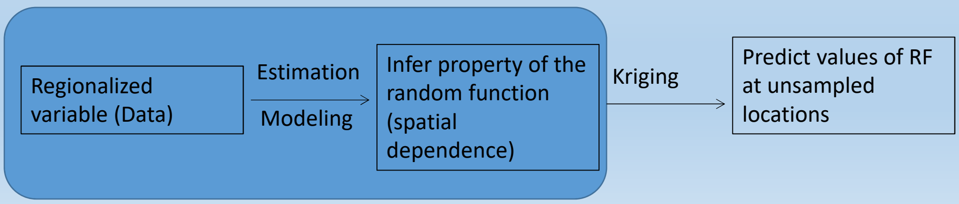
COA 616 Geostatistics in Environmental Sciences

Lecture 5 – Modelling the Variogram

Wei Wu

October 9, 2018



Why models

1. We need a variogram to describe the variance of the region.
2. Estimating/predicting values at unsampled places and in larger blocks of land optimally by kriging requires semivarioances at lags for which we could not calculate directly.

Mathematical consideration

The model we choose must describe random variation, and the function must be such that it will not give rise to negative variance of combination of random variables.

Consider a general linear function of the data

$$y = \sum_{i=1}^{N} \lambda_i z(\vec{x}_i) \quad \lambda_i : \text{Arbitrary weights for each data value}$$

$$\text{var}(y) = \lambda^T C \lambda \quad C : \text{covariance matrix}$$

$$\text{var}(y) = \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j C(\vec{x}_i, \vec{x}_j) \geq 0 \Rightarrow C \text{ must be positive semidefinite}$$

Model consideration cont.

$$\text{var}(y) = \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j C(\vec{x}_i, \vec{x}_j) \geq 0 \Rightarrow C \text{ must be positive semidefinite}$$

$$\text{var}(y) = C(0) \sum_{i=1}^{N} \lambda_i \sum_{j=1}^{N} \lambda_j - \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \gamma(\vec{x}_i, \vec{x}_j)$$

If the variable is intrinsic only, we can eliminate C(0) by making the weights sum to 0 without loss of generality, then

$$\text{var}(y) = -\sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \gamma(\vec{x}_i, \vec{x}_j) = -\lambda^T \gamma \lambda \quad \text{Conditional negative semidefinite (CNSD)}$$

Any function that we use to model the semivariogram must produce a semivariance matrix that is CNSD.

Unbounded model

$$\gamma(h) = \mathrm{wh}^{\alpha} \text{ for } 0 < \alpha < 2$$

$$\mathrm{Z(x+h)} = \beta \mathrm{Z(x)} + \varepsilon$$

$$2\gamma(h) = E\{[Z(x+h) - Z(x)]^2\} = |h|^k$$

$$k = 1 \quad \text{Random walk}$$

$$\alpha = 0 \quad \text{Pure nugget model}$$

Bounded models

Linear model

$$\gamma(\vec{h}) = \begin{cases} c\dfrac{\vec{h}}{a} & \text{when } \vec{\mathrm{h}} \le a \\ \mathrm{c} & \text{when h > a} \end{cases}$$

a – range

c – sill

Only apply to transect data. It is CNSD at 1 dimension only. It may not be used to describe variation in 2 or 3 dimensions.

Bounded models cont.

Circular model

$$\gamma(\vec{h}) = \begin{cases} c\{1 - \frac{2}{\pi}\cos^{-1}(\frac{\vec{h}}{a}) + \frac{2\vec{h}}{\pi a}\sqrt{1 - \frac{\vec{h}^2}{a^2}}\} & \text{when } \vec{h} \le a \\ c & \text{when } \vec{h} > a \end{cases}$$

a – range
c – sill
Gradient at the origin = $\frac{4c}{\pi a}$

It is CNSD in 1 dimension and 2 dimensions, but not 3 dimensions.
Geometric significance: Derived from area of intersection of two circles with diameter = a and separated by **h**.

Bounded models cont.

Spherical model

$$\gamma(\vec{h}) = \begin{cases} c\{\frac{3h}{2a} - \frac{1}{2}(\frac{h}{a})^3\} & \text{when } \vec{h} \le a \\ c & \text{when } \vec{h} > a \end{cases}$$

a – range
c – sill
Gradient at the origin = $\frac{3c}{2a}$
Curves up more gradually
It is CNSD in 1 dimension, 2 dimensions and 3 dimensions
It is one of the most frequently used models.
It represents patches with random locations but similar sizes. The average of diameter of the patches is represented by a.

Bounded models cont.

Exponential model

$$\gamma(\vec{h}) = \quad c\{1-\exp(\frac{\vec{h}}{r})\}$$

c – sill

Another commonly used model

The function approaches its sill asymptotically.

Effective range = 3r

Slope at the origin = $\frac{c}{r}$

Linear at short distance, but rises more steeply and flatterns out more gradually.

Represents patches with random locations and sizes.

Bounded models cont.

Gaussian model

$$\gamma(\vec{h}) = \quad c\{1-\exp(-\frac{\vec{\mathrm{h}}^2}{r^2})\}$$

c – sill

It approaches its sill asymptotically

Effective range = $\sqrt{3}r$

Disadvantage: the parabolic behavior at the origin, which can cause bizarre estimation using kriging.

Solution: Always have a nugget variance, or use

$$\gamma(\vec{h}) = \quad c\{1-\exp(-\frac{\vec{\mathrm{h}}^\alpha}{r^\alpha})\} \quad 1<\alpha<2$$

CNSD in 1, 2 and 3 dimensions

Bounded models cont.

Hole-effect model

$$\gamma(\vec{h}) = \quad W\{1 - \frac{\omega}{2\pi \vec{h}} \sin(\frac{2\pi \vec{h}}{\omega})\}$$

W – amplitude

Ω – Wavelength (average spacing)

Hole-effect models apply when patches are regularly spaced.

Disadvantage: gradient at the origin is 0. CNSD in 1 dimension.

Solution: combined with other semivariogram models

$$\gamma(\vec{h}) = \quad W\{1 - \frac{1}{\theta} \sin(\theta)\}$$

CNSD in 1, 2, 3 dimensions

Model fitting

Visual fitting

Minimize $\sum_{j=1}^{n} w_j(\hat{\gamma}(\bar{h}_j) - \gamma(\bar{h}_j))^2$ or Akaike Information Criterion (AIC)

fit	fit by	weight
0	-	- (no fit)
1	gstat	N_j
2	gstat	$N_j/\{\gamma(h_j)\}^2$
3	gnuplot	N_j
4	gnuplot	$N_j/\{\gamma(h_j)\}^2$
5	gstat	REML
6	gstat	no weights (OLS)
7	gstat	N_j/h_j^2

Table 4.2: values for `fit`