

Short Course

Introduction to Spatial Data Analysis

Luc Anselin

Regional Economics Applications Laboratory (REAL)
Department of Agricultural and Consumer Economics
Department of Economics and Department of Geography
University of Illinois, Urbana-Champaign
Urbana, IL 61801

anselin@uiuc.edu
<http://www.spacestat.com/>

ICPSR-CSISS, University of California, Santa Barbara
June 24-28, 2002

© 2002, Luc Anselin, All Rights Reserved
May not be reproduced without express written permission

CONTENTS

Course Objectives

Outline of Short Course

Brief Guide to Software and Sample Data Sets

Lecture Overheads

COURSE OBJECTIVES

The goal of this 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. The latter is covered in the companion course on “Spatial Regression Analysis” (ICPSR, August 5-9, 2002).

What this course is **not**:

- this is not a GIS training course
- this is not a in-depth course on any of the techniques covered
- this is not a comprehensive survey
- this is not a SpaceStat training course
- this is not an ArcView training course

Course Topics

The course topics are selected to provide an entry into the field, rather than being comprehensive. Also, many more topics are included in the course overheads and exercises than can reasonably be covered in a five day period. This is by design, and allows for some flexibility in the coverage of materials depending on particular audience interest.

The course is organized into six broad topics:

- **concepts**: what makes spatial data analysis different, some basic GIS concepts, understanding of the paradigms in spatial data analysis
- **geovisualization**: the visualization and exploration of spatial data (exploratory spatial data analysis or ESDA), dynamically linked windows, outlier analysis, smoothing of maps for rates (proportions)

- **point pattern analysis:** assessing whether a pattern of locations (points) is clustered, spatial point processes, nearest neighbor statistics, second order statistics, bivariate and space-time point patterns
- **spatial autocorrelation analysis:** descriptive statistics for spatial autocorrelation, constructing spatial weights, visualizing spatial autocorrelation, local indicators of spatial association (LISA), multivariate spatial correlation
- **geostatistics:** the geostatistical perspective, variograms, kriging
- **spatial regression:** specifying spatial econometric models, spatial externalities, estimation methods, specification tests

Organization

The course will meet for lectures in the morning and for laboratory exercises in the afternoon. Lectures will generally be from 9:00 am till 1:00pm, with frequent breaks (the first day's lectures may run into the afternoon, with a shorter lab). There will be open ended group meetings at the end of the day to discuss research problems and methodological issues.

Laboratory Exercises

A set of exercises is provided to gain hands-on experience in the methods covered in class. The exercises consist of a step-by-step tutorial to practice a particular technique using SpaceStat and the SpaceStat extensions, as well as other specialized software, such as CrimeStat, VarioWin and the Geostatistical Analyst for ArcGis. In addition, some assignments are included to gain further familiarity and to stimulate "thinking spatially". These exercises can be completed at your own pace. On average they should take between 30 minutes and an hour each, depending on your familiarity with the methods and software.

You can choose to selectively work through the exercises, or simply go through the tutorials in sequence. The exercises designated in the course outline for use in the lab are a subset from four more extensive collections contained on the CD:

Spatial Data Analysis Laboratory Exercises (2001) [WhartonLab], *Spatial Data Analysis with SpaceStat and ArcView (3rd Edition)* (1999)[Workbook] and its *Addendum for CrimeStat and VarioWin* (2001) [Addexercises], and *Spatial Analysis with ArcGIS 8.1 Extensions, Laboratory Exercises* (2001) [Spatiallab]. These have been used extensively in past workshops and are also available from my web site <http://geog55.geog.uiuc.edu> .

Course CD

Every registered student will receive a CD containing the course materials, readings, data sets, exercises and tutorials. The disk also includes a complete set of documentation for both SpaceStat and CrimeStat, as well as the executables for the SpaceStat extension for ArcView and the DynESDA extension for ArcView. It also holds a copy of an early Beta release of the new DynESDA2 software for spatial data exploration (this software is still in a testing stage and is *guaranteed to contain bugs*). In addition, the latest versions of CrimeStat and Variowin, as well as the executable and several spatial statistical packages for the R software are included. See the readme file on the CD for any last minute additions or changes.

The materials on the CD are provided as is, without any warranties of any kind. They are copyrighted and provided for the personal use of the participants and may not be redistributed without express permission of the respective copyright holders.

Web Resources

A considerable set of additional resources to help with learning spatial data analysis can be found on the web. The list below 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 SpaceStat home site, with tutorials, downloadable data sets and other utilities:
<http://www.spacestat.com/>
- the TerraSeer home site, with tutorials on cluster analysis and boundary analysis:
<http://www.terraseer.com/>
- the ESRI home page, with links to resources for digital maps, data sets, utilities, courses, etc.:
<http://www.esri.com/>
- the long version of this course, my one semester course on spatial analysis, which contains powerpoint class notes, exercises, readings, etc.
<http://geog55.geog.uiuc.edu/sa>
- the long version of my spatial econometrics course, in a one semester format, with exercises, data sets and supporting materials:
<http://geog55.geog.uiuc.edu/ace492se>

OUTLINE OF THE SHORT COURSE

DAY 1 – INTRODUCTION AND GEOVISUALIZATION

1. *Spatial Data and Spatial Data Analysis*

- focus on concepts and jargon
- motivation for spatial analysis
- distinguishing characteristics of spatial analysis
- why spatial data analysis is different
- spatial data models and how they constrain/define spatial data analysis
- classification of spatial autocorrelation analyses

Selected Readings

Goodchild M., Anselin L., Appelbaum R., Harthorn B. (2000). Toward spatially integrated social science. *International Regional Science Review* 23, 139-159. [on CD]

Anselin L. (1999). The future of spatial analysis in the social sciences. *Geographic Information Sciences* 5, 67-76. [on CD]

2. *Geovisualization and ESDA*

- how to lie with maps
- beyond mapping, ESDA
- visualizing spatial distributions
- outlier maps
- dynamically linked windows

Selected Readings

Anselin L. (1999). Interactive techniques and exploratory spatial data analysis. In P. Longley, M. Goodchild, D. Maguire, D. Rhind (eds) *Geographical Information Systems* (2nd ed). New York: Wiley.

Laboratory Exercises

Most of the time this first day will be devoted to becoming familiar with the lab and an introduction to the available software. For those of you not familiar with ArcView or ArcGIS, the Workbook contains a series of tutorials on ArcView, while the Spatiallab introduced ArcGIS. You are encouraged to skim these. If time permits, you may also try the following exercises on ESDA.

- WhartonLab, ESDA exercise
- Spatiallab, Exercise 8 [NOTE: the version of DynESDA2 installed in the lab is *different from the one described here*; some of the interfaces may look different; details will be pointed out during the software demonstration]
- Workbook, Exercise 13, 14

DAY 2 – RATE MAPS AND POINT PATTERN ANALYSIS

3. *Visualizing Rates*

- rate mapping
- events
- risk surface, probability surface
- Relative risk, excess risk maps
- variance instability
- empirical Bayes smoothing
- spatial window smoothing
- model-based smoothing

Selected Readings

Bailey, T and Gatrell A (1995). *Interactive spatial data analysis*. New York: Wiley.
(pp. 299-308)

Anselin L. (2002). *Rate Transformations*. SpaceStat Support Document. Ann Arbor: TerraSeer Inc. [on CD]

Laboratory Exercises

- Workbook Exercise 12
- new exercises using DynESDA2 (to be handed out in class)

4. Point Pattern Analysis

- Pattern
- First order statistics
- Nearest neighbor statistics
- Second order statistics

Selected Readings

Bailey and Gatrell, Chapters 3-4.

Levine N. (2000). *CrimeStat 1.1, A spatial statistics program for the analysis of crime incident locations*. Washington: National Institute of Justice, Chapters 4 and 5

Gatrell A., T. Bailey, P. Diggle, B. Rowlingson (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers* 21, 256-274.

Okabe, A. and I. Yamada (2001). The K function method on a network and its computational implementation. *Geographical Analysis* 33, 271-290.

Laboratory Exercises

Descriptive statistics and nearest neighbor analysis using CrimeStat

- Addexercises : Centrophraphy
- Addexercises: Nearest neighbor statistics

DAY 3 – SPATIAL AUTOCORRELATION

3. *Spatial Autocorrelation*

- spatial autocorrelation terminology
- null and alternative hypothesis
- spatial weights
- join count statistics
- Moran's I statistic, Moran scatterplot
- LISA, Local Moran
- visualizing LISA statistics
- interpretation and limitations
- generalizations: multivariate, space-time

Selected Readings

Cliff A. and Ord J.K. (1981). *Spatial Processes, Models and Applications*.
London: Pion, pp. 17-19, Ch. 2.

Anselin L. (1995). Local indicators of spatial association - LISA. *Geographical Analysis* 27, 93-115.

Messner, S., L. Anselin, R. Baller, D. Hawkins, G. Deane, S. Tolnay (1999). The Spatial Patterning of County Homicide Rates: An Application of Exploratory Spatial Data Analysis, *Journal of Quantitative Criminology* 15, 423–450.

Anselin, L., Syabri I. and Smirnov O. (2002). *Visualizing Multivariate Spatial Correlation with Dynamically Linked Windows*. Proceedings, New Tools in Spatial Data Analysis. [on CD]

Laboratory Exercises

- Whartonlab, Spatial Autocorrelation exercise
- Spatiallab, Lab 9 [NOTE differences with current version of DynESDA2]
- Workbook, Exercise 15

- Workbook, Exercise 20
- Workbook, Exercise 21 (Local Moran only)
- Optional: Workbook, Exercises 18 and 19

DAY 4 – GEOSTATISTICS

5. Geostatistics

- spatial random field
- spatial stationarity
- variogram, semi-variogram
- EDA with a variogram
- correlogram
- range, sill, nugget
- spherical, exponential variogram
- optimal spatial prediction, kriging

Selected Readings

Cressie N. (1993) *Statistics for spatial data*. New York: Wiley, Chapter 2.

Pannatier, Y (1996). *Variowin, software for spatial data analysis in 2D*. Berlin: Springer-Verlag, Chapters 4, 5.

Bailey and Gatrell (1995). Chapter 5.

Goovaerts P. (1997). *Geostatistics for natural resources evaluation*. New York: Oxford, Chapter 5.

Laboratory Exercises

Exploring and modeling variograms

- Addexercises: Variowin Basics
- Addexercises: Exploring Variograms
- Addexercises: Modeling Variograms
- Spatiallab: Lab 10

DAY 5 – SPATIAL REGRESSION ANALYSIS

6. *Spatial Regression*

- specifying regression models with spatial autocorrelation
- spatial multipliers and spatial externalities
- simultaneous and conditional models
- maximum likelihood and instrumental variables estimation
- Moran's I test for regression residuals
- Lagrange Multiplier tests for spatial autocorrelation
- spatial specification searches

Selected Readings

Anselin L. (2001). Spatial econometrics. In Baltagi B. (ed) *A companion to theoretical econometrics.*, pp. 310-330. Oxford: Basil Blackwell.[original long draft on CD]

Anselin L. and Bera A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In Ullah A. and Giles D. (eds) *Handbook of applied economic statistics*, pp. 237-289. New York: Marcel Dekker.

Anselin, L. (2003). Spatial Externalities, Spatial Multipliers and Spatial Econometrics. *International Regional Science Review* [on CD]

Laboratory Exercises

- Whartonlab, Spatial Regression exercise
- Workbook, Exercise 22, 26, 28, 30
- Optional: Workbook, Exercises 29, 31