Module 9: Spatial Statistics

9.1 Spatial Data Analysis

Spatial Autocorrelation

- Spatial autocorrelation is similar to temporal autocorrelation
- Data that are positively autocorrelated in space are more similar the closer together they are
- As data points become further and further apart, they are less alike
- Data that are negatively autocorrelated in space are more similar the more distant they are. That is much less common.

Spatial Autocorrelation

- Examples of positively autocorrelated data:
  - soil characteristics
  - makeup of plant communities
  - contamination that has spread from a point source
  - geology
  - etc

Spatial Autocorrelation

- Like time series analysis, the field of spatial data analysis is large and complex
- This module will give an overview of some of the concepts and techniques as an introduction to the topic

Quadrat Counts

- Some spatial data involves gridding the study area (dividing it into cells) and counting the occurrences of one or more variables in each grid cell
- Interesting questions include:
  - Is the variable of interest randomly distributed across the study area, is it clustering, or is it uniform?
  - Are there correlations between different variables?

Module 9.1

Quadrat Counts

- If the data are randomly distributed, then the counts should follow a Poisson distribution

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

where μ is both the mean and variance of the distribution

Module 9.1

Quadrat Counts

- Then $$R = \frac{S^2}{\overline{X}}$$
- should equal
  - 1 if the counts are randomly distributed
  - less than 1 if they are uniformly distributed
  - greater than 1 if they are clustering

Quadrat Counts

- The null hypothesis is that the data are randomly distributed over the area of study
- The test statistic is $$T = \frac{R-1}{\sqrt{\frac{2}{n-1}}}$$
- which has a t distribution with n-1 degrees of freedom

Quadrat Counts - Example

				Distance East/West				
	30	0	0	8	4	1	0	0
Distance	20	1	2	13	7	3	1	0
North/South	10	2	6	20	14	6	1	0
	0	0	4	12	18	8	2	1
	-10	1	3	9	10	15	2	0
	-20	0	1	3	6	7	9	3
	-30	0	2	1	2	4	3	6
		-30	-20	-10	0	10	20	30
				Mean	4.5			
				Variance	25.38			
				R	5.6			
				T (0.05, 48)	22.67			
				Reject null hypothesis.				
	Counts are not randomly distributed - clusters exist.							

Module 9.1

Spatial Data Analysis

- Calculating the correlation between two spatial data variables is difficult and requires specialized software
- There are several ways to do spatial data analysis that involve computer intensive methods (similar to bootstrapping).
- These are outlined in Manly and won't be covered here.

Module 9.1

Geostatistics

- The most commonly used method for spatial data analysis in the environmental area is geostatistics.
- Geostatistics involves determining a variogram and then using it to carry out kriging
- Kriging allows you to map out a gradient in space and to predict values of variables in between the data points

Variograms

- The variogram (also called a semivariogram) shows the distance between points plotted on the horizontal axis and half of the squared difference in their values on the vertical axis
- If data are positively autocorrelated, then we would expect that the differences would be small for points close together in space and become larger as points are further and further apart.
- Points very close together will probably still not be equal to each other (why?) so even at a distance close to zero there will be some difference

Variograms

- There are several types of variograms
- The variogram cloud simply plots the points.
- An empirical variogram creates a curve through the points using smoothing or averaging techniques
- A model variogram fits a mathematical function to the data and estimates its parameters
- At each step, the variogram becomes smoother and smoother

Variograms - Example

- Our example data set:

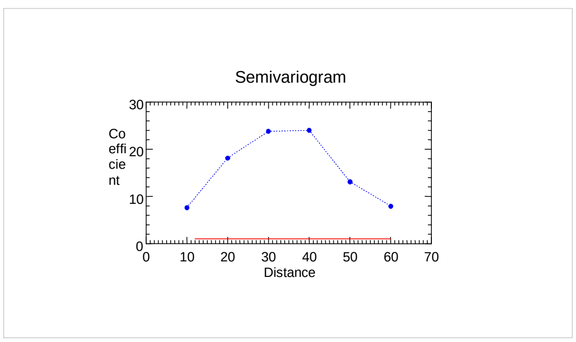
				Distance East/West				
	30	0	0	8	4	1	0	0
Distance	20	1	2	13	7	3	1	0
North/South	10	2	6	20	14	6	1	0
	0	0	4	12	18	8	2	1
	-10	1	3	9	10	15	2	0
	-20	0	1	3	6	7	9	3
	-30	0	2	1	2	4	3	6
		-30	-20	-10	0	10	20	30

Variogram Clouds - Example

- See Manly Figures 9.4 and 9.5 on pages 223 and 224.
- It is often difficult to see the underlying relationship from a variogram cloud.
- Some sort of smoothing is needed.

Empirical Variograms - Example

- The empirical variogram for our example data set is:



Model Variograms

- There are different important characteristics of model variograms (see Manly Figure 9.6, page 225):
  - The nugget is the point where the curve touches the Y axis. It's the expected difference for points very close together.
  - The sill is the maximum value that the curve attains for points far apart
  - The range of influence is the distance at which the points can be thought to be independent

Module 9.1

Model Variograms

- There are several functions commonly used as model variograms
  - Gaussian model
  - Spherical model
  - Exponential model
  - Power model

Module 9.1

Model Variograms

- Gaussian Function

$$\gamma(h) = c + (S - c)(1 - \exp(-3h^2/a^2))$$

- where h is the distance, c is the nugget, S is the sill, and a is the range of influence.

Module 9.1

Model Variograms

- Spherical Function

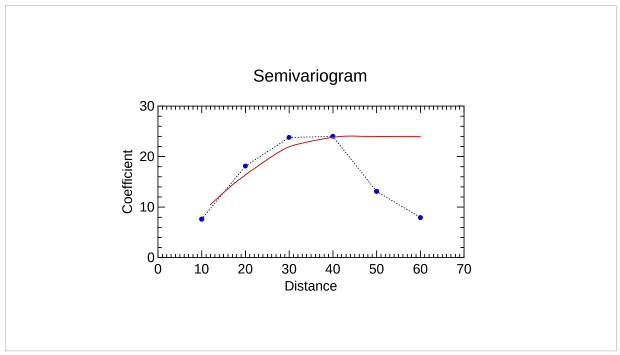
$\gamma(h) = c + (S - c)(1.5(h/a) - 0.5(h/a)^{3)}$ h<=a
c otherwise

- where h is the distance, c is the nugget, S is the sill, and a is the range of influence.

Module 9.1

Model Variograms - Example

The spherical model variogram for our example data set is:



Kriging

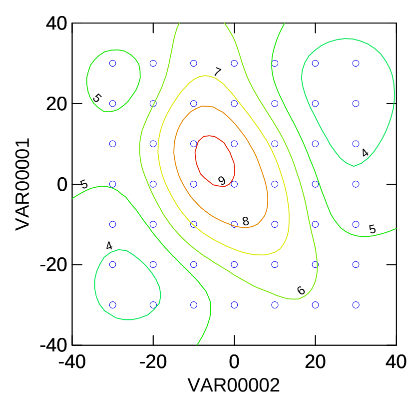
- A common problem in spatial data analysis is to determine the value of a variable at a point in space where it wasn't measured
- Kriging is a technique to do this
- It relies, in part, on the fitted model variogram
- The kriging estimate is a weighted linear combination of the known values where the weights are chosen so that the prediction errors have the minimum variance possible
- Ordinary kriging allows the true mean to vary over the study area

Kriging

$$\hat{y}_o = \sum_{i=1}^{n} a_i y_i$$

where the y_i's are the measured data points and the a_i's are the weights

Kriging - Example



Spatial Data Analysis

Spatial data analysis is a complex area worthy of additional study.

It is gaining in importance due to all of the spatial data being collected to support GIS efforts

The most common type of spatial data analysis comes from geostatistics and involves computing variograms, fitting theoretical models, and using it to support kriging.