4 The Nature of Geographic Data

OVERVIEW

- Elaborates on the *spatial is special* theme
- Focuses on how phenomena vary across space and the general nature of geographic variation
- Describes the main principles that govern scientific sampling, how spatial variation is formalized and measured as spatial autocorrelation, and outlines the concept of fractals.

LEARNING OBJECTIVES

- **How Tobler's First Law of Geography is formalized through the concept of spatial autocorrelation;**
- **The relationship between scale and the level of geographic detail in a representation;**
- **The principles of building representations around geographic samples;**
- **How the properties of smoothness and continuous variation can be used to characterize geographic variation;**
- **How fractals can be used to measure and simulate surface roughness.**

KEYWORDS AND CONCEPTS

Time and space; spatial autocorrelation and the Tobler Law; scale; representation; types of spatial objects; fractals and self-similarity; spatial sampling; distance decay; induction and deduction; isopleth maps; choropleth maps; adjacency; regression; generalization.

OUTLINE

CHAPTER SUMMARY

4.1 Introduction

Reviews the governing principles of the development of representations already covered, and adds three more related to the *nature of spatial variation*:

- that proximity effects are key to understanding spatial variation, and to joining up incomplete representations of unique places;
- that issues of geographic scale and level of detail are key to building appropriate representations of the world;
- that different measures of the world co-vary, and understanding the nature of co-variation can help us to predict

4.2 The fundamental problem revisited

- Distinguishes between *controlled* variation, which oscillates around a steady state, and *uncontrolled* variation.
- Some applications address controlled variation, such as utility management
- Others address uncontrolled, such as those studying longer term processes
- Introduces the concept of *time series* and acknowledges the concept of *temporal autocorrelation*.
- "Our behavior in space often reflects past patterns of behavior", thus it is one-dimensional, need only look in the past
- *Spatial heterogeneity* is the tendency of geographic places and regions to be different from each other.
- Occurs in both form and process

- The principles addressed in this chapter help answer questions about what to leave in and what to take out of digital representations
- Scale and spatial structure help determine how to sample reality and weight sample observations to build representations

Technical Box 4.1 Types of spatial objects

- Classifies geographic objects by their topological dimension
- Points, lines, area objects, volume objects, time (as the 4^{th} dimension)
- Classification of spatial phenomena into object types is dependent fundamentally upon scale

4.3 Spatial autocorrelation and scale

- Spatial autocorrelation measures attempt to deal simultaneously with similarities in the location of spatial objects and their attributes
- Brief discussion about how *neighboring* might be defined, more later in this chapter.
- Measures of spatial and temporal autocorrelation are *scale dependent*
- The issue of *sampling interval* is of direct importance in the measurement of spatial autocorrelation
- When the pattern of spatial autocorrelation at the coarser scale is replicated at the finer scale, the overall pattern exhibits the property of *self-similarity*.

Technical Box 4.2 The many meanings of scale

- Scale is in the details
- Scale is about extent
- Scale of a map – including reference to *representative fraction* and confusion between large and small scales. This book, thus, uses “coarse” and “fine”.

4.4 Spatial sampling

- *Sample frame* is defined as the universe of eligible elements of interest.
- Might be bounded by the extent of the field of interest or by the combined extent of a set of areal objects
- *Sampling* is the process of selecting points from a continuous field or selecting some objects while discarding others
- Any geographic representation is a kind of sample
- Procedures of *statistical inference* allow us to infer from samples to the population from which they were drawn
- Classical statistics often emphasizes the importance of randomness in sound sample design

- Types of sampling designs include: simple random sampling, spatially systematic sampling (problems if the sampling interval and spatial structure coincide so that the sample frame exhibits *periodicity*), stratified random sampling, periodic random changes in the sampling grid, clustered sampling, sampling along transects
- In circumstances where spatial structure is either weak or is explicitly incorporated through clear definition of subpopulations, standard statistical theory provides a robust framework for inferring the attributes of the population from those of the sample
- However, the existence of spatial autocorrelation fundamentally undermines the inferential framework and invalidates the process of generalizing from samples to populations.

4.5 Distance decay

- Discusses the attenuating effect of distance and the need to make to make an informed judgment about an appropriate *interpolation* function and how to *weight* adjacent observations.
- Explains and illustrates the structure of the distance decay equation: *b* as a parameter that affects the rate at which the weight w_{ij} declines with distance
- Discusses linear, negative power, and negative exponential distance decay equations and graphs
- Notes that with these equations, the effects of effects of distance are presumed to be regular, continuous, and *isotropic* (uniform in every direction)
- Several paragraphs discuss how these simple equations may not reflect reality

Technical Box 4.3 Isopleth and choropleth maps

- Explains how isopleth and choropleth maps are constructed and displayed
- Choropleth maps describe properties of non-overlapping areas.
- In this section is the important discussion about spatially extensive and intensive variables:
- Spatially extensive variables are true only of entire areas, such as total population, or total number of children under 5 years of age.
- Spatially intensive variables could potentially be true of every part of an area, if the area were homogeneous – examples include densities, rates, or proportions.
- The figure caption explains that spatially extensive variables should be converted to spatially intensive form.

4.6 Measuring distance effects as spatial autocorrelation

- *Induction* reasons from data to build up understanding, while *deduction* begins with theory and principle as a basis for looking at data.
- Knowledge of the actual or likely nature of spatial autocorrelation can be used *deductively* in order to help build a spatial representation of the world.
- The *measurement* of spatial autocorrelation is a more i*nductive* approach to developing an understanding of the nature of a geographic dataset.
- If the phenomenon is conceived as a field, then spatial autocorrelation measures the smoothness of the field using data from the sample points, lines, or areas that represent the field.
- If the phenomena of interest are conceived as discrete objects, then spatial autocorrelation measures how the attribute values are distributed among the objects, distinguishing between arrangements that are clustered, random, and locally contrasting.
- Figure 4.11 shows examples of each of the four object types, with associated attributes, chosen to represent situations in which a scientist might wish to measure spatial autocorrelation.

Technical Box 4.4 Measuring similarity between neighbors

- Outlines the concept of the weights matrix, w_{ij}
- Measures of spatial autocorrelation compare a set of locational similarities w_{ij} with a corresponding set of attribute similarities c_{ij} , combining them into a single index in the form of a cross-product.
- Explains ways to measure similarity of attributes
- Briefly outlines the Moran Index

4.8 Taming geographic monsters

- Expands the discussion in Technical Box 4.6 into *fractal geometry*
- Fractals can be thought of as geometric objects that are between Euclidean dimensions
- Ascertaining the fractal dimension of an object involves identifying the scaling relation between its length or extent and the yardstick (or level of detail) that is used to measure it.
- Illustrates the log-log relationship between length and step length