Chapter 1 **GIS and Modeling Overview**

MICHAEL F. GOODCHILD

NATIONAL CENTER FOR GEOGRAPHIC INFORMATION AND ANALYSIS

UNIVERSITY OF CALIFORNIA

SANTA BARBARA, CALIFORNIA

ABSTRACT

MODELING CAN BE DEFINED in the context of geographic information systems (GIS) as occurring whenever operations of the GIS attempt to emulate processes in the real world, at one point in time or over an extended period. Models are useful and used in a vast array of GIS applications, from simple evaluation to the prediction of future landscapes. In the past it has often been necessary to couple GIS with special software designed for high performance in dynamic modeling. But with the increasing power of GIS hardware and software, it is now possible to reconsider this relationship. Modeling in GIS raises a number of important issues, including the question of validation, the roles of scale and accuracy, and the design of infrastructure to facilitate sharing of models.

The term modeling is used in several different contexts in the world of GIS, so it would be wise to start with an effort to clarify its meaning, at least in the context of this book. There are two particularly important meanings. First, a data model is defined as a set of expectations about data—a template into which the data needed for a particular application can be fitted. For example, a table is a very simple example of a data model, and in the way in which tables are often used in GIS, the rows of the table correspond to a group or class of real-world features, such as counties, lakes, or trees, and the columns correspond to the various characteristics of the features, in other words, the attributes. This table template turns out to be very useful because it provides a good fit to the nature of data in many GIS applications. In essence, GIS data models allow the user to create a representation of how the world looks. A later section of the chapter provides a more extended discussion of data modeling in the particular context of dynamic models.

Second, a model (without the data qualification) is a representation of one or more processes that are believed to occur in the real world—in other words, of how the world works. A model is a computer program that takes a digital
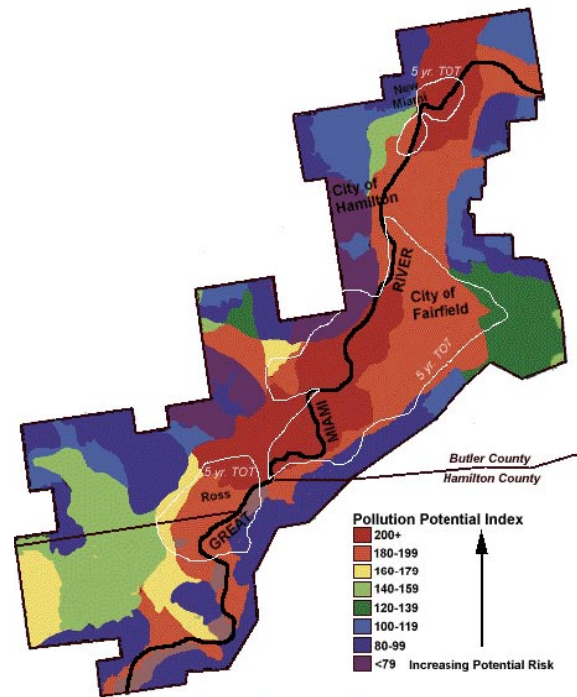
Figure 1. The results of using the DRASTIC groundwater vulnerability model in an area of Ohio. The model combines GIS layers representing factors important in determining groundwater vulnerability and displays the results as a map of vulnerability ratings. (screen shot from http://www.gwconsortium.org/DRASTIC.gif, needs permission)

representation of one or more aspects of the real world and transforms them to create a new representation. Models can be static, if the input and the output both correspond to the same point in time, or dynamic, if the output represents a later point in time than the input. The common element in all of these models is the operation of the GIS in multiple stages, whether they be used to create complex indicators from input layers or to represent time steps in the operation of a dynamic process.

Static models often take the form of indicators, combining various inputs to create a useful output. For example, the Universal Soil Loss Equation (USLE) combines layers of mapped information about slope, soil quality, agricultural practices, and other properties to estimate the amount of soil that will be lost to erosion from a unit area in a unit time (Wischmeier and Smith 1978). The DRASTIC model (fig. 1) estimates geographic variation in the vulnerability of groundwater to pollution, again based on a number of mapped properties (Aller et al. 1987). Dynamic models, on the other hand, represent a process that modifies or transforms some aspect of the Earth's surface through time. Contemporary weather forecasts are based on dynamic models of the atmosphere; dynamic models of stream flow are used to predict flooding from storms; and dynamic models of human behavior are used to predict traffic congestion.

This chapter provides an introductory overview of models and modeling, in the context of GIS. It begins with a discussion of the various types of models that have been implemented in GIS, then describes GIS from a modeling perspective, and finally identifies a series of major issues that confront modelers who use GIS. The chapter serves as an extended introduction to the book, providing a context for the chapters that follow.

All of the models discussed in this book are spatial, meaning that they describe the variation of one or more phenomena over the Earth's surface. The inputs to a spatial model must depict spatial variation, which is why a GIS is a particularly good platform for modeling (this subject is covered in detail in Chapter 2). Moreover, a spatial model's results depend on the locations of the features or phenomena being modeled, such that if one or more of those locations change, the results of the model change.

Modeling can serve a number of purposes. Static models provide indexes or indicators that can provide useful predictors of impacts, sensitivities, or vulnerabilities. The USLE, for example, is widely used to predict soil erosion and to guide management strategies on the part of farmers or county, state, or federal governments to minimize erosion. DRASTIC is widely used as the basis for policies regarding groundwater and to make decisions about the environmental impacts of proposed developments. Dynamic models go further by attempting to quantify impacts into the future and are used to assess different management or development scenarios—what–if scenarios. For example, urban-growth models can be used to predict the impact of land-use controls and future economic conditions on urban sprawl and to devise strategies to

contain sprawl. Atmospheric models are used daily to predict weather conditions as much as seven days into the future.

This experimental aspect of modeling is perhaps its most compelling justification. Aircraft pilots are now routinely trained on simulators, which attempt to emulate the operation of an aircraft in a purely computational environment—as a result, pilots can be brought to a high level of training without the risks associated with the use of real aircraft. Whereas surgeons used to be trained on cadavers, much surgical training now occurs in virtual environments using precise digital representations of the human body. Dynamic modeling of the Earth's environment raises the possibility that we will eventually be able to evaluate the effects of such human activities as the burning of fossil fuels or the release of ozone-destroying chemicals long before such activities actually take place.

TYPES OF MODELS

This section explores the various types of models, placing them in a unifying framework. More detail on several of the contemporary modeling types, including cellular automata, agent-based models, and finite-element and finite-difference models is provided in Chapter 3.

ANALOG AND DIGITAL

Although we rarely consider them in the context of GIS, analog models are even today perhaps the most common type. An analog model is defined as a scale model, a representation of a real-world system in which every part of the real system appears in miniature in the model. For example, architects designing skyscrapers routinely create scale models in order to investigate the effects of high winds on proposed structures, placing the models in wind tunnels to observe deformations under very high stress. Analog models play a key role in the design of aircraft wings, dams and canals, and a host of other engineering projects. Of course the success of analog models depends on the degree to which the system can be scaled—whether the operation of the system in a scaled model is identical to the operation of the real system. A key measure of an analog model is its scale or representative fraction, the ratio of distance between two points in the model to distance between corresponding points in the real world. In an analog model, all aspects of the system must be scaled by the same ratio for the model to be valid.

Ian McHarg, a landscape architect who made many contributions to GIS, originally developed his techniques of ecological planning using an analog version of GIS (McHarg 1969). Each factor important to a decision was represented as a transparent map, with darker areas representing areas of greater impact with respect to that factor. Maps were made for impact on groundwater, human populations, and any other relevant factors. The maps were stacked over a light source, and the areas appearing lightest corresponded to the areas of least impact and were, therefore, the areas most suitable for development. Today,

the same basic principles are embodied in myriad site-suitability analyses conducted using GIS, but with the greater power of the digital computer to vary the weights assigned to each layer and with the mathematical approaches used to combine weighted layers (see Chapter 16).

In a digital or computational model, all operations are conducted using a computer. Data is assembled in a data model and coded using a variety of coding schemes that reduce relevant aspects of the real world to patterns of 0s and 1s. The model itself is also coded in the same limited alphabet, as a computer program or software. Digital models do not have a representative fraction, since there is no distance in the model to compare to distance in the real world (Goodchild and Proctor 1997). Instead, the level of geographic detail is captured in the spatial resolution, or the size of the smallest feature represented in the database. For raster data, this is the size of the individual cell or pixel. When a GIS data set is created by digitizing a paper map, it is helpful to use a simple rule of thumb that the spatial resolution of the data set is approximately 0.5 mm at the scale of the map—in other words, a map at 1:24,000 has a spatial resolution of approximately 12 m. When such information on the lineage of vector data is unavailable, it is difficult to assign a value to spatial resolution since the size of the smallest polygon may be determined by the phenomenon being represented, rather than by the representation. For example, on a map of U.S. states, the smallest state will always be Rhode Island, however detailed the digitized state boundaries.

Besides spatial resolution, temporal resolution is also important in dynamic models since it defines the length of the model's time step. Any dynamic model proceeds in a discrete sequence of such steps, each representing a fixed interval of time, as the software attempts to predict the state of the system at the end of the timestep based on inputs at the beginning of the time step. Both spatial and temporal resolution need to be appropriate to the real nature of the process being modeled. For example, in modeling the atmosphere for weather forecasts, there would be little point in using spatial resolutions as fine as 1 m or temporal resolutions as short as 1 sec because the processes affecting the atmosphere respond to variations that are much coarser than these. On the other hand, 1 m and 1 sec would be quite reasonable resolutions for a model of a small river or stream.

Spatial and temporal resolution determine the relationship between the real world and the model of the real world that is constructed in the computer. The two will never be identical, of course, and any digital representation will leave the user to some extent uncertain about the real world because of the detail that is present in the real world at finer resolutions than those of the model. A model of the atmosphere, for example, is not likely to represent the minute, local, and short-lived fluctuations in pressure caused by the flight of birds. It follows that the predictions of the model will be to some degree uncertain, in the sense that they leave the modeler in the dark about the precise nature of real-world outcomes.

DISCRETE AND CONTINUOUS

Dynamic modelers recognize two very different styles of models. Discrete models emulate processes that operate between discrete entities, such as the forces that operate between celestial bodies and govern their motion, or the behaviors that are exhibited by humans or animals as they interact over space (Chapter 17). Continuous models, on the other hand, are cast in terms of variables that are continuous functions of location, such as atmospheric pressure or temperature, soil acidity or moisture content, or ground elevation. From a GIS perspective, these two possibilities mirror the widely accepted distinction between two conceptualizations of geographic space and geographic variation: the discrete-object view and the continuous-field view (Worboys and Duckham 2004). In the former, geographic space is empty except where it is occupied by point, line, or area objects, which may overlap, do not necessarily exhaust the available space, and are countable. From this viewpoint, the map of U.S. states is a jigsaw puzzle, with 50 pieces (51 including the District of Columbia) that can be moved around at will. The discrete-object view tends to work best in describing and representing biological organisms or human-made features such as buildings, vehicles, or fire hydrants.

In the continuous-field view, the geographic world is described by a series of continuous maps, each representing the variation of a different variable over the Earth's surface. There are no gaps in coverage, and there is exactly one value for each variable at each location. This view tends to work best in describing the variation of physical quantities. Models of the atmosphere are built using this view, though the results are often interpreted in weather forecasts in terms of the behaviors of discrete objects—highs, lows, and fronts. Continuous-field models typically express knowledge of the operation of the physical system in terms of partial differential equations (PDEs) which relate the values, rates of change through time, spatial gradients, and spatial curvatures of the continuously varying quantities. The Navier-Stokes equation, for example, describes the behavior of a viscous fluid, while the Darcy flow equation describes the flow of groundwater through a porous medium. PDEs must be solved through a process of numerical approximation, using either finite-difference methods that represent continuous variation as a raster of fixed spatial resolution or finite-element methods that use polynomial functions over irregular triangles and quadrilaterals (for a discussion of methods for constructing meshes for the solution of PDEs, see Carey 1995).

The so-called gravity or spatial interaction model (Fotheringham and O'Kelly 1989) is an excellent example of a discrete model since it can be used to predict the amount of interaction that will occur in the form of telephone calls, daily journeys to work, numbers of migrants, or numbers of shopping trips between a discrete origin and a discrete destination, arguing by analogy to the gravitational pull that exists between two celestial masses. The model is frequently and easily implemented in a GIS context, using vector representations of the origin and destination features. It is also possible to imagine hybrid models that combine aspects of both approaches, for instance models in which discrete objects representing vehicles or organisms behave in response to local values

of a continuous field. For example, the behavior of an individual in a crowd might be modeled as the response of a discrete object to a continuously varying field of perceived crowding, computed as some form of population density.

INDIVIDUAL AND AGGREGATE

In principle, it is possible to model any system using a set of rules about the mechanical behavior of the system's basic objects. The behavior of a crowd, for example, can be modeled through a series of rules about each individual's behavior, and the development of land-use patterns over an area can be modeled through a series of rules that describe the behavior of each decision maker. But for many systems, the number of basic objects is far too large for this approach to be practical. No coastal geomorphologist would think of modeling the behavior of beaches using rules about the behavior of each individual grain of sand because there would be far too many discrete objects to handle, and it would be far too costly to define the system at time zero—the position and movement of every sand grain at the outset of the simulation, or what are often termed the initial conditions. Similarly, no hydrologist would attempt to model a watershed with rules about the behavior of each molecule of water (Chapter 14).

Continuous-field models address this problem by replacing individual objects with continuously varying estimates of such abstracted properties as density—the density of people in a crowd or the mean velocity and acceleration of water molecules considered as a continuous fluid. Another approach is to aggregate individual objects into larger wholes and to model the system through the behavior of these aggregates. Thus, much modeling of human systems occurs at the aggregate level of census blocks or tracts, and much modeling of hydrologic systems occurs with lumped systems that aggregate areas into entire watersheds or stream reaches. Lumped systems ignore within-lump variation as well as behaviors that modify the variation within lumps, in effect ignoring variation and processes that fall below the implied spatial resolution of the representation.

Over time, the increasing power and storage capacity of computers has made individual-level modeling more practical, and today it is possible to build models involving millions and even billions of objects. The problem of determining initial conditions remains, however, since it is often the result of real constraints on data gathering, which often requires the use of expensive human resources. Technologies such as remote sensing provide a partial solution, allowing the initial conditions over large areas to be characterized at fine spatial resolution, but optical remote sensing is limited in its ability to see through clouds and to differentiate areas based on properties relevant to an investigator's model.

CELLULAR AUTOMATA

In a cellular automaton, spatial variation is represented as a raster of fixed resolution, with each cell being assigned to one of a number of defined states. Such models have been used widely to study processes of urban growth (Chapter 8), in which case the possible states will likely be limited to two: undeveloped and developed. At each time step, the next state of each cell is determined by a number of rules based on the properties of the cell and its neighbors and on the states of the cell and its neighbors. For example, the rules for a simple urban growth model might be as follows:

- If the cell is currently undeveloped, convert to developed with a probability that depends on the slope of the cell, its proximity to a major transportation link (Chapter 10), the zoning of the cell, and the number of its neighbors that are already developed.
- If the cell is currently developed, make no change.

Clarke and his co-workers (e.g., Clarke and Gaydos 1998) have applied models of this type to a number of urban areas in the United States, typically using 30 m spatial resolution and 1 year temporal resolution and forecasting growth for up to 50 years.

The concepts of cellular automata were first explored by John Conway over artificial spaces that were typically uniform and undifferentiated. His interest lay in the sometimes stable properties that emerged after large numbers of time steps, based on particular sets of initial conditions. His Game of Life (Gardner 1970) generates some surprising and intriguing patterns (fig. 2) and was one of the key developments that led to today's strong interest in complex systems and the simple properties that sometimes emerge in such systems, largely independent of initial conditions. Many geographers and others have speculated that similarly surprising patterns might emerge on the Earth's surface through the operation of complex, dynamic processes.

AGENT-BASED MODELS

In an agent-based model, a system's dynamic behavior is represented through rules governing the actions of a number of autonomous agents. Such models can be regarded as generalizations of cellular automata in which agents are able to move around in space, rather than being confined to the cells of a raster—but in other cases the locations of the agents may be irrelevant to the model. Dibble (Dibble and Feldman 2004) has explored the operations of economic agents in simple nonraster worlds similar to the 'small-worlds' popularized by Watts and Strogatz (1998), in which agents occupy locations and can interact both with their spatial neighbors and with certain distant and randomly identified neighbors.
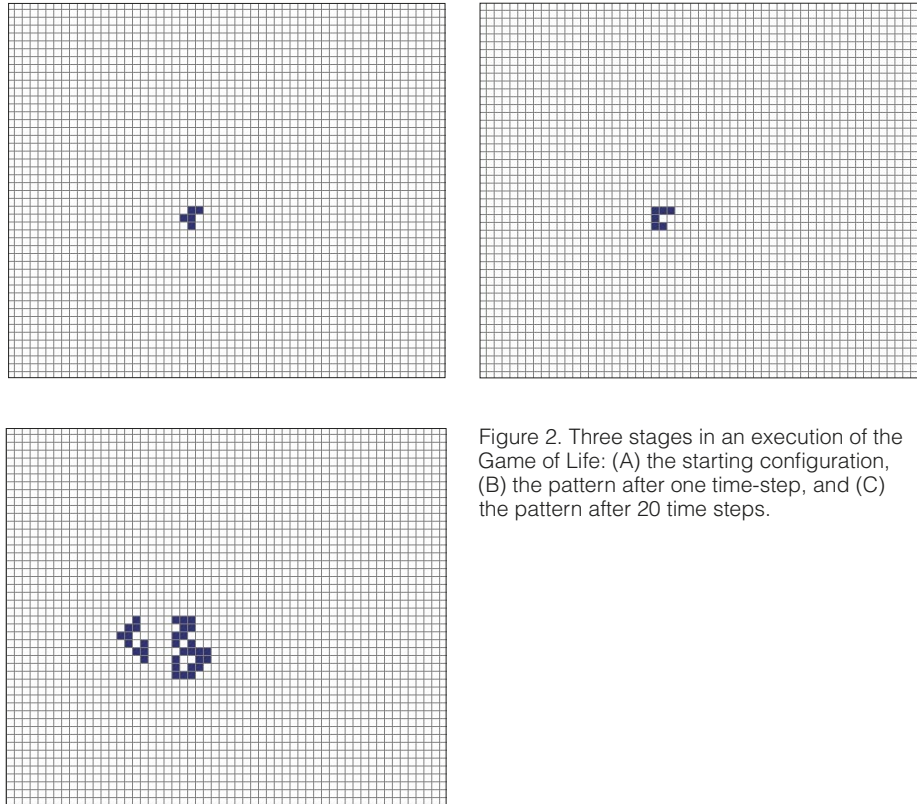
Figure 2. Three stages in an execution of the Game of Life: (A) the starting configuration, (B) the pattern after one time-step, and (C) the pattern after 20 time steps.

Agent-based modeling has found many interesting applications to geographic phenomena. Benenson (2004) has explored the use of such models to represent the behavior of households in cities and the process by which segregation emerges through housing choices. Several efforts have been made to apply agent-based modeling to the emergence of land-use and land-cover patterns (Chapters 6, 18, and 19), with particular emphasis on the processes that lead to greater fragmentation of land cover as a result of development and thus to problems for species that require specialized natural habitat (see, e.g., www.csiss.org/resources/maslucc).

One of the factors that has led to the recent explosion of interest in agent-based models is the emergence of the object-oriented paradigm in software development. Batty (1997) has described the concept of modeling the actions of individuals in a complex geographic landscape through the construction of a set of parallel, independent software modules, each representing the actions and decisions of one actor in the system. Object-oriented languages have made it much easier to conceptualize and build such simulation systems, which are very different in software architecture from the traditional serial approach to computing.

The traditions of GIS are firmly rooted in the map, and even today it is common for GIS to be introduced through the idea of representing the contents of maps in computers. Map-related ideas, such as layers, projections, generalization, and symbolization are still prevalent in GIS and account for a large proportion of the capabilities of a contemporary GIS. So it is by no means clear how a technology built essentially for handling maps can be adapted to the needs of dynamic simulation modeling, and indeed few would think of GIS in that light or suggest that GIS is in any sense the optimum platform for modeling. GIS has never handled time particularly well (Langran 1993, Peuquet 2002), and its representations of continuous variation do not include the irregular meshes of triangles and quadrilaterals that form the basic meshes of finite-element modeling.

On the other hand, there are many good reasons for urging that GIS evolve into an effective platform for spatial modeling, and the technical aspects of doing so are discussed further in Chapter 2. First, GIS is an excellent environment for representing spatial variation, in the initial and boundary conditions of models and in their outputs. GIS also includes numerous tools for acquiring, pre-processing, and transforming data for use in modeling, including data management, format conversion, projection change, resampling, raster–vector conversion, etc.—in fact, all of the tools that would be needed to assemble the data for dynamic simulation. It also includes excellent tools for displaying, rendering, querying, and analyzing model results and for assessing the accuracies and uncertainties associated with inputs and outputs.

Second, much progress has been made recently in the handling of time in GIS. Object-oriented data models have moved the emphasis away from the representation of the contents of maps to a much more general and powerful modeling environment (Zeiler 1999), in which it is possible to represent events, transactions, flows, and other classes of information that would be difficult or impossible to render cartographically.

Third, and perhaps most important, many of the techniques used in GIS analysis would be much more powerful if they could be coupled with an extensive toolkit of methods of simulation. For example, it is widely accepted that the results of GIS analysis are often distorted or biased by the choice of spatial units used in its support. In a classic case study, Openshaw and Taylor (1979) showed that a strong and positive relationship existed between the percentage of people over 65 and the percentage registered as Republicans in each of the 99 counties of Iowa. But by reaggregating the data to units other than counties, in other words by changing the support, they were able to produce correlations ranging from almost perfectly negative (the greater the percentage over 65, the fewer registered Republicans) to almost perfectly positive (the greater the percentage over 65, the more registered Republicans). They coined the term Modifiable Areal Unit Problem (MAUP) for this dependence of analytic results on support and urged that researchers experiment with a range of zoning schemes to determine the specific sensitivity in any actual analysis.

More generally, many of the techniques commonly used for analyzing patterns of points, lines, or areas using GIS (Bailey and Gatrell 1995, Haining 2003, O'Sullivan and Unwin 2003) produce results that are similarly difficult to interpret. An extensive library of simulation methods would allow analysts to compare actual patterns with those expected under a wide range of suitable and interesting conditions. For example, instead of testing whether a map of incidence of cancer displayed a general tendency for clustering, one might test a specific hypothesis relating cancer incidence to data on some known cancer-causing atmospheric or groundwater pollutant.

GIS AND TIME

Over the years, researchers have devised a limited number of ways of handling time within the structures provided by a technology that, as noted earlier, has its roots in the representation of the essentially static contents of maps. The earliest GIS data models were topological, meaning that they included information on such topological properties as adjacency and connectivity. The coverage model—originally developed for the Canada Geographic Information System in the mid-1960s, then for the U.S. Bureau of the Census DIME project for the 1970 census, later for the ODYSSEY project of the Harvard Laboratory for Computer Graphics and Spatial Analysis in the late 1970s, and later still the basis for the original release of ArcInfo in the early 1980s—was designed to represent a partitioning of two-dimensional space into nonoverlapping and space-exhausting polygons. Cartographers know this as the choropleth map, but it also provides an effective representation of any classification of soils, land cover, land use, or surficial geology and also of cadastral maps of land ownership. Many examples of such maps change through time—for example, the map of U.S. county boundaries has changed frequently since Independence as new areas were divided into counties, as county boundaries moved, and as counties were split or merged.

One approach to handling such change is through the concept of a region as an aggregation of smaller areas. All of the county boundaries that ever existed are first mapped, creating a very large number of small basic units. In the coverage model, these are represented as a collection of arcs, each arc defining the boundary between two adjacent units. The counties at any point in time can then be re-created by selecting those arcs that separated counties at that time and assembling them into areas to form that time's regions (Maguire et al. 1992). The same concept of basic units has frequently surfaced in discussions of multiple land classifications, where an integrated terrain unit (ITU) is defined as an area of land that is homogeneous and contiguous with respect to all of the classifications -- all of the original maps can be recreated from a map of ITUs by dissolving appropriate arcs. Regions are also useful for representing events through time that may overlap and do not exhaust space, such as forest fire footprints or land easements.

Another approach consists of tracking the locations of independently moving objects. For example, a collection of individuals might be tracked using GPS, their locations being recorded at every predetermined interval of time. Similar techniques are frequently used to track animals (Chapter 17). In effect, this type of data yields a series of lines in a three-dimensional space formed by the two spatial dimensions (horizontally) and time (vertically), with the restrictions that each line intersects exactly once with any horizontal slice (fixed time) of the model. ESRI Tracking Analyst software has been developed to support simple forms of analysis, summary, and visualization of this type of space–time data. Although it is limited to point-like objects, Agouris and Stefanidis (2003) have developed a version that can be used to represent area objects whose orientation and shape change through time.

A third approach represents each time period as a simple snapshot, typically in raster, and change through time as an ordered sequence of such snapshots. This is the approach inherent in remote sensing. Moving objects are not part of the representation, though they might be detected by some form of image processing and represented using the tracking approach. The approach is used in many raster-based simulation packages, including the GIS PCRaster (Chapter 15; pcraster.geog.uu.nl).

MODELING SOFTWARE

As noted earlier, traditional GIS was designed to support the representation and analysis of maps. Static modeling and the calculation of indicators are classic GIS applications and are well suited to this traditional architecture. Recently, the power of GIS for static modeling has been greatly enhanced by the availability of graphic interfaces that allow the user to interact with the various stages of the modeling process through a simple point-and-click environment. The first of these was perhaps the Imagine software of ERDAS; more recently, ESRI ModelBuilder software is a powerful addition to the spatial analytic capabilities of ArcGIS. These technologies address a fundamental problem of GIS: the vast number of possible transformations and operations that can be performed on geographic data and the complexity in practice of many analysis sequences.

In principle, such software can be used for dynamic modeling through a process of iteration, in which standard GIS functions are used to transform the system at each timestep, and the output of one time step becomes the input for the next. But two problems stand in the way of this. First, the command language of the GIS will not have been designed for iteration, requiring the user to reenter the transformation operations at each step, and second, the poor performance of the system is likely to be frustrating to the user. Scripting languages provide some help in the first regard by supporting the storage and execution of sequences of instructions and by allowing repeated execution of sequences (looping), and today's version of ArcGIS allows scripts to be written in standard

languages such as Microsoft Corporation's Visual Basic for Applications (VBA), Python, and PERL.

PCRaster was perhaps the first GIS designed specifically for simulation, using the ordered-snapshot approach described above. As the name suggests, it is designed to operate on rasters and to implement a range of operations that includes the functions required by a cellular automaton approach to modeling. Tomlin (1990) was the first to systematize the functions that could be performed on raster representations, and his approach has been implemented in numerous raster GIS. Van Deursen (1995) developed the language used by PCRaster to operationalize simple raster functions, through commands that allow entire rasters to be addressed at once—for example, the instruction B = A*2 will take the values in all of the cells of A and double them to create a new raster B. PCRaster includes functions for visualizing its outputs as a movie and has been applied very successfully to the simulation of a range of environmental and social processes (see the examples in Chapter 15 and at pcraster.geog.uu.nl).

Nevertheless, the one-size-fits-all approach that is inherent in GIS and in systems such as PCRaster is unlikely ever to address all possible needs, and instead much attention has been devoted to coupling GIS with packages that are more directly attuned to the needs of modeling (Chapter 6). Matlab is a commonly used toolbox in this context because of its powerful mathematical routines. A prototype linkage between GoldSim and ArcGIS is discussed in Chapter 6. STELLA (www.iseesystems.com) was developed to support dynamic modeling and has the advantage of having a sophisticated visual interface that allows the researcher to express ideas about processes and causality through simple diagrams; STELLA has also been coupled with GIS (Chapter 7). Coupling is also widely used to link standalone models to GIS (Goodchild, Parks, and Steyaert 1993), including models developed to simulate particular environmental processes in areas such as hydrology (Chapter 14).

It is common to distinguish three types of coupling. First, a standalone package might be coupled with GIS by exchanging files—the GIS might be used to prepare the inputs, which are then passed to the modeling package, and after execution, the results of modeling would be returned to the GIS for display and analysis. This approach requires the existence of a format that is understood by both the GIS and the modeling packages or if no such format exists, of an additional piece of software designed to convert formats in both directions. Second, coupling may take the form of integrating the GIS with the modeling packages using standards such as Microsoft's COM and .Net that allow a single script to invoke commands from both packages. This type of integration is now common, based on the compliance to these standards of GIS programs such as ArcGIS and Idrisi, and similar compliance by packages such as Excel and Matlab that have powerful capabilities needed by modelers. The integration occurs through a single script, written in a standard scripting language (Ungerer and Goodchild 2002). Finally, the entire model may be executed by calling functions of the GIS, using a single script (in this option the model is said to be embedded in the GIS). Coupling GIS and modeling systems is discussed at length in Chapter 2.

CALIBRATION AND VERIFICATION

Any attempt to predict the future or to provide indicators of future impact is necessarily problematic, and various techniques are available to assess a model's validity and to build confidence in its results. In general, it seems better to regard a model as a basis for reducing uncertainty about the future from some prior state of complete ignorance to one of more limited uncertainty, rather than to think of a model as failing if its predictions are not perfectly accurate. In other words, and in the language of regression modeling, it would be better to think of a model as improving on $R2=0$ than on failing to achieve $R2=1$.

Many models require some form of calibration, a process of determining appropriate values for one or more parameters that are not specified by theory or past practice. Models are often calibrated and verified using past history, on the grounds that the future will repeat the past. For example, a model of urban growth might be calibrated and verified on the past decades of growth patterns before being applied to forecasting future decades. A common approach is to partition the data into a calibration set and a verification set, using the former to determine the best values of any unknown parameters (by adjusting them to give the best possible fit between the model and the data) and using the latter to verify the model's predictions. Of course, any process of calibration based on past history will only be as valid as its basic assumption that historic trends will continue into the future, at least over the period of the model's forecast.

Alternatively, a model's validity might be assessed based on the validity of each of its component parts. For example, a model that includes rules might be tested by comparing its rules to data on real behavior, rather than by comparing the results of the model as a whole to real data. In practice, this is often the primary basis of assessment, though it is dependent on the assumption that all relevant processes are incorporated in the model.

Sensitivity analysis is also commonly used to assess models. In this approach, the various parameters and inputs are systematically varied to observe their impacts on the model's results. The model might be rerun with the value of a given parameter increased by 10% and then reduced by 10% from its original value. If the impact on the results is substantially less than 10%, the modeler knows that the parameter is not of critical importance and its accuracy is not a major concern. On the other hand, the results may be extraordinarily sensitive to some parameters, and the modeler should therefore invest additional time in ensuring that their values are appropriate.

All geographic data leave their users, to some extent, uncertain about the nature of the real world: because of measurement error, or because detail has been omitted, or because definitions of terms are not rigorous, or because error has crept into the compilation of the data in some way (Zhang and Goodchild 2002). Uncertainty propagation attempts to determine the effects on the results of modeling of known uncertainties in the input data (Chapter 4; Heuvelink 1998). In principle, every prediction of any model should be accompanied by some form of confidence limits, expressing the researcher's uncertainty about the validity of the results.

THE VALUE OF MODELING

At this point, it makes sense to reexamine a question discussed earlier in the introduction to this chapter: why model? From a practical perspective, the answer is surely to reduce uncertainty about the future. But modeling is also conducted for several other reasons. Models may be simply formal representations of belief about process or of how various aspects of the real world work, rather than tools for prediction and forecasting. But formalization has value—in allowing people to communicate in terms that are mutually understood and in allowing knowledge to be expressed in the demanding environment of a digital computer. In court, a model may have great power as an expression of the modeler's willingness to think and operate clearly, to incorporate ideas explicitly, and to address known uncertainties.

Models may also be repositories, structures in which investigators can store knowledge in ways that can be readily executed in what-if scenarios. In this sense, models are not tools for discovering knowledge, but places where discovered knowledge can be brought to bear on real policy questions—models are formal representations of what is known about a system.

But models also contribute to the creation of knowledge, as in the case of the emergent properties discussed in connection with the Game of Life, when the execution of a model reveals something about the real world that was not already known. Batty and Longley (1994) argue that their fractal model of cities led them to a clearer understanding of the processes by which cities develop, and similar arguments are often made about models in other contexts.

MODEL SHARING

Tested, operational models are among the most valuable forms of digital information since they encapsulate a wealth of practical and theoretical scientific knowledge in an easy-to-use form. Thus it is surprising that so much effort has gone into the creation of data repositories, digital libraries, data warehouses, and other sophisticated mechanisms for sharing digital data and so little into the equivalent infrastructure for sharing methods and models. There are no widely accepted methods for describing models in formal, structured terms equivalent to the metadata standards for data sets, and while some collections exist, there is no central clearinghouse for models. Crosier et al. (2003) have proposed such a standard and demonstrated its use in documenting several models. Model and method sharing, or more generally the sharing of process objects, is a core concept of the emerging Grid, the high-performance worldwide network of research computers, and of discussions over cyberinfrastructure, a general name for the use of information technology in the service of collaborative research. There is also increasing interest in providing basic GIS services, such as geocoding, as remotely invokable methods implemented on the Web. In the next few years, dramatic improvements are expected in the availability of techniques for sharing methods and models.

REFERENCES

Agouris P., and A. Stefanidis. 2003. Efficient summarization of spatiotemporal events. Communications of the Association for Computing Machinery 46: 65–6.

Aller L., T. Bennett, J. H. Lehr, R. J. Petty, and G. Hackett. 1987. DRASTIC: A standardized system for evaluating ground water pollution potential using hydrogeological settings. EPA/600/2-87/035. Washington, D.C.: Environmental Protection Agency.

Bailey, T. C., and A. C. Gatrell. 1995. Interactive spatial data analysis. Harlow, UK: Longman.

Batty, M. J. 1997. The computable city. International Planning Studies 2: 155–73.

Batty, M. J., and P. A. Longley. 1994. Fractal cities: A geometry of form and function. San Diego, Calif.: Academic Press.

Benenson, I. 2004. Agent-based modeling: From individual residential choice to urban residential dynamics. In Spatially integrated social science, ed. M. F. Goodchild and D. J. Janelle, 67–94. New York: Oxford University Press.

Carey, G. F., ed. 1995. Finite element modeling of environmental problems: Surface and subsurface flow and transport. New York: John Wiley and Sons.

Clarke, K. C., and L. Gaydos. 1998. Loose coupling a cellular automaton model and GIS: Long-term growth prediction for San Francisco and Washington/Baltimore. International Journal of Geographical Information Science 12: 699–714.

Crosier, S. J., M. F. Goodchild, L. L. Hill, and T. R. Smith. 2003. Developing an infrastructure for sharing environmental models. Environment and Planning B: Planning and Design 30: 487–501.

Dibble, C., and P. G. Feldman. 2004. The GeoGraph 3D Computational Laboratory: network and terrain landscapes for RePast. Journal of Artificial Societies and Social Simulation 7(1). Available: jasss.soc.surrey.ac.uk/7/1/7.html.

Fotheringham, A. S., and M. E. O'Kelly. 1989. Spatial interaction models: Formulations and applications. Boston: Kluwer.

Gardner, M. 1970. Mathematical games: The fantastic combinations of John Conway's new solitaire game "Life." Scientific American, 223: 120–123.

Goodchild, M. F., B. O. Parks, and L. J. Steyaert. 1993. Environmental modeling with GIS. New York: Oxford University Press.

Goodchild, M. F., and J. Proctor. 1997. Scale in a digital geographic world. Geographical and Environmental Modeling 1: 5–23.

Haining, R. P. 2003. Spatial data analysis: Theory and practice. New York: Cambridge University Press.

Heuvelink, G. B. H. 1998. Error propagation in environmental modelling with GIS. London: Taylor and Francis.

Langran, G. 1993. Time in geographic information systems. London: Taylor and Francis.

McHarg, I. L. 1969. Design with nature. Garden City, N.Y.: Natural History Press.

Maguire, D. J., G. Stickler, and G. Browning. 1992. Handling complex objects in geo-relational GIS. Proceedings of the Fifth International Spatial Data Handling Symposium, 652–61.

O'Sullivan, D., and D. J. Unwin. 2003. Geographic information analysis. New York: John Wiley and Sons.

Openshaw, S., and P. J. Taylor. 1979. A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. In Statistical applications in the spatial sciences, ed. R. J. Bennett, N. J. Thrift, and N. Wrigley, 127-144. London: Pion.

Peuquet, D. 2002. Representations of space and time. New York: Guilford.

Tomlin, C. D. 1990. Geographic information systems and cartographic modeling. Englewood Cliffs, N.J.: Prentice Hall.

Ungerer, M. J., and M. F. Goodchild. 2002. Integrating spatial data analysis and GIS: A new implementation using the Component Object Model (COM). International Journal of Geographical Information Science 16: 41–54.

van Deursen, W. P. A. 1995. Geographical information systems and dynamic models: Development and application of a prototype spatial modelling language. Utrecht: Koninklijk Nederlands Aardrijkskundig Genntschap/Faculteit Ruimtelijke Wetenschappen Universiteit Utrecht.

Watts, D. J., and S. H. Strogatz. 1998. Collective dynamics of 'small-world' networks. Nature 393(6684): 440–42.

Wischmeier, W. C., and D. D. Smith. 1978. Predicting rainfall erosion losses: A guide to conservation planning. Agricultural Handbook 537. Washington, D.C.: Department of Agriculture.

Worboys, M. F., and M. Duckham. 2004. GIS: A computing perspective. New York: Taylor and Francis.

Zeiler, M. 1999. Modeling our world: The ESRI guide to geodatabase design. Redlands, Calif.: ESRI Press.

Zhang, J. X., and M. F. Goodchild. 2002. Uncertainty in geographical information. New York: Taylor and Francis.