

# Spatial Data Analysis

---

## OVERVIEW

- This chapter is the first in a set of three dealing with geographic analysis and modeling methods.
- The chapter begins with a review of the relevant terms, and an outlines the major topics covered in the three chapters
- Examines methods constructed around the concepts of location, distance, and area

## LEARNING OBJECTIVES

- **Definitions of spatial data analysis and tests to determine whether a method is spatial.**
- **Techniques for detecting relationships between the various properties of places and for preparing data for such tests.**
- **Methods to examine distance effects, in the creation of clusters, hotspots, and anomalies.**
- **The applications of convolution in GIS, including density estimation and the characterization of neighborhoods.**

## KEY WORDS AND CONCEPTS

Spatial analysis, inductive, deductive, normative, queries, measurements, transformations, , algorithm, metric, buffer, point in polygon, polygon overlay, spurious polygons, coastline weave,

tolerance, Thiessen polygons, Inverse-distance weighting (IDW), Kriging, semivariograms, density estimation

## OUTLINE

14.1 Introduction: What Is Spatial Analysis?

14.2 Analysis Based on Location

14.3 Analysis Based on Distance

14.4 Conclusion

## CHAPTER SUMMARY

### 14.1 Introduction: what is spatial analysis?

- The techniques covered in these three chapters are generally termed spatial rather than geographic, because they can be applied to data arrayed in any space, not only geographic space.
- Spatial analysis the crux of GIS because it includes all of the transformations, manipulations, and methods that can be applied to geographic data to add value to them, to support decisions, and to reveal patterns and anomalies that are not immediately obvious
  - Spatial analysis is the process by which we turn raw data into useful information,
- The term analytical cartography is sometimes used to refer to methods of analysis that can be applied to maps to make them more useful and informative
- In this and the next chapter the authors look first at some definitions and basic concepts of spatial analysis.
- Chapter 16 is devoted to spatial modeling, a loosely defined term that covers a variety of more advanced and more complex techniques, and includes the use of GIS to analyze and simulate dynamic processes, in addition to analyzing static patterns.
- The human eye and brain are also very sophisticated processors of geographic data and excellent detectors of patterns and anomalies in maps and images.
  - So the approach taken here is to regard spatial analysis as spread out along a continuum of sophistication, ranging from the simplest types that occur very

quickly and intuitively when the eye and brain look at a map, to the types that require complex software and sophisticated mathematical understanding.

- Spatial analysis is a set of methods whose results change when the locations of the objects being analyzed, or the frame used to analyze them, changes.

#### 14.1.1 Examples

- John Snow map of cholera
- Openshaw's technique which generates a large number of circles, of random sizes, and throws them randomly over the map. The computer generates and places the circles, and then analyzes their contents, by dividing the number of cases found in the circle by the size of the population at risk. If the ratio is anomalously high, the circle is drawn (Figure 14.4).
- Spatial analysis can be
  - *inductive*, to examine empirical evidence in the search for patterns that might support new theories or general principles, in this case with regard to disease causation.
  - *deductive*, focusing on the testing of known theories or principles against data
  - *normative*, using spatial analysis to develop or prescribe new or better designs

#### 14.2.1 Analysis of Attribute Tables

- Likely that the kinds of factors responsible for the occurrence of an observed phenomenon are contained within the attribute tables of a GIS
  - One way to examine this suspicion is to plot one variable against the other as a scatterplot.
  - Allow us to examine in detail the dependence of one variable on one or more independent variables.
- Regression analysis focuses on finding the simplest relationship indicated by the data.
  - Multiple regression extends this principle to consider the effects of multiple independent variables
- Relationships between variables can vary across space, which is an issue termed spatial heterogeneity
  - Geographers have developed a set of techniques that recognize such heterogeneity explicitly:
  - The example of Geographically Weighted Regression is given

### 14.2.2 Spatial Joins

- One of the most powerful features of a GIS is the ability to join tables based on common geographic location.

### 14.2.3 The Point-in-Polygon Operation

- The point in polygon operation is used to determine whether a point lies inside or outside a polygon.
- Occurs when point-like events must be compared to properties of the surrounding environment

### 14.2.4 Polygon Overlay

- Polygon overlay is similar to the point-in-polygon operation in the sense that two sets of objects are involved
- Exists in two form depending whether discrete or continuous perspective is taken:
  - The complexity of computing a polygon overlay was one of the greatest barriers to the development of vector GIS
  - From the discrete-object perspective, the task is to determine whether two area objects overlap, to determine the area of overlap, and to define the area formed by the overlap as one or more new area objects
  - The continuous-field version of polygon overlay does this by first computing a new dataset in which the region is partitioned into smaller areas that have uniform characteristics on both variables.
- One of the issues that must be tackled by a practically useful algorithm is known as the spurious polygon or coastline weave problem.
  - Although the same boundary line may be represented in both source datasets, its representations will almost certainly not be the same.
  - The most common method for dealing with this is the specification of a tolerance.
    - If two lines fall within this distance of each other, the GIS will treat them as a single line and not create slivers.

### 14.2.5 Raster Analysis

- Overlay in raster is much simpler – the attributes of each cell are combined according to a set of rules

## 14.3 Analysis Based on Distance

- The ability to calculate and manipulate distances underlies many forms of spatial analysis
  - based on the concept that the separation of features or events on the Earth's surface can tell us something useful

### 14.3.1 Measuring Distance and Length

- A metric is a rule for the determination of distance between points in a space.
- Pythagorean or straight-line metric is explained with an equation and diagram (Fig 14.12)
  - Notes this metric does not work for latitude and longitude, must use the spherical metric provided in Section 5.7 to calculate great circles
- Distance along a route represented by a polyline is often calculated by summing the lengths of each segment of the polyline
- Because there is a general tendency for polylines to short-cut corners, the length of a polyline tends to be shorter than the length of the object it represents.
- Length of a 3 dimensional line measured off its planimetric representation will also be shorter than its true length

### 14.3.2 Buffering

- Builds a new object or objects by identifying all areas that are within a certain specified distance of the original objects
- In raster, buffers can be spread outwards from objects to create friction surfaces

### 14.3.3 Cluster Detection

- Points patterns can be identified as clustered, dispersed, or random
- Kinds of processes responsible for point patterns are:
  - First-order processes involve points being located independently
  - Second-order processes involve interaction between points
- Briefly introduces the K function as an example of a descriptive statistic of pattern, and explains a simple example