School of Geography and Earth Science McMaster University

Advanced Topics in Spatial Statistics

Spatially Continuous Data III & IV

This session:

- **Spatially Continuous Data**
  - Exploration: Second Order Properties
    - Covariogram, correlogram and variogram
  - Modeling spatially continuous data
    - Trend surface analysis

Spatially continuous data: Visualization of first order effects

- Visualization
  - Symbol maps
  - Indicator maps
  - Grayscale maps
  - Proportional symbol maps
  - Colors
  - ...

Spatially continuous data: Exploration of first order effects

- Exploration
  - Moving averages
  - Kernel estimation
  - Tesselations
    - Gradient maps
    - Contour maps

Second order effects: Smaller scale variation

- Point patterns
  - Nearest-neighbor distance analysis

- Spatially continuous data
  - Covariation of neighboring observations

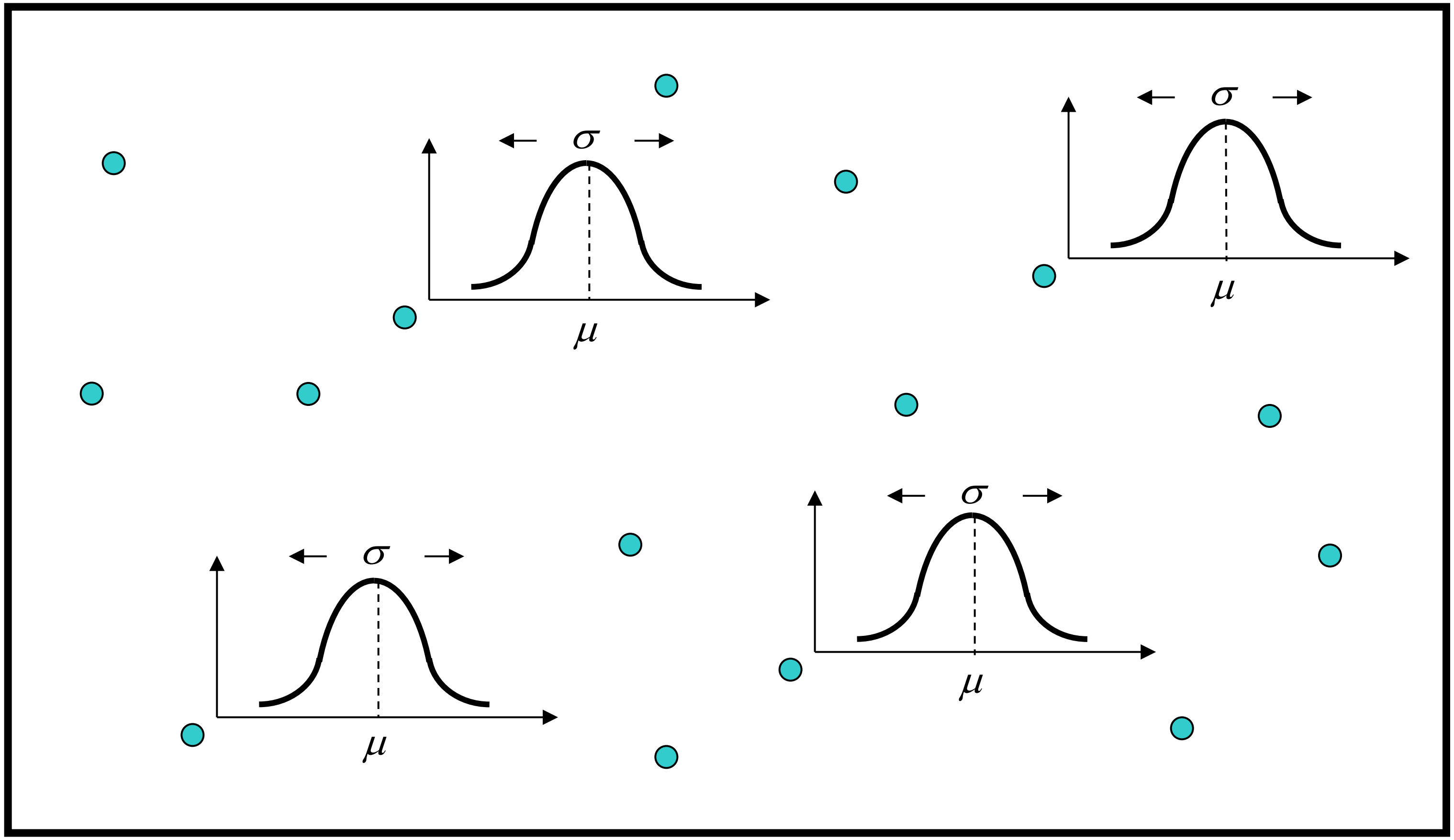
Definitions

- Mean and variance of a spatial random variable

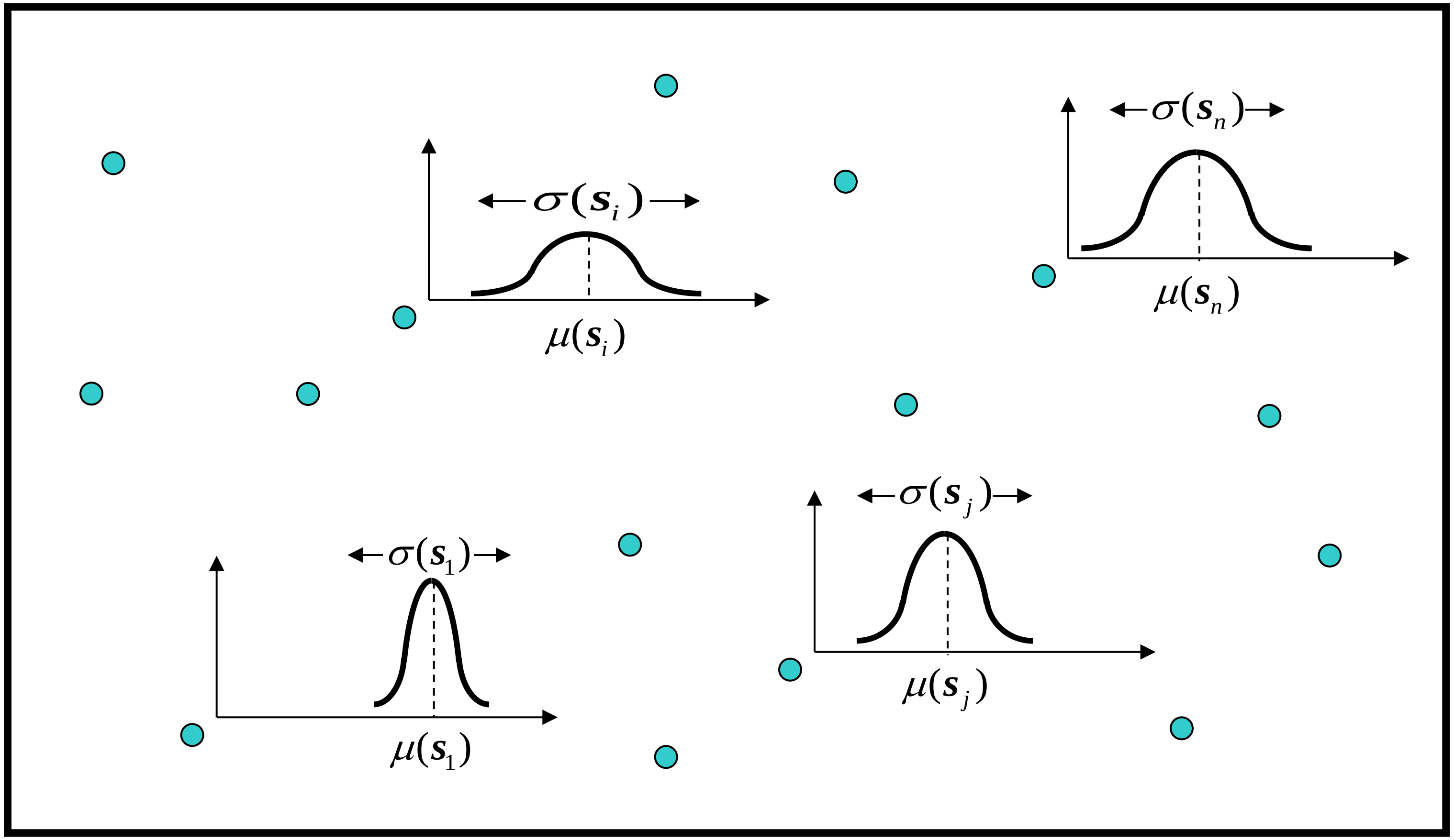
$$E[Y(\boldsymbol{s})] = \mu(\boldsymbol{s})$$

$$VAR[Y(\boldsymbol{s})] = \sigma^2(\boldsymbol{s})$$

Mean and variance in a stationary process



Mean and variance of a spatial random variable



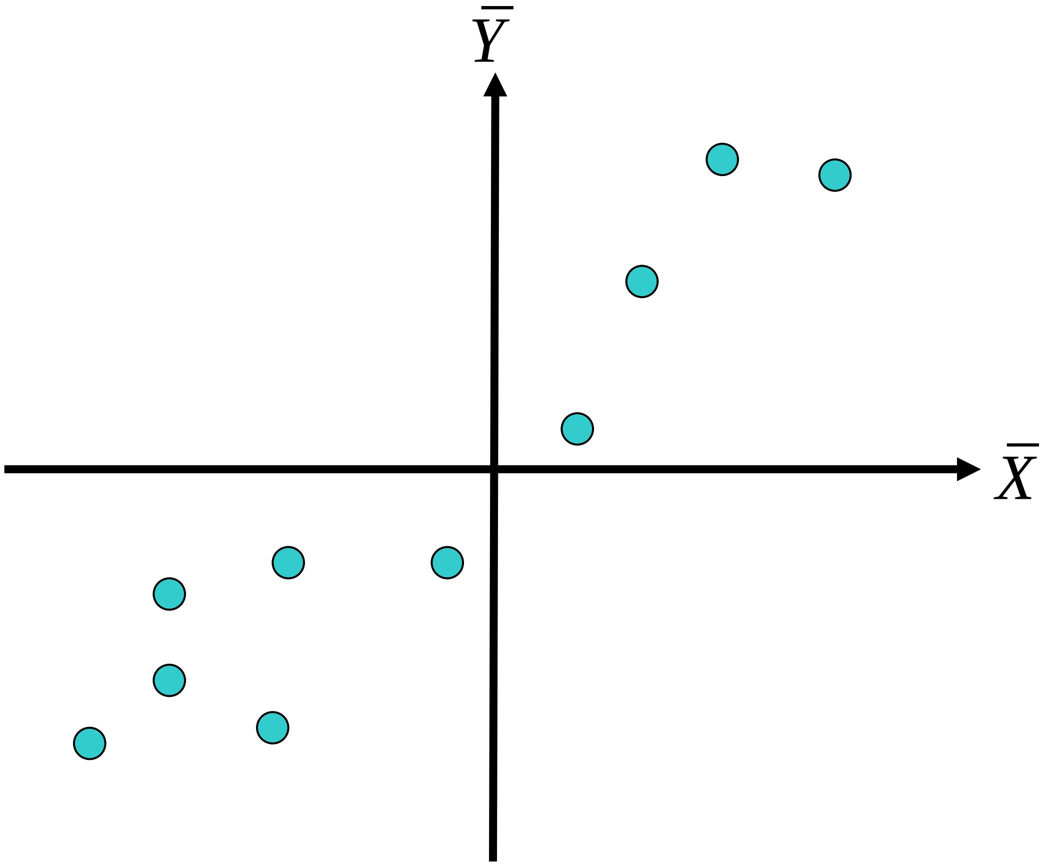
Covariance (two variables)

- Covariance

$$C_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

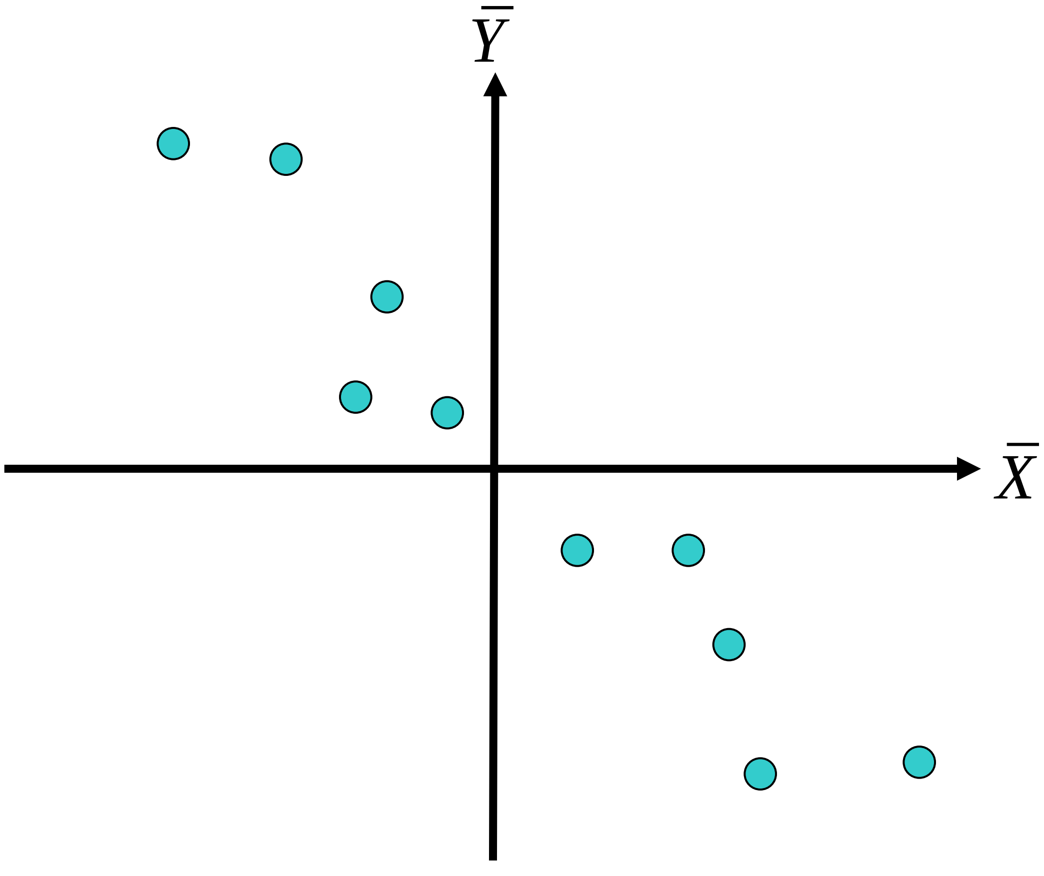
Covariance (two variables)

- Covariance



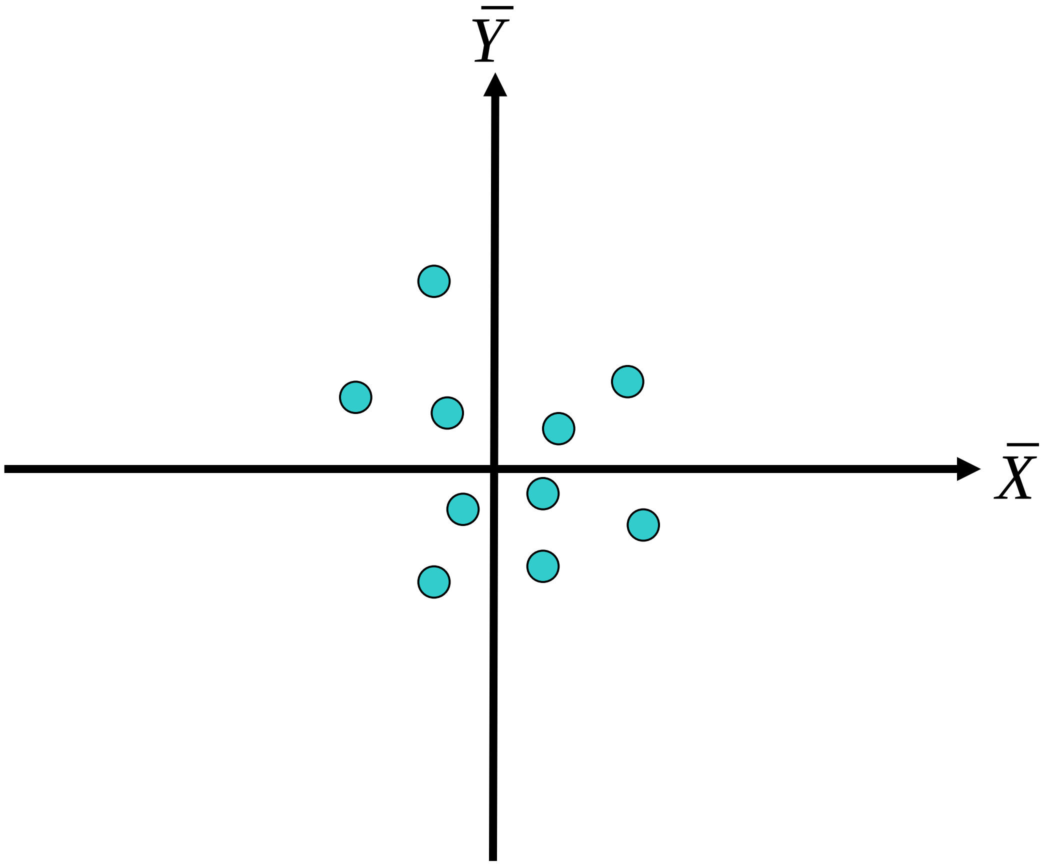
Covariance (two variables)

- Covariance



Covariance (two variables)

- Covariance



Correlation of the process

- Correlation of the process between *X* and *Y*

$$\rho(X,Y)=\frac{C_{XY}}{\sigma_X\sigma_Y}$$

Auto-covariance of the process

- Covariance of the process between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$

$$C(\mathbf{s}_i, \mathbf{s}_j) = E[(Y(\mathbf{s}_i) - \mu(\mathbf{s}_i))(Y(\mathbf{s}_i) - \mu(\mathbf{s}_i))]$$

$$C(\mathbf{s}_i, \mathbf{s}_i) = \sigma^2(\mathbf{s}_i)$$

Covariance of a stationary process

- Covariance of the process between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ *depends on distance and direction only*

$$C(\boldsymbol{s}_i, \boldsymbol{s}_j) = C(\boldsymbol{s}_i - \boldsymbol{s}_i) = C(\boldsymbol{h})$$

- Correlation

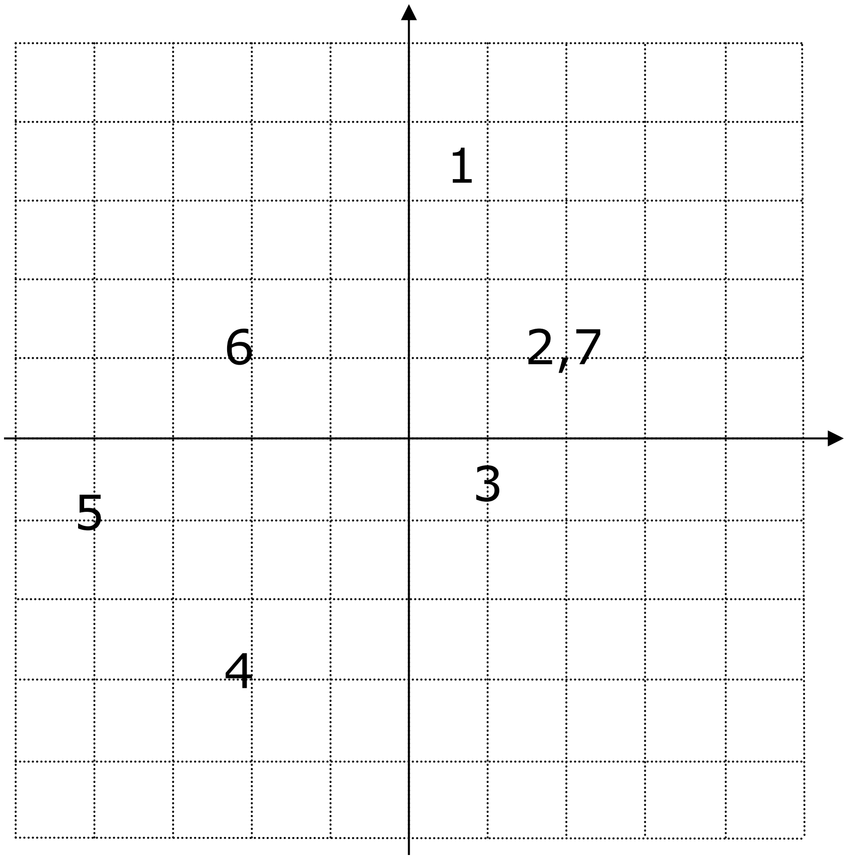
$$\rho(\boldsymbol{s}_i, \boldsymbol{s}_j) = \frac{C(\boldsymbol{h})}{\sigma^2}$$

Auto-covariance at various spatial lags

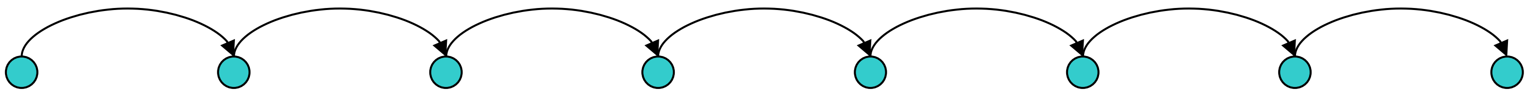
y_1=8 y_2=9 y_3=8 y_4=7 y_5=6 y_6=7 y_7=8 y_8=9

$\bar{y}$=7.75

h=1



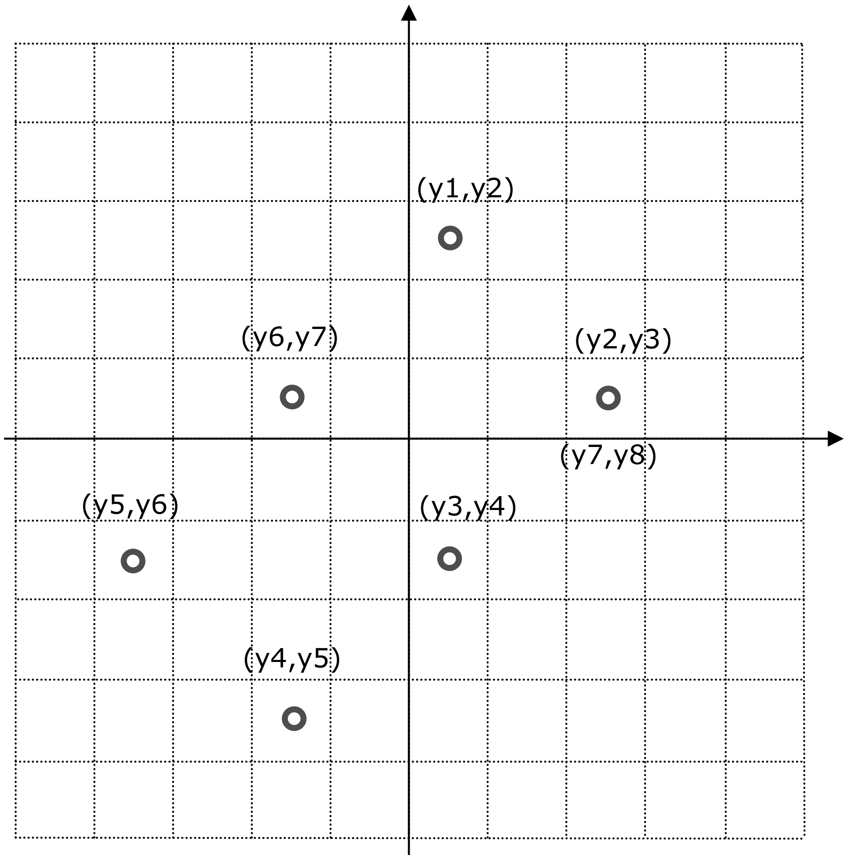
Auto-covariance at various spatial lags



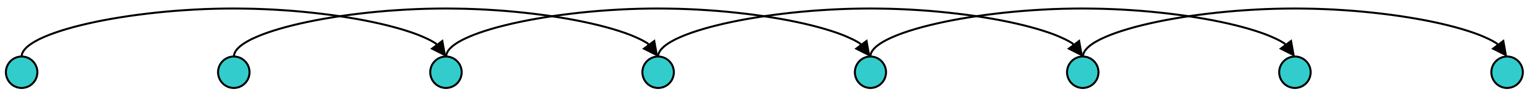
$y_1=8$ $y_2=9$ $y_3=8$ $y_4=7$ $y_5=6$ $y_6=7$ $y_7=8$ $y_8=9$

$\bar{y}=7.75$

h=1



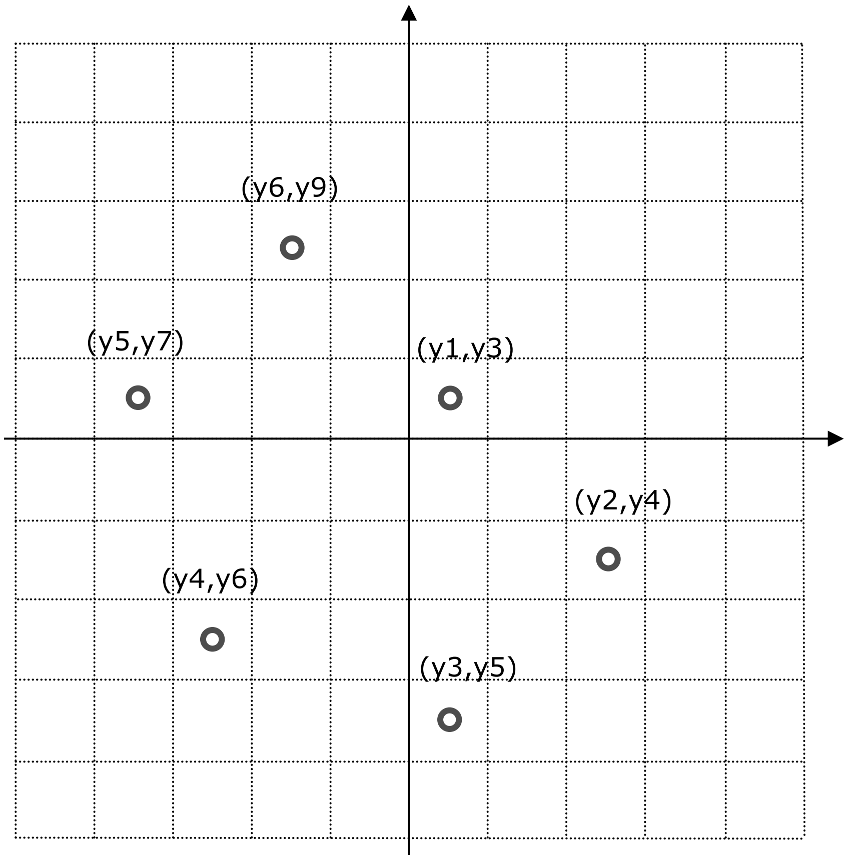
Auto-covariance at various spatial lags



y_1=8 y_2=9 y_3=8 y_4=7 y_5=6 y_6=7 y_7=8 y_8=9

$\bar{y}$=7.75

h=2



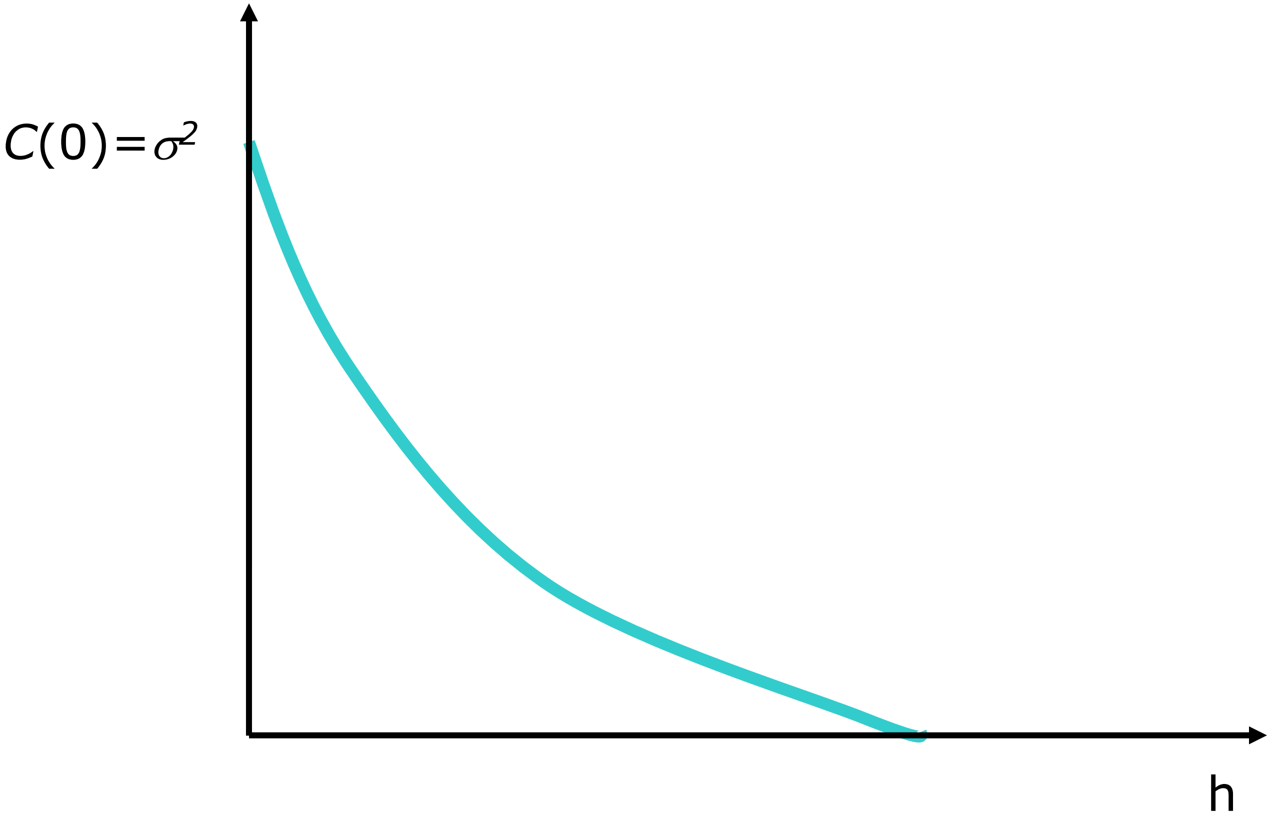
Covariogram

- A plot of the covariance of the process at various spatial lags

$$C(\boldsymbol{h}) = \frac{1}{n(\boldsymbol{h})} \sum_{\boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}} (y_i - \overline{y})(y_j - \overline{y})$$

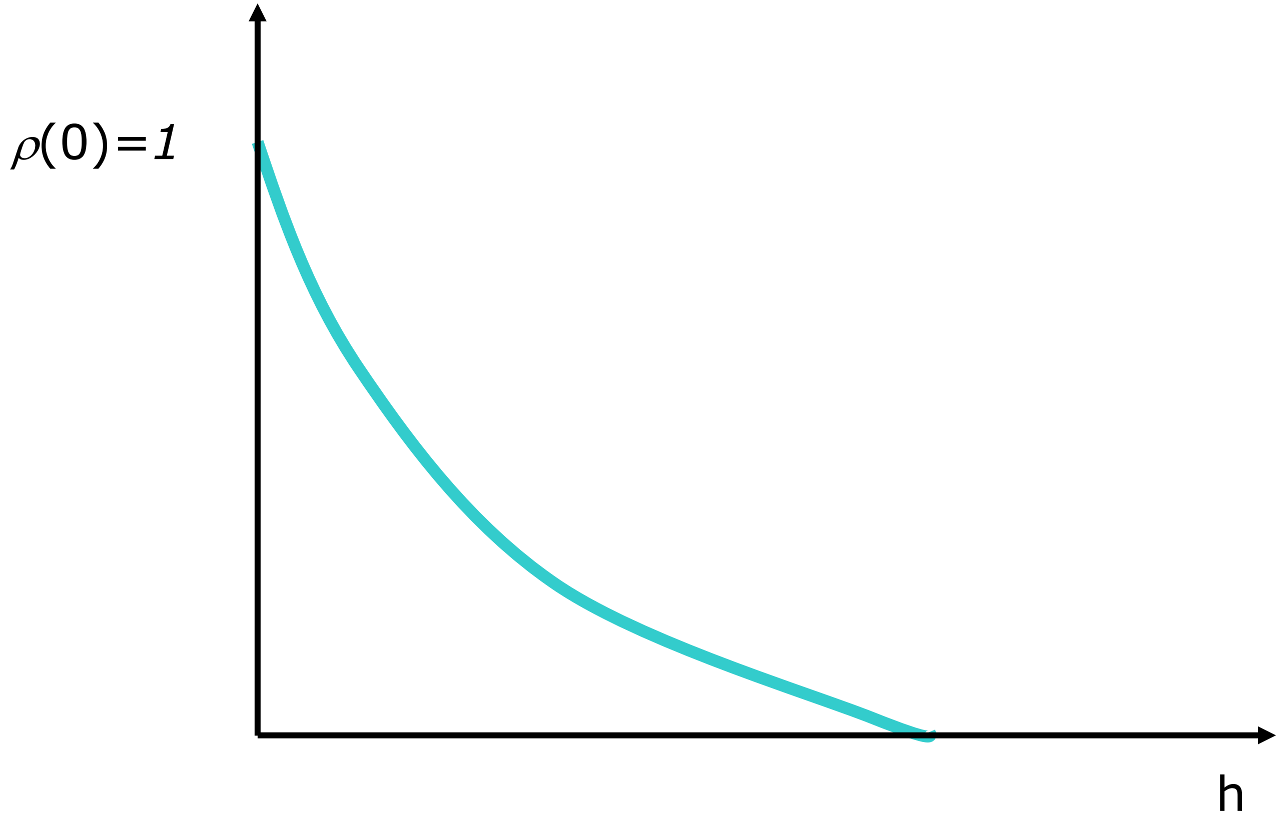
Covariogram

$$C(\boldsymbol{h}) = \frac{1}{n(\boldsymbol{h})} \sum_{\boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}} (y_i - \overline{y})(y_j - \overline{y})$$



Correlogram

$$\rho(\boldsymbol{h}) = \frac{C(\boldsymbol{h})}{\sigma^2}$$



Variogram

- Intrinsic stationarity (weaker assumption)

$$E[Y(\boldsymbol{s}+\boldsymbol{h})-Y(\boldsymbol{s})]=0$$

$$VAR[Y(\boldsymbol{s}+\boldsymbol{h})-Y(\boldsymbol{s})]=2\gamma(\boldsymbol{h})$$

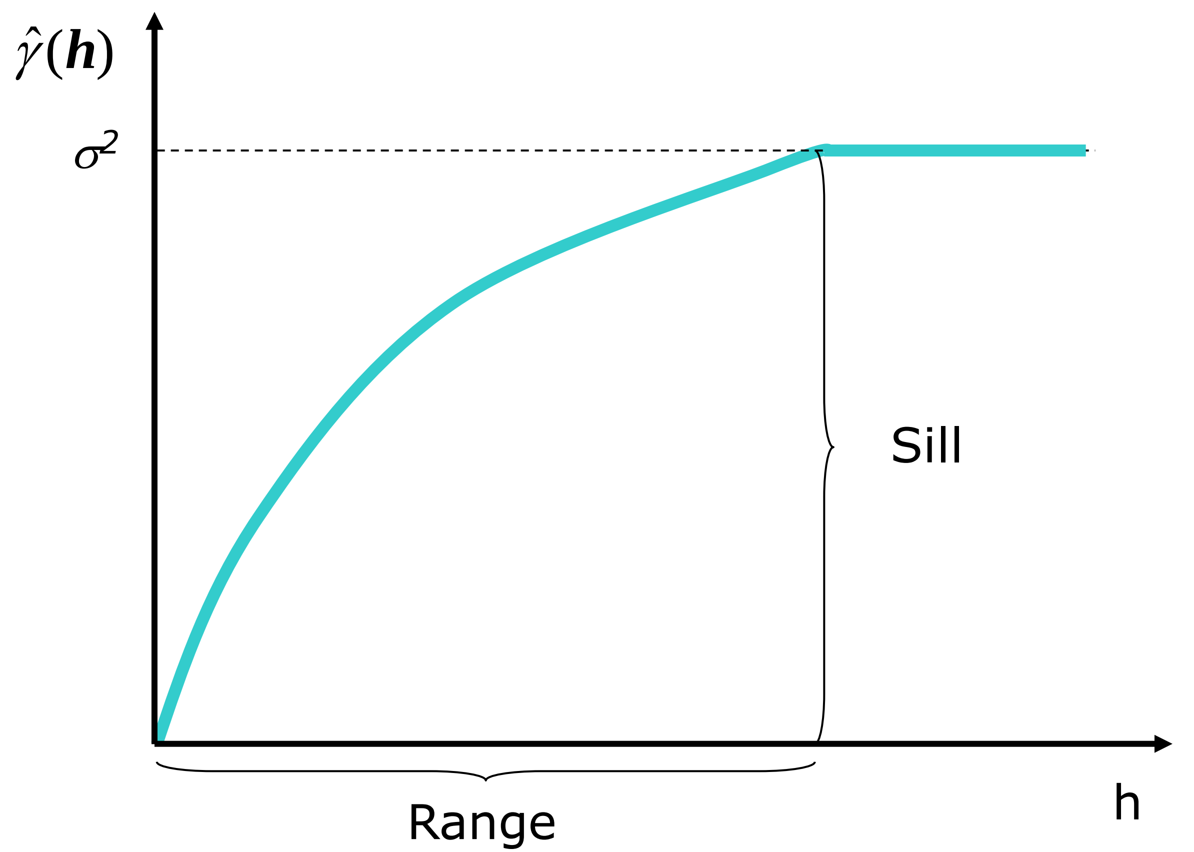
Variogram

- Intrinsic stationarity

$$2\hat{\gamma}(\boldsymbol{h}) = \frac{1}{n(\boldsymbol{h})} \sum_{\boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}} (y_i - y_j)^2$$

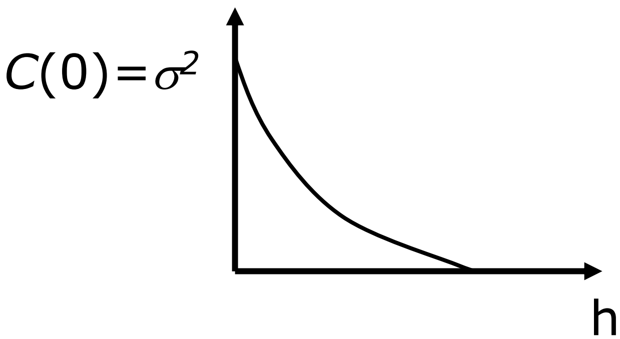
Variogram

$$\hat{\gamma}(\boldsymbol{h}) = \frac{1}{2n(\boldsymbol{h})} \sum_{\boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}} (y_i - y_j)^2$$

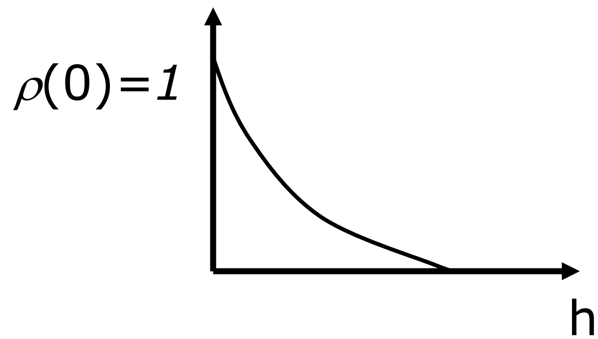


Covariogram, correlogram, variogram

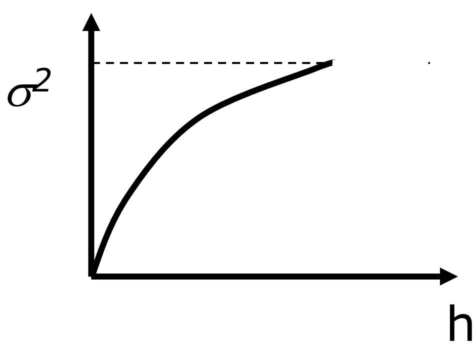
Covariogram $C(\boldsymbol{h})$



Correlogram

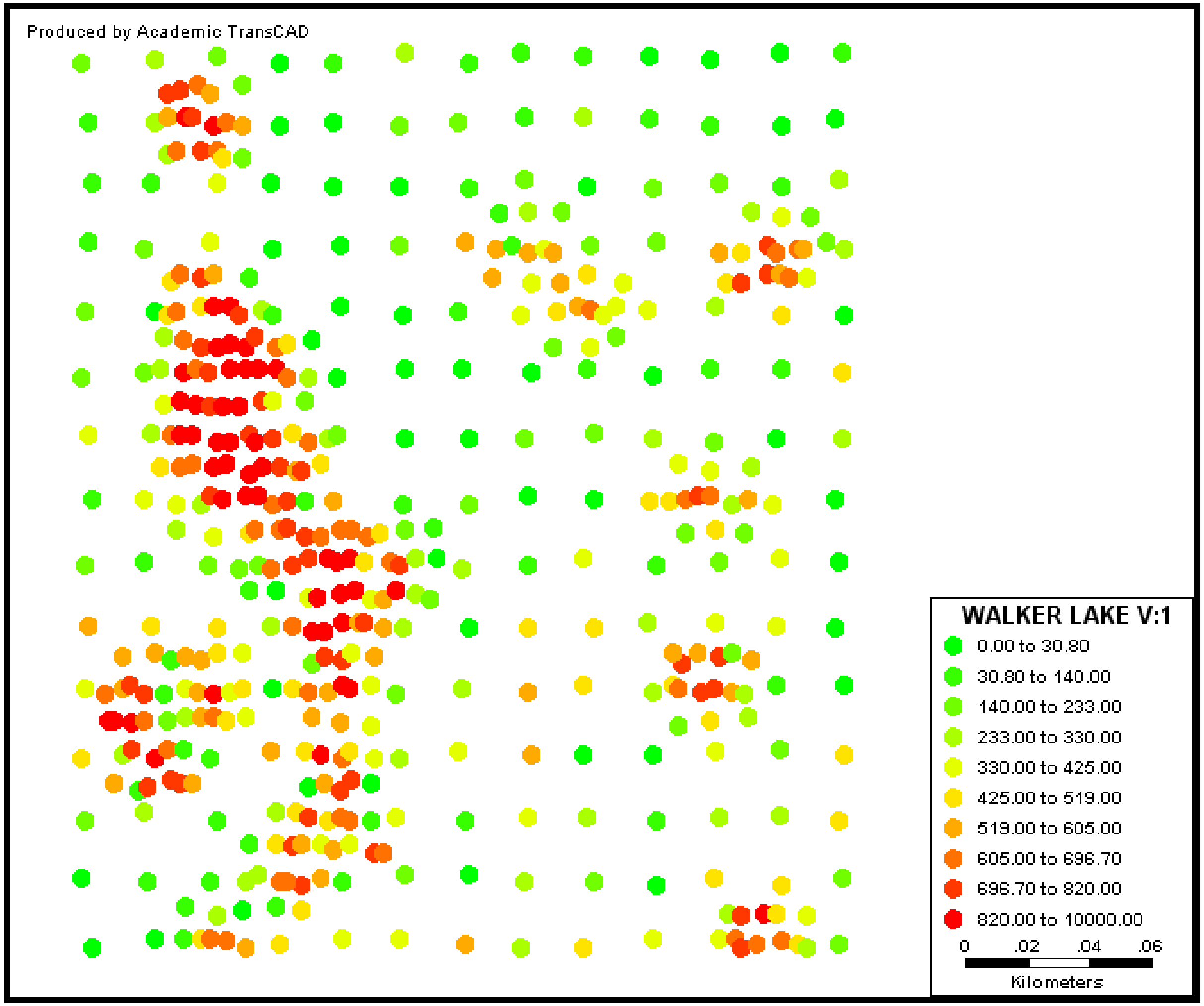


Variogram $\hat{\gamma}(\boldsymbol{h})$



$$\hat{\gamma}(\boldsymbol{h}) = \sigma^2 - C(\boldsymbol{h})$$

How to read the variogram



How to read the variogram

- Example of variogram

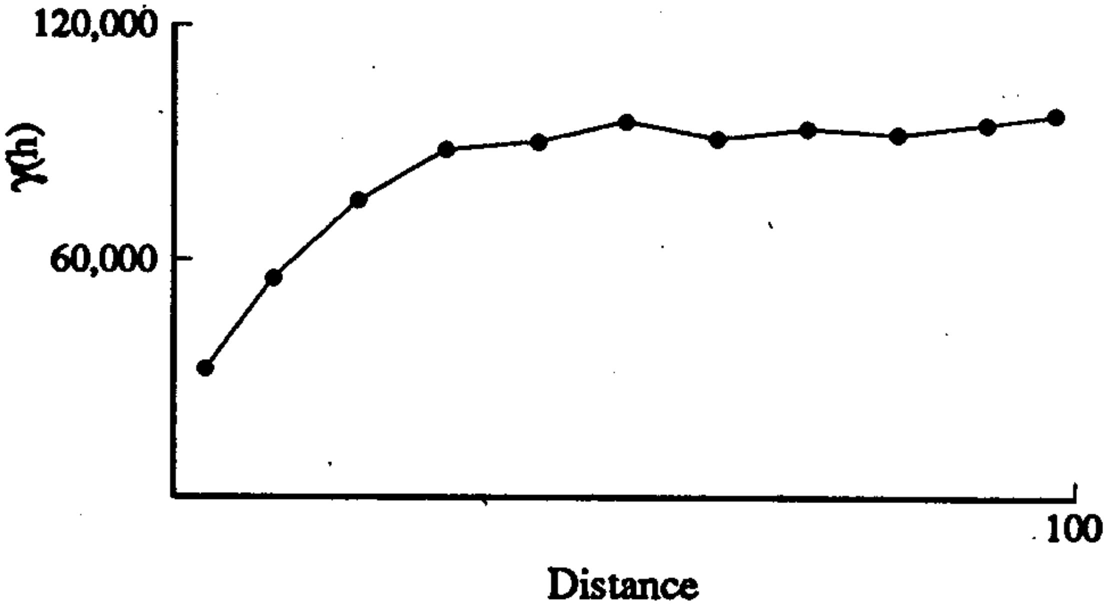


Figure 7.3 Omnidirectional sample variogram for V with a 10 m lag.

How to read the variogram

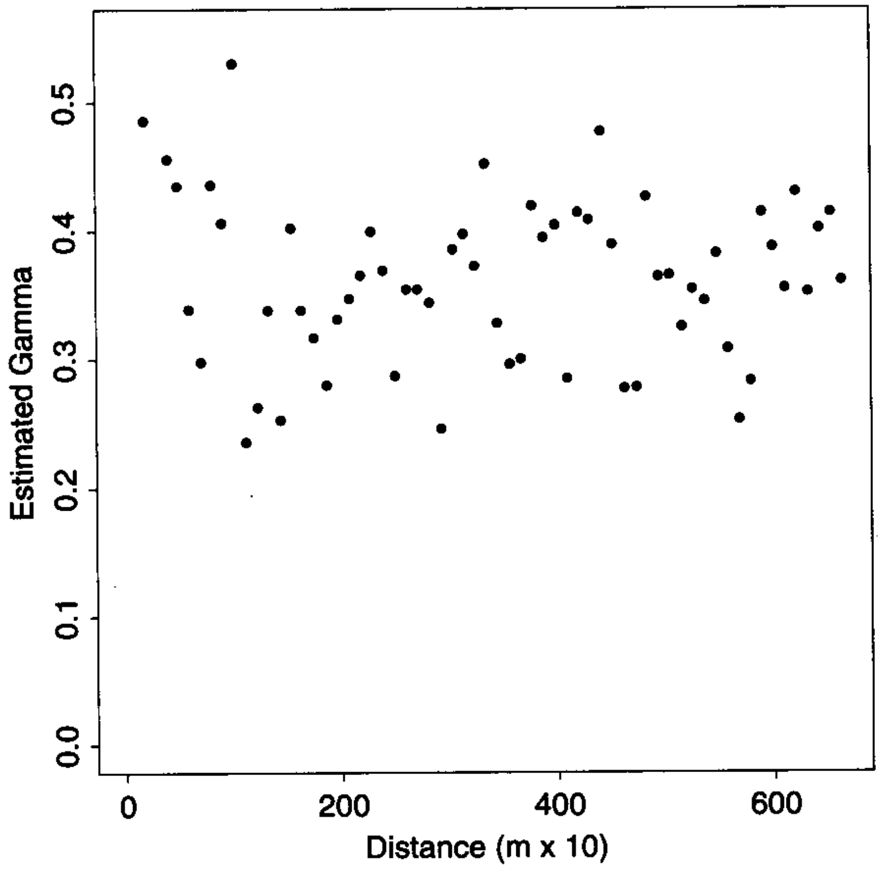


Fig. 5.11 Sample variogram for logarithms of radon levels

How to read the variogram

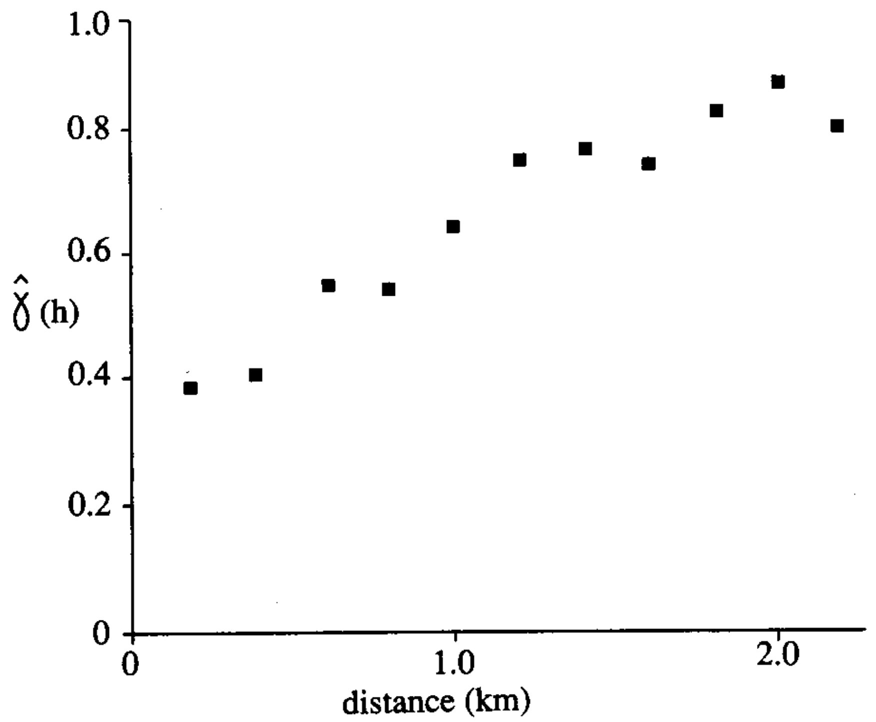
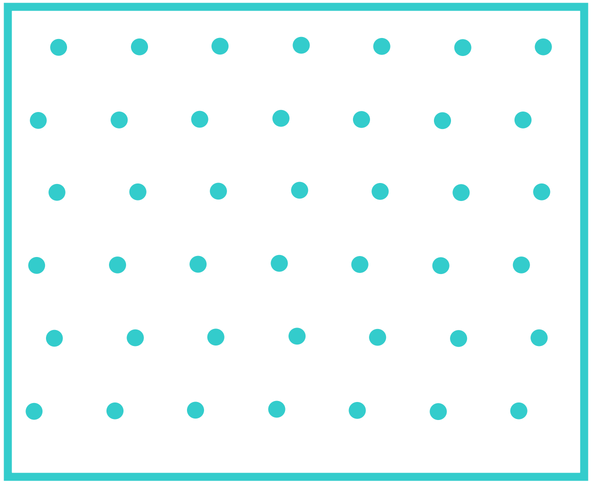
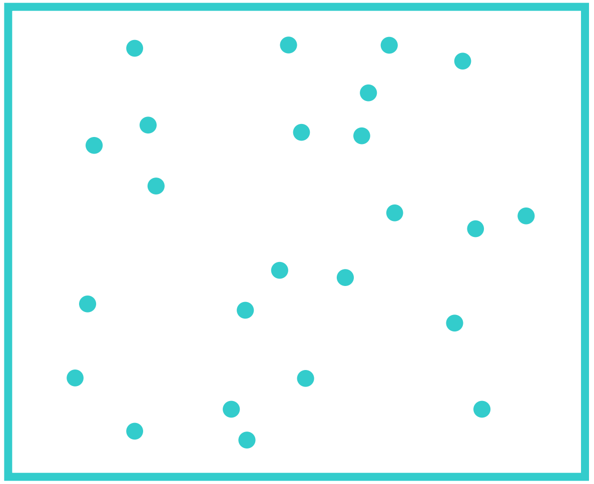


Fig. 5.12 Sample variogram for logarithms of nickel concentrations on north Vancouver Island

Some important issues

- Selection of spatial lags h





Some important issues

- Selection of spatial lags h
  - Number of point pairs
  - Few points, low reliability

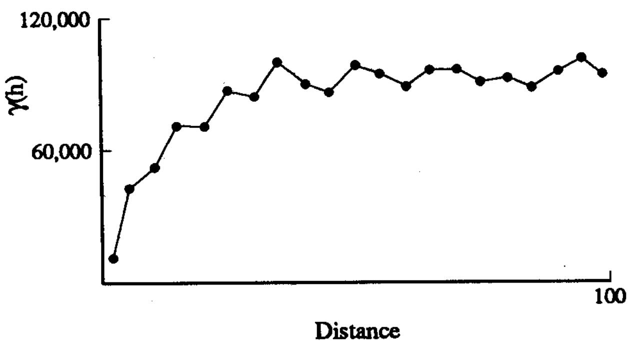


Figure 7.2 Omnidirectional sample variogram for V with a 5 m lag.

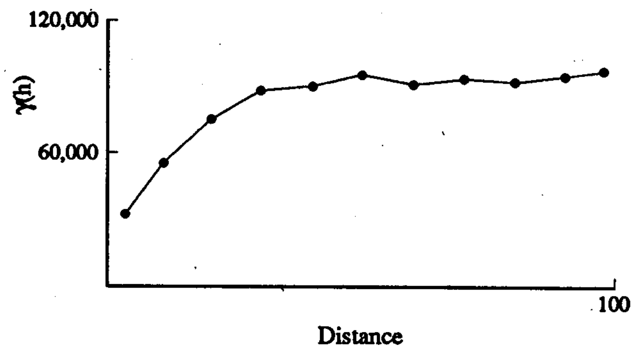


Figure 7.3 Omnidirectional sample variogram for V with a 10 m lag.

Some important issues

- Selection of spatial lags h
  - Number of point pairs
  - Few points, low reliability

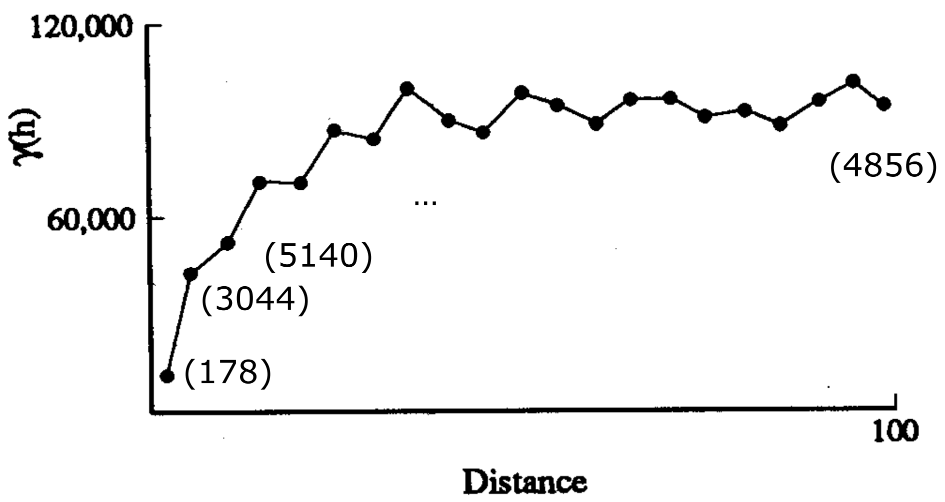
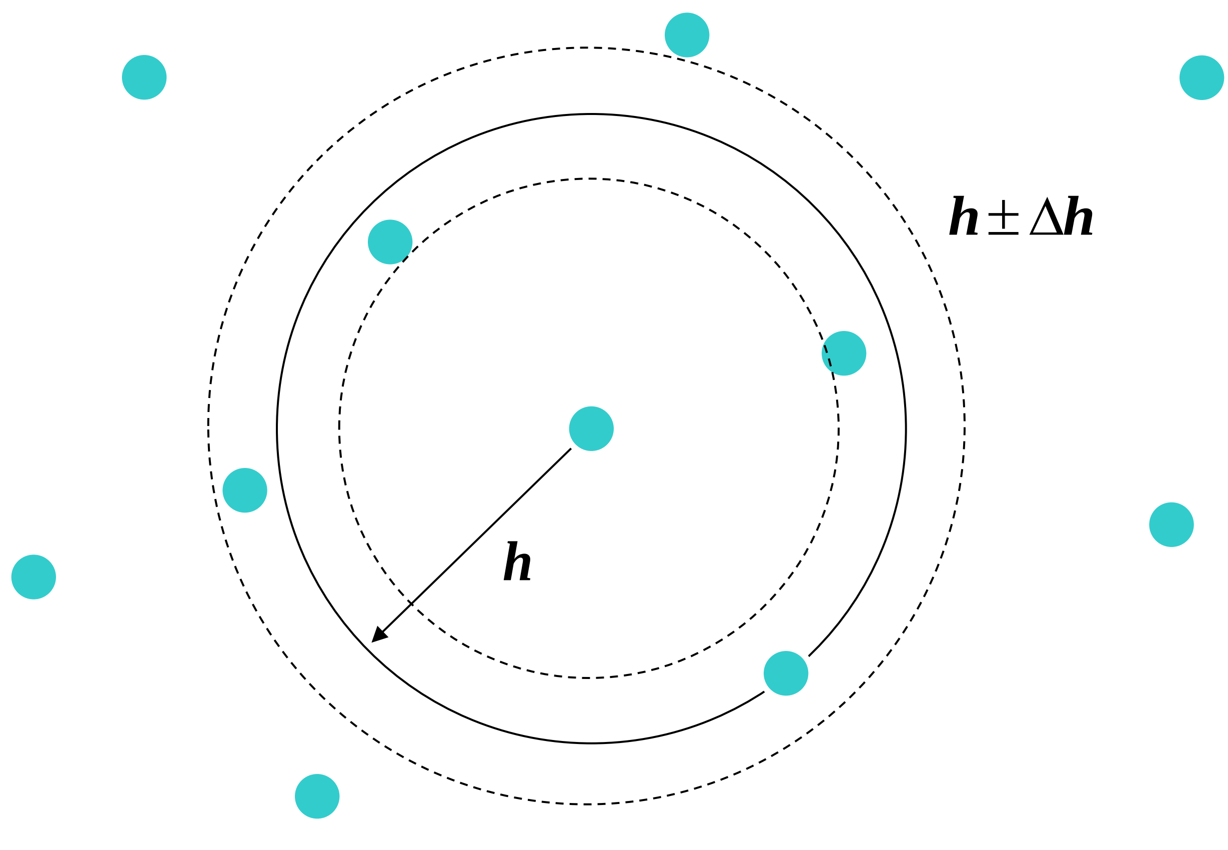


Figure 7.2 Omnidirectional sample variogram for *V* with a 5 m lag.

Some important issues

- Selection of spatial lags h



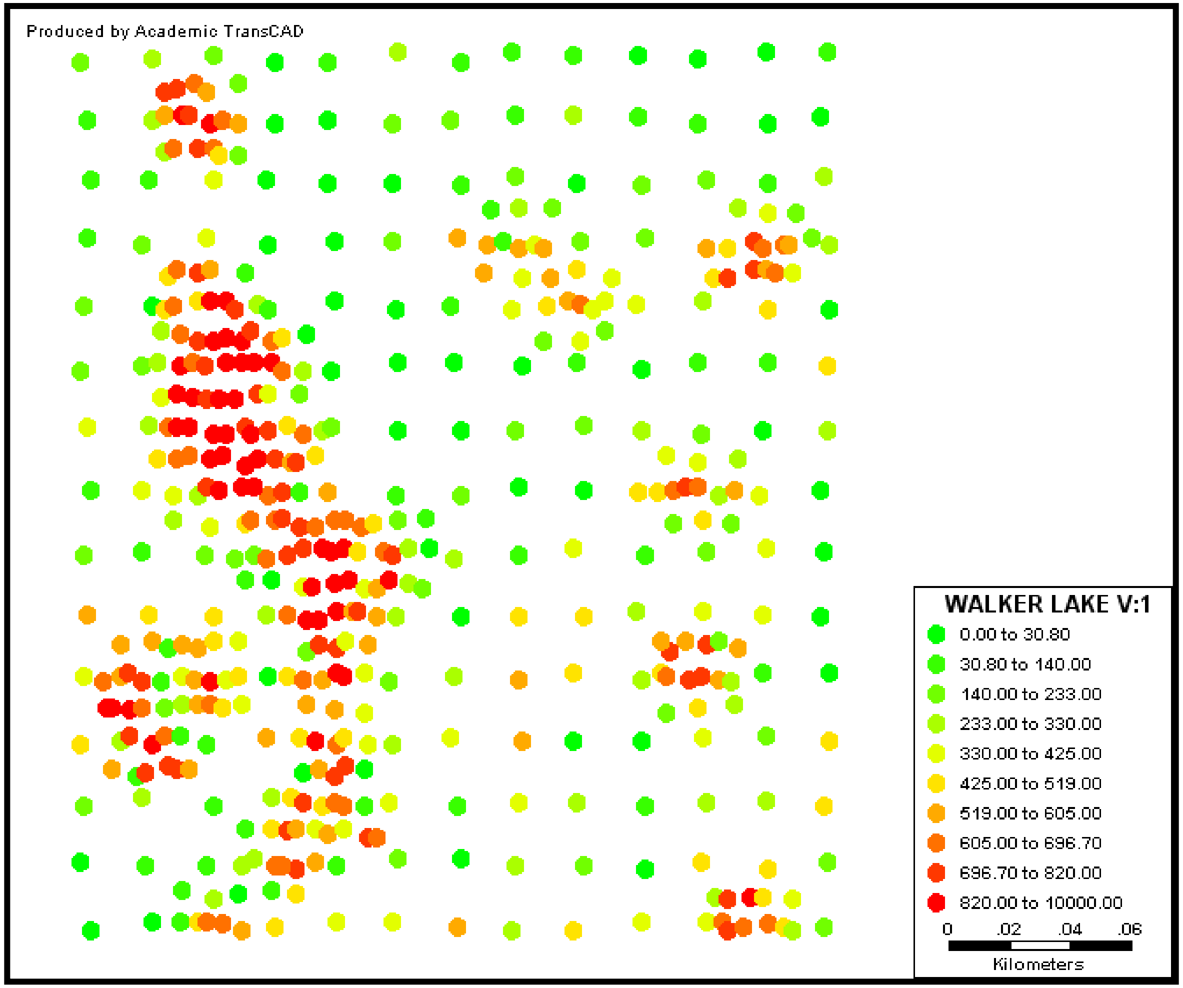
Some Important Issues: Isotropy

- The process depends only on distance, not direction
  - Useful preliminary assumption
    - Omnidirectional variogram
    - Cleaner relationships
  - Must be checked against the data

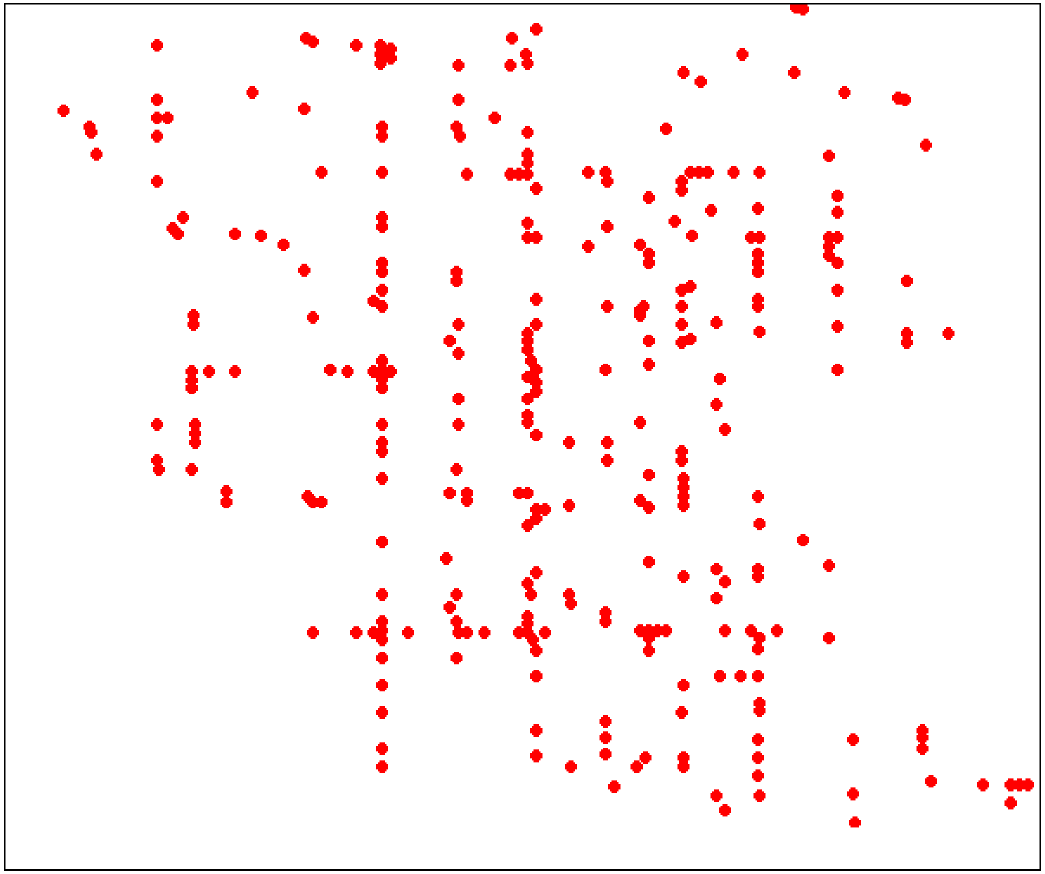
Some important issues

- Anisotropy
  - Isaaks and Srivastava (1989) Applied Geostatistics, chapters 7 and 16

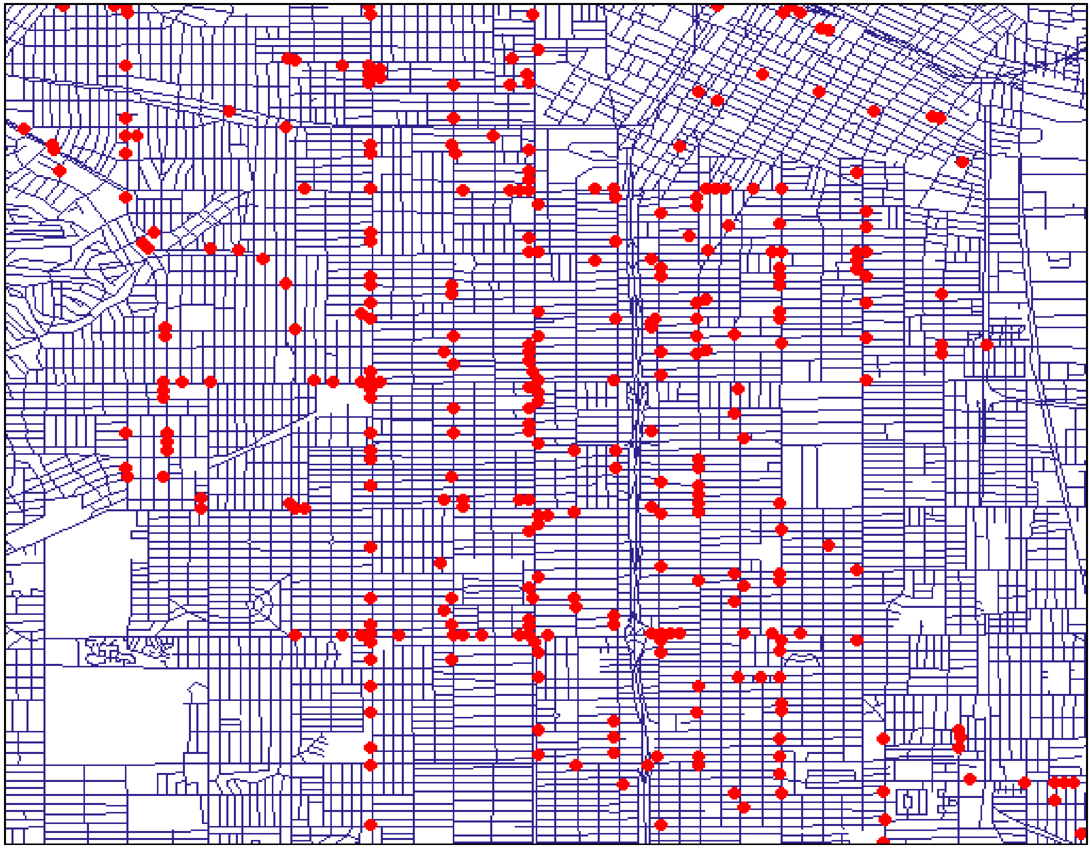
Anisotropy: Walker Lake



Anisotropy: Street Dust

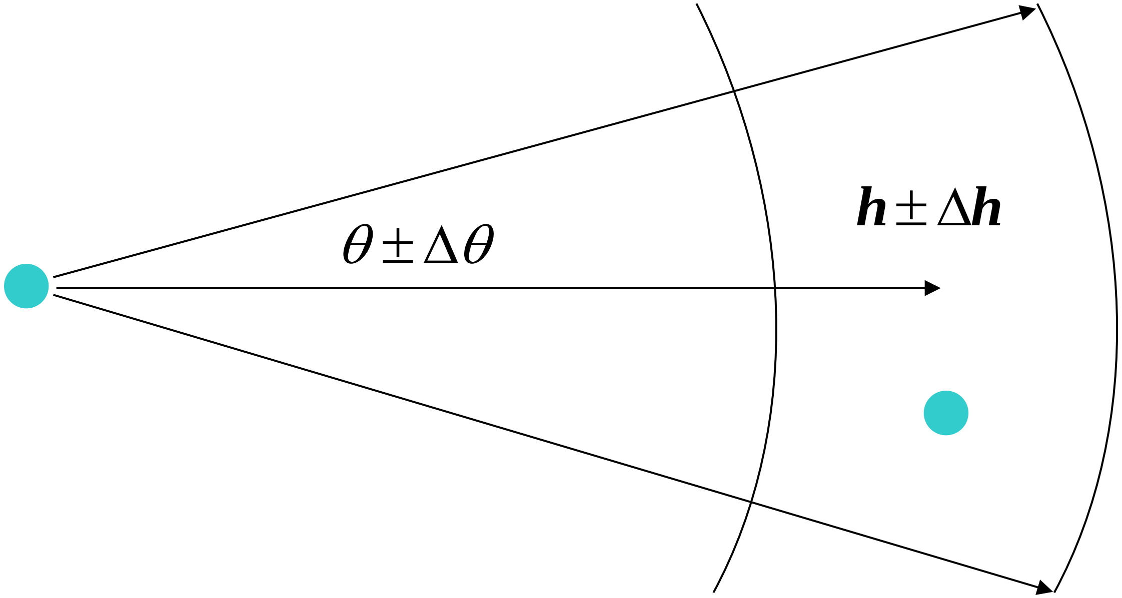


Anisotropy: Street Dust



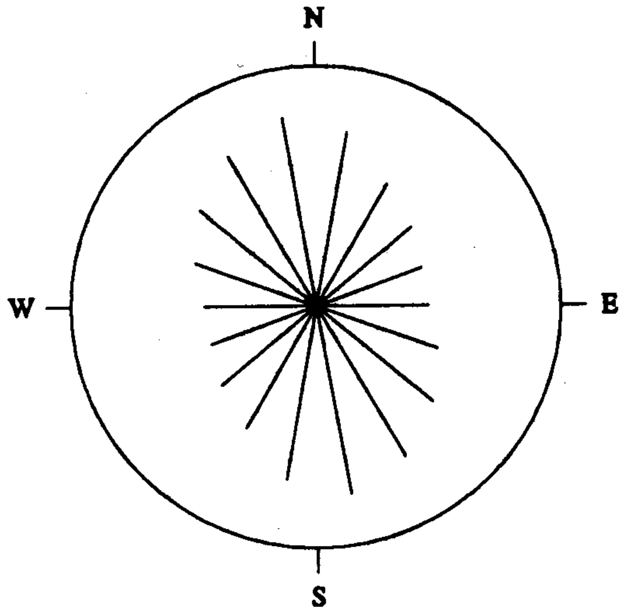
Anisotropy

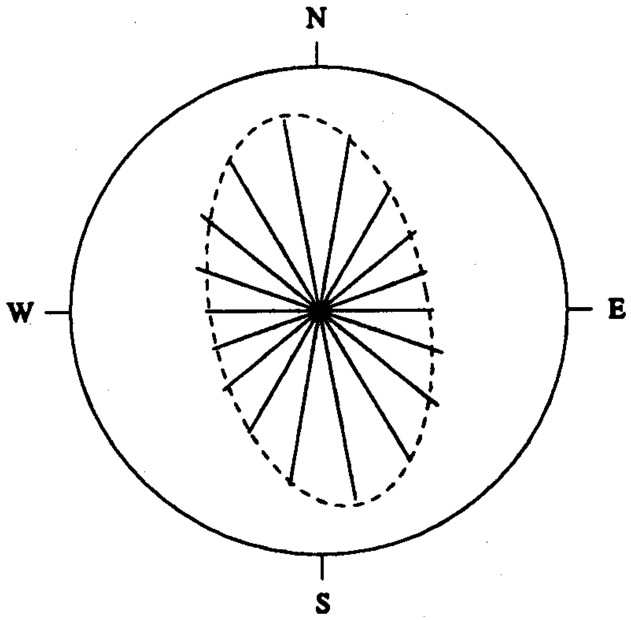
- Directional variogram
  - Selection of spatial lags h



Directional Variograms

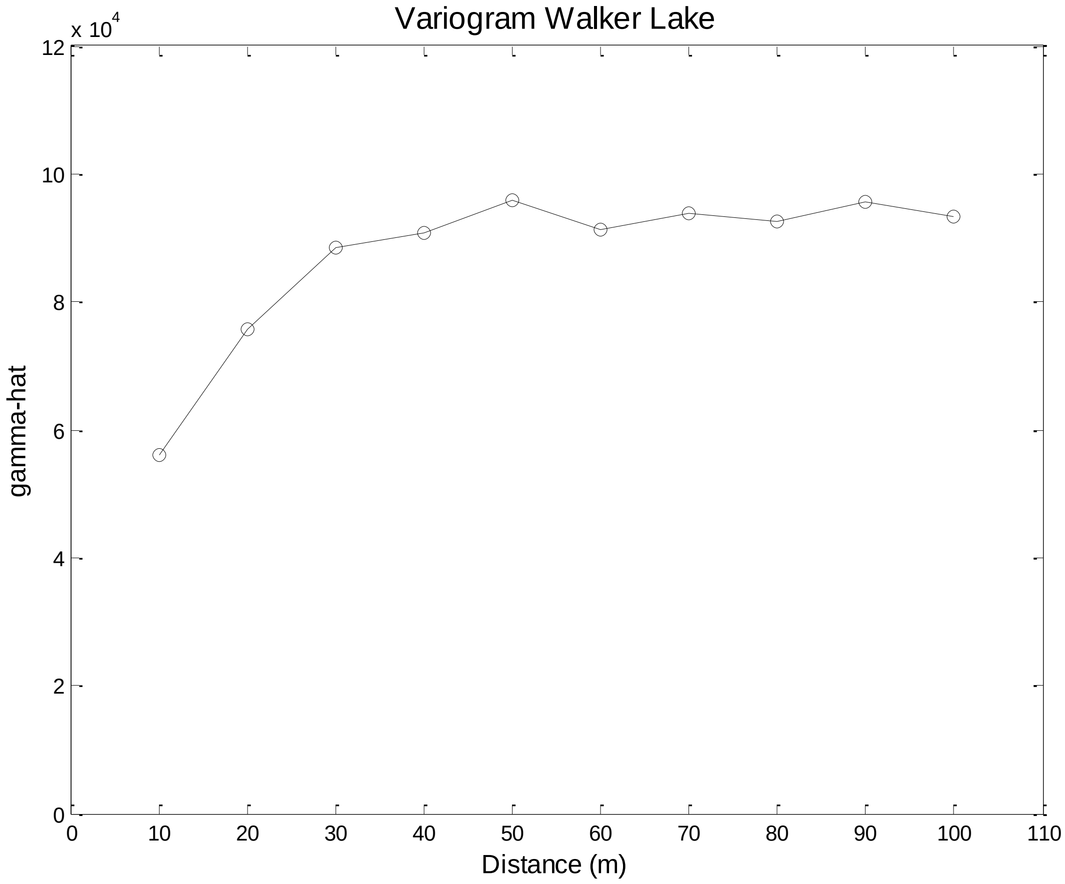
- Walker Lake





Describing Spatial Structure

- Empirical Variogram



Variogram Models

- A theoretical model that fits the empirical variogram
- The variogram model should produce a covariance matrix that is:
  - Symmetric

$$C(\mathbf{s}_i, \mathbf{s}_j) = C(\mathbf{s}_j, \mathbf{s}_i)$$

  - Non-negative definite

Variogram Models

- Spherical
- Exponential
- Gaussian
- Linear
- Bessel
- …

Variogram Models

- Spherical

$$\gamma(h) = \begin{cases} \sigma^2 \left(\dfrac{3h}{2r} - \dfrac{h^3}{2r^3} \right) & \text{for } h \le r \\ \sigma^2 & \text{otherwise} \end{cases}$$

Variogram Models

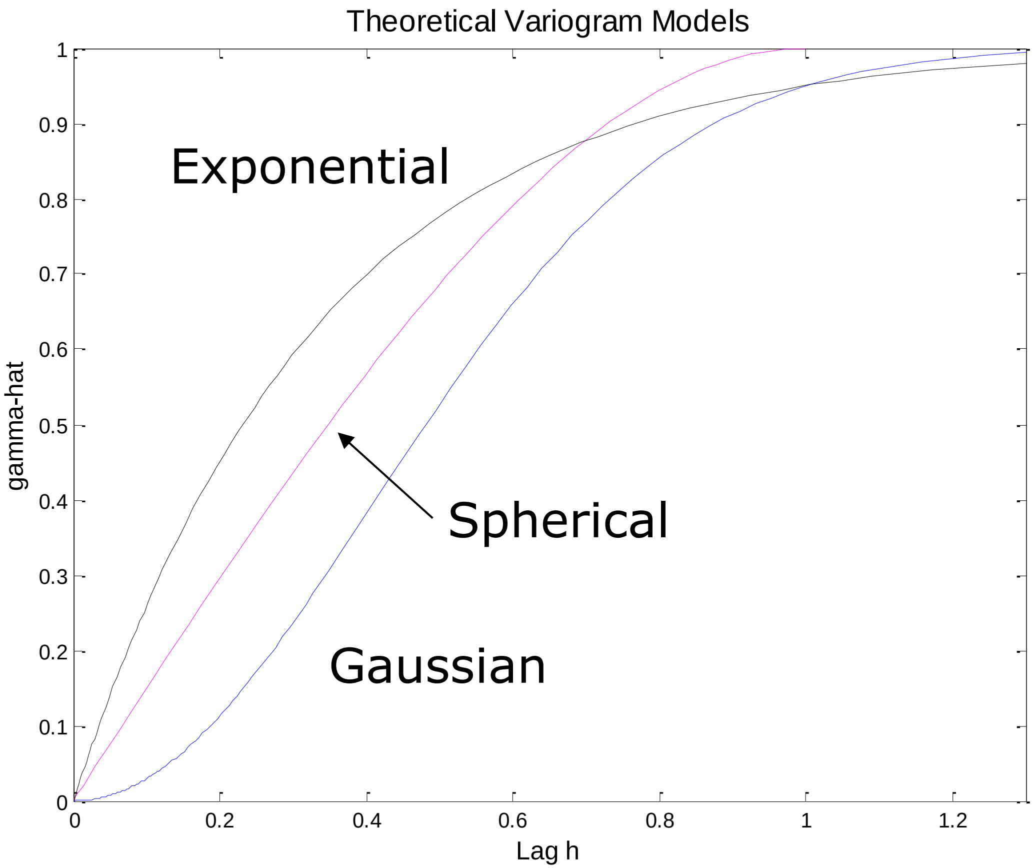
- Exponential

$$\gamma(h) = \sigma^2\left(1 - \mathrm{e}^{-3h/r}\right)$$

- Gaussian

$$\gamma(h) = \sigma^2\left(1 - \mathrm{e}^{-3h^2/r^2}\right)$$

Variogram Models



Variogram Models

- Spherical with nugget effect

$$\gamma(h) = \begin{cases} a + \left(\sigma^2 - a\right)\left(\dfrac{3h}{2r} - \dfrac{h^3}{2r^3}\right) & \text{for } 0 < h \le r \\ 0 & \text{for } h = 0 \\ \sigma^2 & \text{otherwise} \end{cases}$$

Variogram Models

- Exponential with nugget effect

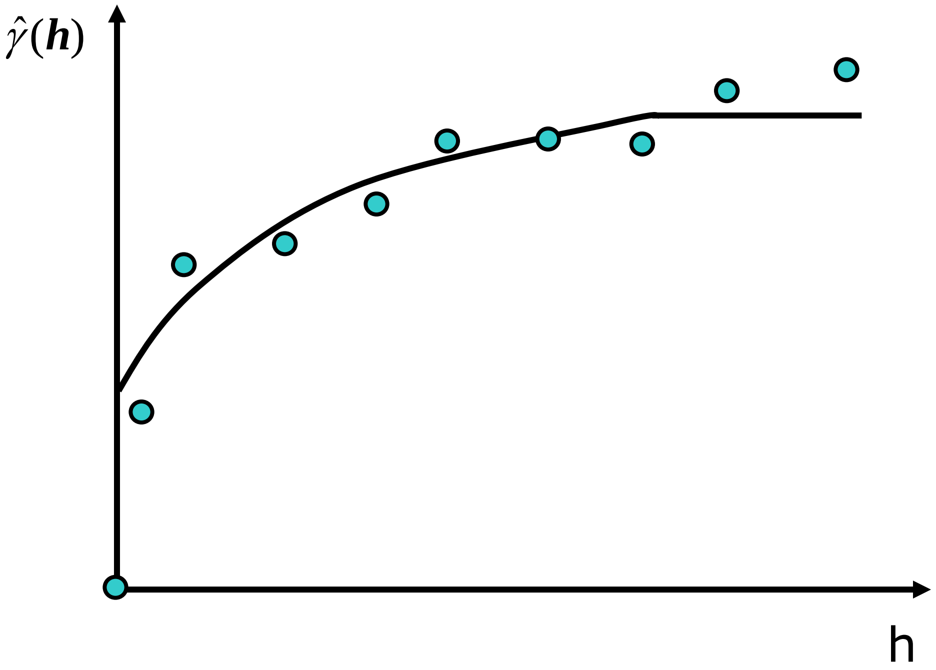
$$\gamma(h) = \begin{cases} \sigma^2\left(1 - \mathrm{e}^{-3h/r}\right) & \text{for } h > 0 \\ 0 & \text{for } h = 0 \end{cases}$$

Variogram Models

- Relative nugget effect
  - The ratio of the nugget to the sill (a/σ^2)
  - Sometimes expressed as percentage
  - A relative nugget of 100% indicates total absence of spatial autocorrelation

Variogram Models

- Fitting a theoretical model: Least squares



$$\sum_h \left(\hat{\gamma}(h) - \gamma(h) \right)^2$$

Second Order Effects

- The covariogram, correlogram and variogram and exploratory tools to study second order properties of spatially continuous data
- The variogram is more resistant to departures from the assumptions
- These tools are important elements in modeling spatially continuous data

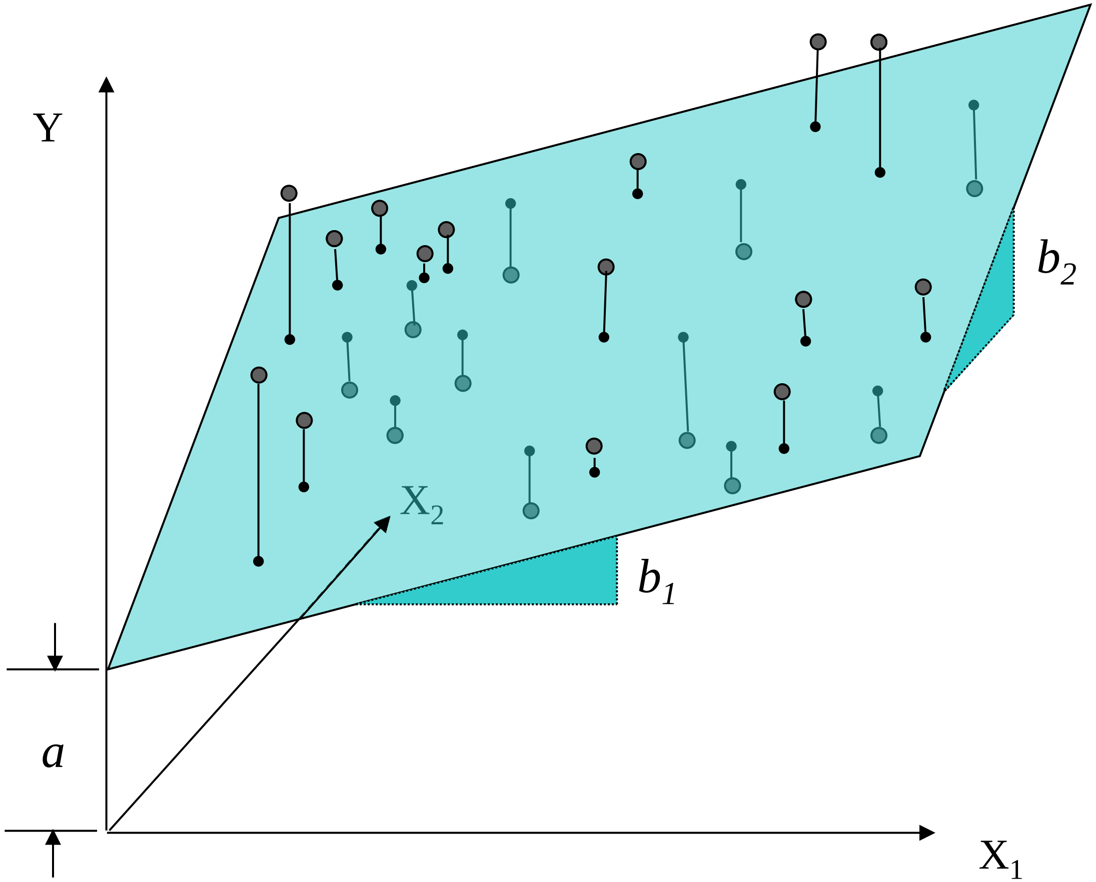
First AND/OR Second Order Effects

- Variogram analysis very effective in the presence of pure second order effects
- Effect of a trend on variogram analysis
- De-trending the data
- Trend surface analysis

Modeling Spatially Continuous Data

- Trend Surface Analysis
  - A method to model the first order component of a spatial process
  - Multiple regression analysis using the coordinates of the points as “explanatory” variables

Multiple Regression Analysis



Trend Surface Analysis

- Linear trend surface

$$z_i = b_1 + x_i b_2 + y_i b_3 + e_i$$

  - z: dependent variable
  - b_1 , b_2 , b_3: regression parameters
  - x , y: coordinates of point i

Regression Analysis: Matrix Notation

- Linear trend surface

$$z_1 = b_1 + x_1 b_2 + y_1 b_3 + e_1$$

$$z_2 = b_1 + x_2 b_2 + y_2 b_3 + e_2$$

$$\vdots$$

$$z_n = b_1 + x_n b_2 + y_n b_3 + e_n$$

Regression Analysis: Matrix Notation

- Linear trend surface

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

Regression Analysis: Matrix Notation

- Linear trend surface

$$\boldsymbol{Z} = \mathbf{X}\boldsymbol{b} + \boldsymbol{e}$$

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

Regression Analysis: Assumptions

- Error terms

Regression Analysis: Assumptions

- Error terms
  - Expected (mean) value = 0

$$E[\boldsymbol{e}] = \boldsymbol{0}$$

Regression Analysis: Assumptions

- Error terms
  - Constant variance
  - Independent

$$E[\boldsymbol{ee}'] = \mathbf{C} = \begin{bmatrix} \sigma^2 & & 0 & 0 \\ & \sigma^2 & & 0 \\ 0 & & \sigma^2 & \\ 0 & 0 & & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & & 0 & 0 \\ & 1 & & 0 \\ 0 & & 1 & \\ 0 & 0 & & 1 \end{bmatrix} = \sigma^2 \mathbf{I}$$

Regression Analysis: Estimation

- Parameters

$$\hat{\boldsymbol{b}} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \boldsymbol{Z}$$

- Variance

$$\hat{\sigma}^2 = \frac{1}{n-k}\left(\boldsymbol{Z} - \mathbf{X}^T \hat{\boldsymbol{b}}\right)^T \left(\boldsymbol{Z} - \mathbf{X}^T \hat{\boldsymbol{b}}\right)$$

Next…

- Spatially continuous data V-VI
  - Spatial prediction
  - Kriging