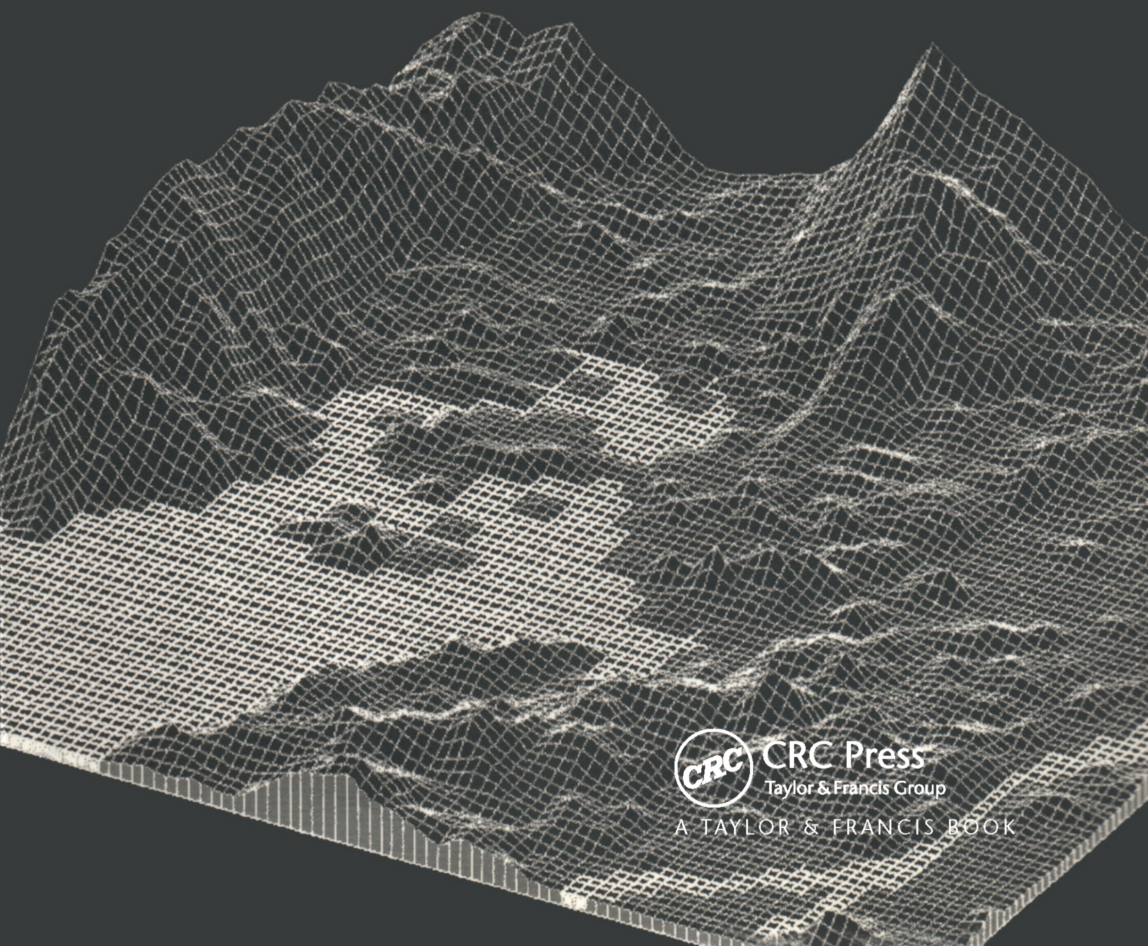
GIS, Environmental Modelling and Engineering

Allan Brimicombe



CRC Press
Taylor & Francis Group
A TAYLOR & FRANCIS BOOK

GIS, Environmental Modelling and Engineering

GIS, Environmental Modelling and Engineering

Allan Brimicombe

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A TAYLOR & FRANCIS BOOK

First published 2003 by Taylor & Francis

Published 2018 by CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

ISBN-13: 978-0-415-25923-1 (pbk)
ISBN-13: 978-0-415-25922-4 (hbk)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

British Library Cataloguing in Publication Data
A catalogue record for this book is available
from the British Library

Library of Congress Cataloging in Publication Data
A catalog record for this book has been requested

Typeset in Sabon by
Integra Software Services Pvt. Ltd, Pondicherry, India

Contents

Acknowledgements

First, a heartfelt thanks to my wife Lily for her unwavering support in this venture and for her hard work in preparing most of the Figures.

Second, I would like to thank my colleague Dr Yang Li for his assistance with some of the Figures and particularly for the preparation of the coastal oil-spill modelling examples.

Third, I would like to thank Professor Li Chuan-tang for his invaluable insights into finite element methods.

Fourth, I would like to thank my sequential employers – Binnie & Partners International (now Binnie Black & Veatch, Hong Kong); Hong Kong Polytechnic University; University of East London – for providing me with the opportunities and space to do so much.

Finally, my thanks to Sarah Kramer at Taylor & Francis for her patience.

Statement on trade names and trademarks

In a book such as this it is inevitable that proprietary or commercial products will be referred to. Where a name is used by a company to distinguish their product which they may claim as a trade name or trademark, then that name appears here with an initial capital or all capital letters. Readers should contact the appropriate companies regarding complete information. Use of such names is to give due recognition to these products in illustrating different approaches and concepts and providing readers with practical information. Mention of proprietary or commercial products does not constitute an endorsement, or indeed, a refutation of these products.

1 Introduction

I wish to begin by explaining why this book has been written. Peter Fleming, in writing about his travels in Russia and China in 1933, put the need for such an explanation this way:

> With the possible exception of the Equator, everything begins somewhere. Too many of those too many who write about their travels plunge straight *in medias res*; their opening sentence informs us bluntly and dramatically that the prow (or bow) of the dhow grated on the sand, and they stepped lightly ashore. No doubt they did. But why? With what excuse? What other and anterior steps had they taken? Was it boredom, business, or a broken heart that drove them so far afield? We have a right to know.
>
> (Peter Fleming (1934) *One's Company*)

The more commercially minded reader would say that this book was written purely because the publisher and I considered there was a market out there. The fact that there is such an audience and because I feel I have something important to tell them, both stem from the same source: a widespread concern for the quality of world we live in, an urgent need for its maintenance and where necessary, repair. The phrase 'quality of world' is left intentionally broad, even ambiguous. It encompasses:

- our natural environment – climate, soils, oceans, biological life (plants, animals, bacteria) – that can both nurture us and be hazards to us;
- the built environment that we have created to protect and house ourselves and to provide a modified infrastructure within which we can prosper;
- the economic environment that sustains our built environment and allows the organisation of the means of production;
- the social, cultural and legal environments within which we conduct ourselves and our interactions with others.

These environments are themselves diverse, continually evolving and have strong interdependence. Each of them varies spatially over the face of the globe mostly in a transition so that places nearer to each other are more likely to be similar than those further apart. Some abrupt changes do of course happen, as for example between land and sea. They also change over time, again mostly gradually but catastrophic events and revolutions do happen. Together they form a complex mosaic, the most direct visible manifestation being land cover and land use – evolved cultural landscapes. Furthermore, the interaction of these different aspects of environment gives enormous complexity to the notion of 'quality of life' for our transient existence on Earth. Globalisation may be a new force for uniformity in business and consumerism, but even then businesses have had to learn to be spatially adaptive. When it comes to managing and ameliorating our world for sustainable quality of life, there is no single goal, no single approach, no theory of it all. Let's not fight about it. Let us celebrate our differences and work towards a common language of understanding on how we (in conjunction with the rest of nature) are going to survive and thrive.

Metaphors of nature

We often use metaphors as an aid to understanding complexity, none more so perhaps than in understanding nature and our relationship within it. These metaphors are inevitably bound up in philosophies of the environment, or knowledge of how the environment works and the technology available to us to modify/ameliorate our surrounding environment. Thus for millennia, environmental knowledge was enshrined in folklore derived from the trial and error experiences of ancestors. Archaeology has revealed patterns of site selection which changed as we developed primitive technologies or adapted to new environments. Places for habitation had to satisfy the needs for water, food, raw materials, shelter and safety, and humans learnt to recognise those sites that offered the greatest potential for their mode of existence. Examples are numerous: caves near the feeding or watering places of animals; Neolithic cultivation of well-drained, easily worked river terraces; early fishing communities on raised beaches behind sheltered bays and so on. Undoubtedly mistakes were made and communities decimated, but those that survived learnt to observe certain environmental truths or inevitabilities.

Successful early civilisations were those that had social structures that allowed them to best use or modify the landforms and processes of their physical environment. Thus the Egyptians, Mesopotamians and Sumarians devised irrigation systems to regulate and distribute seasonally fluctuating water supplies, whilst the Chinese and Japanese included widespread terracing as a means of increasing the amount of productive land. More than 2500 years ago the Chinese developed the Taoist doctrine of nature, in which the earth and the sky had their own 'way' or 'rule' to maintaining harmony. Human beings should follow and respect nature's way or risk punishment in the form of disasters from land and sky. Thus, even at that time there were laws governing, for example, minimum mesh size on fishing nets so that fish would not be caught too young. Of course, our stewardship has not always been a continual upward journey of success. Some human civilisations have collapsed spectacularly through environmental impact and loss of natural resources (Tickell, 1993). These disasters aside, the dominant metaphor was of 'Mother Earth': a benevolent maker of life, a controlling parent that could provide for our needs, scold us when we erred and, when necessary, put all things to right.

The industrial revolution allowed us to ratchet up the pace of development. Early warnings of the environmental consequences, such as from Marsh (1864), were largely ignored as the Victorians and their European and North American counterparts considered themselves above nature in the headlong rush to establish and exploit dominions. Our technologies have indeed allowed us to ameliorate our lifestyle and modify our environment on an unprecedented scale – on a global scale. But from the 1960s the cumulative effect of human impact on the environment and our increasing exposure to hazard finally crept onto the agenda and remains a central issue today. The rise of the environmental movement brought with it a new metaphor 'Spaceship Earth' which was inspired by photographs from the Apollo moon missions of a small blue globe rising above a desolate moonscape. We were dependant on a fragile life-support system with no escape, no prospect of rescue, if it were to irreparably break down. This coincided with the publication of seminal works such as Rachel Carson's (1963) *Silent Spring* which exposed the effects of indiscriminate use of chemical pesticides and insecticides, McHarg's (1969) *Design with Nature* which exhorted planners and designers to conform to and work within the capacity of nature rather than compete with it and Schumacher's (1973) *Small is Beautiful* proposed an economics which emphasised people rather than products and reduced the squandering of our 'natural capital'.

The words fractal, chaos, butterfly effect and complexity (Mandelbrot, 1983; Gleick, 1987; Lewin, 1993; Cohen and Stewart, 1994) have since been added to the popular environmental vocabulary to explain the underlying structure and workings of complex phenomena. Added to these is the Gaia hypothesis (Lovelock, 1988) in which the Earth is proposed to have a global physiology or may in fact be thought of as a superorganism capable of switching states to achieve its own goals in which we humans may well be (and probably are) dispensable organisms.

A solution space?

That we are capable of destroying our life support system is beyond doubt. As a species, we have already been responsible for a considerable number of environmental disasters. If I scan the chapter titles of Goudie's (1997) *The Human Impact Reader*, the list becomes long indeed, including (in no particular order): subsidence, sedimentation, salinisation, soil erosion, desiccation, nutrient loss, nitrate pollution, acidification, deforestation, ozone depletion, climate change, wetland loss, habitat fragmentation and desertification. I could go on to mention individual events such as Exxon Valdez and Chernobyl, but this book is not going to be a catalogue of dire issues accompanied by finger-wagging exhortations that something must be done. At the time of writing this introduction, the President of the United States, George, W. Bush, has already rejected the Kyoto Agreement on the control of greenhouse gas emissions, European leaders appear to be in a dither and eco-warriors alongside anti-capitalists have again clashed with riot police in the streets. One can be forgiven for having an air of pessimism – the environment is in trouble and it's likely to get a lot worse before we start seeing any signs of improvement, if at all. But we are far from empty-handed. We have a rich heritage of science and engineering, a profound knowledge of environmental processes and experience of conservation and restoration. The technologies that have allowed humankind to run out of control in its impact on the environment can surely be harnessed to allow us to live more wisely. Our ingenuity got us here and our ingenuity will have to get us out of it.

As stated above, we need a common language and in this regard we have some specific technologies, drawing upon science, that can facilitate this. Whilst humankind has long striven to understand the workings of the environment, it has only been in the last 30 years or so that our data collection and data processing technologies have allowed us to reach a sufficiently detailed understanding of environmental processes so as to create *simulation* models. I would argue that it is only when we have reached the stage of successful quantitative simulation, can our level of understanding of processes allow us to confidently manage them. This is the importance of *environmental modelling*. Facilitated by this in a parallel development has been *environmental engineering*. Engineering also has a rich history but whilst traditionally engineering has focused on the utilisation of natural resources, environmental engineering has recently developed into a separate discipline that focuses on the impact and mitigation of environmental contaminants (Nazaroff and Alvarez-Cohen, 2001). Whilst most management strategies arising out of environmental modelling will usually require some form of engineering response for implementation, environmental engineering provides solutions for managing water, air and waste. *Engineering* in the title of this book refers to the need to design solutions more so in as much as those designs are often informed by computational or simulation modelling. The youngest technology I would like to draw into this recipe for a common language is *geographical information systems* (GIS). Since environmental issues are inherently spatial – they occur *somewhere*, often affecting a geographical *location* or *area* – their spatial dimension needs to be captured if modelling and engineering are to be relevant in solving specific problems or avoiding future impacts. GIS have proved successful in the handling, integration and analysis of spatial data and have become an easily accessible technology. Whilst the link between simulation modelling and engineering has been longstanding, the link between GIS and these technologies is quite new, offers tremendous possibilities for improved environmental modelling and engineering solutions and can help build these into versatile decision support systems for managing, even saving our environment. And that is why this book has been written.

Scope and plan of this book

From the early 1990s onwards there has been an accelerating interest in the research and applications of GIS in the field of environmental modelling. There have been a few international conferences/workshops on the subject – most notably the series organised by the National Centre for Geographic Information and Analysis (NCGIA), University of California, Santa Barbara in 1991, 1993, 1996 and 2000 – and have resulted in a number of edited collections of papers (Goodchild *et al.*, 1993, 1996; Haines-Young *et al.*, 1993; NCGIA, 1996, 2000) as well as a growing number of papers in journals such as the *International Journal of Geographical Information Science, Hydrological Processes, Computers Environment and Urban Systems, ASCE Journal of Environmental Engineering, Photogrammetric Engineering and Remote Sensing* and so on. But working with GIS and environmental simulation models is not just a case of buying some hardware, some software, gathering some data, putting it all together and solving problems with the wisdom of a sage. Whilst technology has simplified many things, there still remain many pitfalls and users need to be able to think critically about what they are doing and the results that they get from the technology. Thus the overall aim of this book is to provide a structured, coherent text that not only introduces the subject matter but also guides the reader through a number of specific issues necessary for critical usage.

This book is aimed at final year undergraduates, postgraduates and professional practitioners in a range of disciplines from the natural sciences, social sciences to engineering, at whatever stage in their lifelong learning or career they need or would like to start working with GIS and environmental models. The focus is on the use of these two areas of technology together and the issues that arise in so doing. This book is less concerned with the practicalities of software development and the writing of code (e.g. Payne, 1982; Kirkby *et al.*, 1987; Hardisty *et al.*, 1993; Deaton and Winebrake, 2000). Nor does it consider in any detail the practicalities of data collection such as by remote sensing, GPS, data loggers and so on as there are numerous texts that already cover this ground (e.g. Anderson and Mikhail, 1998; Skidmore, 2002).

The overall thrust of this book can be summarised in the mapping:

$$f : \Omega \rightarrow \Re \tag{1.1}$$

where Ω = set of domain inputs, $\Re$ = set of real decisions.

The subject matter is laid out in three sections. Section A concentrates uniquely on GIS: what they are, how data are structured, what the most common types of functionality are. GIS will be viewed from the perspective of a technology, the evolution of its scientific basis and latterly its synergies with other technologies within a geocomputational paradigm. This is not intended to be an exhaustive introduction as there are now many texts that do this (e.g. Chrisman, 1997; DeMers, 1997; Burrough and McDonnell, 1998; Heywood *et al.*, 1998; Longley *et al.*, 2001). Rather, its purpose is to lay a sufficient foundation of GIS for an understanding of the substantive issues raised in Section C. Section B similarly focuses on modelling both from a neutral scientific perspective of its role in simulating and understanding phenomena and from a more specific perspective of environmental science and engineering. Section C is by far the largest. It looks at how GIS and simulation modelling are brought together, each adding strength to the other. There are examples of case studies and chapters covering specific issues such as interoperability, data quality, model validity, space-time dynamics and decision-support systems. Those readers who already have a substantial knowledge of GIS or have completed undergraduate studies in GIS may wish to skip much of Section A and move quickly to Sections B and C. Those readers from a simulation modelling background in environmental science or engineering should read Section A, skim through Section B and proceed to Section C.

In a book such as this, it is always possible to write more about any one topic, there are always additional topics that a reader might consider should be added. There are, for example, as many environmental models as there are aspects of the environment. GIS, environmental modelling and engineering are quite endless and are themselves evolving. Also, I have tried not to focus on any one application of simulation modelling. Given its popularity, there is a temptation to focus on GIS and hydrology, but that would detract from the overall purpose of this book which is to focus on generic issues of using GIS and external models to solve real problems. Presented in the following chapters is what I consider to be a necessary understanding for critical thinking in the usage of such systems and their analytical outputs. Enjoy.

2 From GIS to geocomputation

The cosmological event of the Big Bang created the universe and in so doing space–time emerged (some would say 'switched on') as an integral aspect of gravitational fields. Space and time are closely interwoven and should more properly be thought of as a four-dimensional (4-D) continuum in which time and space, over short durations, are interchangeable. Nevertheless, we conventionally think of separate one-dimensional (1-D) time and three-dimensional (3-D) space. The terrestrial space on which we live, the Earth, is at least 4.5 billion years old and has been around for about 40 per cent of the time since time began. Since our earliest pre-history we have grappled with the problems of accurately measuring time and space. Crude measures of time probably came first given the influences of the regular cycles of the day, tides, the moon, and seasons on our lives as we evolved from forager to agriculturist. With technology, we have produced the atomic clock and the quartz watch. Measuring position, distances and area were less obvious in the absence of the type of benchmark that the natural cycles provided for time. Early measurements used a range of arbitrary devices – the pace, the pole, the chain – and longer distances tended to be equated with the time it took to get to destinations. Even much later, the development of accurate clocks were key to solving the problem of determining longitudinal position when coupled with observations of the sun. Measurement requires numerical systems, and 1-D time requires either a linear accumulation (e.g. age) or a cyclical looping (e.g. time of day). Measurement of 3-D space requires the development of higher order numerical systems to include geometry and trigonometry. Let us not forget that at the root of algebra and the use of algorithms was the need for precise partitioning of space (land) prescribed by Islamic law on inheritance. Calculus was developed with regard to the changing position (in time) of objects in space as a consequence of the forces acting upon them.

Three fundamental aspects of determining position are: a datum, a co-ordinate system (both incorporating units of measurement), and an adequate representation of the curved (or somewhat crumpled) surface of the Earth in the two dimensions of a map, plan or screen. The establishment of a datum and co-ordinate system is rooted in geodetic surveying which aims to precisely determine the shape and area of the Earth or a portion of it through the establishment of wide-area triangular networks by which unknown locations can be tied into known locations. Cartographers aim to represent spatial features and their relationships on a plane. This involves the art of reduction, interpretation and communication of spatial features and the science of transforming co-ordinates from the spherical to the plane through the construction and utilisation of map projections. The production of quality spatial data used to be a time-consuming, expensive task and for much of the twentieth century there was a spatial data 'bottleneck' that held back the wider use of such data. Technology has provided solutions in the form of *global positioning system* (GPS), electronic total stations, *remote sensing* (RS), digital photogrammetry and GIS. GPS, RS and GIS are now accessible to every citizen through inexpensive devices and the Internet. Determining *where* is no longer difficult and perhaps soon the use of position and location will be no more difficult than telling the time.

This chapter will chart the rise of GIS as a *technology*, consider its main paradigms for representing the features of the Earth and structuring data about them. The basic functionality of GIS will be described with examples. A 'systems' view of GIS will then be developed bringing us to the point where GIS can be formally defined. The limitations of modern GIS will be discussed leading us to consider the rise of geocomputation as a new paradigm and the role of GIS within it.

In the beginning . . .

It would be nice to point to a date, a place, an individual and say 'That's where it all started, that's the father of GIS!'. But no. As Coppock and Rhind put it in their History of GIS (Coppock and Rhind, 1991) – 'unhappily, we scarcely know'. In the beginning, of course, there were no 'GIS experts' and nobody specifically set out to develop a new body of technology nor a new scientific discipline for that matter. In the mid-1960s there were professionals from a range of disciplines, not many and mostly in North America, who were excited by the prospect of handling spatial data digitally. There were three main focal points: the Harvard Graduate School of Design, the Canada Land Inventory and the U.S. Census Bureau. In each of these organisations were small groups of pioneers who made important contributions towards laying the foundations for today's GIS industry.

The significance of the Harvard Graduate School of Design lies in its Laboratory for Computer Graphics and Spatial Analysis, a mapping package called SYMAP (1964), two prototype GIS called GRID (1967) and ODYSSEY (c.1978) and a group of talented individuals within the laboratory and the wider graduate school: N. Chrisman, J. Dangermond, H. Fisher, C. Steinitz, D. Sinton, T. Peucker and W. Warntz to name a few. The creator of SYMAP was Howard Fisher, an architect. His use of line printers to produce three types of map – isoline, choropleth and proximal – were a way of visualising or recognising spatial similarities or groupings in human and physical phenomena (McHaffie, 2000). The other leap was a recognition (rightly or wrongly) that just about any such phenomenon, no matter how ephemeral or whether described quantitatively or qualitatively could be represented as a map of surfaces or regions. The printing of these maps using equally spaced characters or symbols line by line naturally resulted in a 'blocky' cell-based map representation (Figure 2.1). David Sinton, a landscape architect, took cell-based (raster) mapping forward with GRID which allowed analyses to include several thematic data sets for a given area (layers). Furthermore, by 1971 a rewrite of GRID allowed users to define their own logical analyses rather than being restricted to limited pre-packaged procedures. A flexible user interface had thus been developed. By the late 1970s ODYSSEY, a line-based (vector) GIS prototype had been written capable of polygon overlay. In this way, it can be seen that the overlay or co-analysis of several thematic layers occupied the heart of early GIS software strategies (Chrisman, 1997).

In 1966, the *Canada Geographic Information System* (CGIS) was initiated to serve the needs of the Canada Land Inventory to map current land uses and the capability of these areas for agriculture, forestry, wildlife and recreation (Tomlinson, 1984). Roger Tomlinson had recognised some years earlier that the manual map analysis tasks necessary for such an inventory over such a large area would be prohibitively expensive and that a technological solution was necessary. Within this solution came a number of key developments: optical scanning of maps, raster to vector conversion, a spatial database management system and a seamless coverage spatially partitioned into 'tiles'. The

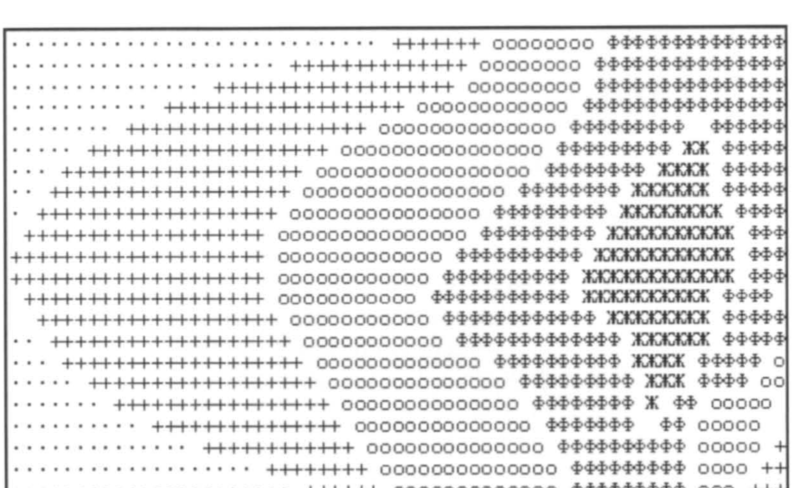


Figure 2.1 Sample of a SYMAP-type line printer contour map showing emphasis on similarities; the contour lines are perceived only through the 'gap' between the areas of printed symbols.

system was not fully operational until 1971 but has subsequently grown to become a digital archive of some 10,000 maps (Coppock and Rhind, 1991).

The significance of the U.S. Bureau of Census in developing its Dual Independent Map Encoding (DIME) scheme in the late 1960s is an early example of adding additional information on spatial relationships to data files through the use of topological encoding. Early digital mapping data sets had been unstructured collections of lines that simply needed to be plotted with the correct symbology for a comprehensible map to emerge. But the demands for analysis of map layers in GIS required a structuring that would allow the encoding of area features (polygons) from lines and their points of intersection, ease identification of neighbouring features and facilitate the checking of internal consistency. Thus DIME was a method of describing urban structure, for the purposes of census, by encoding the topological relationships of streets, their intersection points at junctions and the street blocks and census tracts that the streets define as area features. The data structure also provided an automated method of checking the consistency and completeness of the street block features (U.S. Bureau of Census, 1970). This laid the foundation of applying topology or graph theory now common in vector GIS.

Technological facilitation

The rise of GIS cannot be separated from the developments in information and communication technology that have occurred since the 1960s. A timeline illustrating developments in GIS in relation to background formative events in technology and other context is given in Table 2.1. Most students and working professionals today are familiar at least with the PC or Mac. I am writing this in 2002 on a note-book PC (400 MHz, 160 MB RAM, 5 GB disk), my GIS and environmental modelling workhorse is an NT workstation (860 MHz, 1 GB RAM, 60 GB disks) – they both run the same software with a high degree of interoperability, they both have the same look and feel with toolbars, icons and pulldown menus. Everything at a click of a mouse. I can send files between them by email (I just plug my mobile phone into the laptop) and at the same time I can look up just about anything on the Internet. Even my junk mail has started to arrive on CD, so cheap and ubiquitous has this medium become. It all takes very little training and most of the basic functions have become intuitive. I'm tempted to flex my muscles (well, perhaps just exercise my index finger) for just a few minutes on the GIS in this laptop . . . and have indeed produced Figure 2.2 – a stark contrast to Figure 2.1.

To fully comprehend the technological gulf we have crossed, let me briefly review a late 1970s GIS-based land capability study in South Dakota (Schlesinger *et al.*, 1979). The project was carried out on an IBM 370/145 mainframe computer using ten standalone programme modules written in FORTRAN IV and IBM Assembler. A digitising tablet and graphics terminal were available but all hardcopy maps were produced using a line printer. Maps wider than a 132 character strip had to be printed and glued together.

Table 2.1 Timeline of developments in GIS in relation to background formative events in technology and other context

Year	*GIS*	*Context*	
1962		Carson's *Silent Spring*	
1963	Canadian Geographic Information System		
1964	Harvard Lab for Computer Graphics & Spatial Analysis	GPS specification	
1966	SYMAP	WGS-66	
1967	US Bureau of Census DIME		
1968		Relational database defined by Codd	
1969	ESRI founded; Intergraph founded; Laser-Scan founded	Man on the Moon; NEPA; McHarg's *Design with Nature*	
1970	Acronym GIS born at IGU/UNESCO conference	Integrated circuit	
1971		ERTS/Landsat 1 launched	
1973	UK Ordnance Survey starts digitising		
1974	AutoCarto conference series; *Computers & Geosciences*	UNIX	
1975		C++	
1978	ERDAS founded		
1980	FEMA integrates USGS 1:2m mapping into seamless database		
1981	*Computers, Environment & Urban Systems*; Arc/Info launched		
1982		8088 chip; IBM PC;	
1983		Mandelbrot's *The Fractal Geometry of Nature*	
1984	1st Spatial Data Handling Symposium	80286 chip, RISC chip; WGS-84	
1985		GPS operational	
1986	Burrough's *Principles of Geographical Information Systems for Land Resources Assessment*; MapInfo founded	SPOT 1 launched	
1987	*International Journal of Geographical Information Systems*; GIS/LIS conference series; '*Chorley*' *Report*	80386 chip	
1988	NCGIA; *GIS World*, UK RRL initiative	Berlin Wall comes down	Internet, mobile phones
1989	UK Association for Geographic Information		
1990		Berners-Lees launches W WW	
1991	USGS digital topo series complete. 1st International Symposium on Integrating GIS and Environmental Modelling	Dissolution of Soviet Union	
1992		Rio Earth Summit – Agenda 21	

Table 2.1 (Continued)

Year	*GIS*	*Context*
1993	GIS Research UK conference series	Pentium chip
1994	Open GIS Consortium	HTML
1995	OS finished digitising 230,000 maps	Java
1996	1st International Conference on GeoComputation; *Transactions in GIS*	
1997	*IJGIS* changes '*Systems*' to '*Science*'; last AutoCarto; *Geographical and Environmental Modelling*	
1998	*Journal of Geographical Systems*; last GIS/LIS	GPS selective availability off
2000		'Millennium Bug'

The study area covered 115 km^2; size of cell was standardised at one acre (~0.4 ha). With the objective to identify land use potential, four base data layers were digitised: 1969 and 1976 land use from *aerial photographic interpretation* (API), soils and underlying geology from published map sheets. Through a process of either reclassification of single layers or a logical combination (overlay) of two or more layers with reclassification, a total of 19 new factor maps were created (Table 2.2) to answer a range of spatial questions (*where?*) concerning land suitability for development.

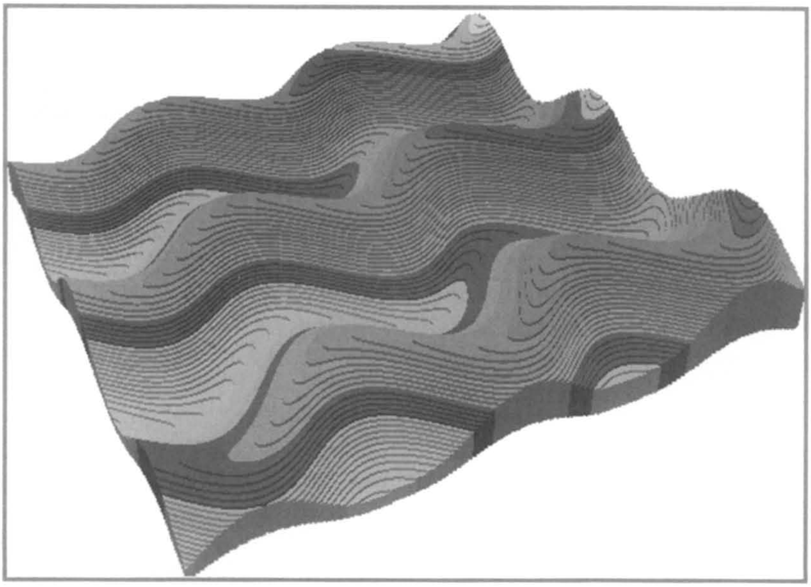


Figure 2.2 Laptop GIS of today: 3-D topographic perspective of a landscape.

Table 2.2 Multiple layer production from three source data sets (based on Schlesinger *et al.*, 1979)

Base maps → *↓ factor maps*	*1969 land use*	*1976 land use*	*Soils*	*Geology*
Slope			✓	
Flood hazards			✓	
Potential for building sites			✓	
Potential for woodland wildlife habitat			✓	
Potential for rangeland habitat			✓	
Potential for openland habitat			✓	
Limitations to road and street construction			✓	
Limitations for septic tank absorption fields			✓	
Soils of statewide importance for farmland			✓	
Sliding hazards				✓
Groundwater recharge areas				✓
Land use change	✓	✓		
Limitations to sewage lagoons			✓	✓
Important farmland		✓	✓	
Important farmland lost to urban development	✓	✓	✓	
Limitations to urban development			✓	✓
Land suitable for urban development but not important agricultural land			✓	✓
Limitations for septic tanks	✓	✓	✓	
Limitations for new urban development	✓	✓	✓	✓

Typical of the many pioneering efforts of the time, this study achieved its goals and was well received in the community despite the rudimentary hardware and software tools available.

Some of the changes are obvious. Over the intervening 20+ years, the action of Moore's Law, by which the hardware price to performance ratio is expected to double every 18 months, means that the laptop I'm writing on far outstrips the IBM mainframe of that time in terms of power, performance and storage by several orders of magnitude at a fraction of the cost in real terms. Instead of using a collection of software modules which may need to be modified and recompiled to satisfy the needs of the individual project, we have a choice of off-the-shelf packages such as MapInfo, ArcView, GeoMedia, AutodeskMap, Geo-Concept and Spans that combine a wide range of functionality with mouse- and icon/menu-driven interfaces. For project-specific needs, most of these packages have object-oriented scripting languages that facilitate customisation and the addition of new functionality with many such scripts available over the Internet. Moreover, analysis can now be vastly extended to include external computational models that communicate either through the scripting or use of common data storage

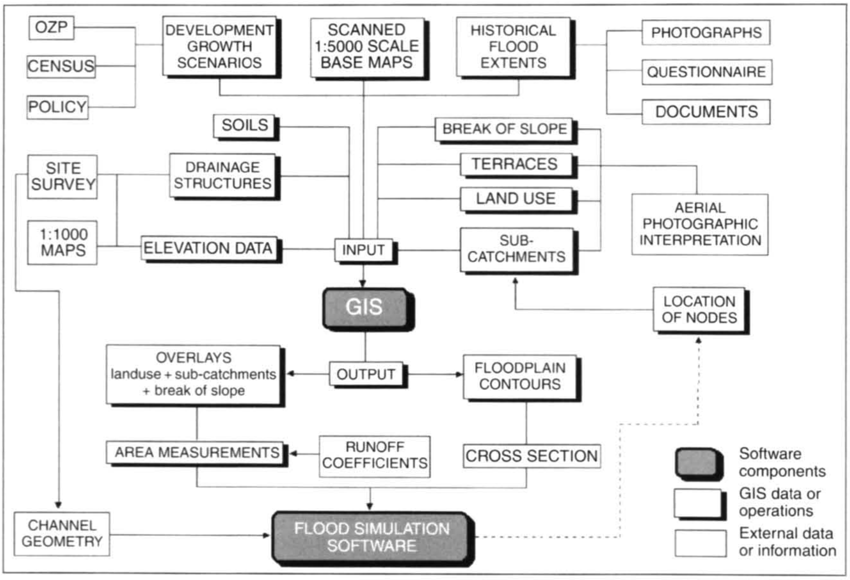


Figure 2.3 An example of data sets and their provenance for a modern study of flood hazard analysis and basin management planning.

formats. Although the availability of digital map data is uneven across the world, particularly when it comes to large scale mapping, off-the-shelf digital data ready for use in GIS are much more common today to the point where certainly for projects in North America and Europe there is hardly the need anymore to manually digitise. As mentioned above, the bottleneck in the production of digital spatial data has been burst not only by technologies such as GPS, RS and digital photogrammetry but through palm-top data loggers, high-speed scanners, digital data transfer standards and above all the computer capacity to cost-effectively store, index and deliver huge data sets. In contrast to Table 2.2 in which only four data sources were used, Figure 2.3 summarises the data sets used in a more recent project where GIS and hydraulic modelling were combined for flood hazard analysis and basin management planning. Nevertheless, despite this technological facilitation that has made GIS more widespread, sophisticated and easier to use, many of the underlying principles have remained largely the same.

Representing spatial phenomena in GIS

The dominant paradigm in the way GIS data are structured comes from the idea that studies of landscape (both human and physical) and the solution to problems concerning the appropriate use of land can be achieved by describing the landscape as a series of relevant factor maps or *layers* which can then be overlaid to find those areas having particular combinations of factors which would identify them as most suited to a particular activity. The methodology in its modern GIS context derives from the seminal work of McHarg (1969) as well as the conventional cartographic tradition of representing spatial phenomena. Although the use of manual overlay of factor maps considerably predates McHarg (Steinitz *et al.*, 1976), he provided a compelling case for the methodology as a means of organising, analysing and visualising multiple landscape factors within a problem-solving framework. Consider the landscape shown in Figure 2.4.

This landscape can be viewed both holistically as a piece of scenery and as a series of constituent elements such as its topography, geology, hydrology, slope processes, flora, fauna, climate and man-made (anthropomorphic)

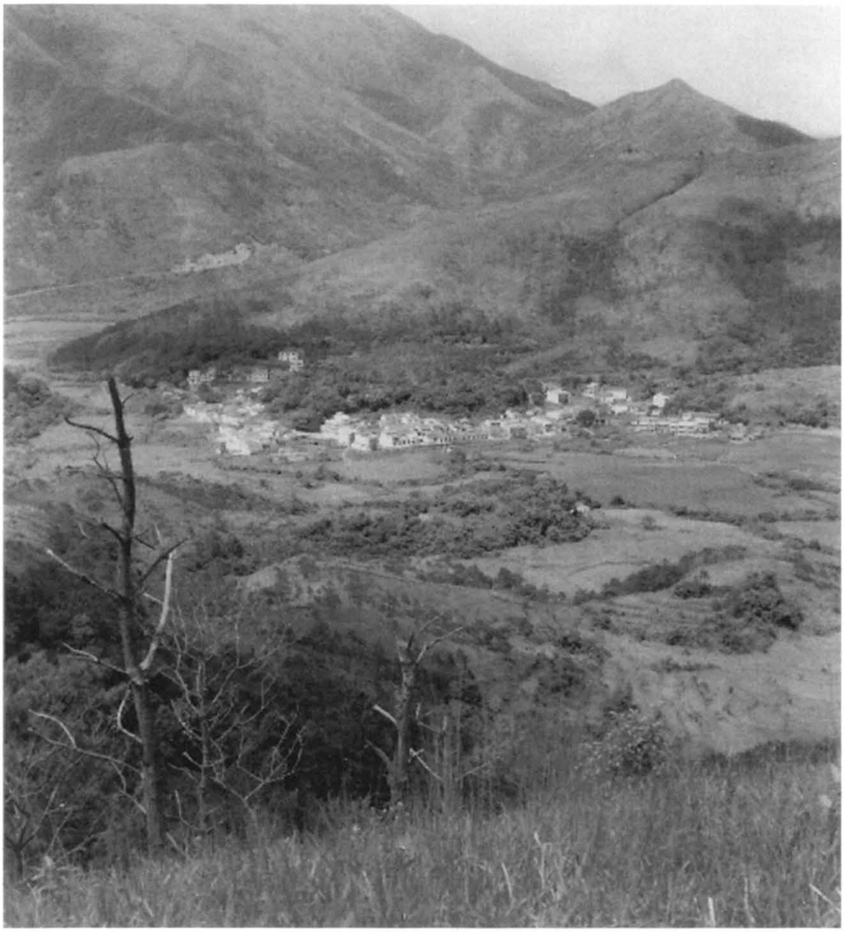


Figure 2.4 A view of a sample landscape (photograph by the author).

features to name but a number that could be separated out. At any place within this landscape there are several or all constituents to be considered: stand on any point and it has its topography, geology, hydrology, climate and so on. Any comprehensive map of all these constituents would quickly become cluttered and complex – almost impossible to work with. So consider then the mapped constituents of a very similar landscape in Figure 2.5(a)–(i).

Although this particular landscape has been artificially created to illustrate a number of issues throughout this book, it shows well a number of aspects of the layer or coverage paradigm and the graphic primitives used in any one layer. First, in order for a selection of layers to be used together, superimposed and viewed as a composite, *they must all conform to the same co-ordinate system and map projection*. This is critically important, otherwise the layers will be distorted and wrongly positioned in relation to one another. Individual layers, however, need not necessarily cover exactly the same area of the landscape in their extent as may happen, for example, if they have been derived from different surveys or source documents. Each layer can nevertheless be clipped to a specific study area as has happened in Figure 2.5. Second, some of the layers are given to represent *discrete objects* in the landscape (e.g. landslides, streams, land cover parcels) whilst others represent a *continuous field* (e.g. topography, gradient, rainfall) which varies in its value across the landscape. What aspects of the landscape should be treated as continuous or discrete and how they should be presented cartographically is an old, but significant problem which can still be debated today (Robinson and Sale, 1969; Peuquet, 1984; Goodchild, 1992a; Burrough, 1992; Burrough and Frank 1996; Spiekermann and Wegener, 2000). To a considerable extent it is a matter of data resolution, scale of representation, convention and convenience. For example, landslides can be quickly mapped at a regional level as individual points representing each scar in the terrain (as in Figures 2.5(h) and 2.6(a)). Another approach would be to represent each landslide as a line starting at the scarp and tracing the down slope extent of the debris to the toe (Figure 2.6(b)). Clearly any laterally extensive landslide in Figure 2.5(h) would represent a methodological problem for which a single point or a line would be an oversimplification. So yet another approach would be to represent either the whole landslide or its morphological elements according to a consistent scheme (e.g. source, transport, deposition) as polygons (Figure 2.6(c)). This latter approach, whilst providing more information, is more time consuming and expensive to produce. Finally, these landslides could be represented as a field of varying numbers (or as densities – Figure 5.11(a)) of landslides within a tessellation of cells (Figure 2.6(d)).

To pursue this issue just a bit further, topography is a continuous field but is conventionally represented by contours which in geometric terms are nested polygons. Gradient on the other hand is also a continuous field but would generally be confusing to interpret if drawn as contours and is thus usually represented by a tessellation of cells each having its own gradient value. Soils are conventionally classified into types and each type is

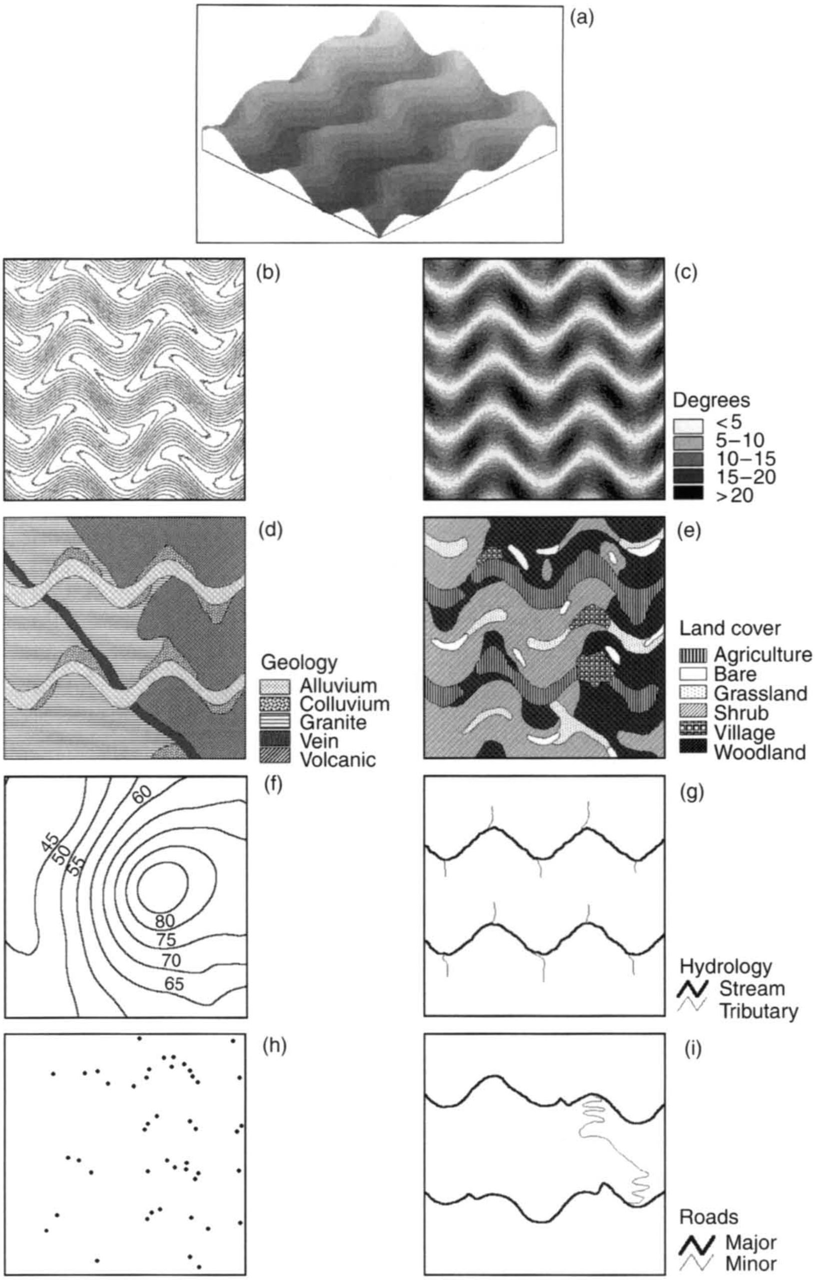


Figure 2.5 Mapped constituents of an example landscape in eight layers (coverages): (a) oblique view of topography, (b) contours, (c) slope gradient, (d) geology, (e) land cover, (f) rainfall isohyets from a storm event, (g) drainage network, (h) landslide scars, (i) transport.

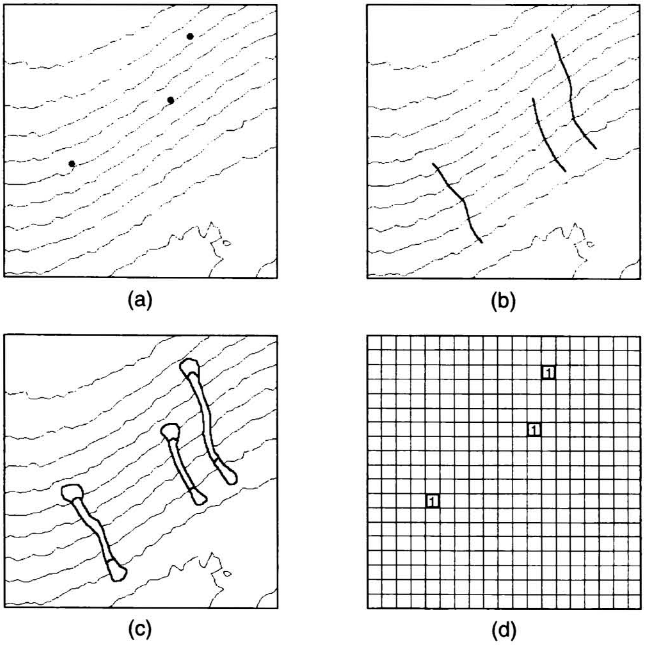


Figure 2.6 Four possible methods of representing landslides in GIS: (a) as points, (b) as lines, (c) as polygons, (d) as a tessellation (raster).

represented by discrete polygons wherever they occur. This is despite the fact that many boundaries between soil types are really gradations of one dominant characteristic (say, clay content or structure of horizons) to another. Land uses are similarly defined as homogenous discrete polygons on the basis of dominant land use type despite perhaps considerable heterogeneity within any polygon. We will return to these issues later in Chapter 8 when we consider the implications of this on spatial data quality.

Fundamentally then, any point within a landscape can be viewed as an array containing the co-ordinates of location $\{x, y\}$ and values/classes for n defined attributes a. The first two of these attributes may be specifically defined as elevation z and time t. The whole landscape L can thus be described by a large number of such points p in a matrix:

$$\mathrm{L} = \begin{bmatrix} x_1\ y_1\ z_1\ t_1\ a_{13}\ a_{14}\ a_{15} \ldots a_{1n} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_p\ y_p\ z_p\ t_p\ a_{p3}\ a_{p4}\ a_{p5} \ldots a_{pn} \end{bmatrix} \tag{2.1}$$

In practical terms, time t is often fixed and the matrix is taken to be a single snapshot of the landscape. Also, since the number of points used to describe the landscape is usually only a tiny proportion of all possible points, **L** is considered to be a sample of one. Elevation z is taken to be an attribute of a location and is therefore not really a third dimension in the traditional sense of an $\{x, y, z\}$ tuple. GIS are thus commonly referred to as $2\frac{1}{2}$-D rather than 3-D. The points themselves can be organised into a series of *points*, *lines* or *polygons*, that is, discrete objects of 0, 1 and 2 dimensions respectively to form *vector layer(s)*. Usually, objects that are points, lines and polygons are not mixed within a layer but are kept separate. This describes the planar geometry and disposition of the objects within the landscape. The attributes of each object are stored in a database and are linked to the graphics via a unique identifier (Figure 2.7). The other approach to **L** is for the landscape to be tessellated, that is, split into a space-filling pattern of *cells* and for each cell to take an attribute value according to the distribution of points to form a *raster layer*. There may thus be n layers, one for each attribute. Although the objective in both vector and raster approaches is to achieve spatially seamless layers that cover an entire area of interest, it may be that for large areas the data volume in each layer becomes too large and cumbersome to handle conveniently (e.g. response times in display and analysis). When this occurs, layers are usually split into a series of non-overlapping *tiles* which when used give the impression of seamless layers.

Thus far I have described the mainstream approach to representing spatial phenomena in GIS. Since the early 1990s an alternative has emerged – the *object-oriented* (OO) view of spatial features – not to be confused with the above *object-based* approach of vector representation. Spatial objects as discernible features of a landscape are still the focus, but rather than splitting their various aspects or attributes into layers (the geology, soils, vegetation, hydrology, etc. of a parcel of land), an object is taken as a whole with its properties, graphical representation and behaviour in relation to other spatial objects embedded within the definition of the object itself (Worboys

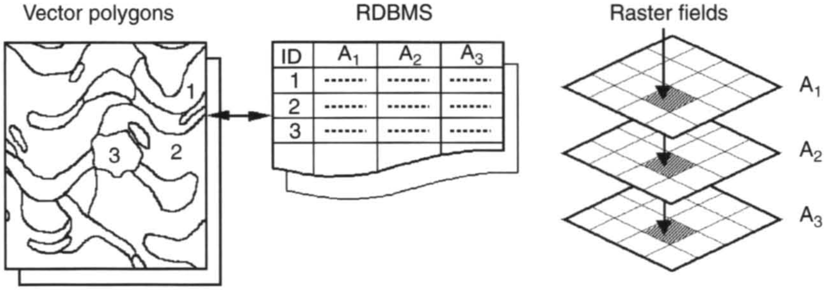


Figure 2.7 Basic organisation of geometry and attributes in layered GIS: vector and raster.

et al., 1990; Milne *et al.*, 1993; Brimicombe and Yeung, 1995; Wachowicz, 1999). Thus the modelling of 'what' is separated from 'where' and in fact both 'where' and whether to use raster or vector (or both, or neither!) as a means of graphical representation can be viewed as attributes of 'what'. This then allows even abstract spatial concepts such as socio-cultural constructs to be included in GIS alongside more traditional physical features of a landscape (see Brimicombe and Yeung, 1995). Although from a personal viewpoint the OO view provides a superior, more robust approach to spatial representation in GIS, the market share for truly OO GIS (e.g. Smallworld, Laser-Scan) and database management systems (e.g. ObjectStore) has remained comparatively small.

Putting the real world on a diskette

Having introduced the representation of geographical phenomena in GIS from a practical 'what you see on the screen' perspective, it is now necessary to do so from a computer science 'what technically underpins it' perspective. Essentially we want to achieve a representation of a landscape that can be stored digitally on a machine in such a way that the representation is convenient to handle and analyse using that machine. Ultimately, the intended purpose of the representation, the nature of software tools available and the types of analyses we wish to undertake will strongly influence the form of representation that is deemed appropriate.

A machine representation of a landscape as a digital stream of binary zeros and ones on a hard disk or diskette necessitates a considerable amount of abstraction to say the least! The process of abstraction and translation into zeros and ones needs to be a formally controlled process if the results are going to be of any use at all. This process is known as *data modelling* and is discussed at some length by Peuquet (1984) and Molenaar (1998). Two diagrammatic views of the data modelling process are given in Figure 2.8.

In general, four levels can be recognised within data modelling:

- The first of these is *reality* itself, that is the phenomena we wish to model as they actually exist or are perceived to exist in all their complexity.
- The second level is the *conceptual model* which is the first stage abstraction and incorporates only those parts of reality considered to be relevant to the particular application. A cartographic map is a good metaphor for the conceptual model as a map only contains those features which the cartographer has chosen to represent and all other aspects of reality are omitted. This provides an immediate simplification though a sense of the reality can still be readily interpreted or reconstituted from it. Just as a cartographer must decide in creating a map which symbologies should be used for the various features, so it is at the conceptual modelling stage that decisions are generally made as to whether to use raster or vector and what the theme for each layer

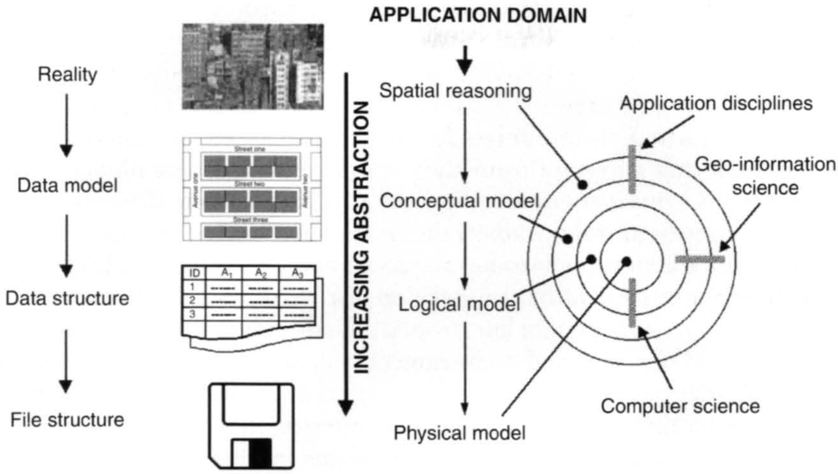


Figure 2.8 Stages in the data modelling process (partly based on Molenaar, 1998).

is going to be. The conceptual model is often referred to as the *data model* which in a data modelling process can give rise to confusion.

- The third level is the *logical model*, often called the *data structure*. This is a further abstraction of the conceptual model into lists, arrays and matrices that represent how the features of the conceptual model are going to be entered and viewed in the database, handled within the code of the software and prepared for storage. The logical model can generally be interpreted as reality only with the assistance of software, such as by creating a display.
- The fourth level is the *physical model* or *file structure*. This is the final abstraction and represents the way in which the data are physically stored on the hardware or media as bits and bytes.

The third and fourth levels, the logical and physical models, are usually taken care of in practical terms by the GIS software and hardware being used. Long gone are the days of programming and compiling your own GIS software from scratch when the design of the logical and physical models were important. *De facto* standards such as Microsoft Windows are even leading to a high degree of interoperability allowing Excel spreadsheets to be accessed in MapInfo as just one example. The challenge then is in creating the conceptual model that will not only adequately reflect the phenomena to be modelled but also lead to efficient handling and analysis. The choice between vector and tessellation approaches can be important as they have their relative advantages and disadvantages. These, however, are not entirely straightforward as the logical model (as offered by the software) used to underpin any conceptual model has important bearing on the ease of handling and 'added intelligence' of the data for particular types of analyses. This issue then needs some further discussion.

Vector

As already discussed, the primitives or basic *entities* of vector representation are *point*, *line* and *polygon* (Figure 2.9) where a point is a zero-dimensional object, a line is a linear connection between two points in one-dimension and a polygon is one or more lines where the end point of the line or chain of lines coincides with the start point to form a closed two-dimensional (2-D) object. A line need not be straight but can take on any weird shape as long as there are no loops. Any non-straight line, from a digital perspective, is in fact made up of a series of *segments* and each segment will, of course, begin and end at a point. In order to avoid confusion then, points at the beginning and end of a line or connecting two or more lines are referred to as *nodes*. Lines connected at their nodes into a series can form a *network*. Polygons (also known as area features) when adjacent to one another will share one or more lines. Since all lines have orientation from their start node to their end node, they have a direction and on the basis of this have a left and right side. Thus within a logical model that records *topology*, that is explicitly recording connectivity (as in a network) or adjacency (as for polygons), the polygon to the left and right of a line can be explicitly recorded in the database (Figure 2.10). In this way, a fully topological database has additional intelligence so that locating neighbouring lines and polygons becomes straightforward. Some desktop GIS do not go so far, leaving each feature to be recorded separately without reference to possible neighbours. These are commonly referred to as *shapefiles*. Finally, by providing a unique identifier to each point, line and polygon (usually done automatically by the software), a join can be made to a database containing relevant attributes for each object (Figure 2.7). Thus by selecting specific map features in a vector-based GIS their attributes can be displayed from the database. Conversely, by selecting specific attributes from the database, their spatial representation on the map can be highlighted.

Tessellations

A tessellation is a space-filling mesh (Figure 2.11) either with explicit boundaries as a mesh of polygons or with an implicit mesh as defined, say, by a matrix of values in the logical model. A tessellation can be either *regular*, in which case mesh elements are all the same size and shape, or *irregular*. Elements of a regular mesh could be isosceles triangles, squares (raster), rectangles or hexagons. One example of an irregular mesh is a *triangulated irregular network* or TIN (Mark, 1975) in which a point pattern is formed into a triangular mesh often as a precursor to interpolating contours. Another is *Theissen polygons* (Theissen, 1911) which is the dual of TIN and represents the area of influence of each point in a point pattern.

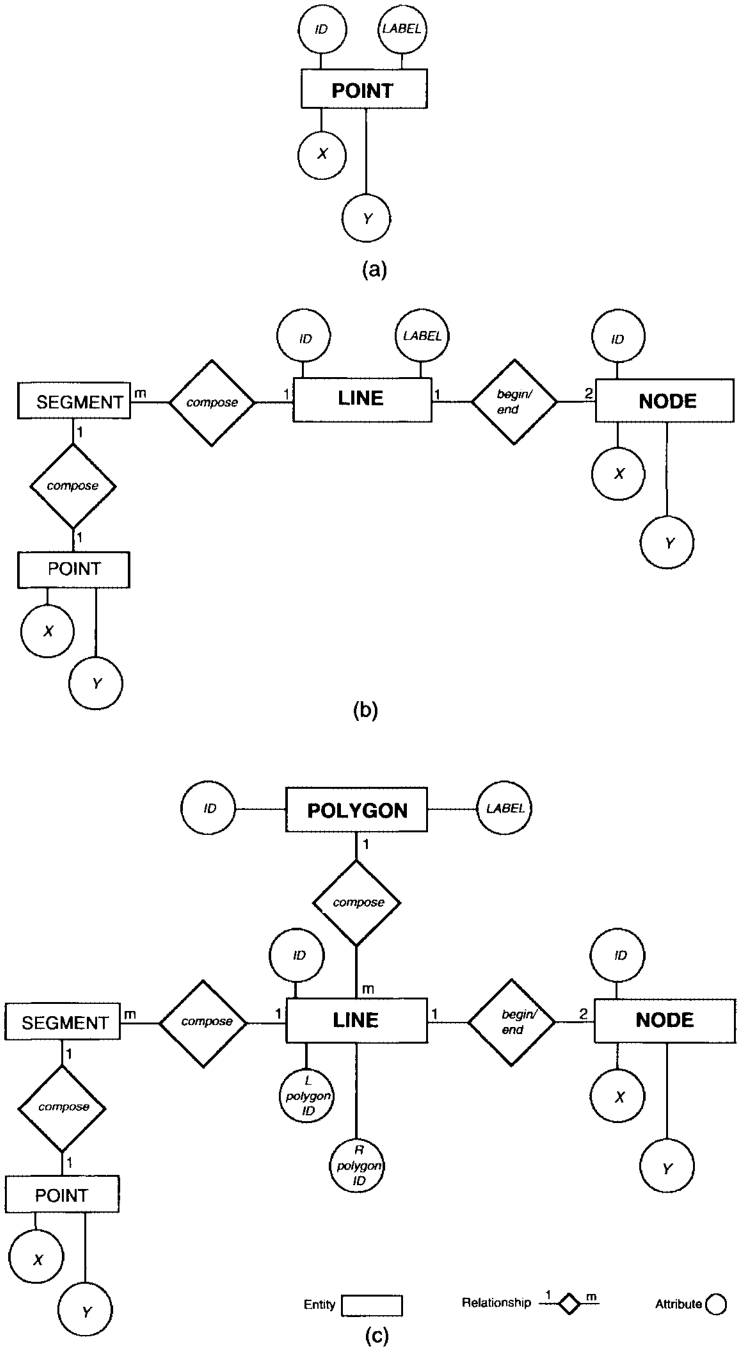
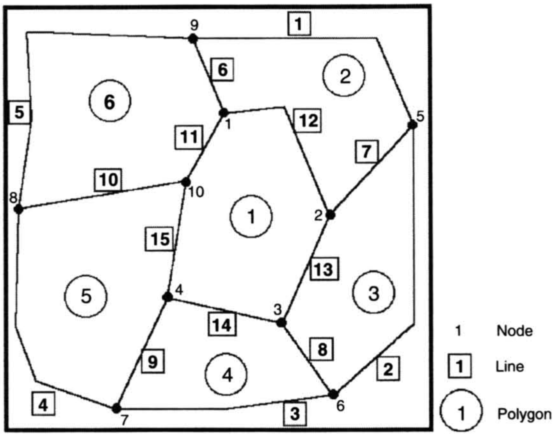


Figure 2.9 Entities of the vector model: (a) point, (b) line, (c) polygon.



Node list

ID	X	Y
1	95	115
2	112	94
3	107	72
4	83	77
5	134	112
6	117	56
7	72	54
8	53	95
9	89	130
10	87	101

Line list

ID	FROM NODE	TO NODE	LEFT POLYGON	RIGHT POLYGON
1	5	9	2	
2	5	6		3
3	6	7		4
4	7	8		5
5	8	9		6
6	1	9	6	2
7	2	5	2	3
8	3	6	3	4
9	4	7	4	5
10	8	10	6	5
11	1	10	1	6
12	1	2	2	1
13	2	3	3	1
14	3	4	4	1
15	4	10	5	1

Segment list

Line ID	X, Y Pairs
1	122, 130
2	134, 72
3	95, 54
4	56, 60 52, 71

Polygon list

ID	Line list
1	11 12 13 14 15
2	1 7 12 6
3	2 8 13 7
4	3 9 14 8
5	4 10 15 9
6	5 6 11 10

Figure 2.10 Building topology into the vector model.

Tessellations can also be *recursive*, that is, the basic mesh shape can be progressively split into a finer mesh in order to represent higher resolution features. An example of this type of tessellation is the *quadtree* (Samet, 1984) which seeks to subdivide in a hierarchy, subject to a pre-defined minimum resolution, in order to achieve homogeneity within cells. One clear advantage of quadtree data structure over the traditional raster approach is that redundancy is reduced and storage is more compact. Topology in tessellations can be either implicit or explicit (Figure 2.12). For regular meshes, neighbours can be easily found by moving one cell to the left, right, up, down or diagonally in which case the topology is implicit. For a TIN, the topology can be made explicit just as it is in the vector model since each triangular element is a polygon. For structures such as quadtree, an explicit