CSISS WORKSHOP

Introduction to Spatial Pattern Analysis in a GIS Environment

Introduction to Pattern Analysis:

Context and Problems

Arthur Getis

Spatial Pattern Analysis: Approaches

- Exploratory Spatial Data Analysis (ESDA)
- Spatial Statistics
- Geostatistics
- Spatial Econometrics

ESDA

- GIS Functionality (buffers, distances, etc)
- Map Pattern Measures
- Histograms, Box Plots, Leaf and Stem Plots
- Multiple Scatter Diagrams, Surfaces
- Residuals from Regression
- Visualization, 3-D, Fly-through

ESDA: Data and Technology

- Technology and Large Data Sets
- Computer-assisted
- Simulation, Scenarios
- Permutations

Spatial Statistics

- Based on Conventional Statistical Theory
- Chance or Non-chance Occurrences
- Measures of Association, Segregation
- All Interevent Distances: K-functions
- Clustering Statistics: I, G, c, etc.
- Specially Developed Tests on Randomness (or Normal) Hypotheses

Geostatistics

Distance Based Models

- Semi-variograms of different types
  Describes relationship of observations
  to each other in terms of distance
- Kriging for Surface Extrapolation
- Kriging Models (Punctual, Universal,
  Co-, Disjunctive) for Prediction

Spatial Econometrics

- Regression Models with One or More Spatial Parameters (that describes the effect of distance -- in conjunction with certain variables -- on dependent variable)
- Development of Spatial Association Matrices (describes hypothesized spatial effects
- Parameter Estimation and Testing (model identification)
- Spatial Filtering (removing spatial effects)
- Study of Model Assumptions Including Sapatially Random Residuals)

Software Developments

New Packages Arrive Often

- GIS (ESRI is working on version 9 of ARCGIS)
- Anselin's SpaceStat being replaced with GEODA
- The Big Stat Packages are including more spatial data manipulators (SPSS, SAS, etc.)
- Geostatistics package (ESRI's Geospatial Analyst), GeoLib
- Aldstadt, Chen, and Getis' PPA
- Jacquez' ClusterSeer
- Kulldorff's Scan Statistics
- Bivand's R Package
- Fotheringham, Brunsdon, and Charlton's GWR

Examples From the Social Sciences

- **Sociology**: Behavior in Space; Ethnic Patterns; Spatial Patterns of Criminal Activities; Spatial Manifestation of Demographic Trends
- **Political Science**: Spatial Patterns of Voting; Redistricting; Diffusion of Political Movements
- **Anthropology and Archaeology**: Patterns of Human Activities (usually local in scale); Re-creation of Past Settlement Patterns
- **Economics**: Spatial Aspects of Economic Variables, Trends, Location Patterns (Sectors), Economic Concentrations, Trade Patterns
- **Geography**: Patterns of Human Spatial Interaction, Distance Decline
- **Transportation**: Movement, Accidents
- **History**: Patterns of Change in Socio-economic Patterns Over Time
- **Public Health**: Disease Diffusion; Patterns of Care; Clustering of Disease

Problems Associated with Spatial Pattern Analysis: Spatial is Special

The Problems Help to Define the Field

- Scale
- Zoning
- Dependence
- Heterogeneity
- Boundaries
- Missing Data
- Large Data Sets

Modifiable Areal Unit Problem (MAUP): Scale

- How do changes in scale change results
- What is the appropriate scale?
- Aggregation and the ecological fallacy
- Multi-scale analyses

MAUP: Zoning

- How do changes in zoning change results
- The political redistricting problem
- Appropriate zoning
- Multiple zonings

The Dependence Problem

- Tobler's Law
- The problem of nearness
- The *value of an observation* problem
- Too many observations
- Spillovers/bisection
- Traditional statistics and independence
- Overcoming the problem

The Heterogeneity Problem

- Uneven distributions at the global scale
- How does heterogeneity affect our results
- Stationarity
- Drift and its effect on analysis
- Some suggested solutions

The Boundaries Problem

- What affect do boundaries have on results?
- How do we take them into account?
- Sampling problems
- Awareness and care

The Missing Data Problem

- “Empty areas”
- Census restrictions, Privacy
- Imputation
- Algorithms and common sense solutions
- TINs, Kriging, etc.

The Large Data Set Problem

- Censuses
- Remotely sensed data
- Dependence and heterogeneity
- Data mining, partitioning and filtering, principal components analysis