the semivariograms in Excel, print results.