

## GEOGRAPHIC INFORMATION SYSTEMS AND SCIENCE: TEACHING MANUAL

Singleton, A.D., D.J., Unwin, D., Kemp, K.

The original version of this Instructor Manual the result of a collaboration between Professors David Unwin (then of Birkbeck College, University of London) and Karen Kemp (then of the University of Redlands). This updated and expanded version has been prepared for the Third Edition of **Geographic Information Systems and Science** by Dr. Alex Singleton (University College London).

**Alex Singleton** (<http://www.alex-singleton.com>)



Alex Singleton is a Research Fellow in the Department of Geography and Centre for Advanced Spatial Analysis at University College London where he completed his PhD in 2007. His research uses a framework of Geographic Information Science to extend the tradition of area classification in geography and planning. This has developed an empirically informed critique of the ways in which geodemographic

methods can be refined for effective yet ethical use in public resource allocation applications. This research develops from substantive interests that investigate those social, spatial and temporal dimensions of access inequalities in Higher Education. In addition to this core research agenda, he has also examined issues of digital exclusion and deprivation, as well as ways in which 'Neogeography' mapping tools and techniques can be utilised in activities related to the public understanding of science.

**David Unwin**



Until his retirement in 2004, and after a short spell (2002-2004) as a Director of the UK eUniversity, Dave Unwin was Professor of Geography at Birkbeck College, University of London where he taught courses in Geographic Information Science and Environmental Science. He retains an Emeritus Chair in the

College. His research concerns spatial statistical analysis and visualization, particularly in environmental applications, but he also has a strong interest in teaching GIS based originally on his membership of the group that founded the *Journal of Geography in Higher Education*. In 2005 he was awarded the US

University Consortium for Geographic Information Science's Educator of the Year prize for his work in furthering the teaching of GIS.

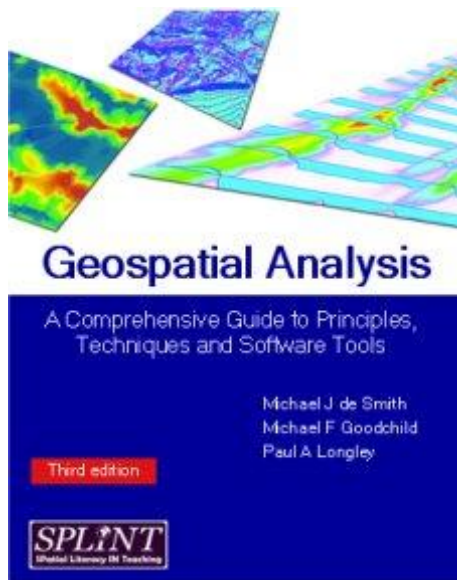
The original exercises in this Instructor's Companion were developed from materials used in his teaching over many years and from his two texts published in 1981 (*Introductory Spatial Analysis*, London: Methuen) and 2003 (*Geographic Information Analysis*, with David O'Sullivan, New York: Wiley) . They were developed in 2005 during a short spell he spent as a guest instructor in the University of Redlands.

**Karen Kemp (<http://www.geokemp.net/>)**

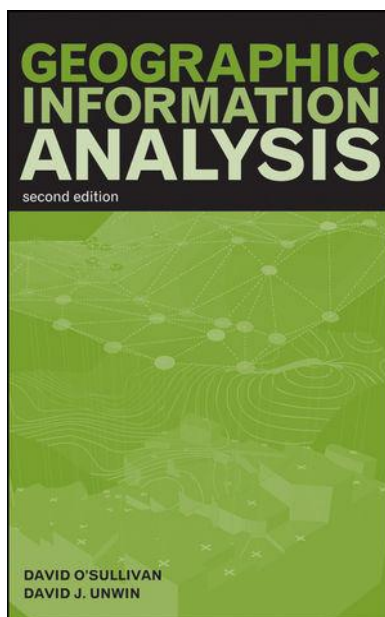


Karen Kemp was the founding Director of the International Masters Program in GIS at the University of Redlands in southern California, having held that position and Professor of Geographic Information Science until December 2005. Dr. Kemp holds geography degrees from the University of Calgary, Alberta (BSc 1976), the University of Victoria, British Columbia (MA 1982) and the University of California Santa Barbara (PhD 1992). Before moving to the US in 1988, she taught Geography, Geology and Microcomputer Applications in the university transfer program at Malaspina College, in Nanaimo, British Columbia. In 1988 she joined the National Center for Geographic Information and Analysis (NCGIA) in Santa Barbara working as Coordinator of Education Programs and co-editor of the internationally recognized *NCGIA Core Curriculum in GIS*. After completing her PhD at UCSB in 1992, she worked at the Technical University of Vienna, Austria, and with Longman GeoInformation in Cambridge, England on various international GIS education projects. She returned to the NCGIA in 1994 to work as Assistant Director and later Associate Director. In January 1999 she moved to the University of California Berkeley to become Executive Director of the Geographic Information Science Center where she helped build the foundation for an innovative campus-wide GIScience initiative. In September 2000, she was invited to join the faculty at the University of Redlands to create and direct their new MS GIS program. In December 2005, Dr. Kemp stepped down as Director of the MS GIS Program and is currently on permanent sabbatical on the Big Island of Hawaii, while continuing to work with the University on various GIS initiatives and as Senior Consultant with the Redlands Institute.

The authors of the book and of this Manual have been involved in two other projects that are intended as a natural extension and progression of the materials covered in the book. The first is



De Smtih M, Goodchild M F, Longley P A 2009  
**Geospatial Analysis: a Comprehensive Guide to Principles, Techniques and Software Tools (third edition). 560(+xxii) pp. Leicester, Troubador** (with contributions by C Castle, A Crooks). This book is intended as an authoritative and independent guide to GIS and spatial analysis in the Internet age. A free-to-access online version is available at [www.spatialanalysisonline.com](http://www.spatialanalysisonline.com), from where the book can also be purchased (with discounts for students). ISBN: 9781848761582; Pages: 560



The second is O'Sullivan D and Unwin D J (2010)  
**Geographic Information Analysis (second edition).** Hoboken, NJ, Wiley, which remains an excellent guide to spatial analysis methods, with links to some of the exercises and activities described in this manual. (Insert web link to the wiley page and add number of pps and ISBN)

The book is available from Wiley:  
[http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470288574\\_descCd-authorInfo.html](http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470288574_descCd-authorInfo.html)

ISBN: 9780470288573; Pages: 432

# Introduction to the Instructor's Manual

The Instructor's Manual is designed to assist you in using this textbook effectively and efficiently. Any instructor wishing to use this text in an educational setting should begin by reading carefully the Preface in which the authors lay out their philosophy for the content and structure of the book.

The Preface gives detailed information about the learning resources to which the text is linked. These include:

- An online course that has been created to accompany the book, 'Turning Data into Information' available through ESRI's Virtual Campus (<http://campus.esri.com>). Each section in the online course is cross-linked to sections in the text and provides learners with the opportunity to practice concepts and techniques they have read about using ESRI's ArcGIS. The materials have been designed to allow learning in a self-paced way, and there are self-test exercises at the end of each section. The course is available, free of charge, to any individual working in an institution that has an ESRI site license. The course catalog for the six-module 'Turning Data into Information' course can be viewed at - [http://training.esri.com/acb2000/showdetl.cfm?DID=6&Product\\_ID=821](http://training.esri.com/acb2000/showdetl.cfm?DID=6&Product_ID=821).
- The four questions that are at the end of each chapter. Each of these has a specific purpose and entail, in the following order:
  1. Student-centered 'learning by doing'.
  2. A review of material contained in the chapter.
  3. A review and research task – involving integration of issues discussed in the chapter with those discussed in additional external sources.
  4. A compare and research task – similar to the review and research task above, but additionally entailing linkage with material from one or more other chapters in the book.
- A website containing this Manual and other resources ([www.wiley.com/college/longley](http://www.wiley.com/college/longley)).

In addition to the materials referenced by the authors, we have given thought to the kinds of resources we would most like to have as support materials when designing and offering a course that uses this text as a foundation. Like the book, these suggestions have been put together in the belief that students do not learn simply by being taught, but rather that learning is essentially an active process, which good instructors at all levels will do their best to facilitate.

In addition to the PowerPoint slides, which are available separately on this website, for each chapter in the book we have provided:

- A summary of the chapter overview
- The intended learning objectives
- Keywords and concepts
- The top level chapter headings
- A bullet point synopsis of the chapter
- Suggested essay style examination questions generally in ascending order of challenge in a Bloom-like taxonomy of educational objectives
- Multiple choice questions, if appropriate to the intended learning outcomes
- In-class and/or individual activities
- Finally, a collected list of further resources taken primarily from the text but with some additions.

# Systems, Science, and Study

---

The chapter introduces the conceptual framework for the book, by addressing several major questions:

- What exactly is geographic information, and why is it important? What is special about it?
- What is information generally, and how does it relate to data, evidence, knowledge, wisdom, and understanding?
- What kinds of decisions make use of geographic information?
- What is a geographic information system (GIS)?
- What is geographic information science, and how does it relate to the use of GIS for scientific purposes?
- How do companies make money from GIS?

## LEARNING OBJECTIVES

- **Know definitions of the terms used throughout the book, including GIS itself;**
- **Be familiar with a brief history of GIS;**
- **Recognize the sometimes invisible roles of GIS in everyday life, and the roles of GIS in business;**
- **Understand the significance of geographic information science, and how it relates to geographic information systems;**
- **Understand the many impacts GIS is having on society, and the need to study those impacts.**

## KEY WORDS AND CONCEPTS

Knowing about 'where'; geographic problems; scale/detail; science and practical problem solving; spatial is special; the sequence data/information/evidence/knowledge/wisdom; knowing how; GIS roles; GIS history; Internet and GIS; GIS as several businesses; 'systems', 'science' and 'studies'

## OUTLINE

- 1.1 Introduction: Why Does GIS matter?
- 1.2 Data, Information, Knowledge, Evidence, Wisdom
- 1.3 Systems and Science
- 1.4 A Brief History of GIS
- 1.5 Views of GIS
- 1.6 The Business of GIS
- 1.7 GISystems, GIScience, and GISudies
- 1.8 GIS and the Study of Geography

## CHAPTER SUMMARY

### 1.1 Introduction: Why does GIS matter?

- GIS are a special class of information systems that keep track not only of events, activities, and things, but also of *where* these events, activities, and things happen or exist.
- Knowing where something happens is of critical importance
- Geographic location is an important attribute of activities, policies, strategies, and plans.
- *Geographic problems*. involve an aspect of location, either in the information used to solve them, or in the solutions themselves
- Three bases for classifying geographic problems

- the scale or level of geographic detail
- the intent or purpose
  - practical problem-solving or *normative* applications, such as finding new locations
  - curiosity driven or *positive* uses that advance science, such as predicting behavior based on location
- the timescale
  - *operational* decisions required for smooth functioning of an operations
  - *tactical* decisions which are medium-term
  - *strategic* decision for long-term direction
  - *transactional* timescales at which databases are updated
- Applications Box 1.1 Hurricane Katrina, August 29, 2005, discusses how location was crucial in the emergency responses to Hurricane Katrina.
- Applications Box 1.2 Where did your ancestors come from? Illustrates curiosity driven research.

### 1.1.1 Spatial is Special

- The basic terms need clarification
  - *Geographic* refers to the Earth's surface and near-surface
  - *Spatial* refers to any space, often used in place of geographic
  - *Geospatial* implies a subset of spatial applied to the Earth's surface and near-surface
- Spatial is special because
  - almost all human activities and decisions involve a geographic component
  - working with geographic information involves unique, complex and difficult choices
- technical reasons are summarized in Technical Box 1.3



## 1.2 Data, Information, Evidence, Knowledge, Wisdom

- No universally agreed definition of these terms, however,
- *Data* are raw facts, neutral and context-free, internal meaning is irrelevant
- *Information* is data refined for some purpose or that have been given some degree of interpretation, often costly to produce but easy to add value to through processing.
- *Knowledge* is information to which value has been added by interpretation based on a particular context, experience, and purpose; it is often acquired over substantial periods and involvement in many projects.
  - *Codified* – can be written and transferred easily to others
  - *Tacit* – is slow to acquire and difficult to transfer accurately
- *Evidence* – is halfway between information and knowledge, a multiplicity of information from different sources whose selection and analysis is focused on specific problems
- *Wisdom* – used in the context of decisions made or advice given
- Table 1.2 compares and illustrates these terms

## 1.3 Systems and Science

- GIS are computer-based systems for storing and processing geographic information.
- GI Science (GISc) is the scientific context and underpinnings of geographic information systems

### 1.3.1 The science of problem solving

- Knowledge about how the world works (*process*) is more valuable than knowledge about how it looks (*form*), because such knowledge can be used to predict
  - *Idiographic* geography focuses on the description of form and emphasizes the unique characteristics of places
  - *Nomothetic* geography seeks to discover general processes
- GIS is useful as a tool for problem solving because it combines the general with the specific
- Software captures and implements general knowledge
- The database represents specific information



- General knowledge for problem solving comes in many forms
  - Classifications
  - Rule sets, some based on statistical generalisations
  - Laws
- Solving problems needs *objectives* which can be expressed in *tangible* (measured on some well-defined scales) and/or *intangible* ways. There may be multiple objectives.

### 1.3.2 The technology of problem solving

- Table 1.3 lists several definitions of GIS and the groups who find them useful

## 1.4 A brief history of GIS

- This section mentions several milestones including
  - The Canada Geographic Information System or CGIS, designed in the mid-1960s as a computerized map measuring system
  - The US Bureau of the Census and the DIME program for the 1970 census
  - Harvard University's Laboratory for Computer Graphics and Spatial Analysis which developed ODYSSEY GIS in the late 1970s
  - UK Experimental Cartography Unit pioneered high quality computer mapping from 1968 up to the mid-1970s
  - Computer mapping at the national mapping agencies
  - Remote sensing developments including military satellites, Landsat and GPS
- "The modern history of GIS dates from the early 1980s, when the price of sufficiently powerful computers fell below a critical threshold."
- Table 1.4 summarizes the major events

## 1.5 Views of GIS

- There are many different perspectives on GIS
  - It is clearly too much for any one software package to handle
- GIS has grown from its initial commercial beginnings as a simple off-the-shelf package to a complex of software, hardware, people, institutions, networks, and activities that can appear very confusing to the novice.

### 1.5.1 Anatomy of a GIS

#### 1.5.1.1 The network

- This section offers a brief history of the Internet and its impacts
- The links between GIS and the Internet are described, as a vehicle for delivering information, applications and services (including location-based services)

#### 1.5.1.2 The other five components of GIS

- Figure 1.16 illustrates the six component parts of a GIS
- Hardware, software, procedures, data, people and network

## 1.6 The business of GIS

- Includes:
  - The software industry
  - The data industry
  - The GISService industry
  - The GeoWeb Service industry
  - The publishing Industry
  - GIS education
- Technical Box 1.4 lists magazines and websites offering GIS news and related services
- Technical Box 1.5 lists scholarly GIS journals
- Technical Box 1.6 lists sites offering Web-based education and training in GIS

## 1.7 GISystems, GIScience and GISudies

- The term *geographic information science* was coined in a paper by Michael Goodchild published in 1992.
- Related terms include geomatics, geoinformatics, spatial information science, geoinformation engineering.

- All suggest a scientific approach to the fundamental issues raised by the use of GIS and related technologies
- Technical Box 1.7 shows the UCGIS GIScience research agenda and ties it to the chapters in the book
- *GIS Studies* can be defined as the systematic study of society's use of geographic information, including its institutions, standards, and procedures

## 1.8 GIS and the Study of Geography

- This section explores this relationship and its sometimes tense characteristics.
- While spatial analysis has a long history, new data handling techniques and rich data sources are moving it strongly to new frontiers
- However, there is enduring unease in some academic quarters about GIS applications and their social implications, including
  - GIS favors certain phenomena and perspectives
    - Often used for purposes that may be ethically questionable or invade individual privacy
    - Concern about a field led by the technology and the marketplace rather than human need (as articulated by academics)
    - GIS as a tool in the hands of the already powerful
- An absence of critical research in GIS
- 'Guilt by implied association' – the uninformed bind GIS to logical positivism, with its restrictive assumptions.

## ESSAY TOPICS

Generally speaking, these get more searching as one progresses down the list:

1. List and outline any four geographical problems to which GIS might be applied.
2. What are the three characteristics that enable distinctions to be made between different geographical problems?
3. The authors state that 'spatial is special'. List seven reasons why this is true.
4. Figure 1.6 shows the 'geography' of the family names of the authors in 1881 and in 1998. What processes over the Twentieth Century do you think account for the similarities and differences between the maps?
5. If you (or any of your classmates) have a fairly common Anglo Saxon family name, visit [gbnames.publicprofiler.org](http://gbnames.publicprofiler.org) and [worldnames.publicprofiler.org](http://worldnames.publicprofiler.org) and investigate (a) the

Great Britain geographies of the family name in 1881 and 1998 and (b) the international dispersal of the name – to North America and Australasia in particular. Perhaps after consulting family members, suggest some of the processes that may have fuelled this dispersal.

6. What are the distinguishing characteristics of the scientific method? Discuss the relevance of each to GIS.
7. Why is a GIS much more than 'a container of maps in digital form'?
8. Is knowing 'where' as important as knowing 'when'?
9. Why should 'knowing about how the world works' be more valuable than 'knowing how it looks'? Illustrate your answer with real world examples.
10. Can a 'single collection of tools' like a GIS ever 'bridge the gap between curiosity-driven science and practical problem solving'?

### **MULTIPLE CHOICE QUESTIONS (MCQ)**

The material, and the relatively high level of the intended learning objectives set for this chapter, do not readily lend themselves to the use of MCQs.

### **ACTIVITIES**

1. Look ahead to Chapter 2, A Gallery of Applications, and select any one of the case studies. Using the information provided, suggest how well it illustrates each step in the sequence: data/information/evidence/knowledge/wisdom.
2. Examine the geographic data available for the area within 50 miles (80 km) of either where you live or where you study. Use it to produce a short (2,500 word) illustrated profile of either the socioeconomic or the physical environment. (See for example <http://www.data.gov/>; <http://www.data.gov.uk>; [www.arcgis.com](http://www.arcgis.com); <http://www.inspire-geoportal.eu/>; or [www.magic.gov.uk](http://www.magic.gov.uk)).
3. Allocate each student the task of researching the career history of a well known protagonist, or detractor, of GIS (e.g. the contributors to the Foresman book referenced below, or one of the authors of the book or of this Instructor Manual). Each student should prepare a ten minute presentation illustrating the importance of software systems, (social, natural or environmental) science and GIS studies to the individual's career history. A summary session might be used to develop the class's thinking as to whether

the academy provides adequate bridgeheads between GISystems, GIScience and GISStudies.

4. A debate on the critical theoretical perspective on GIS. Create two teams, and organize a full debate using as a guide the instructions given in Chapter 5 of Gold et al (1990, see <http://www2.glos.ac.uk/gdn/gold/ch5.htm>). The first team should propose the motion that 'Geographic information systems sell second rate geography' using as evidence relevant essays in Pickles J. 1993 *Ground Truth: The Social Implications of Geographic Information Systems*. New York: Guilford Press. The second team should use the Research Agenda provided by the University Consortium for Geographic Information Science to oppose this motion, see UCGIS 1996 'Research priorities for geographic information science'. *Cartography and Geographic Information Systems* 23(3): 115–127.
5. The fundamental importance of scale can be best illustrated by examination of a series of paper or on-screen maps from a single national mapping agency at scales from, say, 1:2,500 to 1:1,000,000. Using the Google™ search engine to do image searches is an easy way to find suitable maps, but the map extracts sequence provided by the Ordnance Survey of Great Britain at [www.ordnancesurvey.co.uk](http://www.ordnancesurvey.co.uk) are also easy to access.
6. A very good way to introduce the power of Internet mapping is through a project that uses it. A very good source of on-demand thematic mapping can be found <http://www.maptube.org/> and <http://geocommons.com/>. Using these sources, ask students to prepare a presentation of what types of data are contained on the sites, and what this can say about the country in which they live.
7. Figure 1.13, shows the world geography of Internet usage in April 2008 and illustrates a phenomenon that has been called the 'digital divide' – the gap between information rich, Internet connected developed countries and the rest, where even access to electricity is rare. Data from the World Bank website, at [www.worldbank.org/data](http://www.worldbank.org/data) enables a class investigation of this divide more formally, by a correlation analysis of economic indicators for a sample of countries with their reported number of Internet users. Having completed this at the country level, a useful discussion can be held on other regional and local dimensions of the 'divide' such as age, income, gender, and rural/urban differences.
8. Section 1.5.1.2 provides Websites for a number of major GIS vendors. Divide the class into four or more groups, with the objective that each group will represent the interests of

a particular vendor. Each group visits the Websites and produce a summary catalogue of the products on offer. In each case, each groups rates the offering as High/Medium/Low on a series of criteria such as level of user sophistication, intended data volume, commercial applicability, cost, amount of computer power needed, and so on. In a seminar presentation, a spokesperson for each vendor provides a market assessment of the vendor's position in the GIS marketplace.

9. The five M's of GIS are mapping, measuring, management, monitoring and modeling. Ask each student to chose an academic journal from the list given in Technical Box 1.5, or a trade magazine from those in Technical Box 1.4 and using this as source find, say, ten articles that use GIS. In each case classify the principal aim of the article into one or other of these M's. This exercise is valuable in drawing student attention to the literature, and it can be followed by assembling a frequency table using all the class results that should/might provide an interesting commentary on which of the M's is valued by which user community.
10. Either individually or in class, examine the research themes of the UCGIS research agenda listed in Technical Box 1.7. In each case locate the theme on the triangle diagram presented as Figure 1.18. Use a different color symbol for the 'long-term research challenges' and the 'short term research priorities', and comment on any differences that you observe.

## FURTHER READING

Chrisman N.R. 2003 *Exploring Geographical Information Systems* (2nd edn). Hoboken, NJ: Wiley.

At first sight this is a standard text on GIS, but in fact it contains a great deal of wisdom about its social and scientific consequences from one of the pioneers.

Curry M.R. 1998 *Digital Places: Living with Geographic Information Technologies*. London: Routledge.

Together with the set of essays edited by Pickles (1993), Curry's book articulates the 'critical theory' critique of GIS.

Foresman T.W. (ed) 1998 *The History of Geographic Information Systems: Perspectives from the Pioneers*. Upper Saddle River, NJ: Prentice Hall.

An edited series of essays that together show how GIS developed, largely from a North American perspective.

Goodchild M.F. 1992 'Geographical information science'. *International Journal of Geographical Information Systems* 6: 31–45.

THE seminal paper defining the field:

Longley P.A. and Batty M. (eds.) 2003 *Advanced Spatial Analysis: The CASA Book of GIS*. Redlands, CA: ESRI Press.

A series of useful case studies showing the richness of advanced research in spatial analysis.

Pickles J. 1993 *Ground Truth: The Social Implications of Geographic Information Systems*. New York: Guilford Press.

When it was published, the essays in this book, which are almost all critical of GIS as it was then practiced, led to a flurry of counter claims and challenges in what has been called geography's version of the 'science wars'. From the perspective of a decade a truce seems to have been called.

University Consortium for Geographic Information Science 1996 'Research priorities for geographic information science'. *Cartography and Geographic Information Systems* 23(3): 115–127.

Exactly what it says on the tin can. This is also available at [www.ucgis.org](http://www.ucgis.org)

## RELATED READING

Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, NJ: Wiley.

- 3. Geography and GIS, R J Johnston
- 4. Arguments, debates and dialogues: the GIS-social theory debate and the concern for alternatives, J Pickles J
- 40. The future of GIS and spatial analysis, M F Goodchild, P A Longley
- 54. Enabling progress in GIS education, P Forer, D Unwin



Maguire D.J., Goodchild M.F. and Rhind D.W. (eds) 1991 *Geographical Information Systems: Principles and Applications*. Harlow, UK: Longman (text available online at [www.wiley.co.uk/gis/volumes.html](http://www.wiley.co.uk/gis/volumes.html)).

- An overview and definition of GIS, D J Maguire, pp. 9-20
- The history of GIS, J T Coppock and D W Rhind, pp. 21-43
- The technological setting of GIS, M F Goodchild, pp. 45-54
- The commercial setting of GIS, J Dangermond, pp. 55-65
- The academic setting of GIS, D J Unwin, pp. 81-90
- The organizational home for GIS in the scientific professional community, J L Morrison, pp. 91-100

Clarke, K.C. (2010) *Getting Started with Geographic Information Systems*. Fifth Edition. Boston, USA: Prentice Hall.

O'Sullivan, D., Unwin, D. (2010) *Geographical Information Analysis*. Second Edition. Hoboken, USA, John Wiley and Sons.

## ONLINE RESOURCES

For an independent review of the analytical capabilities of GIS software see:

[www.geospatialanalysis.com](http://www.geospatialanalysis.com)

For an excellent summary of the history of GIS see the GIS Timeline at

[www.casa.ucl.ac.uk/gistimeline](http://www.casa.ucl.ac.uk/gistimeline)

NCGIA Core Curriculum in GIScience, 2000 ([www.ncgia.ucsb.edu/giscc](http://www.ncgia.ucsb.edu/giscc))

What is GIS?- Michael Goodchild (<http://www.ncgia.ucsb.edu/giscc/units/u002/>)

Land Information Systems and Cadastral Applications Steve Ventura

(<http://www.ncgia.ucsb.edu/giscc/units/u164/>)

## 2

# A gallery of applications

---

## OVERVIEW

The chapter gives a flavor of the breadth and depth of real-world GIS implementations. It considers:

- How GIS affects our everyday lives;
- How GIS applications have developed, and how the field compares with scientific practice;
- The goals of applied problem solving;
- How GIS can be used to study and solve problems in transportation, the environment, local government, and business.

## LEARNING OBJECTIVES

The chapter teaches students to:

- **Grasp the many ways in which we interact with GIS in everyday life;**
- **Appreciate the range and diversity of GIS applications in environmental and social science;**
- **Be able to identify many of the scientific assumptions that underpin real-world applications;**
- **Understand how GIS is applied in the representative application areas of transportation, the environment, local government, and business.**

## KEYWORDS AND CONCEPTS

Problem solving with GIS; the five M's; government and public service; tax assessment; business and service planning; geodemographics; housing study; logistics and transportation; emergency evacuation; environment; deforestation

## OUTLINE

2.1 Introduction

2.2 Science, geography, and applications

2.3 Representative application areas and their foundations

2.4 Concluding comments

## CHAPTER SUMMARY

### 2.1 Introduction

#### 2.1.1 One day of life with GIS

- A fictitious daily diary highlights how GIS:
  - affects each of us, every day;
  - can be used to foster effective short- and long-term decision making;
  - can be applied to many socio-economic and environmental problems;
  - encourages public participation in decision making
  - supports mapping, measurement, management, monitoring, and modeling operations;
  - generates measurable economic benefits;
  - requires key management skills for effective implementation;
  - provides a challenging and stimulating educational experience for students;
  - can be used as a source of direct income;
  - can be combined with other technologies;
  - is a dynamic and stimulating area in which to work.

#### 2.1.2 Why GIS?

- GIS is being widely implemented because of

- Wider availability of GIS through the Internet, as well as through organization-wide local area networks.
- Increasingly wide availability of low-cost, locationally aware, hand-held devices.
- Reductions in the price of GIS hardware and software because economies of scale are realized by a fast-growing market.
- Greater awareness that decision making has a geographic dimension.
- Greater ease of user interaction, using standard windowing environments.
- Better technology to support applications, specifically in terms of visualization, data management and analysis, and linkage to other software.
- Proliferation of geographically referenced digital data, such as those generated using GPS technology or supplied by value-added resellers (VARs) of data.
- Availability of open-source, free to use software across the Web, in addition to sophisticated packaged GIS solutions that are ready to run out of the box.
- The accumulated experience of applications that *work*.

## 2.2 Science, geography, and applications

### 2.2.1 Scientific questions and GIS operations

- Within the spatial domain, the goals of applied problem solving include,
  - Managing spatial operations and inventories
  - Rational, effective, and efficient allocation of resources
  - Monitoring and understanding observed spatial distributions of attributes
  - Understanding the difference that place makes
  - Understanding of processes in the natural and human environments
  - Prescription of strategies for environmental maintenance and conservation

### 2.2.2 GIScience applications

- GIS applications need to be grounded in sound concepts and theory.

## 2.3 Representative application areas and their foundations

### 2.3.1 Introduction and overview

- Applications generally fulfill the five M's of GIS: mapping, measurement, monitoring, modeling, and management
- Applications can be traditional, developing, and new
- Figure 2.6 shows the classic Rogers model of innovation diffusion and the text applies this to GIS
- Four domains of GIS applications are: Government and public service; Business and service planning; Logistics and transportation; and Environment.

### 2.3.2 Government and public service

- As GIS has become cheaper, so it has come to be used in government decision making at all levels from the nation to the neighborhood.
- Figure 2.6 shows the hierarchy of GIS use in government decision-making
- Table 2.1 summarizes GIS applications in local government including inventory applications, policy analysis, and management/policy-making

#### 2.3.2.2 Case study: GIS in tax assessment

- Scientific foundations (2.3.2.4)
  - is dependent on an unambiguous definition of parcels, and common standards about how different characteristics (such as size, age, and value of improvements) are represented.
  - although the application is driven by results rather than scientific curiosity, it follows scientific procedures of controlled comparison.
- Principles
  - Combines Tobler's First Law with local knowledge

- Techniques
  - Tax assessment requires a good database, a plan for system management and administration, and a workflow design.
- Analysis
  - Tax assessment uses standard GIS techniques such as proximity analysis, and geographic and attribute query, mapping, and reporting.
- Generic scientific questions (2.3.2.5)
  - Once a property database has been created, it becomes a very valuable local asset
  - Public works departments may use it to label access points for repairs and meter reading
  - Housing departments may use it to maintain data on property condition
  - Many other departments may like shared access to a common address list for record keeping and mailings.
- Management and Policy (2.3.2.6)
  - Easy to develop a cost-benefit case for this application
  - GIS is an important tool for efficiency and equitable local government.

### 2.3.3 Business and service planning

- Geodemographic analysis is an important operational tool in market area analysis, where it is used to plan marketing campaigns.
  - *Geodemographics* is a shorthand term for composite indicators of consumer behavior that are available at the small-area level (e.g., census output area, or postal zone).
  - Often combined with *lifestyles* data on the consumption choices and shopping habits of individuals
  - *Market area analysis* describes the activity of assessing the distribution of retail outlets relative to the greatest concentrations of potential customers.

- Is increasingly being adapted to improving public service planning, in areas such as health, education, and law enforcement.
- Tools are used to analyze and inform the range of *operational*, *tactical*, and *strategic* functions of an organization.
  - Operational functions concern the day-to-day processing of routine transactions and inventory analysis in an organization, such as stock management (see Logistics section following)
  - Tactical functions require the allocation of resources to address specific (usually short term) problems, such as store sales promotions.
  - Strategic functions contribute to the organization's longer-term goals and mission, and entail problems such as opening new stores or rationalizing existing store networks

#### **2.3.3.2 Case study: Hierarchical diffusion and convenience shopping (Tesco)**

- Objective is to promote a new 'Express' format Tesco store to encourage repeat patronage.
- Method
  - Identify postal addresses all of the households within a 1 km radius.
  - Overlay geodemographic profiles in order to tailor coupon offerings to the differing consumption patterns (from lifestyles data)
- Scientific foundations
  - Distance decay and Tobler's First Law
    - An assumption that the differences in the observed social and economic characteristics of residents between neighborhoods are greater than differences observed within them.
    - Has the potential to invoke ecological fallacy – individual resident in an area is assigned the characteristics of the area.
    - Text discusses ethical and scientific implications



- Principles
  - Linear distance versus other kinds of distance (network, psychological)
- Techniques
  - Assigning postcode coordinates to catchment area requires point in polygon analysis
- Analysis
  - Need to include analysis of spatial interaction between this and other stores
- Generic scientific questions (2.3.3.5)
  - While it is the mix of individuals with particular characteristics that largely determines the likely store turnover of a particular location, this example illustrates the kinds of simplifying assumptions that we may choose to make using the best available data in order to represent consumer characteristics and store attributes.
  - Even blunt-edged tools can increase the effectiveness of operational and strategic R&D activities many-fold.
- Management and policy (2.3.3.6)
  - Increasing role of SAPs as mainstream managers alongside accountants, lawyers, and general business managers.
  - Key roles in organizational activity such as marketing, store revenue predictions, new product launch, improving retail networks, and the assimilation of pre-existing components into combined store networks following mergers and acquisitions.

### **2.3.4 Logistics and transportation**

- Deal with the movement of goods and people from one place to another, and the infrastructure (highways, railroads, canals) that moves them.
- Has two parts
  - the static part that deals with the fixed infrastructure, and

- the dynamic part that deals with the vehicles, goods, and people that move on the static part.
- GPS is an important technology in this area
- Many applications involve optimization, or the design of solutions to meet specified objectives

#### **2.3.4.2 Case study: Planning for emergency evacuation**

- Discusses a planning tool that allows neighborhoods to rate the potential for problems associated with evacuation, and to develop plans accordingly.
  - The tool uses a GIS database containing information on the distribution of population in the neighborhood and the street pattern.
  - The result is an evacuation vulnerability map and identification of worst-case scenarios for a given locations.
- Method (2.3.4.3)
  - Census data are used to determine population and household counts, and to estimate the number of vehicles involved in an evacuation.
  - The locations of streets are obtained from street centerline files, which give the geographic locations, names, and other details of individual streets
  - Every intersection in the network is tested to see if it presents a bottleneck, by dividing the total number of vehicles that would have to move out of the neighborhood by the number of exit lanes.
- Scientific foundations
  - Census data are aggregated to areas that, while small, nevertheless provide only aggregated counts of population.
  - The street layouts of TIGER and other sources can be out of date and inaccurate, particularly in new developments, although users willing to pay higher prices can often obtain current data from the private sector.
  - the essentially geometric approach cannot deal with many social issues: evacuation of the disabled and elderly, and issues of culture and language that may impede evacuation.

- Principles
  - Central to the analysis is connectivity, an instance of a *topological* property
- Techniques
  - Spatial interpolation performed across the network from the values calculated at intersections
  - Shortest path
- Analysis
  - an excellent example of the use of GIS analysis to *make visible what is otherwise invisible*.
- Generic scientific questions (2.3.4.5)
  - Logistic and transportation applications of GIS rely heavily on representations of networks, and often must ignore off-network movement.
- Management and policy (2.3.4.6)
  - GIS is applied to this area in all three modes - operational, tactical and strategic

### 2.3.5 Environment

- Monitoring land use change
- Assessing the impact of urban settlements
- Simulation of processes in the urban and natural environment

#### 2.3.5.2 Case study: Deforestation in the Philippines

- Objective is to identify a range of different development scenarios that make it possible to anticipate future land use and habitat change, and hence also anticipate changes in biodiversity.
- Method (2.3.5.3)
  - used qualitative data collected through stakeholder interviews in a quantitative GIS-based analysis to calculate the probabilities of land use transition under three different scenarios of land use change

- The three different scenarios not only resulted in different forest *areas* by 2019 but also different *spatial patterning* of the remaining forest.
- Scientific foundations (2.3.5.4)
  - The theme of inferring *process* from *pattern*, or function from form, is a common characteristic of GIScience applications.
  - Contrasts nomothetic and idiographic approaches
- Principles
  - GIS makes it possible to incorporate diverse physical, biological, and human elements, and to forecast the size, shape, scale, and dimension of land use parcels.
  - It makes use of the core GIS idea that the world can be understood as a series of layers of different types of information, that can be added together meaningfully through overlay analysis to arrive at conclusions.
- Analysis
  - Process is inferred not just through size measures, but also through spatial measures of connectivity and fragmentation
- General scientific questions (2.3.5.5)
  - Irrespective of the quality of the measurement process, uncertainty will always creep into any prediction
  - Data are never perfect
  - Simulations are subject to exogenous forces not included
  - GIS users should not think of systems as black boxes,
  - Users of GIS should always know exactly what the system is doing to their data.
  - User awareness of these important issues can be improved through appropriate metadata and documentation of research procedures
  - The results of analysis should always be reported in sufficient detail to allow someone else to replicate them.

## 2.4 Concluding comments

- The principles of the scientific method have been stressed throughout
  - the need to maintain an enquiring mind, constantly asking questions about what is going on, and what it means;
  - the need to use terms that are well-defined and understood by others, so that knowledge can be communicated;
  - the need to describe procedures in sufficient detail so that they can be replicated by others; and
  - the need for accuracy, in observations, measurements, and predictions.

## ESSAY TOPICS

1. Write an account of how you think that GIS impacts on your daily life. Now imagine the world in a decade's time and speculate how things might have changed. (Save your answer for future reference!)
2. Since the mid 1990s there has been an explosive growth in the range of applications of GIS technology. Explain why this has happened.
3. It is often suggested that GIS is an 'applications led' technology. What does it mean to be 'applications led' and what are the likely consequences for the design and development of GI software?
4. Write an account of the ways in which GIS can assist in business logistics.
5. One of the major ethical issues raised by critics of GIS is that of protecting personal privacy. List and review the privacy implications of each of the case study applications discussed in this chapter.
6. How has the emergence of the Global Positioning System (GPS) affected applications in GIS?
7. Geodemographics assumes that where you live is a good predictor of your social and economic status ('birds of a feather flock together'). To what extent is this assumption justified?

## MULTIPLE CHOICE QUESTIONS (MCQ)

The material, and the intended learning objectives set for this chapter, do not readily lend themselves to the use of MCQs.

## CLASS AND INDIVIDUAL ACTIVITIES

1. In Chapter 2 (2.3.1), the five M's of GIS are claimed to be mapping, measuring, management, monitoring and modeling. Examine each of the applications described in Sections 2.3.2, 2.3.3, 2.3.4 and 2.3.5, and in each case classify the application into its principal M. Is there anything to be learned from these differences?
2. Devise a diary for your own activity patterns for a typical (or a special) day, like that described in Section 2.1.1, and speculate how GIS might affect your own daily activities. What activities are not influenced by GIS, and how might its use in some of these contexts improve your daily quality of life?
3. Compare and contrast the operational, tactical, and strategic priorities of the GIS specialists responsible for the specific applications described in Sections 2.3.2, 2.3.3, 2.3.4 and 2.3.5.
4. The Chapter provides detail of four applications of GIS (2.3.2.3, 2.3.3.2, 2.3.4.2 and 2.3.5.2). In each case, identify the 'added value' in using a GIS in preference to a standard data base management system (DBMS).
5. Visit [gbnames.publicprofiler.org](http://gbnames.publicprofiler.org) and type in your own surname and/or those of several famous people in history. You will see that the information returned to you for each search includes an index of the geodemographic profile of people with the selected surname. Does this descriptive profiling tell you anything useful?
6. Assume that you are working in a brand new GIS facility located in a water supply/resources company. Your responsibility extends to the rivers and water distribution mechanisms over a county-sized area. Your GIS is needed to support operations such as monitoring rainfall, reservoir capacity, stream flow and abstraction for public use. Think about, write down, and justify how the GIS could be used. This outline project briefing can be modified according to taste to address almost any applications domain. The books listed under Resources can be used to assist!
7. A role-play simulation (see Gold et al., Section 5.6). In a GIS context a useful simulation to set up is a 'competitive' bid for a contract to create a specific GIS. Divide the class into teams for this, spreading the exercise over at least three formal sessions. Each team takes on the role of a GI consultancy bidding for a contract that addresses a deliberately, but realistically brief and poorly specified request for proposals (RFP) drafted by the instructor. Each bid must specify the number of person days that the work will take, and include costings for completing the work based on labour market rates, supplied by the

instructor. The final session should consist of timed presentations with voting to decide which bid wins. A variant of this is to invite groups to bid for one of a range of consultancy projects, spread across a range of applications domains. This exercise is based on a similar suggestion from Peter Keene: Keene, P. (1988) Teaching physical geographers to talk. *Journal of Geography in Higher Education*, 12 (1), 85-94. It also introduces students to terminology used in the AM/FM Project Life Cycle Model, an introduction to which can be found at the Geographer's Craft website:

<http://www.colorado.edu/geography/gcraft/notes/lifecycle/lifecycl.html>

8. A similar role play, making use of the same sources of reference, is drafting a Request for Information (RFI) for some obvious GI systems. Use four groups of 14 or so, asking them to produce a formal RFI of, say, 1,500 words. Students are required to act as GIS-literate members of an IT section, requested by management to produce such a document. It is assumed that it is in their interests to procure such a system. Topics should be as specific as possible, even down to using agencies that actually exist and that have formal terms of reference or obvious objectives (eg. a supermarket planning system to vary their product mix geographically, a specified River Authority, a police authority). In seminar, each presents the RFI to their 'senior management' for approval (15 minutes/team).
9. Reading a paper. All students read the same paper and isolate what they think to be the two most important issues that arise from their reading. Students then discuss the paper in pairs, isolating the two most important of their four issues. Next, the exercise is repeated in groups of four and so on until a suitable stopping point arises. For obvious reasons, this device is called a pyramid exercise, and its value lies in the student discussion. For this chapter, and although the detail is dated, a suitable paper is that by Goddard and Openshaw: Openshaw, S. & J Goddard (1987) Some implications of the commodification of information and the emerging information economy for applied geographical analysis in the United Kingdom *Environment and Planning A* 19, 1423 - 1439

## **FURTHER READING**

Birkin M., Clarke G.P., and Clarke M. 2002 *Retail Geography and Intelligent Network Planning*. Chichester, UK: Wiley.

Bonham-Carter G F 1994 *Geographic information systems for geoscientists: modeling with GIS*. New York, Pergamon (see the various applications developed throughout the book).



- Chainey S., and Ratcliffe J. 2005 *GIS and Crime Mapping*. Chichester UK: Wiley.
- Goodchild M F, Parks B O, Steyaert L T 1993 *Environmental modeling with GIS*. New York, Oxford University Press (see the various applications developed throughout the book).
- Greene R.W. 2000 *GIS in Public Policy*. Redlands, CA: ESRI Press.
- Haines-Young R, Green D R, Cousins S 1993 *Landscape ecology and geographic information systems*. London, Taylor and Francis
- Harris R., Sleight P., and Webber R. 2005 *Geodemographics, GIS and Neighbourhood Targeting*. Chichester, UK: Wiley.
- Johnston C.A. 1998 *Geographic Information Systems in Ecology*. Oxford: Blackwell.
- Longley P A, Martin D J, Higgs, G 1994 The predictive use of GIS to model property valuations. *International Journal of Geographical Information Systems* 8: 217-35
- Longley P A, Higgs G, Martin D J 1996 The rates revisited? A geographical reassignment of property valuations and local tax burdens under the council tax. *Environment and Planning C* 14: 101-20
- Martin D J, Longley P A 1995 Data sources and their geographical integration. In P Longley, G Clarke (eds) *GIS for Business and Service Planning*. New York, John Wiley: 15-32
- O'Looney J. 2000 *Beyond Maps: GIS and Decision Making in Local Government*. Redlands, CA: ESRI Press.

## RELATED READING

Maguire D.J., Goodchild M.F. and Rhind D.W. (eds) 1991 *Geographical Information Systems: Principles and Applications*. Harlow, UK: Longman (text available online at [www.wiley.co.uk/gis/volumes.html](http://www.wiley.co.uk/gis/volumes.html)).

See all the chapters in Section III Applications

## ONLINE RESOURCES

ESRI Virtual Campus, ([training.esri.com](http://training.esri.com)) - see various courses on applications of GIS

There are numerous blogs detailing many new and innovative applications of GIS:

- <http://www.spatiallyadjusted.com/>

- <http://www.esri.com/blogs>
- <http://blog.gisuser.com/>
- <http://mapperz.blogspot.com/>

# Representing Geography

---

## OVERVIEW

- This chapter introduces the concept of representation, or the construction of a digital model of some aspect of the Earth's surface.
- The geographic world is extremely complex, revealing more detail the closer one looks. So in order to build a representation of any part of it, it is necessary to make choices, about what to represent, at what level of detail, and over what time period.
- Generalization methods are used to remove detail that is unnecessary for an application, in order to reduce data volume and speed up operations.

## LEARNING OBJECTIVES

**After reading this chapter, your students will be able to explain:**

- **The importance of understanding representation in GIS;**
- **The concepts of fields and objects and their fundamental significance;**
- **What raster and vector representation entails and how these data structures affect many GIS principles, techniques, and applications;**
- **The paper map and its role as a GIS product and data source;**
- **The importance of generalization methods and the concept of representational scale;**
- **The art and science of representing real-world phenomena in GIS.**

## KEY WORDS AND CONCEPTS

Digital, binary, representation, Tobler's First Law of Geography, attributes, the fundamental problem (the world is infinitely complex), discrete objects, attribute tables, continuous fields, raster, vector, pixel, polygon, polylines, TIN, representative fraction, generalization, simplification, weeding.

## OUTLINE

- 3.1 Introduction
- 3.2 Digital representation
- 3.3 Representation of what and for whom?
- 3.4 The fundamental problem
- 3.5 Discrete objects and continuous fields
- 3.6 Rasters and vectors
- 3.7 The paper map
- 3.8 Generalization
- 3.9 Conclusion

## CHAPTER SUMMARY

### 3.1 Introduction

- This section mentions the notions of
  - *spatial concepts* such as containment and proximity
  - *data models*, the structures and rules that are programmed into a GIS to accommodate data
  - *ontologies*, the frameworks that we use for acquiring knowledge of the world
- Figure 3.1 is a time-space *representation* of the daily journeys of several people
- Representations are reinforced by the rules and laws that we humans have learned to apply to the unobserved world around us, such as using spatial interpolation to guess the conditions that exist in places where no observations were made.
- Introduces *Tobler's First Law of Geography*: Everything is related to everything else, but near things are more related than distant things.

### 3.2 Digital representation

- This section introduces the notions of *digital* and *binary*.

- Technical Box 3.1 outlines fundamentals about the binary number system, including short (16-bit) and long (32-bit) storage, ASCII, floating point numbers, and BLOBs (binary large object).
- Benefits of digital representations of geography include: can be handled in ways that are independent of meaning; easy to copy and transmit; stored at high density; easy to transform, process, and analyze.

### 3.3 Representation for what and for whom?

- *Geographic* representation is defined as a representation of some part of the Earth's surface or near surface, at scales ranging from the architectural to the global
- Short history of maps and the significance of their use historically
- Like maps, any application of GIS requires attention to questions of *what* should be represented, and *how*.

### 3.4 The fundamental problem

- Geographic data are built up from atomic elements, or facts about the geographic world.
- At its most primitive, an atom of geographic data (strictly, a datum) links a *place*, often a *time*, and some descriptive *property*.
- *Attribute* refers to this descriptive property and may:
  - vary slowly or rapidly
  - be physical/environmental or social/economic
  - identify a place or entity
  - provide a measurement of something at that place
- Technical Box 3.3 defines the types of attributes: nominal, ordinal, interval, ratio, and cyclic
- The *fundamental problem* is “the world is infinitely complex, but computer systems are finite”.

### 3.5 Discrete objects and continuous fields

#### 3.5.1 Discrete objects

- In the *discrete object* view, the world is empty, except where it is occupied by objects with well-defined boundaries that are instances of generally recognized categories.
  - Objects can be counted

- Objects have dimensionality: 0-dimension (points), 1-dimension (lines), 2-dimensions (areas, polygons)
- 3-dimensions are problematic in GIS
- Associating attributes to discrete objects can be expressed in a table (see Table 3.1):
  - each row corresponding to a different discrete object,
  - each column corresponding to an attribute of the object.
- A table does not look like the real world

### 3.5.2 Continuous fields

- The continuous field view represents the real world as a finite number of variables, each one defined at every possible position.
- Continuous fields, on the other hand, can be distinguished by what varies, and how smoothly.
- The challenge of counting the lakes in Minnesota provides an elegant illustration of various ways to think about the conceptual transition from objects to fields
- Technical Box 3.4 explains the concept of 2.5 dimensions

## 3.6 Rasters and vectors

- raster and vector are two methods that are used to reduce geographic phenomena to forms that can be coded in computer databases
- In principle, each can be used to code both fields and discrete objects, but in practice there is a strong association between raster and fields, and between vector and discrete objects.

### 3.6.1 Raster data

- In a raster representation geographic space is divided into an array of cells, each of which is usually square, but sometimes rectangular (Figure 3.9).
- All geographic variation is then expressed by assigning properties or attributes to these cells.
- The cells are sometimes called pixels (short for *picture elements*).
- A few fundamentals of remote sensing are introduced
- When information is represented in raster form all detail about variation within cells is lost, and instead the cell is given a single value based on a rule such as largest share or value at the central point.

### 3.6.2 Vector data

- Usually, in a vector representation, all lines are captured as points connected by precisely straight lines (some GIS software also allows curves).
- Features are captured as a series of points or *vertices* connected by straight lines
  - areas are often called polygons
  - lines are sometimes called polylines
  - the choice between raster and vector is often complex, as summarized in Table 3.3.

### 3.6.3 Representing continuous fields

The six approximate representations of a field used in GIS are (Figure 3.12)

- Regularly spaced sample points
- Irregularly spaced sample points.
- Rectangular cells.
- Irregularly shaped polygons.
- Irregular network of triangles
- Polylines representing contours

## 3.7 The paper map

- Key property of a paper map is its *scale* or *representative fraction*, defined as the ratio of distance on the map to distance on the Earth's surface.
  - when a scale is quoted for a digital database it is usually the scale of the map that formed the source of the data.
  - while the paper map is a useful metaphor for the contents of a geographic database, we must be careful not to let it limit our thinking about what is possible in the way of representation.
- Digital representations can include information that would be very difficult to show on maps.
- Digital databases can represent all three spatial dimensions, including the vertical, whereas maps must always show two-dimensional views.

## 3.8 Generalization

- ways of simplifying the view of the world include:
  - describe entire areas, attributing uniform characteristics to them, even when areas are not strictly uniform;

- identify features on the ground and describe their characteristics, again assuming them to be uniform;
- limit our descriptions to what exists at a finite number of sample points, hoping that these samples will be adequately representative of the whole.
- some degree of generalization is almost inevitable in all geographic data.
- A map's specification defines how real features on the ground are selected for inclusion on the map.

### 3.8.1 Generalization about places

- Geometry generalization methods, see Figure 3.14
- Attribute generalization methods include
  - *Simplification*
  - *Smoothing*
  - *Aggregation*
  - *Amalgamation*
  - *Merging*
  - *Collapse*
  - *Refinement*
  - *Exaggeration*
  - *Enhancement*
  - *Displacement*
- Weeding is the process of simplifying a line or area by reducing the number of points in its representation
- The operation of the Douglas-Poiker weeding algorithm is shown in Figure 3.16.

### 3.8.2 Generalization about properties

- Many representations in GIS bring together multiple properties of places as composite indicators of conditions at particular locations
- Representation may not be fit for its purpose if it is based upon inappropriate indicators of real-world conditions

### 3.8.3 Representation using VGI

Generalizability of representations based on VGI is questioned as follows:

- Are all places equally accessible to volunteers?
- Are volunteers more likely to provide data on places or properties that interest them than those that do not?



- Where volunteers supply information about themselves, is this representative of other non-volunteers?
- Are volunteers at liberty to collect locationally sensitive data?
- Should volunteers supply data that are open to malevolent use?
- Is it socially acceptable to make available any observations of uniquely identifiable individuals?
- Is the information date-stamped, as an indicator of its provenance and current reliability?

## ESSAY TOPICS

Again, these are in rough order of sophistication of necessary response:

1. What do you understand by the terms raster and vector? How would you decide which to use in any specific given project? Describe the situations encountered in this chapter where the distinction between raster and vector becomes blurred.
2. Why is digital map generalization necessary and why does it involve so much more than line simplification?
3. GIS professionals and cartographers represent and understand 'geography' in certain ways. Compare and contrast with your own everyday experience of spaces and places.
4. Distinguish between nominal, ordinal, interval, ratio, and cyclic data types, and explain how this classification does not cover all the attribute types that might be recorded in a GIS
5. Table 3.3 summarized some of the arguments between raster and vector representations. Expand on these arguments, providing examples, and add any others that would be relevant in a GIS application.
6. "No representation of geographic phenomena can ever be perfect". Is this true, are there exceptions, and what implications does this statement have for users of GIS?
7. A map is defined by the International Cartographic Association as 'A conventionalized image representing selected features or characteristics of geographic reality designed for use when spatial relationships are of primary relevance' (Board, 1991). Explain the reasoning behind every element of this definition.
8. An argument has been made that representing geography in a GIS can be improved by understanding human perception and cognition. Review and discuss this argument. In what ways do you agree or disagree with this idea?
9. To what extent do you agree with the two propositions expressed by Goodchild in 1995 that 'It is GIS's supreme conceit that one can structure a useful representation of

- geographical knowledge in the absurdly primitive domain of the digital computer' and that 'it is geography's conceit that one can accomplish the same with pen and paper'?
10. With reference to sound, maps, and photographic images, write an account of the advantages and disadvantages of digital representations over analog ones.

### MULTIPLE-CHOICE QUESTIONS

1. The following lists show a series of numbers and some binary equivalents. Each binary number consists of four binary digits (bits). Match each number to its representation in the binary system by writing its sequence letter in the final column:

Number	Binary	(a)-(e)?
15	0000	
0	1001	
9	0110	
6	1111	
10	1010	

2. Which form of representation does a paper map use?
- a. analog
  - b. digital
  - c. binary
  - d. decimal
3. Which is **NOT** a commonly used coding scheme for images?
- a. JPEG
  - b. GIF
  - c. MP3
  - d. TIFF

4. The table lists some commonly used attributes. For each write down the measurement scale (nominal, ordinal, interval, ratio) most often used to record it:

Attribute	Scale
Building height	
Temperature in °C	
Land capability	
Soil type	
Atmospheric pressure	

5. Which is **not** characteristic of discrete objects?
- they may include points, lines, and areas
  - they completely cover the space
  - they can overlap
  - they can be counted
6. When weather forecasters isolate a cold front on a weather map are they representing this phenomenon as:
- A table
  - A surface
  - A field
  - An object?
7. Of the six ways of representing a field in GIS, which is usually used for remotely sensed images?
- a raster of regularly spaced sample points
  - a raster of rectangular cells
  - irregularly spaced sample points
  - irregularly shaped polygons
8. Ontology is the study of:
- the processes by which knowledge is acquired
  - the processes by which the world is described

- c. properties that are not affected by warping space
- d. the naming of places on the Earth's surface?

9. Figure 3.13 is an extract from a 1:24,000 USGS topographic map. Study it carefully and then complete the table below by specifying how it represents each of the named entities in the appropriate blank cells. Note that there may be more than one method for each entity:

Entity	Field	Point object	Line object	Area object
Relief				
Housing				
Wood/forest				
Roads				
Railway				
Public buildings				
Rivers				
Lakes				

## CLASS AND INDIVIDUAL ACTIVITIES

1. In class, one of the best ways of stimulating discussion is to look critically at a series of paper maps. Understanding the nature of maps and mapping is a vital step towards understanding GIS. Several of these suggested projects demand access to an extensive map library, or its copyright cleared digital equivalent. Appropriate examples for individual study will be found using the 'image' search at [www.google.com](http://www.google.com).
2. What fraction of the Earth's surface have you experienced in your lifetime? Consider how you would make diagrams at appropriate levels of detail, to show a) where you have lived in your lifetime, b) how you spent last weekend. How would you describe what is missing from each of these diagrams? How could these be shared?

3. The early explorers had limited ways of communicating what they saw, but many were very effective at it. Examine the published diaries, notebooks, or dispatches of one or two early explorers and look at the methods they used to communicate with others. What words did they use to describe unfamiliar landscapes and how did they mix words with sketches?
4. The following exercise can be conducted in class as a sort of collective 'thought experiment'. The 'answers' don't matter. What it does is to look at and discuss a variety of 'paper' maps and 'pre-maps'. It tries to:
  - get students into looking at maps and map-like products and so introduce them to the art and science of cartography;
  - help put the products that can be obtained from a GIS into a cartographic perspective.

The idea is to find 'sequences' from what all can agree are definitely 'maps' to what are equally definitely images and/or 'plans'.

For example, all maps are scaled representations of the real world but at what point does a map become what we normally think of as a 'plan' as we progress from say, 1:250,000 road map, through 1:50,000 topographic map, 1:10,000 map, to 1:2,500, a map of a room, and even smaller scale? Students should be encouraged to find their own examples at websites, in a local library, or in their city or university map store.

For links to URLs concerned with cartography, see

<http://oddens.geog.uu.nl/index.html>.

Similar sequences can be assembled to illustrate how maps distort (is an area cartogram a 'map'? How does this distortion differ from that on a UTM projection? What can be gleaned about how they generalize, the viewpoint and datum they adopt, their subject matter (is a 'map' of DNA still a 'map'?)). The book by Dodge, Kitchin and Perkins (2009) *Rethinking Maps: New Frontiers in Cartographic Theory* (Routledge) also has some good source materials.

5. This is an in-class or project variant on Multiple-Choice Question 10. Take map extracts from the Ordnance Survey of GB or a similar National Mapping Agency, maps at five different scales (1:10,000, 1:25,000, 1:50,000, 1:250,000 & 1:625,000),

and spend, say, 15 minutes making notes and thinking about how these maps become more or less generalized as the scale changes. Individual students can limit what they do to one or two examples of how specific features are represented as point, line, or area objects. They should be encouraged to think about how objects change their representation with scale and the issue of planimetrically 'correct' representation vs the use of symbolism. The implications of these findings for drawing maps from a single digital database should be explored.

6. The word 'mapping' occurs in many contexts. Visit your preferred Internet book vendor (for example [www.amazon.com](http://www.amazon.com)) and do a search for all the books on offer that have the words 'map' or 'mapping' in their titles. How many different contexts can you find, and what does this tell you about geographic maps?
7. Fields are represented in many ways on a typical national-series topographic map. Take a typical map at a scale around 1:20,000 through to 1:50,000 and list and describe all the cartographic methods used. Your list might include, for example, spot heights, contours, hypsometric tinting in which height bands are shown in a graded series of colors, hill shading (shadowing), even hachuring. Having completed this exercise, speculate on how easily each method could be implemented from the six ways of representing a field shown in Figure 3.12
8. Use tracing paper to make a tracing from a map of a limited number of objects, preferably illustrating point, line, and area object classes (see exercise 6). Then overlay on this some transparent graph paper with a regular grid of square cells. Use this grid to hand code the map as a vector data structure and as a raster, both coded according to what is being recorded. This exercise can be done 'on screen', but there is value in using pencil and paper. Its point is to illustrate the nature of a digital caricature of the underlying geography.

In most critiques of GIS, one issue that repeatedly arises is the tendency of GIS to promote or privilege certain types of representation at the expense of alternatives. Organize two teams to debate the motion that 'This house believes that in representing geography in the ways they do, GIS promote a one-eyed view of the world'. A good reference is Schuurman, N. (2004) *GIS: a short introduction*, Oxford: Blackwell.

## FURTHER READING

- Chrisman N R 2002 *Exploring Geographic Information Systems* (2nd edn). New York: Wiley.
- Fisher P, Unwin D (2005) *Re-Presenting Geographical Information Systems*. Chichester: Wiley.
- McMaster R B, Shea K S 1992 *Generalization in Digital Cartography*. Washington, DC: Association of American Geographers.
- National Research Council 1999 *Distributed Geolibraries: Spatial Information Resources*. Washington, DC: National Academy Press. Available: [www.nap.edu](http://www.nap.edu)
- Gore A 1992 *Earth in the Balance: Ecology and the Human Spirit*. Boston: Houghton Mifflin
- Mardia K V, Jupp P E 2000 *Directional Statistics*. New York: Wiley
- Ryerson R A editor 1998 *Manual of Remote Sensing*. New York: Wiley.

## RELATED READING

Longley P A, Goodchild M F, Maguire D J, Rhind D W, editors 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, NJ: Wiley.

- Spatial representation: the scientist's perspective, J F Raper
- Spatial representation: the social scientist's perspective, D J Martin
- Spatial representation: a cognitive view, D M Mark
- Time in GIS and geographical databases, D J Peuquet
- Representation of terrain, M F Hutchinson, J C Gallant
- Digital remotely sensed data and their characteristics, M Barnsley

Maguire D J, Goodchild M F, Rhind D W editors 1991 *Geographical Information Systems: Principles and Applications*. Harlow, UK: Longman

Concepts of space and geographical data, A C Gatrell, pp. 119-34

- GIS and remote sensing, F W Davis and D S Simonett, pp. 191-213
- High-level spatial data structures for GIS, M J Egenhofer and J R Herring, pp. 227-37

## ONLINE RESOURCES

- ESRI Virtual Campus course, *Turning Data into Information* by Paul Longley, Michael Goodchild, David Maguire, and David Rhind ([training.esri.com](http://training.esri.com))
  - Module 1: Basics of Data and Information, Unit: Representing geography
- Section 3.1, Module 1: Basics of Data and Information  
Unit: Representing geography

- Section 3.3, Module 1: Basics of Data and Information  
Unit: Representing geography  
Sub-unit: How are geographic data represented?  
Sub-unit: Problems with representing geographic data
- Section 3.4, Module 1: Basics of Data and Information  
Unit: Representing geography  
Sub-unit: What are geographic data?
- Section 3.5, Module 1: Basics of Data and Information  
Unit: Representing geography  
Sub-unit: Discrete objects and fields  
Sub-unit: Comparing the discrete object and field views for lakes in Minnesota
- Section 3.6, Module 1: Basics of Data and Information  
Unit: Representing geography  
Sub-unit: Rasters and vectors
- NCGIA Core Curriculum in GIScience, 2000 ([www.ncgia.ucsb.edu/giscc](http://www.ncgia.ucsb.edu/giscc))
  - The World in Spatial Terms (005), ed. Reg Golledge
  - Human Cognition of the Spatial World (006), Dan Montello
  - 1.4.2. Maps as Representations of the World (020), Judy Olson
- NCGIA Core Curriculum in GIS, 1990 ([www.ncgia.ucsb.edu/pubs/core.html](http://www.ncgia.ucsb.edu/pubs/core.html))
  - 2. Maps and map analysis
  - Models of reality
  - 22. Object/layer debate
  - 73. GIS and spatial cognition



# 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

- 4.1 Introduction
- 4.2 The fundamental problem revisited
- 4.3 Spatial autocorrelation and scale
- 4.4 Spatial sampling
- 4.5 Distance decay
- 4.6 Measuring distance effects as spatial autocorrelation
- 4.7 Taming geographic monsters
- 4.8 Induction and deduction and how it all comes together

## 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<sup>th</sup> 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 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 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 *inductive* 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

### Technical Box 4.5 The Strange Story of the Lengths of Geographic Objects

- Introduces the method of measuring a line by counting the number of steps of a given length
- Using these measurements to determine fractal dimension is covered in the regular text following this box

### ESSAY TOPICS

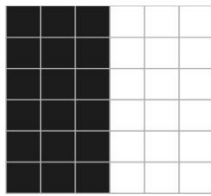
1. Why, and under what circumstances, do GI scientists sample?
2. Chapter 4 is mostly about how geography is represented in a digital computing environment, but what does 'representation' actually mean and what are the limits on what it can achieve?
3. In what ways is the advent of mass data storage in online environments (e.g. 'the Cloud') having an effect on representation?
4. If it is the case that many geographic objects have fractal characteristics, what are the consequences for how we might represent them in a GIS?
5. Section 4.2 develops the idea of spatial heterogeneity. How can this be seen to be in conflict with the Tobler Law (Section 3.1)?
6. What are some of the determinants of how phenomena are represented on maps?
7. Compare and contrast the representation problems given by data about 'where' with those about 'when'.
8. An argument has been made that how we represent geography in GIS can be improved by understanding human perception and cognition. In what ways do you agree or disagree with this idea?
9. Define what in science is meant by the words 'induction' and 'deduction' and give a reasoned account of why most geographic knowledge has been derived by induction.

### MULTIPLE CHOICE QUESTIONS (MCQ)

1. To a cartographer, is a 1:250,000 map conventionally referred to as large (A) or small (B) scale?
2. The Tobler Law is formalized by which of the following concepts?
  - a. spatial heterogeneity
  - b. fractal dimensions
  - c. spatial heterogeneity
  - d. spatial autocorrelation

3. Study the three 6 by 6 'chess boards':

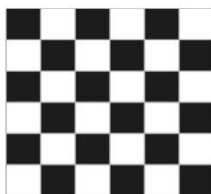
(a)



(b)



(c)



- a. Which of these illustrates positive spatial autocorrelation?
  - b. Which of these illustrates negative spatial autocorrelation?
4. Using the so-called 'Rook's Case' definition of adjacency in which only links along rows and down columns are allowed, how many black to black joins (BB) are there in (b). Is it 27, 0 or 6?
  5. Using the so-called 'Queens Case', in which diagonal links are also allowed, how many black to black joins (BB) are there in (b)? Is it 47, 14 or 25?
  6. A more than expected number of black to white joins indicates positive spatial autocorrelation. TRUE or FALSE?
  7. Of the following alternatives ranges, in which range is the typical fractal dimension of sinuous lines used to map coastlines such as that of Norway likely to lie?
    - a. Below 1.0
    - b. Between 1.5 and 1.9
    - c. Between 1.1 and 1.5
    - d. Above 2.0
  8. For each of the following real geographic features, write down the length dimension of the database object that would best represent them in a GIS developed for mapping at the 1:50,000 scale. (Answer L0, L1, L2 or L3)

- a. Public building
  - b. Highway
  - c. Housing
  - d. Woodland
  - e. Surface relief
  - f. Census tract
  - g. Railway
9. Choropleth maps should only be used to display ...?..... spatial variables.
- a. extensive
  - b. nominal
  - c. ordinal
  - d. intensive
10. Induction can be defined as .....?
- a. beginning with theory and principles as a basis for looking at data
  - b. a way of measuring spatial autocorrelation
  - c. reasoning from data to build up understanding
11. For each of the listed datasets, state which type of map would be most appropriate for its display (select isopleth, choropleth or neither of these):
- a. Earth surface relief
  - b. % of senior citizens in the population
  - c. height of the 500mb atmospheric pressure surface
  - d. land use
  - e. population density in people/km<sup>2</sup>
12. Which of the following is not one of the many meanings of the word 'scale'?
- a. extent
  - b. fraction
  - c. detail

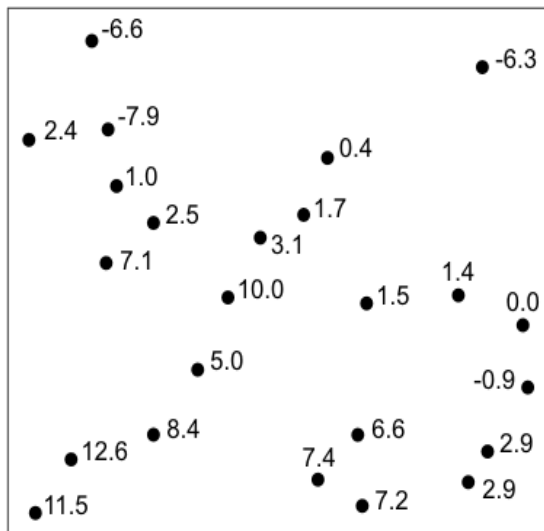
## CLASS AND INDIVIDUAL ACTIVITIES

1. Figure 4.4 presents seven possible point sampling schemes. Figure 3.13 showed a 1:24,000 USGS topographic map on which the wooded area is shown as a light green wash. Use at least three of the sampling designs (random, grid and cluster are suggested) estimate the actual wooded area. Each point sample can be recorded as 'wooded/not wooded' and the percentage of the total map area estimated in each case by the proportion of 'hits' on the woodland. Initially use a low number of points, but then double this and repeat the exercise. A simple way to generate a random



point sample is to lay transparent graph paper over the map and then use tables of random numbers for both the x and y co-ordinates of each point.

2. Many students assume that it is easy to create a continuous isopleth map (Technical Box 4.3) from a field of point attribute values. This exercise, from O'Sullivan and Unwin (2010, pages 251-253) will show them otherwise. It is also a useful introduction to automated contouring.



The point attributes are 'spot heights' of the average January temperature (°F) in a part of Alberta, Canada; your task is simple: Get a pencil (you may also need an eraser!) and produce a continuous surface representation of these data, by drawing 'contours' of equal temperature (in climatology these are called iso-therms). In threading these through the data, bear in mind three things. First, don't 'join the dots'. Remember that the data are unlikely to be exact and, even with a 0.1° resolution, each isotherm is likely to have substantial spatial width. In many applications it is wildly optimistic to assume that the data are exact. Second, experience suggests that it pays to start the process with a contour value in the middle of the data range and to work up and down from this value. Third, you should try to make the resulting surface of average temperatures as smooth as you can, consistent with it honoring all the data. By honoring the data we mean that the interpolated surface should pass through all the measured data exactly. In practice, this means that there should be no inconsistencies where measured temperatures lie on the 'wrong' side of relevant isotherms.

3. Establishing spatial relationships between zones. Figure 4.12 shows a simplified mosaic of zones, which has been used to generate Table 4.1, a matrix W of

adjacencies in which each element  $w_{ij}$  is coded '1' if the zones are adjacent and '0' otherwise. Find a small pattern of zones, such as the Standard Economic Regions of England and Wales and use it to create a similar matrix, noting that it is symmetrical. Next, develop a definition of adjacency that is a ratio-scaled number, such as the shared boundary length (measure this by stepping with dividers). Develop this further for each zone by re-expressing this as a proportion of the total boundary length of that zone, thus creating an asymmetric W matrix. O'Sullivan and Unwin (2010, pages 48-49 and 200-205 develop this further).

4. At only a slightly more advanced level, this exercise can be developed by replacing the adjacency relationship with a distance relation, as shown by Worboys, M.F., (1996) Metrics and Topologies for Geographic Space, in *Advances in Geographic Information Systems Research II: Proceedings of the International Symposium on Spatial Data Handling, Delft*, Kraak, M.J. and Molenaar, M. (eds.), Taylor & Francis, pp. 365-376. The original paper can be found at: [www.spatial.maine.edu/~worboys/mywebpapers/sdh1996.pdf](http://www.spatial.maine.edu/~worboys/mywebpapers/sdh1996.pdf)
5. Study Technical Box 4.3 and Figure 4.10 carefully. It should be clear that creating a choropleth map from 'raw' census data involves a series of decisions, any and all of which can affect the appearance of the final map. Create a simple 'workflow' in which you list each step in the process in its correct sequence, identifying at each step the range of choices available to you.
6. The best way to understand work in GIS using the concept of fractal objects and its implications for the measurement of objects that display this property is to compute the fractal dimension of an example, in this case a cartographic line using the Richardson plot. Find a reasonably detailed topographic map at a scale something like 1:24,000 or 1:50,000 and select a length of river as your object of study. Obviously, since we are interested in the sinuosity of linear objects it makes sense to choose a river that shows meandering behavior! Around a 20km length is about right and will not involve you in too much work. Now set a pair of dividers at a large equivalent distance on the ground, say 1km and 'walk' them along the river counting the number of steps. Record the yardstick length and number of segments as in example given in the lesson. Repeat using a halved yardstick, 500m; repeat again and again until the yardstick becomes so short that the experiment is impractical. You should now have a table of paired values. Ideally you should have 5 or 6 such pairs of measures. Plot a graph of the estimated length (on the vertical Y axis,

computed as the number of steps multiplied by the yardstick size) against yardstick size (on X) and comment on its shape. Convert both the numbers of steps and the yardstick lengths to their common logarithms and plot the resulting numbers with the  $\log(\text{number of steps})$  on the vertical axis and  $\log(\text{yardstick length})$  on the horizontal. If you have access to a spreadsheet you should be able to do this using that software. Hopefully, the points will fall roughly along a straight line but success isn't guaranteed.

Use the spreadsheet (or a straight edge and a good eye) to fit a linear regression line to your data and so estimate the fractal dimension of your river. What result did you get?

The regression equation can be used to estimate the fractal dimension. It is of the form:

$$\log[L(s)] = (1-D)\log(s) + b$$

In which  $Y = \log(L)$ , the total estimated length)

$b$  = the 'intercept constant', which isn't of concern

$D$  = the fractal dimension

$s$  = the step length (yardstick)

Note that since the length estimated  $L$  gets less the longer the length you set for ' $s$ ', the slope of the line, shown by Mandelbrot to be  $(1-D)$ , is a negative one.

7. Demonstrating a key theorem in statistics, Gold et al (1990, Figure 4.2) has a very old-fashioned computer program to simulate the generation of as many samples of selected size,  $n$ , as required from a uniform distribution in which every number has an equal chance of occurring. You would probably want to update the code, but the idea is to use it to demonstrate that repeated sampling from this distribution produces a sampling distribution of the mean that is both normal and peaked around the true population value. Computation of the standard deviations of these means will also give a value close to the standard error of the mean of each sample. Finally, the exercise illustrates that the gain in precision of the estimate rises as the square root of  $n$ . The basic idea is again that of Silk (1979).
8. Imagine you are the director of the market research group of a large supermarket chain, tasked to find a location for a new store in the suburbs of a large city. Prepare

a presentation for your company Board of Directors of the factors you have considered in locating the store and estimating the likely store turnover (See Section 4.4, pages 93-95)

## FURTHER READING

Batty M. and Longley P.A. 1994 *Fractal Cities: A Geometry of Form and Function*. London: Academic Press. Available for free download at [www.fractalcities.org](http://www.fractalcities.org).

Mandelbrot B.B. 1983 *The Fractal Geometry of Nature*. San Francisco: Freeman.

O'Sullivan D. and Unwin D. 2002 *Geographic Information Analysis*. Hoboken, NY: Wiley

Goodchild M F 1986 *Spatial Autocorrelation*. Concepts and Techniques in Modern Geography 47. Norwich: GeoBooks

Goodchild M F 2001 Models of scale and scales of modelling. In Tate N J, Atkinson P M (eds) *Modelling scale in geographical information science*. Chichester: Wiley

Quattrochi D.A. and Goodchild M.F. (eds) 1996 *Scale in Remote Sensing and GIS*. Boca Raton, Florida: Lewis Publishers.

Silk, J.A. 1979 The use of classroom experiments and the computer to illustrate statistical concepts. *Journal of Geography in Higher Education*, 3:13-25

Tate N.J. and Atkinson P.M. (eds) 2001 *Modelling Scale in Geographical Information Science*. Chichester: Wiley.

Wright D. and Bartlett D. (eds) 2000 *Marine and Coastal Geographical Information Systems*. London: Taylor and Francis.

Wright D. 2002 *Undersea with GIS*. Redlands, CA: ESRI Press.

## RELATED READING

Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, NJ: Wiley.

2. Space, time, geography, H Couclelis

16. Spatial statistics, A Getis

17. Interactive techniques and exploratory spatial data analysis, L Anselin

Maguire D.J., Goodchild M.F. and Rhind D.W. (eds) 1991 *Geographical Information Systems: Principles and Applications*. Harlow, UK: Longman (text available online at [www.wiley.co.uk/gis/volumes.html](http://www.wiley.co.uk/gis/volumes.html)).

9. Concepts of space and geographical data, A C Gatrell, pp. 119-34

30 Generalization of spatial databases, J-C Muller

## ONLINE RESOURCES

Fractal Cities - <http://www.fractalcities.org/>

ESRI Virtual Campus course, *Turning Data into Information* by Paul Longley, Michael Goodchild, David Maguire, and David Rhind ([training.esri.com](http://training.esri.com) )

Module 1: Basics of Data and Information

Module 4: Transformations and Descriptive Summaries

Module 5: Optimization and Hypothesis Testing

Module 6: Uncertainty

Section 4.2, Module 1: Basics of Data and Information,

Unit: The nature of geographic data

Sub-unit: Spatial autocorrelation

Section 4.4, Module 1: Basics of Data and Information,

Unit: The nature of geographic data

Sub-unit: Spatial sampling

Module 5: Optimization and Hypothesis Testing,

Unit: Hypothesis testing

Sub-unit: Sampling

Module 6: Uncertainty

Unit: Measuring uncertainty of nominal and ordinal values

Sub-unit: Spatial sampling

Sub-unit: Difficulties in sampling natural areas

Section 4.5, Module 4: Transformations and Descriptive Summaries

Unit: Spatial interpolation and density estimation

Sub-unit: What is spatial interpolation?

Section 4.6, Module 4: Transformations and Descriptive Summaries

Unit: Spatial dependence and fragmentation

NCGIA Core Curriculum in GIScience, 2000 ([www.ncgia.ucsb.edu/giscc](http://www.ncgia.ucsb.edu/giscc))

. 1.6.1. [Sampling the World](#) (031)

NCGIA Core Curriculum in GIS, 1990 ([www.ncgia.ucsb.edu/pubs/core.html](http://www.ncgia.ucsb.edu/pubs/core.html))

6. Sampling the world

47. Fractals

# Georeferencing

---

## OVERVIEW

- Lays out the principles of georeferencing
- Discusses commonly used systems, including placenames and street addresses, and moving to the more accurate scientific methods that form the basis of geodesy and surveying.
- Considers conversions between georeferencing systems, GPS, georeferencing of computers and cellphones, and gazetteers.

## LEARNING OBJECTIVES

- **Know the requirements for an effective system of georeferencing;**
- **Be familiar with the problems associated with placenames, street addresses, and other systems used every day by humans;**
- **Know how the Earth is measured and modeled for the purposes of positioning;**
- **Know the basic principles of map projections, and the details of some commonly used projections;**
- **Know about conversion between different systems of georeferencing**
- **Understand the principles behind GPS, and some of its applications.**

## KEY WORDS AND CONCEPTS

Georeference, placenames, linear referencing systems, cadaster, PLSS, latitude, longitude, spheroid, ellipsoid, WGS84, NAD83, NAD27, great circle, projections, easting, northing, Cartesian coordinates, conformal property, equal-area property, tangent projections, secant projections, Plate Carrée, UTM, State Plane Coordinates, GPS

## OUTLINE

- 5.1 Introduction
- 5.2 Placenames
- 5.3 Postal addresses and postal codes
- 5.4 IP addresses
- 5.5 Linear referencing systems
- 5.6 Cadasters and the US Public Land Survey System
- 5.7 Measuring the Earth: latitude and longitude
- 5.8 Projections and coordinates
- 5.9 Measuring latitude, longitude, and elevation: GPS
- 5.10 Converting georeferences
- 5.11 Geotagging and mashups
- 5.12 Georegistration
- 5.13 Summary

## CHAPTER SUMMARY

### 5.1 Introduction

Several terms are used to describe the act of assigning locations including georeference, geolocate, geocode, or tag with location.

- Authors use *georeference* in this chapter

Lists several characteristics that georeferences must have

Table 5.1 Some commonly used systems of georeferencing - summarizes on dimensions of domain of uniqueness, metric, and spatial resolution.

### 5.2 Placenames

- Any distinctive feature on the landscape can serve as a point of reference, and these are often named
- Language extends the power of placenames through words like 'between' or 'near' or by the addition of directions and distances
- Placenames are limited as georeferences

### 5.3 Postal addresses and postal codes

Postal addresses rely on several assumptions

- Postal addresses fail in locating anything that is not a potential destination for mail, such as a natural feature

- Canadian FSAs, US ZIP code areas and UK postcodes can be changed whenever the post office decides. However, they are sufficiently constant to be used for many purposes

## 5.4 IP Addresses

- Every device (computer, printer, etc.) connected to the Internet has a unique IP (Internet Protocol) address
- The IP address of the user's computer is provided whenever the computer is used to access a Web site, allowing the operators of major sites to determine the user's location

## 5.5 Linear referencing systems

- A linear referencing system identifies location on a network by measuring distance from a defined point of reference along a defined path in the network
- Linear referencing systems are widely used in managing transportation infrastructure and in dealing with emergencies
- However, there may be difficulties in defining distances along the network accurately, especially if roads include steep sections where horizontal distance is different from three-dimensional distance

## 5.6 Cadasters and the U.S. Public Land Survey System

- The *cadaster* is defined as the map of land ownership in an area, maintained for the purposes of taxing land, or of creating a public record of ownership.
- The US Public Land Survey System (PLSS) defines land ownership over much of western North America, and is a useful system of georeferencing.
  - The structure of the PLSS is explained in this section

## 5.7 Measuring the Earth: latitude and longitude

- The system of latitude and longitude is often called the *geographic* system of coordinates
- Explains how latitude and longitude are determined and defines various terms such as axis of Earth's rotation, the Equator, meridian, parallel
- Before defining latitude further, this section briefly touches on the topics of the figure of the Earth, spheroid, ellipsoid, flattening of the spheroid
  - The distance between the Poles is about 1 part in 300 less than the diameter at the Equator



- WGS84, NAD83, and NAD27 are mentioned
- Latitude is often symbolized by the Greek letter phi ( $\varphi$ ) and longitude by the Greek letter lambda ( $\lambda$ )
- Some statistics
  - two points on the same north–south line of longitude and separated by one degree of latitude are 1/360 of the circumference of the Earth, or about 111 km, apart
  - One minute of latitude corresponds to 1.86 km, and also defines one nautical mile
  - One second of latitude corresponds to about 30 m.
  - One degree of longitude along 60 degrees North is 55 km
- On a spherical Earth the shortest path between two points is a *great circle*, or the arc formed if the Earth is sliced through the two points and through its center

*Some trivia: the French originally defined the meter in the late 18th century as one ten millionth of the distance from the Equator to the Pole. On the spherical Earth, the actual value Equator to Pole would be 10,018 km.*

## 5.8 Projections and coordinates

- Projections are needed because many technologies for working with geographic data are flat, including paper and printing, which evolved over many centuries long before the advent of digital geographic data and GIS.
- The Cartesian coordinate system assigns two coordinates to every point on a flat surface, by measuring distances from an origin parallel to two axes drawn at right angles.
- We often talk of the two axes as  $x$  and  $y$ , and of the associated coordinates as the  $x$  and  $y$  coordinate, respectively.
- Because it is common to align the  $y$  axis with North in geographic applications, the coordinates of a projection on a flat sheet are often termed *easting* and *northing*.
- A map projection transforms a position on the Earth's surface identified by latitude and longitude ( $\varphi, \lambda$ ) into a position in Cartesian coordinates ( $x, y$ ).
- Every recognized map projection can be represented as a pair of mathematical functions:
 
$$x = f(\varphi, \lambda)$$

$$y = g(\varphi, \lambda)$$
- Two datasets can differ in both the projection and the datum, so it is important to know both for every dataset.

- Two important properties of projections are
  - The conformal property ensures that the shapes of small features on the Earth's surface are preserved on the projection
  - Useful for navigation because a straight line has a constant bearing
- The equal area property ensures that areas measured on the map are always in the same proportion to areas measured on the Earth's surface.
- Useful for analyses involving areas
- Major classes of physical models used are cylindrical, azimuthal, and conic
- Where the developable surface coincides with the surface of the Earth, the scale of the projection is 1
  - Tangent projections are where the developable surface coincides with the surface of the Earth along only one line (or at one point)
  - Secant projections attempt to minimize distortion by allowing the paper to cut through the surface, so that scale can be both greater and less than 1
- The grid of latitude and longitude is known as the *graticule*

#### 5.8.1 The Plate Carrée or Cylindrical Equidistant projection

- The simplest of all projections maps longitude as x and latitude as y, and for that reason is also known informally as the unprojected projection.
- Discusses the serious problems that can occur when doing analysis using this projection and talks about the Tissot indicatrix (without using that term)

#### 5.8.2 The Universal Transverse Mercator projection

- Based on the Mercator projection, but in transverse rather than Equatorial aspect,
- There are 60 zones in the system, each zone being 6 degrees wide
- The UTM system is secant, with lines of scale 1 located some distance out on both sides of the central meridian.
- Scale is 0.9996 at the central meridian and at most 1.0004 at the edges of the zone
- The system for defining the coordinates in meters measured from the Equator and central meridian is described
- Mentions the problem that maps will not fit together across a zone boundary
- Are good for analysis because distances can be calculated for points within the same zone with no more than 0.04% error

### 5.8.3 State Plane Coordinates and other local systems

- In the 1930s each US state agreed to adopt its own projection and coordinate system, generally known as State Plane Coordinates (SPC), in order to support high-accuracy applications.
- The UK uses a single projection and coordinate system known as the National Grid that is based on the Oblique Mercator projection.

## 5.9 Measuring latitude, longitude and elevation: GPS

- Very briefly explains the constellation of satellites
- Positioning in three dimensions (latitude, longitude, and elevation) requires that at least four satellites are above the horizon
- If elevation is not needed then only three are needed
- Simple GPS can get horizontal accuracies within 10m, differential GPS can improve it to 1m or better
  - Vertical accuracies are much poorer
- A variety of vertical datums are used, even within single countries.

## 5.10 Converting georeferences

### Technical Box 5.3 Geocoding: conversion of street addresses

Requires a database containing records representing the geometry of street segments between consecutive intersections, and the address ranges on each side of each segment

## 5.10 Summary

### ESSAY TOPICS

1. From Lewis Carroll's *Hunting of the Snark*:  
*'What's the good of Mercator's North Poles and Equators,  
Tropics, Zones, and Meridian Lines?'*  
*So the Bellman would cry: and the crew would reply  
They are merely conventional signs!'*
  - a. How would you answer the Bellman and crew?
  - b. What are the desirable properties of a georeference?
2. What the practical problems associated with the use of latitude and longitude as a global georeference?
3. What are the deficiencies of postal codes when they are used as georeferences?

4. What are the deficiencies of IP addresses when they are used as georeferences?
5. What features of some linear referencing systems allow geographical analysis, and what are the limitations of these systems?
6. Summarize the arguments for and against use of a single ellipsoid such as WGS84.
7. How would you go about identifying the projection used by a common map source, such as the weather maps shown by a TV station or in a newspaper?
8. Identify the map projections that would be best for measurement of (1) area, (2) length, (3) shape.
9. To what extent does the use of unusual projections present cartographers with another 'visual variable'?

### MULTIPLE CHOICE QUESTIONS (MCQ)

1. Eastings in map co-ordinates are conventionally associated with which axis of a Cartesian system, x or y?
2. What is the value of the polar flattening of the Earth relative to a perfect sphere? Is it 1 part in (a) 500, (b) 300 or (c) 200?
3. Which of the following is NOT an essential requirement of a georeference?
  - a. Uniqueness
  - b. Shared meaning
  - c. Persistence
  - d. A time and date
10. Which of the following are not metric georeferences?
  - a. Post/ZIP codes
  - b. Latitude/longitude
  - c. Postal address
  - d. House number
  - e. State plane co-ordinates
  - f. National grid reference
11. What are the two most important characteristics of a geographical data set that must be known before any analysis?
  - a. the datum
  - b. the scale
  - c. the projection
  - d. the key
12. Complete the following sentence by choosing from the bracketed alternatives. "UTM has (180/100/60/10) zones each of 3/6/12 degrees width; from 174E to 180E which is Zone

number (1/30/60); from 180W to 174W which is Zone number (1/30/60) The scale at the central meridian is (1.004/0.9996) and at most (1.0004/0.9996) at the zone edges”.

13. A cadaster is

- a. a land subdivision
- b. the boundary of some property
- c. part of the Public Land Survey System
- d. a map of land ownership

14. Match each of the places in the table to its latitude and longitude (rounded to the nearest integer):

City		Latitude	Longitude
Narsarssuaq, Greenland		51°N	0°W
Hong Kong		49°N	2°E
Paris, France		61°N	45°W
London, England		37°S	175°E
Auckland, New Zealand		22°N	114°E

15. In the US PLSS a township has what area?

- a. 40 square miles
- b. 36 square miles
- c. 1 square mile

16. Which of the following ‘surfaces’ does not define a class of map projections?

- a. sphere
- b. cylinder
- c. plane
- d. cone

## CLASS AND INDIVIDUAL ACTIVITIES

1. Where do you live? Ask a class to take five minutes to write down how they would answer someone who asks ‘where do you live?’ Compare the responses to the methods listed in Table 5.1. Two simple additions to this exercise are (a) to use a WWW location-finding utility such as in UK the one at [www.streetmap.co.uk](http://www.streetmap.co.uk) to perform searches using each and every georeference the system uses (street name, telephone code, OS (x,y), post code, place name, and latitude/longitude), and (b) qualify the question by some imagined location, such as overseas, in your country, and in your locality. The first will demonstrate that the same place can be georeferenced in many ways, whilst the second shows that the one used is frequently context dependent.

2. Visit your local map library, and determine: (1) the projections and datums used by selected maps; and (2) the coordinates of your house in several common georeferencing systems
3. The advent of GPS has revolutionized the geo-location process, but using the system is not without its problems. Even simple, low-cost hand-held units without differential correction can be used by students in projects that collectively will give them a good idea of the technology and its limitations. Most GPS either come with, or have as an optional extra, a cable that connects to either the USB Serial or COMs port of a PC. For some useful shareware to up- and down-load data to and from most easily obtainable GPS see <http://www.gpsu.com>.
  - a) Track a location over time.
  - b) Log a self-evidently 'straight' line
  - c) Develop a database of where places are in your locality.
  - d) Visit <http://confluence.org> to catch up with some very, curious folk who want to visit and photograph every point of intersection of whole number lines of latitude and longitude. Actually, as a record of the landscape at the beginning of the century, there is some purpose to the project
  - e) Visit <http://www.gpsdrawing.com/gallery/land/oxdon.htm> for some real fun. This is for people who have nothing better to do than use various forms of transport and their GPS to trace out on the ground huge drawings of the outlines of familiar objects. One example, chosen semi-randomly, is the outline of the face of an Oxford Don, 'drawn' over a length of over 100km and superimposed on an aerial photograph.
  - f) Make your own maps. It's possible to make crude maps, using a hand-held GPS as a rough surveying instrument but, for anything other than a fairly large piece of terrain, the inaccuracies might well be embarrassing.
4. In addition to the examples given in the chapter, there are projections that vary locally in scale (sometimes called magnifying-glass projections because the effect is similar to running a magnifying glass across the mapped area). For a good example see: [Fairbairn, D. and Taylor, G.](#) (1995) Developing a variable scale map projection for urban areas. *Computers & Geosciences*, 21(9): 1053-1064.

5. It is also possible to reference on the sphere by recursive subdivision using a shape that nests onto it, so that the spatial location is a list of the subdivisions in which the point resides. In his Quaternary Triangulated Mesh (QTM) system, the American cartographer Geoff Dutton has developed an approach that does just this. It is useful to visit his website to examine the principles involved, which is at <http://www.spatial-effects.com>. Having visited the site, make a list of the 'pros' and 'cons' of his method.
6. A very good way to introduce the distortion properties of map projections is to use the computer to examine the graticules (grids of parallels and meridians), continental outlines, and, if the software used allows, the Tissot Indicatrix for a variety of projections. A simple 'upside-down' Mercator projection, or moving the central meridian from the Atlantic to somewhere else, radically alters one's perception of the planet. Software to support such an exercise can be found in many GIS.
7. During the preparation of this supplement, one of the team flew from London England round the world by way of Los Angeles (CA), Auckland (New Zealand), Sydney, Perth (Australia) and Singapore. Assuming that on every leg of this journey the aircraft took the shortest route, how far was the journey? This can be done using the usual formula,

$$Distance = R \arccos[\sin \varphi_1 \sin \varphi_2 + \cos \varphi_1 \cos \varphi_2 \cos(\lambda_1 - \lambda_2)]$$

in which  $\varphi_1$  and  $\varphi_2$  are the latitudes and  $\lambda_1$  and  $\lambda_2$  the longitudes.  $R$  is an accepted value for the Earth radius such as 6378km. You should get a distance of around 39,507nm, which is, of course, considerably more than the equatorial circumference of the Earth. Many globetrotting travelers keep to the service networks of airline alliances, in order to accrue frequent-traveler benefits. Would our author have been able to travel more nautical miles with members of the Oneworld Alliance ([www.oneworld.com](http://www.oneworld.com), including American Airlines, British Airways and Qantas) or the Star Alliance ([www.staralliance.com](http://www.staralliance.com), including Air New Zealand, Singapore Airlines, and United Airlines), assuming he minimized any distance traveled on non-alliance airlines?

8. Most metric coordinate systems use a Euclidean distance as their basis. Such distances have five key properties:

They are *non-negative*. You cannot have a minus distance.

If they are zero, then the distance from a place to itself is zero

There is one and only one *shortest* distance

They are symmetrical: the distance from A to B is the same as that from B to A.

They obey the so-called *triangle inequality*. This says that if we imagine a triangle with vertices labeled A, B, and C then one side (say AC) will be no longer than the sum of the other two.

Spend a few moments thinking about other 'distances' that might be used instead, and then search your map library and the literature for examples of maps based on other sorts of 'distance' such as cost and time. The book by Anthony Gatrell (1983) *Distance and Space: a geographical perspective* (Oxford: OUP) is a good place to start. Is it legitimate to call these graphics 'maps'?

9. A debate/seminar. Ask the house to debate the motion that 'Euclidean geometry is not a good candidate for representing geographic information, since it relies on the existence of complete co-ordinate n-tuples' (Egenhofer and Mark, 1995). Arguments in favor of the motion will be found in the original paper on what is called 'Naïve Geography', written for a meeting of the COSIT group by Max Egenhofer and David Mark, published in: A. Frank and W. Kuhn (eds.), *Lecture Notes in Computer Science*, Vol. 988, Springer-Verlag, pp. 1-15. The same paper can be downloaded from <http://www.spatial.maine.edu/~max/RC20.html>. Arguments against the motion can be found in almost all of the materials in Chapter 4. This exercise is a very good way to introduce students to ideas at the research frontier in GIScience.
10. Chapter 5 is all about georeferencing using placenames, various post codes, linear systems, lat/longitude and then coordinates in some projection. A useful thought exercise is to match these to the equivalents that might be used to locate ourselves in time (a tempo-reference?).

## FURTHER READING

Bugayevskiy L M, Snyder J P 1995 *Map Projections: A Reference Manual*. London: Taylor and Francis.

A profusely illustrated catalog.

Burrough P A, Frank A U editors 1996 *Geographic Objects with Indeterminate Boundaries*. London: Taylor and Francis ('dip into' all chapters)

Fotheringham A S, Wong D W S 1991 The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 23: 1025-1044

Keay J 2001 *Great Arc: The Dramatic Tale of How India was Mapped and Everest was Named*. New York: Harper Collins



Tells the heroic, but bizarre, tale of how William Lambton set out to determine the shape of the Earth. Mapping the sub-continent was a side issue! Reading this alongside Dava Sobell's better known book (q.v.) should be compulsory for all those who habitually use latitude and longitude georeferences!

Kennedy M 1996 *The Global Positioning System and GIS: An Introduction*. Chelsea, Michigan: Ann Arbor Press.

Maling D H 1992 *Coordinate Systems and Map Projections* (2nd edn). Oxford: Pergamon.

This is very much the standard work.

Sobel D 1995 *Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time*. New York: Walker.

A surprise best seller that led to several TV dramatizations, and that tells the story of Harrison and his chronometers.

Snyder J P 1997 *Flattening the Earth: Two Thousand Years of Map Projections*. Chicago: University of Chicago Press.

Steede-Terry K 2000 *Integrating GIS and the Global Positioning System*. Redlands, CA: ESRI Press.

## RELATED READING

Longley P A, Goodchild M F, Maguire D J, Rhind D W (eds) 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition).

Hoboken, NJ: Wiley.

30. Spatial referencing and coordinate systems, H Seeger

Maguire D J, Goodchild M F, Rhind D W (eds) 1991 *Geographical Information Systems: Principles and Applications*. Harlow, UK: Longman (text available online at

[www.wiley.co.uk/gis/volumes.html](http://www.wiley.co.uk/gis/volumes.html)).

10. Coordinate systems and map projections for GIS, D H Maling, pp. 135-46

## ONLINE RESOURCES

ESRI Virtual Campus course, *Turning Data into Information* by Paul Longley, Michael Goodchild, David Maguire, and David Rhind ([training.esri.com](http://training.esri.com))

Module 1: Basics of Data and Information, Unit: The nature of geographic data

Section 5.5, Module 1: Basics of Data and Information,

Unit: The nature of geographic data,

Sub-unit: Distance decay and spatial interpolation

NCGIA Core Curriculum in GIScience, 2000 ([www.ncgia.ucsb.edu/giscc](http://www.ncgia.ucsb.edu/giscc))

- 1.3. [Position on the earth](#) (012), ed. *Ken Foote*
- 1.3.1. [Coordinate Systems Overview](#) (013), *Peter Dana*
- 1.3.2. [Latitude and Longitude](#) (014), *Anthony Kirvan*
- 1.3.3. [The Shape of the Earth](#) (015), *Peter Dana*
- 1.3.4. [Discrete Georeferencing](#) (016), *David Cowen*
- 1.3.5. [Global Positioning Systems Overview](#) (017), *Peter Dana*
- 1.4.1. [Projections and transformations](#) (019)
- 1.4.2. [Maps as Representations of the World](#) (020), *Judy Olson*

NCGIA Core Curriculum in GIS, 1990 ([www.ncgia.ucsb.edu/pubs/core.html](http://www.ncgia.ucsb.edu/pubs/core.html))

- 26. Common coordinate systems
- 27. Map projections
- 29. Discrete georeferencing

## OVERVIEW

- Uncertainty in geographic representation arises because almost all representations of the world are incomplete.
- This chapter identifies many of the sources of geographic uncertainty and the ways in which they operate in GIS-based representations.
- Uncertainty arises from the way that GIS users conceive of the world, how they measure and represent it, and how they analyze their representations of it.
- This chapter investigates a number of conceptual issues in the creation and management of uncertainty, before reviewing the ways in which it may be measured using statistical and other methods.
- The propagation of uncertainty through geographical analysis is considered.

## LEARNING OBJECTIVES

- **Understand the concept of uncertainty, and the ways in which it arises from imperfect representation of geographic phenomena;**
- **Be aware of the uncertainties introduced in the three stages (conception, measurement and representation, and analysis) of database creation and use;**
- **Understand the concepts of vagueness and ambiguity, and the uncertainties arising from the definition of key GIS attributes;**
- **Understand how and why scale of geographic measurement and analysis can both create and propagate uncertainty**

## KEY WORDS AND CONCEPTS

Uncertainty, error, ambiguity, vagueness, data quality, fuzzy membership, accuracy, precision, RMSE, Gaussian or Normal distribution, error propagation, aggregation, simulation, ecological fallacy, concatenation, conflation, induction, deduction, MAUP

## OUTLINE

- 6.1 Introduction
- 6.2 U1: Uncertainty in the conception of geographic phenomena
- 6.3 U2: Further uncertainty in the measurement and representation of geographic phenomena
- 6.4 U3: Further uncertainty in the analysis of geographic phenomena
- 6.5 Consolidation

## CHAPTER SUMMARY

This chapter is very dense in places. Many students will need extra mentoring to make it through the thorough descriptions of the effect of error and the implications of various reported error measurements. Alternatively, instructors of introductory courses may choose to use the 'U1, U2, U3' schema in order to provide a selective overview of the principal issues.

### 6.1 Introduction

- This chapter uses the term *uncertainty* as an umbrella term to describe the problems that arise out of our incomplete representations of the world.
- Various terms are used to describe differences between the real world and how it appears in a GIS
- The established scientific notion of measurement *error* focuses on differences between observers or between measuring instruments.
- *Ambiguity* and *vagueness* identify further considerations which need to be taken into account in assessing the *quality* of a GIS representation.
- The US Federal Geographic Data Committee's various standards list five components of quality: attribute accuracy, positional accuracy, logical consistency, completeness, and lineage.

- Uncertainty may thus be defined as a measure of the user's understanding of the difference between the contents of a dataset, and the real phenomena that the data are believed to represent.
- In GIS, the term uncertainty has come to be used as the catch-all term to describe situations in which the digital representation is simply incomplete, and as a measure of the general quality of the representation.
- The chapter structures the discussion of uncertainty through a consideration of the chain of events in which *conception* prescribes *measurement and representation*, which in turn prescribes *analysis*. This is summarized in Figure 6.1.

## 6.2 U1: Uncertainty in the conception of geographic phenomena

A characteristic that sets geographic information science apart from most every other science is that it is only rarely founded upon *natural* units of analysis.

### 6.2.2 Conceptions of attributes: Vagueness and ambiguity

#### 6.2.2.1 Vagueness

- Given the lack of natural units of analysis, we often transform point-like events into area objects. This leads to two important questions
  - Is the defining boundary of a zone crisp and well-defined?
  - Is our assignment of a particular label to a given zone robust and defensible?
- Thus, uncertainty can exist both in the positions of the boundaries of a zone and in its attributes.
- The questions have statistical implications, cartographic implications, and cognitive implications.
- Box 6.1 introduces school catchments as functional zones

#### 6.2.2.2 Ambiguity

- Many linguistic terms used to convey geographic information are inherently ambiguous.
- Many objects are assigned different labels by different national or cultural groups, and such groups perceive space differently.
- Object names and the topological relations between them may thus be inherently *ambiguous*.
- GIS cannot present a value-neutral view of the world, yet it can provide a formal framework for the reconciliation of different worldviews
- Ambiguity is introduced when imperfect *indicators* of phenomena are used instead of the phenomena themselves.

- *Direct* indicators are deemed to bear a clear correspondence with a mapped phenomenon.
- *Indirect* indicators are used when the best available measure is a perceived surrogate link with the phenomenon of interest.
- Conception of the linkage between any indicator and the phenomenon of interest is subjective, hence ambiguous.
- Our ability to generalize about spatial distributions is constrained by the different taxonomies that are conceived and used by data-collecting organizations within our overall study area.
- How may mismatches between the categories of different classification schema be reconciled?
- The process of reconciling the semantics of different classification schema is an inherently *ambiguous* procedure

#### **Applications Box 6.2 Vagueness, ambiguity, and the geographies of family names**

- Provides an interesting discussion of how surnames can be used as indicators of regional identity and diversity

#### **6.2.3 Fuzzy approaches to attribute classification**

- *Frequentist* approaches to assigning values to areas are based on the notion that the probability of a given outcome can be defined as the proportion of times the outcome occurs in some real or imagined experiment, when the number of tests is very large.
- However, in the geographic situation, there is only one field with precisely these characteristics, and one observer
- The *subjectivist* conception of probability represents a judgment about relative likelihood of a single occurrence and is best illustrated through the concept of *fuzzy membership*
- One of the major attractions of fuzzy sets is that they appear to let us deal with sets that are not precisely defined, and for which it is impossible to establish membership cleanly.
- Box 6.2 (Fuzziness in classification: description of a soil class) shows a typical extract from the legend of a soil map, including frequent use of terms such as ‘very’, ‘moderate’, ‘about’, ‘typically’, and ‘some’.
  - Figure 6.8 shows an example of mapping classes using fuzzy methods. The final map shows how these can be converted to crisp categories
- Researchers have struggled with the question of whether fuzzy methods are more *accurate*.

- If we are uncertain about which class to choose then it is more accurate to say so, in the form of a fuzzy membership, than to be forced into assigning a class without qualification.

## 6.3 U2: Further uncertainty in the representation of geographic phenomena

### 6.3.1 Representation of place / location

- The conceptual models (fields and objects) impose very different filters upon reality, and their usual corresponding representational models (raster and vector) are characterized by different uncertainties
- The vector model requires *a priori* conceptualization of the nature and extent of geographic individuals and the ways in which they nest together into higher-order zones.
  - In the vector model, point-like objects often appear only as aggregate counts for apparently *uniform* zones.
- The raster model defines individual elements as square cells, with boundaries that bear no relationship at all to natural features
- Discusses mapping coastline as a field
- Introduces the concept of *mixel*, a pixel whose area is divided among more than one class

### 6.3.2 Statistical models of uncertainty in attribute measures

A geographic database is a collection of measurements of phenomena on or near the Earth's surface

#### 6.3.2.1 Nominal case

- Describes the structure and interpretation of the *confusion matrix*
- Provides the equation for the *kappa index*
- Identifies some problems with this method
- Notes that in vector area model cases, error has two forms: misallocation of an area's class and the mislocation of an area's boundary

#### 6.3.2.2. Interval/ratio case

- Here, error is best thought of not as a change of class, but as a change of value such that the observed value  $x$  is equal to the true value  $x$  plus some distortion  $\delta x$ .

- If the average distortion is zero, with positive and negative errors balanced out, the observed values are said to be *unbiased*
- Distinguishes between *accuracy*, which has to do with the magnitude of the error, and *precision* which is defined in two ways
  - The variability among repeated measurements
  - The number of digits used to report a measurement (see Technical Box 6.4 that summarizes rules that are used to ensure reported measurements do not mislead)
- Discusses in detail the calculation and relevance of RMSE
- Explains the structure and interpretation of the Gaussian (or Normal) probability distribution

### 6.3.3 Statistical models of uncertainty in location measures

- This section is particularly detailed in its examination of the implications of positional error in our spatial databases. Many students will need extra mentoring to understand this section.
- A two-dimensional measured position (x,y) is subject to errors in both x and y
- In three dimensions, we expect the RMSEs of x and y to be the same, but z is often subject to errors of quite different magnitude.
- National Map Accuracy Standards often prescribe the positional errors that are allowed in databases.
- The 1947 US National Map Accuracy Standard specified that 95% of errors should fall below 1/30 inch (0.85 mm) for maps at scales of 1:20,000 and finer (more detailed), and 1/50 inch (0.51 mm) for other maps
- A useful rule of thumb is that features on maps are positioned to an accuracy of about 0.5mm. Table 6.2 shows the corresponding distance on the ground at different scales

## 6.4 U3: Further uncertainty in the analysis of geographic phenomena

### 6.4.1 Internal and external validation through spatial analysis

- *Internal* validation can be achieved through simulation of different possible outcomes (i.e. error propagation)
- *External* validation can be achieved by merging diverse data sources



#### 6.4.2 Validation through autocorrelation: The spatial structure of errors

- Error propagation measures the impacts of uncertainty in data on the results of GIS operations.
- There are two strategies available for evaluating error propagation
  - Obtain a complete description of error effects based upon known measures of likely error. This is discussed in detail by the use of some examples
  - Simulate the impacts of uncertainty on results which requires the generation of a series of realizations

#### 6.4.3 Validation through investigating the effects of aggregation and scale

- The measurement of geographic individuals is unlikely to be determined with the end point of particular spatial analysis applications in mind.
- As a consequence, we cannot be certain in ascribing even dominant characteristics of areas to true individuals or point locations *in* those areas.
- This source of uncertainty is known as the *ecological fallacy* (inappropriate inference from aggregate data about the characteristics of individuals)
- Gives rise to the related *aggregation* or *zonation* problem, in which different combinations of a given number of geographic individuals into coarser-scale areal units can yield widely different results.
- The effects of scale and aggregation are generally known as the *Modifiable Areal Unit Problem* (MAUP).

#### 6.4.4 Validation with reference to external sources: Data integration and shared lineage

- *Concatenation* is used to describe the integration of two or more different data sources, such that the contents of each are accessible in the product.
- *Conflation* attempts to replace two or more versions of the same information with a single version that reflects the pooling, or weighted averaging, of the sources.
- Yet such different datasets are likely to have been collected at a range of different scales and for a range of areal units
- Established procedures of statistical inference can only be used to reason from representative samples to the populations from which they were drawn.

#### 6.4.5 Internal and external validation; induction and deduction

- This section provides several areas of caution that need to be considered
  - The Modifiable Areal Unit Problem can be investigated through simulation of large numbers of alternative zoning schemes.

- However, zone design experiments are merely playing with the MAUP, and most of the new sources of external validation are unlikely to sustain full scientific scrutiny, particularly if they were assembled through non-rigorous survey designs.
- In measuring the distribution of all possible zonally averaged outcomes, there is no tenable analogy with the established procedures of statistical inference and its concepts of precision and error.
- The way forward seems to be to complement our new-found abilities to customize zoning schemes in GIS with external validation of data and clearer application-centered thinking about the likely degree of within-zone heterogeneity that is concealed in our aggregated data.
- Notes that within the socio-economic realm, the act of defining zones can also be self-validating if the allocation of individuals affects the interventions they receive

## 6.5 Consolidation

- Briefly lists the key points made
- Gives some rules for how to live with uncertainty
  - Acknowledge that uncertainty is inevitable
  - Assemble all that is known about the quality of data and use this to assess whether the data are fit for use
  - Gain some impression of the impacts of input uncertainty on outputs
  - Rely on multiple sources of data
  - Be honest and informative in report the results of GIS analysis

## ESSAY TOPICS

1. Why are error and uncertainty in the results of a GIS-based analysis not the same thing?
2. Review the ways by which continuous fields can be represented in a GIS.
3. Giving specific examples, explain what is meant by the term 'ecological fallacy', how it arises and why it can lead to false conclusions.
4. In assembling objects such as trees into area objects such as a 'forest' what are the major characteristics that would be required of these areas? (see Section 6.2.2.1)
5. Outline and contrast the available methods for evaluating error in 'field' data with those for object-based representations.
6. Explain why soil classes are archetypal examples of fuzzy objects with uncertain boundaries.

7. Outline a justification for the assertion that ‘Modifiable areal units aren’t just a technical issue for GI science: they have profound implications for society’.
8. According to at least one GIS authority combinations of maps “... unravel more questions about data quality and boundary mismatch than they solve” (Leung Y and Leung K S (1993) An intelligent expert system shell for knowledge based geographical information systems 1. The tools. *International Journal of Geographic Information Systems*, 7(3): 189-99, page 193). Why?

## MULTIPLE CHOICE QUESTIONS (MCQ)

1. Arrange the following notions into the order in which they would commonly be addressed by a GIS analyst:
  - a. Representation
  - b. Analysis
  - c. The real world
  - d. Conception
  - e. Measurement
2. For each of the definitions given state whether it is of ‘error’, ‘accuracy’ or ‘precision’:

Definition	Concept?
Difference between observers or between instruments	
Difference between reality and our representation of reality	
Number of decimal digits in a measurement	

3. Section 6.2.3 contrasts uncertainty about clearly defined objects, handled by probability theory with ‘fuzziness’ in the objects themselves. Say whether each of the following statements is neither uncertain nor fuzzy, probabilistic, fuzzy or both:

Statement	Status
Joe Lobley is about 65	
I think maybe that Joe Lobley is about 65	
Joe Lobley is 65	
I think maybe that Joe Lobley is 65	

4. Which of the following are NOT ways of representing a field variable?

- a. Triangulated network
  - b. Contours
  - c. Planar enforcement
  - d. Elevation matrix
  - e. Pixel
5. A digital elevation matrix set has a claimed RMSE of 2.0m. If the matrix is 100 by 100, how many of the height values might be expected to be in error by over this amount?
- a. 6,800
  - b. 3,200
  - c. 500
6. If a variable can be assumed to be normally distributed, with sample mean of 12.34 and best estimate of the standard deviation of 3.44, what percentage of all sample observations can be expected to have values in the range 8.90 to 15.78?
- a. 34
  - b. 95
  - c. 68
  - d. 5
7. Round the following numbers to the specified number of decimal places:

Number	Decimal places	Answer
323.55	0	
25.05	1	
45.7896	2	
34.4999	0	

8. Rate each of the following area objects on a scale of ambiguity that goes from 'high' though 'medium' to 'low':

Object	Ambiguity Rating
--------	------------------

Wetland	
Oak forest	
Urban park	
Rain area	
Fog area	
Census tract	
State or County	

## CLASS AND INDIVIDUAL ACTIVITIES

1. What tools do GIS designers build into their products to help users deal with uncertainty? Take a look at your favorite GIS from this perspective. Does it allow you to associate metadata about data quality with datasets? Is there any support for propagation of uncertainty? How does it determine the number of significant digits when it prints numbers? What are the pros and cons of including such tools?
2. Visit [http://eros.usgs.gov/#/Find\\_Data/Products\\_and\\_Data\\_Available/DLGs](http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/DLGs) and examine the US Geological Survey Digital Line Graph User Guide. Discuss and debate its adequacy in the light of the issues raised in this Chapter.
3. Find out about the five components of data quality used in GIS standards, from the information available at [www.fdgc.gov](http://www.fdgc.gov). How are the five components applied in the case of a standard mapping agency data product, such as the US Geological Survey's Digital Orthophoto Quarter-Quadrangle program?
4. Imagine that you are a senior retail analyst for *Safemart*, which is contemplating expansion from its home US state to three others in the Union. Assess the relative merits of your own company's store loyalty card data (which you can assume are similar to those collected by any retail chain with which you are familiar) and of data from the 2000 Census in planning this strategic initiative. Pay particular attention to issues of survey content, the representativeness of population characteristics, and problems of scale and aggregation. Suggest ways in which the two data sources might complement one another in an integrated analysis.
5. The NCGIA Core Curriculum in GI Science has a Unit by Veregin that lists eight components of data quality (spatial accuracy, temporal accuracy, thematic accuracy, spatial resolution, temporal resolution, thematic resolution, consistency and completeness), see [www.ncgia.ucsb.edu/education/curricula/giscc/units/u100/u100\\_f.html](http://www.ncgia.ucsb.edu/education/curricula/giscc/units/u100/u100_f.html). Explore some digital data sets known to you and assess each data set on an ordinal scale good/average/poor on each component. Summarize your results by creating a table

in which the columns represent each of these components and rows represent each dataset.

6. Illustrate the effects of aggregation on basic descriptive statistics using the following 8 by 8 'raster' of cell values:

87	95	72	37	44	24	24	45
40	55	55	38	88	34	24	33
41	30	26	35	38	24	16	12
14	56	37	34	8	18	9	13
49	44	51	67	17	37	45	63
55	25	33	32	59	54	54	76
58	56	37	24	45	58	64	87
70	67	44	34	13	46	67	88

- a. First, compute the mean and standard deviation of the entire set of 64 values. Next, create a 4 by 4 matrix by amalgamating sets of four neighboring cells, replacing the four cell values by their arithmetic mean, and repeat the operation using these 16 values. Finally, repeat again, re-amalgamating to create a 2 by 2 grid. Discuss the answers to the following questions:
  - b. Why doesn't the mean change?
  - c. What is the effect on the standard error of the mean and why?
  - d. What is the effect on the range of values and the maximum and minimum values?
  - e. How do the variance and standard deviation change?
  - f. What does this exercise tell you about the nature of geographical data aggregates?
  - g. How do you think this effect will influence the results of any analysis using aggregated data?
7. One of the major problems with the idea of 'quality' lies in its definition, with at least four possible approaches based on
  - a. minimum quality, as with the US National Map Accuracy Standard;
  - b. conformance to an explicit metadata description, sometimes called 'truth in labeling' or 'what it says on the can', as with the US Spatial Data Transfer Standard;
  - c. market feedback
  - d. fitness for purpose

Organize four teams and ask them to prepare and present a case for the use of each of these approaches for spatial data. This exercise can be given a vaguely competitive edge by having an independent adjudicator.

8. Cohen's Kappa index is a standard measure of between ratings reliability in, for example, educational research. Its use in assessing attribute accuracy in classified satellite imagery is described at the Geographer's Craft website: <http://www.colorado.edu/geography/gcraft/notes/manerror/html/attribut.html>. It can be introduced either by using a classified image of your locality for which ground truth data are available, or by asking students to perform a classification at the same set of grid points and to compare their estimates of the preponderance of each land use class.
9. In today's marketplace, it is probable that there will be many possible sources of data for any specific GIS analysis, so that it is important to become what the *Geographer's Craft* (see website below) team call a 'smart shopper'. Imagine you are a GI consultant analyst, tasked to develop a GIS for a specific project. Use the Web to find the data that you would use, then justify your selection. A useful series of questions that a 'smart data shopper' might ask are given at the following website under the heading 'What to Look For and When to Quit': [http://www.colorado.edu/geography/gcraft/notes/sources/sources\\_f.html](http://www.colorado.edu/geography/gcraft/notes/sources/sources_f.html). The topic of the project and its region of application can be varied by the instructor, but outside of US care needs to be taken to ensure that a sufficiently rich list of data sets is available.
10. A geodemographic classification of the people living in some small zone is one of the most commonly used data source in business use of GIS. These classifications essentially collapse a very large number of variables in a single nominal category that is in some sense held to characterize that zone, such as 'laptops and latte' or 'empty nesters'. Examine at least two of these products and list the sources and likely magnitude of the uncertainties that their use will introduce into any analysis using them.

## FURTHER READING

Burrough P.A. and Frank A.U. (eds) 1996 *Geographic Objects with Indeterminate Boundaries*. London: Taylor and Francis.

This edited collection is the classic book on uncertainty in area objects.

Unwin, D.J. (1995) Geographical information systems and the problem of error and uncertainty. *Progress in Human Geography*, 19(4): 549-558.

Develops the idea of an error sensitive GIS.

Heuvelink G.B.M. 1998 *Error Propagation in Environmental Modelling with GIS*. London: Taylor and Francis.

It is almost certain that many of the issues raised in this Chapter will only be solved by rigorous use of concepts from geostatistics. The Dutch mathematician-geographer Gerard Heuvelink has been a pioneer in this effort, and his book should be essential reading for all serious GI analysts.

Zhang J.X. and Goodchild M.F. 2002 *Uncertainty in Geographical Information*. New York: Taylor and Francis.

Very much the standard work in which the idea of uncertainty is developed in considerable detail.

## RELATED READING

Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, NJ: Wiley.

13. Models of uncertainty in spatial data, P F Fisher, pp. 191–205. Much of the research on uncertainty is reviewed in this chapter.

18. Applying geocomputation to the analysis of spatial distributions, S Openshaw and S Albanides, pp. 267–282. If geostatistical theory is incapable of giving analytical solutions, then simulation of the uncertainty in results may well prove to be the best way of addressing the issue.

40. The future of GIS and spatial analysis, M F Goodchild and P A Longley, pp. 567–80

Maguire D.J., Goodchild M.F. and Rhind D.W. (eds) 1991 *Geographical Information Systems: Principles and Applications*. Harlow, UK: Longman (text available online at [www.wiley.co.uk/gis/volumes.html](http://www.wiley.co.uk/gis/volumes.html)).

11. Language issues for GIS, A U Frank and D M Mark, pp. 147-63

12. The error component in spatial data, N R Chrisman, pp. 165-74

24. Spatial data integration, R Flowerdew, pp. 375-87



## ONLINE RESOURCES

ESRI Virtual Campus course, *Turning Data into Information* by Paul Longley, Michael Goodchild, David Maguire, and David Rhind ([campus.esri.com](http://campus.esri.com) )

Module 1: Basics of Data and Information

Module 6: Uncertainty

Section 6.1, Module 1: Basics of Data and Information

Unit: Uncertainty

Section 6.2, Module 1: Basics of Data and Information

Unit: Uncertainty,

Sub-unit: Uncertainty in the conception of geographic phenomena

Section 6.2.3, Module 6: Uncertainty

Unit: Uncertainty issues for spatial data

Sub-unit: Fuzzy approaches

Section 6.3, Module 1: Basics of Data and Information

Unit: Uncertainty

Sub-unit: Uncertainty in the measurement of geographic phenomena

Section 6.3.2, Module 6: Uncertainty

Unit: Measuring uncertainty of nominal and ordinal values

Section 6.3.4, Module 6: Uncertainty

Unit: Uncertainty issues for spatial data,

Sub-unit: The spatial structure of errors

Section 6.4, Module 1: Basics of Data and Information

Unit: Uncertainty

Sub-unit: Uncertainty in the analysis of geographic phenomena

Section 6.4.2, Module 6: Uncertainty

Unit: Uncertainty issues for spatial data

Sub-unit: Error propagation

Section 6.4.4, Module 6: Uncertainty

Unit: Measuring uncertainty of interval or ratio values,

Sub-unit: Uncertainty in spatial data

Section 6.4.5, Module 6: Uncertainty

Unit: Uncertainty issues for spatial data,

Sub-unit: Living with uncertainty

Section 6.5, Module 6: Uncertainty

Unit: Uncertainty issues for spatial data

NCGIA Core Curriculum in GIScience, 2000 ([www.ncgia.ucsb.edu/giscc](http://www.ncgia.ucsb.edu/giscc))

2.10. Handling uncertainty (096), ed. *Gary Hunter* (see also [GC notes](#))

2.10.1. [Managing Uncertainty in GIS](#) (187), *Gary Hunter*

2.10.2. [Uncertainty Propagation in GIS](#) (098), *Gerard Heuvelink*

2.10.3. [Detecting and Evaluating Errors by Graphical Methods](#) (099), *Kate Beard*

2.10.4. [Data Quality Measurement and Assessment](#) (100), *Howard Veregin*

NCGIA Core Curriculum in GIS, 1990 ([www.ncgia.ucsb.edu/pubs/core.html](http://www.ncgia.ucsb.edu/pubs/core.html))

45. Accuracy of spatial databases

46. Managing error

## OVERVIEW

This chapter provides an overview of GIS software from various perspectives including architecture, functionality, history and key vendors.

## LEARNING OBJECTIVES

By the end of this chapter students should:

- Understand the architecture of GIS software systems, including organization by project, department, or enterprise; and, the three-tier architecture of software systems (graphical user interface; tools, and data access);
- Describe the process of GIS customization;
- Describe the main types of commercial software
  - Desktop
  - Web mapping
  - Server
  - Virtual globe
  - Developer
  - Hand-held
  - Other
- Outline the main types of commercial GIS software products currently available.

## KEY WORDS AND CONCEPTS

GIS software, graphical user interfaces (GUIs), COTS, Web services, software architecture, three-tier architecture, desktop and internet GIS, enterprise GIS, client-server, thick and thin clients, customization, API, components, thick and thin servers, middleware, hand-held GIS.

### Outline

- 7.1 Introduction
- 7.2 The evolution of GIS software
- 7.3 Architecture of GIS software
- 7.4 Building GIS software systems
- 7.5 GIS software vendors
- 7.6 Types of GIS software systems
- 7.7 Conclusion

## CHAPTER SUMMARY

### 7.1 Introduction

- Focus is on the different ways in which GIS capabilities are realized in GIS software products and implemented in operational GIS
- Takes a fairly specific view of GIS software, concentrating on systems with a range of generic capabilities to collect, store, manage, query, analyze, and present geographic information.
- The discussion is also restricted to GIS software products – well-defined collections of software and accompanying documentation, install scripts, etc. – that are subject to multi-versioned release control. By definition it excludes specific-purpose utilities, unsupported routines, and ephemeral codebases.
- Software can be:
  - COTS – commercial-off-the-shelf software
  - Shareware - (often intended for sale after a trial period)
  - Liteware - shareware with some capabilities disabled
  - Freeware - free software but with copyright restrictions
  - Public domain software - free with no restrictions
  - Open source software - the source code is provided and users agree not to limit the distribution of improvements

## 7.2 The evolution of GIS software

- Provides a very brief summary of the evolution of GIS from command line to GUI.
- Two key developments in the late 1980s were GUIs and customization capabilities
- The recent emergence of the Web services paradigm is mentioned.
- A Web service is an application that exposes its functions via a well-defined published interface that can be accessed over the Web from another program or Web service.

## 7.3 Architecture of GIS software

### 7.3.1 Project, departmental, and enterprise GIS

- Project GIS – a single, fixed term, one-off project, data are collected specifically for the project and little is thought of reuse, absence of organizational vision, sharing data and experience is a low priority
- Department GIS – several projects in the same department are amalgamated, creation of common standards, development of a focused GIS team, procurement of new GIS capabilities
- Enterprise GIS – standards are accepted across multiple departments, resources are centrally-funded and managed
- Societal implementation – tens of thousands of users become engaged in GIS and connected over the Web – e.g. Google Earth, Microsoft Virtual Earth, and ESRI's ArcGIS

### 7.3.2 The three-tier architecture

Three key parts of a GIS are

- The user interface – via a GUI, an integrated collection of menus, tool bars and other controls
- The tools – define the capabilities or functions available
- The data management system
- Describes the classical three tier IS architecture – presentation, business logic and data server – and maps it to the key parts of a GIS
- Explains the four types of computer system architecture configurations – desktop, client-server, centralized desktop, and centralized server.
- Introduces the terms client-server, thick and thin clients

### 7.3.3 Software data models and customization

- A data model defines how the real world is represented in a GIS
- A software data model defines how the different tools are grouped together, how they can be used, and how they interact with data.
- Customization is the process of modifying GIS software
- A number of industry standard programming languages (such as Visual Basic, Java, and Python) are available for customizing GIS software systems.
- Integrated development environments (IDEs) combine several software development tools including a visual programming language; an editor; a debugger, and a profiler.
- To support customization using open, industry-standard IDEs, a GIS vendor must expose details of the software package's functionality using a set of application programming interfaces (APIs).
- Components are important to software developers because they are the mechanism by which reusable, self-contained, software building blocks are created.
- Allow many programmers to work together
- Can be easily assembled into larger systems
- Can be reused
- Support multiple third party extensions

### 7.3.4 GIS on the desktop and on the Web

Describes the concepts of thick and thin servers

Network GIS use the cross-platform Web browser to host the viewer interface

## 7.4 Building GIS software systems

- Start with a formal design for a software system and then build each part of component separately before assembling the whole system.
- Core GIS software systems are usually written in a modern programming language like Visual C++, C# or Java.
- A key choice that faces all software developers is whether to design a software system by buying in components, or to build it more or less from scratch.

## 7.5 GIS software vendors

This section very briefly describes the origins and product environment of four leading GIS vendors

- Autodesk
- Bentley

- ESRI
- Intergraph

## 7.6 Types of GIS software systems

### 7.6.1 Desktop GIS software

- Owes its origins to the personal computer and Microsoft Windows operating system
- Are the mainstream workhorses of GIS today
- Free GIS viewers are able to display and query popular file formats
- These help to establish market share and can create *de facto* standards
- Focus on use rather than data creation
- Professional GIS are full-featured desktop GIS with a superset of capabilities

#### **Applications Box 7.2 Desktop GIS: Intergraph GeoMedia**

Outlines the components of this full featured product

#### **Applications Box 7.4 Desktop GIS: ESRI ArcGIS ArcInfo**

Summarizes the history of the development of this modern GIS software

### 7.6.2 Web Mapping

- There are signs that the next decade will in turn be dominated by Web and server GIS products
- Web mapping is taken to mean integrated Web-accessible software, a 2-D database comprising one or more base maps, and an associated collection of services.
- Web access is provided via easily accessible, open interfaces running in web browsers and returning image tiles (fragments of the total map)
- Website functions can easily be accessed programmatically via well-defined application programming interfaces (APIs)

### 7.6.3 Server GIS

- Server GIS runs on a computer server that can handle concurrent processing requests from a range of networked clients
- Initially, server GIS were nothing more than ports of desktop GIS products
- Second-generation systems were subsequently built using a multiuser services based architecture that allows them to run unattended and to handle many concurrent requests from remote networked users

### **Applications Box 7.4 Server GIS: Autodesk MapGuide**

Describes the components and key functionality of this server GIS

#### **7.6.4 Virtual Globes**

- Virtual globes allow users to visualize geographic information on top of 3-D global base maps
- Virtual Globes include: Google Earth, Microsoft Virtual Earth
- Neogeography is the “new” geography that among other things includes the overlay or mashing up two or more sources of geographic information

#### **7.6.5 Developer GIS**

- Toolkits of GIS functions (components) that a reasonably knowledgeable programmer can use to build a specific-purpose GIS application

#### **7.6.6 Hand-held GIS**

- GIS for mobile and personal use on hand-held systems
  - development of low-cost, lightweight location positioning technologies and wireless networking has further stimulated this market
- Recently, “smartphones” can deal with comparatively large amounts of data and sophisticated applications

#### **7.6.7 Other types of GIS software**

- Mentions raster-based and CAD-based GIS and explains how they have merged with more general GIS
- Middleware offer centralized management of data, the ability to process data on a server and control over database editing and update
- Standard DBMS have extensions to store and process geographic information efficiently
- Also mentions public-domain, open source and free software
- Seamless GIServices offer functionality delivered packaged along with data

## **7.7 Conclusion**

### **ESSAY TOPICS**

1. With reference to specific examples, outline the main types of commercial GIS products currently available.



2. Section 7.1 lists some 'GIS-like' products, which includes mapping systems, image processors and spatial extensions to standard database management systems. In each case, give a reasoned justification for the view that these products are not GIS.
3. Discuss the role of each of the three tiers of software architecture in an enterprise GIS implementation.
4. With reference to a large organization that is familiar to you, describe the ways in which its staff might use GIS, and evaluate the different types of GIS software systems that might be implemented to fulfill these needs.
5. Compare and contrast the roles of GIS software used in 'project', 'departmental' and 'enterprise' frameworks.
6. Distinguish between 'desktop', and both 'thin' and 'thick' client serving in GIS and explain how they GIS would be implemented in each environment.
7. What are the advantages and disadvantages of component architecture in a GIS?
8. From the perspective of both user and supplier, how has the coming of the Internet and World Wide Web affected GIS?
9. What do you understand by the phrase 'client server architecture' and why is it so popular in GIS implementations?
10. You are working in the research department of a hospital, responsible for processing data on the health of a small city. List and describe the factors that would influence your choice of software configuration and supplier.

### MULTIPLE CHOICE QUESTIONS (MCQ)

1. For each of the listed software distribution models, match it to its most appropriate description:

Distribution Model	ANSWER	Description
'shareware'		a) for direct sale
Open source		b) free with no restrictions on use
Commercial off the shelf (COTS)		c) code provided free but with an agreement not to distribute improvements
'freeware'		d) for sale after a trial period
Public domain		e) free, but with copyright restrictions

2. Which of the following programming languages is not currently used in customizing GIS?
  - a. Python
  - b. Java
  - c. FORTRAN
  - d. Visual Basic
  
3. Starting with the simplest, arrange the following types of GIS implementation into their logical order (a) – (d):

Type of Implementation	Answer (a) to (d)
Workgroup	
Societal	
Project	
Enterprise	

4. Which of the following are standards for defining and re-using software components?

Standard	Answer (Yes or No)
HTML	
.NET	
Java	
XML	

5. For each of the following features, chose the alternative feature that most characterizes desktop and network GIS:

Feature	Choices	Desktop GIS	Network GIS
Client size	Thick/Thin		
Client platform	Windows/Browser		
Server size	Thin/Thick		
Server platform	.Net/Java		
Component standard	.Net/Java		
Network	LAN/WAN		

6. Rank the three systems listed in section 7.5 by their 2007 market share
- Intergraph
  - ESRI
  - Autodesk
  - Bentley
7. Arrange the four types of computer architectural configurations used to build operational GIS implementations in order of their probable software complexity from (a, most) to (d, least):

Configuration	Answer (a) to (d)
Desktop	
Client server	
Centralized desktop	
Centralized server	

8. Which is the most important component of a full GIS implementation?
- Software
  - People
  - Data
  - Hardware
9. From the user's viewpoint, which two of the following were to two key GIS software developments of the 1980s?
- Web services
  - Graphical user interfaces
  - Integrated development environments
  - Customization capabilities

## CLASS AND INDIVIDUAL ACTIVITIES

Design a GIS architecture that 25 users in 3 cities could use to create an inventory of recreation facilities.

Use an Internet search engine to find the websites of the main GIS software vendors and compare and contrast what you infer is their individual product strategies. In what ways are they different?

Section 7.6.6 describes some hand-held GIS based primarily on GPS as their vocational technology. Find appropriate descriptions of these devices and then collectively assess the problems of delivering maps and map like information to their users. Useful information on 'wearable' GIS can be found in the Project Battuta website at <http://dq.statlab.iastate.edu/dq/>.

Re-read 7.5. on popular GIS, then visit the website of a popular image processing system such as ERDAS Imagine at <http://www.erdas.com/tabid/84/currentid/1050/default.aspx>. Now use the 'brainstorming' technique to develop a list of the key differences between the GIS. The technique is described by Gold et al (1992) at <http://www2.glos.ac.uk/gdn/gold/ch5.htm>

There are three views on the suitability of GIS for educational use. One offers an open source and fully functional GIS (for example Quantum GIS from <http://www.qgis.org/>), the second offers a 'cut down' version of a full GIS (e.g. ArcGIS Explorer - <http://www.esri.com/software/arcgis/explorer/index.html>), and the third (exemplified by AEGIS-3, see [www.advisory-unit.org.uk](http://www.advisory-unit.org.uk)) offers strongly curriculum-linked functions. Organize a debate or seminar on the relative strengths and weaknesses of each.

As discussed in Section 7.6.7, it is possible to create a GIS-like environment using essentially free of charge systems such as GeoDa, gvSIG, PostGIS and GRASS. Download, install and use one, but keep a journal of your experiences during this work. Your log may show that getting this type of software 'up and running' isn't as easy as might at first sight be thought.

A full GIS implementation consists of appropriately configured hardware, geospatial data, the software and the 'liveware' to make it all work. Debate the proposition that 'This house believes that it is the people who are the most important component of any GIS implementation'.

## FURTHER READING

Bishr Y. and Radwan M. 2000 'GDI Architectures'. In Groot R. and McLaughlin J. (eds) *Geospatial Data Infrastructure: Concepts, Cases, and Good Practice*. New York: Oxford University Press, pp. 135–50.

Covers the main architectures to 2000.

Peng Z-H. and Tsou M-H. 2003 *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks*. Hoboken, NJ: Wiley.

A clear account of the issues in delivering GIS over the Internet

Haklay, M., Singleton, A.D., Parker, C. (2008) Web Mapping 2.0: the Neogeography of the Geospatial Internet. *Geography Compass*, 2(6), 2011–2039

## RELATED READING

Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, NJ: Wiley.

22. GIS in networked environments, D J Coleman, pp. 317–29.

23. Desktop GIS software, S Elshaw Thrall and G I Thrall, pp. 331–45

25. GIS customization, D J Maguire, pp. 359–69

Maguire D.J., Goodchild M.F. and Rhind D.W. (eds) 1991 *Geographical Information Systems: Principles and Applications*. Harlow, UK: Longman (text available online at [www.wiley.co.uk/gis/volumes.html](http://www.wiley.co.uk/gis/volumes.html)).

3. The technological setting of GIS, M F Goodchild, pp. 45-54

4. The commercial setting of GIS, J Dangermond, pp. 55-65

## ONLINE RESOURCES

NCGIA Core Curriculum in GIScience, 2000 ([www.ncgia.ucsb.edu/giscc](http://www.ncgia.ucsb.edu/giscc))

2.1.1. [Fundamentals of Data Storage](#) - Carol Jacobson (037)

2.14.5. [WebGIS](#) (133), Kenneth Foote and Anthony Kirvan

NCGIA Core Curriculum in GIS, 1990 ([www.ncgia.ucsb.edu/pubs/core.html](http://www.ncgia.ucsb.edu/pubs/core.html))

3. Introduction to computers

4. Raster GIS

13. Vector GIS

- 18. Modes of user/GIS interaction
- 24. GIS marketplace

# Geographic Data Modeling

---

## OVERVIEW

Describes the process of data modeling and the various data models that have been used in GIS.

## LEARNING OBJECTIVES

By the end of this chapter students should:

- Define what geographic data models are and discuss their importance in GIS;
- Understand how to undertake GIS data modeling;
- Outline the main geographic models used in GIS today and their strengths and weaknesses;
- Understand key topology concepts and why topology is useful for data validation, analysis, and editing;
- Read data model notation;
- Describe how to model the world and create a useful geographic database.

## KEY WORDS AND CONCEPTS

Topological relationships, planar enforcement, georelational model, linear referencing, dynamic segmentation, object classes, topology rules, encapsulation, inheritance, polymorphism, UML, CASE tools, database schema

## OUTLINE

- 8.1 Introduction
- 8.2 GIS data models
- 8.3 Example of a water-facility object data model
- 8.4 Geographic data modeling in practice

## CHAPTER SUMMARY

### 8.1 Introduction

- Focuses on how geographic reality is modeled
- Differentiates between
  - *Representation* which can be considered to denote the conceptual and scientific issues
  - *Model* which is used in practical and database concepts
  - *Data model* used to distinguish it from process models

#### 8.1.1 Data model overview

- Data model is a set of constructs for representing objects and processes in the digital environment of the computer
- Decisions about the type of data model to be adopted are vital to the success of a GIS project.

#### 8.1.2 Levels of data model abstraction

- Four different levels of abstraction (levels of generalization or simplification) are
  - Reality is made up of real-world phenomena and includes all aspects that may or may not be perceived by individuals, or deemed relevant to a particular application.
  - The conceptual model is a human-oriented, often partially structured, model of selected objects and processes that are thought relevant to a particular problem domain.
  - The logical model is an implementation-oriented representation of reality that is often expressed in the form of diagrams and lists.
  - The physical model portrays the actual implementation in a GIS, and often comprises tables stored as files or databases



- The conceptual modeling phase begins with definition of the main types of objects to be represented in the GIS and concludes with a conceptual description of the main types of objects and relationships between them.
- The logical modeling phase leads to the creation of diagrams and lists describing the names of objects, their behavior, and the type of interaction between objects.
- The physical modeling phase involves describing the exact files or database tables used to store the data, the relationships between objects types, and the precise operations that can be performed.
- A data model provides
  - Developers with the means to represent an application domain in terms that may be translated into a design and implemented
  - Users with a description of the structure of the system, independent of specific items of data or details of the particular application

## 8.2 GIS data models

- The key types of geographic data models and their main areas of application are listed in Table 8.1.
- A collection of entities of the same geometric type (dimensionality) is referred to as a class or layer.
- It should also be noted that the term layer is quite widely used in GIS as a general term for a specific dataset.

### 8.2.1 CAD, graphical, and image GIS data models

- In a CAD system, real-world entities are represented symbolically as simple point, line, and polygon vectors.
- There are three severe problems with CAD models
  - Use local drawing coordinates
  - Individual objects do not have unique identifiers
  - Are focused on graphical representation of objects, cannot store relationships
- Computer cartography (graphical) data models store entities as points, lines and polygons, with annotation used for placenames
- Image data models store scanned aerial air photos and digital satellite images as rasters or grids

### 8.2.2 Raster data model

- The raster data model uses an array of cells, or pixels, to represent real-world objects

- The cells can hold any attribute values based on one of several encoding schemes including categories, and integer and floating-point numbers
- In some systems, multiple attributes can be stored
- Techniques for compressing rasters are described in Box 8.1, includes run-length encoding, block encoding and wavelet

### 8.2.3 Vector data model

- In the vector data model each object in the real world is first classified into a geometric type: in the 2-D case point, line, or polygon (Figure 8.7).
- Points are recoded as single coordinate pairs
- Lines as a series of ordered coordinate pairs
- Polygons as one or more line segments that close
- In some systems, curves can be defined by a mathematical function

#### 8.2.3.1 Simple features

- Geographic entities encoded using the vector data model are usually called features
- Features of the same geometric type are stored in a geographic database as a feature class, or when speaking about physical representation the term feature table is preferred
- Simple feature datasets are sometimes called *spaghetti* because lines and polygons can overlap and there are no relationships between any of the objects.
- Simple features lack more advanced data structure characteristics, such as topology

#### 8.2.3.2 Topological features

- Topological features are essentially simple features structured using topological rules.
- Topology is the mathematics and science of geometrical relationships.
- Topological relationships are non-metric (qualitative) properties of geographic objects that remain constant when the geographic space of objects is distorted.
- Topological structuring of layers forces all line ends that are within a user-defined distance to be snapped together so that they are given exactly the same coordinate value. A node is placed wherever the ends of lines meet or cross.
- *Data validation* topology tests include network connectivity, line intersection, overlap, duplicate lines
- Many objects share common locations and partial identities. These situations can be modeled in a GIS database as either
  - single objects with multiple geometry representations, or

- multiple objects with separate geometry integrated for editing, analysis, and representation.
- Topologically aware tools include: manipulate common, shared polylines and nodes as single geometric objects; rubberbanding; snapping; auto-closure; tracing
- Optimized queries from topological relationships include network tracing, polygon adjacency, containment, intersection
- In a topologically structured polygon data layer each polygon is defined as a collection of polylines that in turn are made up of an ordered list of coordinates (vertices).
- Figure 8.8 shows an example of a topologically structured polygon dataset
- Storing common boundaries between adjacent polygons avoids the potential problems of gaps (slivers) or overlaps
- Downside is that drawing polygons from multiple polylines is time intensive
- *Planar enforcement* means that all the space on a map must be filled and that any point must fall in one polygon alone, that is, polygons must not overlap.
- Planar enforcement implies that the phenomenon being represented is conceptualized as a field.
- In the *georelational* model, the feature geometries and associated topological information are stored in regular computer files, whereas the associated attribute information is held in relational database management system (RDBMS) tables.
- Geometry and topology were not placed in RDBMS because until relatively recently RDBMS were unable to store and retrieve geographic data efficiently

#### 8.2.3.3 Network data model

- Network topological relationships define how lines connect with each other at nodes and define rules about how flows can move through the network
- In *linear referencing systems*, the location of geographic entities are stored as distances along a network from a point of origin
- *Dynamic segmentation* is a special case of linear referencing in which data values are added dynamically to the route each time the user queries the database

#### 8.2.3.4 TIN data model

- A TIN is a topological data structure that manages information about the nodes comprising each triangle and the neighbors of each triangle.
- Briefly explains how Delaunay triangulation is carried out
- Advantages of TINs are
  - the density of sampled points can be adjusted to reflect relief
  - they incorporate the original sample points

- easy to calculate elevation, slope, aspect, and line-of-sight
- Limitations of TINs are
  - They are especially susceptible to extreme high and low values since there is no smoothing of the original data
  - Unable to deal with discontinuity of slope across triangle boundaries
  - Difficult to calculate optimum routes

#### 8.2.4 Object data model

- Each geographic object is an integrated package of geometry, properties, and methods.
- Geometry is treated like any other attribute of the object
- Geographic objects of the same type are grouped together as *object classes*
- Individual objects in the class are *instances*
- Three types of *relationships* in object data models are
  - *Topological* relationships are built into the class definition such as network and polygon structures
  - *Geographic* relationships are based on geographic operators that determine the interaction between objects
  - *General* relationships define other types of relationships such as between land parcels and ownership data, light pole IDs and attributes
- *Rules* are a means of maintaining data integrity during editing tasks. They include
  - *Attribute* rules are used to define the possible attribute values that can be entered, includes both range and coded values
  - *Connectivity* rules specify the valid combinations of features
  - *Geographic* rules define what happens to the properties of objects when an editor splits or merges them

#### Technical Box 8.2 Object-oriented concepts in GIS

- An *object* is a self-contained package of information describing the characteristics and capabilities of an entity under study.
- An interaction between two objects is called a *relationship*.
- A collection of objects of the same type is called a *class*.
- A class can be thought of as a template for objects.
- There are three key facts of object data models
  - *Encapsulation* which describes the fact that each object packages together a description of its state and behavior

- *Inheritance* is the ability to reuse some or all of the characteristics of one object in another object.
- *Polymorphism* describes the process whereby each object has its own specific implementation for operations like draw, create, and delete.

### 8.3 Example of a water-facility object data model

- The goal of this section is to describe an example of a geographic object model and discuss how many of the concepts introduced earlier in this chapter are used in practice.
- Figure 8.18 shows a possible object model using Unified Modeling Language (UML) to show objects and the relationships between them
- In UML models each box is an object class and the lines define how one class reuses (inherits) part of the class above it in a hierarchy.
- Figure 8.19 shows how a computer-aided software engineering (CASE) tool is used to specify the logical model

### 8.4 Geographic data modeling in practice

- No step in data modeling is more important than understanding the purpose of the data modeling exercise, gained by collecting user requirements from the main users
- Once an implementation-independent logical model has been created (using CASE tools and UML, for example), this model can be turned into a system-dependent physical model
- A physical model will result in an empty *database schema* – a collection of database tables and the relationships between them

### ESSAY TOPICS

1. Define what is meant by the term ‘data model’ and explain why its design is vital to the success of any GIS application.
2. Describe, with examples, five key differences between topological vector and raster geographic data models. It may be useful to consult Figure 8.3 and refer back to Chapter 3.
3. It has been suggested that the differences between the vector and raster conceptual computer models will be eliminated by technological advances. What do you think are these advances? Have the vector and raster models converged?
4. What are the deficiencies of CAD, graphical and image GIS data models in geographic data analysis?

5. In the hybrid georelational model, use is made of a standard relational data base management system. What are the advantages of this approach to both data modeler and GIS user?
6. Study Section 8.4.2 and Technical Box 8.2. Why do GI scientists favor the object oriented approach?
7. 'There is no such thing as the correct geographic data model' (page 196). Explain why.
8. What do you understand by the terms encapsulation, inheritance and polymorphism and why are they important in object oriented data modeling?
9. The Chapter lists eight different approaches to data modeling in GIS. Consider how each might be extended to create either (a) a truly three-dimensional system for use in geology or (b) a temporal GIS for use in tracking land use change over time.

### MULTIPLE CHOICE QUESTIONS (MCQ)

1. Place the following terms into their correct ordering by level of abstraction involved:

Term	Write 1 – 4
Conceptual model	
Logical model	
Reality	
Physical model	

2. For each of the eight types of model, state whether it is best for representing a field or an object:

Geographic Data Model	Field (F) or Object (O)?
Computer-aided design	
Graphical, non topological	
Image	
Raster/grid	
Vector/Georelational topological	
Network	
Triangulated Irregular Network (TIN)	
Object	

3. For each of the eight models, state which applications area it is most suited to support, choosing your answers from the list of the right hand column (each application should be

selected once only):

Geographic Data Model	Answer (a)-(h)	Application
Computer-aided design		a) Surface terrain analysis
Graphical, non topological		b) Computer cartography
Image		c) Spatial analysis
Raster/grid		d) Socio-economic data analysis
Vector/Georelational topological		e) Natural hazards assessment
Network		f) Estate management
Triangulated Irregular Network (TIN)		g) River management
Object		h) Land ownership information

4. Tick the following GIS operations in which topological information is useful:

- a. Data validation
- b. Modeling integrated features
- c. Editing
- d. Spatial interpolation
- e. Query optimization

5. Give the full name of each of the acronyms listed below:

UML	
VRML	
XML	
HTML	

6. Tick which of the following statements describe a TIN model in GIS:

- a. An alternative to raster coding of surface terrain
- b. A way of coding objects in GIS
- c. A way of modeling continuous data in a vector GIS environment
- d. An image data format.

7. A data model that separates the attribute information from the feature geometries and stores it separately is called

- a. Object oriented
  - b. Georelational
  - c. Topological
  - d. Network
8. Which of the following are not topological relationships?
- a. Containment
  - b. Direction
  - c. Adjacency
  - d. Intersection
  - e. Distance
9. A 'class' in data base modeling is best defined as:
- a. A grouping of similar object types
  - b. A template for creating objects
  - c. A set of features

## CLASS AND INDIVIDUAL ACTIVITIES

1. Figure 8.3 has a SPOT satellite image together with a vector map of the same area. Make a list of the ways in which these differ, thinking about resolution, selectivity and so on.
2. Read Section 8.2, and then refer back to Figure 3.14 or an equivalent topographic map of an area known to you. Following the processes outlined in the chapter develop a data model for the information displayed on this map. An accessible guide to database design in GIS will be found at <http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u10.html#UNIT10>.
3. Assume that you are working for a road maintenance agency. Your spatial objects are the roads over, say, a small city. The GIS is to support operations such as:
  - a. Surface renewal as required;
  - b. Avoiding clashes with other agencies that might want to dig holes in your roads; and
  - c. Improvements to the road structure.
4. How would you record the *network* of roads in your database and what attributes would you attempt to collect?
5. Now assume that you are working for a bus company for the same city, and suggest a GIS to support operations such as:



Timetabling and timing

Predicting demand for the buses

6. This exercise explores the dependence of lossless compression of a raster on the nature of the data involved. Run length encoding (RLE) is an obvious way to compress a raster of values. In it we record each run of values as a pair of numbers, one giving the number itself, the other the number of repeats. Given the raster of integers:

12111234

11112233

11111223

44112233

44412223

44442233

44422223

7. What is the saving in store if we run length encode this?

Now, given the raster of floating-point values:

1.1 2.1 1.1 1.4 1.6 2.1 3.6 4.2

1.1 1.4 1.5 1.8 2.9 2.7 3.4 3.1

1.1 1.4 1.2 1.1 1.4 2.4 2.8 3.3

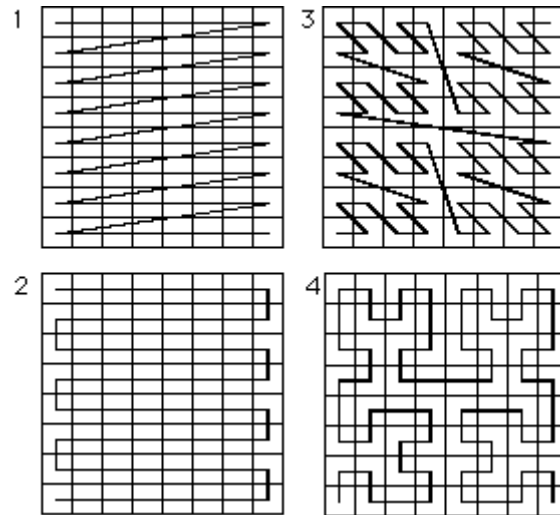
4.1 4.2 1.1 1.9 2.2 2.4 3.1 3.9

4.2 4.3 4.4 1.4 2.2 2.3 2.4 3.3

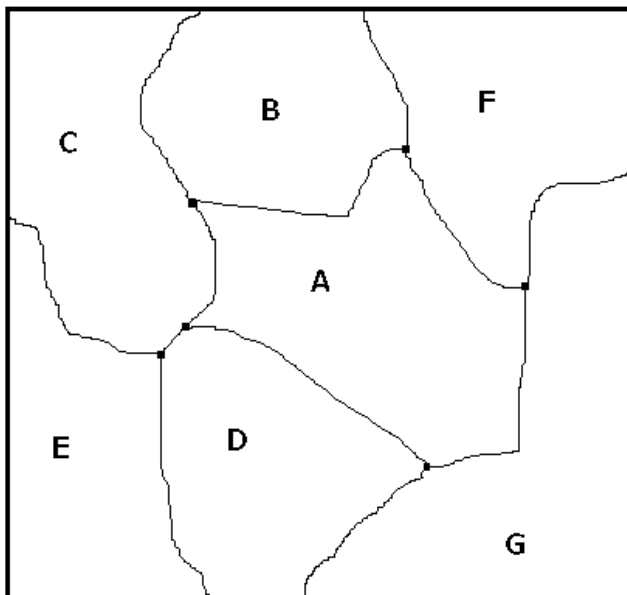
4.4 4.1 4.9 4.7 2.1 2.2 3.1 3.3

4.2 4.4 2.1 2.2 2.3 2.4 3.6 4.0

What is the saving in store if we use the same approach? To be effective RLE relies on like values being next to each other in the sequence, and in the two examples so far, this has been as in (1) below. It is possible to do better by wrapping round at the end of each row of the raster, as in (2). Repeat the exercise. Does it compress the data even further?

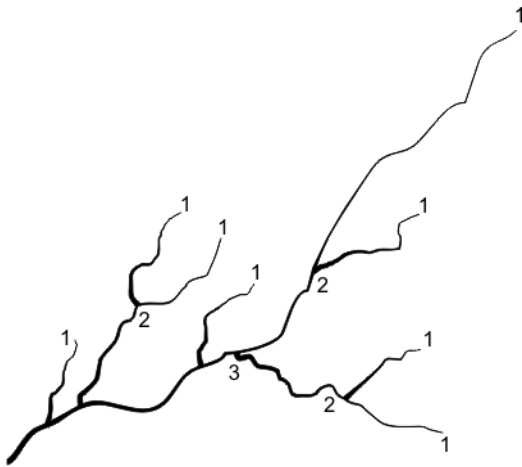


8. Boxes (3) and (4) show scan orders that attempt to improve things further, this time by scanning the raster in an order that, if it is present, will make maximal use of any spatial autocorrelation in the grid. Order the entries according to these schemes and repeat the exercise. Method (3) is the so-called Morton Ordering, and it has been much debated in GIS. What do you conclude from this exercise?
  
9. Study Figures 8.8. For the planar enforced areas A – E below number the polylines and create two tables that describe the structure. There is no need to complete the digitization of each polyline.



10. How big is mainland Australia? This exercise is taken from O'Sullivan and Unwin (2002, page 175). It can be done using a semi-automatic digitizer, but it is useful to do it by hand.
11. Trace the shoreline of Australia from a map of the continent, taking care to ensure that the source is drawn on an equal area map projection. Record the shoreline as a series of (x, y) co-ordinates. How many vertices do you need to represent the shape of Australia so that it is instantly recognizable? What is the minimum number you can get away with? How many do you think you need to ensure that you get a reasonable estimate of the total area of the continent?
- Use Simpson's method (see O'Sullivan and Unwin, 2010, page 191-194) to compute its area. This is easily done using any spreadsheet program. Enter your co-ordinates into the first two columns and copy these from row 2 onwards into the next two columns, displacing them 'upwards' by one row as you do so. Make a copy of the first co-ordinate pair into the last row of the copied columns. The next column can then be used to enter and calculate the trapezoid formula. The sum of this column then gives your estimate of the continent's area. You will have to scale the numbers from co-ordinate units into real distances and areas on the ground. Compare your result with the 'official' value, which is 2,974,581 km<sup>2</sup>.
  - O'Sullivan and Unwin thought that the minimum number of coordinate pairs needed to make the result recognizably Australia is nine. Using a 1:30 000 000 map as a source they recorded 45 co-ordinates for the shoreline and got an area of 2,964,185 km<sup>2</sup>, which is about 0.3% too low. In fact, the closeness of this result is likely to be a happy accident. What conclusions do you draw from this exercise?
11. Illustrating structure in a tree network. This exercise is taken from O'Sullivan and Unwin (2010). From a topographic map at a scale of 1:25,000 or 1:50,000 (or nearest equivalents), trace off a drainage network represented by (usually) the 'blue lines'. Try to find a reasonably large basin with, say, about 50 or so 'sources' where streams begin.
- Now 'order' this network using Strahler's method. To do this, give all the fingertip tributaries the order '1'. Where two such 1<sup>st</sup> order streams meet, the stream that results becomes 2<sup>nd</sup> order, and should be labeled as such. Note that any 1<sup>st</sup> order tributaries that join a 2<sup>nd</sup> order stream will not change its order until it meets another 2<sup>nd</sup> order stream, when the stream becomes '3<sup>rd</sup> order'.

- b. Continue until all the streams have been ordered in this way and there is just one stream with the highest order reached. This ordering scheme is :



### The Strahler ordering scheme

- c. Now count the number of streams in each order. In the example this count is eight 1<sup>st</sup>, three 2<sup>nd</sup> and one 3<sup>rd</sup>.
- d. Plot a graph of the *logarithm of the number of streams in each order* on the vertical axis against stream order (1 , 2 , 3 , ... *n*) on the horizontal axis.

Usually this results in a straight line plot, evidence of what was called the 'law of stream numbers', one of several laws due to a US engineer called Horton. This structure seems to be a property of most branching networks.

12. Representation in raster and vector codings. The basic objective of this assignment is to code a series of area polygons into vector and raster codings, and then use appropriate algorithms to calculate the total area of the object. We suggest that you proceed as follows:

- a. Choose an area object such as a lake, city built over area or forest. Initially make this small enough so that considerations of Earth curvature are not significant.
- b. Establish an origin at the bottom left of the plan and create an appropriate plane co-ordinate system.
- c. Use this system to record the co-ordinates in an exact, vector data structure. You could do this directly into a GIS package, by digitizing off screen, but it is just as easily done using a spreadsheet.
- d. Use the appropriate algorithm to compute the area of each and hence the total area.
- e. Using the same origin (bottom left), create an appropriate raster grid for the entire area. Use two resolutions, a 16 by 16 grid and a 32 by 32 grid. Record

these grids using both run length encoding based on a row ordering, and compare this with the same run length encoding generated using the Morton ordering.

- f. Estimate the total area from both grids.
- g. Finally, write an account saying which of the two approaches is most appropriate for this problem and why.

13. As a final task, imagine that the given map is not depicting a single, small area, but an extensive, continental scale series of areas, such as the lakes of Manitoba, Canada. What implications would this have on measurement techniques and results, in particular in reference to map projection and map reference system?

At slightly more advanced level this exercise can be extended to include coding the quadtree for each of the rasters.

14. Investigating compression. Although raster compression can be achieved using run length and related encodings, use is often made of proprietary compression using, for example, discrete cosine, fractal, and wavelet approaches. Using the websites and a selection of these approaches to compress some of the files on your computer. Which seems to work most effectively for which type of data? Can the approach be used to enhance data for display? Is the method lossy or lossless? Some sites to explore are: [www.lizardtech.com](http://www.lizardtech.com), [www.jpeg.org](http://www.jpeg.org), and [www.pkzip.com](http://www.pkzip.com).
15. As the chapter indicates, obtaining a clear set of user requirements is a sine qua non of data modeling. Use a role play simulation involving a data modeling team with a series of 'clients'. In this, it is helpful to develop the client needs as a set of 'use cases', each defining a specific, defined potential use of the GIS when it is established. A useful website describing the approach is at [www.pols.co.uk/use-case-zone/](http://www.pols.co.uk/use-case-zone/) but there are many other sites that have similar materials.

## FURTHER READING

Abel, D. and Mark D.M. (1990) A comparative analysis of some two-dimensional orderings International Journal of Geographical Information Systems, 4(1): 21-31.

A classic attempt to show how raster ordering can help compression.

Fisher, P. and D. Unwin (Eds) 2005 Re-presenting GIS, London: Wiley

A set of research papers that together provide numerous arguments for full object orientation in geographical data bases.

Gahegan, M.N. and S. Roberts (1988) An intelligent, object- oriented geographical information system, *International Journal of Geographical Information Systems* 2: 101-10. An attempt to produce a fully object oriented GIS.

Goodchild, M.F., and A.W. Grandfield (1983) Optimizing raster storage: an examination of four alternatives, *Proceedings, AutoCarto 6*, Ottawa, 1: 400-7. More on compression.

Peucker, T.K., and N. Chrisman (1975) Geographic Data Structures, *American Cartographer* 2(1): 55-69.

Peuquet, D.J. (1984) A conceptual framework and comparison of spatial data models, *Cartographica* 21(4): 66-113. Two papers that explore the standard arc-node data structure.

Raper, J.F. (2000) *Multidimensional GIS*, London: Taylor & Francis

An advanced discussion of the need to incorporate additional dimensions, such as height/depth and time, into geographic data modeling. A case study of research at Scolt Head in England provides a powerful argument for full object orientation in data base design.

Worboys, M.F. (1992) A generic model for planar spatial objects. *International Journal of Geographical Information Systems*, 6(5), 353-372. The classic paper on object types in GIS by a logician/computer scientist.

Worboys, M.F., Hearnshaw H.M. and D.J. Maguire, (1990) Object-oriented data modeling for spatial databases. *International Journal of Geographical Information Systems* 4: 369-383. Another classic paper, one of the earliest calls for object orientation in GIS.

Zeiler, Michael (1999) *Modeling our World: The ESRI Guide to Geodatabase Design* (Redlands CA: ESRI Press). An extremely useful survey of the approaches taken by the world-leading GIS software.

## RELATED READING

Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, NJ: Wiley.

26. Relational databases and beyond, M R Worboys

Maguire D.J., Goodchild M.F. and Rhind D.W. (eds) 1991 *Geographical Information Systems: Principles and Applications*. Harlow, UK: Longman (text available online at [www.wiley.co.uk/gis/volumes.html](http://www.wiley.co.uk/gis/volumes.html)).

16. High-level spatial data structures for GIS, M J Egenhofer and J R Herring, pp. 227-37

19. Digital terrain modeling, R Weibel and M Heller, pp. 269-97

## ONLINE RESOURCES

NCGIA Core Curriculum in GIScience, 2000 ([www.ncgia.ucsb.edu/giscc](http://www.ncgia.ucsb.edu/giscc))

- 2.3.1. [Information Organization and Data Structure](#) (051), Albert Yeung
- 2.3.2. [Non-spatial Database Models](#) (045), Thomas Meyer
- 2.4.1. [Rasters](#) (055), Michael Goodchild
- 2.4.2. [TINs](#) (056)
- 2.4.3. [Quadrees and Scan Orders](#) (057), Michael Goodchild
- 2.6. [Representing networks](#) (064), Benjamin Zhan
- 2.9.1. [Transportation Networks](#) (183), Val Noronha

NCGIA Core Curriculum in GIS, 1990 ([www.ncgia.ucsb.edu/pubs/core.html](http://www.ncgia.ucsb.edu/pubs/core.html))

- 4. Raster GIS
- 11. Spatial objects and database models
- 12. Relationships among spatial objects
- 13. Vector GIS
- 21. Raster/vector debate
- 30. Storage of complex spatial objects
- 31. Storage of lines: chain code
- 35. Raster storage
- 36. Hierarchical data structures
- 37. Quadtree algorithms, spatial indexes
- 38. Digital elevation models
- 39. TIN data model
- 42. Temporal and 3D databases
- 43. Database concepts I
- 44. Database concepts II

## OVERVIEW

This chapter reviews the main methods of GIS data capture and transfer and introduces key practical management issues.

It distinguishes between primary (direct measurement) and secondary (derivation from other sources) data capture for both raster and vector data types.

## LEARNING OBJECTIVES

- Describe data collection workflows;
- Understand the primary data capture techniques in remote sensing and surveying;
- Be familiar with the secondary data capture techniques of scanning, manual digitizing, vectorization, photogrammetry, and COGO feature construction;
- Understand the principles of data transfer, sources of digital geographic data, and geographic data formats;
- Analyze practical issues associated with managing data capture projects.

## KEY WORDS AND CONCEPTS

Data capture, data transfer, primary and secondary data sources, resolution (spatial, spectral and temporal), scanning, digitizing, error, photogrammetry, COGO, data transfer, data formats, ISO, CEN, OGC, OCR



## OUTLINE

- 9.1 Introduction
- 9.2 Primary geographic data capture
- 9.3 Secondary geographic data capture
- 9.4 Obtaining data from external sources (data transfer)
- 9.5 Capturing attribute data
- 9.6 Citizen-centric Web based Data Collection
- 9.7 Managing a data collection project

## CHAPTER SUMMARY

### 9.1 Introduction

In this chapter, data collection is split into *data capture* (direct data input) and *data transfer* (input of data from other systems).

- Two main types of data capture are
  - *Primary data sources* are those collected in digital format specifically for use in a GIS project.
  - *Secondary sources* are digital and analog datasets that were originally captured for another purpose and need to be converted into a suitable digital format for use in a GIS project.
- This chapter describes the data sources, techniques, and workflows involved in GIS data collection.
- The processes of data collection are also variously referred to as data capture, data automation, data conversion, data transfer, data translation, and digitizing.
- Table 9.2 shows a breakdown of costs for two typical client-server GIS implementations.
- Data collection is a time consuming, tedious, and expensive process.
- Typically it accounts for 15–50% of the total cost of a GIS project
- If staff costs are excluded from a GIS budget, then in cash expenditure terms data collection can be as much as 60–85% of costs.

#### 9.1.1 Data collection workflow

- Figure 9.1 shows the stages in data collection projects
- *Planning* includes establishing user requirements, garnering resources, and developing a project plan.

- *Preparation* involves obtaining data, redrafting poor-quality map sources, editing scanned map images, removing noise, setting up appropriate GIS hardware and software systems to accept data.
- *Digitizing and transfer* are the stages where the majority of the effort will be expended.
- *Editing and improvement* covers many techniques designed to validate data, as well as correct errors and improve quality.
- *Evaluation* is the process of identifying project successes and failures.

## 9.2 Primary geographic data capture

### 9.2.1 Raster data capture

- Remote sensing is a technique used to derive information about the physical, chemical, and biological properties of objects without direct physical contact
- Information is derived from measurements of the amount of electromagnetic radiation reflected, emitted, or scattered from objects.
- Figure 9.2 shows the spatial and temporal characteristics of commonly used remote sensing systems and their sensors
- *Resolution* is a key physical characteristic of remote sensing systems.
- *Spatial resolution* refers to the size of object that can be resolved and the most usual measure is the pixel size.
- *Spectral resolution* refers to the parts of the electromagnetic spectrum that are measured.
- *Temporal resolution*, or repeat cycle, describes the frequency with which images are collected for the same area.
- A paragraph describes SPOT imagery
- Aerial photography is equally important in medium- to large-scale projects
- Photographs are normally collected by analog optical cameras and later scanned
- Aerial Photographs are usually collected on an ad hoc basis
- Can provide stereo imagery for the extraction of digital elevation models
- Advantages are
  - Consistency of the data
  - Availability of systematic global coverage
  - Regular repeat cycles
- Disadvantages are
  - Resolution is often too coarse

- Many sensors are restricted by cloud cover

### 9.2.2 Vector data capture

- Two main branches are ground surveying and GPS
  - Distinction is increasing blurred

#### 9.2.2.1 Surveying

- Ground surveying is based on the principle that the 3-D location of any point can be determined by measuring angles and distances from other known points.
- Traditional equipment like transits and theodolites have been replaced by total stations that can measure both angles and distances to an accuracy of 1 mm
- Ground survey is a very time-consuming and expensive activity, but it is still the best way to obtain highly accurate point locations.
- Typically used for capturing buildings, land and property boundaries, manholes, and other objects that need to be located accurately.
- Also employed to obtain reference marks for use in other data capture projects

#### 9.2.2.2 LiDAR

- Relatively new technology that employs a scanning laser rangefinder to produce accurate topographic surveys
- Typically carried on a low-altitude aircraft that also has an inertial navigation system and a differential GPS to provide location.

## 9.3 Secondary geographic data capture

### 9.3.1 Raster data capture using scanners

Three main reasons to scan hardcopy media are

- Documents are scanned to reduce wear and tear, improve access, provide integrated database storage, and to index them geographically
- Film and paper maps, aerial photographs, and images are scanned and georeferenced so that they provide geographic context for other data
- Maps, aerial photographs and images are scanned prior to vectorization

### 9.3.2 Vector data capture

- Secondary vector data capture involves digitizing vector objects from maps and other geographic data sources.

#### 9.3.2.1 Heads-up digitizing and vectorization

- Vectorization is the process of converting raster data into vector data.

- The simplest way to create vectors from raster layers is to digitize vector objects manually straight off a computer screen using a mouse or digitizing cursor.
- Describes how automated vectorization is performed

#### 9.3.2.2 Measurement error

- Figure 9.10 presents some examples of human errors that are commonly introduced in the digitizing procedure including overshoots, undershoots, invalid polygons, and sliver polygons
- Discussion of how errors may arise by the use of rubbersheeting which assumes that spatial autocorrelation exists among errors

#### 9.3.2.3 Photogrammetry

- Is the science and technology of making measurements from pictures, aerial photographs, and images.
- Measurements are captured from overlapping pairs of photographs using stereo plotters.
- Figure 9.13 shows a typical workflow in digital photogrammetry
- Orientation and triangulation are fundamental photogrammetry processing tasks.
  - Orientation is the process of creating a stereo model suitable for viewing and extracting 3-D vector coordinates that describe geographic objects.
  - Triangulation (also called 'block adjustment') is used to assemble a collection of images into a single model so that accurate and consistent information can be obtained from large areas.
- Orthoimages are images corrected for variations in terrain using a DEM.
- Photogrammetry is a very cost-effective data capture technique that is sometimes the only practical method of obtaining detailed topographic data

#### 9.3.2.4 COGO data entry

- COGO is a contraction of the term *coordinate geometry*, a methodology for capturing and representing geographic data.
- COGO uses survey-style bearings and distances to define each part of an object
- COGO data are very precise measurements and are often regarded as the only legally acceptable definition of land parcels.

### 9.4 Obtaining data from external sources (data transfer)

A small selection of key data sources is listed in Table 9.3

The best way to find geographic data is to search the Internet

### 9.4.1 Geographic data formats

- One of the biggest problems with data obtained from external sources is that they can be encoded in many different formats.
- Many tools have been developed to move data between systems and to reuse data through open application programming interfaces (APIs).
- More than 25 organizations are involved in the standardization of various aspects of geographic data and geoprocessing
- ISO (International Standards Organization) is responsible for coordinating efforts through the work of technical committees TC 211 and 287
- In Europe, CEN (Comité Européen de Normalisation) is engaged in geographic standardization.
- OGC (Open Geospatial Consortium) is a group of vendors, academics, and users interested in the interoperability of geographic systems
- Geographic data translation software must address both syntactic and semantic translation issues.
- Syntactic translation involves converting specific digital symbols (letters and numbers) between systems.
- Semantic translation is concerned with converting the meaning inherent in geographic information.
- While the former is relatively simple to encode and decode, the latter is much more difficult and has seldom met with much success to date.

## 9.5 Capturing attribute data

- Attributes can be entered by direct data loggers, manual keyboard entry, optical character recognition (OCR) or, increasingly, voice recognition.
- An essential requirement for separate data entry is a common identifier (also called a key) that can be used to relate object geometry and attributes together following data capture

## 9.6 Citizen centric web based data collection

- Describes how a raft of new Web 2.0 technologies has enabled organizations and individual projects to use citizens to collect data across a wide variety of thematic and geographic areas

## 9.7 Managing a data collection project

- Most of the general principles for any GIS project apply to data collection: the need for a clearly articulated plan, adequate resources, appropriate funding, and sufficient time.
- A key decision facing managers of such projects is whether to pursue a strategy of incremental or very rapid collection.
- A further important decision is whether data collection should use in-house or external resources.

### ESSAY TOPICS

1. Distinguish between primary and secondary data and give examples of each. In what circumstances is this distinction difficult to maintain?
2. Why is data maintenance often a far more difficult and expensive activity than the initial data collection?
3. What do you understand by the terms 'active' and 'passive' satellite sensor systems and what are the relative advantages of each?
4. Why is it often necessary to scan paper documents for data entry into a GIS?
5. Describe the necessary steps in a workflow for manual digitizing using a semi-automatic digitizer. How and why does this process introduce 'error' into the database?
6. You are required to merge together in your GIS database digital cartographic data with some satellite imagery. What are the necessary steps in this process and the likely sources of difficulty?
7. How does national and international legislation on freedom of information and copyright affect the market for geospatial data?
8. What are the difficulties in translating between different data formats, and what software solutions have been suggested?
9. It is often suggested that in satellite imagery there is a trade off between spatial, spectral and temporal resolution. Outline and illustrate what is meant by these properties. To what extent do the data in Table 9.2 support this idea?
10. Describe the various ways by which 'error', defined as the difference between reality and our representation of it, can be introduced in the process of data collection and integration into a GIS.

## MULTIPLE CHOICE QUESTIONS (MCQ)

1. The table below lists examples of geographic data used in GIS. For each example, state whether it is usually (a) raster or vector and (b) a primary or a secondary source and (c) digital or analogue. In some cases your answer can be 'both':

Geographical data	Raster or vector?	Primary or secondary?	Digital or analogue?
Satellite imagery			
Aerial photography			
Printed maps			
Elevation models			
GPS measurements			
Survey measurements			
Place name data			
Census bureau records			

2. Rank the following satellite platforms in increasing order (i.e. least frequent = 1) of temporal resolution of their imagery:

Platform	Answer
Landsat	
SPOT	
Meteosat	
IKONOS	

3. For each of the four listed elements of the cost of a client-server GIS to serve 100 users, match it to its estimated percentage of the total costs of the system:

Cost element	ANSWER a, b, c or d	Choose from:
Hardware		a) 15.5
Software		b) 69.0
Data		c) 6.9
Staff		d) 8.6

4. If we exclude staff costs from a GIS budget, what is the likely maximum cash expenditure on data acquisition? Is it a) 15% b) 30% or c) 85% of the total budget?  
ANSWER ....

5. In the data collection cycle, place the following items in their logical sequence:

Process	Order (1) to (4)
Editing and improvement	
Planning	
Digitizing/transfer	
Planning	
Evaluation	

6. Which of the following methods of sensing the environment are remote and which are direct?

Method	Answer Remote or direct
Aerial Photography	
Satellite imagery	
Standard thermometer	
GPS	
pH meter	

7. You have a scanner capable of recording using 8 binary digit 'bit's for each pixel in its output. Which of the following hardcopy media would it be capable of recording accurately:

Medium	Answer YES or NO
Newspaper printed gray scale	
Color aerial photograph	
CAD drawing	
USGS 1:24000 Quad sheet (Figure 3.14)	



8. Match each of the listed earth observation systems to the spatial resolution of its data:

System	Answer (a)-(f)	Choose from:
GEOS (Viz)		a) 0.25 x 0.25m
Digital aerial photography		b) 5 x 5km
Landsat 4 MSS		c) 80 x 80m
Meteosat		d) 1 x 1 km
Landsat 4 TM		e) 30 x 30m
SPOT Panchromatic		f) 10 x 10m

9. You need to acquire data to populate a GIS using a variety of secondary data providers. In the table below, match each data type to the most probable source:

Type	Answer (a)-(f)	Possible providers
Lifestyle classifications		a) National census agency
Population		b) Military/National mapping agency
Toponymy		c) National mapping agency
Elevation		d) Commercial and military
Satellite imagery		e) National government
Administrative areas		f) Private specialist agency

10. Which of the following is the most essential in linking attribute and geographic components in a GIS database?
- Complete records
  - Metadata describing both
  - A key field
  - Lineage information

## CLASS AND INDIVIDUAL ACTIVITIES

- Use the web to discover what satellite imagery and digital aerial photography are available for a 20 x 20km area centered on your home or place of study. Make a table listing the spatial, spectral and temporal resolution of each source together with its ease of acquisition, georeferencing and other relevant metadata.

2. Table 9.4 lists a dozen data aggregation or indexing websites. Make a systematic visit to all to determine what data are available for your home area.
3. Choose either a high resolution aerial photograph or satellite image of a small area known to you and select a few polygonal objects that appear on it, such as woodland, field boundaries and the like. Use an appropriate tool to 'heads up' digitize the outline from the screen. Pixel co-ordinates can be obtained using a simple picture editing program such as Microsoft's Paint™. Make a list of the errors you have introduced into the data. Having obtained these picture values, consider how you would register them onto some known map projection. The next activity is a possible extension.
4. A critical and sometimes neglected step in many data integration exercises is the co-registration of the map and image data onto the same co-ordinate system. O'Sullivan and Unwin (2010, pages 221-324) provide more detail. Choose any suitable satellite image of an area for which you also have good fine scale mapping, and use the 'tick point' approach to co-register a series of points on the image onto the map co-ordinates. This exercise can be done using GIS software, but it is very instructive to follow all the steps using more basic tools such as a spreadsheet or statistical analysis program. It is probable that a simple affine transformation will not be adequate, and it is desirable that you select and co-register at least 20 ground control points. Some of the complications and considerations in taking this approach are detailed in the following papers:
  - Mather, P.M. (1995) Map-image registration using least-squares polynomials. *International Journal of Geographical Information Systems*, 9(5), 543-545.
  - Morad, M., Chalmers, A.I. and O'Regan, P. R. (1996) The role of root-mean-square error in geo-transformation of image in GIS. *International Journal of Geographical Information Systems*, 10(3), 347-353.
  - Unwin, D.J. & P.M.Mather (1998) Selecting and Using Ground Control Points in Image Rectification and Registration, *Geographical Systems*, 5(3), 239-260.
5. This graphical exercise is intended to help fix ideas about co-ordinate transformation in integrating digitizer data into a GIS. First study the outline of affine transformation in O'Sullivan and Unwin (2010, pages 322-324). To do the exercise, you will need some lined graph paper, some tracing paper (ideally transparent lined graph paper), a pencil and, perhaps, a calculator or spreadsheet.
  - a) Create a grid on your graph paper with X and Y axes each going from 0 to 100.
  - b) On this grid mark eight randomly located points and read off their (x, y) co-ordinates.
  - c) Use the tracing paper to prepare an identical set of axes, but do not mark any points on it.

- d) Place the transparent grid on your original one with its origin exactly on the origin of the original one and rotate it by a small known angle (say  $15^\circ$ ).
- e) Next shift (translate) the origin by a known small amount, and then mark on the transparent paper the positions of your eight points.
- f) Read off the co-ordinates of the eight points in this new system.

If mathematically inclined, compute and use the affine transformation matrix for this operation.

*If you have the time and resources, then there is no better way to introduce the trials and tribulations of data collection than by doing it yourself in the field. This, and the next two activities, outline possible field projects appropriate to a GIS class. They can be adapted to suit almost any disciplinary background and educational level.*

6. Random points survey of attribute data. Divide the class into small teams and ask each to generate a small number of randomly selected points in a small area of country they can visit in a day's field work. It helps, and introduces the notorious traveling salesman problem, if you ask them to decide on a minimum distance route that links the points. Students then visit the points, recording a series of attributes that characterize the environment in that locality. Suitable attributes might include soil type, slope and aspect, and land use. Back at base merge all these data and use them to develop, for example, estimates of the actual land use and soil type proportions, the overall frequency distributions of slope and aspect. This can be extended to use suitable statistical analyses to examine interactions between these variables such as the impact of slope on land use. Taking it further, it is very useful to have available a GIS for the same area with an 'official' version of these same attributes taken from standard map and imagery sources. There is a useful resource data base of similar projects at the GEES Resource Database: <http://www.gees.ac.uk/db/>
7. Surveying by leveling to establish the z co-ordinate. Although use of a simple device such as an Abney or 'Quickset' level is a long way from current professional survey practice, asking students to undertake a simple leveling transect along a clearly defined feature (such as a beach) is a good way to introduce key ideas in GIS data collection, especially measurement error and the implicit discretization introduced by selection of the stations used.
8. The third exercise is equally far from current survey practice, but has similar educational objectives. In small student teams, ask them to use simple plane table survey with basic table and tripod, alidade and ranging rods to map a small feature, such as the boundary

of a field or short reach of a river. For maximum effect, students should have to handle the errors in their survey. A useful extension is to use basic and/or differential GPS to capture co-ordinates of the same points used in the survey. Suitable basic equipment may be hard to come by in these modern times but is manufactured for educational use. You can, of course, do exactly the same exercise using any more modern surveying equipment that is appropriate and available.

9. Organize a debate around the motion that 'This house believes that geographic data should be freely available to all the citizens of country x'. See Gold *et al.* (1992), Chapter 5 at [www2.glos.ac.uk/gdn/gold/](http://www2.glos.ac.uk/gdn/gold/) for the usual organizational details.
10. Another suitably contentious topic for debate is that of the need for mandated standards in geographic data. A suitable motion is that 'This house believes that standards in GI data are a bad idea'. A very good place to find the case for standards is at the UK Association for Geographic Information (AGI) website: [www.agi.org.uk](http://www.agi.org.uk). As detailed in the text, page 214, similar lobby organization materials will be found at: [www.opengeospatial.org](http://www.opengeospatial.org). As an example of the problems that adherence to a standard can generate, an instructive case is the British Standard BS 7666 on street addresses. There are many others.

## FURTHER READINGS

Hohl P. (ed) 1997 *GIS Data Conversion: Strategies, Techniques and Management*. Santa Fe, NM: OnWord Press.

Jones C. 1997 *Geographic Information Systems and Computer Cartography*. Reading, MA: Addison-Wesley Longman.

Lillesand T.M., Kiefer R.W. and Chipman R.W. 2003 *Remote Sensing and Image Interpretation* (5th edn). Hoboken, NJ: Wiley.

Paine D.P. and Kiser J.D. 2003 *Aerial Photography and Image Interpretation* (2nd edn). Hoboken, NJ: Wiley.

Walford N. 2002 *Geographical Data: Characteristics and Sources*. Hoboken, NJ: Wiley.

Jensen, J.R. 2007. *Remote Sensing of the Environment: An Earth Resource Perspective* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

## RELATED READINGS

Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, NJ: Wiley.

31. Encoding and validating data from maps and images, I Dowman
32. Digital remotely-sensed data and their characteristics, M Barnsley
33. Using GPS for GIS data capture, A Lange and C Gilbert

Maguire D.J., Goodchild M.F. and Rhind D.W. (eds) 1991 *Geographical Information Systems: Principles and applications*. Harlow, UK: Longman (text available online at [www.wiley.co.uk/gis/volumes.html](http://www.wiley.co.uk/gis/volumes.html)).

17. GIS data capture hardware and software, M J Jackson and P A Woodsford, pp. 239-49
34. Spatial data exchange and standardization, S C Guphill, pp. 515-30

## ONLINE RESOURCES

NCGIA Core Curriculum in GIScience, 2000 ([www.ncgia.ucsb.edu/giscc](http://www.ncgia.ucsb.edu/giscc))

2.9.2. [Natural Resources Data](#) (090), Peter Schut

2.9.2.1. [Soil Data for GIS](#) (091), Peter Schut

NCGIA Core Curriculum in GIS, 1990 ([www.ncgia.ucsb.edu/pubs/core.html](http://www.ncgia.ucsb.edu/pubs/core.html))

7. Data input
8. Socio-economic data
9. Environmental data
66. Database creation

# Creating and Maintaining Geographic Databases

---

## OVERVIEW

- After people, the database is arguably the most important part of a GIS because of the costs of collection and maintenance, and because the database forms the basis of all queries, analysis, and decision making.
- Today, virtually all large GIS implementations store data in a database management system (DBMS), a specialist piece of software designed to handle multi-user access to an integrated set of data.
- Databases need to be designed with great care, and to be structured and indexed to provide efficient query and transaction performance.
- A comprehensive security and transactional access model is necessary to ensure that multiple users can access the database at the same time.
- On-going maintenance is also an essential, but very resource-intensive, activity.

## LEARNING OBJECTIVES

**By the end of this chapter students should:**

- **Understand the role of database management systems in GIS;**
- **Recognize structured query language (SQL) statements;**
- **Understand the key geographic database data types and functions;**
- **Be familiar with the stages of geographic database design;**
- **Understand the key techniques for structuring geographic information, specifically creating topology and indexing;**
- **Understand the issues associated with multi-user editing and versioning.**

## KEY WORDS AND CONCEPTS

DBMS, (RDBMS, ODBMS, ORDBMS), parsers, middleware, object classes, database tables, keys, normal forms, SQL, SQL/MM, database design, indexes, B-tree indexes, grid indexes, quadtree indexes, R-tree indexes, database editing and update, transactions, long transactions, versioning

## OVERVIEW

- 10.1 Introduction
- 10.2 Database management systems
- 10.3 Storing data in DBMS tables
- 10.4 SQL
- 10.5 Geographic database types and functions
- 10.6 Geographic database design
- 10.7 Structuring geographic information
- 10.8 Editing and data maintenance
- 10.9 Multi-user editing of continuous databases
- 10.10 Conclusion

## CHAPTER SUMMARY

### 10.1 Introduction

- A database can be thought of as an integrated set of data on a particular subject.
- Geographic databases are simply databases containing geographic data for a particular area and subject.
- Lists the advantages of the database approach to storing geographic data over traditional file-based datasets including reducing redundancy, decreasing costs, allowing multiple applications, transfer of knowledge, data sharing, security and standards, concurrent users
- Disadvantages include cost, complexity, single user performance decreased
- Describes how to create and maintain geographic databases, and the concepts, tools, and techniques that are available to manage geographic data in databases.

### 10.2 Database management systems

- A DBMS is a software application designed to organize the efficient and effective storage and access of data.

- Briefly outlines the capabilities of DBMS which include a data model, a data load capability, indexes, a query language, security, controlled update, backup and recovery, database administration tools, applications and APIs
- This list of DBMS capabilities is very attractive to GIS users and so, not surprisingly, virtually all large GIS databases are based on DBMS technology.

### 10.2.1 Types of DBMS

- Three main types of DBMS are available to GIS users today: relational (RDBMS), object (ODBMS), and object-relational (ORDBMS).
- A relational database comprises a set of tables, each a two-dimensional list (or array) of records containing attributes about the objects under study.
- Object database management systems (ODBMS) were initially designed to address weaknesses of RDBMS, including the inability to store complete objects directly in the database (both object state and behavior), poor performance for many types of geographic query
- ODBMS have not proven to be as commercially successful as some predicted because of the massive installed base of RDBMS. Thus appeared...
- Hybrid object-relational DBMS (ORDBMS) can be thought of as an RDBMS engine with an extensibility framework for handling objects.
- The ideal geographic ORDBMS is one that has been extended to support geographic object types and functions through the addition of a geographic query parser, a geographic query optimizer, a geographic query language, multidimensional indexing services, storage management for large files, long transaction services, replication services

### 10.2.2 Geographic DBMS extensions

- Two of the commercial DBMS vendors have released spatial database extensions to their standard ORDBMS products
  - IBM – DB2 Spatial Extender and Informix Spatial Datablade
  - Oracle Spatial
  - spatial capabilities in the core of Microsoft SQLServer
  - Opensource DBMS PostgreSQL has also been extended with spatial types and functions (PostGIS).
- None is a complete GIS software system
- Focus is on data storage retrieval and management
- Technical box 10.1 details Oracle Spatial



## 10.3 Storing data in DBMS tables

- The lowest level of user interaction with a geographic database is usually the *object class* (also called a layer or feature class), which is an organized collection of data on a particular theme
- Object classes are stored in a standard database *table*, a two-dimensional array of rows and columns.
  - Rows contain objects (*instances* of object classes)
  - Columns contain object properties or attributes
- The data stored at individual row, column intersections are usually referred to as values.
- Geographic database tables are distinguished from non-geographic tables by the presence of a geometry column (often called the shape column).
- To save space and improve performance, the actual coordinate values may be stored in a highly compressed binary form.
- Tables are joined together using common row/column values or *keys*.
- Following joins, all tables can be treated as a single table
- Lists Codd's five principles for the efficient and effective design of tables and introduces the concept of *normal forms*
- Normal forms improve the simplicity and stability of a database and reduce redundancy of tables by splitting them into sub-tables that are re-joined at query time
- Notes that large tables common in geographic applications leads to tendency for non-normalized table designs in GIS
- Includes a worked example of normalization of a simple land parcel tax assessment table

## 10.4 SQL

- The standard database query language adopted by virtually all mainstream databases is SQL (Structured or Standard Query Language: ISO Standard ISO/IEC 9075).
- May be used directly via command line, compiled in a general purpose programming language or via a GUI
- The third major revision of SQL (SQL 3) which came out in 2004 defines spatial types and functions as part of a multi-media extension called *SQL/MM*.
- There are three key types of SQL statements:
- DDL (data definition language) used to create, alter and delete relational database structures

- DML (data manipulation language) used to retrieve and manipulate data
- DCL (data control language) handle authorization and access
- Text briefly walks through simple examples of the first two of these

## 10.5 Geographic database types and functions

- Working together, ISO and OGC have defined the core geographic types and functions to be used in a DBMS and accessed using the SQL language.
- Figure 10.5 shows the geometry class hierarchy
- There are nine methods for testing spatial relationships between these geometric objects.
- Each takes as input two geometries and evaluates whether the relationship is true or not.
- Figure 10.6 illustrates two examples
- The full set of Boolean operators to test the spatial relationships between geometries is: Equals, Disjoint, Intersects, Touches, Crosses, Within, Contains, Overlaps, Relate
- Seven methods support spatial analysis on these geometries: Distance, Buffer, ConvexHull, Intersection, Union, Difference, SymDifference

## 10.6 Geographic database design

### 10.6.1 The database design process

- All GIS and DBMS packages have their own core data model that defines the object types and relationships that can be used in an application and which drive how data types will be implemented and accessed and how more advanced types of feature types and relationships are created
- Database design involves the creation of conceptual, logical, and physical models in the six practical steps shown in Figure 10.9
- The next sections summarize each of these steps and their products

#### 10.6.1.1 Conceptual model

- Model the user's view
- Define objects and their relationships
- Select geographic representation

#### 10.6.1.2 Logical model

- Match to geographic database types
- Organize geographic database structure

### 10.6.1.3 Physical model

- Define database schema

## 10.7 Structuring geographic information

### 10.7.1 Topology creation

- Two database-oriented approaches have emerged in recent years for storing and managing topology: Normalized and Physical.
- Normalized Model focuses on the storage of an arc-node data structure
  - Is said to be normalized because each object is decomposed into individual topological primitives for storage in a database and then subsequent reassembly when a query is posed.
  - Normalized approach advantages are: similarities to the familiar arc-node concept, geometry is only stored once, access can be via an SQL API
  - Normalized approach disadvantages are: query performance suffers, standard referential integrity rules in DBMS have no provision for the complex topological relationships, updates are problematic due to cascading effects
- Physical Model topological primitives are not stored in the database and the entire geometry is stored together for each object.
  - Only other things required to be stored are the specific set of topology rules
  - Topological relationships are then computed on-the-fly whenever they are required by client applications.
  - Requires an external client or middle-tier application for validating topological integrity
- Figures 10.10 and 10.11 illustrate these forms for the same geometry

### 10.7.2 Indexing

- Geographic databases tend to be very large and geographic queries computationally expensive
- Indexes speed up searching by allowing random instead of sequential access.
- A database index is, conceptually speaking, an ordered list derived from the data in a table.
- Figure 10.12 and the related paragraphs explain a simple example of the standard DBMS one-dimensional B-tree (Balanced Tree) index that is found in most major commercial DBMS.
- Since these 1D indexes are very poor at indexing geographic objects, several geographic indexing techniques have been developed

### 10.7.2.1 Grid index

- A grid index can be thought of as a regular mesh placed over a layer of geographic objects.

### 10.7.2.2 Quadtree indexes

- Quadtrees are data structures used for both indexing and compressing geographic database layers, although the discussion here relates only to indexing.
- Several paragraphs and diagrams explain quadtrees

### 10.7.2.3 R-tree indexes

- R-trees group objects using a rectangular approximation of their location called a minimum bounding rectangle (MBR) or minimum enclosing rectangle (see Box 10.2).
- Groups of point, line, or polygon objects are indexed based on their MBR.
- Technical Box 10.2 explains that a MBR essentially defines the smallest box whose sides are parallel to the axes of the coordinate system that encloses a set of one or more geographic objects.

## 10.8 Editing and data maintenance

- *Editing* is the process of making changes to a geographic database by adding new objects or changing existing objects as part of data load or database update and maintenance operations.
- A database *update* is any change to the geometry and/or attributes of one or more objects or any change to the database schema.
- Contemporary GIS come equipped with an extensive array of tools for creating and editing geographic object geometries and attributes.

## 10.9 Multi-user editing of continuous databases

- It is relatively easy to provide multiple users with concurrent read and query access to a continuous shared database, but more difficult to deal with conflicts and avoid potential database corruption when multiple users want write (update) access.

### 10.9.1 Transactions

- A group of edits to a database is referred to as a *transaction*.
- Many geographic transactions extend to hours, weeks, and months, and are called *long transactions*

### 10.9.2 Versioning

- Identifies two kinds of versioning

- *Pessimistic locking* locks out all but one user during an update operation
- *Optimistic* versioning allows multiple users to update at the same time
- Versioning addresses the locking concurrency problem and supports alternative representations of the same objects in the database.
- Within a versioned database, the different database versions are logical copies of their parents (base tables). Only the modifications are stored in the database
- *Branching* and *merging* occur as the two versions are managed
- Includes an example illustrating how versioning works

## 10.10 Conclusion

DBMS require a database administrator (DBA) to control database structure and security, and to tune the database to achieve maximum performance.

## ESSAY QUESTIONS

1. What are the main advantages and disadvantages of storing geographic data in a DBMS?
2. Is SQL a good language for querying geographic databases?
3. Why are there multiple methods of indexing geographic databases? Compare and contrast at least two indexing methods.
4. What are the main capabilities that are required of a database management system?
5. 'All right in theory, no good in practice'. To what extent does this statement summarize the state of play in object oriented database work?
6. Outline Codd's five principles of good relational database design. Why are geographic table designs often left in an un-normalized form?
7. What are the main conceptual stages in database design, and what is accomplished in each?
8. Outline what is meant by the term minimum bounding rectangle and explain why it can be useful to compute and store it in a GIS database.
9. What do you understand by the physical and normalized models for storing topology in a geographic database and how do these differ?
10. What are the issues that have to be tackled in the management of geographic databases open to simultaneous use by many users?

## MULTIPLE CHOICE QUESTIONS (MCQ)

1. What are the two main reasons why a geographic database is a critical part of an operational GIS?
  - a. The cost of its creation
  - b. The impact it has on analysis
  - c. Bringing all the data together
  - d. It is large and complex
2. For each of the listed items, state whether it is as good or a bad feature of the use of a database management system:

Item	Good or Bad?
Single user performance	
Maintenance costs	
Application independence	
Data sharing	
Management complexity	
Transfer between applications	
Multi-user capability	
Cost of acquisition	
Maintenance cost	
Data redundancy	

3. Write out the full version of the following acronyms:
  - a. ORDBMS .....
  - b. RDBMS .....
  - c. ODBMS .....
  - d. DBMS .....
4. Which of the two following SELECT queries of a geographic database is likely to be fastest in execution: (a) or (b)?
  - a. Select all households living within 100m of a railway line, then from this group select all those who do not have a motor car
  - b. Select all households without a motor car, then from this group select all those living within 100m of a railway line

5. Classify each of the following proprietary products as ORDBMS, geographic middleware or RDBMS:

Product	Nature
Oracle	
Informix Dynamic Server	
ArcSDE	
IBM DB2	
Geomedia Transaction Server	
Versant	

6. Write out the full version of the following acronyms:

- DDL .....
- DCL .....
- SQL .....

7. Order the following steps into correct sequence in the creation of a geographic database.  
Write (1) to (6) as appropriate:

Stage	Order (1)-(6)
Geographic database structuring	
Geographic representation	
Object and relationship modeling	
Geographic database typing	
User view analysis	
Database schema definition	

8. Complete the verbal equation: OBJECT = STATE + .....?
9. Which of the following principles are NOT among those suggested by Codd as important in all databases?
- Only one value per cell
  - All cells have numbers
  - All values in a column are about the same subject
  - Each row is unique
  - Every table must be normalized
  - Rows can be in any order

- g. Columns can be in any order
10. Which of the following grid dimensions would not allow the generation of a grid quadtree?
- a. 16 by 16
  - b. 100 by 100
  - c. 256 by 256
  - d. 1024 by 1024

### CLASS AND INDIVIDUAL ACTIVITIES

1. Identify a geographic database with multiple layers and draw a diagram showing the tables and the relationships between them. Which are the primary keys, and which keys are used to join tables? Does the database have a good relational design?
2. Extract from a topographic map (or section of a topographic map) a pattern of point objects of interest to you. According to the mapping scale, these might be towns, public houses ('pubs' or bars), telephone boxes, railway stations and so on, but do not attempt to use more than, say, 15-30. Create a layer of point objects from this source. Using the method outlined in section 10.7.2.2 and Figure 10.14 create the point quadtree for these data.
3. The classic quadtree index is a compact way of storing some raster grids. Find a raster representation of a polygonal object. For simplicity, restrict your map to less than, say, four objects. Now create a raster grid of size 16 by 16 the area and code each cell '1' if it is more than 50% covered by one of these areas. Using the method outlined in the text, create the grid quadtree for these data. How effectively does the quadtree compress these data? How would you scan the quadtree to access any specific datum?
4. Finally, in this sequence of three, find or create a layer consisting of a few area objects. Using the methods outlined in section 10.7.2 create:
  - a. A simple grid index
  - b. a region quadtree
  - c. an R-tree index using the minimum bounding rectangles
5. How do these indexing devices compare?
6. Section 10.5 lists a set of nine Boolean operators to test the spatial relationships between various geometries. Figure 10.6 (a) and (b) illustrates the 'contains' and 'touch' relationships. Complete this figure by sketching the remaining seven relationships for all possible pairings of point, line and area base geometries. Note that some of the operations will be impossible in some geometries.



7. Find out as much as you can about extensions to SQL that enable basic geographic operations. Make a list of the key additional features and show why each is necessary. A good place to start is R.V. Kothuri, A Godfrind and E. Beinat (2004) *Pro Oracle Spatial*, Apress: Berkeley, CA, but the paper by M. Egenhofer (1992) Why not SQL? *International Journal of Geographical Information Systems*, 6:71-85 is a less daunting.
8. Figure 10.7 illustrates four geometric spatial analysis operations (buffer, convex hull, intersection and difference) for each geometric type. In each of the twelve illustrated cases suggest how the operation might be used in a real analysis.
9. Following the sequence illustrated in Figure 10.3, and Section 10.6, devise a database structure to store information about a large collection of landscape photographs to include subject, a classification of the subject, place, locational co-ordinates, direction of camera and related information. Having devised the structure, write out SQL commands to CREATE the table and a selection of possible queries involving joined.
10. [Frantisek Brabec](#) and [Hanan Samet](#) at the University of Maryland have devised a wonderfully comprehensive website with interactive demonstrations of numerous spatial indexing techniques and associated algorithms for deletion and insertion of new data. The applets require the Java Development Kit 1.1 and are based on algorithms published in Samet's classic books: H. Samet (1990) *Applications of Spatial Data Structures: Computer Graphics, Image Processing, and GIS*, Addison-Wesley, Reading, MA, and (1990) *The Design and Analysis of Spatial Data Structures*, Addison-Wesley, Reading, MA. Visit the website at [www.cs.umd.edu/~brabec/quadtree](http://www.cs.umd.edu/~brabec/quadtree) and use this resource both to cement and to extend your knowledge of indexing. Pay particular attention to the algorithms for editing the indexes.
11. A key issue in almost all this chapter has been industry attempts to handle the geometric/spatial components of geospatial data using standard database technologies. This has obvious commercial and related advantages, but is arguably not what practitioners of the science of handling such data need if progress is to be made in, for example, handling time varying or three- and more dimensional representations. Organize a debate around the proposal that 'This house believes that use of standard relational database technology is a major impediment to research in representing geographic information'. Use the text for arguments against the motion and for arguments for it almost every essay in the book by Fisher, P.F. and D.J. Unwin (eds, 2005) *Re-presenting GIS*, Wiley: Chichester. The readings also list papers by Egenhofer et al (1999) and an early work by Worboys et al (1990) that have relevant materials. It will also help the 'for' case if the debating team visits websites that describe experimental GIS, such as DOGIS, at [www.cse.ohio-state.edu/~prasun/publications/conf/dogis.pdf](http://www.cse.ohio-state.edu/~prasun/publications/conf/dogis.pdf). There are others.

## FURTHER READING

Date C.J. 2003 *Introduction to Database Systems* (8th edn). Reading, MA: Addison-Wesley.  
Egenhofer, M., Glasgow, J., Gunther, O., Herring, J.R. and D. J. Peuquet 1999 Progress in computational methods for representing geographic objects. *International Journal of Geographical Information Science*, 13: 775-796 (available at [www.spatial.maine.edu/~max/PiCM.pdf](http://www.spatial.maine.edu/~max/PiCM.pdf))

Is a very good overview of database issues in GIS

Hoel E., Menon S. and Morehouse S. 2003 'Building a robust relational implementation of topology.' In Hadzilacos T., Manolopoulos Y., Roddick J.F. and Theodoridis Y. (eds) *Advances in Spatial and Temporal Databases. Proceedings of 8th International Symposium, SSTD 2003 Lecture Notes in Computer Science*, Vol. 2750.

OGC 1999 OpenGIS simple features specification for SQL, Revision 1.1. Available at [www.opengis.org](http://www.opengis.org)

Samet H. 1990 *The Design and Analysis of Spatial Data Structures*. Reading, MA: Addison-Wesley.

Worboys M.F. and Duckham M. 2004 *GIS: A Computing Perspective* (2nd edn). Boca Raton, FL: CRC Press.

Worboys, M.F., Hearnshaw, H.M. and D.J. Maguire 1990 Object-oriented data modelling for spatial databases. *International Journal of Geographical Information Systems*, 4(4): 369-383  
Very much the manifesto in favor of object orientation written by a team from UK's ESRC Regional Research Laboratory.

Zeiler M. 1999 *Modeling our World: The ESRI Guide to Geodatabase Design*. Redlands, CA: ESRI Press.

## RELATED READING

Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, NJ: Wiley.

27. Spatial access methods, P van Oosterom, pp. 385-400

Maguire D.J., Goodchild M.F. and Rhind D.W. (eds) 1991 *Geographical Information Systems: Principles and Applications*. Harlow, UK: Longman (text available online [www.wiley.co.uk/gis/volumes.html](http://www.wiley.co.uk/gis/volumes.html)).

18. Database management systems, R G Healey, pp. 251-267

## ONLINE RESOURCES

ESRI Virtual Campus course, *Turning Data into Information* by Paul Longley, Michael Goodchild, David Maguire, and David Rhind ([campus.esri.com](http://campus.esri.com) )

Module 3: Query and Measurement

Section 10.4, Module 3: Query and Measurement

Unit: Advanced queries,

Sub-unit: Tabular queries

NCGIA Core Curriculum in GIScience, 2000 ([www.ncgia.ucsb.edu/giscc](http://www.ncgia.ucsb.edu/giscc))

3.1. [Making it work](#) (136), Hugh Calkins and others

NCGIA Core Curriculum in GIS, 1990 ([www.ncgia.ucsb.edu/pubs/core.html](http://www.ncgia.ucsb.edu/pubs/core.html))

60. System planning overview

66. Database creation

67. Implementation issues

68. Implementation strategies for large organizations

## OVERVIEW

- This chapter describes current and capabilities in distributed GIS, and looks to a future in which GIS is increasingly mobile and available everywhere. The GeoWeb is a vision for the future.
- It is organized into three major sections, dealing with distributed data, distributed users, and distributed software.

## LEARNING OBJECTIVES

**After studying this chapter, students will understand**

- **How the parts of GIS can be distributed instead of centralized.**
- **Geoportals, and the standards and protocols that allow remotely stored data to be discovered and accessed.**
- **The technologies that support real-time acquisition and distribution of geographic information.**
- **The service-oriented architectures and mashups that combine GIS services from different Web sites. The capabilities of mobile devices, including mobile phones and wearable computers.**
- **The concepts of augmented and virtual reality.**

## KEYWORDS AND CONCEPTS

Interoperate, Open Geospatial Consortium, cyberinfrastructure, object-level metadata, FGDC, CSDGM, geolibraries, collection-level metadata, virtual and augmented reality, location-based services, GIServices

## OUTLINE

- 11.1 Introduction
- 11.2 Distributing Data
- 11.3 The Mobile User
- 11.4 Distributing the Software: GIServices
- 11.5 Prospects

## CHAPTER SUMMARY

### 11.1 Introduction

- This chapter describes how the network, to which almost all computers are now connected, has enabled a new vision of *distributed* GIS, in which the component parts no longer need to be co-located.
- There are four distinct locations of significance to distributed GIS:
  - the location of the user and the user interface, *U*
  - the location of the data, *D*
  - the location where the data are processed, *P*
  - the area that is the focus of the project, the subject location, *S*
- Critical to distributed GIS are the standards and specifications that make it possible for devices, data, and processes to interoperate
- Some are universal like ASCII and XML
- Others are specific to GIS and were developed through OGC
- Open Geospatial Consortium (OGC) is an organization set up to promote openness and interoperability in GIS.
- Successes include simple feature specification, Geography Markup Language (GML), and specifications for Web services
- Mentions the notion that today's computing is extended beyond the desktop
  - Cyberinfrastructure describes a new approach to the conduct of science, relying on high-speed networks, massive processors, and distributed networks of sensors and data archives
  - Grid computing is a generic term for such a fully integrated world-wide network of computers and data.
  - GPS data of real-time locations increasingly fed to the Web

- RFID allows the tracking of objects that have been implanted or tagged with small sensors - widely used in retailing, livestock management, and building construction
- Makes the point that Mashup (see Chapter 5), and the linking of services in general, is a key concept of the GeoWeb

## 11.2 Distributing the Data

- Many private citizens have become involved in the distributed creation and dissemination of geographic information, in a process known as volunteered geographic information (VGI).
- The vision of distributed GIS goes well beyond the ability to access and retrieve remotely located data, however, because it includes the concepts of search, discovery, and assessment
- Three concepts are important in this respect: object-level metadata, geolibraries, and collection-level metadata.

### 11.2.1 Object-Level Metadata

- *Object-level* metadata (OLM) describe the contents of a single dataset by providing essential documentation.
- Need OLM for many reasons:
  - to automate the process of search and discovery over archives
  - to assess the fitness of a discovered dataset for a given use
  - to provide the information needed to handle the dataset effectively
  - to provide useful information on the dataset's contents
- It is easy for the complete description of a dataset to generate a greater volume of information than the actual contents.
- OLM are expensive to generate because they represent a level of understanding of the data that is difficult to assemble, and requires a high level of professional expertise.
- The most widely used standard for OLM is the US Federal Geographic Data Committee's (FGDC) Content Standards for Digital Geospatial Metadata (CSDGM) first published in 1993 and now the basis for many other standards worldwide.
  - Box 11.2 lists some of its major features
- Alternatively, the Dublin Core (see Box 11.3) is the outcome of an effort to find the minimum set of properties needed to support search and discovery for datasets in general
  - treats both space and time as instances of a single property, *coverage*

- The principle of establishing a minimum set of properties is sharply distinct from the design of CSGDM, which was oriented more toward the capture of all knowable and potentially important properties of geographic datasets.

### 11.2.2 Geolibraries and Geoportals

- The term geolibrary has been coined to describe digital libraries that can be searched for information about any user-defined geographic location.
- Since both the spatial and temporal dimensions are continuous, it is impossible to capture them in a single property analogous to author that can then be sorted numerically or alphabetically.
  - Searching based on location or time is possible in a digital system
- A geoportal is defined as a single point of entry to a distributed collection of geolibraries

## 11.3 The Mobile User

- Briefly summarizes the emergence of mobile computing

### 11.3.1 Virtual reality and augmented reality

- Research environments that place what humans normally gather through their senses into a database are termed virtual realities
- In most GIS applications only one of the senses is used, sight, and the view is usually from above rather than from the ground
- More elaborate VR systems are capable of immersing the user, by presenting the contents of a database in a three-dimensional environment, using special eyeglasses or by projecting information onto walls surrounding the user, and effectively transporting the user into the environment represented in the database.
- The idea of combining information from a database with information derived directly through the senses is termed augmented reality, or AR.
- Discusses two examples of AR systems
- Gollledge's AR for helping visually impaired people perform the simple task of navigation
- Feiner's AR that superimposes historic images and other information directly on the user's field of view

### 11.3.2 Location-based services

- A location-based service (LBS) is defined as an information service provided by a device that knows where it is and is capable of modifying the information it provides based on that knowledge.
- GPS receivers and cellphones are discussed as means of determining a user's location
- Several applications of LBS are described

### 11.3.3 Issues in mobile GIS

- GIS in the field or 'on the road' is very different from GIS in the office.
  - The location of the user is important, and directly relevant to the application.
  - The field environment may make certain kinds of interaction impractical
- Battery (or other wireless energy) technology has not advanced as rapidly as other components of mobile systems, such as processors and storage devices.
- Wireless communication using WiFi is still problematic

## 11.4 Distributing the software: GIS services

- This section addresses distributed processing, the notion that the actual operations of GIS might be provided from remote sites
- A GIService is defined as a program executed at a remote site that performs some specific GIS task
- Asks how would a GIService pay for itself
- The characteristics that would make a GIS function suitable for offering as a service appear to be
  - Reliance on a database that must be updated frequently, and is too expensive for the average user to acquire -- includes geocoding, wayfinding, and gazetteer services
  - Reliance on GIS operations that are complex and can be performed better by a specialized service than by a generic GIS.

### 11.4.1 Service-Oriented Architecture

- An illustration of the capabilities of distributed GIS and the GeoWeb technology is given for emergency management.



## 11.5 Prospects

### ESSAY TOPICS

1. Define the U, D, P and S of distributed GIS and give examples of applications that 'locate' them in different (distributed) places.
2. Why does the vision of distributed, hand-held, and wearable GIS rely heavily on the existence of mandated standards, and where in the entire process are these necessary?
3. If 'the network' really is 'the machine' of the future, how is grid computing likely to change GIS?
4. What do you understand by the term 'Geoportal', and what services should one offer?
5. How do virtual reality applications in GIS differ from advanced geovisualization (see Chapter 13)?
6. Outline, with examples, what is meant by the term 'object-level metadata' and explain why it is important in geographic resource discovery.
7. Attempt and illustrate a classification of 'location-aware' computing applications.
8. How and why is GIS in the field different from GIS in the office, and what are its major present-day limitations?
9. Outline the business logic behind the development of geographic information services and suggest how companies providing them can develop an income stream.
10. George Orwell's 1949 vision of the world in *Nineteen Eighty Four* featured a Big Brother who was always watching. To what extent, twenty years after the predicted date, has Orwell's nightmare been realized?

### MULTIPLE CHOICE QUESTIONS (MCQ)

1. What do most of the world's computers do most of the time? Tick one answer.
  - a. Word process
  - b. Perform GIS
  - c. Nothing
  - d. Run games
  - e. Analyse non-geographic databases

2. Roughly, how many CDs are need to store a petabyte of data? Circle your answer:

1	10	100	1000	10,000	1,000,000	10,000,000
---	----	-----	------	--------	-----------	------------

3. Which of the following are examples of location-based computing?

Activity	Yes or no?
Yell/Yellow Pages	
Onstar	
PlayUnderCover	
Virtual London	
Wearable GIS	

4. Expand the following acronyms:

Acronym	Stands for ...
OLM	
CSDGM	
GML	
CLM	

5. According to the T-Mobile provider, in 2004 how many wireless Internet 'hot spots' were within 1 mile of 1600 Pennsylvania Ave., Washington DC? Circle your answer:

10	37	78	200	1000
----	----	----	-----	------

6. Which of the following devices is normally 'location aware'? Tick all as appropriate:

Device	Yes or no?
Personal digital assistant	
Cell phone/Mobile phone	
Laptop	
WiFi enabled laptop	
GPS	

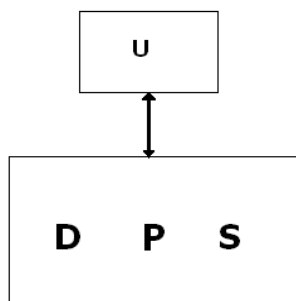
7. Which two of the following characteristics of a business most make it a candidate for outsourcing to a GI service company?

Characteristic	Yes or no?
Reliance on a large expensive database	
Use of complex GI operations	
Need for frequent database updates	
Large number of concurrent users	
Lack of in-house expertise	

8. Which of the following statements best describes 'grid computing'?
  - a. Sharing data between several Internet hosts
  - b. Autonomous communicating computers
  - c. Using the Internet to allow distributed computation
  - d. Parallel processing
9. Which of the following is the major current limitation on mobile computing?
  - a. Lack of software
  - b. Limited access
  - c. Noisy signals
  - d. Lack of security

## ACTIVITIES

1. The figure shows the location of the User, Data, Processor, and Subject in a desk top GIS. Draw similar diagrams to illustrate possible alternative configurations in a distributed GIS and attempt to find an example of each.



2. Visit [www.opengeospatial.org](http://www.opengeospatial.org) and use it to identify the open standards and initiatives relevant to the development of distributed GIS and LBS. To add structure to this activity, create a list of each and write down a succinct description of what it entails.

Peng Z.-H. and Tsou M.-H. (2003) provide an academic and practical guide to these materials.

3. The Dublin Core Metadata Initiative (DCMI) is fully documented at [dublincore.org](http://dublincore.org) and has a stated mission 'to make it easier to find resources using the Internet'.

Examine the basic standard in order to answer the following questions:

- a. How can geographic metadata be incorporated within it?
  - b. Why is it a poor model for such data?, and
  - c. How might it be extended for geographic use?
4. Either from direct experience, or using advertising materials, find examples of in-vehicle navigation systems that integrate GPS location awareness with digital mapping and route-finding software. In each case comment on the system usability, accuracy, geographic resolution, and quality of the mapping used.
  5. A classroom 'competition' involving a hypothetical response to a business request. Organize teams to respond to the following requirement:

'The 'Good-Eat' <sup>TM</sup> fast food chain has over 1000 outlets in towns and cities across the USA, but is losing business to an upstart rival. The rival makes use of distributed GIS to guide potential customers to their nearest outlet using a basic Web mapping facility driven by the customer's ZIP code captured either from their Web page or by way of a text message from a mobile phone. Good-Eat <sup>TM</sup> intend to outsource an equivalent service to a GIServices facility. Your company is invited to bid for the contract to design, build, and run the system

Each team has to produce a design for a system, with as full a specification of the necessary components and system architecture as possible, with the 'contract' being awarded after each team has presented to an adjudicating panel of company directors. Past experience with this kind of role play suggests that it is useful if the instructor acts as 'intermediary' to the company to provide additional information about current company data holdings and GIS capabilities.

6. Current capabilities in Internet GIS did not arrive overnight, but are a stage in an evolutionary sequence. Use the Internet to discover examples of the following steps on the way:
  - a. Static map publishing
  - b. 'Clickable' maps that act as a key to other information
  - c. Interactive Web mapping

7. In each case list the technologies and software involved. Peng Z.-H. and Tsou M.-H. (2003, Chapter 4) will be found to be a useful source for this activity.
8. Organize a formal debate on the motion that 'This house believes that Big Brother, in the guise of geospatial technologies, is now watching us and that this is an invasion of our citizens' liberties that has to be curtailed'.
9. 'Brainstorming' supplies a formal framework for group-based creative thought and is a useful way to capture ideas. The technique is described by Gold et al (1991, Chapter 5). Either in teams, or as a class, brainstorm a list of possible applications of hand-held and wearable GIS. When the group is exhausted, attempt to formalize the suggestions by creating posters that develop each idea.

## FURTHER READING

Câmara A S 2002 *Environmental Systems: A Multidimensional Approach*. New York: Oxford University Press.

National Research Council 1999 *Distributed Geolibraries: Spatial Information Resources*. Washington, DC: National Academy Press.

National Research Council. 2007. *Successful Response Starts with a Map: Improving Geospatial Support for Disaster Management*. Washington, DC: National Academy Press.

Peng Z-H, Tsou M-H 2003 *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks*. Hoboken, NJ: Wiley.

This book is a tour de force, covering all of the materials in this chapter and much, much more.

Scharl A, Tochtermann K. (eds) 2007 *The Geospatial Web: How Geobrowsers, Social Software and the Web 2.0 Are Shaping the Network Society*. Berlin: Springer.

# Cartography and map production

---

## OVERVIEW

- This chapter reviews the nature of cartography and the ways that users interact with GIS in order to produce digital and hard-copy reference and thematic maps.
- Standard cartographic conventions and graphic symbology are discussed, as is the range of transformations that are used in map design.
- Map production is reviewed in the context of creating maps for specific applications and also map series.

## LEARNING OBJECTIVES

After studying this chapter, students will understand:

- The nature of maps and cartography;
- Key map design principles;
- The choices that are available to compose maps;
- The many types of map symbology;
- Concepts of map production flow lines.

## KEY WORDS AND CONCEPTS

Cartography, reference maps, thematic maps, maps vs GIS, map design principles, symbolization, graphic variables, choropleth maps, dot density maps, classification schemes, digital landscape model (DLM)

## OUTLINE

- 12.1 Introduction
- 12.2 Maps and cartography
- 12.3 Principles of map design
- 12.4 Map series
- 12.5 Applications
- 12.6 Conclusions

## CHAPTER SUMMARY

### 12.1 Introduction

- Maps are a very effective way of summarizing and communicating the results of GIS operations to a wide audience.
- Is useful to distinguish between two types of GIS output:
- Formal maps, created according to well established cartographic conventions, that are used as a reference or communication product
- Transitory maps and map-like visualizations used simply to display, analyze, edit, and query geographic information
- This chapter focuses on maps and the next on visualization.
- Cartography concerns the art, science, and techniques of making maps or charts.
- Paper maps remain in widespread use because of their transportability, their reliability, ease of use, and the straightforward application of printing technology that they entail.
- Today's mapping must be capable of communicating an extensive array of messages and emulating the widest range of 'what if' scenarios

### 12.2 Maps and cartography

- There are many possible definitions of a map; here the authors use the term to describe digital or analog (soft- or hardcopy) output from a GIS that shows geographic information using well-established cartographic conventions.
- There are two basic types of map: reference maps and thematic maps
- Maps fulfill two very useful functions as both storage and communication mechanisms for geographic information.
- A major function of a map is not simply to marshal and transmit known information about the world, but also to create or reinforce a particular message.

- Maps have several limitations
  - Can have the effect of miscommunication, accidentally or on purpose
  - Are a single realization of a spatial process
  - Use complex rules, symbology and conventions that may be difficult to understand and interpret by the untrained viewer

### 12.2.1 Maps and media

- Digital cartography and GIS frees map-makers from many of the constraints inherent in traditional (non-GIS) paper mapping.
- Paper maps:
  - are fixed scale, zoom facility of GIS allows viewing at a range of scales
  - are fixed extent, GIS can provide a seamless coverage
  - present a static view of the world, GIS representations can be animated
  - are flat, GIS can provide 3D visualization
  - provide a complete view of the world, GIS allows supplementation of additional data
  - provide a map producer-centric view, GIS allows users to create their own view

## 12.3 Principles of map design

- Primary goals in map design are to share information, highlight patterns and processes, and illustrate results.
- A secondary objective is to create a pleasing and interesting picture, but this must not be at the expense of the message inherent in the data
- Robinson et al (1995) define seven controls on the map design process: purpose, reality, available data, map scale, audience, conditions of use and technical limits

### 12.3.1 Map composition

- Map composition is the process of creating a map comprising of several closely interrelated elements: map body, inset/overview map, title, legend, scale, direction indicator, map metadata
- A key requirement for a good map is that all map elements are composed into a layout that has good visual balance.

### 12.3.2 Map symbolization

- The data to be displayed on a map must be classified and represented using graphic symbols that conform to well-defined and accepted conventions.



- Measurement scales and spatial object types are thus one set of conventions that are used to abstract reality.

#### **12.3.2.1 Attribute representation and transformation**

- Attribute mapping entails use of graphic symbols
  - Basic point, line, and area symbols are modified in different ways in order to communicate different types of information.
- The nature of these modifications was first explored by Bertin in 1967, and was extended to the typology illustrated in Figure 12.9 by MacEachren.
- Some of the common ways in which these graphic variables are used to visualize spatial object types and attributes are shown in Table 12.1.
- Automating placement of symbols and labels presents some challenging analytical problems.
- Most GIS packages include generic algorithms for positioning labels and symbols in relation to geographic objects
- As a general rule, the typical user is unable to differentiate between more than seven (plus or minus two) ordinal categories, and this provides an upper limit on the normal extent of an ordinal hierarchy.
- There are a variety of conventions used to visualize interval- and ratio-scale attributes.
  - Ascribing interval or ratio scale attribute data to areal entities that are pre-defined.
- The standard method of depicting areal data is in zones. However, the choropleth map brings the dubious visual implication of within-zone uniformity of attributes
  - Moreover, conventional choropleth mapping also allows any large (but possibly uninteresting) areas to dominate the map visually.
- Dot density map, which uses points as a more aesthetically pleasing means of representing the relative density of zonally averaged data – but not as a means of depicting the precise locations of point events.
- Proportional circles
- There is no natural ordering implied by use of different colors and the common convention is to represent continuous variation on the red-green-blue (RGB) spectrum.
- There are four basic classification schemes to divide interval and ratio data into categories:

- Natural (Jenks) breaks in which classes are defined according to apparently natural groupings of data values. Best used when the breaks are relevant to a particular application or threshold values
- Quantile breaks in which each of a predetermined number of classes contains an equal number of observations. Well suited to the spatial display of uniformly distributed data
- Equal interval breaks. Best applied if the data ranges are familiar to the user, such as temperature ranges
- Standard deviation classifications
- The choice of classification is very much the outcome of choice, convenience, and the accumulated experience of the cartographer and is aided by the ease with which different classifications can be tested in GIS

#### **12.3.2.2 Multivariate mapping**

Multivariate maps show two or more variables for comparative purposes

This sections illustrates how this can be done with a few examples

### **12.4 Map series**

- Map series by definition share a number of common elements (for example, projection, general layout, and symbology) and a number of techniques have been developed to automate the map series production process.
- The heart of map production through GIS is a geographic database covering the area and data layers of interest
- Such a base cartographic data model is often referred to as a Digital Landscape Model (DLM) because its role is to represent the landscape in the GIS as a collection of features that is independent of any map product representation.
- For each DLM, one or more cartographic data models (DCMs) can be created
- Thus it will be possible to create multiple different map products from this DLM database.
- Many similar maps can be created efficiently from a common map template that includes any material common to all maps (for example, inset/overview maps, titles, legends, scales, direction indicators, and map metadata).

### **12.5 Applications**

- The goal of this section is to highlight a few examples that raise some interesting cartographic issues

- In some instances, prevailing conventions will have evolved over long periods of time, while in others the new-found capabilities of GIS entail a distinct break with the past.
- As a general rule, where accuracy and precision of georeferencing are important, the standard conventions of topographic mapping will be applied
- Utility applications often use schematic maps and now hybrid schematic and geographic maps, *geoschematics*
- Transportation applications use linear referencing
- Military uses of maps have special cartographic conventions of multivariate symbols that have operational and tactical significance

## 12.6 Conclusions

### ESSAY TOPICS

1. In what ways does the printing of a map on paper differ from display on a computer screen and how does this impact on the cartographic techniques that might be used?
2. Describe some methods for the display on a single map of multivariate data.
3. Over the entire range of cartography, to what extent is it true to assert that 'color is a cartographic quagmire'?
4. With reference to specific examples from the history of mapping, is a picture really 'worth a thousand words'?
5. Outline the major uses and limitations of maps and mapping.
6. Cartography is often described as both an art and a science. In the production of a map, when is it 'art' and when 'science'?
7. An objective of many national mapping agencies is to replace multiple map scales and series with a single geospatial database from which a very wide range of maps can be produced. To what extent does the nature of cartography doom this objective to failure?
8. What are the additional cartographic considerations in the production of a series of maps rather than a single product?
9. To what extent do you agree with Wooldridge and East's assertion that "In Geography we may take it as an axiom that which cannot be mapped cannot be described." (Spirit and Purpose of Geography, 1951)?
10. You have been asked to generate maps to be served to user's cell phones, such that they are centered at phone's location and show the road network in the

surrounding square mile. Describe and justify the cartographic methods you would use.

### MULTIPLE CHOICE QUESTIONS (MCQ)

1. For each of the listed properties, say whether it is more characteristic of a computer display or a printed paper map:

Property	Paper, display or both?
Uses symbolism	
Has a single purpose	
Is an end product	
Is selective	
Has many purposes	
Is used to find the unknown	
Can be photo realistic	
Is intended for many viewers	
Is a means to an end	
Can show all the data	
Is used many times	
Demonstrates what's known	
Is used by one person	
Is used once	

2. Which of the following cartographic techniques are best suited to showing the shape of a field such as that of the Earth's surface relief? Select just two from the list.
  - a. Spot heights
  - b. Hachures
  - c. Hypsometric tinting
  - d. Contours
  - e. Hill shading
3. For each of the graphic variables illustrated in Figure 12.9 state the type of attribute (nominal, ordinal, interval, ratio or cyclic, see Box 3.3) it is best suited to display:
- 4.

Graphic Variable	Type of attribute
Size	
Value	
Hue	
Saturation	
Orientation	
Shape	
Arrangement	
Texture	
Focus	

5. Match each of the variable frequency distributions listed to what is likely to be the best choropleth map classing scheme:

Frequency Distribution	Answer (a)-(e)	Scheme
J curve		Standard deviation intervals
Normal distribution		Equal intervals
Multimodal distribution		Geometric series intervals
Uniform distribution		Natural breaks

6. For each of the technical properties listed, state whether it is a characteristic of a paper printed map or a computer display:

Capability	Paper, display or both?
Limited to two dimensions	
Cannot easily show change	
Must use a projection	
Can be animated	
Have limited choice of drafting tools	
Can provide depth cues	
Wide range of drafting technologies	
Can wrap round the globe	

7. Which of the following is generally held to be the maximum number of ordinal categories, such as classes in a choropleth that human beings can readily differentiate?
- 3
  - 5
  - 7
  - 11
8. Which of the following are not essential elements in a balanced map design?
- Title
  - Legend
  - Insets
  - Scale
  - Direction indicator
  - Metadata
  - Map body
9. Of the written word and mapping, which is the oldest form of human communication?
10. Arrange the following processes into a logical cartographic workflow by labeling them (1) TO (5):
- Data collection
  - Analysis
  - Output
  - Editing and maintenance
  - Management

11. Give a suggested example of a map showing each of the following combinations of spatial object and attribute types:

Spatial Object Type & Attribute Type	Suggested Example Map
Point/Nominal	
Area/Nominal	
Surface/Ordinal	
Line/Ordinal	
Surface/Ratio	

## ACTIVITIES

1. In addition to the body of the map itself, Section 12.3.1 lists a further six map requirements (title, legend, scale, north arrow, metadata and possible insets) that should be balanced into a good map design. Although these can be added to the main body of any map using a full GIS, for publication and for presentations, use is often made of generic software such as Adobe Illustrator™, Microsoft Paint™ or PowerPoint™. Incorporate a section of a scanned map into one of these programs using the JPEG format, and then use the tools provided to create a balanced design. The ability to incorporate and annotate maps into presentations using this approach is a very useful skill. A useful restriction to this exercise is to create a single page showing where your home is, together with photographic and text inserts specifying how to find it.
2. Visit the collection of images of maps at [www.esri.com/mapmuseum/index.html](http://www.esri.com/mapmuseum/index.html) and select a volume and applications area of interest to you. For at least five maps rate the quality of each of the seven elements of map composition discussed in Section 12.3.1 (use a simple scale such as “bad” through “good”). Do these vary according to the stated purpose of the map?
3. Find examples of either paper or web maps that illustrate each of the following ways of representing ‘fields’ cartographically: ‘spot heights’, contours, hypsometric tints, relief shading, and hachuring. In each case comment on the effectiveness of the method and its combination with other approaches. A good, but dated, source of examples is the old Tourist Editions of the OSGB’s 1 inch mapping.

4. Experiments with choropleth color. Read the paper by Cynthia A. Brewer (1994) Colour use and guidelines for mapping and visualization. In: MacEachren, A.M. and D.R.F. Taylor (eds) *Visualization in Modern Cartography*, Pergamon: Oxford, pages 123-147 and then run 'ColorBrewer' at <http://colorbrewer2.org/> . Try as many of the available choropleth color schemes as you can and comment on their suitability.
5. Make a list of the possibly different design criteria necessary for publishing maps over the web rather than on sheets of paper. The website at [kartoweb.itc.nl/webcartography/webbook/index1.htm](http://kartoweb.itc.nl/webcartography/webbook/index1.htm) has useful explanatory materials.
6. Classing a choropleth map. If you have not already done so, read the classic paper on choropleth class intervals by I.S. Evans (1976) The selection of class intervals, *Transactions of the Institute of British Geographers*, New Series, 2: 98-124
7. Below are the frequencies distributions of three variables recorded in 1991 over the 403 Districts of England and Wales. Each variable has been standardised in some way, as a % 'rate' or as an areal density, but each has a very different frequency distribution. For each:
  - a) Outline the difficulties that the distribution has for choropleth mapping.
  - b) Suggest and justify a choropleth classification scheme involving five classes.
  - c) Suggest and justify an appropriate symbolism (i.e. set of shade patterns) for the map to be produced. Should you think it necessary, you are allowed the luxury of colour reproduction with an available gamut of either 256 shades of the same colour or 256 different colours.

Software to assist the classification process has also been developed, see Xiao, N., Armstrong, M.P. and D.A. Bennett, Choroware: a software tool for choropleth map classification, at : [www.csiss.org/events/meetings/spatial-tools/papers/xiao.pdf](http://www.csiss.org/events/meetings/spatial-tools/papers/xiao.pdf)

- a) Population Change (% of 1981 value) to 1991



Class	-2	-1.5	-1.0	-0.5	0	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5
Mid-point														
No:	1	0	10	60	127	100	71	29	2	2	0	0	0	1

b) Unemployment in January 1991 as % of total 1989 population

Mid	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0
No:	1	17	60	73	58	55	45	30	36	7	7	4	5	2	2	1

c) Population density (number/km<sup>2</sup> of District area, 1991)

Mid	500	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000	11000
No:	208	82	44	29	21	4	4	1	4	2	2	2

- Several authors have attempted to define some form of choropleth 'error' index of which that proposed by Jenks, G.F. and F.C. Caspall (1971) Error on choroplethic maps: definition, measurement, reduction. *Annals, Association of American Geographers*, 61:217-244 is the best known. A simpler, area weighted version, which can be easily computed using a spreadsheet program, has been suggested by Unwin, D. (1982) *Introductory Spatial Analysis*, Methuen: London. This is based on the idea that when a choropleth zone value is allocated to a class, an error equal to the difference between the actual value for the zone and that of the class mid-point is introduced. Squaring and summing over all zones, with or without a weighting for the zone area, gives a simple index. For some choropleth data known to you suggest and calculate this, or a similar, index. On completing this, and the previous exercise, we suspect that you might well have a view on a famous debate on the desirability of classless choropleths between two authorities, see Tobler, W. R. (1973) Choropleth maps without class intervals *Geographical Analysis* 5, 26–28. and Dobson, M. W. (1973) Choropleth maps without class intervals? A comment, *Geographical Analysis* 5, 358–60. On a classless choropleth the error index is precisely zero, but, as Dobson points out, this may not be what we should be trying to achieve!
- Organize a debate on the motion that 'This house believes that in the world of mobile and ubiquitous GIS, paper maps have no future'.

8. For your national topographic map series, attempt to classify all the symbolism used into one or other of the 'cartographic variables' recognized by Jacques Bertin and extended by Alan MacEachren (Figure 12.9). Summarize your results in a simple table.
9. What is the most beautiful and/or effective map ever created? There is no answer to this, of course, but an amusing and instructive exercise is to ask each member of a class to find one map that they think fits the bill, and then ask them to present their case using just two PowerPoint slides, one of the map itself, the other with bulleted reasons why. The results should be merged into one presentation and stored as a map gallery. Students should vote for the 'winner'.
10. Two of the world's leading authorities on the history of cartography are of the opinion that Harry Beck's justly famous 1933 cartogram of the London Underground network (see Figure 13.9) is 'the most outstandingly successful practical map of all times' (C. Delano Smith & R. Kain (1999) *English Maps: a History* (British Museum:London). The website [//tube.tfl.gov.uk/content/history/map.asp](http://tube.tfl.gov.uk/content/history/map.asp) provides some background on this very influential product, which has been copied in many other cities.

## FURTHER READING

- Board, C. 1967. Maps as models, in Chorley, R. J. and Haggett, P. (eds) *Models in Geography*, (London: Methuen) pages 671–725.
- Dent B.D. 1999 *Cartography: Thematic Map Design* (5<sup>th</sup> edn). Dubuque, Iowa: WCB/McGraw-Hill.
- Downs, R. M. 1998. The geographic eye: seeing through GIS. *Transactions in GIS*, 2(2), 111–121.
- Fisher, P.F., Dykes, J., Wood J. (1993) Map design and visualisation. *The Cartographic Journal*, 30: 36–42.
- Hearnshaw, H. and D. J. Unwin (eds) 1994. *Visualisation and GIS* (London: Wiley)
- Kraak M.-J. and Ormerling F. 1996. *Cartography: Visualization of Spatial Data*. Harlow: Longman.
- Robinson A.H., Morrison J.L., Muehrcke P.C., Kimerling A.J. and Guptill S.C. 1995 *Elements of Cartography* (6th ed). New York: Wiley.
- Slocum T.A. 1999. *Thematic Cartography and Visualization*. Upper Saddle River, NJ: Prentice-Hall Inc.

Tufte E.R. 2001. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

Every GIS library should have this book and others by Tufte.

## ONLINE RESOURCES

ESRI Virtual Campus course, *Turning Data into Information* by Paul Longley, Michael Goodchild, David Maguire, and David Rhind ([training.esri.com](http://training.esri.com) )

Module 1: Basics of Data and Information

Module 2: Cartography, Map Production, and Geovisualization

Section 12.1, Module 2: Cartography, Map Production, and Geovisualization

Unit: GIS-based visualization

Section 12.2, Module 1: Basics of Data and Information

Unit: Creating and visualizing information

Sub-unit: Visualization and interaction

Section 12.2.1, Module 2: Cartography, Map Production, and Geovisualization

Unit: GIS-based visualization

Sub-unit: Properties of GIS-based visualization

Section 12.3, Module 1: Basics of Data and Information

Unit: Creating and visualizing information

Sub-unit: Visualization and interaction

Module 2: Cartography, Map Production, and Geovisualization

Unit: Representing attributes and spatial objects

Section 12.3.2.2, Module 2: Cartography, Map Production, and Geovisualization, Unit:

Advanced methods for improving visualizations

Sub-unit: Multivariate mapping

Sub-unit: Multivariate visualization in Australia

## OVERVIEW

- Using techniques of geovisualization, GIS provides a far richer and more flexible medium for portraying attribute distributions than the paper mapping which is covered in Chapter 12.
- First, through techniques of spatial query, it allows users to explore, synthesize, present (communicate), and analyze the meaning of any given representation.
- Second, it facilitates transformation of representations using techniques such as cartograms and dasymetric mapping.
- Third, GIS-based geovisualization allows the user to interact with the real world from a distance, through interaction with and even immersion in artificial worlds.
- Together, these functions broaden the user base of GIS, and have implications for public participation in GIS (PPGIS).

## LEARNING OBJECTIVES

**After studying this chapter, students will understand:**

- **How GIS facilitates visual communication.**
- **The ways in which good user interfaces can help to resolve spatial queries.**
- **Some of the ways in which GIS-based representations may be transformed.**
- **How 3-D geovisualization and virtual worlds can improve our understanding of the world.**

## KEY WORDS AND CONCEPTS

Geovisualization, cartographic transformation, cartograms, dasymetric maps, scientific visualization (ViSC), PPGIS

## OUTLINE

- 13.1 Introduction: Uses, Users, Messages, and Media
- 13.2 Geovisualization, Spatial Query, and User Interaction
- 13.3 Geovisualization and Interactive Transformation
- 13.4 Participation, Interaction, and Immersion
- 13.5 Consolidation

## CHAPTER OUTLINE

### 13.1 Introduction: Uses, Users, Messages, and Media

- Effective decision support through GIS requires that computer-held representations (discussed in Chapters 3 and 6) are readily interpretable in the minds of users who need them to make decisions.
- Critiques of GIS have suggested that digital maps and virtual Earths present a privileged' or "God's Eye" view of the world because the selectivity inherent in the representation
  - predates the GIS age: it is evident in the early motivations for creating maps
  - Web mapping and digital spatial data infrastructures expose politically sensitive issues and, in the era of VGI, are still explicitly contested as a consequence
- Geovisualization entails multiple representations of large and complex datasets, on the fly.
  - Different from cartography and map production (see Chapter 12) in that it typically uses an interactive computer environment for data exploration
- Historically, the paper map was the only available interface between the mapmaker and the user
  - it was not possible to differentiate between the needs of different users
  - Application Box 13.1 describes the OpenStreetMap project and "tag wars" occurring as a result of the contested geography of Cyprus.
- Geovisualization builds on the established tenets of map production and display.

- It is the creation and use of visual representations to facilitate thinking, understanding, and knowledge construction about human and physical environments, at geographic scales of measurement.
- It is a research-led field that integrates approaches from visualization in scientific computing (ViSC), cartography, image analysis, information visualization, exploratory data analysis (EDA), as well as GIS.
- Its motivation is to develop theories, methods, and tools for visual exploration, analysis, synthesis, and presentation of geospatial data.

## 13.2 Geovisualization, Spatial Query, and User Interaction

### 13.2.1 Overview

- Fundamental to effective geovisualization is an understanding of how human cognition shapes GIS usage, how people think about space and time, and how spatial environments might be better represented using computers and digital data.
- Four principal purposes of geovisualization are explanation, synthesis, presentation and analysis
- The most straightforward way in which reformulation and evaluation of a representation of the real world can take place is through posing *spatial queries*
- These questions are articulated through the GUI paradigm called a WIMP interface, based upon **W**indows, **I**cons, **M**enus and **P**ointers
- The familiar actions of pointing, clicking, and dragging windows and icons are the most common ways of interrogating a geographic database and summarizing results in map and tabular form.

### 13.2.2 Spatial Query Online and the Geoweb

- Spatial query functions are also central to many Internet GIS applications
  - For many users, spatial query is the end objective of a GIS application
  - In other applications, spatial query is a precursor to more advanced spatial analysis
- Web 2.0 has enabled bidirectional collaboration between web sites and has important implications for geovisualization
  - Applications Box 13.2 describes an online mapping portal

## 13.3 Geovisualization and Interactive Transformation

### 13.3.1 Overview

- The term ‘transformation’ is used here in the cartographic sense of taking one vector space (the real world) and transforming it into another (the geovisualization).
- Table 13.1 gives examples of coordinate and cartographic transformations

### 13.3.2 Cartograms

- Cartograms are maps that lack planimetric correctness, and distort area or distance in the interests of some specific objective.
- The usual objective is to reveal patterns that might not be readily apparent from a conventional map or, more generally, to promote legibility.
- Thus, the integrity of the spatial object, in terms of areal extent, location, contiguity, geometry, and/or topology, is made subservient to an emphasis upon attribute values or particular aspects of spatial relations.
- One of the best known linear cartograms is the London Underground map, devised in 1933 by Harry Beck to fulfil the specific purpose of helping travelers to navigate across the network.
- Figure 13.11 presents an equal population cartogram that ensures that every area is drawn approximately in proportion to its population

### 13.3.3 Re-modeling spatial distributions as dasymetric maps

- Dasymetric mapping intersects two datasets to obtain more precise estimates of a spatial distribution.
- Examples of this technique are described and illustrated in Figures 13.12 and 13.13
- It is important to remain aware that the visualization of reality is only as good as the assumptions that are used to create it.
- Inference of land use from land cover, in particular, is an uncertain and error-prone process, and there is a developing literature on best practice for the classification of land use using land cover information and classifying different (e.g., domestic versus non-domestic) land uses.

## 13.4 Participation, Interaction, and Immersion

### 13.4.1 Public Participation in GIS (PPGIS)

Geovisualization has a range of uses in PPGIS, including:

- Making the growing complexity of land-use planning intelligible
- Transforming the planning profession through use of new tools for community design and decision making.
- Unlocking the potential of the digital data
- Helping communities shift land-use decisions from being more proactive and less reactive.
- Improving community education about local environments
- Improving the feed of information between public and government in emergency planning and management

#### 13.4.2 Interaction and 3-D Representation

- Three dimensional representations may be used to reveal aspects of data that are not easily observable in two dimensions
- The third dimension can also be used to represent built form
  - Popular in globe offerings from Microsoft and Google and often simple to augment through free software such as Google Sketchup
- Virtual models of cities provides a valuable forum for planning through consultation on a range of environmental and economic development issues
  - Biographical Box 13.3 describes the Virtual Kyoto (Japan) project

#### 13.4.3 Hand-Held Computing and Augmented Reality

- These are discussed, along with some of the geovisualization conventions they entail, in Sections 7.6.4 and 11.3.

#### 13.4.4 Scientific Visualization (ViSC) and Virtual Reality

- Visualization in Scientific Computing (ViSC) provides new and more sophisticated ways of visualizing and interacting with the world than conventional mapping alone.
  - Involves wide range of computer devices in different settings
- New computer technologies can be used to partially or wholly immerse users in artificial worlds.
- The advent of immersive and semi-immersive systems has important implications for participation by a broad user base since they allow:
  - Users to access virtual environments and select different views of phenomena.
  - Real-time fly-throughs.
  - Repositioning or rearrangement of the objects that make up virtual scenes.
  - Users to be represented graphically as avatars



- Engagement with avatars connected at different remote locations, in a networked virtual world.
  - The development, using avatars, of new kinds of representation and modeling.
- The linkage of networked virtual worlds with virtual reality (VR) systems.

## 13.5 Consolidation

- Geovisualization can make a powerful contribution to decision making and can be used to simulate changes to reality
  - the limitations of human cognition mean that it necessarily provides a further selective filter on the reality that it seeks to represent.

## ESSAY TOPICS

1. For any one application known to you, describe how modern geovisualization has led to new insights or efficiencies.
2. Identify and describe the major drivers of the move to geovisualization as a means of analyzing data.
3. What are the three functional dimensions of geovisualization recognized by MacEachren et al (2000)? Locate some examples known to you in the cube space created by these dimensions.
4. 'All analysis involves transformation'. To what extent is this idea illustrated by geovisualization?
5. A neglected aspect of geovisualization is the impact of the way that we chose to project our geography. Develop, with examples, a case for regarding projection as a very basic graphic variable.
6. What do you understand by the term dasymetric mapping, and why may it offer improvements on conventional choropleth displays? Illustrate your answer with examples.
7. Distinguish between desktop, semi-immersive and immersive virtual reality and describe at least one application of each in GIS
8. Is the development of public participation in GIS frustrated principally by technical, social, educational, or economic impediments? Give reasons for your answer.
9. One of the objectives of many virtual reality applications in geography is to create photorealistic scenes from an original digital representation. Under what circumstances and in which types of application might this actually reduce the viewer's understanding? (see Fisher and Unwin, 2002)

10. Can geovisualization help us handle uncertainty in the results of geographic information analysis?

### MULTIPLE CHOICE QUESTIONS (MCQ)

The emphasis in the chapter on conceptual materials does not readily generate simple MCQs, but here is a collection of five possible questions:

1. The geovisualization process can be broken down into a series of steps. Order these into their logical sequence, and number them starting with '1' through to '6':
  - a. Analysis
  - b. Measurement and representation
  - c. Feedback
  - d. Conception
  - e. Real world
  - f. Exploration
2. Write one sentence descriptions of the basis of the following geovisualization techniques:
  - a. Linear cartogram .....
  - b. Equal population cartogram .....
  - c. Parallel co-ordinates plot .....
  - d. Area cartogram .....
3. Rank the following sources of information in order of their probable suitability for producing a map of the true population densities of an area (Answer 1 to 4, with 1 as the most suitable):
  - a. Land use classified from aerial photography
  - b. Low resolution imagery classified by an unsupervised process
  - c. Point pattern of ZIP or Post Code reference points
4. Expand the following acronyms used in geovisualization work:
  - a. ViSC .....
  - b. EDA .....
  - c. WIMP .....
  - d. PPGIS .....

## ACTIVITIES

1. There is a number of highly graphic websites that describe geovisualization research. One of the best is that developed by the GeoVISTA Center at Penn State University (see Box 13.3) at [www.geovista.psu.edu](http://www.geovista.psu.edu). From the wealth of materials at this site, we suggest that you navigate to the 'software demos' section and examine the effectiveness of any or all of the following advanced visualization tools:
  - a. Scatter plot matrices
  - b. Dynamic parallel co-ordinates plots
  - c. Simulation applied to spatial statistics (advanced)
  - d. Temporal legends
2. Why not also download and examine the geovisualization environment called GeoVISTA Studio? The system is well-documented. A suitable topic for discussion is to assess the important ways that it differs from a standard GIS. Alternatively, compare and contrast the Penn. State approach with that of Wood's Landserf system (see Activity 6).
3. Visualizing health data. Organize the class into teams, and ask each to develop and describe the methods they would use to visualize the pattern of incidence of a rare disease such as TB or cancer across a large city. In doing this, attention needs to be drawn to the nature of the available disease data, which can be assumed to be some form of ZIP or Post Code, and the baseline population at risk data, which should be assumed to be from a standard decennial census of population. One approach, using the regular GIS method of kernel density estimation is outlined at [www.agocg.ac.uk/wshop/sosci/atkinson.htm](http://www.agocg.ac.uk/wshop/sosci/atkinson.htm) .
4. A classroom discussion. Ask the class to read the famous paper by J. Brian Harley (1989) Deconstructing the map, *Cartographica*, 26(2): 1-20. First summarize Harley's arguments and then discuss their merits. How do you think that Harley would have 'deconstructed' current geovisualization activity?
5. One of the best free geovisualization systems yet produced is that written over the past decade by Jo Wood, of the City University, London, England GIS group. His system can be downloaded from [www.soi.city.ac.uk/~jwo/landserf](http://www.soi.city.ac.uk/~jwo/landserf) and then used to experiment with a number of geovisualizations. Although originally developed for morphometric analysis of terrain from DEM data, it is now a very serious, if specialized, GIS. Even if you do not

download the system, visit the image gallery at the website to explore some of its geovisualization capabilities.

6. Visit [www.ncgia.ucsb.edu/projects/Cartogram\\_Central/gallery.html](http://www.ncgia.ucsb.edu/projects/Cartogram_Central/gallery.html) and examine the example cartograms, then attempt a classification of the various types that have been recognized. Whilst at the site, follow the link to [www.mapresso.com](http://www.mapresso.com) to see an animation of Dorling's circle growing algorithm for cartogram construction. MAPresso also has code to create continuous choropleth maps. Other examples of Dorling's (collaborative) work can be viewed at [www.worldmapper.org](http://www.worldmapper.org).

## FURTHER READING

Craig W.J., Harris T.M. and Weiner D. 2002 *Community Participation and Geographic Information Systems*. Boca Raton, FL: CRC Press.

Dorling D. and Thomas B. 2004 *People and Places: a 2001 Census Atlas of the UK*. Bristol, UK: Policy Press.

Hearnshaw H.M. and Unwin D.J. (eds) 1994 *Visualization in Geographical Information Systems*, Wiley: London

Readings on scientific visualization in GIS. The visually stunning cover of this book was produced by Jo Wood, using the GRASS system

MacEachren A. 1995 *How Maps Work: Representation, Visualization and Design*. New York: Guilford Press.

MacEachren A.M., Gahegan M., Pike W., Brewer I., Cai G., Lengerich E., and Hardisty F. 2004 'Geovisualization for knowledge construction and decisionsupport'. *Computer Graphics and Applications* **24**(1), 13–17.

Raubal M. 2009. Cognitive engineering for geographic information science. *Geography Compass* 3: 1–18.

Tufte E.R 2001 *The Visual Display of Quantitative Information* (2nd edn). Cheshire, CT: Graphics Press.

Unwin D. and Fisher P. 2002 *Virtual Reality in Geography*, Wiley: London

An edited collection of early experiments with virtual reality in geography. The essay by Mark Gillings makes a case against photo-realism in archaeological applications.

## RELATED READING

Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, NJ: Wiley.

28 Interacting with GIS, M J Egenhofer and W Kuhn. pp. 401–12.

52 Managing public discourse: towards the augmentation of GIS with multimedia, M J Shiffer, pp. 723–732.

## ONLINE RESOURCES

ESRI Virtual Campus course, *Turning Data into Information* by Paul Longley, Michael Goodchild, David Maguire, and David Rhind ([campus.esri.com](http://campus.esri.com) )

Module 2: Cartography, Map Production, and Geovisualization

Section 13.2, Module 2: Cartography, Map Production, and Geovisualization

Unit: Scientific visualization,

Sub-unit: Purposes of visualization

Sub-unit: Internet GIS as a site selection tool

Section 13.2.2, Module 2: Cartography, Map Production, and Geovisualization, Unit:

Scientific visualization,

Sub-unit: Interacting with representations on the Internet

Section 13.3, Module 2: Cartography, Map Production, and Geovisualization, Unit:

Advanced methods for improving visualizations

Unit: Representing attributes and spatial objects

Sub-unit: Representing spatial objects

Section 13.4.1, Module 2: Cartography, Map Production, and Geovisualization, Unit:

Scientific visualization

Section 13.4.2, Module 2: Cartography, Map Production, and Geovisualization, Unit:

Scientific visualization

Sub-unit: Interacting with representations to support decisions

# 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 is 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

#### 14.3.4 Dependence at a Distance

- The Moran statistic (introduced in Section 4.5) is a global measure that distinguishes between positively (clustered) and negatively (dispersed) autocorrelated patterns.
- Local measures of clustering can be used to identify hot spots

#### 14.3.5 Density Estimation

- Convolution is described as the attenuating effect of distance on a function
- Potential functions have many uses in spatial analysis and are intended to measure influence at a distance.
- Kernel function is a central idea in density estimation
  - In density estimation, each point is replaced by its kernel function and the various kernel functions are added to obtain an aggregate surface, or continuous field of density.

#### 14.3.6 Spatial Interpolation

- Spatial interpolation is a process of intelligent guesswork, in which the investigator (and the GIS) attempt to make a reasonable estimate of the value of a continuous field at places where the field has not actually been measured.
- Notes that the one principle that underlies all spatial interpolation is the Tobler Law

##### 14.3.6.1 Thiessen polygons

- To estimate value at any point take the value measured at the closest point. This leads to a map in which the value is constant within polygons surrounding each point.

##### 14.4.4.2 Inverse-distance weighting

- IDW is the method that is most often used by GIS analysts.
- It estimates unknown measurements as weighted averages over the known measurements at nearby points, giving the greatest weight to the nearest points.
- The mathematical notation is given
- There are various ways of defining weights, but the option most often employed is to compute them as the inverse squares of distances
- IDW is an exact method of interpolation because its interpolated results honor the data points exactly

- Cautions that IDW uses weights that are never negative, values are always calculated between the limits of the measured values, and it tends to regress to the mean outside the area of the data points (Figure 14.25)

#### 14.3.6.3 Kriging

- The basic idea is to discover something about the general properties of the surface, as revealed by the measured values, and then to apply these properties in estimating the missing parts of the surface
- Provides a good but general explanation of the development of the semivariogram, isotropic and anisotropic semivariograms, range, sill, and nugget.
- Notes that a non-zero nugget occurs when there is substantial error in the measuring instrument
- To make estimates using Kriging, need to reduce the semivariogram to a mathematical function, usually by selecting one from a set of standard functional forms and fitting that form to the observed data points in the semivariogram

## 14.4 Conclusion

### Essay topics

1. What is meant by the process of density estimation, and why is it the logical twin of spatial interpolation?
2. Why does the interpolation of real land-surface heights challenge standard methods of automated interpolation?
3. Statistically-minded medical epidemiologists were very critical of the original Geographical Analysis Machine (Figure 14.4) analysis. Outline the cases for and against the approach it took.
4. Evaluate the view that, although 'grounded in good theoretical principles', in practice Kriging is as arbitrary as any other automated contouring approach.
5. Differentiate between first and second order effects in geospatial data and provide illustrative examples of each. Why is it almost always virtually impossible to differentiate between them in practical analysis?
6. Attempt a review and classification of approaches to spatial point pattern analysis.
7. What are the key problems in the application of statistical hypothesis testing to geospatial data?

8. Evaluate the effectiveness of the Moran's I statistic for measuring global spatial autocorrelation paying particular attention to the input data, ease of computation, sampling distributions, and interpretation.
9. Why is polygon overlay such a central operation in a vector GIS environment, and what are the essential steps in a good algorithm to accomplish it?
10. Compare and contrast Kriging with inverse distance weighting as methods for spatial interpolation from control point samples.
11. How and why does geospatial analysis using geographical data differ from the 'spatial analysis' conducted by other sciences? Which is the more difficult and why?

### MULTIPLE CHOICE QUESTIONS (MCQ)

1. In spatial analysis in a GIS environment, which of the following is the most important resource?
  - a. A really powerful computer
  - b. Good functionality in the software
  - c. An intelligent user
  - d. High quality data
2. If we estimate the length of a real world feature using its representation as a polyline in a GIS, is the result almost certain to be (a) shorter or (b) longer than the real world feature?
3. For each of the techniques listed below, state whether or not it being a sensible thing to do depends on the truth of the Tobler 'First law' of geography:

Technique	Law? Y/N
Spatial queries	
Buffering	
Interpolation	
Density estimation	

4. For which type of data is kernel density estimation appropriate?
  - a. A point pattern

- b. A sample of heights
  - c. Values aggregated over areas
5. Tick the true answer or answers to the following five statements. 'Inverse distance weighting and Kriging are similar interpolators because they both:
- a. specify a spatial structure in the data
  - b. are objective
  - c. compute a distance-weighted sum of neighboring data values
  - d. provide estimates of the error at every point in the field
  - e. have a good theoretical basis'.
6. A metric is a .....?

## ACTIVITIES

1. Deconstructing John Snow: 2004 marked the 150<sup>th</sup> anniversary of John Snow's celebrated work on cholera in London, and Biographical Box 14.1 presents the standard story. His work has been presented as pioneering spatial epidemiology and disease mapping, geovisualization, and even GIS 'overlay'. Create a poster that de-bunks the myth by examining the timing of his intervention, the relationship to a priori theory, and the detail of the map itself. Possible sources of information can be found in:
  - a. Brody, H et al, (2000) Map-making and myth making in Broad Street: the London cholera epidemic, 1854 *The Lancet*, 356(9223) 64-68
  - b. [www.jsi.com](http://www.jsi.com)
  - c. [www.ph.ucla.edu/epi/snow.html](http://www.ph.ucla.edu/epi/snow.html), which has numerous useful materials

Reporting results by means of a poster is much more challenging than at first sight it appears. Students will need direction on how to do this effectively and have access to appropriate resources. Some advice can be found at [www2.glos.ac.uk/gdn/abstracts/a141.htm](http://www2.glos.ac.uk/gdn/abstracts/a141.htm).

Nowadays, the same objectives can be addressed by production of either a PowerPoint presentation or a Web page, with the supreme advantage that these can easily be archived and made available to subsequent classes as examples of good and bad practice.

2. The data file below, taken from Davis (2002, Figure 5.66, page 373), has three columns for the (x, y, z) co-ordinates of a sample of survey control points of some topographic (relief) data:

x	y	Height
0.3	6.1	870
1.4	6.2	793
2.4	6.1	755
3.6	6.2	690
5.7	6.2	800
1.6	5.2	800
2.9	5.1	730
3.4	5.3	728
3.4	5.7	710
4.8	5.6	780
5.3	5	804
6.2	5.2	855
0.2	4.3	830
0.9	4.2	813
2.3	4.8	762
2.5	4.5	765
3	4.5	740
3.5	4.5	765
x	y	Height
4.1	4.6	760
4.9	4.2	790
6.3	4.3	820
0.9	3.2	855
1.7	3.8	812
2.4	3.8	773
3.7	3.5	812
4.5	3.2	827
5.2	3.2	805
6.3	3.4	840
0.3	2.4	890

2	2.7	820
3.8	2.3	873
6.3	2.2	875
0.6	1.7	873
1.5	1.8	865
2.1	1.8	841
2.1	1.1	862
x	y	Height
3.1	1.1	908
4.5	1.8	855
5.5	1.7	850
5.7	1	882
6.2	1	910
0.4	0.5	940
1.4	0.6	915
1.4	0.1	890
2.1	0.7	880
2.3	0.3	870
3.1	0	880
4.1	0.8	960
5.4	0.4	890
6	0.1	860
5.7	3	830
3.6	6	705

- a. Import these data into your GIS and apply a selection of interpolation techniques (such as IDW and Kriging) to produce alternative visualizations;
- b. Select the result that you think best reconstructs the entire field from the sample survey point data and in less than 250 words, explain why you think this is 'best';
- c. Visualize this 'best' result using alternative display techniques.

Initially, you might attempt to isoline these data by hand (see Chapter 4).

3. Using a standard road atlas for your country, select any five towns. To complete the exercise the atlas must include a table/matrix of the road distances between towns and your five should be among the listed places.
  - a) Locate your five places on a simple grid placed over the map and read off their (x, y) co-ordinates.
  - b) Use the standard formula for the straight line distance given in Section 14.3.1 to compute the ten unique distances between these places and assemble these into a matrix.
  - c) O'Sullivan and Unwin (2010, Chapter 5) show how this type of matrix can be used to develop almost all of the standard point pattern statistics, as well as 'adjacency' matrices that implement differing conceptions of 'next to'. Using a suitable threshold distance, convert your matrix into a 0/1 binary matrix of adjacencies.
  - d) Examine each of the measures you have created in relation to the three properties that any metric distance should have. Simply stated these are (1) that the distance between points must be a positive number unless the points are the same, in which case the distance will be zero; (2) that the distance between two points is independent of which way round it is measured; and (3) the *triangle inequality*, which states that it must always be at least as far to travel between two points via a third point rather than to travel directly.
  - e) Now assemble a different matrix, of the observed road travel distances as given in the source atlas and compare this with your 'straight line' values. There will be a general similarity, but what reasons are there for the differences? How do these distances measure up to the necessary properties?
  - f) Use the road distance matrix to compute for each row (place) a relative distance, defined as the road distance from that place divided by the average of all the distances from that place (i.e. the row total divided by the number of places). Do the metric properties now hold?



- g) Finally, a nearness statistic can be computed as  $1 / (\text{the relative distance} + 1)$  and again summarized in matrix form. Comment on what it shows.

The second half of this experiment is detailed by Worboys, M.F., (1996) [Metrics and topologies for geographic space](#), in *Advances in Geographic Information Systems Research II: Proceedings of the International Symposium on Spatial Data Handling*, Delft, Kraak, M.J. and Molenaar, M. (eds.), Taylor & Francis: London, pages 365-376, and at [www.spatial.maine.edu/~worboys/mywebpapers/sdh1996.pdf](http://www.spatial.maine.edu/~worboys/mywebpapers/sdh1996.pdf)

4. Investigating 'naïve' notions of distance: Section 14.3.1 discusses the metric distances used in almost all spatial analysis with GIS. As we saw in Activity 5, Chapter 4, this is a notion held mostly by spatially aware professionals (SAPs) and it may not always represent the ideas that individuals have about distance. A 2001 paper by Michael Worboys (Biographical Box 10.3) "Nearness relations in environmental space", *International Journal of Geographical Information Science*, 15(7):633-651 outlines a simple experiment that investigates the idea of distance measured as nearness or proximity. In the paper he presents an experiment with human subjects concerning the vague spatial relation 'near' in environmental space. Three approaches to experimental analysis are presented and discussed: nearness neighborhoods as regions with broad boundaries, fuzzy nearness and distance measures, and four-valued logic. The text of the paper is available at [www.spatial.maine.edu/~worboys/mywebpapers/ijgis2001.pdf](http://www.spatial.maine.edu/~worboys/mywebpapers/ijgis2001.pdf). Using a class of volunteers and the area around your location, repeat the experiment, and compare your findings to those of Worboys. To what extent can GIS be modified to incorporate this, and similar measures of 'distance'?
5. Study the paper: Gatrell, A.C., Bailey, T.C., Diggle, P.J. and B.S. Rowlingson (1996) *Spatial point pattern analysis and its application in geographical epidemiology*. Transactions, Institute of British Geographers, NS 21: 256-274. This deals with three example studies of the use of spatial statistical analysis in medical geography.
  - a. A study of 325 post-coded cases of childhood leukemia in west central Lancashire, 1954–92. Unlike studies by others of the entire northern region, or of the area of West Cumbria immediately to the north of this area, using the  $K(d)$  and  $D(d)$  functions, they fail to detect any evidence of purely spatial clustering;

- b. A study of Burkitt's lymphoma in Uganda using 174 cases that shows some evidence of space-time clustering as evaluated using the space-time difference function  $D(d, t)$ ;
- c. A study of lung cancer in Chorley and South Ribble (Lancashire) using a so-called raised incidence model designed to test whether or not there is a significant effect in distance from a waste incinerator. This suggests that there may well be an effect that would repay detailed medical scrutiny.

In each case summarize the problem, available data, reasons for the choice of technique and result of the analysis.

6. Visit [www.csiss.org](http://www.csiss.org) and download the GeoDa software and associated data and tutorial files. Use it to illustrate the creation of Thiessen polygons from a point data set such as the locations of juvenile crimes set taken from Bailey, T. and Gatrell, A. (1995) *Interactive Spatial Data Analysis*, Wiley: NY, page 95.
7. The concepts involved in ordinary and universal Kriging are not easily explained, and there is no substitute for a structured activity that covers at least some of the ground. The entire process is best introduced in three distinct phases (variogram cloud, semi-variogram and model, and estimate computation) each treated on its merits, rather than as a single 'button click' in a GIS. Use some suitable point-valued data, such as those provided in Activity 3 above, to compute a variogram 'cloud' and then summarize this as an experimental semi-variogram by dividing the distance axis into 'bins' and computing a series of means. Your GIS may be able to do this (check), but if not download and use one or other of the readily available packages for geostatistical analysis such as GSLIB ([www.gslib.com](http://www.gslib.com)), GS+ ([www.geostatistics.com](http://www.geostatistics.com)) or VarioWin (<http://www-sst.unil.ch/research/variowin/>). Failing this, a few simple lines of code are all that is needed to compute the values and send them to a file, which can then be entered into any spreadsheet or statistical analysis program for further analysis and visualization. Using the data from Activity 3, O'Sullivan and Unwin (2002, pages 45-49 and 265-273) illustrate a typical analysis. Modeling the semi-variogram by fitting an appropriate model is likely to be computationally more difficult, but again is worthwhile doing as a distinct step. Finally, computation of the Kriging estimates and their mapping will certainly need computer power (see O'Sullivan and Unwin, pages 274-281).
8. A phrase that is often used is that 'people who play with sharp tools often get cut'. Examine the defaults in any GIS known to you for a selection of the methods introduced

in this chapter, list them and then indicate whether or not you think that a) a beginner, b) an average user and c) a knowledgeable user would 'get cut' using them.

## FURTHER READING

- Cliff A D, Ord J K 1973 *Spatial Autocorrelation*. London: Pion.
- Bailey T C, Gatrell A C 1995 *Interactive Spatial Data Analysis*. Harlow, UK: Longman Scientific and Technical.
- Burrough P A, McDonnell R A 1998 *Principles of Geographical Information Systems*. New York: Oxford University Press.
- De Smith M J, Goodchild M F, Longley P A 2009. *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. Third Edition. Winchelsea: Winchelsea Press. [www.spatialanalysisonline.com](http://www.spatialanalysisonline.com).
- Isaaks E H, Srivastava R M 1989 *Applied Geostatistics*. New York: Oxford University Press.
- O'Sullivan D, Unwin D J 2010 *Geographic Information Analysis*. Hoboken, NJ: Wiley.
- Silverman B W 1986 *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall.

## RELATED READING

Longley P A, Goodchild M F, Maguire D J, Rhind D W (eds) 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, NJ: Wiley.

16. Spatial statistics, A Getis

17. Interactive techniques and exploratory spatial data analysis, L Anselin

19. Spatial analysis: retrospect and prospect, M M Fischer

Maguire D J, Goodchild M F, Rhind D W (eds) 1991 *Geographical Information Systems: Principles and Applications*. Harlow, UK: Longman (text available online at [www.wiley.co.uk/gis/volumes.html](http://www.wiley.co.uk/gis/volumes.html)).

21. The functionality of GIS, D J Maguire and J Dangermond, pp. 319-35

22. Information integration and GIS, I D H Shepherd, pp. 337-60

23. Cartographic modeling, C D Tomlin, pp. 361-74

24. Spatial data integration, R Flowerdew, pp. 375-87

25. Developing appropriate spatial analysis methods for GIS, S Openshaw, pp. 389-402

26. Spatial decision support systems, P J Densham, pp. 403-12

## ONLINE RESOURCES

ESRI Virtual Campus course, *Turning Data into Information* by Paul Longley, Michael Goodchild, David Maguire, and David Rhind ([training.esri.com](http://training.esri.com) )

Module 1: Basics of Data and Information

Module 3: Query and Measurement

Module 4: Transformations and Descriptive Summaries

Module 1: Basics of Data and Information

Unit: Creating and visualizing information

Sub-unit: Types of spatial analysis

Sub-unit: What is spatial analysis?

Module 3: Query and Measurement, Unit: Querying views of a GIS

Section 14.2, Module 1: Basics of Data and Information

Unit: Creating and visualizing information,

Sub-unit: Types of spatial analysis

Module 3: Query and Measurement

Unit: Querying views of a GIS

Unit: Advanced queries

Module 4: Transformations and Descriptive Summaries

Unit: Histograms, pie charts, and scatterplots

Section 14.3, Module 1: Basics of Data and Information

Unit: Creating and visualizing information,

Sub-unit: Types of spatial analysis

Module 3: Query and Measurement

Unit: Querying for measurements

Section 14.3.1, Module 3: Query and Measurement

Unit: Querying for measurements,

Sub-unit: Distance and length

Section 14.3.2, Module 3: Query and Measurement

Unit: Querying for measurements,

Sub-unit: Shape

Section 14.3.3, Module 3: Query and Measurement

Unit: Querying for measurements,

Sub-unit: Slope and aspect

Section 14.4, Module 1: Basics of Data and Information

Unit: Creating and visualizing information,

Sub-unit: Types of spatial analysis

## Module 4: Transformations and Descriptive Summaries

Unit: Buffering, point-in-polygon, and polygon overlay

Section 14.4.1, Module 4: Transformations and Descriptive Summaries

Unit: Buffering, point-in-polygon, and polygon overlay,

Sub-unit: Buffering

Section 14.4.2, Module 4: Transformations and Descriptive Summaries

Unit: Buffering, point-in-polygon, and polygon overlay,

Sub-unit: Point-in-polygon

Section 14.4.3, Module 4: Transformations and Descriptive Summaries

Unit: Buffering, point-in-polygon, and polygon overlay,

Sub-unit: Polygon overlay

Section 14.4.4, Module 4: Transformations and Descriptive Summaries

Unit: Spatial interpolation and density estimation

Section 14.4.4.2, Module 4: Transformations and Descriptive Summaries

Unit: Spatial interpolation and density estimation

Sub-unit: Inverse Distance Weighting

Section 14.4.4.3, Module 4: Transformations and Descriptive Summaries

Unit: Spatial interpolation and density estimation

Sub-unit: Kriging

Section 14.4.5, Module 4: Transformations and Descriptive Summaries

Unit: Spatial interpolation and density estimation

Sub-unit: Calculating density

NCGIA Core Curriculum in GIScience, 2000 ([www.ncgia.ucsb.edu/giscc](http://www.ncgia.ucsb.edu/giscc))

2.1.2.1. [Simple Algorithms for GIS I: Intersection of Lines](#) (184)

2.1.2.2. [Simple Algorithms for GIS II: Operations on Polygons](#), (185)

2.1.2.3. [The Polygon Overlay Operation](#) (186)

2.14.2. [Exploratory Spatial Data Analysis](#) (128), *Robert Haining and Stephen Wise*

NCGIA Core Curriculum in GIS, 1990 ([www.ncgia.ucsb.edu/pubs/core.html](http://www.ncgia.ucsb.edu/pubs/core.html))

5. Raster GIS capabilities

14. Vector GIS capabilities

15. Spatial analysis

32. Simple algorithms I - line intersection

33. Simple algorithms II - polygons

34. Polygon overlay

40. Spatial interpolation I

41. Spatial interpolation II

# Spatial Analysis and Inference

---

## OVERVIEW

- This chapter focuses on five areas: analyses that address concepts of area and centrality, analyses of surfaces, analyses that are oriented to design, and statistical inference.

## LEARNING OBJECTIVES

- **Methods for measuring properties of areas.**
- **Measures that can be used to capture the centrality of geographic phenomena.**
- **Techniques for analyzing surfaces and for determining their hydrologic properties.**
- **Techniques for the support of spatial decisions and the design of landscapes according to specific objectives.**
- **Methods for generalizing from samples, and the problems of applying methods of statistical inference to geographic data.**

## KEY WORDS AND CONCEPTS

Central tendency, centroid, minimum aggregate travel (MAT), standard deviation, clustered, dispersed, and random patterns, normative methods, optimization, p-median and coverage problems, spatial interaction modal, shortest path problem, traveling-salesman problem, heuristics, rook's or queen's case, confidence limits, hypothesis testing, randomization tests, slope, aspect, friction surface

## OUTLINE

15.1 The Purpose of Area-Based Analysis

15.2 Centrality

15.3 Analysis of Surfaces

15.4 Design

15.5 Hypothesis Testing

15.6 Conclusion

## 15.1 The Purpose of Area-Based Analyses

- One way in which humans simplify geography of the Earth's infinite complexity is by ascribing characteristics to entire areas rather than to individual points.

### 15.1.1 Measurement of Area

- Refers to the origins of the Canada GIS in the need to accurate measurements of area and mentions the manual dot-counting and planimeter methods.
- Figure 15.1 gives the algorithm for calculation of the area of a polygon

### 15.1.3 Measurement of Shape

- This section talks mostly about gerrymandering and the need for measures of compact shape

## 15.2 Centrality

- Section reviews numerical summaries

### 15.2.1 Centers

- Lists several measures of *central tendency* including mean, median, and mode
- Introduces *centers* which are the two-dimensional equivalent of the mean and lists some of their properties
- The *centroid* or *mean center* is the most convenient way of summarizing the locations of a set of points
  - Found by taking the weighted average of the x and y coordinates
- The point of minimum aggregate travel (MAT) is the point that minimizes the total straight-line distance from a set of points

- Discusses the stability of MAT given relocation of a point further away on the same line as an illustration of the point that humans can be very poor at guessing the answers to optimization problems in space
- Notes that all methods in this and the next section are based on plane geometry

### 15.2.2 Dispersion

- The measure of choice for numbers with interval or ratio properties is the standard deviation, or the square root of the mean squared difference from the mean
  - Standard deviation and variance are considered more appropriate measures of dispersion than the range (difference between the highest and lowest numbers) because as averages they are less sensitive to the specific values of the extremes.
  - RMSE is a similar measure of dispersion
- A simple measure of dispersion in two dimensions is the mean distance from the centroid

## 15.3 Analysis of Surfaces

### 15.3.1 Slope and Aspect

- Derivative measures such as slope and aspect are defined
- Discusses non-differentiable surfaces and the effects of grid resolution on the calculation of these measures
  - Notes that the spatial resolution used to calculate slope and aspect should always be specified.
- Discusses that there are several alternative measures of slope, and it is important to know which one is used in a particular software package and application.
  - Slope can be measured as an angle and as rise over run
  - Unfortunately there are two different ways of defining run. Figure 14.16 shows the two options, depending on whether run means the horizontal distance covered between two points, or the diagonal distance (the adjacent or the hypotenuse)
- Also, the number of surrounding points used to calculate slope and aspect will vary, as do the weights (see Box 15.2)



### 15.3.2 Modeling Travel on a Surface

- Finding paths across continuous surface easier to solve on a raster as follows
  - Each cell is assigned a friction value equal to the cost or time associated with moving across the cell (the cost layer)
  - A set of allowable moves is selected: rook's case or queen's case (Figure 15.9)
  - Given a cost layer, and a defined origin and destination, a least cost-path is determined

### 15.3.3 Computing Watersheds and Channels

- Describes how a DEM provides an easy basis for predicting how water will flow through a river catchment

### 15.3.4 Computing Visibility

- Describes the process of calculating a viewshed analysis

## 15.4 Design

- This section looks at the analysis of spatial data with the objective of creating improved designs.

### 15.4.1 Point Location

- Explains Hakimi's theorem that states that for the problem of minimizing distance, the only locations that have to be considered are the nodes
- Location-allocation problems which involve where to locate and how to allocate demand to central facilities include
  - The  $p$ -median problem which seeks optimum locations for any number  $p$  of central facilities such that the sum of the distances between each weight and the nearest facility is minimized.
  - The coverage problem which seeks to minimize the furthest distance traveled
- Models that make predictions of how consumers will choose among available options are spatial interaction models

### 15.4.2 Routing Problems

- Routing and scheduling involves decisions about the optimum tracks followed by vehicles

- The shortest path problem finds the path through the network between a defined origin and destination that minimizes distance or some other measure based on distance
- The traveling-salesman problem (TSP) is to select the best tour out of all possible orderings of places to visit, in order to minimize the distance (or other measure) traveled
- The orienteering problem is similar to TSP but the objective is to maximize the rewards associated with visiting a selection of the stops while minimizing the total distance traveled
- In these problems, solution methods usually use heuristics which are algorithms designed to work quickly and to come close to the optimum answer

## 15.5 Hypothesis Testing

- Much work in statistics is *inferential* which uses information obtained from samples to make general conclusions about a larger population, on the assumption that the sample came from that population
- Very briefly introduces the concepts *confidence limits* and *hypothesis-testing*

### 15.5.1 Hypothesis tests on geographic data

- Although inferential tests are standard practice in much of science, they are very problematic for geographic data
- Many inferential tests propose the existence of a population from which samples are obtained independently. However,
  - A geographic dataset is often all there is of a given area – it is the population
  - If we could regard a geographic dataset as a sample, the pervasiveness of spatial dependence means samples would not be independent
  - The Earth's surface is heterogeneous, making it difficult to take samples that are truly representative of any large region
- Before using any inferential tests on geographic data, must ask
  - Can I conceive of a larger *population* about which I want to make inferences?
  - Are my data acceptable as a *random* and *independent* sample of that population?
    - If the answer to either of these questions is no, then inferential tests are not appropriate

- *Randomization* tests which simulate a large number of random arrangements of the data offer a good alternative
- Also, new versions of inferential tests that cope effectively with spatial dependence and spatial heterogeneity are now being developed.

## 15.6 Conclusion

### ESSAY TOPICS

1. Parks and other conservation areas have geometric shapes that can be measured by comparing park perimeter length to park area, using the methods reviewed in this chapter. Discuss the implications of shape for park management in the context of a) wildlife ecology and b) neighborhood security.
2. Use tracings of a number of country borders to compute at least one common measure of shape such as the circularity index  $S = P/3.54\sqrt{A}$  (in which  $P$ =perimeter and  $A$ = area) detailed in the text. Other indices are mentioned by O'Sullivan and Unwin (2002, 177-179). What can be concluded from this exercise about the difficulty of measuring area, the dependence of derived measures (such as shape) on such basics, and the discriminatory power of the indices used?
3. You have been hired to suggest optimum locations for five schools in a new town that is being developed. Suggest how you might approach the problem using location-allocation methods.
4. Besides being the basis for useful measures, fractals also provide interesting ways of simulating geographic phenomena and patterns. Browse the Web for sites that offer fractal simulation software, or investigate one of many commercially available packages. What other uses of fractals in GIS can you imagine?
5. 'Shape is an extremely difficult property to measure, or even to define in a precise manner' (Davis, 2002, page 355). Why?

## MULTIPLE CHOICE QUESTIONS (MCQ)

- Give short definitions of each of the following 'centers':
  - Centroid .....
  - Mean center .....
  - Median center .....
  - MAT point .....
- Rank the following GIS operations in order of their probable computational complexity from 1 (easiest) to 4 (hardest):

Operation	Rank
$p$ -median problem in location-allocation modeling	
Finding the point of minimum aggregate travel	
Traveling salesman solution for 10 places	
Mean center of a point pattern	

- Which statement best completes the sentence? "An heuristic is a computational device for...:
  - Finding correct solutions
  - Finding solutions that might be correct but we can't be sure
  - Speeding up an otherwise lengthy computation
  - Solving location-allocation problems
- Write down the two basic approaches to statistical hypothesis testing:
  - .....
  - .....
- What are the two necessary requirements for a valid sample drawn from a population?
  - .....
  - .....
- What percentage of the Earth's surface that is land has an antipodal point (the point that would be reached by drilling a straight line through the Earth's center and out to the other side) that is also land?
  - 25
  - 50
  - 2
  - 18?

## ACTIVITIES

1. Mean centre and standard distance. Davis, J.C. (2002) *Statistics and Data Analysis in Geology*, Wiley: NY, Figure 5.107 page 444, has a map of two bush species, creosote bush and brittlebush, in a 100 x 100m area in southern Arizona. The associated data file is available from the website at [www.wiley.com/college/davis](http://www.wiley.com/college/davis), under the file name THERMAL.TXT (the full reference is [www3.interscience.wiley.com:8100/legacy/college/davis/0471172758/datafiles/ascii/thermal.txt](http://www3.interscience.wiley.com:8100/legacy/college/davis/0471172758/datafiles/ascii/thermal.txt)). There are 190 records and the data has (x, y) co-ordinates in the range from 0 to 90, with the bush type indicated by a '1' or a '2' in the third column, as in the first 10 items:

x Axis	y Axis	Type
21.97	6.56	1
25.59	8.41	1
36.9	7.79	1
42.01	4.1	1
17.91	10.05	1
16.42	10.87	1
12.37	10.66	1
13.44	12.3	2
15.14	12.71	1
18.98	12.1	1
Etc	...	...

- a) Visit the website and download the full data file. The data can be imported into a spreadsheet for the analysis;
- b) Map the distributions as a dot map;
- c) Compute the mean center ('centroid') and standard distance of the two distributions;
- d) Do these differ significantly?

These types of measure are best used in applications such as that shown in Figure 15.3 for the mean center of US population from 1790 onwards, where there is a comparison over time or between different types of object.

2. Point pattern analysis: The values below give coordinates for the distribution of 65 Japanese pine tree seedlings in a small 23m by 23m area:

0.09 0.09	0.42 0.49	0.62 0.97
0.59 0.02	0.37 0.68	
0.86 0.13	0.76 0.66	
0.42 0.22	0.97 0.86	
0.02 0.41	0.29 0.84	
0.08 0.59	0.58 0.83	
0.31 0.53	0.39 0.96	
0.94 0.58	0.39 0.18	
0.59 0.67	0.73 0.13	
0.94 0.78	0.02 0.18	
0.17 0.95	0.73 0.23	
0.39 0.79	0.52 0.42	
0.36 0.97	0.12 0.66	
0.29 0.02	0.52 0.52	
0.65 0.16	0.47 0.67	
0.89 0.08	0.73 0.73	
0.48 0.13	0.12 0.84	
0.03 0.44	0.32 0.83	
0.08 0.63	0.69 0.93	
0.32 0.52	0.43 0.96	
0.34 0.68	0.48 0.03	
0.66 0.68	0.79 0.03	
0.98 0.79	0.11 0.31	
0.21 0.79	0.89 0.23	
0.52 0.93	0.64 0.43	
0.36 0.96	0.17 0.58	
0.38 0.03	0.91 0.52	
0.67 0.13	0.52 0.67	
0.98 0.02	0.89 0.74	
0.62 0.21	0.11 0.94	
0.07 0.42	0.35 0.86	
0.12 0.63	0.77 0.93	

These data are used by Diggle, P.J.(1991) *Statistical Analysis of Spatial Point Patterns* (London: Academic) and have been analysed many times (see pages 128-9).

Map them as a dot map and overlay a suitable grid, using it to perform a quadrat census of the sort illustrated by O’Sullivan and Unwin (2010, pages 127-130). Note how the counts might change if a different grid was used, and that an alternative approach would be to generate the proportions by a true sampling process in which use was made of a sequence of, say, 100 randomly located quadrats. Our count for these data gave results as follows:

Number of events in quadrat, k	Number of quadrats, x
0	49
1	39
2	10
3	2

Use the variance/mean ratio test to assess the probability that these events are a realization of the independent random process (O’Sullivan and Unwin, 2010, pages 142-143 provide the details).

3. Choose an example of some planar-enforced area objects such as, for example, the lower 48 states of the USA, or the counties in a state. In doing this, you should find data in which the basic ‘geography’ of the co-ordinates and topology have already been constructed for you in a format suitable for ArcGIS or similar GIS format. The US Bureau of Census website at [www.census.gov/](http://www.census.gov/) has such data readily available.
  - a. Select some variables that describe characteristics of the people resident in this area. It makes sense to choose maybe 3-4 variables that together help understanding;
  - b. Visualize these data using maps and other statistical graphics. It is most probable that you will choose to use the choropleth technique, but this should not rule out other approaches;

Compute and use an appropriate global measure of spatial autocorrelation to verify that the apparent patterns are globally significantly different from independent random

process/complete spatial randomness process (see O'Sullivan and Unwin, 2010, pages 142-143).

- c. Write a brief (less than 500 words) account of what the study shows about the spatial variation in social/economic conditions across the region.
7. Organize a debate on the motion that 'This house believes that statistix inferens has no place in the analysis of geographical data'. Preliminary reading for this can be found in:
  - a. Gould, P. R. (1970), Is statistix inferens the geographical name for a wild goose?, *Economic Geography* 46: 439-448.
  - b. Harvey, D. W. (1966), Geographical processes and the analysis of point patterns, *Transactions of the Institute of British Geographers* 40: 81-95.
  - c. Harvey, D. W. (1967), Some methodological problems in the use of Neyman type A and negative binomial distributions for the analysis of point patterns, *Transactions of the Institute of British Geographers* 44: 81-95.
  - d. O'Sullivan and Unwin (2002, pages 53-58) develop the argument that any one map can be treated as a realization of a 'random' process and there is similar material in the excellent text: Stewart Fotheringham, Chris Brunsdon and Martin Charlton (2000) *Quantitative Geography: Perspectives on Spatial Analysis*, Sage: London
7. What exactly are multicriteria methods? Examine one or more of the methods in the chapter referenced below, summarizing the issues associated with a) measuring variables to support multiple criteria, b) mixing variables that have been measured on different scales (e.g., dollars and distances), ) and c) finding solutions to problems involving multiple criteria. Eastman J R 1999 'Multicriteria methods.' In Longley P A, Goodchild M F, Maguire D J, Rhind D W (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition (2005)). Hoboken, NJ: Wiley.
8. Algorithmics is the study of the basic properties of computer algorithms, often by assessment of their computational complexity, as defined by the relationship between their notional time to completion as a function of the inputs. Thus an algorithm that is complete in direct proportion to the number of inputs has linear time order, denoted  $O(n)$ , whereas computation of a matrix of all possible distances between objects has order  $O(n^2)$  and so on. A good introduction to the topic that examines the traveling salesman and other geographical problems is Harel, D.



(2000) *Computers Ltd: What They Really Can't Do*, Oxford: Oxford University Press.  
Examine each of the problems outlined in the chapter and attempt to assess its computational complexity.

## FURTHER READING

Davis J C 2002 *Statistics and Data Analysis in Geology*, Wiley: Hoboken, NJ, 3rd edition  
Fotheringham A S, Brunson C, Charlton M 2000 *Quantitative Geography: Perspectives on Spatial Data Analysis*. London: Sage.  
Fotheringham A S, O'Kelly M E 1989 *Spatial Interaction Models: Formulations and Applications*. Dordrecht: Kluwer.  
Ghosh A, Rushton G (eds) 1987 *Spatial Analysis and Location-Allocation Models*. New York: Van Nostrand Reinhold.  
Mardia K V, Jupp P E 2000 *Directional Statistics*. New York: Wiley.

## RELATED READING

Longley P A, Goodchild M F, Maguire D J, Rhind D W (eds) 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, NJ: Wiley.

35. Multi-criteria evaluation and GIS, J R Eastman

Maguire D J, Goodchild M F, Rhind D W (eds) 1991 *Geographical Information Systems: Principles and Applications*. Harlow, UK: Longman (text available online at [www.wiley.co.uk/gis/volumes.html](http://www.wiley.co.uk/gis/volumes.html)).

24. Spatial data integration, R Flowerdew, pp. 375-87

25. Developing appropriate spatial analysis methods for GIS, S Openshaw, pp. 389-402

26. Spatial decision support systems, P J Densham, pp. 403-12

27. Knowledge-based approaches in GIS, T R Smith and Ye Jiang, pp. 413-25

## ONLINE RESOURCES

ESRI Virtual Campus course, *Turning Data into Information* by Paul Longley, Michael Goodchild, David Maguire, and David Rhind ([campus.esri.com](http://campus.esri.com))

Module 1: Basics of Data and Information

Module 4: Transformations and Descriptive Summaries

Module 5: Optimization and Hypothesis Testing

Section 15.2, Module 1: Basics of Data and Information

Unit: Creating and visualizing information,

Sub-unit: Types of spatial analysis

Module 4: Transformations and Descriptive Summaries

Unit: Centers and dispersion

Section 15.2.1, Module 4: Transformations and Descriptive Summaries

Unit: Centers and dispersion

Sub-unit: Centers

Sub-unit: The Varignon Frame experiment

Section 15.2.2, Module 4: Transformations and Descriptive Summaries

Unit: Centers and dispersion,

Sub-unit: Dispersion

Section 15.2.5, Module 4: Transformations and Descriptive Summaries

Unit: Spatial dependence and fragmentation

Sub-unit: Fragmentation

Section 15.3, Module 1: Basics of Data and Information

Unit: Creating and visualizing information

Sub-unit: Types of spatial analysis

Module 5: Optimization and Hypothesis Testing

Unit: Optimization

Section 15.3.1, Module 5: Optimization and Hypothesis Testing

Unit: Optimization

Sub-unit: Point location

Section 15.3.2, Module 5: Optimization and Hypothesis Testing

Unit: Optimization,

Sub-unit: Routing problems

Section 15.3.3, Module 5: Optimization and Hypothesis Testing

Unit: Optimization,

Sub-unit: Optimum paths

Section 15.4, Module 1: Basics of Data and Information

Unit: Creating and visualizing information,

Sub-unit: Types of spatial analysis

Module 5: Optimization and Hypothesis Testing

Unit: Hypothesis testing

Section 15.4.1, Module 4: Transformations and Descriptive Summaries

Unit: Spatial dependence and fragmentation,

Sub-unit: Spatial dependence

Module 5: Optimization and Hypothesis Testing

Unit: Hypothesis testing,

Sub-unit: Hypothesis tests on geographic data

NCGIA Core Curriculum in GIScience, 2000 ([www.ncgia.ucsb.edu/giscc](http://www.ncgia.ucsb.edu/giscc))

2.14.1. [Spatial Decision Support Systems](#) (127), Jacek Malczewski

2.14.6. [Artificial Neural Networks for Spatial Data Analysis](#) (188), Suchi Gopal

NCGIA Core Curriculum in GIS, 1990 ([www.ncgia.ucsb.edu/pubs/core.html](http://www.ncgia.ucsb.edu/pubs/core.html))

57. Multiple criteria methods

58. Location-allocation on networks

59. Spatial decision support systems

74. Knowledge based techniques

# Spatial Modeling with GIS

---

## OVERVIEW

- This chapter begins with the necessary definitions and presents a taxonomy of models
- The alternative software environments for modeling are reviewed, along with capabilities for cataloging and sharing models, which are developing rapidly.

## LEARNING OBJECTIVES

- **Know what modeling means in the context of GIS;**
- **Be familiar with the important types of models and their applications;**
- **Be familiar with the software environments in which modeling takes place;**
- **Understand the needs of modeling and how these needs are being addressed by current trends in GIS software.**

## KEY WORDS AND CONCEPTS

Computational models, geocomputation, spatial and temporal resolution, static models, indicators, cellular models, cartographic modeling, map algebra, model coupling, multicriteria methods, cross-validation, agent-based models (ABM), scripts, uncertainty propagation, sensitivity analysis

## OUTLINE

- 16.1 Introduction
- 16.2 Types of Models
- 16.3 Technology for Modeling
- 16.4 Multicriteria Methods
- 16.5 Accuracy and Validity: Testing the Model
- 16.6 Conclusion

### 16.1 Introduction

- A clear distinction needs to be made between *data models* and the *spatial models* that are the subject of this chapter.
  - A data model is a template for data, a framework into which specific details of relevant aspects of the Earth's surface can be fitted. It is a statement about how the world looks
  - Models in this chapter are expressions of how the world is believed to *work*, in other words they are expressions of process
- All of the models discussed in this chapter are *digital* or *computational* models, meaning that the operations occur in a computer
  - The term *geocomputation* is often used to describe the application of computational models to geographic problems
- All of the models discussed in this chapter are also *spatial* models. There are two key requirements of such a model:
  1. There is variation across the space being manipulated by the model
  2. The results of modeling change when the locations of objects change
- The level of detail in computational models is measured as *spatial resolution* which is defined as the shortest distance over which change is recorded.
- *Temporal resolution* is defined as the shortest time over which change is recorded
- Spatial and temporal resolution determine
  - What is left out of the model
  - What is the level of uncertainty between the model and the real processes
  - The cost of acquiring data
  - The cost of running the model

#### 16.1.1 Why model?

- To support a design process
- To experiment on a replica of the world

- To examine dynamic outcomes

### 16.1.2 To analyze or to model?

- Analyzing results in static outcomes
- Modeling can result in dynamic animations

## 16.2 Types of models

### 16.2.1 Static models and indicators

- A static model represents a single point in time and typically combines multiple inputs into a single output. There are no time steps or loops.
- USLE is given as an example. Why use GIS to evaluate the USLE?
  - Some of the inputs require GIS for their calculation
  - The inputs and outputs are best expressed, visualized, and used in map form
  - The inputs and outputs are often integrated with other types of data

### 16.2.2 Individual and aggregate models

- Models of physical systems are forced to adopt aggregate approaches because of the enormous number of individual objects involved, whereas it is much more feasible to model individuals in human systems, or in studies of animal behavior
- Even when modeling the movement of water as a continuous fluid it is still necessary to break the continuum into discrete pieces
  - Some models adopt a raster approach and are commonly called *cellular* models
  - Other models break the world into irregular pieces or polygons
- Models of individuals are often termed *agent-based* models (ABM) or *autonomous agent* models

### 16.2.3 Cellular models

- Each cell in the raster has a number of possible states which change through time as a result of the application of transition rules.
- Typically the rules are defined over each cell's neighborhood, and determine the outcome of each stage in the simulation based on the cell's state, the states of its neighbors, and the values of cell attributes.
- One of the most important issues in such modeling is calibration and validation

#### 16.2.4 Cartographic modeling and map algebra

- Classifies all GIS transformations of rasters into four basic classes: local, focal, global, and zonal
- Provides a simple language in which to express a model as a script

### 16.3 Technology for modeling

#### 16.3.1 Operationalizing models in GIS

- Models can be defined as sequences of operations which can be expressed as either graphic flowcharts or *scripts*
- Scripts express models as sequences of commands
- While scripting used to be specific to individual software products, today it is common for scripts to be written in industry-standard languages

#### 16.3.2 Model coupling

- GIS will often fail to provide adequate performance as a modeling environment for very large datasets and large numbers of iterations
- Therefore, spatial modeling is often done by coupling GIS with other software
  - *Loosely coupled* models are run separately with data exchanged as files
  - *Closely coupled* models use the same files
  - *Embedded* models are implemented as GIS scripts or through Graphic User Interfaces (GUI)

#### 16.3.2 Cataloging and sharing models

- Very little effort to date has gone into making it possible to share *process objects*, digital representations of the process of GIS use
- GIServices describe GIS functions being offered by a server for use by any user connected to the Internet
  - Reasons for offering these are described in Section 7.2 and repeated here
  - Examples of services include gazetteer service and geocoding service

### 16.4 Multicriteria methods

- Are used when there are a number of factors that influence vulnerability and there are different stakeholders with different views about what is important, how that importance should be measured, and how the various important factors should be combined

- Multicriteria decision making (MCDM) is commonly used whenever decisions are controversial
- A common framework is the Analytical Hierarchy Process (AHP), which focuses on capturing each stakeholder's view of the appropriate weights to give each impact factor
  - These weights are inserted as parameters in the spatial model to produce a final result

## 16.5 Accuracy and validity: testing the model

- Models can often be tested by comparison with past history, by running the model not into the future, but forwards in time from some previous point.
  - But these are often the data used to *calibrate* the model, to determine its parameters and rules, so the same data are not available for testing.
- Many modelers use *cross-validation*, a process in which a subset of data is used for calibration, and the remainder for validating results.
- Models of real-world processes can be validated by experiment, by proving that each component in the model correctly reflects reality.
- However, the outputs of models must always be taken advisedly, bearing in mind a model:
  - May reflect behavior under ideal circumstances and therefore provide a norm against which to compare reality
  - Should not be measured by how closely its results match reality but by how much it reduces uncertainty about the future
  - Is a mechanism for assembling knowledge from a range of sources and presenting conclusions based on that knowledge in readily used form
- Models are subject to uncertainty
- Present in their inputs
  - *Uncertainty propagation* concerns the impacts of input uncertainty on the uncertainty of the outputs
- Present in their parameters
  - *Sensitivity analysis* examines each parameter in turn to see how much influence it has on the results
- Related to the labels used to express the results



## 16.6 Conclusion

### ESSAY TOPICS

1. What do you understand is the scientific meaning of the word 'model', and how does this differ, if at all, from everyday use of the word?
2. Give a reasoned definition of the term 'geocomputation', and provide examples of its use in GIS analysis.
3. Why model? Illustrate your answer by reference to examples from GIScience.
4. Differentiate between analog and digital models. What are the main difficulties in validating both types of model?
5. What is a cartographic model, and what distinguishes it from other types of model used in GIScience?
6. Compare and contrast 'modeling' with 'analysis'. Why and in what ways do these activities normally interact?
7. With reference to specific examples, explain how dynamic models might be coupled to a GIS. What are the advantages and disadvantages of each approach?
8. You have just completed a climate forecast using a large-scale model of the atmosphere for a situation in which there is a doubling of atmospheric carbon dioxide. The model predicts a global warming over the next 50 years of around 4C. How might you validate this finding?
9. A possible classification of mathematical models is into 'deterministic' and 'stochastic'. Giving examples of each, illustrate what is meant by these terms.
10. A famous meteorologist, Ed Lorenz, once distinguished two types of atmospheric model. The first were designed to predict the future behavior of the present weather for use in weather and climate forecasting, the second to paint 'what if' scenarios for how the climate might change when disturbed in some way such as by increased carbon dioxide, nuclear war, deforestation and so on. Of the two types of modeling activity, he was of the opinion that the second is much easier than the first. Can you state why, and give other examples drawn from both human and physical sciences?

### MULTIPLE CHOICE QUESTIONS (MCQ)

1. Classify each of the following model as 'analog' or 'digital'
  - a) Printed map
  - b) Varignon frame

- c) Atmospheric general circulation
  - d) Universal Soil Loss Equation
2. Attempt a definition of the following model properties:
- a) Spatial resolution is .....
  - b) Temporal resolution is .....
3. For each of the following models, classify them by the number of dimensions used to represent space (Point =  $L^0$ , Section/Line =  $L^1$ , Area =  $L^2$ , Volume =  $L^3$ ) and whether or not there is a time dimension:

Model	Spatial Dimension?	Time?
Dam Busters model, Application Box 16.1		
Richard Church's evacuation model, Biographical Box 16.2		
DRASTIC groundwater model, Figure 16.7		
Mammoth Cave groundwater vulnerability, Application Box 16.3		
Batty's crowd movement models, Technical Box 16.4		
Conway's Game of Life, Technical Box 16.5		
Keith Clarke's urban growth model, Figure 16.3		
MCD model, Figure 16.15		

4. Draw diagrams to illustrate each of the four basic types of spatial operation in the Map Algebra. These are 'local', 'focal', 'global', and 'zonal'.
5. What does the term analog mean when applied to a scientific model?
- Which of the following words best describes analysis and which modeling? STATIC, MULTIPLE STAGED, IMPLEMENTING HYPOTHESES, EXPERIMENTING, MANIPULATING DATA, and GENERATING HYPOTHESES. Write your answers under the headings provided.

Modeling	Analyzing

6. Which of the following are valid media for defining a dynamic model:
  - a) Computer programs
  - b) Wood and glue
  - c) Mathematical equations
  - d) Flowcharts
  - e) Databases
  
7. In a sentence, say what is meant by each of the following when applied to coupling between a model and a GIS:
  - a) Loose .....
  - b) Close .....
  - c) Embedded .....
  
8. Define each of the following acronyms used in the Chapter:
  - a) ABM
  - b) AHP
  - c) MCDM
  - d) USLE
  - e) VBA

## ACTIVITIES

1. John R. Gold *et al.* (1992) *Teaching Geography In Higher Education: A Manual of Good Practice*, Chapter 7 has some useful, if dated, materials on computer simulation modeling, which give useful ideas on teaching with, and about, models. They can be used to:
  - provide in class demonstrations;
  - carry out experiments;
  - replicate a known feature or event;
  - relate model behavior and outcomes to the real world;
  - explore alternative realities;
  - provide an environment for problem solving;
  - act as a focus for evaluating geographic theory.

For many of these activities, the portal developed at [www.ncgia.ucsb.edu/projects/metadata/web\\_models.html](http://www.ncgia.ucsb.edu/projects/metadata/web_models.html) will be found to be extremely useful.

2. Building a cartographic model. As suggested in the text, Application Box 16.2, visit [www.esri.com/news/arcuser/0704/files/modelbuilder.pdf](http://www.esri.com/news/arcuser/0704/files/modelbuilder.pdf) and follow the sequence of steps taken by Rhonda Pfaff and Alan Glennon in using ESRI's *ModelBuilder*™ to create their model of groundwater vulnerability in the Mammoth Cave watershed.
3. The anatomy of a model. This exercise can be developed using any suitable model that is within the student's discipline. Simply examine the detail of a model, concentrating on aspects such as its formulation in mathematical terms, how it is operationalized, the necessary inputs, what the outputs are used for, and so on. The SOLIM model of A. Xing Zhu at [solim.geography.wisc.edu](http://solim.geography.wisc.edu) is an accessible example. Examples of other computational models in a GIS frame will be found at [www.ncgia.ucsb.edu/projects/metadata/](http://www.ncgia.ucsb.edu/projects/metadata/).
4. Model sensitivity to inputs. An important feature of many complex models is their sensitivity to changes in their inputs. This exercise examines this by way of a simple example:
  - a) Visit the Dane County website at [www.co.dane.wi.us/landconservation/uslepg.htm](http://www.co.dane.wi.us/landconservation/uslepg.htm) and examine the materials presented on the Universal Soil Loss Equation (USLE). This has been used many times in GIS to create cartographic models of soil loss over a region. It has as its inputs 'factors' developed from analysis of rainfall, erodibility, slope, crop type, and management. Download the model spreadsheet and experiment to discover the output sensitivity to changing inputs.
  - b) Now examine the Temperature Urban Run-Off (TURM) model at the same website. How sensitive are its three outputs to changes in the inputs?
  - c) Can you develop indices that would provide a systematic assessment of model sensitivity?
  - d) Can you suggest a useful way of measuring model output sensitivity?
5. Models can be classified in many ways. Climatologists, such as Kendal McGuffie and Ann Henderson-Sellers (2005), often classify models by their spatial dimensionality. In a way that will be familiar to GIS students, this starts at the base of a pyramid by point models ( $L^0$ ) that also do not have time variation and are thus 'equilibrium models'.

Watson and Lovelock's (1985) famous 'Daisyworld' model is of this type. Keeping the same spatial dimensionality and thus treating the Earth-atmosphere system as a single point, Budyko (1969) and Myrup (1969) developed a model that varied its input over time. Myrup's model, designed to teach about urban heat islands, was developed on an analog computer and in digital code by Outcalt (1972). The next step is to make space 1-dimensional ( $L^1$ ), usually either by looking at variation with height or by latitude (that is, by averaging conditions over longitude), as in a global climate model by Sellers (1973). A two-dimensional spatial model ( $L^2$ ) without time variation creates what Section 16.2.4 refers to as a cartographic model. Allow time to vary and we have models like the cellular automata discussed in the text. Add a third spatial dimension, the height ( $L^3$ ) and allow time to vary and you have, for example, a full atmospheric general circulation model (AGCM) of the sort used routinely to predict tomorrow's weather or to experiment in order to investigate global warming.

Either from the literature, or using your imagination, find or suggest examples of every kind of model that arises from a classification of the sort given in the table below:

Dimensionality of space	Time invariant	Time varying
$L^0$	?	?
$L^1$	?	?
$L^2$	?	?
$L^3$	?	?

### References for this question

Budyko M I (1969) The effect of solar radiation variations on the climate of the Earth, *Tellus*, 21, 611-619.

Kendal McGuffie and A Henderson-Sellers 2005 *A Climate Modelling Primer*, Wiley: NY, 3rd Edition A very comprehensive text with software.

Myrup LO 1969 A numerical model of the urban heat island, *Journal of Applied Meteorology*, 8, 908-910.

Saaty T L 1980 *The Analytical Hierarchy Process: Planning, Priority Setting, Resource Allocation*. New York: McGraw-Hill.

Outcalt S I 1972 A numerical surface climate simulator, *Geographical Analysis*, 3, 379-398.

Sellers W D 1973 A new global climate model, *Journal Applied Meteorology*, 12, 241-254

Watson A J, Lovelock J E 1983 Biological homeostasis of the global environment: the parable of Daisyworld, *Tellus*, 35, 285-288. Proof that to influence decisions and debate,

models do not have to be complicated. In terms of its influence on thought, Daisyworld is possible the most influential model of all time.

6. John Conway's *Game of Life* (Technical Box 16.5 and Figure 16.11) is great fun to play, but it also has a serious objective in demonstrating how simple rules, played out over a 'geography', can generate enormously complex behaviors. First, visit the website at [www.math.com/students/wonders/life/life.html](http://www.math.com/students/wonders/life/life.html) and run the game itself. Next, consider how this sort of simple model might be used in geography. Examples include Michael Batty's simulation of crowd behavior in crowded spaces, as detailed in Technical Box 16.4. One fruitful area of work has been to develop cellular automaton models of urban land use. In this, each cell had a classified land use, with various rules that at each time step say whether it will change its use or not. What rules do you think might be appropriate for such a model? Roger White and Guy Engelen 2000 have developed some very realistic urban simulations using this approach.

7. Dana Tomlin's Map algebra (see Section 16.2.4 and, for example [www.quantdec.com/SYSEN597/GTKAV/section9/map\\_algebra.htm](http://www.quantdec.com/SYSEN597/GTKAV/section9/map_algebra.htm)) is an attempt to define a language that enables cartographic models to be quantified, at least in a raster GIS environment. In this exercise we list extensions to the basic command set:

To introduce time variation and some general operators of use primarily in environmental science modeling using GIS. The best example of this is PcRaster, developed by Willem van Deursen, Cees Wesseling, Peter Burrough, Derek Karssenberg, Edzer Pebesma and Kor de Jong of the Department of Physical Geography, Utrecht University, The Netherlands. There is a very good website at <http://pcraster.geo.uu.nl>

8. Model user interfaces. Visit [www.spatial.maine.edu/~max/MapAlgebraSurvey.pdf](http://www.spatial.maine.edu/~max/MapAlgebraSurvey.pdf) and examine the paper by H Thomas Bruns and Max J Egenhofer (1997) User interfaces for map algebra. *Journal of the Urban and Regional Information Systems Association*, 9(1): 44-54. In this they note the importance of interface design, recognizing four basic designs: (a) command line (b) form-based (c) flowchart and (d) 'stack'. Collect as many examples of models as you can, and in each case classify their user interfaces using this approach.
9. The role play exercise takes you through a hypothetical siting study involving a garbage landfill site close to the City of 'Greensea', in which a cartographic model using overlay of a series of GIS products is the basic approach. Although this is technically very easy in a

GIS, the analysis has considerable potential for error in the final locational decision arising from the original data, generalizations in the inputs and outputs, the data pre-processing, and the procedures used.

We first imagine that the City's own GIS team has performed a sieve analysis developing a cartographic model for the decision from four allegedly 'objective' criteria (see below). As a result of this analysis, they have selected a suitably sized site that meets these criteria. Possibly catastrophic for property values and for the inhabitant's well-being, this site is close to the small community known locally as 'East-Central'. East-Central's inhabitants are understandably upset by this decision and are to appeal the decision. As part of their evidence, they wish to discredit the City GIS unit by revealing as many sources of uncertainty as they can in the study. To do this they have hired a newly graduated GIS analyst (you) in which they ask you to:

*Produce a short report on the possible sources of error in the final overlay map to be included as part of the community's submission to the City Council*

A brief overview of what the City GIS team did in their study and the data sources they used is as follows. Their objective was to site a garbage landfill near to Greensea according to the following four agreed criteria. The site has to be:

Within 250m of the main urban area of Greensea to provide easy access. This layer was found by buffering using the US Bureau of Census City Block data;

Within 100m of a road, for easy access by garbage trucks. This used a layer locally surveyed in 2000 using GPS to record the road center-lines, again developed by buffering these lines.

On slopes of less than  $2^{\circ}$ , in order to avoid water-pollution problems with drainage and allow easy working conditions. This was derived from standard USGS 30m digital elevation matrices using the supplied 'slope' command in the GIS;

On land of low agricultural potential. This layer was developed by classification based on the USGS GIRAS series using aerial photography shot in the 1970s and 1980s. At source, the data were projected onto an Albers Equal Area Projection referenced to the 1927 NAD and, to co-register these with the other layers used, they were re-projected onto UTM using NAD

83. Land-cover categories were amalgamated so as to exclude currently productive agricultural land.

These requirements cannot be challenged, since they were set by the City Council at an earlier meeting, and, as typical in this environment, some of the details of what was done are not available to you. However, it is admitted that the analytical strategy the GIS team used was to create a binary (0/1) map for each of these four criteria. On each map areas were coded '1' if, and only if, they met the stated criterion. These four maps were then overlaid so that only those cells that were coded "1" on every one of these maps could be considered as possible sites. In cartographic modeling terms this is a favorability model in which the output / isn't a weighted sum of the input  $x$ , as in Section 16.4, but is the function:

$$I = \prod_{m=1}^M w_m f(x_m)$$

In this all the input map values  $x_m$  are coded 0/1, all the  $m = 4$  layers have a weight  $w_m$  of 1.0, there is an additional function ( $f$ ) to generate the inputs, and the capital  $\prod$  symbol means take the product of the  $m$  values, not as in the text, their summation.

10. Organize a debate on the motion that 'This house believes that because of its limited data structures, a GIS is an unsuitable framework for any modeling activity'. Almost all of the previous activities are relevant to developing both a case for and a case against this motion.
11. Bidding for a research contract. In teams, develop a bid for a research project to map the risk from landsliding in southern California using GIS techniques together with a dynamical model. The objective is to produce a map at a scale of 1:50,000 of areas so much at risk that they cannot be zoned for housing. The contract, with the Governor's Office of the State can be for up to \$500,000, but the work has to be completed within two years. Each team should produce a fully costed and illustrated proposal, specifying its approach and anticipated 'deliverables'. Two famous Californian slides are documented at

[www.geog.ucsb.edu/~jeff/projects/la\\_conchita/apcg2001\\_article/apcg2001\\_article.html](http://www.geog.ucsb.edu/~jeff/projects/la_conchita/apcg2001_article/apcg2001_article.html) and [www.ci.fremont.ca.us/Environment/MissionPeakLandslide/Default.htm](http://www.ci.fremont.ca.us/Environment/MissionPeakLandslide/Default.htm)



California residents will know that the La Conchita site had a second disastrous slide on January 10<sup>th</sup> in the very wet winter of 2005.

## FURTHER READING

Goodchild M F, Parks B O, Steyaert L T editors 1993 *Environmental Modeling with GIS*. New York: Oxford University Press, p. 488

Heuvelink G B M 1998 *Error Propagation in Environmental Modelling with GIS*. London: Taylor and Francis.

Kirkby M J, Naden P S, Burt T P, Butcher D P 1987 *Computer Simulations in Physical Geography*, Wiley: Chichester

Essentially a teaching oriented text with some valuable materials.

Kirkby M, Naden P 1988 The use of simulation models in teaching geomorphology and hydrology, *Journal of Geography in Higher Education*, 12 (1): 31-49

See above!

Marble D F, Anderson B M 1972 *LANDUSE: A Computer Program for Laboratory Use in Economic Geography Courses*, Washington, DC, Association of American Geographers, Technical Paper 8, Commission on College Geography.

A very forward-looking teaching simulation program, now of historical interest, that implemented the 'simple' Von Thunen land use model, but then by relaxing the assumptions rapidly showed how even a simple model could generate seemingly complex land use patterns. The NCGIA develop a similar model for use with IDRISI, see Dodson R F 1991 *VT/GIS: The von Thunen GIS Package*, see

<http://www.ncgia.ucsb.edu/pubs/pubslst.html#91-27>

Tomlin C D 1990 *Geographic Information Systems and Cartographic Modeling*. Englewood Cliffs, NJ: Prentice Hall.

Unwin D J 1981 Teaching a model-based climatology using energy balance simulation, *Journal of Geography in Higher Education*, 5, 133-138

Describes teaching units in a local context, but has ideas about the educational benefits of modeling.

White R, Engelen G 2000. High Resolution Integrated Modelling of the Spatial Dynamics of urban and Regional Systems. *Computers, Environment, and Urban Systems*, 24: 383-400  
Making a simple cellular automaton emulate city growth.

## RELATED READING

Maguire D J, Goodchild M F, Rhind D W (eds) 1991 *Geographical Information Systems: Principles and applications*. Harlow, UK: Longman (text available online at [www.wiley.co.uk/gis/volumes.html](http://www.wiley.co.uk/gis/volumes.html)).

- 24. Spatial data integration, R Flowerdew, pp. 375-87
- 25. Developing appropriate spatial analysis methods for GIS, S Openshaw, pp. 389-402
- 26. Spatial decision support systems, P J Densham, pp. 403-12
- 27. Knowledge-based approaches in GIS, T R Smith and Ye Jiang, pp. 413-25

## ONLINE RESOURCES

NCGIA Core Curriculum in GIScience, 2000 ([www.ncgia.ucsb.edu/giscc](http://www.ncgia.ucsb.edu/giscc))

- 2.14.1. [Spatial Decision Support Systems](#) (127), Jacek Malczewski - [GC notes](#)
- 2.14.6. [Artificial Neural Networks for Spatial Data Analysis](#) (188), Suchi Gopal

NCGIA Core Curriculum in GIS, 1990 ([www.ncgia.ucsb.edu/pubs/core.html](http://www.ncgia.ucsb.edu/pubs/core.html))

- 57. Multiple criteria methods
- 58. Location-allocation on networks
- 59. Spatial decision support systems
- 74. Knowledge based techniques

## OVERVIEW

- This chapter addresses the essential ‘front end’: the procurement and operational management of GIS
- This chapter describes how to choose, implement, and manage operational GIS.
- It involves four key stages: the analysis of needs, their formal specification, the evaluation of alternatives, and the implementation of the chosen system.
- Implementing GIS requires consideration of issues such as planning, support, communication, resource management, and funding.
- Successful on-going management of an operational GIS has five key dimensions: support for customers, operations, data management, application development, and project management.

## LEARNING OBJECTIVES

- **Return on investment concepts.**
- **How to go about choosing a GIS to meet your needs.**
- **Key GIS implementation issues.**
- **How to manage an operational GIS effectively with limited resources and ambitious goals.**
- **Why GIS projects fail—some pitfalls to avoid and some useful tips about how to succeed.**
- **The roles of staff members in a GIS project.**
- **Where to go for more detailed advice.**

## KEY WORDS AND CONCEPTS

Return on investment (ROI), prototyping, stages of choosing, implementing and managing a GIS

## OUTLINE

17.1 Introduction

17.2 The case for GIS: ROI

17.3 The process of developing a sustainable GIS

17.4 Sustaining a GIS – the people and their competences

17.5 Conclusions

### 17.1 Introduction

- This chapter is concerned with the practical aspects of managing an operational GIS.
- Describes high-level management concepts
- The authors argue that success comes from combining strategy and implementation
- Success involves constant sharing of experience and knowledge with other people, keeping good records, and making numerous judgments where the answer is not preordained.

### 17.2 The Case for GIS: ROI

- Outlines the strategic questions that senior executives are likely to ask before implementing a GIS
- To be successful:
  - GIS strategies must be aligned with business strategies,
  - GIS processes must reflect business processes.
- GIS that exists in a vacuum and is disconnected from an organization's business processes may be more of a business concern than no GIS at all.
- ROI uses a combination of qualitative and quantitative measures to assess the utility that an organization will obtain from an investment.
- Figure 17.1 outlines the ROI methodology comprising of 10 steps
- Table 17.1 describes some examples of the main types of tangible and intangible benefits that have been used in the past to justify GIS projects for government and utility organizations.

## 17.3 The Process of Developing a Sustainable GIS

- This section provides a short, but very useful summary of the formal stages and steps in the development of a GIS
- GIS projects comprise four major lifecycle phases: business planning; system acquisition; system implementation; and operation and maintenance.

### 17.3.1 Choosing a GIS

- Choosing a GIS involves four stages: analysis of requirements; specification of requirements; evaluation of alternatives; and implementation of the system.
- Figure 17.4 shows these stages and the 14 steps
- This section briefly describes each of these steps in sufficient detail to understand what they entail
- Stage 1: Analysis of requirements
  - Step 1: Definition of objectives
  - Step 2: User requirements analysis
  - Step 3: Preliminary design
  - Step 4: Cost-benefit analysis
  - Step 5: Pilot study
- Stage 2: Specification of requirements
  - Step 6: Final design
  - Step 7: Request for proposals
- Stage 3: Evaluation of alternatives
  - Step 8: Short-listing
  - Step 9: Benchmarking
  - Step 10: Cost-effectiveness evaluation
- Stage 4: Implementation of system
  - Step 11: Implementation plan
  - Step 12: Contract
  - Step 13: Acceptance testing
  - Step 14: Implementation

#### 17.3.1.1 Discussion of the classical acquisition model

- While widely employed, this model has some significant shortcomings including:
  - It is expensive and time-consuming
  - Proposals can become technologically obsolete
  - Short-listing requires multiple vendors
  - Evaluation process often focuses under attention on price

- This type of procurement can be highly adversarial
- Many organizations have little idea about what they really need
- A less complex and formal selection method is *prototyping*
- Here a vendor or two are selected early on and funded to build a prototype in close collaboration with the user organization
- Works best for those procurements where there is some uncertainty about the most appropriate technical solution

### 17.3.2 Implementing a GIS

- Plan effectively
- Obtain support
- Communicate with users
- Anticipate and avoid obstacles
- Avoid false economies
- Ensure database quality and security
- Accommodate GIS within the organization
- Avoid unreasonable timeframes and expectations
- Funding
- Prevent meltdown
- Table 17.3 lists many tools and techniques available for use in implementation projects including SWOT analysis, data flow diagrams, project management tools and object model diagrams

### 17.3.3 Managing a sustainable, operational GIS

- Success in operational management of GIS requires customer support, effective operations, data management, and application development and support.

#### 17.2.3.1 Customer support

- Key tasks include including technical support and problem logging plus meeting requests for data, maps, training, and other products.

#### 17.2.3.2 Operations support

- Operations support includes system administration, maintenance, security, backups, technology acquisitions, and many other support functions.

#### 17.2.3.3 Data management support

- Large, multi-user geographic databases use database management system (DBMS) software to allocate resources, control access, and ensure long-term usability.

A database administrator (DBA) is responsible for ensuring that all data meet all of the standards of accuracy, integrity, and compatibility required by the organization.

#### **17.2.3.4 Application development and support**

- Sources of application development work include improvements/enhancements to existing applications, as well as new users and new project areas starting to adopt GIS.

## **17.4 Sustaining a GIS—The People and Their Competences**

### **17.4.1 GIS staff and the teams involved**

- In addition to management, day-to-day GIS work involves three key groups of people: the GIS team itself, headed by a GIS manager; the GIS users; and external consultants.
- Figure 17.9 shows the GIS staff roles in a medium to large GIS project

### **17.4.2 Project managers**

The role of the project manager is to establish user requirements, to participate in system design, and to ensure that projects are completed on time, within budget, and according to an agreed quality plan.

### **17.4.3 Coping with uncertainty**

- The chapter ends with a strong note of caution about the importance of understanding uncertainty
- Organizations must determine how much uncertainty they can tolerate before information is deemed useless.
- Managing error requires use of quality assurance techniques to identify them and assess their magnitude.

## **17.5 Conclusions**

### **ESSAY QUESTIONS**

1. 'Does my company really need a GIS'? Outline the arguments you would present to the Chairman of the Board of a utilities (gas, electricity or water) supply company by way of answer to this question.
2. List and describe a series of reasons why GIS implementations can 'fail'. Why are 'success' and failure often hard to determine in GIS and similar projects?

3. What do you understand by the term 'cost benefit analysis'? Identify and describe the major elements of such an analysis as applied to a possible GIS implementation.
4. How would you evaluate the impact of a corporate GIS implementation on an organization?
5. GIS can be introduced into an organization either 'top down' or 'bottom up'. Compare and contrast these approaches, commenting on their strengths and weaknesses.
6. What are the general management strategies that can be adopted to minimize the risk associated with a large scale GIS acquisition and roll out into a company?
7. The classical system acquisition model is widely used in the GI industry and has proved its worth many times. What are its shortcomings?
8. Many studies of GIS introductions isolate the importance of a 'champion' in the enterprise. What are the advantages and disadvantages of being a 'GIS champion'?
9. What are the key needs of a 'sustainable' GIS?
10. Analyses of the labor force in the GI industry show that by far the largest proportion have a background in academic geography and related disciplines. Is this a good or a bad thing for GIS project management?

### **MULTIPLE CHOICE QUESTIONS (MCQ)**

The materials in this Chapter do not lend themselves into creating MCQ or related examinations questions.

### **ACTIVITIES**

Perhaps for three reasons, formal project management is often dismissed by GIS instructors. First, it is sometimes seen as 'not GIS', or someone else's responsibility, with the result that, if it is covered at all, it is treated in a cursory manner. Second, many instructors seem to think that it is in some sense 'trivial', not worthy of attention relative to all that exciting GI theory. Third, when formal techniques are addressed, the opinion often expressed is that they are 'common sense'.

We empathically reject all of these arguments, and the text devotes three extended chapters to examination of a number of related human issues in GIS. Although there is a reluctance to discuss it, what evidence we have in the literature of 'failure' in GIS shows that it is almost always not associated with the technology but with various complex failures of management. Formal project management will not in itself guarantee project success, but it will lessen the chance of 'failure', and is almost always worth the additional effort and cost. It is a discipline



in its own right, with an associated body of both theory and practice that is relevant to GIS. The 'common sense' argument ignores the fact that, once it has been pointed out, a great deal of well found scientific theory can be dismissed in the same terms.

There are numerous institutions dedicated to furthering the art and science of project management, for example:

[www.pmi.org](http://www.pmi.org)

[www.apm.org.uk](http://www.apm.org.uk)

[www.pmforum.org](http://www.pmforum.org)

1. Implementation is critical to the spread of any technological innovation such as GIS, but what are the conditions that favor it? How do organizations interact with technology?

Campbell and Masser (1995, pages 25-50) suggest three main approaches:

- Technological determinism
- Managerial rationalism
- Social interactionalism

Working in three teams, study their Tables 3.1, 3.2 and 3.4 and then outline the implications for GIS implementations of each approach. A possible approach makes use of a so-called **PEST** analysis (Political, Economical, Social, and Technical) of the environment.

2. A key skill in developing any project is the determination and articulation of the objectives. The standard acronym used is that these should be **SMART** (Specific, Measurable, Achievable, Relevant, Timely). An effective in-class pyramid exercise is to define a specific possible GIS implementation and then ask each member of the class to list just four **SMART** objectives for the system. Next, students join into groups of two and negotiate their 'best' four from the eight available. Then meet as groups of four and repeat the exercise, 'pyramiding' in this way until you have, perhaps 12-16 agreed objectives identified. These can then be discussed and moderated by the entire class. Examples and criteria for making objectives SMART can be found, for example, at [www.learnmarketing.net/smartobjectives.htm](http://www.learnmarketing.net/smartobjectives.htm)

- a. Stage 1: Requirements analysis. This exercise, and those that follow, is suitable for an extended class activity, associated perhaps with a formal course on GIS management. Using any favored formal approach to project management, and for a selected applications area, simulate by role play the

management of the implementation, starting with the requirements analysis. In each step identify the roles that are involved, assign these to individuals or groups, and provide the necessary time and related resources for them to complete the tasks. The opportunity should be taken to introduce the actors to relevant materials from a selected formal approach (such as **PRINCE 2**) and to appropriate aids (such as Microsoft Project <sup>™</sup>, **GANTT** charts and associated critical pathways). In this stage deal with:

- Objectives definition
- User requirements analysis
- Preliminary design
- Cost-benefits analysis
- Pilot study

- b. Stage 2: Formal specification of requirements (see also Activity 2.9). The objective here is to create a formal request for proposals (RFI) for the project that can be sent out as part of an invitation to tender to system vendors. Changing the teams and individuals around, address the next two steps in the management model suggested in the Chapter:

- Final design
- Request for proposals

- c. Stage 3: Evaluation of alternatives. Either generate initial expressions of interest from possible suppliers, or ask student teams to prepare vendor responses (see Activity 2.8) capable of formal evaluation using a weighted scoring system and undertake the next two steps

- Short-listing
- Benchmarking
- Cost-effectiveness evaluation

- d. Stage 4: Implementation of system. Not every part of this is possible, but adopt the same approach to the final four steps:

- Implementation plan
- Contract
- Acceptance testing
- Implementation

3. Almost all writers on IT implementations agree that the 'people ware' is the most important and potentially costly element of the system. It follows that developing appropriate role and job descriptions is an extremely important part of any implementation. Either alone, or working in teams, use a template with headings such as 'job name', 'conditions and salary', 'required educational level and qualifications', 'experience', 'necessary skills', and 'person profile', develop job descriptions for the following roles in a GIS:
- manager
  - administrator
  - applications programmer
  - data base analyst
  - geographic information analyst
  - data entry technician
  - group secretary and assistant
4. In the UK and elsewhere, the standard approach to project management is called PRINCE 2, and possession of its basic qualification is a sine qua non for any IT manager. PRINCE <sup>TM</sup>, which stands for **P**rojects **i**n **C**ontrolled **E**nvironments, is a project management method covering the organization, management and control of projects. It was first developed by the Central Computer and Telecommunications Agency (CCTA), now part of the Office of Government Commerce (OGC), in 1989 as a UK Government standard for IT project management. The website at [www.ogc.gov.uk/prince](http://www.ogc.gov.uk/prince) describes the formal process and usefully provides no less than 28 blank templates for appropriate documentation. Nowadays, no IT literate business would actually use the printed medium to create or store these documents, but would instead maintain them via a shared and networked computer facility. Design an appropriate system to do this for a large number of projects, using either a simple basic hierarchical folder structure or standard relational database such as Microsoft Access <sup>TM</sup>.

Another excellent, if commercial, resource for information about PRINCE 2 methods is at <http://www.crazycolour.com/p2/>. Note that this lists no less than 45 formal steps in a full application!

5. Visit [www.ogc.gov.uk/prince2/downloads/template\\_case.htm](http://www.ogc.gov.uk/prince2/downloads/template_case.htm) and examine the listed case studies of major IT implementations. In each case, identify the critical management decisions and actions.
11. Evaluating and using project management tools. Starting at the website [www.pcworld.com/downloads](http://www.pcworld.com/downloads), examine a selection of the available tools for project management that go by names such as Project, Business Plan Master, Plan Builder, Proficient Project Manager and so on. Make a formal evaluation of each, using a check list you devise in advance.
12. Imagine you are the GIS manager in a large retailing corporation. Currently your unit provides services solely for the market research team, but you would like your unit to grow. Prepare a presentation intended for your Board of Directors to persuade them to make a substantial investment in developing the current system into an enterprise wide GIS.

## FURTHER READING

Campbell, H. and I. Masser (1995) GIS and organizations: how effective are GIS in practice? Taylor & Francis: London

In addition to a detailed analysis of GIS diffusion into UK local government, this text has some very useful materials setting the process into its general context.

Harmon J.E. and Anderson S. J. 2003 The Design and Implementation of Geographic Information Systems. Hoboken, NJ: Wiley (e-book).

Heywood I., Cornelius S. and Carver S. 2002 An Introduction to Geographical Information Systems (2nd edn). Harlow: Prentice-Hall.

Tomlinson R. 2003 Thinking about GIS: Geographic Information System Planning for Managers. Redlands, CA: ESRI Press.

By definition, Roger Tomlinson has more experience in GIS management than anyone else. This book distills all he has learnt over a long and distinguished career about how to manage a GIS project. As a 'how to do it' guide, it is required reading.

Thomas C. and Ospina M. 2004 Measuring Up: The Business Case for GIS. Redlands, CA: ESRI Press.

## RELATED READING

Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) 2005 Geographical Information Systems: Principles, Techniques, Management and Applications (abridged edition). Hoboken, NJ: Wiley.

42. Measuring the benefits and costs of GIS, N J Obermeyer, pp. 601-610

43. Managing an operational GIS, L J Sugarbaker, pp. 611-620

Maguire D.J., Goodchild M.F. and Rhind D.W. (eds) 1991 Geographical Information Systems: Principles and applications. Harlow, UK: Longman (text available online at [www.wiley.co.uk/gis/volumes.html](http://www.wiley.co.uk/gis/volumes.html)).

31. GIS specification, evaluation and implementation, A L Clarke, pp. 477-88

33. Managing an operational GIS: the UK National On-Line Manpower Information System (NOMIS), M J Blakemore, pp. 503-13

44. Counting the people: the role of GIS, D W Rhind, pp. 127-37

## ONLINE RESOURCES

NCGIA Core Curriculum in GIScience, 2000 ([www.ncgia.ucsb.edu/giscc](http://www.ncgia.ucsb.edu/giscc))

3.1. [Making it work](#) (136), *Hugh Calkins and others*

NCGIA Core Curriculum in GIS, 1990 ([www.ncgia.ucsb.edu/pubs/core.html](http://www.ncgia.ucsb.edu/pubs/core.html))

60. System planning overview

61. Functional requirements analysis

62. System evaluation

63. Benchmarking

64. Pilot project

65. Costs and benefits

# Operating Safely with GIS

---

## OVERVIEW

- Describes how best to use GI science and technology to make safe decisions
- Involves seizing opportunities without losing public trust or incurring excessive costs
- Describes the role nation-states play in setting the ground rules and the extent to which they are active participants and regulators in GI markets

## LEARNING OBJECTIVES

- **The nature of trade-offs in GIS-based decision making.**
- **The common and different drivers for organizations and employees using GIS.**
- **The characteristics of information in general and GI in particular.**
- **Some important elements of the operating environment for safe GIS-based decision making—the law, trading in GI and the role of governments, protecting or undermining privacy, the ethics of behavior, and public trust.**
- **How the operating environment can vary through time and in different jurisdictions.**
- **Where to go for more detailed information and advice.**

## KEYWORDS

Copyright, intellectual property and liability; government; open data; trust and ethics

## OUTLINE

- 18.1 Introduction
- 18.2 GIS and Decision making
- 18.3 Organizational Context
- 18.4 Geographic Information
- 18.5 GIS, GI, and Key Management Issues
- 18.6 Conclusion

### 18.1 Introduction

- Managers exist to make decisions and implement them successfully
- every organization -businesses, governments, voluntary organizations, and everything else (even universities) -has managers, and everyone at some stage is going to be a manager in some function
- GIS and GI are central to many management functions and much decision making.

### 18.2 GIS and Decision Making

- Any policy or management action taken has consequences, some of which are unforeseen
- Everything is interconnected: normally, some gain and some lose in decision making.
  - As example is given in Applications Box 18.2 which details the trade-off between geographic detail and privacy

### 18.3 Organizational Context

- The GIS world is driven by organizational and personal objectives, scientific understanding, and raw materials (GIS, skills, and GI).

### 18.4 Geographic Information

#### 18.4.1 The Characteristics of Information

- From a management perspective information does not wear out through use, though it may diminish in value as time passes
- Information can be considered a public good where:
  - The marginal cost of providing an additional unit is close to zero
  - Use by one individual does not reduce availability to others (“nonrivalry”)
  - Individuals cannot be excluded from using the good or service (“nonexcludability”)

- In practice, information is an optional or partial public good- it is possible to opt to take it or not and individuals can be excluded from using it in some situations
- Information in digital form can generally be copied and distributed via the Internet at near-zero marginal cost and creates: producer externalities; network externalities and consumer externalities

#### 18.4.2 Additional Characteristics of GI

- Comparing one dataset for an area with another often permits the infilling of missing data
- indication of the quality of individual GI datasets can be assessed through overlay
- It provides added value, almost for nothing.
- It is difficult to quantify the quality of some types of GI

### 18.5 GIS, GI, and Key Management Issues

- Details five different areas in which safe GIS use is impacted by management or where GIS can contribute to ameliorating management's difficulties

#### 18.5.1 The Law

- Laws of various sorts have several roles to: regulate and incentivize the behaviour of citizens; to help resolve disputes; and protect the individual citizen
- Protecting innovation and liability are considered in relation to GIS

##### 18.5.1.1 Fostering Innovation and Protecting Exploitation

- Two types of innovation are identified:
  - Where it is created internal to an organization and its exploitation is protected by law
  - Where its creation occurred through an open innovation process where the innovation results may not be protected (though there may still be some form of licensing—see Section 18.5.1.2).
- Applications box 18.4 details intellectual property rights

##### 18.5.1.2 Owning and Exploiting GI

- The relationship between GIS and GI with IPR is complex because:
  - National or regional IPR laws may vary
  - The formal interaction between the private and public sector varies between countries.



- Global databases are being built in some domains while many other geographical databases are national in scope, content, ontology, and reference systems. •
- Can geographic data, information, evidence, and knowledge be regarded as property?
  - Yes, but who owns the data is sometimes difficult to define unequivocally
- Can geographic information always be legally protected?
  - Whether the products and contents of GIS are in the public domain, are “facts”, or are creative works can be disputed
- Can information collected directly by machine, such as a satellite sensor, be legally protected under copyright?
  - Possibly not because of the need for originality, however this is not the view taken by the major players, such as GeoEye.
- How can tacit geographic and process knowledge – such as that held in the heads of employees and gained by experience – be legally protected?
  - The only formal way to protect tacit knowledge is to write some appropriate obligations into the contracts of all members of staff – but it rarely works
- How can you prove theft of your data or information?
  - Data may be watermarked or finger-printed by the careful insertion of small bits of data which are carefully documented for this purpose or by incorporation of highly time-dependent forms
- Who owns information derived by adding new material to source information produced by another party?
  - The answer is the originators of both the first and resulting datasets
- Almost all GI is licensed somehow. Read the license before using the GI!

### **18.5.1.3 GIS, GI, and Legal Liability**

- Liability is a creation of the law to support a range of important social goals, such as avoidance of injurious behavior, encouraging the fulfillment of obligations established by contracts, and the distribution of losses to those responsible for them.
- Other liability burdens may also arise under legislation relating to specific substantive topics such as intellectual property rights, privacy rights, anti-trust laws (or non-competition principles in a European context), and open records laws.
- Reducing liability exposure for creators and distributors of geographic software and data products is achieved primarily through performing competent work and keeping all parties informed of their obligations.

## 18.5.2 Trading in GI and the Role of Government

- GI is now a necessary part of delivering services

### 18.5.2.1 The Role of Governments

- Governments have been and are major collectors (and sometimes providers) of certain types of GI and could provide much more information that is presently “locked up” inside public-sector organizations.
- Remit of government data providers varies considerably across the world

### 18.5.2.2 Free Our Data?

- Reform of government has led to two approaches in regard to GI
  - “charging for everything”
  - “free access”
  - The advantages and disadvantages of these different approaches are outlined in applications box 18.6

### 18.5.2.3 Competition versus Collaboration

- the relationship between the state as a provider of GI and the private sector as a value-adder sometimes breaks up

## 18.5.3 GIS and Privacy

### 18.5.3.1 Geoslavery

- Constant monitoring of an individual’s position via mobile (or cell) phones
  - Term ‘geoslavery’ coined by Dobson and Fisher
- The widespread availability of GI as in Street View and as handled in GIS is manifestly a factor in the loss of anonymity

### 18.5.3.2 Privacy versus the Case for Individual Data

- Most socioeconomic raw data are collected for individuals, and so privacy matters.
- The potential loss of privacy through misuse or loss of personal data has to be traded against the considerable benefits

## 18.5.4 GIS Ethics and Decision Making

- Ethics refers to principles of human conduct, or morals, as well as to the systematic study of such human values.

## 18.5.5 Public Trust

- The authors discuss how much trust is appropriate?
  - The GIS community seems to enjoy high public trust—but it could evaporate

## 18.6 Conclusions

### ESSAY TOPICS

1. Differentiate between intellectual property rights such as copyright, database rights and patent rights. Which parts of a GIS are regulated by which legal instruments?
2. Outline a strategy for ensuring that you could detect and prove that some geospatial data had been stolen or your copyright infringed.
3. What are the ethical implications of GIS analysis at the level of the individual citizen, as might be obtained from a 'life style' survey?
4. When invited to comment on the role of GIS in society, a famous GI scientist wrote a 'position paper' entitled 'GIS and society: a lot of fuss about very little that matters and not enough about that which does' (see [www.geo.wvu.edu/i19/papers/openshaw.html](http://www.geo.wvu.edu/i19/papers/openshaw.html)). Write a reasoned justification for this view (Note: if setting this as an examination question provide copies of the paper)
5. To what extent might a data supplier be liable in the event of a GIS induced commercial 'disaster' that can be shown to be in part a consequence of errors in data supplied to a second party? What are the legal arguments you would bring forward to support the case that the company is liable?
6. How would you go about reassuring the public that they can retain trust in what you are doing and that their privacy will be maintained?
7. What do you understand by the term 'framework data'? Give examples of national and international attempts to create them.
8. How can the legal risks associated with routine use of a GIS be minimized?
9. Find out from national and local (e.g., state) government Web sites what constraints you will have to operate within if you use GI produced by them.
10. How would you price GI and services based on it if you were running a business?

### MULTIPLE CHOICE QUESTIONS (MCQ)

There are none for this chapter.

## ACTIVITIES

1. This simple exercise has been used many times, and forms part of the UK Association for Geographic Information's (AGI) Continuing Professional Development (CPD) scheme. It is reproduced here by permission of the School of Geography, Birkbeck College's GISOnline program:

The Oxford English Dictionary has a long entry for the word *professional* that reads:

*Adjective. Of, belonging to, connected with, a profession ... usually or properly pursued from higher motives; maintaining a proper standard, business like, not amateurish.*

Listed below are ten 'jobs'. First, can you rank them in increasing order of *professionalism as assessed by the lay public*? The jobs are:

- Bank Manager
- Realtor/Estate Agent
- GI Officer for a Local Authority/US County
- Policeman or woman
- High school teacher
- Encyclopedia salesperson
- Used car salesperson
- Senior religious figure (Bishop?)
- Attorney/Solicitor
- University Professor

It isn't possible to do this straight off, so what we suggest you do is a mental version of the algorithm known as a 'ripple sort'. Take the first pair of jobs and decide which is the most professional. If it's the first, then swap them round so that the more professional job is now second in the list. Next compare this second job with the third and repeat.

Proceed down the list making comparisons until the most professional job is in the last (tenth) position. Now repeat until you have the second most professional job in the second to last place, and so on until no more moves are possible. It helps to do this in a spreadsheet, or, alternatively a class can do it as a 'line up' in which ten people each represent a profession, but then line up in professional order as agreed and negotiated). Based simply on the pair-wise comparisons you should have a list in the required increasing order.

Number your list from 1 (least professional) to 10 (most professional).

Now repeat the same process, but this time rank in increasing order by the extent to which you expect the person in that job to behave ethically. In the context in which the word is being used here, ethics refers to a set of principles or morals and associated rules of conduct. Again number the jobs from 1 (least ethical) to 10 (most ethical) and now write this second ranking against the list relating to professionalism. Do they look similar?

We can correlate these rankings using Spearman's Rank Correlation coefficient,  $\rho$  (rho), based on the differences in the ranks. First, for each job find the difference in the ranks,  $d_i$ . Next use these numbers to find  $\rho$  as:

$$\rho = 1 - [ 6 \sum d^2 / (n^3 - n) ]$$

In this,  $n$  is the number of ranked pairs, in our case 10, and  $d^2$  is the difference in rank squared. A value of +1.0 indicates that you have ranked them in the same order, so that the correlation between the ranks is perfectly positive. If you get a value of -1.0 it indicates a perfect inverse ranking. Most likely, you got a value somewhere in between and if you're into statistics you might like to note that at the 95% confidence level, you need a value in excess of 0.65 to be sure that the relation is significant.

Draw some conclusions from this simple thought experiment such as:

Whatever you think about the words when applied to the GI industry, you can identify something called *professionalism* and something called *ethics*. If you can't, how else were you able to rank them?

The key to both is what the people concerned do in their jobs.

Different jobs have different ethical and professional expectations.

There is a correlation between ethics and professionalism. When we speak of someone as a 'professional' we imply that they also behave ethically.

Many legal systems are adversarial, and rely on argument to establish the rights and wrongs of a case. Consider the following hypothetical sequence of activities:

2. A GI data aggregator company uses GPS to create a road navigation and related attractions data base for a capital city. It then creates a database from these

data, spending almost two years to do this.

At the end of the two years, it decides that it can no longer afford to continue the data collection and acquires additional data, under a license to use, from a second company.

However, a third company that operates using the original data under licenses from the first produce their own data product that includes, predominantly, the information originally provided by the second company.

At this point the second company alleges that this third company has abused their intellectual property rights and demand recompense.

Have they? In which company is the IPR invested? What is the position of the original (first) company in all this? Arguing through this case should convince you that we have entered a legal minefield!

This is another exercise taken by permission of the School of Geography, Birkbeck College's GISOnline program and the AGI CPD scheme:

3. Investigate the published ethical codes for the following GIS-aware organisations:
  - a. Britain's Royal Institution of Chartered Surveyors at:  
<http://www.rics.org/ethics>
  - b. The dominantly US-based Urban and Regional Studies Association (URISA) at: <http://www.urisa.org/about/ethics>
  - c. American Society for Photogrammetry and Remote Sensing (ASPRS) at:  
[www.asprs.org/membership/certification/appendix\\_a.html](http://www.asprs.org/membership/certification/appendix_a.html)

Do these codes differ in any meaningful way? How are they enforced and what would be penalty be if a practitioner broke the relevant code?

4. Organize a debate on whether or not it should be necessary to be in some way licensed in order to practice using a GIS. Arguments for, and details of schemes that lead to certification, will be found at the ASPRS website:  
<http://www.asprs.org/membership/certification/>

The GIS Certification Institute at <http://www.gisci.org/> has relevant materials, including a database listing of all those individuals that have gained certification, which could form the basis of an interesting study.

For a robust, dissenting view, see:

Barr, R. (1995) Excuse me, do you have a licence to drive that GIS? *GIS Europe*, February 1995, 16-17.

## FURTHER READING

Cho G. 2005 Geographic Information Science: Mastering the Legal Issues. John Wiley & Sons Ltd., Chichester, UK.

Required reading for anyone with more than a passing interest in the legal implications of work with a GIS.

Photogrammetric Engineering and Remote Sensing. Special Issue on The National Map. 2003 (69).

Shapiro C. and Varian H.R. 1999 Information Rules. Cambridge, MA: Harvard Business School Press. The classic book on the nature of information and how to exploit it.

## RELATED READING

Maguire D.J., Goodchild M.F. and Rhind D.W. (eds) 1991 Geographical Information Systems: Principles and Applications. Harlow, UK: Longman (text available online at [www.wiley.co.uk/gis/volumes.html](http://www.wiley.co.uk/gis/volumes.html)).

Epilogue, D W Rhind, M F Goodchild and D J Maguire, pp. 313-27

## ONLINE RESOURCES

NCGIA Core Curriculum in GIScience, 2000 ([www.ncgia.ucsb.edu/giscc](http://www.ncgia.ucsb.edu/giscc))

3.2.1. [Public access to geographic information](#) (190), A Yeung

NCGIA Core Curriculum in GIS, 1990 ([www.ncgia.ucsb.edu/pubs/core.html](http://www.ncgia.ucsb.edu/pubs/core.html))

70. Legal issues

# GIS Partnerships

---

## OVERVIEW

- No single organization or individual can provide all the skills, tools, or knowledge required to carry out significant GIS-based projects.
- Describes the evolution of spatial data infrastructures

## LEARNING OBJECTIVES

- **The necessity for GIS partnerships and their potential benefits.**
- **The nature of the “big idea”—spatial data infrastructures (SDIs).**
- **The history of and differences in SDIs at global, continental, national, and local levels.**
- **The evolution in SDIs away from a technical solution to a technically supported social and institutional network.**
- **Other forms of partnership in GIS, notably commercial ones and combinations of individual volunteers exploiting technology to achieve critical mass.**

## KEYWORDS

PPGIS, NSDI, FGDC, standards, National Geospatial Data Clearinghouse, National Digital Geospatial Data Framework, INSPIRE, PCGIAP, DCW, ISCGM, NMO (National Mapping Organization), SRTM, OGC, GSDI and public access to GI.

## OUTLINE

### 19.1 Forcing Functions



## 19.2 Commercial Partnerships

## 19.3 Spatial Data Infrastructures

## 19.4 Partnerships of Individual Volunteers

## 19.5 Have SDIs Been a Success?

## 19.6 Nationalism, Globalization, Politics, and GIS

### 19.1 Forcing Functions

- There are multiple drivers behind the formation of partnerships
- Partnerships are easier if someone is in overall charge

### 19.2 Commercial Partnerships

- Massive consolidation of the industry has taken place since the early 2000s
  - This is illustrated with Nokia purchasing Navteq and TomTom purchasing TeleAtlas
  - Despite these acquisitions many GIS companies still operate partnerships
    - E.g. the Google business model, where data are acquired free or through license (see Section 18.5.1.2), processed, and then served to Google Earth and Google Maps users
- The authors describe hybrid forms of public-private partnership using the example of the Open Geospatial Consortium (OGC: [www.opengeospatial.org](http://www.opengeospatial.org))

### 19.3 Spatial Data Infrastructures

- SDI is the dominant conceptual model of GIS Partnerships for over the past 15 years
- The meaning of the term differs in different countries, though in essence it normally describes a widely available GIS search and mapping engine, GI and additional institutional and legal frameworks

#### 19.3.1 How it all began

- The authors describe the history of SDI from their emergence in the early 1990s

#### 19.3.2 SDI Partnerships at the Global Level

- Global-level partnerships comprise those:
  - based on executive mandate and finances to carry out particular tasks, and
  - that are in some sense voluntary.

### 19.3.2.1 Global Organizations and Their Partnerships

- Examples are given for Food and Agriculture Organization (FAO), the World Meteorological Organization (WMO), and the World Health Organization (WHO)

### 19.3.2.2 Global Voluntary Partnerships

- Describes a series of voluntary partnerships including:
  - Global Spatial Data Infrastructure Association (GSDI)
  - International Society for Digital Earth ([www.digitalearthisde.org](http://www.digitalearthisde.org))
  - International Steering Committee for Global Map
  - OpenStreetMap

### 19.3.3 SDI Partnerships at the Multicountry Level

#### 19.3.3.1 The European Dimension

- The status of SDIs in Europe as of 2007 is summarized in Figure 19.8
  - Most European countries have a fairly well coordinated SDI approach at the national level and also one or more of the SDI components at an operational level
- An EU directive introduced by the European Commission and approved by the European Parliament and the Council of Ministers must be implemented in national law within a defined period (normally two years).
  - The example of the INSPIRE directive is given which is designed to establish a legal framework for the establishment and operation of an INfrastructure for SPatial InfoRmation in Europe
- INSPIRE is designed to reduce the following barriers:
  - Inconsistencies in spatial data collection
  - Lack of documentation
  - Incompatible spatial datasets
  - Incompatible geographic information initiatives
  - Barriers to data sharing
- INSPIRE aims to overcome these barriers by:
  - Creation of metadata
  - Harmonizing key spatial data themes to support policy
  - Forming agreements on network services
  - Making policy agreements on sharing and access
  - Devising coordination and monitoring mechanisms
  - Creating the implementation process and procedures.
- The data types to be made available by all EU countries are listed in Table 19.1

#### **19.3.3.2 The Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP)**

- The authors describes the structure and activities of PCGIAP

#### **19.3.4 SDI Partnerships at the National Level**

- This section reviews the evolution of three national level SDI in the U.S., Singapore and the UK

#### **19.3.5 SDI Partnerships at the Subnational or Local Level**

- This section reviews the evolution of three sub national or local level SDI in Catalonia and Australia.

### **19.4 Partnerships of Individual Volunteers**

- Web 2.0 has transformed the potential for certain types of GI collection, leading to the concept of volunteered GI
  - The example of GISCorps is given which coordinates short-term, volunteer-based GIS services to underprivileged communities and MapAction which supports emergency aid relief using GIS

### **19.5 Have SDIs Been a Success?**

- The authors evaluate the success of SDI in this section. The key points are:
  - Something as diffuse as NSDI will never be seen as a success by everyone, but it has been a catalyst for many positive developments.
  - The impetus for greater government efficiency and effectiveness and the advent of the Department of Homeland Security reignited the cause of NSDI.
  - Highly partisan views exist on what NSDI has achieved and what should be done next. This is not surprising given its ambitious scope, its nature, and the lack of simple measures of success.

### **19.6 Nationalism, Globalization, Politics, and GIS**

- Other than for imagery and some roads data, global GI is currently little more than the sum of the highly varied national parts, and detailed consistent information rarely is readily available. This is a consequence of historical methods of data collection, plus nationally-focused policies and funding

## ESSAY TOPICS

1. Is there anything special about geographic information that makes partnership an attractive proposition?
2. Why might you wish to act in GIS partnership with other organizations? What are the likely benefits and disbenefits?
3. To what extent does the uptake and use of public participation GIS depend on free access to framework data?
4. It is one thing to share geospatial data, quite another to do so in such a way as to make these data effective. Why should data collected by one agency be of little or no use to another?
5. Describe what is meant by spatial data infrastructures?
6. To what extent do the governance systems of a country influence its ability to create a national geospatial infrastructure?
7. What are the prospects for the creation of a globally consistent set of framework data?
8. In the creation of spatial data infrastructures, it is possible to categorize the strategies adopted as 'top down' and 'bottom up'. With examples, outline the relative merits of each.
9. Has EITHER (a) NSDI in USA or (b) National Land Information System in UK or (c) an equivalent initiative in your own country been a success?
10. Does it seem likely to you that the most effective SDIs are those for parts of a country (e.g., states) rather than whole countries? If so, why?

## MULTIPLE CHOICE QUESTIONS (MCQ)

There are none for this chapter.

## ACTIVITIES

1. A structured comparison. Visit the website of the Geospatial Data Infrastructure Association at [www.gsdi.org](http://www.gsdi.org). There is a useful 'cookbook' about GSDI at [www.gsdi.org/gsdicookbookindex.asp](http://www.gsdi.org/gsdicookbookindex.asp). This has several case studies. Make a poster in which you compare and contrast these using as sub-headings:
  - a. Background, Context, and Rationale
  - b. Organizational Approach
  - c. Implementation - Approach
  - d. Components

- e. Implementation
  - f. Issues
2. In the late 1980s, there were a number of academic initiatives to develop partnerships that would build GIS expertise and infrastructure. In UK the Economic and Social Research Council initiated its Regional Research Laboratory Initiative and in USA the superficially-similar National Center for Geographic Information and Analysis (NCGIA) was established. See, for example, J Shepherd et al, 1989, "The ESRC's Regional Research Laboratories: An Alternative Approach to the NCGIA?," *AutoCarto 9*, Sydney, Australia. See also [www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u71.html](http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u71.html) on the Development of National GIS Policy. Use a search engine to trace how each initiative fared, and compare and contrast their individual legacies.
  3. There are many ways by which SDI can be classified, but one of the most useful looks as their status and scope. In terms of status, a distinction is made between those that have a legal mandate (as in USA and Portugal) and those that are an outgrowth of existing collaborations (as in Australia and the Netherlands). In terms of scope a distinction can be made between those that are broad (USA) and those that are narrow (Malaysia). Visit as many website as you can and 'populate' the following table with examples:

Legally mandated, narrow in scope	Legally mandated, broad in scope
Outgrowth, narrow in scope	Outgrowth, broad in scope

What local factors explain each entry in your table?

4. Using as many sources of public domain data as you are able, create a 'spatial data infrastructure' for the area around your home or place of study. 'Journal' the steps you take and the difficulties you encounter. If you are resident in USA, it is almost certain that your state has a geospatial data clearinghouse to which you should direct attention. In the UK Ordnance Survey makes available multiple sets of data under its OpenData scheme; much other data are available from [data.gov.uk](http://data.gov.uk)

## FURTHER READING

Masser I. 2005 Spatial Data Infrastructure: An Introduction. Redlands, CA: ESRI Press.

NRC 2004 Licensing geographic data and services. National Research Council of the National Academies, Division of Earth and Life Sciences. National Academies Press, Washington DC.

Craglia M. and Campagna M. 2009. *Advanced Regional Spatial Data Infrastructures in Europe*. Joint Research Center of the European Union; available at [sdi.jrc.ec.europa.eu/ws/Advanced\\_Regional\\_SDIs/arsdi\\_report.pdf](http://sdi.jrc.ec.europa.eu/ws/Advanced_Regional_SDIs/arsdi_report.pdf).

Masser I., Rajabifard A., and Williamson I. 2007. Spatially enabling governments through SDI implementation. *International Journal of Geographical Information Science*, 21: 1–16.

## RELATED READING

Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, NJ: Wiley.

56. National and international geospatial data policies, D W Rhind, pp. 767–787  
Maguire D.J., Goodchild M.F. and Rhind D.W. (eds) 1991 *Geographical Information Systems: Principles and applications*. Harlow, UK: Longman (text available online at [www.wiley.co.uk/gis/volumes.html](http://www.wiley.co.uk/gis/volumes.html)).

47. Integration of geoscientific data using GIS, G F Bonham-Carter, pp. 171-84

56. Integrated planning information systems, D J Cowen and W L Shirley, pp. 297-310

# Epilogue: GIS&S in the Service of Humanity

---

## OVERVIEW

- Relates the key ideas introduced in the book to the grand challenges of GIS&S
- Describes the importance and value of inculcating spatial thinking in all strands of education

## LEARNING OBJECTIVES

- **Just how differentiated world geography is, notably in the incidence of poverty and disease.**
- **The biggest challenges facing humanity.**
- **The interdependency between many factors causing these problems.**
- **How GI systems and science (GIS&S) can help us tackle these challenges.**
- **A number of issues that we need to resolve to progress.**

## KEYWORDS

Poverty, Education, Social Equality, Health and Wellbeing, Global Partnership

## OUTLINE

- 20.1 Introduction
- 20.2 The Differentiated World
- 20.3 Grand Challenges
- 20.4 Seeking the Root Causes
- 20.5 Meeting the Challenges
- 20.6 Conclusions

### 20.1 Introduction

- The authors ask what GIS can, should, or will be able to do for humanity faced with profound global challenges
- The four big contributions that geographical information systems and science (GIS&S) can make are identified as:
  - The ability to help us discover and share new understandings in the physical, environmental, and social sciences.
    - From these, the means to help us devise new products and services that improve the quality of life, especially for the disadvantaged of the world.
    - Use of these new products or services to enhance the efficiency of public and private tasks so as to release resources for other valued things.
    - Achievement of all this in as sustainable a way as possible.

### 20.2 The Differentiated World

- The authors argue that the world is a very differentiated, heterogeneous, and hence unequal place.
  - GIS&S is fundamental to the conception, representation, and analysis of geographic variations in ways that are robust, transparent, and potentially open to scrutiny by every stakeholder

### 20.3 Grand Challenges

#### 20.3.1 The Global View of Governments

- Millennium Development Goals are discussed which include:
  - Goal 1: Eradicate extreme poverty and hunger.
  - Goal 2: Achieve universal primary education.
  - Goal 3: Promote gender equality and empower women.
  - Goal 4: Reduce child mortality.



- Goal 5: Improve maternal health.
- Goal 6: Combat HIV/AIDS, malaria, and other diseases.
- Goal 7: Ensure environmental sustainability (this is expanded since it is key to what follows).
- Goal 8: Develop a Global Partnership for Development.
- Progress in achieving these goals has been slower than anticipated...

### 20.3.2 Challenges Amenable to Use of GIS&S

- The authors describe interdependencies between the challenges and how GIS practitioners might make a contribution to ameliorating them.

#### 20.3.2.1 Poverty and Hunger

- Almost half the world—over 3 billion people—live on less than \$2.50 a day, while at least 80% of humanity lives on incomes of less than \$10 per day.
- 80% of the world's population lives in countries where income differentials are widening
- Mapping using GISystems adds greater richness and diversity to statistics such as these as illustrated in Figure 20.4
- Poverty also generally means poor levels of nutrition
- GIS contributes to our understanding of poverty and inequality because it improves our ability to integrate, analyze, and portray multiple datasets and to support logistics through the use of our tools
  - enables us to move from what-is descriptions to what-if analysis

#### 20.3.2.2 Population Growth

- The global population as of January 2010 was estimated by the U.S. Bureau of Census to be 6.796 billion.
  - Many demographers now expect the total to level off to somewhere around 10 billion after the middle of the Twenty-First Century
- The geography of population change, as well as population density, is far from uniform
- GIS provides an excellent instrument for measuring the ebbs and flows of population movements
  - When measuring people by surveys getting reliable numbers is becoming more difficult; using multiple data sources is becoming highly sensible
  - GIS can be used to provide better linkage of data
- GIS can also be used to help understand why people move and examine the impacts of changed laws, other barriers, or incentives

### **20.3.2.3 Disease**

- The spread of disease, access to care, and the treatment and prevention of illness are unevenly distributed across the globe
  - For example, every year there are 350–500 million cases of malaria in the world, with 1 million fatalities: Africa accounts for 90% of malarial deaths, and African children account for over 80% of malaria victims worldwide
- The science underlying GIS is once again important for what-if analysis of diffusion processes and remedial action scenarios

### **20.3.2.4 Access to Potable Water**

- Water problems affect half of humanity: some 1.1 billion people in developing countries have inadequate access to water, and 2.6 billion lack basic sanitation.
- These problems are like to increase as a result of competing uses and climate change
- GIS has been used in climate projections using the best available data and models
- GIS practitioners have made significant contributions to the definition of international boundaries, many of which have followed the course of rivers and are foci of tension

### **20.3.2.5 Natural Disasters**

- GIS technology has imposed itself as an essential tool for all actors involved in the different sectors of the disaster and conflict management cycle. For example:
  - Probabilistic statements of the likelihood of a disaster
  - Improving the effectiveness of post-disaster recovery work

### **20.3.2.6 Environmental Sustainability**

- GIS plays a significant role in monitoring the state of the environment both globally and locally
  - For example, examining the magnitude of biodiversity and ecosystem loss and in developing possible coping strategies
  - To succeed, GIS&S specialists must work with other natural, environmental, and social scientists, and with decision makers.

### **20.3.2.7 Terrorism and Crime**

- Attempts to overthrow governments or ways of life by violence are commonplace in many parts of the world.
- Most organizations related to security or defence are now assiduous users of GIS technology and related sciences.
  - For example increasing use of surveillance technology and sensors

- Surveillance of individuals can reduce risk of harm but endangers privacy.

## 20.4 Seeking the Root Causes

- It is normal in science to look for the root causes rather than the symptoms of problems
- However, the vast range of causal factors of the issues described above, the uncertainty involved in characterizing their geographic nature is considerable, and their only partly understood interdependencies and feedback loops make defining root causes difficult.
- Today's world is tightly interlinked: a disaster in one geographic area can trigger global impacts. This has major governance implications given the responsibilities of nation states.

## 20.5 Meeting the Challenges

- The authors describe how global problems face us all and that, while some challenges might seem consistently threatening and long-lasting, sudden changes of a catastrophic nature will occur
- Addressing these challenges requires collaborations that must be multidisciplinary, and geographically extensive in scale.

### 20.5.1 Why GIS&S Should Enable Us to Make a Difference

- The nature of the Grand Challenges enables those active in GIS&S to make particular contributions for the following reasons:
  - Many problems are manifested initially through geographical variations
  - Studying the geographical manifestation can help us to propose and test causal factors and hence identify possible solutions to the problems
  - The mechanics by which we have to tackle problems are normally geographically structured
  - There has been a significant change, at least in Western democracies, toward the requirement for quantifiable evidence to support and justify policymaking
- The potential contribution of GIS&S to the Grand Challenges is simply formalised in a 5 stage model highlighted in Figure 20.18

## 20.6 Conclusions

The technology and scientific understanding are necessary but not sufficient conditions for progressing. We need other skills (e.g. skilled and trusted communicators and educators)

and we need to embed spatial thinking as part of the curricula of many subjects across the world. Some specific challenges we face are:

- How do we support decision making where both quantitative and qualitative information needs to be combined in the analysis?
- How do we help to break down the silo thinking between different scientists and others who are necessarily working on meeting these challenges?
- How do we encourage all GIS professionals to become ambassadors to their clients and to the public at large for the contributions we can make?

Notwithstanding the difficulties, the authors conclude that GIS&S professionals can make a real difference in tackling the Grand Challenges by virtue of their skills, tools and commitment. They invite you to play your part.

### ESSAY QUESTIONS

1. Why is the world so differentiated?
2. What do you think are the biggest challenges to humanity where GIS can make a contribution?
3. What is spatial thinking? Can you give some examples?
4. At the moment, only a small fraction of the global population is impacted by GIS. Would a world in which everyone had access to GIS be a better or a worse place and why?
5. Why should GIScience researchers bother about issues to do with GIS curricula?

### MULTIPLE CHOICE QUESTIONS (MCQ)

There are none for this chapter.

#### Activities

1. You have been asked by the *International Journal of Geographic Information Science* to write a review of this book intended for the readership of that journal. The review editor is prepared to allow you 800 words of text. Provide him or her with a review.
2. Devise a curriculum for use in higher education for a year's course in geographic information science aimed at the senior undergraduate level. The class should meet formally for around three hours in each of, say, 24 weeks. In doing this, note that the concept of a 'curriculum' is wider than a list of lesson topics, and resist the temptation to list 'content', but instead take an approach based on intended learning objectives. See:

[www.ncgia.ucsb.edu/giscc/units/u159/u159\\_f.html](http://www.ncgia.ucsb.edu/giscc/units/u159/u159_f.html) and Chapter 10 of Gold et al (1992), at [www2.glos.ac.uk/gdn/gold/ch10.htm](http://www2.glos.ac.uk/gdn/gold/ch10.htm)

3. Repeat Activity (2), but for an introductory course aimed at the first year in college in geographic information studies. The idea of GISudies is developed in Forer, P. and D. Unwin (2005) Enabling progress in GIS and education, in: Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, NJ: Wiley., pages 745-756
4. The table below lists the grand challenges facing humanity as identified by the Millennium Development Goals. Attempt to rank them in order from most (1) to least (8) in orders of their difficulty and importance to humanity – and also on which ones are most amenable to help from GIS&S.

Challenge	Difficulty	Importance
Goal 1: Eradicate extreme poverty and hunger.		
Goal 2: Achieve universal primary education.		
Goal 3: Promote gender equality and empower women.		
Goal 4: Reduce child mortality.		
Goal 5: Improve maternal health.		
Goal 6: Combat HIV/AIDS, malaria, and other diseases.		
Goal 7: Ensure environmental sustainability.		
Goal 8: Develop a Global Partnership for Development.		

It is best to do this by systematic pair-wise comparison in a ‘ripple sort’

Use rank correlation to correlate these scores. What do you conclude from this analysis?

Do you agree with our choice of challenges?

5. On the Sunday morning of 26 December 2004 an undersea earthquake triggered a tsunami that devastated many coastal areas surrounding the Indian Ocean and killing

around a quarter of a million people. 'Brainstorm' a list of the ways by which GIS could supply information to help mitigate the worst effects of this disaster.

6. Organize a debate on the proposition that 'This house believes that GIS has no place in a modern university department of geography'.
7. It is always useful to make a personal evaluation of your learning. This text has brought you a long way along a sometimes twisty, but always interesting, road. For each of the 20 chapters re-examine the stated learning objectives and evaluate the extent to which you think you have achieved them.

### **FURTHER READING**

Nyerges T. 2001 *Geographic Information Systems for Group Decision Making*. London: Taylor and Francis.

Friedman T.L. 2007. *The World Is Flat: A Brief History of the Twenty-First Century* (3rd ed.). London: Picador.

NRC. 2009. *Understanding and Adapting to a Changing Planet: Strategic directions for the Geographical Sciences in the Next Decade*. Washington, DC: National Research Council, National Academy of Sciences.

Sachs J.D. 2008. *Common Wealth: Economics for a Crowded Planet*. London: Penguin.

### **RELATED READING**

Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, NJ: Wiley.

3. Geography and GIS, R J Johnston, pp. 39–47.

54. Enabling progress in GIS and education, P Forer and D Unwin, pp. 745-756