Variograms

Modeling & interpolating
spatial dependencies

Ginger Allington - Analyses in R - 17 March 2010
Spatial patterns

• Most ecological field studies are inherently spatial, but this factor is not incorporated into analysis (e.g. anova)

• Standard statistics can miss important trends in the data

• Spatial methods *use* the underlying spatial variations to create better estimates of differences among treatments or sites

(Scheiner & Gurevitch 2001)
Similar mean, but the patterns of spatial dependency are different

(Images from R. Barnes, Golden Software)
Variograms of the two datasets

Figure 1.5 Data Set A
Variogram and Model

Figure 1.6 Data Set B
Variogram and Model
• A method to characterize spatial variance / quantify spatial dependencies in the data

• Semivariograms present the mean variance found in comparisons of samples of increasing lag intervals (distance).

• The semi-variogram is a function that relates semi-variance (or dissimilarity) of data points to the distance that separates them.

\[ \gamma(\Delta x, \Delta y) = \frac{1}{2} \mathbb{E} \left[ (Z(x + \Delta x, y + \Delta y) - Z(x, y))^2 \right] \]
Interpreting variograms

- The distance between the origin and the sill is known as the **range**, and this represents the general distance over which the samples are autocorrelated.

- We can only model lag distances to the smallest distance between pairs of samples, and variance that exists at an even smaller scale is represented by the **nugget**.

- The **expected curve** for random distribution is a straight line.

- **Expected curve** when samples show auto-correlation over a certain range (Ao)
Interpreting variograms (con’t)

When data are randomly distributed we can expect that there will be little difference in the variance ($\gamma$) at any distance comparison.

However, when there is a pattern present in the distribution, we can expect that variance will increase with comparisons of close, autocorrelated samples, but will level off to form a sill when samples become independent.

The nugget:sill ratio indicates what percent of the overall variance is found at a distance smaller than the smallest lag interval, and gives a sense of how much variance you have successfully accounted for in the model.
Variogram components

- Nugget variance: a non-zero value for $\gamma$ when lag distance ($h$) = 0. Produced by various sources of unexplained error (e.g. measurement error).
- 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).
- In general, 30 or more pairs per point are needed to generate a reasonable sample variogram.
- The most important part of a variogram is its shape near the origin, as the closest points are given more weight in the interpolation process.

© Arthur J. Lembo, Jr. Cornell University
Variogram models

A. Spherical

\[ \gamma = C_0 + C_1 \left[ 1.5 \frac{h}{a} - 0.5 \left( \frac{h}{a} \right)^3 \right]; \quad h \leq a \]
\[ \gamma = C_0 + C_1; \quad h > a \]

B. Linear

\[ \gamma = C_0 + C_1 \frac{h}{a} \]

C. Exponential

\[ \gamma = C_0 + C_1 \left[ 1 - \exp \left( -\frac{h}{b} \right) \right] \]

D. Gaussian

\[ \gamma = C_0 + C_1 \left[ 1 - \exp \left( \frac{-h^2}{2b^2} \right) \right] \]

© Arthur J. Lembo, Jr.  
Cornell University
Spatial patterns across treatments, or sites

Figure 2. Schematic of sampling design. 9x12m plot subdivided into a 1.5 m² grid. Four 2x2 m plots subdivided into a 0.5m² grid. Soil sample locations randomly cast into each grid square.
Anisotropy

- Higher spatial autocorrelation in one direction than in others
- Variation in continuity in different directions
- Can be accounted for by calculating variograms in different directions.
Interpolation via kriging