STAT 6560
Graphical Methods

Spring Semester 2009

Project Work

James B. Odei

Utah State University

Department of Mathematics and Statistics

3900 Old Main Hill

Logan, UT 84322–3900

April 10, 2009

Graphics for Spatial Continuous Data

Example from R packages – *gstat*.

*g*stat Package in R

Much of this project focuses on the R package *g***stat**. This is because the *g***stat** package provides a wide range of functionality in geostatistics curriculum for R, and for larger datasets: – it covers variogram cloud diagnostics, variogram modeling, everything from global simple kriging to local universal cokriging, multavariate geostatistcs, block kriging, indicator and Gaussian conditional simulation, and many combinations. Other R packages that provide additional geostatistical functionality will be mentioned where relevant and discussed. Table 1 shows a summary of some user functions in the *g***stat** package.

Example 1: *The Variogram (or Covariogram) and its Modeling.*

The R help page for the function `variogram` under this pacakge indicates:

"Calculates the sample variogram from data, or in case of a linear model is given, for the residuals, with options for directional, robust, and pooled variogram, and for irregular distance intervals."

Formally, if we have a spatial stochastic process $\{Y(\mathbf{s}), \mathbf{s} \in \Re\}$, where we denote $E[Y(\mathbf{s})]$ as $\mu(\mathbf{s})$ and $VAR[Y(\mathbf{s})]$ as $\sigma^2(\mathbf{s})$, then the covariance process at any two particular points $\mathbf{s}_i$ and $\mathbf{s}_j$ is defined in Bailey & Gatrell (1995) as:

$$C(\mathbf{s}_i, \mathbf{s}_j) = E[(Y(\mathbf{s}_i) - \mu(\mathbf{s}_i))\,(Y(\mathbf{s}_j) - \mu(\mathbf{s}_j))] \tag{1}$$

with corresponding correlation defined as:

$$\rho(\mathbf{s}_i, \mathbf{s}_j) = \frac{C(\mathbf{s}_i, \mathbf{s}_j)}{\sigma(\mathbf{s}_i)\sigma(\mathbf{s}_j)} \tag{2}$$

NOTE:

- $C(\mathbf{s}, \mathbf{s}) = \sigma^2(\mathbf{s})$.
- If $\mu(\mathbf{s}) = \mu$, and $\sigma^2(\mathbf{s}) = \sigma^2$, process is said to be *stationary*.

- In addition, $C(\mathbf{s}_i, \mathbf{s}_j) = C(\mathbf{s}_i - \mathbf{s}_j) = C(\mathbf{h})$, often referred to as *covariogram*.
- For stationary process, the variogram,$\gamma(\mathbf{h})$, and covariogram, $C(\mathbf{h})$, are relataed as $\gamma(\mathbf{h}) = \sigma^{\mathbf{2}} - \mathbf{C}(\mathbf{s})$.
- The variogram has the same shape as the covariogram, except that it is "inverted". The variogram plots semivariance as a function of distance. It starts at 0 and increases, at a distance referred to as the *range*, to a maximum of σ^2, often referred to as the *sill*.

The *variogram* or *covariogram* is used simply as an exploratory device to examine spatial dependence in the observed data. They also play a major role in the modeling of such data. For purpose of illustration, we will consider the eight-hour average ozone (O_3) concentration measurements for a given day in the summer of 1987 based on the dataset from 21 ambient ozone monitoring stations in the Chicago area (Figure 1). This data can be found at Dr. Mevin Hooten's webpage: `http://www.math.usu.edu/~hooten/stat6410fall08/`.

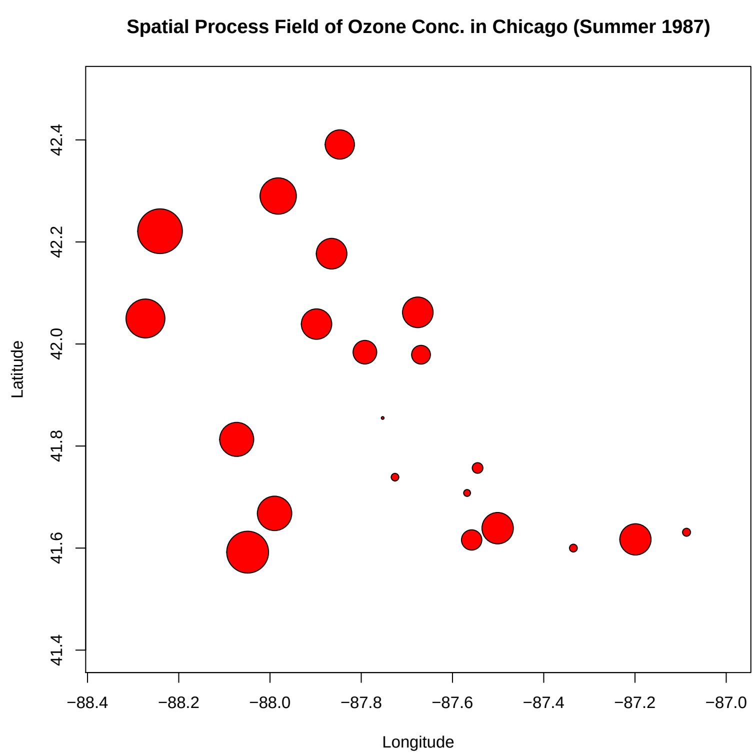

Figure 1: Ozone concentration measurements from 21 ambient monitoring stations in the Chicago area.

Table 1: Taken from Pebesma (2004), p.686.

User functions in package gstat

Gstat	Add variable definition to gstat object
Variogram modelling	
`variogram`	Calculate sample variogram, directional sample variograms, or direct and cross variograms
`fit.variogram`	Fit variogram model coefficients to sample variogram
`fit.lmc`	Fit a linear model of coregionalisation to direct and cross variograms
`variogram.line`	Calculates variogram values from a variogram model
Prediction/simulation	
`predict.gstat`	Spatial prediction or simulation, see also Fig. 3
`krige`	Univariable wrapper around `gstat` and `predict.gstat`
`krige.cv`	Leave-one-out or *n*-fold cross-validation wrapper for `krige`
`zerodist`	Detect observation pairs with identical locations
Plotting	
`bubble`	Bubble scatter plot for data or residuals (using colour for sign, size for value)
`plot.variogram`	Plot sample variogram (optional with number of point pairs) and fitted model; uses conditioning plots for directional or multivariable variograms (Fig. 2)
`plot.variogram.cloud`	Plot variogram cloud, with options for interactive point pairs identification
`plot.point.pairs`	Plot point pairs, identified by `plot.variogram.cloud`, in a map
`image.data.frame`	Draw image for (x,y,z) values, stored in columns of a data frame
`map.to.lev`	Stack data in the form $(x,y,z_1,z_2,\ldots,z_n)$ to a form, suitable for plotting with `levelplot`
`mapasp`	Calculate aspect ratio for geographically correct levelplot

Example 2: *Spatial Prediction (Universal, Ordinary, and Simple Kriging)*

The R help page for the function `krige` under this pacakge indicates:

"Function for simple, ordinary or universal kriging (sometimes called external drift kriging), kriging in a local neighbourhood, point kriging or kriging of block mean values (rectangular or irregular blocks), and conditional (Gaussian or indicator) simulation equivalents for all kriging varieties, and function for inverse distance weighted interpolation."

Spatial prediction refers to the prediction of unknown quantities $\{Y(\mathbf{s}_0)\}$, based on sample data $\{Y(\mathbf{s}_i)\}$ and assumptions regarding the form of the trend of Y and its variance and spatial correlation, Bivand et al. (2008). Suppose the trend can be written as a linear regression function as:

$$Y(\mathbf{s}) = \mathbf{X}\beta + \mathbf{e}(\mathbf{s}) \tag{3}$$

with $X_j(\mathbf{s})$, the known spatial regressors, forming the columns of the $n \times (p+1)$ design matrix $\mathbf{X}$ and β_j, unknown regression coefficients. If the predictor values for $\mathbf{s}_0$ are available in the $1 \times p$ row-vector $x(\mathbf{s}_0)$, $\mathbf{V}$ is the covariance matrix of $Y(\mathbf{s})$ and v the covariance vector of $Y(\mathbf{s})$ and $Y(\mathbf{s}_0)$, then the best linear unbiased predictor of $Y(\mathbf{s}_0)$ is

$$\hat{Y}(\mathbf{s}_0) = x(\mathbf{s}_0)\hat{\beta} + v'\mathbf{V}^{-1}(Y(\mathbf{s}) - \mathbf{X}\hat{\beta}) \tag{4}$$

with $\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}Y(\mathbf{s})$ and $v'\mathbf{V}^{-1}$ is known as *simple kriging weights*.

NOTE:

- The instances of the best linear unbiased prediction method with the number of predictors $p > 0$ are usually called *universal kriging*.
- For the case where $p = 0$ and $X_0 \equiv 1$, the corresponding prediction is called *ordinary kriging*.
- *Simple kriging* is obtained when, for whatever reason, β is a *priori* assumed to be known.

To demonstrate spatial predictions, we will rely on the rabbit burrow dataset. This dataset is part of efforts underway to monitor the Pygmy Rabbit (*Brachylagus idahoensis*, petitioned to be listed under the Endagered Species Act) at a large field site in Rich County, Utah. The spatially explicit data (both response and covariate) have been collected in order to learn more about Pygmy Rabbit occupancy, detection, and possibly, abundance. Burrows and rabbit pellets were sampled along five 200m line transects at each thirty eight 4-ha sites within a 13,500-ha study area. Covariates were chosen based on those that have been shown to be important for predicting pygmy rabbit occurrence in previous studies. This is data was made available by Tammy Wilson of the Department of Wildland Resources, Utah State University.

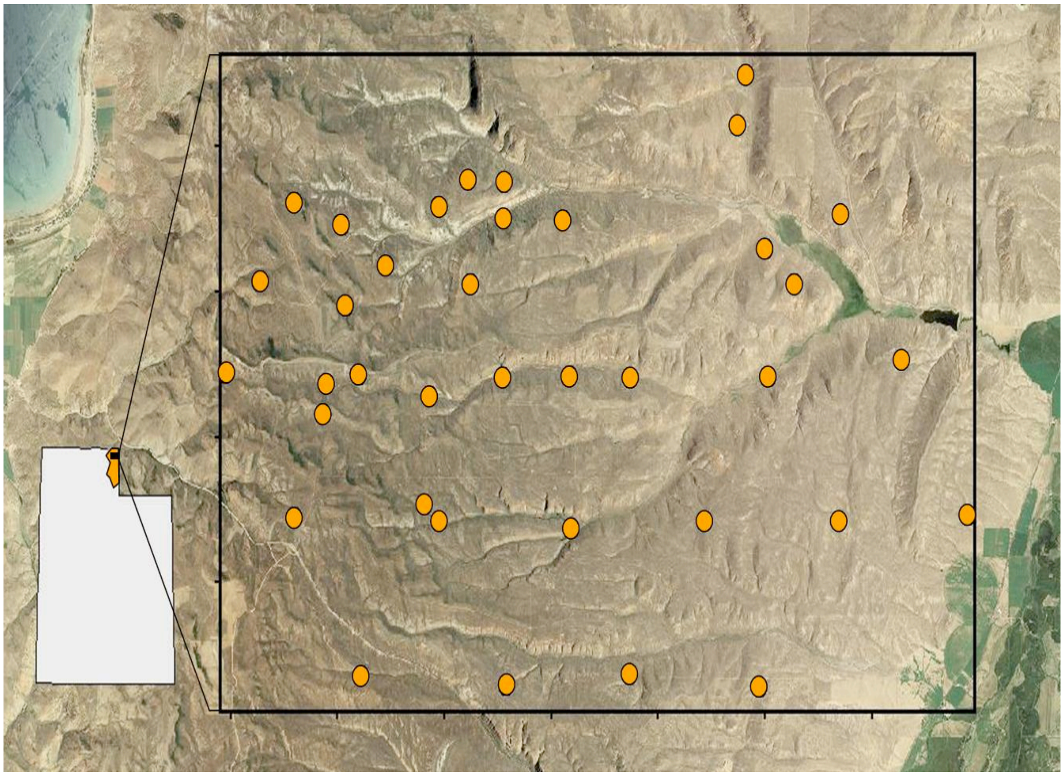

Figure 2: The 13,500-ha location in Rich County, Utah, where distance sampling were used to evaluate burrows and the presence of rabbit pellets.

Other R Packages for Interpolation and Geostatistics

- `geoR`, `geoRglm`, `spatial`, `RandomFields`, `mgcv`, `fields`, `akima`, `stinepack`.

Links to Data file and R-code:

- `http://www.math.usu.edu/~symanzik/teaching/2009_stat6560/RDataAndScripts/odei_james_project2_ozone.txt`
- `http://www.math.usu.edu/~symanzik/teaching/2009_stat6560/RDataAndScripts/odei_james_project2_rabbit.txt`
- `http://www.math.usu.edu/~symanzik/teaching/2009_stat6560/RDataAndScripts/odei_james_project2_rabbit.grid.txt`
- `http://www.math.usu.edu/~symanzik/teaching/2009_stat6560/RDataAndScripts/odei_james_project2.R`

References

Bailey, T. C. & Gatrell, A. C. (1995), *Interactive Spatial Data Analysis*, Prentice Hall, Harlow, England.

Bivand, R. S., Pebesma, E. J. & Goméz-Rubio, V. (2008), *Applied Spatial Data Analysis with R*, Springer, New York, NY.

Pebesma, E. J. (2004), 'Multivariable Geostatistics In S: The gstat Package', *Computers & Geosciences* **30**, 683–691.

- `http://cran.nedmirror.nl/web/views/Spatial.html`