

Applied Spatial Statistics in R, Section 5

Geostatistics

Yuri M. Zhukov

IQSS, Harvard University

January 16, 2010

Outline

- ➊ Introduction
 - Why use spatial methods?
 - The spatial autoregressive data generating process
- ➋ Spatial Data and Basic Visualization in R
 - Points
 - Polygons
 - Grids
- ➌ Spatial Autocorrelation
- ➍ Spatial Weights
- ➎ Point Processes
- ➏ Geostatistics
- ➐ Spatial Regression
 - Models for continuous dependent variables
 - Models for categorical dependent variables
 - Spatiotemporal models

Geostatistics

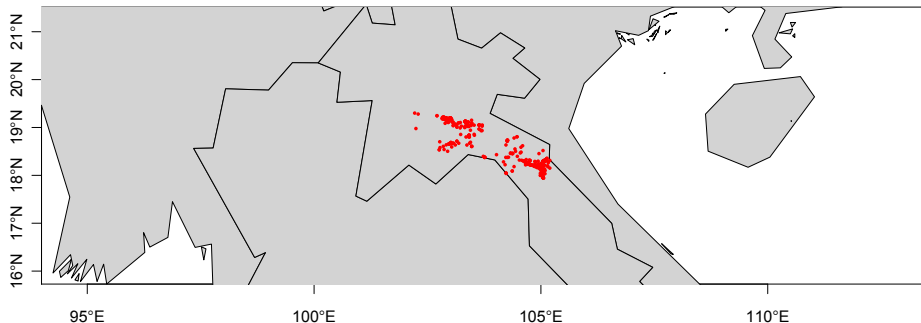
What if the pattern of point locations is not of primary interest? You may wish to...

- determine where new data should be collected,
- identify which observations are spatial outliers,
- perform spatial prediction,
- interpolate missing data from nearby observed locations,
- estimate local averages of spatially autocorrelated variables.

These problems are the domain of a subfield called geostatistics.

Geostatistics

- Let's take a look at some data on U.S. Air Strikes in Laos during the Vietnam War.
- The variable we are interested is `LOAD_LBS`, the payload of each bomb dropped.



Geostatistics: IDW Interpolation

- Spatial interpolation is the prediction of values of attributes at unsampled locations x_0 from existing measurements at x_i .
- This procedure converts a sample of point observations into an alternative representation, such as a contour map or grid.
- One approach to interpolation is to use a locally-weighted average of nearby values.
- Inverse-distance weighted (IDW) interpolation computes one such weighted average:

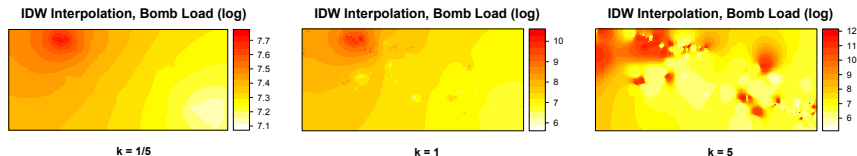
$$\hat{Z}(x_0) = \sum_{i=1}^n w_{i0} Z(x_i)$$

where weights w_{ij} are determined according to the distance between points x_i and x_j , and scaled by parameter k .

$$w_{ij} = \frac{1}{d_{ij}^k}$$

Geostatistics: IDW Interpolation

- Predicted values and variance for Laos bombing data are shown below.
- Values of $k > 1$ reduce the relative impact of distant points and produce a peaky map.
- Values of $k < 1$ increase the impact of distant points and produce a smooth map.



Geostatistics: Variogram

- In geostatistics, spatial autocorrelation has traditionally been modelled by a variogram, which describes the degree to which nearby locations have similar values.
- A variogram cloud is a scatterplot of data pairs, in which the semivariance is plotted against interpoint distance.
- The semivariance $\gamma(d)$ is formally defined as the squared difference in height between locations:

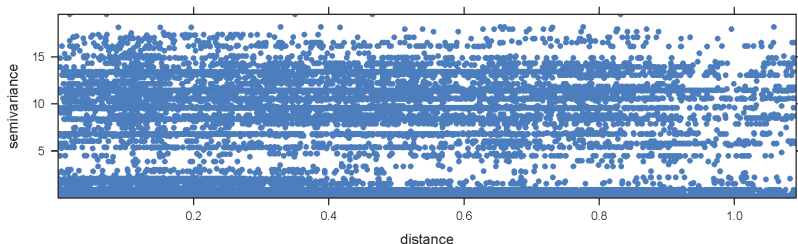
$$\hat{\gamma}(d) = \frac{1}{2n(d)} \sum_{d_{ij}=d} (Z(x_i) - Z(x_j))^2$$

- where $n(d)$ is the number of point pairs separated by distance d , and $Z(x_i)$ is the value of a variable at location x_i .

Geostatistics: Variogram

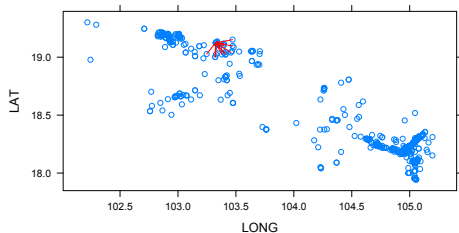
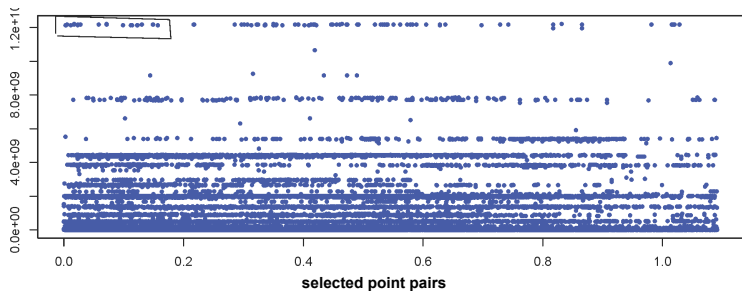
Below is the variogram cloud for Laos bomb load (natural log).

- **Upper left corner:** point pairs are close together, but have very different values.
- **Lower left corner:** close together, similar values.
- **Upper right corner:** far apart, different values.
- **Lower right corner:** far apart, similar values.



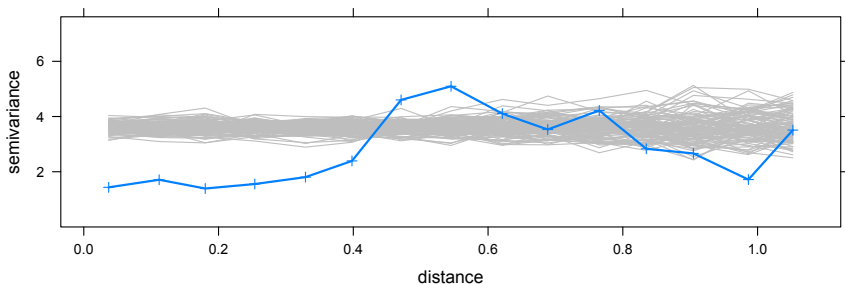
Geostatistics: Variogram

A variogram can be used to identify outliers...



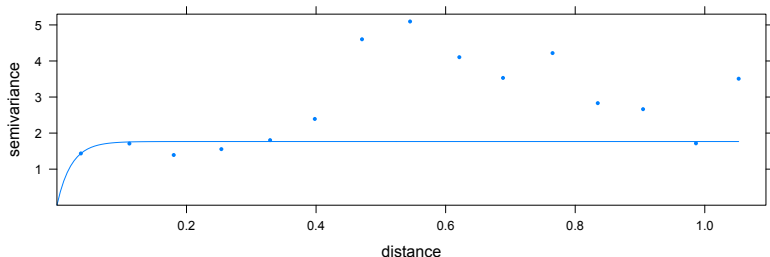
Geostatistics: Variogram

To test the null hypothesis that an increase in semivariance with distance is due to chance, we can use simulation to generate 100 spatially random datasets and check whether the sample variogram falls within the range of the random variograms. As shown below, the CSR hypothesis seems unlikely for the Laos bombing data.



Geostatistics: Variogram

- The variogram can be used for spatial prediction.
- This can be done by fitting a parametric model to the variogram.
- In the Laos example below, an exponential model was used.
- The shape of the curve indicates that at small separation distances, the variance in z is small. After a certain level of separation (.5 degrees), the variance in z values becomes somewhat random and the model flattens out to a value corresponding to the average variance.



Geostatistics: Ordinary Kriging

- Kriging is used to interpolate a value $Z(x_0)$ of a random field $Z(x)$ at unobserved location x_0 , using data from observed location x_i .
- Allows variance to be non-constant, dependent on distance between points as modeled by the variogram $\gamma(d)$.
- The kriging estimator is given by

$$\hat{Z}(x_0) = \sum_{i=1}^n w_i(x_0) Z(x_i)$$

where $w_i(x_0)$, $i = 1, \dots, n$ is a spatial weight.

- Kriging is very similar to IDW interpolation, expect that the weights used in kriging are based on the model variogram, rather than an arbitrary function of distance.

Geostatistics: Ordinary Kriging

- To interpolate at a point x_0 based on points x_1, \dots, x_n , the weights w_1, \dots, w_n must be found. This can be done by solving the system of linear equations:

$$\begin{bmatrix} \gamma(d_{11}) & \gamma(d_{12}) & \cdots & \gamma(d_{1n}) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(d_{n1}) & \gamma(d_{n2}) & \cdots & \gamma(d_{nn}) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \lambda \end{bmatrix} \begin{bmatrix} \gamma(d_{10}) \\ \vdots \\ \gamma(d_{n0}) \\ 1 \end{bmatrix}$$

where $\gamma(d_{ij})$ is the semivariance for the distance between points x_i and x_j , and λ is the trend parameter.

- Ordinary kriging assumes an unknown constant trend: $\lambda(x) = \lambda$.

Geostatistics: Ordinary Kriging

Once weights are estimated, interpolation by ordinary kriging is given by:

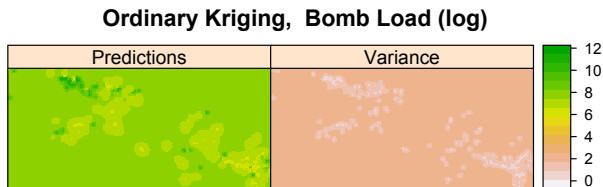
$$\hat{Z}(x_0) = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}' \begin{pmatrix} Z(x_1) \\ \vdots \\ Z(x_n) \end{pmatrix}$$

The ordinary kriging error is given by:

$$\text{var} \left(\hat{Z}(x_0) - Z(x_0) \right) = \begin{pmatrix} w_1 \\ \vdots \\ w_n \\ \lambda \end{pmatrix}' \begin{pmatrix} \gamma(d_{10}) \\ \vdots \\ \gamma(d_{n0}) \\ 1 \end{pmatrix}$$

Geostatistics: Ordinary Kriging

- Predicted values and variance for ordinary kriging is shown below for the Laos bombing data.



Examples in R

Switch to R tutorial script. Section 5.