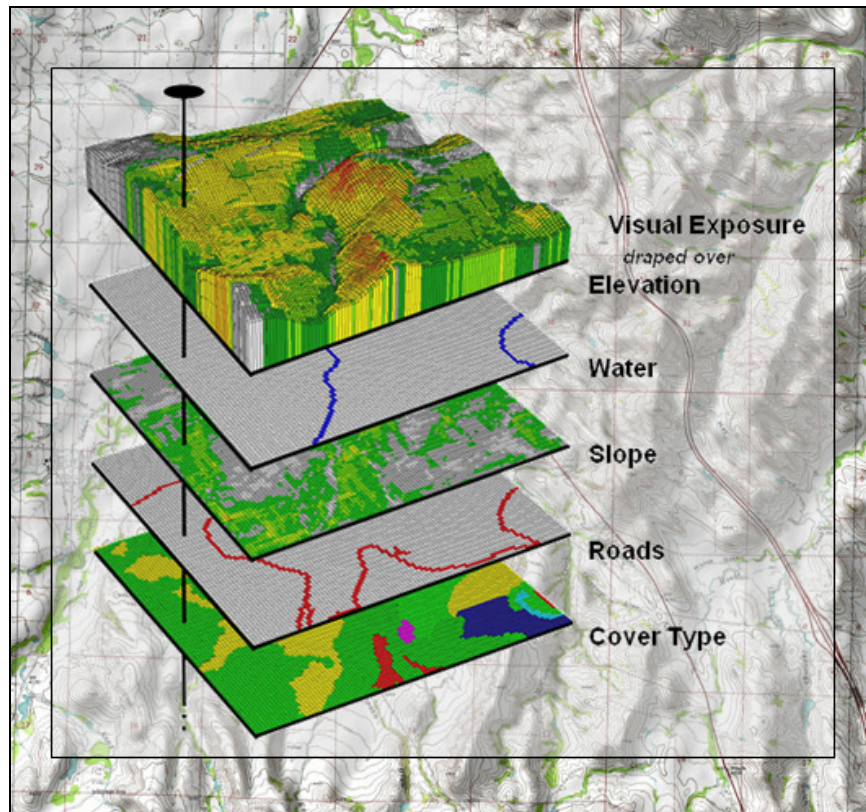


---

# Analyzing Geospatial Resource Data

## A Hands-on Case Study in Spatial Analysis and Data Mining



**Joseph K. Berry**

*W.M. Keck Scholar in Geosciences  
University of Denver  
Denver, Colorado*



---

# Table of Contents

## **1. Overview**

- 1.1 Introduction
- 1.2 Mapping versus Analysis
- 1.3 Suitability Modeling
- 1.4 Some Assembly Required
- 1.5 Exercises
  - Examples of Analysis Capabilities
  - Viewing Full Color Figures
  - Installing Companion Software

## **2. Maps as Data**

- 2.1 Visualizing Map Surfaces
- 2.2 Numbers First, Pictures Later
- 2.3 Framing Maps
- 2.4 Organizing Raster Data
- 2.5 Visualizing Map Values
- 2.6 Exercises
  - Interacting with Grid Maps
  - Grid-based Thematic Mapping
  - Map Summary Statistics and Charts
  - Map Normalization

## **3. Reclassifying and Overlaying Maps**

- 3.1 A Map-ematical Framework
- 3.2 Reclassifying Map Values
- 3.3 Overlaying Maps
- 3.4 Characterizing Spatial Coincidence
- 3.5 A Simple Habitat Model
- 3.6 Exercises
  - Spatial Analysis Operations
  - Reclassifying Map Features
  - Calculating Change
  - Summarizing Map Regions
  - Characterizing Map Coincidence
  - Evaluating a Simple Habitat Model

## **4. Mapping Visual Exposure**

- 4.1 Identifying and Using Visual Exposure
- 4.2 There's More That Meets the Eye
- 4.3 Assessing Visual Impact
- 4.4 Visual Vulnerability
- 4.5 Visual Aesthetics
- 4.6 Exercises
  - Calculating Viewsheds
  - Calculating Visual Exposure
  - Accounting for Screens
  - Calculating Weighted Visual Exposure
  - Modeling Visual Exposure Impacts
  - Extending Visual Analysis to Other Areas

---

## **5. Creating Variable-Width Buffers**

- 5.1 Variable-Width Buffer
- 5.2 Line-of-Sight Buffer
- 5.3 Effective Distance Buffer
- 5.4 Effective Distance Algorithm
- 5.4 Exercises
  - Calculating Simple Proximity
  - Calculating Uphill Proximity
  - Calculating Effective Proximity

## **6. Applying Effective Distance and Connectivity**

- 6.1 Calculating Hiking-Time
- 6.2 Incorporating Effective Distance
- 6.3 Basics of Surface Flow
- 6.4 Characterizing Overland Flow
- 6.5 Constructing Realistic Downhill Flows
- 6.6 Calculating Flow Time and Quantity
- 6.7 Determining Spill Impacts
- 6.8 Exercises
  - Simple Proximity
  - Effective Proximity
  - Identifying Optimal Paths
  - Extending the Analysis
  - Surface Flow Analysis

## **7. Identifying Optimal Routes**

- 7.1 Identifying the Most Preferred Route
- 7.2 Considering Multi-Criteria in Routing
- 7.3 Calibrating and Weighting Criteria
- 7.4 Evaluating Alternative Routes
- 7.5 Exercises
  - Routing Model
  - Extended Routing Model

## **8. Characterizing Terrain Features**

- 8.1 Identifying Micro-Terrain Features
- 8.2 Characterizing Local Terrain Conditions
- 8.3 Assessing Terrain Slope and Roughness
- 8.4 Calculating Realistic Areas
- 8.5 Identifying Valley Bottoms
- 8.6 More on Slope's Slippery Slope
- 8.7 Exercises
  - Deriving Slope
  - Establishing Roughness
  - Identifying Convex/Concave Features

## **9. Analyzing Landscape Patterns**

- 9.1 Summarizing Spatial Context
- 9.2 Calculating Nearby Neighbor Statistics
- 9.3 Analyzing Landscape Structure
- 9.4 Getting to the Core
- 9.5 Assessing Forest Fragmentation
- 9.6 Exercises
  - Calculating Size
  - Establishing Diversity
  - Determining Proportion Similar

---

Deriving Configuration Indices

**10. GIS Modeling Procedures**

- 10.1 Constitution of a GIS Model
- 10.2 A Classification Guide for GIS Models
- 10.3 Expressing Model Structure and Logic
- 10.4 Recipes for Solving Spatial Problems
- 10.5 Extending Model Solutions
- 10.6 Infusing Science into GIS Models
- 10.7 Exercises
  - Campground Suitability Model
  - Extended Campground Model
  - Editing and Using Scripts

**11. Surface Modeling**

- 11.1 From Point Samples to Map Surfaces
- 11.2 The Keystone Concept
- 11.3 Benchmarking Interpolation Results
- 11.4 Assessing Interpolation Results
- 11.5 Exercises
  - Interpolating Point data (IDW)
  - Interpolating Point Data (Kriging)
  - Comparing Interpolation Results
  - Evaluating Interpolation Performance

**12. Spatial Data Mining**

- 12.1 Comparing Yield Maps (Discrete)
- 12.2 Comparing Yield Surfaces (Continuous)
- 12.3 Calculating Field Similarity
- 12.4 Identifying Data Zones
- 12.5 Mapping Data Clusters
- 12.6 Exercises
  - Comparing Yield Maps
  - Comparing Yield Surfaces
  - Characterizing Map Similarity
  - Identifying Data Zones
  - Mapping Data Clusters

**13. Predictive Modeling**

- 13.1 Predicting Yield Maps
- 13.2 Assessing Prediction Model Results
- 13.3 Stratifying Maps for Better Predictions
- 13.4 Exercises
  - Predictive Modeling
  - Calculating Error
  - Deriving a Stratified Model

**Epilog**

**Appendix A. Quick Set of Exercises**

**Appendix B. Creating Your Own Grid-based Database**

---

# Preface

## The GIS Evolution/Revolution and What GIS Modeling Is (and Isn't)

### Where Is GIS Technology

Though GIS is less than three decades old, the approach of its software has evolved as much as its capabilities and practical expressions. In the 70's software development primarily occurred on campuses and its products relegated to library shelves of theses. These formative years provided the basic organization (both data and processing structures) we find in the modern GIS. Raging debate centered on "vector vs. raster" formats and efficient algorithms for processing—techy stuff with minimal resonance outside of the small (but growing) group of innovators.

For a myriad of reasons, this early effort focused on GIS technology itself rather than its applications. First, and foremost, is the necessity of building a viable tool before it can be taken on the road to practical solutions. As with most revolutionary technologies, the "chicken and the egg" parable doesn't apply—the tool must come before the application.

This point was struck home at a recent visit to Disneyland. The newest ride subjects you to a seemingly endless harangue about the future of travel while you wait in line for over an hour. The curious part is that the departed Walt Disney himself is outlining the future through video clips from the 50's. The dream of futuristic travel (application) hasn't changed much and believe me the current practical reality (tool), as embodied in the herky-jerky ride, is a long way from fulfilling the vision.

What impedes the realization of a technological dream is rarely a lack of vision, but the nuts and bolts needed in its construction. In the case of GIS, the hardware and software environments of the 70's constrained its use outside of academia. Working with 256K memory and less than a megabyte of disk made a GIS engine perform at the level of a skateboard. However, the environments were sufficient to develop "working prototypes" and test their theoretical foundations. The innovators of this era were able to explore the conceptual terrain of representing "maps as numbers," but their software products were woefully impractical.

With the 80's came the renaissance of modern computers, and with it the hardware and software environments needed by GIS. The research-oriented software gave way to operational systems. Admittedly, the price tags were high and high-end, specialized equipment often required, but the suite of basic features of a modern GIS became available. Software development switched from specialized programs to extensive "toolboxes," and subsequently spawned a new breed of software specialists.

Working within a GIS macro language, such as ARC/INFO's Arc Macro Language (AML), customized applications could be addressed. Emphasis moved from programming the "tool" within general languages (e.g., FORTRAN and Pascal) to programming the "application" within a comprehensive GIS language. Expertise broadened from geography and computers to an understanding of the context, factors and relationships of spatial problems. Programming skills were extended to spatial reasoning skills—the ability to postulate problems, perceive patterns and interpret spatial relationships.

From an application developer's perspective the floodgates had opened. From an end user's perspective, however, a key element still was missing—the gigabytes of data demanded in practical applications. Once again GIS applications were frustrated. This time it wasn't the programming environment as much as it was the lagging investment in the conversion of paper maps to their digital form.

But another less obvious impediment hindered progress. As the comic strip character Pogo might say, "we have found the enemy and it's us." By their very nature, the large GIS shops established to collect, nurture and

---

process spatial data intimidated their potential customers. The required professional sacrifice at the GIS altar “down the hall and to the right” kept the herds of dormant users away. GIS was more often seen within an organization as an adversary for corporate support (a.k.a., money pit) than as a new and powerful capability one could use to improve workflow and address complex issues in entirely new ways.

The 90’s saw both the data logjam and GIS mystique erode. As Windows-based mapping packages appeared on individuals’ desks, awareness of the importance of spatial data and its potential applications flourished. Direct electronic access enabled users to visualize their data without a GIS expert as co-pilot. For many the thrill of “visualizing mapped data” rivaled that of their first weekend with the car after the learner’s permit.

So where are we now? Has the role of GIS developers been extinguished, or merely evolved once again? Like a Power Rangers transformer, software development has taken two forms that blend the 70’s and 80’s roles. These states are the direct result of changes in software programming approaches in general and “object-oriented” programming in particular.

MapInfo’s MapX and ESRI’s MapObjects are tangible GIS examples of this new era. These packages are functional libraries that contain individual map processing operations. In many ways they are similar to their GIS toolbox predecessors, except they conform to general programming standards of interoperability, thereby enabling them to be linked easily to the wealth of non-GIS programs.

Like using a Lego set, application developers can apply the “building blocks” to construct specific solutions, such as a real estate application that integrates a multiple listing geo-query with a pinch of spatial analysis, a dab of spreadsheet simulation, a splash of chart plotting and a sprinkle of report generation. In this instance, GIS functionality simply becomes one of the ingredients of a solution not the entire recipe.

In its early stages, GIS required “bootstrap” programming of each operation and was the domain of the computer specialist. The arrival of the GIS toolbox and macro languages allowed an application specialist to develop software that tracked the spatial context of a problem. Today we have computer specialists generating functional libraries and application specialists assembling the bits of software from a variety of sources to tailor comprehensive solutions.

The distinction between computer and application specialist isn’t so much their roles, as it is characteristics of the combined product. From a user’s perspective the entire character of a GIS dramatically changes. The look-and-feel evolves from a generic map-centric view to one of a few tailored buttons that walk users through analysis steps that are germane to an application. Instead of presenting users with a generalized set of map processing operations as a maze of buttons, toggles and pull-down menus, only the relevant ones are integrated into the software solution. Seamless links to non-spatial programming “objects,” such as pre-processing and post-processing functions, are automatically made.

As the future of GIS unfolds, it will be viewed less as a distinct activity and more as a key element in a thought process. No longer will users “break shrink-wrap” on stand-alone GIS systems. They simply will use GIS capabilities within an application and likely unaware of the underlying functional libraries. GIS technology will finally come into its own by becoming simply part of the fabric of software solutions.

### **Moving from Mapping to Analysis of Mapped Data**

The evolution (or is it revolution?) of GIS technology has certainly taken it well beyond traditional roles of mapping. For over 8,000 years maps were graphic representations of physical features primarily for the purpose of navigation. With the advent of GIS they have changed form to digital representations that are linked to databases and new processing capabilities.

The accompanying figure identifies two key trends in the movement from mapping to map analysis. *Traditional GIS* treats geographic space in a manner similar to our paper map legacy. Points, lines and polygons are used to define discrete spatial objects, such as houses, streams and lakes. In turn, these objects are linked to attributes in a database that describe their characteristics and conditions. The result is a tremendously useful system enabling users to make complex geo-queries of the information and then map the results.

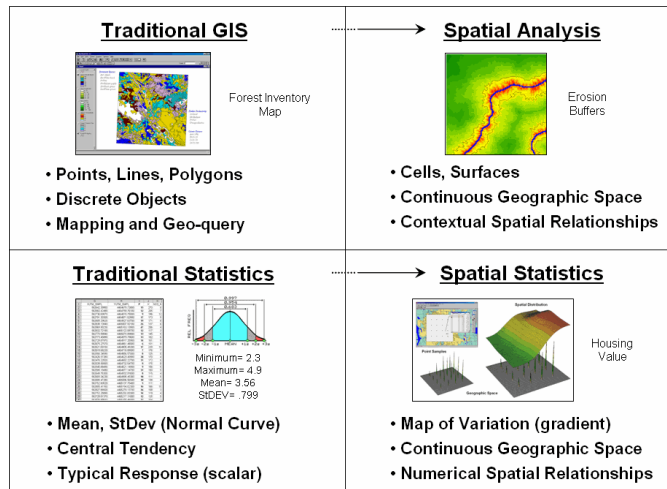


Figure P-1. Spatial Analysis and Spatial Statistics are extensions of traditional ways of analyzing mapped data.

**Spatial Analysis** extends the basic set of discrete map features of points, lines and polygons to surfaces that represent continuous geographic space as a set of contiguous grid cells. The consistency of this grid-based structuring provides a wealth of new analytical tools for characterizing “contextual spatial relationships”, such as effective distance, optimal paths, visual connectivity and micro-terrain analysis.

In addition, it provides a mathematical/statistical framework by numerically representing geographic space. **Traditional Statistics** is inherently non-spatial as it seeks to represent a data set by its typical response regardless of spatial patterns. The mean, standard deviation and other statistics are computed to describe the central tendency of the data in abstract numerical space without regard to the relative positioning of the data in real-world geographic space.

**Spatial Statistics**, on the other hand, extends traditional statistics on two fronts. First, it seeks to map the variation in a data set to show where unusual responses occur, instead of focusing on a single typical response. Secondly, it can uncover “numerical spatial relationships” within and among mapped data layers, such as generating a prediction map identifying where likely customers are within a city based on existing sales and demographic information.

This book investigates the important aspects and procedures used in Spatial Analysis and Spatial Statistics through discussion of the fundamental concepts, considerations and issues involved in the analysis of geospatial data. In addition, “hands-on” experience with the theory is provided through numerous practical exercises at the end of each topic using the MapCalc Learner software accompanying the book.

## Book Organization

*Analyzing Geospatial Resource Data* is organized into thirteen topics and two appendices that lead the reader from an understanding of the fundamental nature of mapped data through a series of basic procedures used in deriving, analyzing and applying spatial information. A case study approach is used with each topic area describing the application of a set of related analysis techniques. The discussion is followed by a series of hands-on exercises providing practical experience in applying the techniques. The exercises include step-by-step instructions that are thoroughly annotated.

## Companion Software

*Surfer* software by Golden Software is used for spatial interpolation and advanced graphics. A demonstration version of the software is included with this book. Information about the full *Surfer* package is available at [www.ssg-surfer.com/](http://www.ssg-surfer.com/).



---

*MapCalc* software by Red Hen Systems is used for grid-based map analysis exercises throughout the book. A 14-day evaluation version of *MapCalc* is included with this book. The software includes the basic set of data and operations needed to complete the hands-on exercises; **the evaluation version expires after a two-week period.**

The full *MapCalc Learner* software for individual use contains additional materials and provides for extended experience in map analysis procedures and applications. It contains a basic set of functions for import/export of your own data that is constrained to a field-level (100row x 100col) analysis grid configuration. *MapCalc Academic* is a full, multi-seat licensed educational version for classroom/lab use and contains an additional instructor CD with supporting teaching materials including PowerPoint slides, exercises and quizzes for a several workshop and college-level course offerings. *MapCalc Professional* is licensed for commercial use and includes a full range of features including image data routines, coordinate/datum transformation and an extended set of import/export formats for data exchange.

*SnagIt* software by TechSmith is used for screen capture and creation of composite presentation graphics. A 45-day evaluation version of the software is included with this book. Information about the full *SnagIt* package is available at [www.techsmith.com/](http://www.techsmith.com/).

For more information on the Learner, Academic and Professional versions of *MapCalc*, visit [www.redhensystems.com/mapcalc/](http://www.redhensystems.com/mapcalc/). Pricing for the *MapCalc* versions is US\$21.95 for Learner, \$495 for Academic and \$695.00 for Professional plus shipping and handling.  
(prices subject to change)

## Case Study Data Sets

Four data sets are used in the exercises to provide experience with different application areas—Tutorial, Natural Resources, Precision Agriculture and Geo-business. The *Tutor25.rgs* is a hypothetical dataset used to quickly demonstrate map analysis operations. The *Bighorn.rgs* natural resource data set is located in southwestern Wyoming; the *Agdata.rgs* precision agriculture data set characterizes a central-pivot field located in central Colorado; and the *Smallville.rgs* geo-business data set contains geographic features for a small northern Colorado city. All of the data sets contain both actual and hypothetical information to illustrate map analysis procedures. While the analytical procedures described are valid, the outcomes do not represent legitimate outcomes and should not be used in decision-making for the project areas.

## Intended Audience

This book is ideal for professionals and students interested in a basic understanding of the concepts, procedures and considerations in analyzing geospatial data. The material is presented in an informal manner designed so the reader can grasp the broad issues and then delve into hands-on exercises for practical experience in applying the techniques. In a classroom setting, portions of the book can be used for an exposure to grid-based map analysis in an introductory GIS course, or in its entirety for an intermediate course on map analysis with individual/group projects.

## Internet Extended Environment

There are several benefits for readers with a self-published book. Traditionally published texts have inordinately slow publishing cycles compared to the speed of software development and application innovation. This book is printed in small batches and frequently revised to keep the material current. The Internet is a great mechanism for feedback so the book can evolve *with your help*. Please help in identifying sections that need correction or clarification, as well as providing suggestions for extending the current topics or adding entirely new ones—email the author at [jberry@innovativegis.com](mailto:jberry@innovativegis.com).

---

In addition, the author's website at [www.innovativegis.com/basis/](http://www.innovativegis.com/basis/) serves as a mechanism for updates, enhancements, extensions and related articles. It is recommended that you periodically check this website for up-to-the-minute information pertaining to the book. Notices and updates for the companion MapCalc software are posted at [www.redhensystems.com/mapcalc/](http://www.redhensystems.com/mapcalc/).

## Acknowledgements

Preparation of a book as complete as this one requires input from several individuals and organizations. Collaboration with *Dr. Kenneth L. Reed* with Accenture Technology Solutions over many years sparked several of the map analysis approaches described in this book. This enduring friendship and interaction dates back to the 1970s when we were junior faculty at Yale University and GIS technology was just a glint in researchers' eyes. Adams Business Media, publishers of *GeoWorld* has supported the "Beyond Mapping" column for over a decade. Several of the topics presented were first introduced through this authoritative resource for spatial information. *Red Hen Systems* provided a critical ingredient—powerful yet inexpensive educational software that is very easy to use. Supporting software by *Golden Software*, *Environmental System Research Institute* and *TechSmith* greatly extend the grid-based exercises and enhance their utility.

Many Thanks!

*Joseph K. Berry*  
*University of Denver*  
*Denver, Colorado*  
*August 2005*