GEOG574 Introduction to Geostatistics

Variogram

Daoqin Tong

School of Geography &Development
408 Harvill Building
Email: daoqin@email.arizona.edu

Properties of Geostatistical Data

- First order effects
  - The mean value $E(Y(s)) = \mu(s)$
  - Global trend, large scale variation
- Second order effects
  - Spatial dependence between two points

    $COV(Y(s_i), Y(s_j))$
  - Local or small scale variation

Covariance and Correlation

- For a spatial stochastic process $\{Y(s), s \in \Re\}$, denoting $E(Y(s))$ as $\mu(s)$, $Var(Y(s))$ as $\sigma^2(s)$, the covariance of the process at two locations s_i and s_j is defined as

$$C(s_i, s_j) = E[(Y(s_i) - \mu(s_i))(Y(s_j) - \mu(s_j))]$$

- This also implies that $C(s_i, s_i) = \sigma^2(s_i)$

- The correlation is defines as

$$\rho(si, sj) \frac{C(s_i, s_j)}{\sigma(s_i)\sigma(s_j)}$$

Strictly Stationary

- To estimate the covariance, we need to make some assumptions of stationarity to achieve repetition

$$(Y(s_1), \ldots, Y(s_m)) \overset{d}{=} (Y(s_1 + h), \ldots., Y(s_m + h))$$
$$\text{Any } m, \text{any } \{s_i\}, \text{any } h$$

Second-order Stationarity

- A process is said to be second-order stationary if the following three conditions are satisfied

$$\mu(s) = \mu$$
$$C(s_i, s_j) = C(s_i - s_j) = C(h) \Rightarrow C(0) = \sigma^2$$

- Note that, for Gaussian processes,
  2^{nd}-order stationary≡strictly stationary
- $C(h)$ is often referred to as the covariance function or covariogram
- $\rho(h) = C(h)/\sigma^2$ is referred to as correlogram

Covariogram and Correlation

- A process is isotropic if the dependence is purely a function of distance between two locations but not their direction
- The assumption will give us more repetitions for covariance estimation
- We will have

$$C(s_i, s_j) = C(|s_i - s_j|) = C(|h|)$$
$$\rho(s_i, s_j) = \rho(|s_i - s_j|) = \rho(|h|)$$

Intrinsic Stationarity

- A weaker assumption than stationarity is intrinsic stationarity
- A process is intrinsically stationary if the differences between two values at locations separated by a given distance and direction have a constant mean and constant variance

$$E(Y(s+h)-Y(s))=0$$
$$VAR(Y(s+h)-Y(s))=2\gamma(h)\Rightarrow\gamma(0)=0$$

- $\gamma(h)$ is referred as (semi) variogram
- For an isotropic process, variogram becomes $\gamma(|h|)$

Semivariogram

- 2nd-order stationarity $\rightarrow$

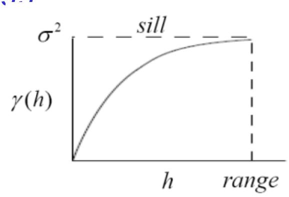
$$2\gamma(h)=2\{C(0)-C(h))$$

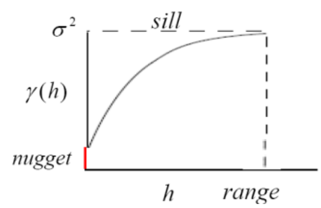
- More generally,

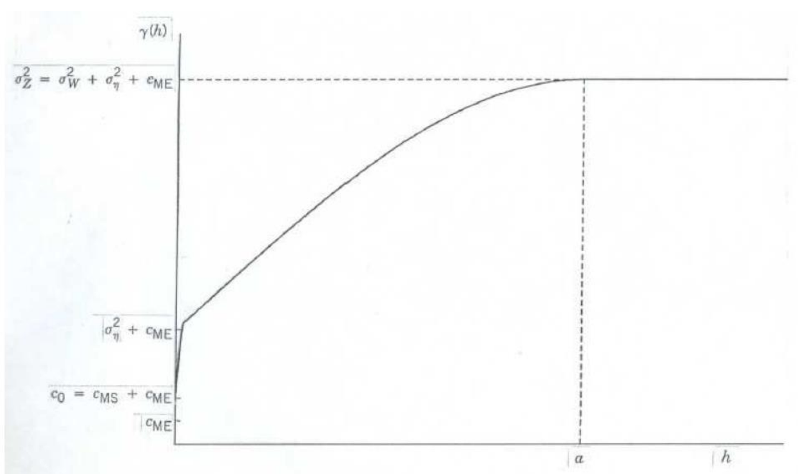
$$2\gamma(u-v)=VAR(Y(u))+Var(Y(v))-2COV(Y(u),Y(v))$$

Variogram

- For a stationary process, variogram should reach an upper bound, called sill, which is equal to σ^2
- Distance at which the upper bound is reached is referred to as the range
- Failure to exhibit an upper bound indicates some nonstationarity
- Theorectically $\gamma(0)=0$, but sampling errors and small variability may cause sample values with small separations to be quite dissimilar, which is known as nugget effect





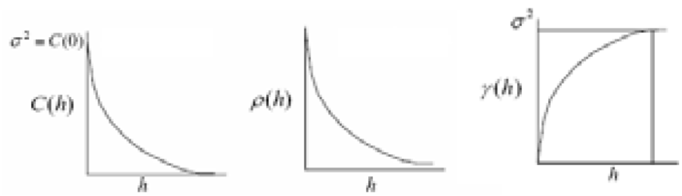


Covariogram, Correlogram and Variogram

- For a stationary process, the following relationships can be found

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

$$\gamma(h) = \sigma^2 - C(h)$$



covariogram correlogram variogram

Properties

$$C(0) = \sigma^2 \qquad \gamma(0) = 0$$

$$C(h) = C(-h) \qquad \gamma(h) = \gamma(-h)$$

$$as\ h \to 0 : C(h) \to \sigma^2 - c_0 \qquad \gamma(h) \to c_0$$

$$as \|h\| \to \infty : if\ \ C(h) \to 0 \qquad then\ \gamma(h) \to \sigma^2 [sill]$$

Properties

C positive-definite

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j C(\mathbf{s}_i - \mathbf{s}_j) \geq 0\,;$$

any m, any $\{\mathbf{s}_i\}$, any $\{a_i\}$

$\Rightarrow$ var$(\Sigma a_i Z(\mathbf{s}_i)) \geq 0$

γ conditionally negative-definite

$$-\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \geq 0\,;$$

any m, any $\{\mathbf{s}_i\}$, any $\{a_i\}$

s.t. $\sum_{i=1}^{m} a_i = 0$

$\Rightarrow$ var$(\Sigma a_i Z(\mathbf{s}_i)) \geq 0$

s.t. $\Sigma a_i = 0$

Conditional Negative-Definiteness

$$\begin{aligned}\sum_i a_i = 0 &= \sum_j b_j \Rightarrow \{\sum_i a_i Z(\mathbf{s}_{1i})\}\{\sum_j b_j Z(\mathbf{s}_{2j})\}\\ &= -(1/2)\sum_i\sum_j a_i b_j\{Z(\mathbf{s}_{1i}) - Z(\mathbf{s}_{2j})\}^2\end{aligned}$$

Suppose $Z(\cdot)$ intrinsically stationary and $\Sigma a_i = \Sigma b_j = 0$

$$\Rightarrow \mathrm{cov}(\sum_i a_i Z(\mathbf{s}_{1i}), \sum_j b_j Z(\mathbf{s}_{2j})) = -\sum_i\sum_j a_i b_j \gamma(\mathbf{s}_{1i} - \mathbf{s}_{2j})$$

Hence

$$\mathrm{var}(\sum_i a_i Z(\mathbf{s}_i)) = -\sum_i\sum_j a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j)\,,$$

provided $\sum_i a_i = 0$.

Example

$Z(\cdot)$ intrinsically stationary. Data $Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)$. Predict $Z(\mathbf{s}_0)$.
Use predictor $\sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i)$ S.T. $\sum_{i=1}^{n} \lambda_i = 1$
Then,

$$\begin{aligned} \text{MSPE} = E(Z(\mathbf{s}_0) - \Sigma\lambda_i Z(\mathbf{s}_i))^2 &= \text{var}\left(\sum_{i=0}^{n} a_i Z(\mathbf{s}_i)\right) \\ &= -\sum_{i=0}^{n}\sum_{j=0}^{n} a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j)\,; \end{aligned}$$

$a_0 \equiv 1$, $a_i \equiv -\lambda_i$. Notice that $\sum_{i=0}^{n} a_i = 1 - \sum_{i=1}^{n} \lambda_i = 0$

Sample Variogram

- Sample estimator of the variogram is

$$2\hat{\gamma}(h) = \frac{1}{n(h)} \sum_{s_i - s_j = h} (y_i - y_j)^2$$

- For a isotropic process, the sample variogram is estimated

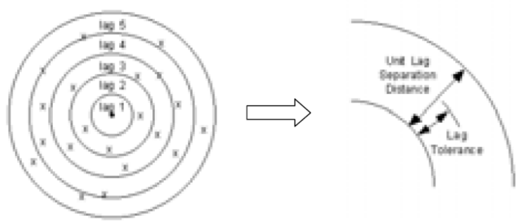
$$2\hat{\gamma}(h) = \frac{1}{n(|h|)} \sum_{s_i - s_j = |h|} (y_i - y_j)^2$$

- Sample covariogram is estimated as

$$\hat{C}(h) = \frac{1}{n(h)} \sum_{s_i - s_j = h} (y_i - \bar{y})(y_j - \bar{y})$$

Sample Variogram

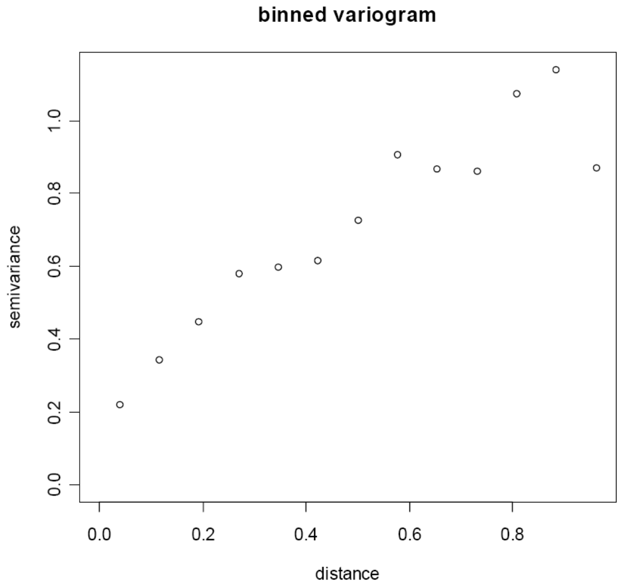
- In general, *n(h)* will increase as |*h*|, so the reliability of the sample estimate of variogram or covariogram decreases as |*h*| increases
- For irregularly spaced sample points, there will rarely be enough observations with an exact separation vector *h*. Tolerance regions are usually used



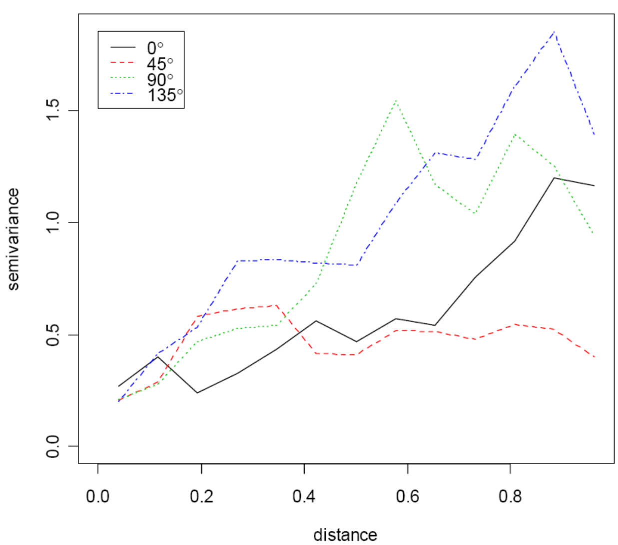
Exploratory Variogram Analysis

- Start with the estimation of isotropic variogram or covariogram
  - To get appropriate distances for variogram
- Then proceed to calculate directional variograms in several broad directions
  - To check the possible directional effects or anisotropy
- Further analysis on the plot of $(y_i - y_j)^2$ vs. distances h, which is called variogram cloud
  - To detect possible outliers and skewness

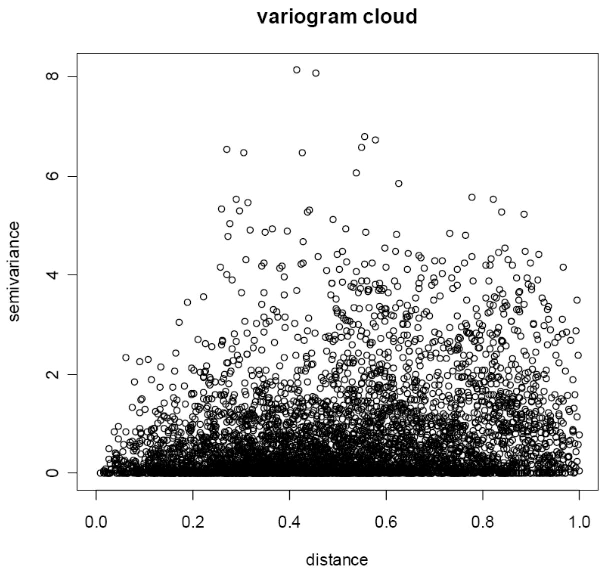
Sample Variogram



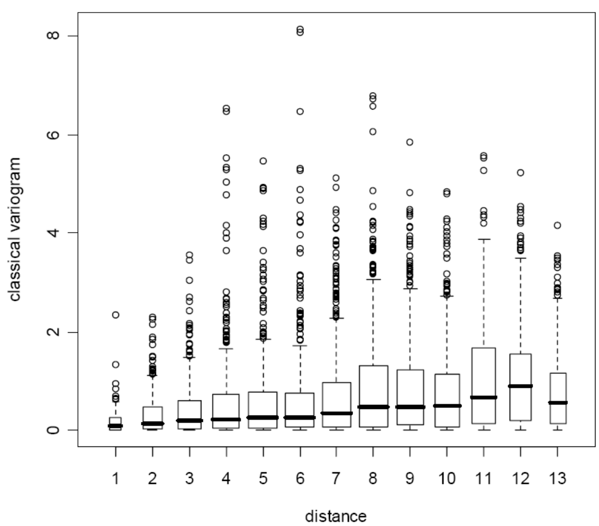
Directional Variograms



Variogram Cloud



Variogram Cloud (Binned)



Models for Variogram

- The assumption for variogram and covariogram is (intrinsic) stationary. If non-stationary exists, the estimate will be biased. However, variogram is more robust to minor departure from first order stationarity
- To fully model the first and second order effects, we neeed the knowlede of covariogram to fit the generalized linear model
- In practice we model variogram and then derive the covariogram based on the relationship between them under the assumption of stationarity

Models for Variogram

- Need to assume some of stationarity to obtain repetition
  - First order can be adjusted
  - Second order stationarity must be assumed
- If first order effect exists, the trend should be removed using ordinary regression models and fit the variogram model to the sample variogram of the residuals. This is usually implemented in an interactive mode
  - Remove trend by ordinary least squares
  - Use residuals to model covariance structure
  - Re-estimate trend using generalized least squares with the estimated covariance structure

Models for Variogram

- Necessary and sufficient conditions for covariance function (c) of a general spatial stochastic process
  - Symmetry

$$C(s_i, s_j) = C(s_j, s_i)$$

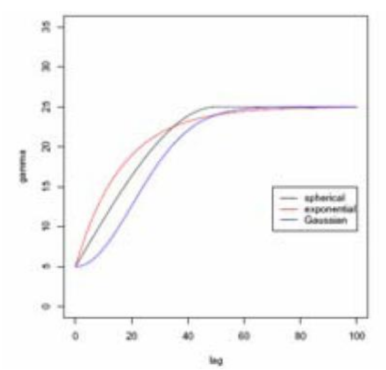
  - Non-negative definiteness

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j C(s_i, s_j) \geq 0$$

- For stationary process, the valid covariance function is a subset of all the functions satisfying the above conditions

Models for Variogram

- Families of valid variogram models
  - Spherical model
  - Exponential model
  - Gaussian model
  - Linear model



Models for Variogram

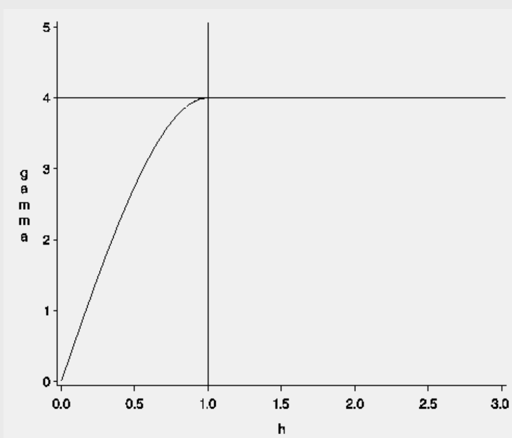
- Spherical model without nugget effect

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

- Spherical model with nugget effect

$$\begin{cases}sill: & \sigma^2\\ range: & r\\ nugget: & a\end{cases}$$

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



$r=1$

$\sigma^2=4$

Exponential Model

- Without nugget effect

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

$$\begin{cases} sill: & \sigma^2 \\ range: & r \\ nugget: & a \end{cases}$$

- With nugget effect

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

Gaussian Model

- Without nugget effect

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

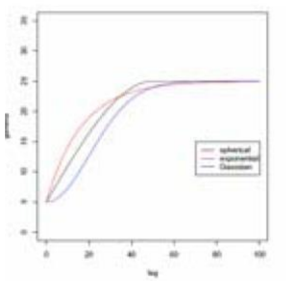
$$\begin{cases} sill: & \sigma^2 \\ range: & r \\ nugget: & a \end{cases}$$

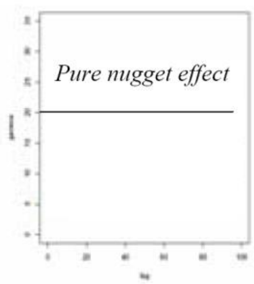
- With nugget effect

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

Nugget effect

- Relative nugget effect
  - The ratio of nugget effect to the sill
- Pure nugget effect
  - Complete lack of spatial dependence

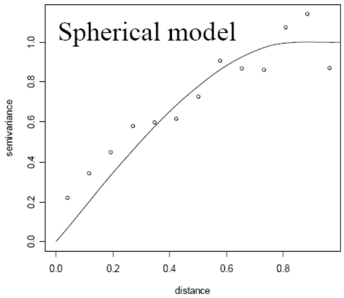




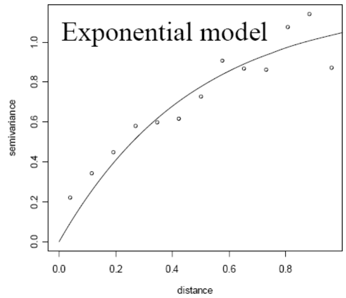
Models for Variogram

- Anisotropic covariance structures
  - Replace the |h| in the previous models with the vector h
- Geometric anisotropy
  - Range changes with direction but sill remains constant
- Zonal anisotropy
  - Sill changes with direction but range remains constant

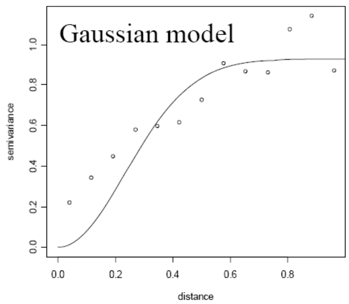
Examples



Spherical model
semivariance
distance



Exponential model
semivariance
distance



Gaussian model
semivariance
distance