PREFACE

Michael F. Goodchild
National Center for Geographic Information and Analysis, and Department of Geography, University of California, Santa Barbara, CA 93106-4060. Phone 805 893 8049 FAX 805 893 3146 Email good@ncgia.ucsb.edu.

In late 1989 I met Brad Parks at an EPA-sponsored workshop near Detroit, and we persuaded each other that much could be gained by encouraging greater interaction between the geographic information systems (GIS) and environmental modeling communities. GIS was gaining ground rapidly in the federal agencies and among policy makers, and was being widely accepted as a tool for the analysis of problems in their spatial context. Environmental models were also proliferating and winning acceptance as tools for numerical analysis and prediction. If they could be integrated, these two computer applications seemed to offer great promise in bridging some of the gaps that we perceived between science and policy. Out of that conversation grew the series of International Conferences/Workshops on Integrating GIS and Environmental Modeling, in Boulder (1991), Breckenridge (1993), Santa Fe (1996), and Banff (2000). The 1991 conference was tightly structured, with sessions and invited speakers on each of the major domains of modeling (hydrological, land surface/subsurface, biological/ecological, and integrated) and cross-cutting issues, and resulted in a text (Goodchild *et al.*, 1993). The 1993 conference proceedings are available in a book (Goodchild *et al.*, 1996); the 1996 proceedings are on the Web (www.ncgia.ucsb.edu under Meetings) and on CD (see the NCGIA Web site); and details of the 2000 meeting are also available via the NCGIA web site (under GIS/EM4) and at www.colorado.edu/research/cires/banff.

Besides the obvious community-building objective, we also hoped that the conferences would address a series of technical objectives. Modern science and problem solving require extensive collaboration between specialists, and the high costs of data acquisition and model development argue strongly for effective mechanisms for sharing information. Sharing is particularly difficult across the boundaries of disciplines and information communities, who may use the same terms in different ways, and one important potential benefit of community-building is the construction of a common language (see particularly the arguments presented by Michael Morrison and Linnea Hall in Chapter 2 of this book). The Web has helped enormously to foster a culture of collaboration and sharing over the past seven years, but at the same time these developments have pointed to some critical weaknesses, in our ability to deal effectively with a list of pervasive problems that includes scale, representation, accuracy, visualization, and many more. The conferences helped to focus attention on these problems, which form the core agenda of the emerging field of geographic information science (see www.ucgis.org, and also UCGIS, 1996). Two in particular—accuracy and scale, and their importance to the prediction of species~~occurrence—~~ occurrence—form the theme of this new book.

These problems all stem from two more-fundamental ones. First, the real geographic world is infinitely complex, such that it would require an infinite and therefore impossible amount of information to characterize it completely. Any description of any aspect of the world must therefore be an approximation, generalization, or abstraction, that almost certainly omits much of the detailed information that organisms sense and respond to. This missing information constitutes uncertainty, in the sense that a user of a database has a degree of uncertainty about the true conditions existing in the real world (for a review of uncertainty in spatial ecology see Hunsaker *et al.*, 2000). It also follows that an infinite number of potential approximations exist, or in other words that an infinite number of ways can be found to represent any aspect of the world in digital form. Some of these will do much more damage to our ability to model and predict than others, because their associated uncertainty translates or propagates into uncertainty over model predictions, so finding the metrics that allow us to evaluate a representation's performance in environmental modeling and prediction is a vitally important task. The metric long favored by cartographers and producers of geographic databases—the scale or representative fraction of a paper map—is almost certainly not sufficient and not the best in any given task; and there are still an infinite number of possible digital representations at any given scale, depending on how the objects represented in the database are specified. Advances in computing power over the past few decades have made it possible to handle much larger data volumes and to model more accurately, as Dean Stauffer argues in Chapter 3, and as many other chapters demonstrate, but in the final analysis data noise will always limit our ability to reduce uncertainty.

Second, GIS has many of its roots in cartography. In the age of the printing press and paper, maps were an ideal way of sharing what was known about the world. Because maps take a long time to draw, and are essentially static once printed, the information shown on maps tended to reflect what was constant, and to omit what was transitory, relegating the latter to the 'uncertainty basket'. Because there are massive economies of scale in the printing process, maps tended to reflect information that was generally useful to large numbers of people, rather than the specialized needs of a few. But all of this changes in a digital world, in which information can be copied and distributed at zero cost, and at electronic speed. Goodchild (2000) details eight ways in which maps constrain or filter our ability to know about the geographic world, and shows how all eight constraints disappear when the medium of communication is digital. Computers are also able to capture not only how the world *looks*, in the form of largely static geographic data, but also how it *works*, in the form of codes that implement environmental process models.

But GIS also inherited the representations of the world developed by cartographers. The data that populate our GIS databases today were largely created to appear on maps, and many of them were obtained by digitizing or scanning those same maps. They use the scales and accuracies that were devised long ago to serve the needs of mapping. Vegetation-cover mapping practice, which emphasizes the delineation of homogenous areas of cover class separated by sharp boundaries, made sense to users of maps (and Kristina Hill and Michael Binford argue in Chapter 7 that this approach is fundamental to human reasoning), but it may be totally inappropriate for today's GIS uses, including dynamic modeling of ecological processes. The concept of a patch as an ecologically meaningful unit may appear superficially similar to the concept of an area of uniform cover class, but at a more fundamental level they may have little relationship. Similarly, the pixels of a remotely sensed image may appear to be convenient units for modeling ecological process, but they originated in the geometry of an instrument on a satellite, and were in no way informed by ecological reality. So bringing GIS and environmental modeling together offered promise at a number of levels: improved support for environmental modeling through better tools for sharing and managing data, and for visualizing and disseminating results; and improved representations in GIS that were more appropriate for a new generation of requirements.

In her chapter in the book from the first conference (Johnston, 1993), Carol Johnston presents an appealing conceptual structure for spatially explicit modeling of ecological populations that is firmly grounded in theory. *K* strategists base their survival and success on local resources, and are relatively easy to model in GIS, provided local resources are represented in suitable ways. Since organisms respond to resources over a local neighborhood, in effect integrating through a spatial convolution function, it is essential that the representation have an appropriate spatial resolution that is no coarser than the convolution if the convolution is precomputed, and much finer if the convolution must be computed on the fly by the model. Vector representations, which lack explicit spatial resolution, are likely to be unsuitable in these models.

r strategists present much more difficult problems. Here, success is determined by the intrinsic rate of natural increase of the population, and hence on spatial interaction between organisms. But spatial interaction presented enormous problems to cartographers, and very little progress has been made in the development of effective representation of interactions in GIS databases. Multiagent models attempt to model interaction at the individual level, and clearly do not scale effectively to large populations. Island models deal with interaction by assuming it to be perfect within the island, and absent otherwise, and it is tempting to think that landscapes can be similarly fragmented into isolated patches. But simple rules for the identification of patches (e.g., every watershed is a patch, every area of homogeneous land cover class is a patch, every Landsat pixel is a patch) are unlikely to have much grounding in ecological reality. Jason Dunham, Bruce Rieman, and James Peterson explore this theme in relation to modeling fish occurrences in Chapter 26; in Chapter 58 David Theobald and Thompson Hobbs present an alternative approach that explicitly addresses gradients; and in Chapter 62 Thomas Sisk, Barry Noon, and Haydee Hampton create a patch model that explicitly recognizes within-patch heterogeneity.

Almost a decade has elapsed since the first GIS and environmental modeling conference, and several changes have occurred in that time. First, the growth of the Web and associated data archives, digital libraries, and clearinghouses has created an environment that is much more conducive to data sharing and access—environmental modelers are now blessed with a much more abundant supply of geographic data. Second, this has raised awareness of the weaknesses of much of the data supply. Too often the rules used to create data are not known, or too subjective, or data have not been effectively validated, or quality is too uncertain and variable. Too often the high cost of accurate data forces us to accept low-resolution data without adequate ways of assessing what has been lost (see, for example, the analysis presented by Kathryn Thomas and her coauthors in Chapter 10). Using data that are too coarse can affect not only the goodness of fit of the model, but also its structure, as Catherine Johnson and her co-authors show in Chapter 12. In the best of all possible worlds questions of structure would be resolved by resort to theory, but in Chapter 15 Claudine Tobalske shows how alternative theories can often lead in conflicting directions.

New acquisition systems are solving some of these problems—for example, the data becoming available from the Shuttle Radar Topography Mission of 2000 clearly have some advantages over previous digital elevation data, and new satellite sensors offer improved spatial, temporal, and spectral resolution. Third, there has been substantial improvement in software support for modeling. GISs aimed specifically at dynamic modeling have appeared, such as the University of Utrecht's PCRaster (www.geog.uu.nl/pcraster/), and powerful scripting languages have been defined (Van Deursen, 1995). Many of the chapters in this volume illustrate the benefits of these changes, and the software environments described in the chapters of Part 4 are very different from those available ten years ago.

Much more important as we look to the future, however, are the changes that are still needed. Old habits die hard, and the world has invested vast resources over the centuries in creating representations of geographic knowledge that are in many cases highly inappropriate for ecological modeling. We need a new generation of representations that are specifically designed to make ecological sense, and to do minimal damage to the accuracy of ecological models that use them for prediction. Specifically, such representations could:

- maintain explicit spatial resolution (rasters are preferable in this respect to vectors);
- support multiple representations (e.g., preserving both raw data and more-generalized interpretations based on them), to implement reasoning and modeling across spatial scales, a theme explored by Brian Maurer in Chapter 9, and in many of the other chapters in Part 2;
- interrupt continuity at ecologically significant barriers, rather than in areas of steep gradient (e.g., bounding areas of approximately homogeneous cover);
- support the representation of interaction [e.g., through vector fields, metamaps (Takeyama and Couclelis, 1997), or attributed relationships];
- support the measurement and modeling of uncertainty, and its propagation through models to confidence limits on outputs;
- support the observation and modeling of change through representations that are fully spatio-temporal (e.g., define objects as regions in three-dimensional space-time, or as space-time trajectories).

All of these examples are part of a much-longer-term transition as GIS evolves away from its cartographic roots, and moves to center stage in support of environmental modeling. They are also highly relevant to the successful prediction of species occurrence. But the strong scientific basis of these suggestions is in sharp contrast to other trends in the GIS community, towards a more human-centric approach that emphasizes the vagueness of everyday geographic knowledge, the fuzzy nature of human reasoning, the importance of GIS as a technology of communication between people, and the representation of subjective knowledge. Efforts to integrate fuzzy logic into mapping practice are also in some respects contrary to the principles of objective science. It remains to be seen how this emerging and fundamental tension in the GIS community will be resolved, and how it will affect progress in environmental modeling in general, and the prediction of species occurrence in particular.

References

Goodchild M F 2000 Communicating geographic information in a digital age. *Annals of the Association of American Geographers* 90(2): 344–355

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

Goodchild M F, Steyaert L T, Parks B O, Johnston C A, Maidment D, Crane M, Glendinning S editors 1996 *GIS and environmental modeling: progress and research issues.* Fort Collins: GIS World Books

Johnston C A 1993 Introduction to quantitative methods and modeling in community, population, and landscape ecology. In M F Goodchild, B O Parks, and L T Steyaert editors *Environmental modeling with GIS.* New York: Oxford University Press, pp. 276–283.

Takeyama M, Couclelis H 1997 Map dynamics: integrating cellular automata and GIS through Geo-Algebra. *International Journal of Geographical Information Science* 11(1): 73–91

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

Van Deursen W P A 1995 *Geographical information systems and dynamic models: development and application of a prototype spatial modelling language.* Nederlandse Geografische Studies 190.Utrecht: Koninklijk Nederlands Aardrijkskundig Genntschap/Faculteit Ruimtelijke Wetenschappen Universiteit Utrecht.