15 Spatial Modeling with GI Systems

Models are used in many different ways, ranging from simulations of how the world works to evaluations of planning scenarios to the creation of indicators. In all these cases the geographic information (GI) system is used to carry out a series of transformations or analyses of geographic space, either at one point in time or at a number of intervals. This chapter begins with the necessary definitions and presents a taxonomy of models, with examples. It addresses the difference between analysis, the subject of Chapters 13 and 14, and this chapter's focus upon modeling. The alternative software environments for modeling are reviewed, along with capabilities for cataloging and sharing models, which are developing rapidly. The chapter ends with a look into the future of modeling and associated GI system developments.

LEARNING OBJECTIVES

After studying this chapter you will:

- Know what modeling means in the context of GI systems.
- 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 GI system software.

15.1 Introduction

This chapter identifies many of the distinct types of models supported by GI systems and gives examples of their applications. After *system* and *object*, *model* is probably one of the most overworked terms in the English language, especially in the language of GI systems, with many distinct meanings in that and even in this book. So first it is important to address the meaning of the term as it is used in this chapter.

> ***Model* is one of the most overworked terms in the English language.**

A clear distinction needs to be made between *data models* discussed in Chapter 7 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, and a set of assumptions about the nature of data. For example, the raster data model forces all knowledge to be expressed as properties of the cells of a regular grid lain on the Earth. Data models are closely related to the concept of ontology, which is best understood as the study of the basic elements of description, or "what we talk about."

A data model is in essence a statement about form or about how the world looks, limiting the options open to the data model's user to those allowed by its template. Models in this chapter are very different and can be expressions of how the world is believed to work, or how a task is broken down into a sequence of operations; in other words they are expressions of process (see Section 1.3 on how these both relate to the science of problem solving). They may include dynamic simulation models of natural processes such as erosion, tectonic uplift, the migration of elephants, or the movement of ocean currents. They may include models of social processes, such as residential segregation or the movements of cars on a congested highway. They may include processes designed by humans to search for optimum alternatives, for example in finding locations for a new retail store. They may include standard workflows that are executed every day in organizations. Finally, they may include simple calculations of indicators or predictors, such as happens when layers of geographic information are combined into measures of groundwater vulnerability or social deprivation.

The common element in all these examples is the manipulation of geographic information in multiple stages, especially if these stages must be run repeatedly. In some cases these stages will perform a simple transformation or analysis of inputs to create an output, and in other cases the stages will loop to simulate the development of the modeled system through time, in a series of iterations. Only when all the loops or iterations are complete will there be a final output. There will be intermediate outputs along the way, and it is often desirable to save some or all of these, in case the model needs to be rerun or if parts of the model need to be changed.

All the models discussed in this chapter are digital or computational models, meaning that the operations occur in a computer and are expressed at the most fundamental level in a language of 0s and 1s. In Chapter 3 representation was seen as a matter of expressing geographic form in 0s and 1s; this chapter looks for ways of expressing geographic process in 0s and 1s. The term *geocomputation* is often used to describe the application of computational models to geographic problems.

All the models discussed in this chapter are also spatial models (and deal, of course, with geographic space). There are two key requirements of such a model:

1. There is variation across the space being manipulated by the model (an essential requirement of all GI system applications).
2. The results of modeling change when the locations of objects change—location matters (this is also a key requirement of spatial analysis as defined in Section 13.1).

Models do not have to be digital, and it is worth spending a few moments considering the other type, known as analog. The term was defined briefly in Section 3.7 as describing a representation, such as a paper map, that is a scaled physical replica of reality. Analog models can be very efficient, and they are widely used to test engineering construction projects and proposed airplanes. They have three major disadvantages relative to digital models, however: they can be expensive to construct and operate, their accuracy is limited by the effects of scaling the real world, and unlike digital models, they are virtually impossible to copy, store, or share. Nevertheless they can be extremely efficient at modeling complex systems, such as airflow over proposed buildings.

An analog model is a scaled physical representation of some aspect of reality.

The level of detail of any analog model is measured by its *representative fraction* (Section 3.7, Box 2.2), the ratio of distance on the model to distance in the real world. Like digital data, computational models do not have well-defined representative fractions; instead, level of detail is measured as *spatial resolution*, defined as the shortest distance over which change is recorded. *Temporal resolution* is also important for models of processes, being defined as the shortest time over which change is recorded and corresponding in the case of many dynamic models to the time interval between iterations.

Spatial and temporal resolutions are critical factors in models. They define what is left out of the model, in the form of variation that occurs over distances or times that are less than the appropriate resolution. They also therefore define one major source of uncertainty in the model's outcomes. Uncertainty in this context can be best defined through a comparison between the model's outcomes and the outcomes of the real processes that the model seeks to emulate. Any model leaves its user uncertain to some degree about what the real world will do; a measure of uncertainty attempts to give that degree of uncertainty an explicit magnitude. Uncertainty has been discussed in the context of geographic data in Chapter 5; its meaning and treatment in the context of spatial modeling are discussed later in this chapter.

Any model leaves its user uncertain to some degree about what the real world will do.

Spatial and temporal resolution also determine the cost of acquiring the data because in general it is more costly to collect a fine-resolution representation than a coarse-resolution one. More observations have to be made, and more effort is consumed in making and compiling them. They also determine the cost of running the model because execution time expands as more data have to be processed and as more iterations have to be made, and fine spatial and temporal resolution may also raise ethical questions for many types of socioeconomic data. One benchmark is of critical importance for many dynamic models: They must run faster than the processes they seek to simulate if the results are to be useful for planning. For example, a model of the atmosphere is only useful if it runs faster than changes in the weather, and if the model can be rerun as changes occur in order to update predictions. One expects this to be true of computational models, but in practice the amount of computing can be so large that the model simulation slows to unacceptable speed. Supercomputers, and indeed the entire field of high-performance computing, must sometimes be used in order for models to run sufficiently fast, on sufficiently detailed data. This and other aspects of advanced computing, often termed *cyberGIS*, are addressed more extensively in Chapter 10.

Spatial and temporal resolutions are major factors in the cost both of acquiring data for modeling, and of actually running the model.

15.1.1 Why Model?

Models are built for a number of reasons. First, a model might be built to support a decision or design process in which the user wishes to find a solution to a spatial problem in support of a decision (Box 15.1 discusses some of the spatial problems that Budhendra Bhaduri

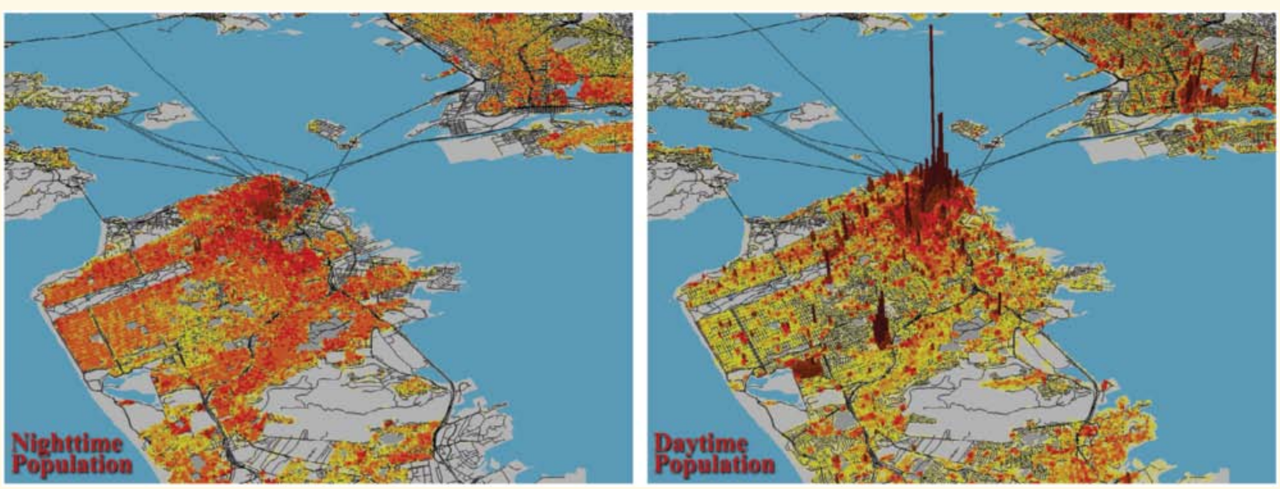
Biographical Box 15.1

Budhendra Bhaduri, Leading Spatial Modeler

Budhendra Bhaduri (Figure 15.1) is a Corporate Research Fellow at Oak Ridge National Laboratory (ORNL) in Tennessee, where he leads the Geographic Information Science and Technology (GIST) group. He is a founding member of the U.S. Department of Energy's Geospatial Sciences Steering Committee and holds professorial appointments with the University of Tennessee, Knoxville. He is passionate about translating research into practice for broad societal impact in such areas as population dynamics, energy resource assessment, and disaster management, using spatially explicit modeling and simulation of complex urban systems. He is a principal member of the LandScan modeling team that develops fine-resolution, dynamic models of population distributions at local to global scales. For example, LandScan USA provides 90 m-resolution nighttime and daytime population-distribution data for the United States (Figure 15.2). LandScan USA utilizes numerous data sources besides the traditional census, including data on commuting patterns and national databases of schools, prisons, and business locations for demographic categorization based on daily activities. Most recently, he has been inspired by volunteered geographic information (VGI), its impact in advancing science and the practice of science, and its potential for engaging and empowering individuals in reshaping the open data economy.

Courtesy: Budhendra Bhaduri

Figure 15.1 Budhendra Bhaduri.



Courtesy: Budhendra Bhaduri

Figure 15.2 Two time slices of the LandScan USA data for San Francisco: night-time (left) and daytime (right), both at 90 m resolution.

and his team at Oak Ridge National Laboratory are working to solve), perhaps a solution that optimizes some objective. This concept was discussed in Section 14.4. Often the decision or design process will involve multiple criteria, an issue discussed in Section 15.4. Second, a model might be built to allow the user to experiment on a replica of the world rather than on the real thing. This is a particularly useful approach when the costs of experimenting with the real thing are prohibitive, or when unacceptable impacts would result, or when results can be obtained much faster with a model, allowing impacts to be anticipated. Medical students now routinely learn anatomy and the basics of surgery by working with digital representations of the human body rather than with expensive and hard-to-get cadavers. Humanity is currently conducting an unprecedented experiment on the global atmosphere by pumping vast amounts of CO_2 into it. How much better it would have been if we could have run the experiment on a digital replica and understood the consequences of CO_2 emissions before the experiment began.

Experiments embody the notion of *what-if scenarios*, or policy alternatives that can be plugged into a model in order to evaluate their outcomes. The ability to examine such options quickly and effectively is one of the major reasons for modeling. Figure 15.3 shows an example of modeling to evaluate possible planning scenarios in Los Angeles.

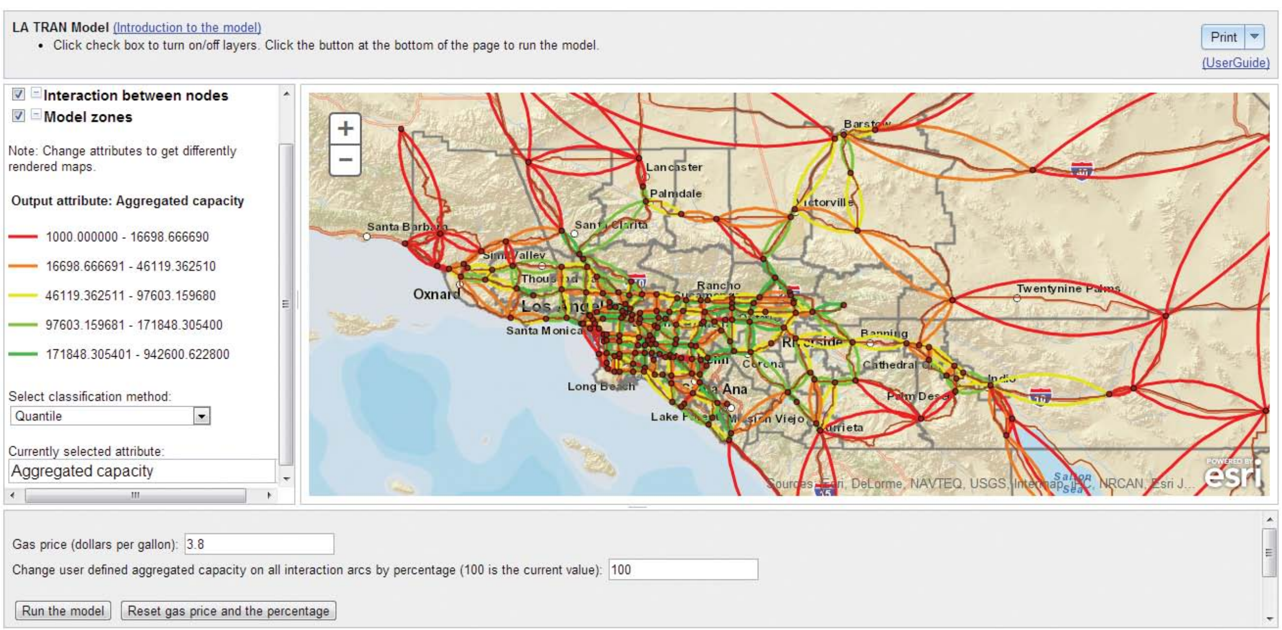
Models allow planners to experiment with what-if scenarios.

Finally, models allow the user to examine dynamic outcomes by viewing the modeled system as it evolves and responds to inputs. As discussed in the next section, such dynamic visualizations are far more compelling and convincing than descriptions of outcomes or statistical summaries when shown to stakeholders and the public. Scenarios evaluated with dynamic models are thus a very effective way of motivating and supporting debates over policies and decisions and of communicating scientific findings to the general public.

15.1.2 To Analyze or to Model?

The problem of emergency evacuation provides a good example of GI system applications in the areas of transportation and logistics. Organizing the evacuation of neighborhoods illustrates a fundamental difference of approach that explains the relationship between Chapters 13 and 14 on the one hand, and Chapter 15 on the other. Using the methods of Chapters 13 and 14 we might analyze the pattern of streets and population, with the aim of mapping difficulty of evacuation, measured as the number of people evacuated per lane of road. This analysis is inherently static, capturing what planners might need to know in a single map display. On the other hand, we might simulate the process of evacuation using a program designed to replicate what would actually happen to individual vehicles and drivers in an emergency, using

Figure 15.3 The graphical user interface to the Virtual Co-Laboratory for Policy Analysis in Greater Los Angeles, a project funded by the University of California's Office of the President. In this illustration, traffic between nodes in the Los Angeles basin is being simulated under conditions controlled by the user, such as the price of gasoline in dollars per U.S. gallon.



rules of human behavior that have been extracted from analysis of real-world traffic. In this way, a GI system would replicate a dynamic *process*. These simulations would allow the researcher to examine what-if scenarios by varying a range of conditions, including zoning controls and new highways. Simulations like these can galvanize a community into action far more effectively than static analysis.

Models can be used for dynamic simulation, providing decision makers with dramatic visualizations of alternative futures.

In summary, analysis as described in Chapters 13 and 14 is characterized by:

- A static approach, at one point in time or an average over time
- The search for patterns or anomalies, leading to new ideas and hypotheses, and perhaps predictions
- Manipulation of data to reveal what would otherwise be invisible

By contrast, modeling, as we define it in this chapter, is characterized by:

- Multiple stages, perhaps representing different points in time
- Implementing ideas and hypotheses about the behavior of the real world
- Experimenting with policy options and scenarios

15.2 Types of Models

15.2.1 Static Models and Indicators

A static model represents a single point in time, or a system that does not change through time, and typically combines multiple inputs into a single output. There are no time steps and no loops in a static model, but the results are often of great value as predictors or indicators. For example, the Universal Soil Loss Equation (USLE) falls into this category. It predicts soil loss at a point, based on five input variables, by evaluating the equation:

$$A = R \times K \times LS \times C \times P$$

where A denotes the predicted erosion rate, R is the Rainfall and Runoff Factor, K is the Soil Erodibility Factor, LS is the Slope Length Gradient Factor, C is the Crop/Vegetation and Management Factor, and P is the Support Practice Factor. Full definitions of each of these variables and their measurement or estimation can be found in descriptions of the USLE (see, for example, www.danewaters.com/business/stormwater.aspx).

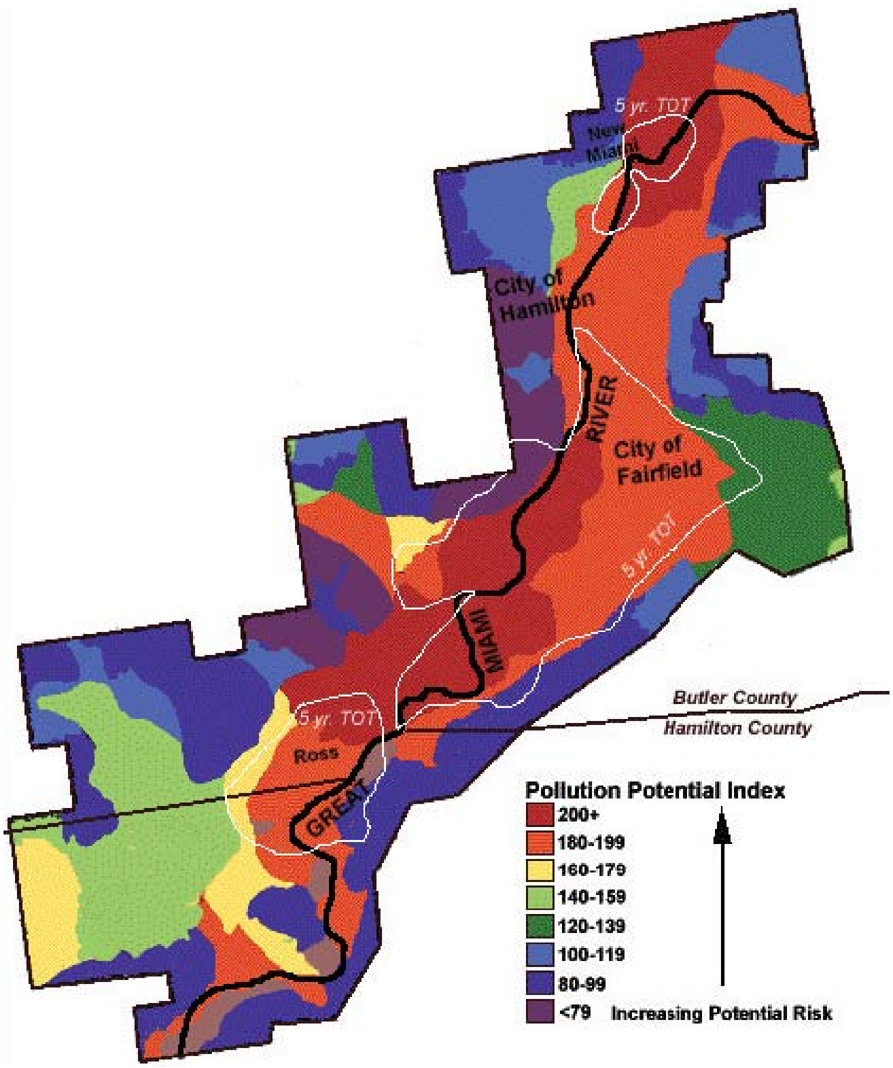
A static model represents a system at a single point in time.

The USLE passes the first test of a spatial model in that many if not all of its inputs will vary spatially when applied to a given area. But it does not pass the second test because moving the points at which A is evaluated will not affect the results. Why, then, use a GI system to evaluate the USLE? There are four good reasons: (1) because some of the inputs, particularly LS, require a GI system for their calculation from readily available data, such as digital elevation models (see Section 14.3.1 for a discussion of the calculation of slope); (2) because the inputs and outputs are best expressed, visualized, and used in map form rather than as tables of point observations; (3) because the inputs and outputs of the USLE are often integrated with other types of data, for further analysis that may require a GI system; and (4) because the data management capabilities of a GI system, and its ability to interface with other systems, may be the best immediately available. Nevertheless, it is possible to evaluate the USLE in a simple spreadsheet application such as Excel; the Web site cited in the previous paragraph includes a downloadable Excel macro for this purpose.

Models that combine a variety of inputs to produce an output are widely used, particularly in environmental modeling. The DRASTIC model calculates an index of groundwater vulnerability from input layers, by applying appropriate weights to the inputs (Figure 15.4). Box 15.2 describes another application, the calculation of a groundwater protection model from several inputs in a karst environment (an area underlain by potentially soluble limestone, and therefore having substantial and rapid groundwater flow through cave passages). It uses Esri's ModelBuilder software, which is described later in Section 15.3.1.

15.2.2 Individual and Aggregate Models

The simulation models used by transportation planners to examine traffic patterns work at the individual level by attempting to forecast the behavior of each driver and vehicle in the study area. By contrast, it would clearly be impossible to model the behavior of every molecule in the Mammoth Cave watershed (Box 15.2). Instead, any modeling of groundwater movement must be done at an aggregate level by predicting the movement of water as a continuous fluid. In general, models of physical systems are



Source: Hamilton to New Baltimore Ground Water Consortium

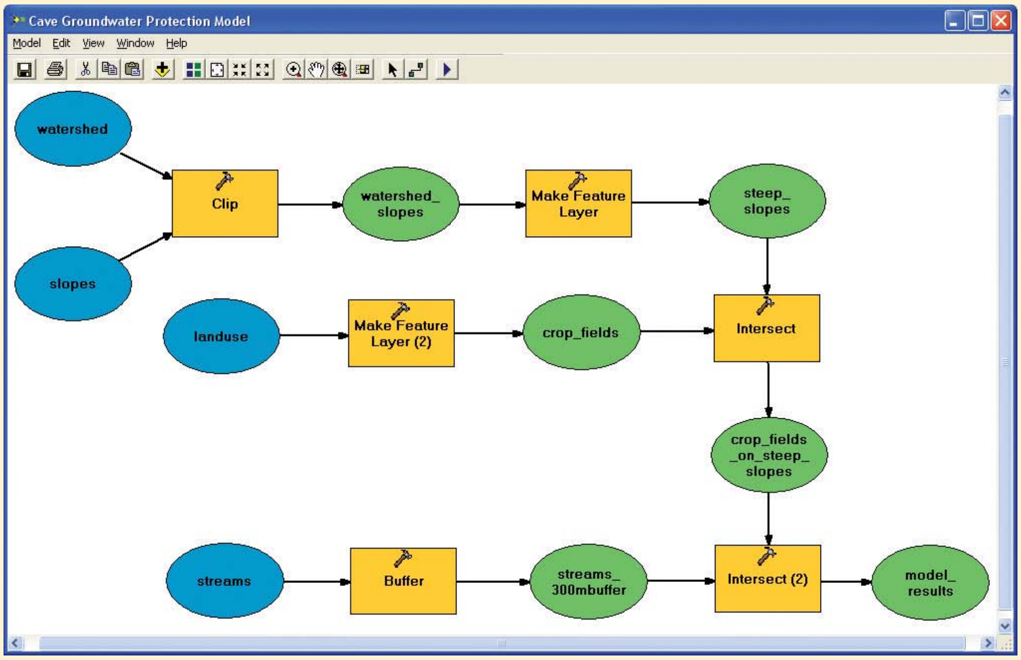
Figure 15.4 The results of using the DRASTIC groundwater vulnerability model in an area of Ohio. The model combines data layers representing factors important in determining groundwater vulnerability and displays the results as a map of vulnerability ratings.

Application Box 15.2

Building a Groundwater Protection Model in a Karst Environment

Rhonda Pfaff (Esri staff) and Alan Glennon (recently graduated PhD from the University of California, Santa Barbara) describe a simple but elegant application of modeling to the determination of groundwater vulnerability in Kentucky's Mammoth Cave watershed. Mammoth Cave is protected as a national park, containing extensive and unique environments, but is subject to potentially damaging runoff from areas in the watershed outside park boundaries and therefore not subject to the same levels of environmental protection. Figure 15.5 shows a graphic rendering of the model, in Esri's ModelBuilder. Each operation is shown as a rectangle and each dataset as an ellipse. Reading from top left, the model first clips the slope layer to the extent of the watershed, then selects slopes greater than or equal to 5 degrees. A land-use layer is analyzed to select fields used for growing crops, and these are then combined with the steep-slopes layer to identify crop fields on steep slopes. A dataset of streams is buffered to 300 m, and finally this is combined to form a layer identifying all areas that are crop fields, on steep slopes, within 300 m of a stream. Such areas are particularly likely to experience soil erosion and to generate runoff contaminated with agricultural chemicals, which will then impact the downstream cave environment with its endangered populations of sightless fish. Figure 15.6 shows the resulting map.

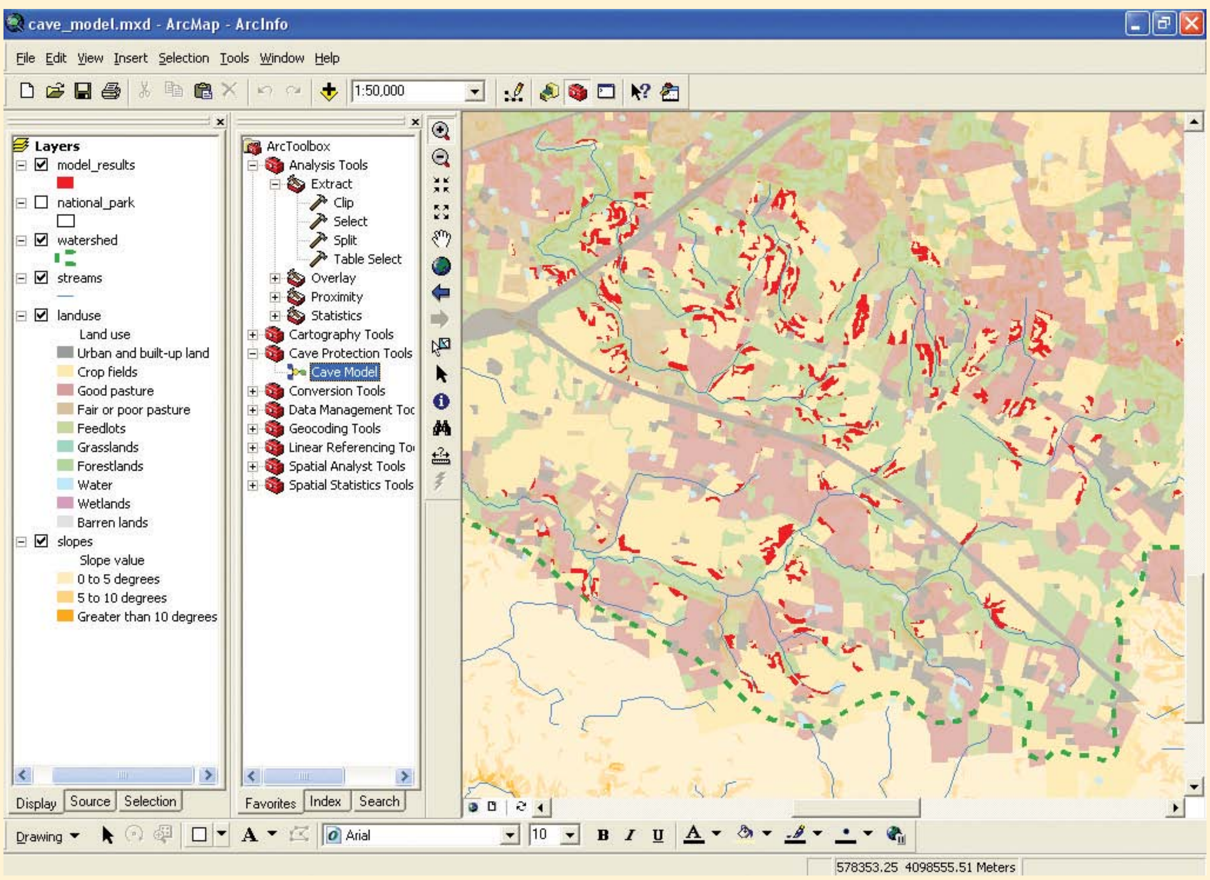
A detailed description of this application is available at www.esri.com/news/arcuser/0704/files/modelbuilder.pdf, and the datasets are available at www.esri.com/news/arcuser/0704/summer2004.html.



Source: Esri

Figure 15.5 Graphic representation of the groundwater protection model developed by Rhonda Pfaff and Alan Glennon for analysis of groundwater vulnerability in the Mammoth Cave watershed, Kentucky.

Figure 15.6 Results of the groundwater protection model. Highlighted areas are farmed for crops, on relatively steep slopes and within 300 m of streams. Such areas are particularly likely to generate runoff contaminated by agricultural chemicals and soil erosion and to impact adversely the cave environment into which the area drains.



Source: Esri

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, as it is in the representation of continuous fields (see Figure 3.7). Some models adopt a raster approach and are commonly called *cellular* models (see Section 15.2.3). Other models break the world into irregular pieces or polygons, as in the case of the groundwater protection model described in Box 15.2.

Aggregate models are used when it is impossible to model the behavior of every individual element in a system.

Models of individuals are often termed *agent-based* models (ABM) or *autonomous agent* models, implying the existence of discrete agents with defined decision-making behaviors (Box 15.3). Each agent might represent an individual, a vehicle and driver, a

Technical Box 15.3

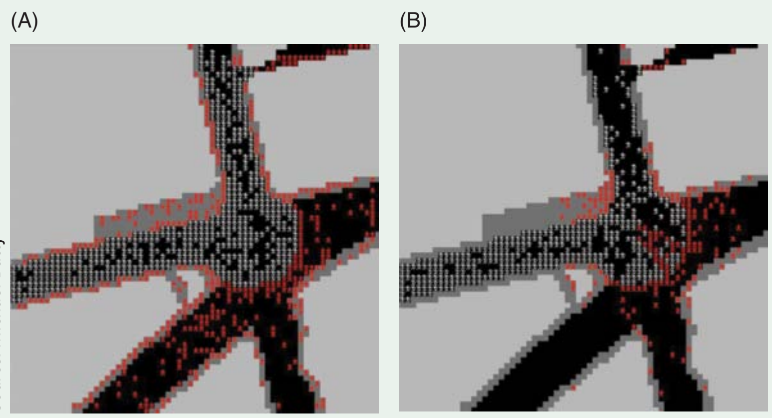
Agent-Based Models of Movement in Crowded Spaces

Working at the University College London's Centre for Advanced Spatial Analysis, geographer and planner Michael Batty uses GI systems to simulate the disasters and emergencies that can occur when large-scale events generate congestion and panic in crowds. The events involve large concentrations of people in small spaces that can arise because of accidents, terrorist attacks, or simply the buildup of congestion through the convergence of large numbers of people into spaces with too little capacity. He has investigated scenarios for a number of major events, such as the movement of very large numbers of people to Mecca to celebrate the Hajj (a major holy event in the Muslim calendar; Figure 15.7) and the Notting Hill Carnival (Europe's largest street festival, held annually in West Central London; Figure 15.8). His work uses agent-based models that take GI science to a finer level of granularity and incorporate temporal processes as well as spatial structure. Fundamental to agent-based modeling of such situations is the need to understand how individuals in crowds interact with each other and with the geometry of the local environment.

© Rabi Karim Photography/Getty Images, Inc.

Figure 15.7 Massive crowds congregate in Mecca during the annual Hajj. On January 12, 2006, 346 pilgrims were trampled when panic stampeded the crowd. This was unfortunately not a unique occurrence: 244 pilgrims were killed in 2004, 50 in 2002, 35 in 2001, and 107 in 1998, and 1425 were killed in a pedestrian tunnel in 1990.

Figure 15.8 shows one of Batty's simulations of the interactions between two crowds. Here a parade is moving around a street intersection—the central portion of the movement (walkers in white) is the parade and the walkers around this in gray/red are the watchers. This model can be used to simulate the buildup of pressure through random motion, which then generates a breakthrough of the watchers into the parade, an event that often leads to disasters of the kind experienced in festivals, rock concerts, and football (soccer) matches as well as ritual situations like the Hajj.



Source: Michael Batty

Figure 15.8 Simulation of the movement of individuals during a parade. Parade walkers are in white, watchers in red. The watchers (A) build up pressure on restraining barriers and crowd control personnel, and (B) break through into the parade.

corporation, or a government agency and would have a defined location at any point in time. With the massive computing power now available on the desktop, together with techniques of object-oriented programming, it is comparatively easy to build and execute ABMs, even when the number of agents is itself massive. Such models have been used to analyze and simulate many types of animal behavior, as well as the behavior of pedestrians in streets, shoppers in stores, and drivers on congested roads. They have also been used to model the behavior of decision makers, not through their movements in space but through the decisions they make regarding spatial features. ABMs have been constructed to model the decisions made over land use in rural areas by formulating the rules that govern individual decisions by landowners in response to varying market conditions, policies, and regulations. Of key interest in such models is the impact of these factors on the fragmentation of the landscape, with its implications for the destruction of wildlife habitat. Figure 15.9 shows an example of the modeling of land-use transition in the Amazon Basin.

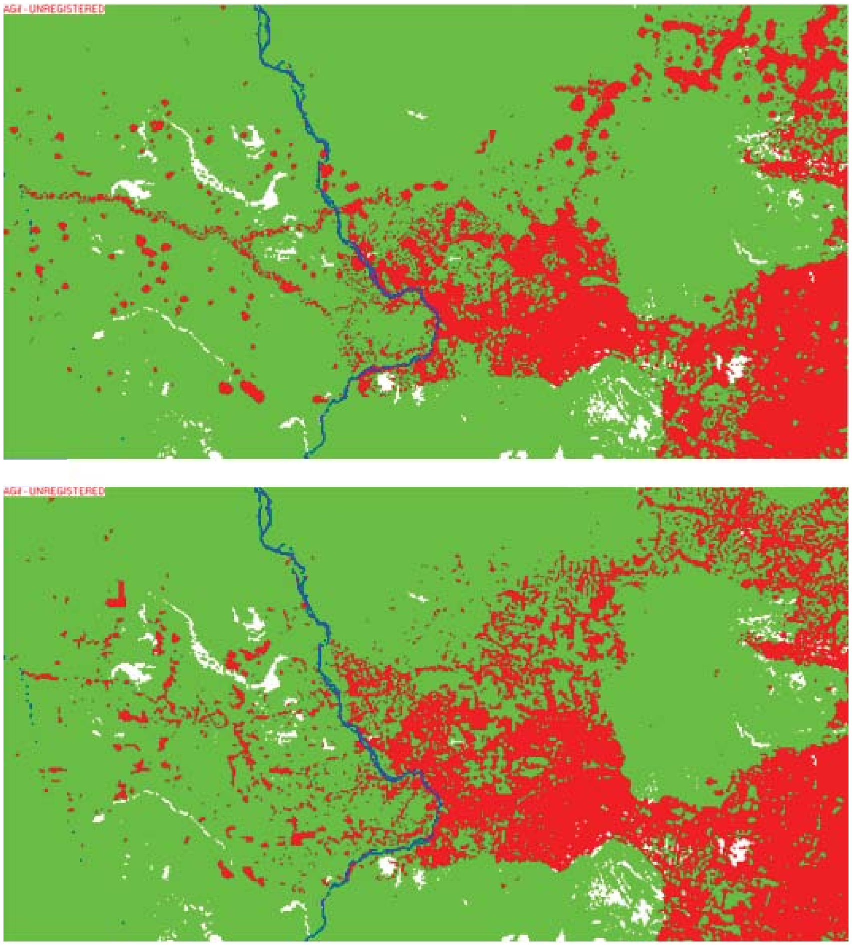
Models such as these find rich applications in the study of disease transmission and the evaluation of alternative policies and interventions aimed at reducing disease prevalence. Indy Hurt, for example, has constructed a model of tuberculosis transmission within a town in Kenya, based on the social behavior of individuals, their journeys to work and shop, and the stages of development and transmission of the disease. Figure 15.10 is an example of her work, visualizing the distribution of susceptibles, people with active disease, and people with latent forms of the disease. Although the accuracy of the predictions (Section 15.5) cannot be guaranteed, models such as these are nevertheless very useful in helping planners, decision makers, and citizens to examine the future impacts of alternative decisions and policy frameworks.

15.2.3 Cellular Models

Cellular models represent the surface of the Earth as a raster (note, however, the difficulties of doing this when the study area is a substantial fraction of the Earth's surface, as discussed in Chapter 4). Each cell in the fixed 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. The study of cellular models was first popularized by the work of John Conway, professor of mathematics at Princeton, who studied the properties of a model he

(A)

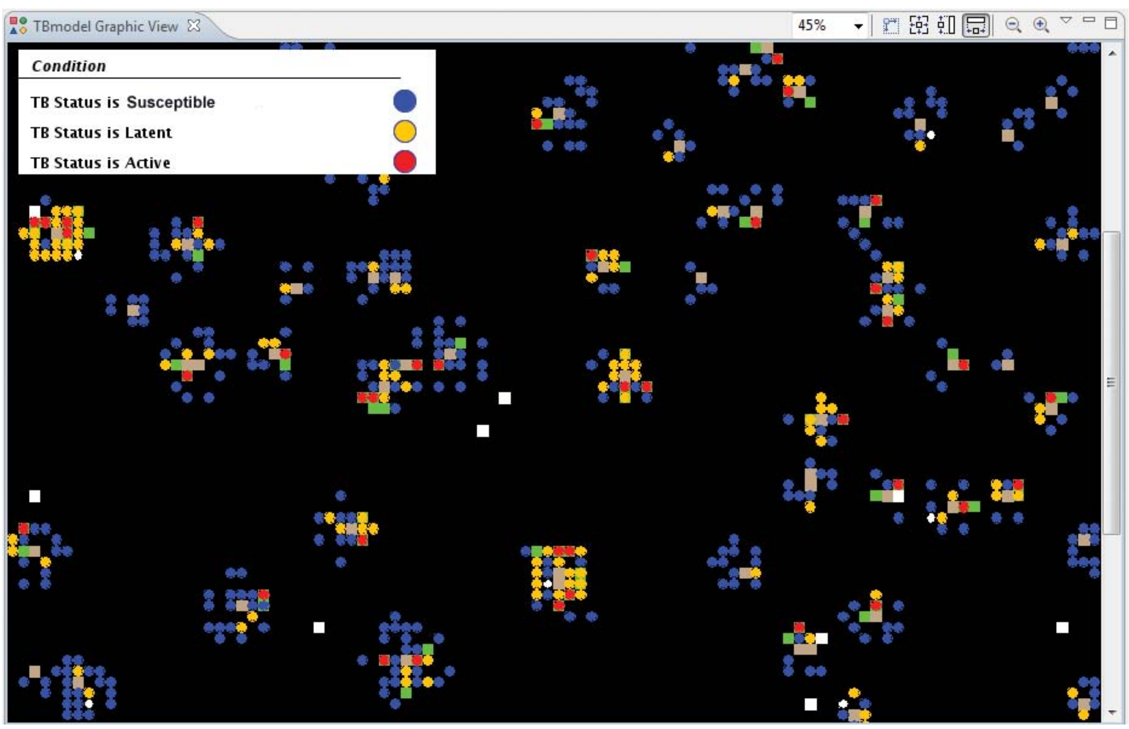
(B)



Source: Gilberto Câmara

Figure 15.9 Simulation of land-cover transition in part of the Amazon Basin. (A) Predictions of a model based on eight years of transitions of individual cells starting with the observed pattern in 1997, using rules that include proximity to roads, changing agricultural conditions, and so on. (B) Observed pattern after eight years of transitions.

Figure 15.10 An illustration from the work of Indy Hurt (Apple), using an agent-based model to simulate the spatial distribution of tuberculosis in a town in Kenya.



Courtesy of Indy Hurt

called the Game of Life (Box 15.4). This has been used by geographer Keith Clarke to model urban growth, using appropriate data and rules of behavior.

Cellular models represent the surface of the Earth as a raster, each cell having a number of states that are changed at each iteration by the execution of rules.

15.2.4 Cartographic Modeling and Map Algebra

In essence, modeling consists of combining many stages of transformation and manipulation into a single whole, for a single purpose. In the example in Box 15.2, the number of stages was quite small—only six—whereas in the Clarke urban growth model

Technical Box 15.4

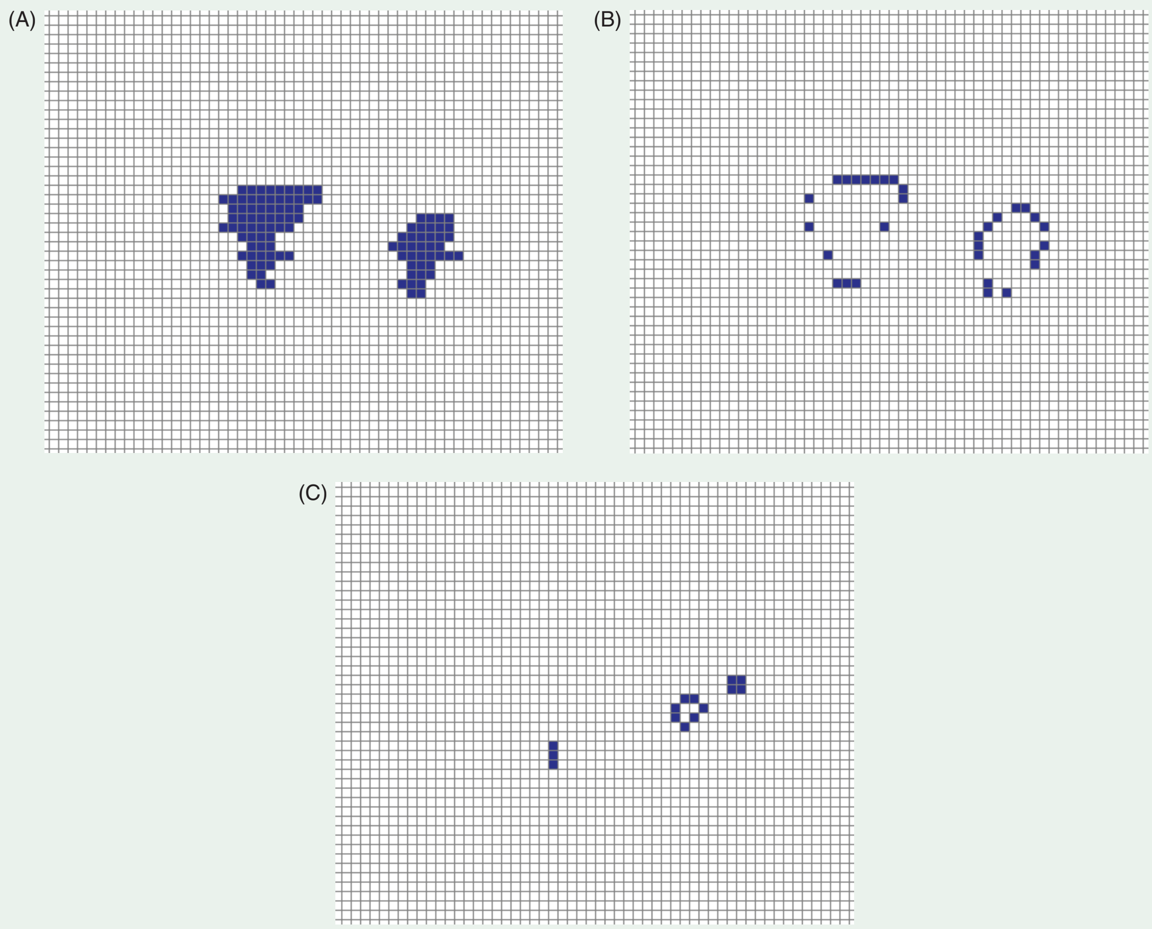
The Game of Life

The game is played on a raster. Each cell has two states, *live* and *dead*, and there are no additional cell attributes (no variables to differentiate the space), as there would be in GI system applications. Each cell has eight neighbors (the Queen's case, Figure 14.9). There are three transition rules at each time-step in the game:

1. A dead cell with exactly three live neighbors becomes a live cell.
2. A live cell with two or three live neighbors stays alive.
3. In all other cases a cell dies or remains dead.

With these simple rules it is possible to produce an amazing array of patterns and movements, and some surprisingly simple and elegant patterns emerge from the chaos. (In the field of agent-based modeling these unexpected and simple patterns are known as

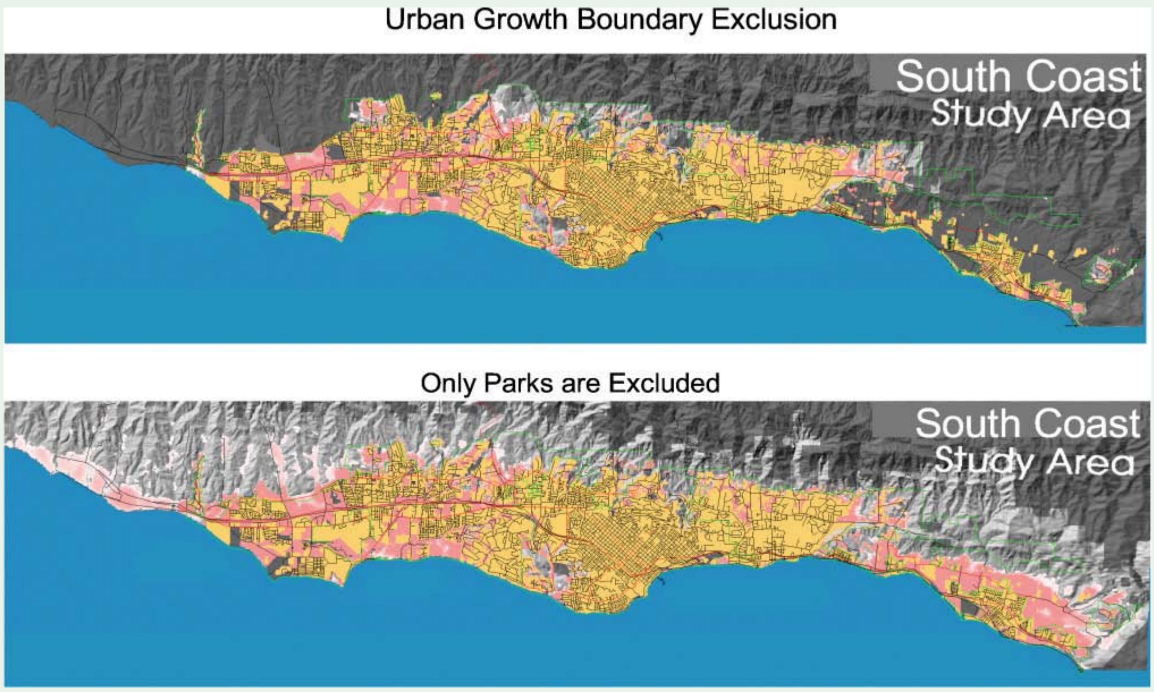
Figure 15.11 Three stages in an execution of the Game of Life: (A) the starting configuration, (B) the pattern after one time-step, and (C) the pattern after fourteen time-steps. At this point all features in the pattern remain stable.



emergent properties.) The page www.math.com/students/wonders/life/life.html includes extensive details on the game, examples of particularly interesting patterns, and executable Java code. Figure 15.11 shows three stages in an execution of the Game of Life.

Several interesting applications of cellular methods have been identified, and particularly outstanding are the efforts to apply them to urban growth simulation. The likelihood of a parcel of land developing depends on many factors, including its slope, access to transportation routes, status in zoning or conservation plans, but above all its proximity to other development. These models express this last factor as a simple modification of the rules of the Game of Life—the more developed the state of neighboring cells, the more likely a cell is to make the transition from undeveloped to developed. Figure 15.12 shows an illustration from one such model, that developed by Keith Clarke and his coworkers at the University of California, Santa Barbara, to predict growth patterns in the Santa Barbara area through 2040. The model iterates on an annual basis, and the effects of neighboring states in promoting infilling are clearly evident. The inputs for the model—the transportation network, the topography, and other factors—are readily prepared in GI systems, and the GI system is used to output the results of the simulation.

One of the most important issues in such modeling is calibration and validation—how do we know the rules are right, and how do we confirm the accuracy of the results? Clarke's group has calibrated the model by *brute force*, that is, by running the model forward from some point in the past under different sets of rules, and comparing the results to the actual development history from then to the present. This method is extremely time consuming because the model has to be run under vast numbers of combinations of rules, but it provides at least some degree of confidence in the results. The issue of accuracy is addressed in more detail, and with reference to modeling in general, later in this chapter.



Source: Keith Clarke

Figure 15.12 Simulation of future urban growth patterns in Santa Barbara, California. (Upper) Growth limited by current urban growth boundary. (Lower) growth limited only by existing parks.

several stages are executed many times, as the model iterates through its annual steps.

The individual stages can consist of a vast number of options, encompassing all the basic transformations of which a GI system is capable. Various ways of organizing the options have been proposed, including that used to structure Chapters 13 and 14. Perhaps the most successful is the one developed by Dana Tomlin and known as *cartographic modeling* or *map algebra*. It classifies all transformations of rasters into four basic classes, and it is used in several raster-centric GI systems as the basis for their analysis languages:

- *Local* operations examine rasters cell by cell, examining the value in a cell in one layer and perhaps comparing it with the values in the same cell in other layers.
- *Focal* operations compare the value in each cell with the values in its neighboring cells—most often eight neighbors (see the discussion of neighborhoods in Chapter 14).
- *Global* operations produce results that are true of the entire layer, such as its mean value.
- *Zonal* operations compute results for blocks of contiguous cells that share the same value, such as the calculation of shape for contiguous areas of the same land use, and attach their results to all the cells in each contiguous block.

With this simple schema, it is possible to express any model, such as Clarke's urban growth simulation, as a series of processing functions and to compile a sequence of such functions into a *script* in a well-defined language, allowing the sequence to be executed repeatedly with a simple command. The only constraint is that the model inputs and outputs be in raster form.

Map algebra provides a simple language in which to express a model as a script.

A more elaborate map algebra has been devised and implemented in the PCRaster package, developed for spatial modeling at the University of Utrecht (pcraster.geog.uu.nl). In this language, a symbol refers to an entire map layer, so the command $A = B + C$ takes the values in each cell of layers B and C, adds them, and stores the result as layer A.

15.3 Technology for Modeling

15.3.1 Operationalizing Models in GI Systems

Many of the ideas needed to implement models at a practical level have already been introduced. In this section they are organized more coherently in a review of the technical basis for modeling.

Models can be defined as sequences of operations, and we have already seen how such sequences can be expressed either as graphic flowcharts or as scripts. The graphic flowchart has already been illustrated in Figure 15.5. In these interfaces datasets are typically represented as ellipses, operations as rectangles, and the sequence of the model as arrows. The user is able to modify and control the operation sequence by interacting directly with the graphic display.

Scripts allow a user to program combinations of GI system operations or to code new operations that are not possible within the standard GI system. Today Python has become a very popular scripting language for GI systems; many GI system tools are now coded in Python, and many users exploit Python to adapt and extend GI system functions to meet specific needs. Models such as the ones being discussed in this chapter can be expressed as sequences of Python commands, allowing the user to execute an entire sequence simply by invoking a script.

Any model can be expressed as a script or visually as a flowchart.

15.3.2 Model Coupling

The previous section described the implementation of models as direct extensions of an underlying GI system, through either graphic model-building or scripts. This approach makes two assumptions: first, that all the operations needed by the model are available in the GI system (or in another package that can be called by the model), and second, that the GI system provides sufficient performance to handle the execution of the model. In practice, a GI system will often fail to provide adequate performance, especially with very large datasets and large numbers of iterations, because it has been designed as a general-purpose software system, rather than specifically optimized for modeling. Instead, the user is forced to resort to specialized code, written in a lower-level language such as C, C++, C#, or Java. Clarke's model, for example, is programmed in C, and the GI system is used only for preparing input and visualizing output. Other models may be spatial only in the sense of having geographically differentiated inputs and outputs, as discussed in the case of the USLE in Section 15.2.1, making the use of a GI system optional rather than essential.

In reality, therefore, much spatial modeling is done by *coupling* a GI system with other software. A model is said to be *loosely coupled* to a GI system when it is run as a separate piece of software, and data are exchanged to and from the GI system in the form of files. Because many GI formats are proprietary, it is common for the exchange files to be in openly published interchange format and to rely on translators or direct-read technology at either end of the transfer. In extreme cases it may be necessary to write a special program to convert the files during transfer, when no common format is available. Clarke's is an example of a model that is loosely coupled to a GI system. A model is said to be *closely coupled* to a GI system when both the model and the GI system read and write the same files, obviating any need for translation. When the model is executed as a GI system script or through the graphic user interface of a GI system, it is said to be *embedded*. Section 6.3.3 describes the process of GI system customization in more detail.

15.3.3 Cataloging and Sharing Models

In the digital computer everything—the program, the data, the metadata—must be expressed eventually in bits. The value of each bit or byte (Box 3.1) depends on its meaning, and clearly some bits are more valuable than others. A bit in Microsoft's operating system Windows is clearly extremely valuable, even though there are hundreds of millions of bits in a single copy of the operating system, and hundreds of millions of copies in computers worldwide. On the other hand, a bit in a remote-sensing image that happens to have been taken on a cloudy day has almost no value at all, except perhaps to someone studying clouds. In general, one would expect bits of programs to be more valuable than bits of data, especially when those programs are usable by large numbers of people.

This line of argument suggests that scripts, models, and other representations of process are potentially valuable to many users and well worth sharing. But almost all the investment society has made in the sharing of digital information has been devoted to data and text. In Section 10.2 we saw how geolibraries, data warehouses, and metadata standards have been devised and implemented to facilitate sharing of geographic data, and parallel efforts by libraries, publishers, and builders of search engines have made it easy nowadays to share text via the Web. But comparatively little effort to date has gone into making it possible to share *process objects*, that is, digital representations of the process of GI system use.

A process object, such as a script, captures the process of GI system use in digital form.

Notable exceptions that facilitate sharing include Esri's ArcGIS Resource Center (resources.arcgis.com), a large collection of resources, including scripts, many of them contributed by users. The library is easily searched for scripts and other resources to perform specific types of analysis and modeling. For example, a search of the online center for the Storm Water Management Model (SWMM) led immediately to tools to implement the model in ArcGIS.

The term *GI service* is often used to describe a function being offered by a server, for use by any user connected to the Internet (see Chapters 6 and 10). In essence, GI services offer an alternative to locally installed GI systems, allowing the user to send a request to a remote site instead of to his or her GI system. GI services are particularly effective in the following circumstances:

- When it would be too complex, costly, or time consuming to install and run the service locally.
- When the service relies on up-to-date data that would be too difficult or costly for the user to constantly update.
- When the person offering the service wishes to maintain control over his or her intellectual property. This would be important, for example, when the function is based on a substantial amount of knowledge and expertise.
- When it is important for several users to collaborate simultaneously in the solution of a problem. By using online services it is possible for many users to view and interact with the problem at the same time.

To date, several standard functions are available as GI services, satisfying one or more of the preceding requirements, and the list is growing rapidly as online services expand. There are several instances on the Web of *gazetteer services*, which allow a user to send a place-name and to receive in return one or more coordinates of places that match the place-name. Several companies offer *geocoding services*, returning coordinates in response to a street address (try www.geocoder.us). Esri's ArcGIS Online (www.arcgis.com) provides many services, plus a searchable catalog to GI services. Such services are important components of the *service-oriented architecture* concept (Section 10.1), in which a user's needs are satisfied by chaining together sequences of such services.

15.4 Multicriteria Methods

The model developed by Pfaff and Glennon and depicted in Figure 15.5 rates vulnerability to runoff based on three factors—cropland, slope, and distance from stream—but treats all three as simple binary measures and produces a simple binary result. Land is vulnerable if slope is greater than 5%, land use is cropping, and distance from stream is less than 300 m. In reality, of course, 5% is not a precise cutoff between nonvulnerable and vulnerable slopes, and neither is 300 m a precise cutoff between vulnerable distances and nonvulnerable distances. Yet such clear cutoffs are essential if a plan is to be administratively workable. The issues surrounding rules such as these, and their fuzzy alternatives, were discussed at length in Section 5.2.

A more general and powerful conceptual framework for models like this would be constructed as follows. A number of factors influence vulnerability, denoted by X_1 through X_n. The impact of each factor on vulnerability is determined by a transformation of the factor $f(X)$. For example, the factor *distance* would be transformed so that its impact *decreases* with increasing distance (as in Section 2.6), whereas the impact of *slope* would be *increasing*. Then the combined impact of all of the factors is obtained by weighting and adding them, each factor i having a weight w_i:

$$I = \sum_{i=1}^{n} w_i f(x_i)$$

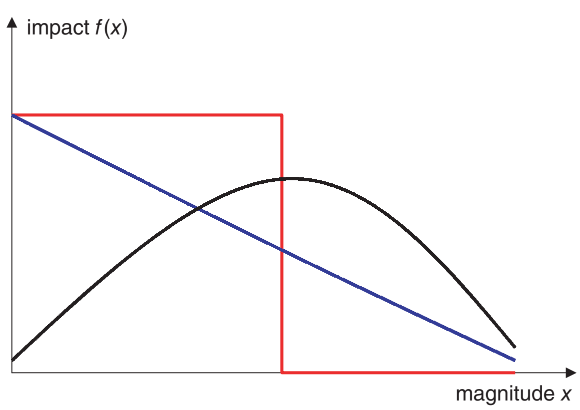
In this framework both the functions *f* and the weights *w* need to be determined. In the example the *f* for slope was resolved by a simple step function, but it seems more likely that the function should be continuously decreasing, as shown in Figure 15.13. The U-shaped function also shown in the figure would be appropriate in cases where the impact of a factor declines in both directions from some peak value (for example, smoke from a tall smokestack may have its greatest impact at some distance downwind—less at shorter distances where the smoke is still descending from the top of the smokestack toward the ground, and less at longer distances where it has become diluted).

Many decisions depend on identifying relevant factors and adding their appropriately weighted values.

This approach provides a good conceptual framework both for the indicator models typified by Box 15.2 and for many models of design processes. In both cases it is possible that multiple views might exist about appropriate functions and weights, particularly when modeling a decision over an important development with impact on the environment. Different *stakeholders* in the design process can be anticipated to have different views about what is important, how that importance should be measured, and how the various important factors should be combined. Such processes are termed *multicriteria decision making*, or MCDM, and are commonly encountered whenever decisions are controversial. Some stakeholders may feel that environmental factors deserve high weight, others that cost factors are the most important, and still others that impact on the social life of communities is all-important.

An important maxim of MCDM is that it is better for stakeholders to argue in principle about the merits of different factors and how their impacts should be measured than to argue in practice about alternative decisions. For example, arguing about whether slope is a more important factor than distance, and about how each should be measured, is better than arguing about the relative merits of Solution A, which might color half of John Smith's field red, over Solution B, which might color all of David Taylor's field red. Ideally, all the controversy should be over once the factors, functions, and weights are decided, and the solution they produce should be acceptable to all because all accepted the inputs. Would it were so!

Figure 15.13 Three possible impact functions: (red) the step function used to assess slope in Figure 15.5; (blue) a decreasing linear function; and (black) a function showing impact rising to a maximum and then decreasing.

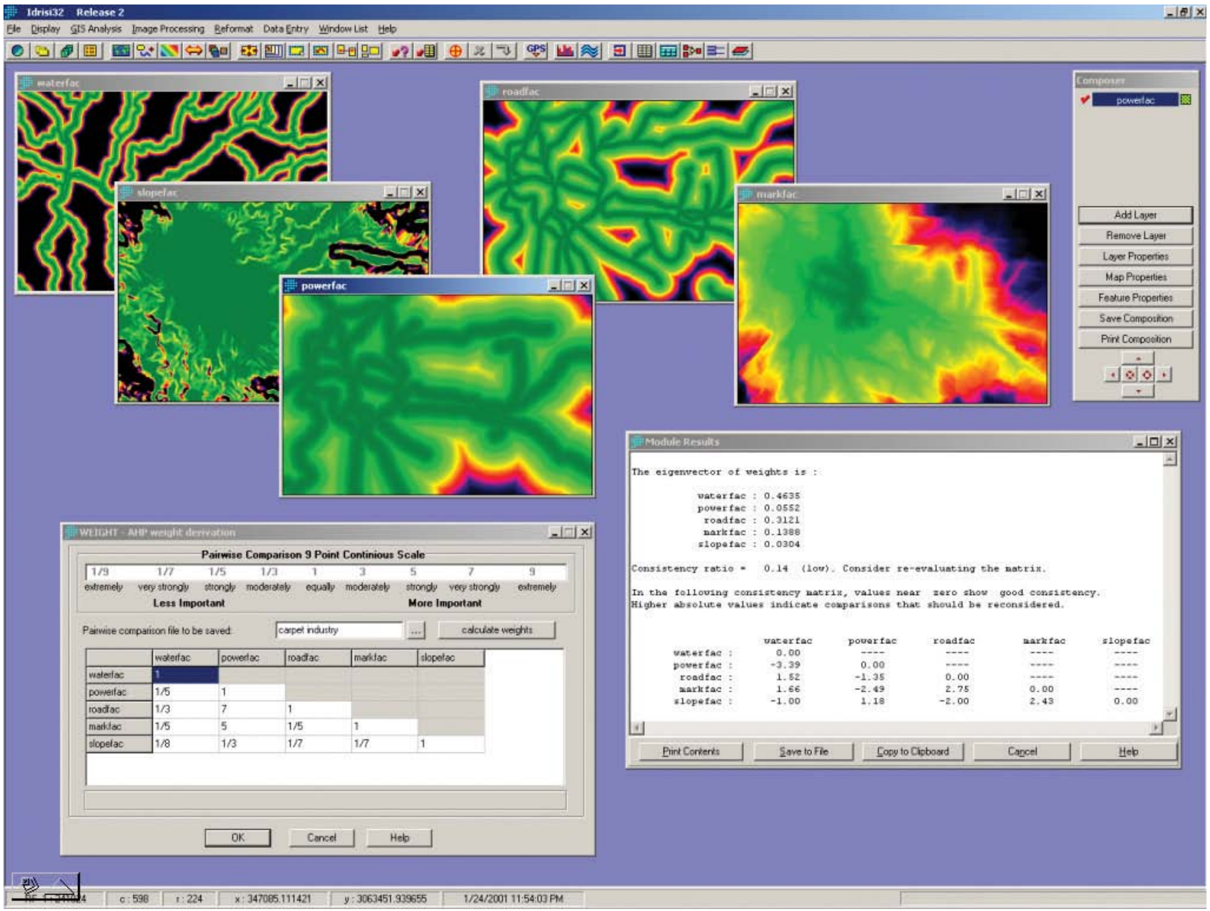


Each stakeholder in a decision may have his or her own assessment of the importance of each relevant factor.

Although many GI systems have implemented various approaches to MCDM, Clark University's Idrisi offers some of the most extensive functionality, as well as detailed tutorials and examples (www.clarklabs.org). Developed as nonprofit software for GI and image processing, Idrisi has many tens of thousands of users worldwide. One example of Idrisi's support for MCDM is the Analytical Hierarchy Process (AHP) devised by Thomas Saaty, which focuses on capturing each stakeholder's view of the appropriate weights to give to each impact factor. The impact of each factor is first expressed as a function, choosing from options such as those shown in Figure 15.13. Then each stakeholder is asked to compare each pair of factors (with n factors there are $n(n-1)/2$ pairs) and to assess their relative importance in ratio form. A matrix is created for each stakeholder, as in the example shown in Table 15.1. The matrices are then combined and analyzed, and a single set of weights is extracted that represent a consensus view (Figure 15.14). These weights would then be inserted as parameters in the spatial model, to produce a final result. The mathematical details of the method can be found in books or tutorials on the AHP.

Table 15.1 An example of the weights assigned to three factors by one stakeholder. For example, the entry "7" in Row 1 Column 2 (and the 1/7 in Row 2 Column 1) indicates that the stakeholder felt that Factor 1 (slope) is seven times as important as Factor 2 (land use).

	Slope	Land use	Distance from stream
Slope		7	2
Land use	1/7		1/3
Distance from stream	1/2	3	



Courtesy: Clark Labs

Figure 15.14 Screenshot of an AHP application using Idrisi (www.clarklabs.org). The five layers in the upper left part of the screen represent five factors important to the decision. In the lower left the image shows the table of relative weights compiled by one stakeholder. All of the weights' matrices are combined and analyzed to obtain the consensus weights shown in the lower right, together with measures to evaluate consistency among the stakeholders.

15.5 Accuracy and Validity: Testing the Model

Models are complex structures, and their outputs are often forecasts of the future. How, then, can a model be tested, and how does one know whether its results can be trusted? Unfortunately, there is still an inherent tendency to trust the results of computer models because they appear in numerical form (and numbers carry innate authority) and because they come from computers (which also appear authoritative). Scientists normally test their results against reality, but in the case of forecasts reality is not yet available, and by the time it is there is likely little interest in testing. So modelers must resort to other methods to verify and validate their predictions.

Results from computers tend to carry innate authority.

Models can often be tested by comparison with past history by running the model not into the future, but forward 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. Instead, many modelers resort to *cross-validation*, a process in which a subset of data are used for calibration and the remainder for validating results. Cross-validation can be done by separating the data into two time periods or into two areas, using one for calibration and one for validation. Both are potentially dangerous if the process being modeled is also changing through time or across space (a statistician would call such a process *nonstationary*), but forecasting is dangerous in these circumstances as well.

Models of real-world processes can be validated by experiment by proving that each component in the model correctly reflects reality. For example, the Game of Life would be an accurate model if some real process actually behaved according to its rules, and Clarke's urban-growth model could be tested in the same way by examining the rules that account for each past land-use transition. In reality, it is unlikely that real-world processes will be found to behave as simply as model rules, and it is also unlikely that the model will capture every real-world process impacting the system.

If models are no better than approximations to reality, then are they of any value? Certainly, human society is so complex that no model will ever fit perfectly. As Ernest Rutherford, the experimental physicist and Nobel Laureate, is said to have once remarked, perhaps in frustration with social-scientist colleagues, "The only result that can possibly be obtained in the social sciences is, some do, and some don't." Neither will a model of a physical system ever perfectly replicate reality. Instead, the outputs of models must always be taken advisedly, bearing in mind several important arguments:

- A model may reflect behavior under ideal circumstances and therefore provide a norm against which to compare reality. For example, many economic models assume a perfectly informed, rational decision maker. Although humans rarely behave this way, it is still useful to know what would happen if they did, as a basis for comparison.
- A model should not be measured by how closely its results match reality, but by how much it reduces uncertainty about the future. If a model can narrow the options sufficiently, then it is useful. It follows that any forecast should also be accompanied by a realistic measure of uncertainty (see Chapter 5).
- A model is a mechanism for assembling knowledge from a range of sources and presenting conclusions based on that knowledge in readily used form. It is often not so much a way of discovering how the world works, as a way of presenting existing knowledge in a form helpful to decision makers.
- Just as Winston Churchill is said to have once remarked, that capitalism was the worst possible economic system "apart from all of the others," so modeling often offers the only robust, transparent analytical framework that is likely to garner any respect among decision makers with competing objectives and interests.

Any model forecast should be accompanied by a realistic measure of uncertainty.

Several forms of uncertainty are associated with models, and it is important to distinguish between them. First, models are subject to the uncertainty present in their inputs. Uncertainty *propagation* was discussed in Section 5.4.2. Here it refers to the impacts of uncertainty in the inputs of a model on uncertainty in the outputs. In some cases propagation may be such that an error or uncertainty in an input produces a proportionate error in an output. In other cases a very small error in an input may produce a variable, sometimes small and sometimes massive change in output; and in other cases outputs can be relatively insensitive to errors in inputs. It is important to know which of these cases holds in a given instance, and the normal approach is through repeated numerical simulation (often called Monte Carlo simulation because it mimics the random processes of that city's famous gambling casino), adding random distortions to inputs and observing their impacts on outputs. In some limited instances it may even be possible to determine the effects of propagation mathematically.

Second, models are subject to uncertainty over their parameters. A model builder will do the best possible job of calibrating the model, but inevitably there will be uncertainty about the correct values, and no model will fit the data available for calibration perfectly. It is important, therefore, that model users conduct some form of *sensitivity analysis*, examining each parameter in turn to see how much influence it has on the results. This can be done by raising and lowering each parameter value, for example by +10% and –10%, rerunning the model, and comparing the results. If changing a parameter by 10% produces a less-than-10% change in results, the model can be said to be relatively insensitive to that parameter. This allows the modeler to focus attention on those parameters that produce the greatest impact on results, making every effort to ensure that their values are correct.

Sensitivity analysis tests a model's response to changes in its parameters and assumptions.

Third, uncertainty is introduced because a model is run over a limited geographic area or *extent*. In reality, there is always more of the world outside the extent, except in a few cases of truly global models. The outside world both impacts the area modeled and is impacted by it. Changes in land-use practices within the extent will influence areas downstream through water pollution and downwind through atmospheric pollution; in turn they will be impacted within the extent by what happens in other areas upstream or upwind. National strategies will be influenced by in-migration, whether legal or illegal. In all such cases the effect is to make the results of any modeling uncertain.

Fourth, and most important, models are subject to uncertainty because of the simplified communication or *labeling* of their results. Consider the model of Box 15.2, which computes an indicator of the need for groundwater protection. In truth, the areas identified by the model have three characteristics: slope greater than

5%, used for crops, and less than 300 m from a stream. Described this way, there is little uncertainty about the results, though there will be some uncertainty due to inaccuracies in the data. But once the areas selected are described as *vulnerable*, and in need of management to ensure that groundwater is *protected*, a substantial leap of logic has occurred. Whether or not that leap is valid will depend on the reputation of the modeler as an expert in groundwater hydrology and biological conservation; on the reputation of the organization sponsoring the work; and on the background science that led to the choice of parameters (5% and 300 m). In essence, this third type of uncertainty, which arises whenever labels are attached to results that may or may not correctly reflect their meaning, is related to how results are described and communicated, in other words to their metadata, rather than to any innate characteristic of the data themselves.

15.6 Conclusion

Modeling was defined at the outset of this chapter as a process involving multiple stages, often a repeated workflow, or perhaps an emulation of some real physical process. Modeling is often dynamic, and current interest in modeling is stretching the capabilities of GI system software, most of which was designed for the comparatively leisurely process of analysis, rather than the intensive and rapid iterations of a dynamic model. In many ways, then, modeling represents the cutting edge of GI systems, and the next few years are likely to see very rapid growth both in users' interest in modeling and in vendors' interest in software development. The results are certain to be interesting.

This chapter has provided a very small and limited sampling of the very rich potential of models, and of the many ways they are being used in a spatial (and temporal) context. Although much work with GI systems is concerned with how the geographic world looks, modeling of processes drives to the very heart of science, in asking how the social and environmental worlds work, and why they evolve as they do. The work of Denise Pumain provides a fitting, final illustration of the fundamental questions spatial modelers are asking (Box 15.5) using today's sophisticated tools and abundant data.

Biographical Box 15.5

Denise Pumain, Demographer, Urban Systems Modeler, and Quantitative Geographer

Denise Pumain (Figure 15.15) is Professor at Université Paris I Panthéon-Sorbonne, Director of the European Research Group "Spatial Simulations for Social Sciences" at CNRS (www.S4.parisgeo.cnrs.fr), and Science Editor of *Cybergeo, European Journal of Geography* (www.cybergeo.eu). She specializes in urban modeling and theoretical geography. Her main scientific contribution is an evolutionary theory of urban systems, transferring concepts and models from self-organizing complex systems to the social sciences. Her latest books *Hierarchy in Natural and Social Sciences* and *Complexity Perspectives in Innovation and Social Change* were published by Springer in 2006 and 2009. She has received several awards (e.g., International Prize for Geography Vautrin Lud 2010, Corresponding Member of the Austrian Academy of Science, Corresponding Fellow of the British Academy) and is the holder of an Advanced Grant (2010) from the European Research Council to research urban system dynamics in the world by analyzing and modeling the geographical diversity of cities and systems of cities. She writes:


Courtesy: Denise Pumain

Figure 15.15 Denise Pumain.

> Spatial analysis is essential for understanding and predicting how cities are expanding their buildings and material or their financial and informational networks at local as well as regional and global scales. Learning about this evolution not only means reconstructing the dynamics of this complex evolution but also looking at specific features of history and culture in several countries and regions whose effects are perceptible in their morphologies and destinies. This knowledge can be inferred both from calibration of analytic models and from experiments with simulation models. When teaching or doing research with students I enjoy trying to explain the surprising long-term existence and stability of cities all over the world, despite the dramatic variations in their appearance. I think that geographic diversity in our way of inhabiting the earth is a lesson about the creativity of humankind and a major resource for future generations!

Questions for Further Study

1. Write down the set of rules that you would use to implement a simple version of the Clarke urban growth model, using the following layers: transportation (cell contains road or not), protected (cell is not available for development), already developed, and slope.
2. Review the steps you would take and the arguments you would use to justify the validity of a GI system-based model.
3. Select a domain of environmental or social science, such as species habitat prediction or residential segregation. Search the Web and your library for published models of processes in this domain, and summarize their conceptual structure, technical implementation, and application history.
4. Compare the relative advantages of the different types of dynamic models discussed in this chapter when applied to a selected area of environmental or social science.

Further Reading

De Smith, M. J., Goodchild, M. F., and Longley, P. A. 2009. *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools* (3rd ed.). Leicester: Troubador and www.spatialanalysisonline.com.

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

Maguire, D., Batty, M., and Goodchild, M. F. (eds.). 2005. *GIS, Spatial Analysis, and Modeling*. Redlands, CA: Esri Press.

Saaty, T. L. 1980. *The Analytical Hierarchy Process: Planning, Priority Setting, Resource Allocation*. New York: McGraw-Hill.

Tomlin, C. D. 1990. *Geographic Information Systems and Cartographic Modeling*. Englewood Cliffs, NJ: Prentice Hall.