

COA 616 Geostatistics in Environmental Sciences

## Lecture 3 Semivariogram

Wei Wu  
September 25, 2018

### Data posting

- A map on which each data location is plotted along with its corresponding data value.

## Spatial continuity – h-scatterplot

An h-scatterplot shows all possible pairs of data values whose locations are separated by a certain distance in a particular direction.

$$Z(s) - Z(s+h)$$

The x-coordinate of a point corresponds to the V value at a particular location and y-coordinate to the V value at a distance and direction **h** away.

The shape of the cloud of points on an h-scatterplot shows how continuous the data values are over a certain distance in a particular direction.

## Spatial continuity cont.

**h**: straight line distance

$\Theta$ : direction (angle) from one point to another  $\Theta=0$  is north, clockwise increase

**h** or  $\vec{h}$ : vector (distance and direction) separating two locations

head: the end of the vector

tail: the beginning of the vector

H-scatterplot is a plot of the head value vs. tail value for all pairs of data points that are separated by  $\vec{h}$ .

$$h_{12} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

$$\theta_{12} = \tan^{-1} \frac{y_2 - y_1}{x_2 - x_1} \quad (x_2 > x_1)$$

$$\theta_{12} = \tan^{-1} \frac{y_2 - y_1}{x_2 - x_1} + 180 \quad (x_2 < x_1)$$

## Spatial continuity cont.

- Indicator for dispersion of cloud on an h-scatterplot

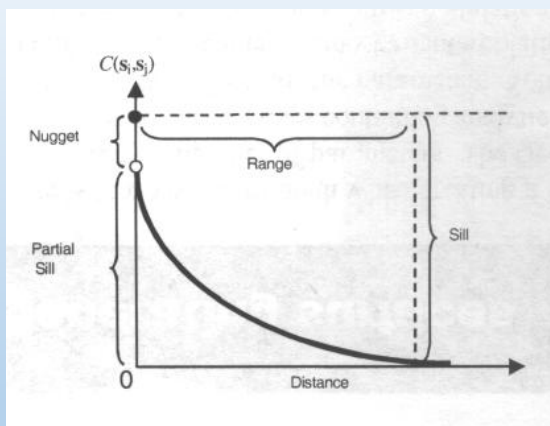
Covariance: The relation between the covariance of an h-scatterplot and  $\mathbf{h}$  is covariance function.

Correlation coefficient: The relation between the coefficient correlation of an h-scatterplot and  $\mathbf{h}$  is called correlation function or correlogram.

Moment of inertia:  $\frac{1}{2n} \sum_{i=1}^n (z(\vec{s}_i) - z(\vec{s}_i + \vec{h}))^2$

Semivariogram: The relationship between the moment of inertia of an h-scatterplot and  $\mathbf{h}$ .

## Spatial covariance



$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$\hat{c}(\vec{h}) = \frac{1}{m(\vec{h})} \sum_{i=1}^{m(\vec{h})} (z(\vec{s}_i) - \bar{z}_1)(z(\vec{s}_i + \vec{h}) - \bar{z}_2)$$

$\hat{c}(\vec{h})$  : spatial covariance

$m(\vec{h})$  : # of pairs of points separated by  $\vec{h}$

$z(\vec{s}_i)$  : data value at location  $\vec{s}_i$  (tail value)

$z(\vec{s}_i + \vec{h})$  : data value at location  $\vec{s}_i + \vec{h}$  (head value)

$$\bar{z}_1 = \frac{1}{m(\vec{h})} \sum_{i=1}^{m(\vec{h})} z(\vec{s}_i) : \text{mean of tail value}$$

$$\bar{z}_2 = \frac{1}{m(\vec{h})} \sum_{i=1}^{m(\vec{h})} z(\vec{s}_i + \vec{h}) : \text{mean of head value}$$

## Spatial Covariance cont.

$$\hat{c}(0) = \frac{1}{m(0)} \sum_{i=1}^{m(0)} (z(\vec{s}_i) - \bar{z}_1)(z(\vec{s}_i + 0) - \bar{z}_2)$$

$$m(0) = n$$

$$\bar{z}_1 = \bar{z}_2 = \bar{z}$$

$$\text{So } \hat{c}(0) = \frac{1}{n} \sum_{i=1}^n (z(\vec{s}_i) - \bar{z})(z(\vec{s}_i) - \bar{z}) \cong \text{var}(z) \frac{n-1}{n}$$

Assume stationarity if 1) the mean value does not vary significantly in space

2) the separation  $\vec{h}$  is small relative to the size of the study area

then  $\bar{z}_1 \cong \bar{z}_2 \cong \bar{z}$  thus

$$\hat{c}(\vec{h}) = \frac{1}{m(\vec{h})} \sum_{i=1}^{m(\vec{h})} (z(\vec{s}_i) - \bar{z})(z(\vec{s}_i + \vec{h}) - \bar{z}) = \frac{1}{m(\vec{h})} \sum_{i=1}^{m(\vec{h})} [(z(\vec{s}_i)z(\vec{s}_i + \vec{h})) - \bar{z}^2]$$

## Correlation function

Normalize the spatial covariance by dividing standard deviations

$$\hat{\rho}(\vec{h}) = \frac{\hat{c}(\vec{h})}{sd_1 sd_2}$$

$sd_1$  : standard deviation of  $z(\vec{s}_i)$  values (tails)

$sd_2$  : standard deviation of  $z(\vec{s}_i + \vec{h})$  values (heads)

$$sd_1 = \sqrt{\frac{1}{m(\vec{h})-1} \sum_{i=1}^{m(\vec{h})} (z(\vec{s}_i) - \bar{z}_1)^2}$$

Correlogram is a plot of  $\hat{\rho}(\vec{h})$  vs.  $\vec{h}$ , should be similar in form to  $\hat{c}(\vec{h})$  plot.

Simpler form if  $\bar{z}_1 \cong \bar{z}_2 \cong \bar{z}$  and  $s_1 = s_2 = s_z$  then

$$\hat{\rho}(\vec{h}) = \frac{\hat{c}(\vec{h})}{sd_z^2}$$

(semi)Variogram: Half of the average squared difference between the paired data values – measures dissimilarity of values separated by a vector

$$\hat{\gamma}(\vec{h}) = \frac{1}{2m(\vec{h})} \sum_{i=1}^{m(\vec{h})} (z(\vec{s}_i) - z(\vec{s}_i + \vec{h}))^2$$

If  $z(\vec{s}_i) = z(\vec{s}_i + \vec{h}) \Rightarrow \hat{\gamma}(\vec{h}) = 0$

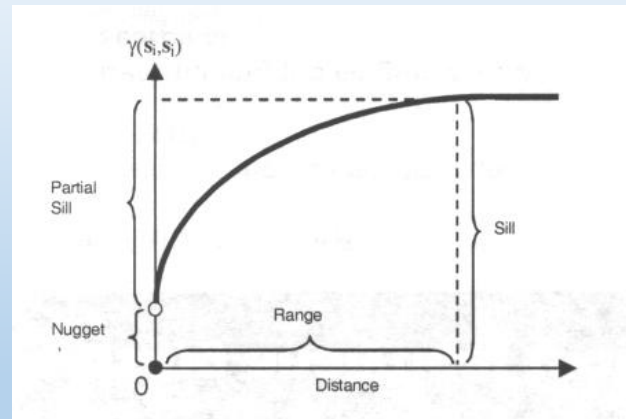
If  $z(\vec{s}_i) - z(\vec{s}_i + \vec{h})$  is large  $\Rightarrow \hat{\gamma}(\vec{h})$  is large

In principle  $\hat{\gamma}(\vec{h}) = 0$  at  $\vec{h} = 0$

In practice  $\hat{\gamma}(\vec{h})$  rarely converges to 0 as  $\vec{h} \rightarrow 0$

Nugget effect : 1) Sampling error

2) Spatial variation at scales smaller than the smallest  $\vec{h}$



## Medogram; Rodegram

Semivariances are sensitive to extreme values,  
medogram and rodegram are alternatives to semivariogram

Define a family of variogram

$$\hat{\gamma}_w(\vec{h}) = \frac{1}{2m(\vec{h})} \sum_{i=1}^{m(\vec{h})} |z(\vec{s}_i) - z(\vec{s}_i + \vec{h})|^w$$

$w \in (0, 2]$

$w = 2$  : semivariogram

$w = 1$  : medogram

$w = 1/2$  : rodegram

## Geometric interpretation of h-scatterplot in relation to semivariance

1. What if all data fall on the 1:1 line in h-scatterplot?

$$\hat{\gamma}(\vec{h}) = 0$$

2. How far is any point from the 1:1 line?

Semivariance = average of squared distance of points to the 1:1 line on an h-scatterplot

## Variogram cloud/map

1. Isotropic
2. Anisotropic