1
Introduction

MICHAEL F. GOODCHILD and TED J. CASE

In 1650 Elsevier Press published *Geographia Generalis* by the young Bernhard Varenius. It represented one half of Varenius's view of the world, which was divided into general geography (i.e., the principles of the earth's geometry and the processes that are responsible for creating and modifying its human and physical landscape), and special geography (i.e., the description and inventory of the unique characteristics of places). The book came to the attention of Sir Isaac Newton, who had it translated, "improved" and illustrated, and republished in London (Warntz 1989). The prevailing view of the surface of the earth was then and remains today essentially Newtonian, with space and time as rigid frames, and quantum and relativistic effects well outside the range of relevant scales.

This book is about ecology, which focuses on one of the more important of the sets of dynamic processes that are responsible for the patterns and phenomena we find on the earth's surface. *Ecology* studies the dynamics of populations of living organisms, including the interactions that occur both within populations, and between populations and their environment. Ecological knowledge is thus part of Varenius's general geography because all other things being equal the behavior of a population of living organisms is invariant under relocation anywhere on the earth's surface.

1.1 Ecological Models

Dynamic processes are driven by interaction, and the opportunity for interaction typically declines with physical separation. Animals are able to overcome the effects of limited physical separation through motion, and plants through transport mechanisms such as wind, so early attempts to model ecological dynamics often ignored physical separation. Although each organism takes up a finite amount of space and interacts with other individuals only in some neighborhood defined by its movements, space itself and movements across it were largely left out of theoretical thinking. It was implicitly assumed that qualitative results would be insensitive to the locations of individuals. The state variables in these theories were abundances of different species or levels of composite factors, without assigning individuals or populations to specific points in space. Models addressed the existence and stability of various species associations.

It was gradually recognized that several important conclusions were critically linked to the neglect of spatial movements. For example, poorly competitive species that could not coexist locally could nevertheless coexist globally in a wider region, if dispersal rates to unoccupied patches were high enough to compensate for their lack of competitive ability. An understanding of *fugitive species* thus requires a spatial backdrop and limits to dispersal, as well as an understanding of population extinction. For such a fugitive species to persist and colonize, the appearance of new vacant patches clearly must be a relatively common occurrence. Vacant patches can arise by extinctions of competitors in existing patches, or because physical processes (i.e., floods, earthquakes, landslides, fires, or hurricanes) create new openings in the landscape.

For purposes of discussion we can distinguish two types of models. *Patch occupancy models* deal with an archipelago of habitat islands surrounded by a sea of less hospitable terrain. Patch occupancy models use as the state variable the fraction of these patches that are occupied by a species, regardless of its density within. In fact, an assumption is made that the densities in each patch are equivalent, which might be the case if the species quickly reached a carrying capacity that was approximately the same for all patches.

A *meta-population* is simply a population of populations (Levins 1970). In the simplest meta-population model patches already occupied by a species are the source of colonists to unoccupied patches. Furthermore, another very restrictive assumption is typically invoked: The spatial arrangement of occupied and empty patches makes no difference to the colonization rate for empty patches. All empty patches are equally accessible to individuals from all occupied patches regardless of their distance away. Most animals have more limited dispersal abilities than this, and are more likely to find vacant patches nearby than far away.

Patch occupancy models can be modified to incorporate distance-mediated dispersal in an implicit way. Distance is irrelevant for each individual within a particular subpopulation of the meta-population, but movements between populations are affected by some simple function of the physical distance between them. Outcomes of these spatially implicit meta-population models are invariant under relocation of individuals within meta-populations, but not invariant under relocation of meta-populations on the earth's surface.

The second type of model we will call *spatially explicit*. Two subclasses can again be distinguished. In individual-based models, a finite number of individuals are tracked as they move, give birth, and die. Movement algorithms may range from a simple random walk to correlated walks or to more sophisticated learning using neural nets. Parameters of movements, births and deaths can be explicitly tagged to landscape and habitat features. On the other hand, the state variables may be the numbers of individuals at spatial points (continuous space) or in spatial grid cells (discrete space).

A *reaction-dispersal* equation results when space and time are continuous and movements are simple random walks. For two-dimensional space (the plane) the model for species I in space x, y takes the form:

$$\begin{aligned}\frac{dN_i(x,y,t)}{dt} =& F_i(N_1(x,y,t), N_2(x,y,t), \ldots, N_n(x,y,t)) \\ &+ D_i\left[\frac{\partial^2 N_i(x,y,t)}{\partial x^2} + \frac{\partial^2 N_i(x,y,t)}{\partial y^2}\right]\end{aligned}$$

In other words, the growth rate of species i at position x, y and at time t is equal to the sum of a growth function for species i that depends on the densities of all of the species at position x, y at time t (the *reaction*) part of the equation, and a term representing *diffusion* across two-dimensional space x, y with diffusion coefficient D_i. With several interacting species the reaction-diffusion equations are coupled through the reaction F_i functions, and spatial locations are coupled through the diffusion term. The analysis and solution of such equations can be quite formidable. The important feature to note, however, is that the F_i functions will typically contain parameters that affect birth and death rates, whose magnitude depends on spatial position x, y. Finally, the dispersal component may be more complicated than diffusion; for example, movements downhill or downwind may be more likely than uphill or upwind. Movement rules may change in different habitat patches, and obstacles on the landscape may prevent movements for some species but not for others.

By giving each organism a location and expressing movements directly through a spatial metric, both individual-based, patch-occupancy models with distance-mediated dispersal and reaction-dispersal models create situations where ecological outcomes are no longer invariant under relocation of the patch or organisms on the earth's surface.

In the history of dynamical models in ecology, spatially explicit models were born out of purely theoretical concerns, yet, as the models developed and as they stimulated more interest in the movement rules of real organisms and how the texture of the landscape influenced these movements, ecologists began applying these models to real species in real places, not just to abstract processes on hypothetical landscapes. Ecologists addressed environmentally relevant questions such as: How will human alteration of the landscape affect the persistence of legally protected sensitive species? Where are pollutants likely to end up on the terrain? How much soil and nutrients will be lost from a watershed and where will they end up? With sea-level rises due to global warming, which parts of the landscape may become submerged and when? The chapters in this book attest to the several real-world applications.

Two technical innovations also greatly accelerated the application of spatial models to landscape-level problems. Advances in remote sensing allowed a much higher resolution of variation across real landscapes to be provided in readily accessible data, and geographic information systems (GIS) provided a spatially tagged record-keeping system for managing, visualizing, and analyzing these data (for general introductions to GIS see, for example, Burrough and McDonnell 1998; Longley et al. 1999; and for their applications in ecology see, for example, Goodchild et al. 1993, 1996; Haines-Young et al. 1993; Johnston 1998). For the first time, ecologists could test the hypotheses that they had developed at the landscape scale. Ecology embraced geography: remote sensing and GIS allowed maps at large spatial scales to become dynamic in ecologically relevant time scales.

With the application of ecological research to legislatively driven environmental problems came a greater need to quantify the likelihood of alternative outcomes predicted by dynamic models. Error and uncertainty enter at various stages and mix into the synthesis of ecological thinking and into the formats of models. If ecology was to be relevant it would have to quantify its uncertainty about spatial dynamics, and do so urgently, within the time scales of environmental decision making. This book is a direct outgrowth of this urgent need.

1.2 Uncertainty in Spatial Processes

The outcomes of any dynamic ecological process depend directly on the initial or boundary conditions that provide its inputs. Even if the earth were a flat, rotationless object in a fixed position relative to a constant sun, ecological and other processes would have created nonuniformity, which would then influence the outcome of subsequent processes. Termites have to build their mounds somewhere, even on a flat, featureless plain; and the basic instability of many processes ensures, for example, that rivers will form and erode valleys in response to the slightest imperfection in the surface. In reality, however, the boundary conditions provided by the earth's surface are extremely diverse. Some processes serve to smooth, with outcomes that reduce the high-resolution variation in their inputs (i.e., glaciation removes much of the high-resolution variation in terrain). Other processes simply preserve the spatial patterns of their inputs; and many ecological processes create or enhance high-resolution variation, as in the case of the termite mounds or the landscape architecture of the beaver. Outcomes of one set of processes provide inputs to others, ensuring that any process on the earth's surface will have complex spatial variation in its boundary conditions.

The boundary or initial conditions of ecological processes are Varenius's special geography, captured today in the spatial databases that are increasingly available in digital form through such technologies as remote sensing, GIS, and the World Wide Web. To someone concerned with planning, managing, or conserving, knowledge of both forms of geography is essential in order to predict outcomes, and make rational decisions. In Figure 1.1, ecological processes operate under boundary conditions to create outcomes, which are then analyzed in order to support decisions. It would clearly be absurd to try to make decisions based on knowledge of processes alone without understanding what outcomes those processes caused under the boundary conditions present at some location on the earth's surface.

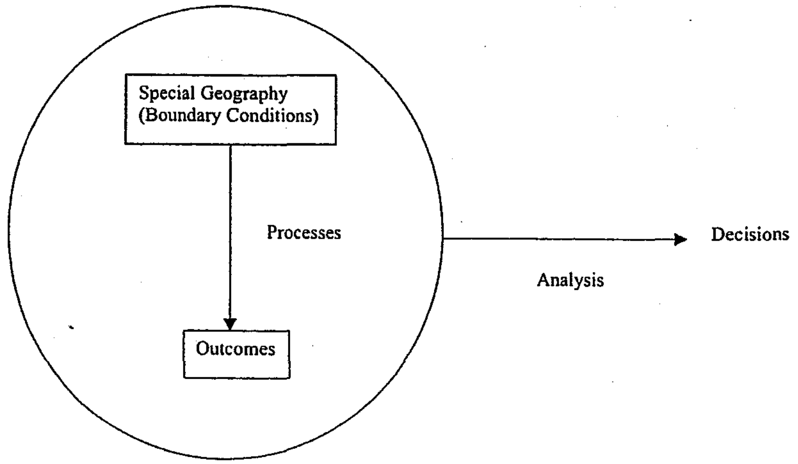

FIGURE 1.1. The role of spatial data, process models, and predictions in decision making.

This book is about the problems that uncertainties create in this simple framework. It is virtually impossible to have certainty about the analysis process that leads to decisions because in today's world decisions almost always involve multiple stakeholders with multiple viewpoints. Knowledge of ecological processes is similarly never perfect, so uncertainty is present whenever outcomes are predicted. Finally, we can never have perfect knowledge of the boundary conditions because the real world is far too complex to be fully measured, observed, or represented. This book covers all aspects of uncertainty, although its heaviest emphasis is on the third kind: uncertainty in the spatial data that provide the boundary conditions for ecological processes. We have chosen to place the emphasis there for several reasons: spatial data uncertainty is a rich area of research in several fields, as the chapters of this book illustrate, but this literature is relatively unknown in ecology, and the methods are not often applied; application in ecology will help to enrich the set of examples and motivating applications for the general study of spatial data uncertainty; and we believe the problems posed by spatial data uncertainty are among the most challenging, the most important for decision makers, and may in some cases be the largest in magnitude.

Spatial data uncertainty arises for many reasons. Complete measurement of the real world may be impossible because of cost, or because the capacity to store and process the resulting volume of data is simply not available. Generalization is almost inevitable, with high-resolution data either not being collected or being discarded at some point through a process of generalization. One of the largest sources of uncertainty lies in the lack of formal methods of generalization; instead, generalization in fields like cartography often has elements of subjectivity and art that have proven very resistant to formalization, and it may be difficult to determine exactly what objective relationship exists between a generalized map (i.e., a map of soils) and the reality that it is intended to represent. Uncertainty arises from measurement errors, notably measurements of position, and from the spurious changes introduced by format conversion and digitizing. It is also present in the essential fuzziness and vagueness associated with much geographic data. We define *uncertainty* for the purposes of this book as:

The state of knowledge about a relationship between the world, and a statement about the world.

Uncertainty defined in this way exists, for example, when an item of data indicates that the elevation at a given point is 200.5 m, but we know that the expected difference between that item of data and the true elevation is 2.4 m; or we may know that the difference between the true vegetation class in an area, and the database record for that area, is characterized by a probability of correct classification of 0.85.

1.3 The Book's Organization

The book has two major parts. The first section includes many demonstrations of the use of spatial data in ecological analysis, and instances of the framework shown in Figure 1.1. The chapters in this section demonstrate the importance of all three kinds of uncertainty and its essentially multidimensional nature.

The second section presents a collection of models of uncertainty, framed within a variety of paradigms that range from spatial statistics to cognitive science. It is clear to us that no single paradigm can fully frame the complex multidimensional issue of uncertainty, and that some paradigms may be more useful in some circumstances than others. Each of them, though, possesses elements of the four-stage structure shown in Table 1.1, which provides an expanded view of the processes implicit in Figure 1.1, and is commonly used to structure discussions of spatial data uncertainty and its impacts (see, e.g., Goodchild and Gopal 1989; Guptill and Morrison 1995; Burrough and McDonnell 1998; Heuvelink 1998; Goodchild and Jeansoulin 1998).

TABLE 1.1. Issues at various stages of application of spatial data.

Stage	Uncertainty issue
Data	Description of uncertainty
Display	Visualization of uncertainty
Analysis, modeling	Models of uncertainty
Decision making	Propagation of uncertainty, confidence limits

Attention initially focuses on the description of the uncertainty present in data. Although various generic methods of uncertainty description exist, and are reviewed in many of the chapters of this book, the problem is always to describe and measure uncertainty in generic ways that are not specific to applications. We hope, for example, that in describing the positional uncertainty of features in a data set we will later be able to analyze the impacts of this uncertainty on the results of analysis and modeling in numerous applications, but that may not always be true. In that context, the uncertainties described and measured by a collector or disseminator of data represent a compromise designed to satisfy as many as possible of the diverse needs of the data's users, just as the standard deviation represents a compromise among the numerous possible ways of describing variation in a sample of measurements of a scalar.

The second stage addresses the vital issue of communication of uncertainty to the users of data. Data increasingly pass between numerous custodians in their life cycle, and the person who eventually uses the data may be unable to contact or communicate with the person who created them. Moreover, that person may be a member of the general public or of a nontechnical discipline who is unfamiliar with the complexities of spatial statistics, and likely to reject any description that is couched in such terms. Visualization is one of the most important tools for communicating effective information on uncertainty and for giving the user some useful basis for assessing its impacts.

Models of uncertainty (e.g., the statistical models exemplified by the classical Gaussian distribution) are essential if we are to understand the impacts of uncertainty on the results of modeling and analysis. They provide the basis for an analytic derivation of confidence limits in a specific instance, or the basis of simulation models which will be used to explore the impacts of uncertainties, especially for users who are not technical experts.

Finally, in the best of all possible worlds every user of spatial data would be provided with a set of confidence limits on any results of analysis, predictions of models, or inputs to decisions. Those confidence limits might then be incorporated into decision-making processes to assess the risks associated with each option. Increasingly, the results of decisions made using spatial data, via such technologies as GIS, are being tested in court under legalistic standards of rigor. Today, maps of wetlands or vegetation cover are being used to implement regulations regarding land use, and those regulations are clearly open to challenge if it can be shown that the regulator knew about uncertainties in the maps, and failed to deal with them appropriately or responsibly. Even though uncertainties may be no more than an inconvenience to the scientific community, they may be the Achilles' heel of attempts to turn the results of science into public policy.

References

Burrough, P.A., and R.A. McDonnell. 1998. Principles of geographical information systems. Oxford University Press, New York.

Goodchild, M.F., and S. Gopal, eds. 1989. Accuracy of spatial databases. Taylor & Francis, London.

Goodchild, M.F., and R. Jeansoulin, eds. 1998. Data quality in geographic information. Editions HERMES, Paris.

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

Goodchild, M.F., L.T. Steyaert, B.O. Parks, C.A. Johnston, D. Maidment, M. Crane, et al., eds. 1996. GIS and environmental modeling: progress and research issues. GIS World, Fort Collins, CO.

Guptill, S.C., and J.L. Morrison, eds. 1995. Elements of spatial data quality. Elsevier Science, New York.

Haines-Young, R., D.R. Green, and S.H. Cousins, eds. 1993. Landscape ecology and GIS. Taylor & Francis, London.

Heuvelink, G.B.M. 1998. Error propagation in environmental modelling with GIS. Taylor & Francis, London.

Johnston, C.A. 1998. Geographic information systems in ecology. Blackwell Science, Oxford.

Levins, R. 1970. Extinction. Pages 77–107 in M. Gerstenhaber, ed. Some mathematical problems in biology. American Mathematical Society, Providence, RI.

Longley, P.A., M.F. Goodchild, D.J. Maguire, and D.W. Rhind, eds. 1999. Geographical information systems: principles, techniques, applications and management. John Wiley & Sons, New York.

Warntz, W. 1989. Newton, the Newtonians, and the *Geographia Generalis Varenii.* Annals of the Association of American Geographers 79(2):165–91.