Introduction to Spatial Data Analysis

Glenn Deane
University at Albany, SUNY

Deirdre Oakley
Northern Illinois University

COURSE OBJECTIVES

The goal of this five-day course is to provide an overview of and introduction to the range of statistical techniques used in the analysis of spatial (geographic) data. The emphasis is on gaining insight into the overall framework for analysis and developing an understanding of the various concepts, rather than an in-depth technical treatment of specific statistical techniques. Also, the focus in this course is on exploration and description, rather than modeling per se. This course does not cover spatial regression analysis.

Course Topics

The course topics are selected to provide an entry into the field, rather than being comprehensive. In addition to an introduction to spatial analysis terminology, four broad topics will be considered, including geovisualization (exploratory spatial data analysis), and the analysis of spatial autocorrelation for different spatial data types: points as events (point pattern analysis), points as samples (geostatistics), and polygons/areas (regional/lattice data). An important aspect of the course is to gain hands on experience with spatial data analysis software, including ArcGIS 9, CrimeStat 3.0, GeoDa 0.9.5.i, SaTScan, and PPA; as well as ArcView 3.x (and various extensions). As such, the extensive lab exercises and instruction provided by Professor Oakley are a key component of this course. The morning lectures will introduce new material and give a substantive grounding to that material, but we will devote most of the course time to the lab applications.

No previous GIS or other spatial analysis background is required, however, course participants will benefit from a familiarity with fundamental statistical concepts and intermediate-level statistical methods (e.g., correlation, multivariate analysis).

The course is organized around the analysis of spatial autocorrelation for different spatial data types. The first day is devoted to an introduction to geovisualization and working knowledge of ArcGIS 9.

• concepts: what makes spatial data analysis different, some basic GIS concepts, understanding of the paradigms in spatial data analysis, spatial scale, analogy of spatial autocorrelation that is found in the more familiar concept of temporal autocorrelation.
• geovisualization: mapping, the visualization and exploration of spatial data (exploratory spatial data analysis or ESDA), outlier analysis, rate mapping
The second day begins point pattern analysis using ArcGIS Spatial Statistics Tools and CrimeStat.

- point pattern analysis: assessing whether a pattern of locations (points) is clustered, spatial point processes, nearest neighbor statistics, second order statistics, model-based analysis of point patterns
- measuring geographic distributions in ArcGIS and creating distance buffers for analysis

Day Three continues point pattern analysis with a discussion of scan statistics and space-time analysis using the Knox statistics (PPA software) and SaTScan.

Day Four addresses spatial analysis using polygon/area data with ArcGIS Spatial Statistics Tools (e.g., local Moran) and then with GeoDa.

- spatial autocorrelation analysis: descriptive statistics for spatial autocorrelation in regional (areal) data, constructing spatial weights, visualizing spatial autocorrelation, local indicators of spatial association (LISA), bivariate spatial correlation (including space-time correlation)
- ESDA: basics of exploratory spatial data analysis, statistical graphs, linking and brushing, visualizing multivariate association

The fifth day introduces the geostatistical approach through surface interpolation using ArcGIS and ArcView.

- geostatistics: the geostatistical perspective, exploring the variogram, kriging (spatial prediction)

Organization
The course will meet for lectures in the morning and for laboratory exercises in the afternoon. Lectures will generally be from 9:30-11:30. Participants are encouraged to meet individually or in groups with instructors in the afternoon to discuss research problems and methodological issues.

Readings
The readings cited in this syllabus are intended to be “entries” into the field. They often contain extensive reviews of the literature and many other references to explore the field further. The readings are available as .pdf files. Some readings are provided mainly for follow up study. We do not expect that you will read through all the materials provided in the five day time period!

Laboratory Exercises
An extensive set of exercises is provided for you in a workbook to gain hands-on experience in the methods covered in class. The exercises are organized according to the order of material covered in the daily lectures with exercises covering ArcGIS, CrimeStat, GeoDa, SaTScan, and PPA; as well as ArcView 3.x. The tutorials can be completed at your own pace. You may choose to selectively work through the exercises, or simply go through them in sequence.

Web Resources
A considerable set of additional resources to help with learning spatial data analysis can be found on the web. The list below and the links contained in them should get you started:

• the Center for Spatially Integrated Social Science (CSISS) main site, especially its learning materials, syllabi and search engines http://www.csiss.org/
• the CSISS spatial tools clearinghouse site, with a specialized tools search engine, links to portals and selected links to specific software http://www.csiss.org/clearinghouse/index.php3
• the GeoDa home site, with links to GeoDa installation, manuals, tutorials, data sets and other supporting materials https://www.geoda.uiuc.edu/
• the R-Geo Project page with links to programs and utilities to carry out spatial data analysis using the R language http://sal.uiuc.edu/csiss/Rgeo/
• the ESRI home page, with links to resources for digital maps, data sets, utilities, courses, etc. http://www.esri.com/
• access shape files at http://www.proximityone.com
• SaTScan v7.0.3 software, manual, and sample data available at http://www.satscan.org
• Point Pattern Analysis v.1.0a A web-based interface on which you can upload ASCII files for analysis is available at http://www.nku.edu/~longa/cgi-bin/cgi-tcl-examples/generic/ppa/ppa.cgi. General documentation by DongMei Chen and Arthur Getis is available at http://zappa.nku.edu/~longa/geomed/ppa/doc/html/ppa.html

OUTLINE OF THE SHORT COURSE

DAY 1 – SPATIAL DATA AND GEOVISUALIZATION

1. Spatial Data and Spatial Data Analysis
   • overview of the course
   • software tools
   • focus on concepts and jargon
   • GIS and spatial analysis
   • why spatial data analysis is different
   • spatial data models and how they constrain/define spatial data analysis
   • classification of spatial data analyses

Selected Reading

2. Introduction to Spatial Autocorrelation
3. Geovisualization

- concepts: mapping, geovisualization, ESDA, knowledge discovery
- map basics
- visualizing extreme values, outlier maps
- statistical maps: conditional maps, map movie

Selected Readings

4. Rate Mapping

- mapping risk, standardized mortality rate, excess risk
- smoothing rates, variance instability, Bayesian principles
- Empirical Bayes (EB) smoothing
- spatial smoothing

Selected Readings

Laboratory Exercises

This first day will be devoted to becoming familiar with the lab and to learning the basic features of ArcGIS 9, including ArcCatalog, ArcMap, Arctoolbox, thematic mapping, and data manipulation.

DAY TWO – POINTS AS EVENTS, POINT PATTERN ANALYSIS

1. Neighbors and Weights Matrices

Selected Readings
2. Describing Spatial Distributions: type of distance measurement
   • spatial distribution I: spherical or projected coordinate system
   • spatial distribution II: polar coordinate system
   • spatial distribution III: spatial autocorrelation
   • spatial distribution IV: distance analysis

3. Hot Spot Analysis
   • mode
   • fuzzy mode
   • nearest neighbor hierarchical clustering
   • risk-adjusted nearest neighbor hierarchical clustering
   • K-mean clustering Anselin’s local Moran test

Selected Readings


4. Measuring Geographic Distributions in ArcGIS 9
   • central feature
   • directional distribution (standard deviational ellipse)
   • linear directional mean (mean center, standard distance)
   • creating distance buffers for analysis

Laboratory Exercises
The tutorial for this day continues instruction in ArcGIS with creating distance buffers and Geocoding; and Point Pattern Analysis with ArcGIS and CrimeStat.

DAY THREE – SPATIAL AND TEMPORAL ANALYSIS OF POINT PATTERNS

1. Scan Statistics: detecting temporal, spatial, and space-time clusters using SaTScan

2. Spatial and Temporal Analysis of Crime Routine (STAC)

3. Space-Time Knox (PPA)

Selected Readings


Laboratory Exercises
The tutorial for this day covers Point Pattern Analysis with PPA software; and temporal, spatial, and space-time analyses in the SaTScan software package.

DAY FOUR – SPATIAL ANALYSIS OF COUNT (LATTICE) DATA

1. Spatial Analysis of Attributes of Polygons
   • choices of weights in the GeoDa software
   • issues (matrix density, edge effects)

Selected Readings

2. Exploratory Spatial Data Analysis
   • Moran’s I statistic
   • relevant quantification of spatial dependence (univariate vs. residual distributions)
   • Moran scatterplot
   • LISA, Local Moran
   • visualizing LISA statistics
   • interpretation and limitations
   • generalizations: multivariate Moran

Selected Readings

3. Spatial Dependence vs. Spatial Heterogeneity
   • spatial clusters vs. spatial regimes
   • introduction to GWR (geographically weighted regression) as an exploratory technique to quantify and visualize spatial heterogeneity
Selected Readings


“What is GWR” and “GWR Primer” available at [http://ncg.nuim.ie/ncg/GWR/index.htm](http://ncg.nuim.ie/ncg/GWR/index.htm)

Laboratory Exercises

Exploratory Spatial Analysis with ArcGIS and GeoDa: analyzing spatial autocorrelation with count (lattice) data. Using GeoDa for data manipulation, spatial weights, construction and other spatial analytic functions.

**DAY 5 – POINTS AS SAMPLES, GEOSTATISTICS**

1. Geostatistics
   - spatial random field
   - spatial stationarity
   - variogram, semi-variogram
   - EDA with a variogram
   - range, sill, nugget
   - distributional forms
   - optimal spatial prediction, kriging

Selected Readings


2. From Grids to Polygons

Laboratory Exercises

Introduction to ArcGIS Geostatistical extension: exploratory spatial analysis and spatial interpolation of continuous data