GIS MODELING IN RASTER

Michael N. DeMers

New Mexico State University

JOHN WILEY & SONS, INC.

ACQUISITIONS EDITOR Rayan Flahive
MARKETING MANAGER Clay Stone
SENIOR PRODUCTION EDITOR Patricia McFadden
SENIOR DESIGNER Harry Nolan
PRODUCTION MANAGEMENT SERVICES Hermitage Publishing Services

This book was set in Cheltenham by Hermitage Publishing Services and printed and bound by Courier Westford. The cover was printed by Lehigh Press, Inc.

This book is printed on acid-free paper. ♾

Library of Congress Cataloging in Publication Data:
DeMers, Michael N.
GIS modeling in Raster / Michael N. DeMers.

ISBN 0-471-31965-1

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

specific measure associated with that character, a potential cause, and possible effect. The cause can be thought of as an hypothesis of the reasons behind the development of the patterns. As a hypothesis, it should be testable at least against some random process (effectively, the null hypothesis). Here is where statistical testing can be very effective in identifying the functional relationships between existing patterns and past processes. A similar approach could be taken in examining the effects of existing patterns on ongoing processes. In each case, the pattern is tested against some observable resulting process. For example, if the overall length of a hedgerow increases, a reasonable hypothesis is that there will be a measurable, predictable increase in the numbers of species that favor edges (edge species) (DeMers et al. 1996). The null hypothesis would be that such an increase in edge species will not result. Again, this shows the utility of applying statistical testing to provide a working knowledge of functional relationships between pattern and process prior to creating a GIS model.

Of course, this does not negate the possibility of using the GIS itself to test these hypotheses. In fact, employing statistical testing within a GIS using a small subset of empirical thematic data is an effective way of using GIS as an inferential statistical testing tool. Once spatial relationships are established for a sample of the data and the confidence limits are known, you can then implement the model for the entire database, on the basis of the established relationships. This also provides you with some measure of the confidence limits for your GIS model itself. A now classic example of using raster GIS to implement a predictive statistical model is the timber breakage model used by Tomlin (1981). By employing a regression equation from within the Map Analysis Package, Tomlin extended the typical regression equation to a spatial domain, thus removing the necessity of performing the regression test prior to GIS implementation.

Although this brief discussion of statistical analysis does not enumerate all the possible ways in which such tests can be performed, it does point out the importance of establishing the functional relationships among patterns within or prior to GIS modeling. Additional tools such as logit modeling, sensitivity analysis, and autocorrelation can be used to identify and quantify functional relationships (e.g., Algarni 1996, Clark et al. 1993, Johnston 1992, Lowell 1991, Pereira and Itami 1991). The advent of a burgeoning set of spatial descriptors, especially those found in the landscape ecology, health care mapping, and spatial crime analysis literature, has prompted an equally growing need to identify and quantify the causes and results of these patterns. The quantitative measures of pattern are a necessary first step, but without linking them to causation, we are unable to build effective, real-world GIS models, whether they are designed to describe a situation or to predict new ones.

TYPES OF GIS MODELS

Introduction

A classification of GIS models, like a classification of anything, is based on a preselected set of criteria. There are many ways of classifying GIS models—so many that it can get very confusing. In fact, Berry (1987, 1997) has separated spatial models from cartographic models, whereas many authors do not. My purpose here is not to add confusion to your already complex modeling tasks by creating yet another set of classifications but rather to examine some basic terminology that is being used among GIS modelers so you can communicate effectively with them. Additionally, the classifications will provide a structure for the modeling tasks by describing substantially different ways of modeling based on selective purpose, varying methodologies, and often fundamentally different logics. It is important to note that although some of these classifications are unique, many cross over and intermix, resulting in a relative inability to create a classification hierarchy that was originally attempted by Tomlin (1990). This confusion can best be assuaged by determining the specific utility of each classification scheme and treating each in turn, as was suggested by Berry (1995). The following paragraphs use this approach, providing important considerations for modeling within each model class. As you read them, you should spend less time on the classification itself and more on the modeling tasks and attendant thought processes specific to each. I limit the classification of GIS models to three fundamental approaches: (1) purpose, (2) methodologies or techniques, and (3) logics. Again, remember that none of these classifications is entirely independent of any of the others.

Models Based on Purpose

If there were a single well-accepted hierarchy to GIS modeling tasks, it would most likely begin with a focus on the overall purpose for which the model is to be developed. As with most other classification methods, one based on purpose is neither discrete nor binary but rather demonstrates extremes in a continuum of possibilities. On the one extreme, we have models whose sole purpose is to describe. These are called *descriptive* models. At the other extreme are GIS models whose primary purpose is to prescribe best uses of existing land resources on the basis of evaluation of known or predicted circumstances. These are *prescriptive* models. These two seemingly discrete model types, although no less a continuum than others, seem to have become accepted as the two most basic types of GIS models within the literature. We begin with the one that is most often the more basic of the two—descriptive—and increase in complexity to prescriptive models.

Descriptive Models As the term implies, descriptive models are *passive,* primarily designed to provide a description of parts or all of a study area under examination. The description can be simple or complex, single or multitheme, preparatory to final modeling or a solution in itself. It is sometimes difficult to separate the term *descriptive GIS model* from other model types because the map itself often describes, in explicitly spatial terms, conditions as they exist, could exist, or should exist. The terminology depends more heavily on the purpose for which the final outcome map, graphic, or other output is to be used than on the output itself. Descriptive models, then, describe conditions as they exist. In other words, they most often answer the "what is" question, rather than the "what should be" question. In some cases, the descriptive model may describe conditions that might also fit selective uses of the land. In such cases, the descriptive model might be answering the "what could be" question by simply describing the conditional fit rather than suggesting actual use.

In their simplest forms, descriptive models attempt to quantify an existing map or set of maps on the basis of any of the functional operations we examined in Chapter 4. In the first case, this type of model attempts to describe the geometry of the components map or maps. This geometry can range from simple measures of length, width, perimeter, area, circuitry, and many more, to more complex, integrative measures such as perimeter-to-area ratios, nearest-neighbor metrics, isolation, and other more topological measures inherent in the cartographic document. Many of these were described earlier in this chapter. The importance of describing or quantifying the geometry of a map is that it allows us an opportunity to isolate patterns within the overall map; to discover patterns that may not, without quantification, be visible;

or to compare patterns from one map to another or from one part of the map to another. Because each pattern is a result of one or more underlying processes, the description of pattern also helps us gain insights into these processes. Dispersion patterns, for example, are often used to relate pattern with process. Clustered dispersion patterns are often a result of nonrandom processes, as are uniform dispersion patterns as one might find in an orchard or row crops, whereas random dispersion patterns are most often a result of stochastic or statistically random processes. These patterns and their associated processes may also be related to other patterns on other themes. And the processes that formed the additional patterns may also be related to the original patterns.

Among the more powerful capabilities of the descriptive GIS model is its ability to go beyond descriptions of geometry to integrate or synthesize often seemingly disparate spatial data. In this context, the descriptive model could also be called a *synthetic* GIS model not because it attempts to describe a situation by examining a single map element or even a single map but because it often merges multiple themes to evaluate possible spatial relationships. In the synthetic approach, successive themes are combined, one at a time, to determine the degree of spatial association each might have in an overall description of existing conditions. Descriptive models are a mainstay within the scientific community. Scientists' training encourages them to follow a pattern of behavior, called the scientific method, that begins by first making observations of patterns, the cause of which will later be hypothesized about, will be rigorously tested, and may evolve into theory or, if proven immutable, become scientific law. Although the GIS is not, in its current form, a particularly good tool for testing hypotheses, it is a very adaptable tool for creating testable spatial hypotheses (Aspinall 1994).

An alternative to the synthetic type of descriptive model is the *deconstructive* type. For determining the sensitivity of certain factors in a descriptive model, it might prove useful to remove each one at a time, examining the final result as each part is removed. This is similar to a stepwise backward approach to regression modeling as opposed to stepwise forward. In stepwise backward regression, one attempts to remove individual independent variables to ascertain the impact each has on the final regression coefficient of the model. Although descriptive GIS models do not currently have a coefficient as a final result, it is still possible to at least see whether particular spatial variables correspond spatially by using a simple test of spatial correspondence (Muerhcke and Muerhcke 1999). And, of course, as with statistical correlation and regression analysis, the mere existence of spatial correspondence does not prove causation. It shows simply that the selected variables occupy some portion of the same geographic space. Such spatial correspondences can, however, be highly suggestive of such causation if carefully chosen (Sauer 1925).

Prescriptive Models At the other extreme of our classification continuum is the more *active* prescriptive model. In its purest form, the prescriptive model is designed to impose a best solution for problems in which a description of existing conditions is insufficient as a decision aid (Tomlin 1991). Just as a physician would first describe symptoms of a disease or other medical condition and then, on categorizing these symptoms (assigning a name, usually), would take of the next step of prescribing the best medicine or treatment to cure the problem, our next step would be to "prescribe" the best solution to geographic problems. In GIS, such scenarios might more aptly be applied to answer such questions as: (1) What is the best location in which to site a factory? (2) Where is the most likely location for finding a serial killer? (3) What is the most likely place to reintroduce aplomado falcons in the southwestern United States? In short, the prescriptive model is more closely associated with answering the "what should be" type of question.

As with any prescription, there isn't always a perfect solution for a given question. In this case, there are generally two approaches. One is to select a best solution on the basis of the best available data and the constaints that currently exist or are expected to exist (if the model is to be *predictive*). This first type is most often used when the individual constraints driving the model can be or are limited to Boolean conditions (i.e., goodsoils versus badsoils or goodzoning versus badzoning). These models are actually pretty rare and are most often implemented in the absence of specific factors with a range of conditions. The second approach is similar, except that it provides a range of possible solutions, some better fitting certain criteria than others. This approach is best applied when more information is known about the conditions of each included factor. Because there is a range of possible factor ratings and weightings, there is a wider range of factor sensitivity and thus a greater opportunity for effective, if not optimal, solutions. In this way, if some unforseen economic or political power should preclude the use of the best site, others would be available for use.

Among the most important characteristics of prescriptive models is their ability to derive a solution, not just to describe what is already there. As such, the prescriptive model tends to be more adept at prediction. Stated differently, if you have a *predictive* GIS model, it is more likely to be a prescriptive one than a descriptive one. But this does not preclude descriptive GIS models from having at least some predictive qualities. It is also important to understand that not all prescriptive models are predictive. Generally, for a model to be effectively predictive (prescriptive), it is very important that the processes that link the themes are explicitly and very completely understood. Such models also typically contain some dynamic elements and may require special database structures (e.g., the cellular automata) or even special computer processors for more complex types (Costanza and Maxwell 1991). Perhaps the classic examples of *dynamic* predictive maps are those that include dispersion or movements of ideas, creatures, or processes. Fire modeling is among the best known and most obvious of such predictive models.

In his original design, Tomlin separated prescriptive models into two types: *holistic* versus *atomistic*. Holistic models are those that evaluate a scenario in its entirety. They require a complete, overall understanding of both the processes and the thematic contents of the maps. Such models are rare, partly because there are few situations in which the overall complexities of systems are fully understood, and partly because they are very difficult to verify and validate. The more common type of prescriptive model is the atomistic type that breaks its processes and its themes into categorical and functional groups. By its very nature, it is the kind of prescriptive GIS model that readily lends itself to compartmentalization. It proceeds step by step, isolating individual elements as it proceeds. Because of that, it is far easier to conceptualize, to formulate, to flowchart, to implement, and to verify and validate.

In turn, atomistic prescriptive GIS models can be broken down into two additional categories: *heuristic* and *algorithmic*. Heuristic model types require either called-on experience or seat-of-the-pants experience. These types of experiential knowledge are often poorly documented, rarely formalized, and very difficult to obtain. Earlier, we talked about the use of knowledge acquisition strateties for obtaining GIS modeling information. This is the classic type of GIS model requiring this knowledge, necessitating unique knowledge acquisition strategies. Actually, on acquiring these heuristics, one most often needs to formalize them into a more explicit recipe for decision making. In other words, heuristic models are actually easier to solve if they are transformed in to *algorithmic* models.

Algorithmic prescriptive models most often take the form of a set of rules that explicitly relate elements of each thematic map. They relate these elements and these maps in a specific order or sequence, often in a hierarchical fashion, that both is representative of the real-world processes and allows for reversal of the process for model verification. In fact, these properties nearly define what a GIS model is—an ordered set of map operations designed to represent real-world settings.

Models Based on Methodology

As with virtually all other model types, the methodology of cartographic models is either *stochastic* (based on statistical probabilities) or *deterministic* (based on known functional linkages and interactions). Stochastic models are linked to the statistical criteria most often used in non-GIS models. For example, models based on statistical measures of central tendency are of necessity driven by the central limit theorem. An extension of this theorem for predictive modeling is regression analysis. One classic example of the use of regression modeling is Tomlin's (1981) predictive GIS model of timber breakage during harvest. The model employs a regression equation on a cell-by-cell basis. This is effectively a spatial regression model. Another example of effective stochastic modeling is the use of logistic regression (categorical regression) to predict the presence or absence of creatures in an underlying environment. Models like this include those examining squirrels (Pereira and Itami 1991), bears (Agee 1989, Clark et al. 1993), desert longhorn sheep (Dunn 1996), deer (Chang et al. 1995), and birds (Miller et al. 1989), each of which exemplifies how statistical techniques can be linked with other GIS functionalities.

Although stochastic GIS models assume things are distributed on the basis of statistical likelihood, deterministic models assume direct functional linkages. Models such as those involving hydrological flow prediction (Chase 1991), pollution evaluation (Gros et al. 1988, Haddock and Jankowski 1993), and soil loss modeling using the universal soil loss equation (Battad 1993) are good examples of how a knowledge of the environment can be modeled with deterministic methods. A primary consideration for these and all deterministic models is that a well-defined cause-and-effect relationship exists and can be identified.

Models Based on Logic

We have examined how GIS models can be based on purpose as well as on the methodology applied to their creation, but the method of logic applied in the conceptualization and formulation of the model is equally fundamental. There are two primary forms of logic that are traditionally employed: *inductive* and *deductive*. The inductive method attempts to build general models based on individual data or instances. For example, by collecting or obtaining data and information on mountain lion habitat use at a number of individual locations, and by summarizing those localized conditions, one can begin to create a general model of mountain lion habitat use in a region. In short, the inductive approach moves from specific elements or instances to a general model, usually employs many *empirical* tests to gauge the viability of each factor, and typically uses a trial and error approach.

This approach is often useful if we are unaware of the general conditions or rules under which our subject or subjects operate. In some circumstances, especially within a data-rich environment, many previously unknown and important factor interactions can be identified by such a *data mining* aproach. To some, the inductive method can be a bit disconcerting because it seems to eliminate the hypothesis testing we are so familiar with in the scientific method. This not true, however, because each of the factors employed is tested. Its advantage over the deductive approach is that it is far less common that we know how all the processes function than is practical for many modeling environments.

But models that build on deductive logic are straightforward, easier to understand, and far more purely algorithmic (atomistic) than inductive models. Deductive logic moves from general to specific. In GIS modeling, what this implies is that we have a substantial preliminary knowledge of what factors are important, how they interact, and which are most important before we conceptualize, formulate, flowchart, and even implement the model. The best candidates for such modeling are those that already have a somewhat formalized set of criteria, weights assigned to each, thematic map data that represent them, and standards for how they are combined, algorithmic models such as the simple additive LESA (Land Evaluation and Site Assessment) model (Williams 1985), the statistically based timber breakage harvest model (Tomlin 1981), and models employing the universal soil loss equation (Battad 1993). GIS models that employ existing aspatial mathematical or statistical models and add the spatial dimension through the use of a raster-based GIS are some of the more obvious examples. Some of these can become quite complex. Some even require the implementation of parallel computer processors to complete their calculations (Costanza and Maxwell 1991), yet they still remain more easily understood than many less algorithmic inductive model types. In fact, deductive models—even very complex deductive models—are easier to explain to clients and are therefore much easier to verify and to validate than are inductive GIS models.

Chapter Review

An essential skill for geographic information system (GIS) modeling is the ability to recognize, identify, and interpret geographic patterns. In essence, it requires spatial thinking. The GIS model may be used to evaluate, describe, or combine either visible patterns (those that are fairly obvious) or functional patterns (those that may be visible only through special means of sampling or unique tools of observation. The primary tools for identifying patterns include visits to the landscape, examination of other research through literature review, interviews and other knowledge engineering techniques, examination of maps, aerial photography, satellite imagery, and descriptive and predictive statistical techniques for relating dependent and independent spatial variables. Statistical techniques, frequently combined with GIS analysis itself, can often link the quantified geometry of individual or multiple thematic maps with functional processes that might be responsible for them. These hypotheses can be an end in themselves or they may be used as the starting point for a GIS model.

There are many kinds of GIS models, depending on how they are classified. On a simple level, the model types are divided into three classification schemes. The first is based on purpose and includes descriptive models and prescriptive models. Although these two types of models are actually part of a continuum, descriptive models are primarily designed to answer the "what is" question, whereas prescriptive GIS models most often answer the "what should be" type of question. The second GIS model classification scheme is based on methodology that includes stochastic (statistical) versus deterministic models, where cause-and-effect relationships can be effectively identified. The final class of GIS model types is based on the type of logic used to implement the model. In this case, some models are inductive, in that they try to make generalizations based on subsets or samples of the population of the data, or deductive, where knowledge of the overall situation can be used to predict individual conditions.

Discussion Topics

1. The U.S. Forest Service is building a generalizable predictive geographic information system (GIS) model of forest fire potential based on factors such as forest use,

fuel types and buildup, locations of campgrounds, surveys of visitors, and many others. Aditionally, the model is also designed to predict the spread of fire on the basis of wind speeds and directions, topography, humidity, and many other factors. Discuss the types of GIS models we have just examined and categorize this model in general and its component parts in particular.

2. A new acquaintance does not understand what the problem is with creating GIS models. After all, it is just like using any other piece of software. Once you learn which buttons to push, the rest is just rote. Provide a modeling scenario in which your new acquaintance will be forced to think in explicitly spatial terms. The purpose of the exercise is to make you think spatially to develop the scenario and for your friend to do so to discuss its solution.
3. Discuss the role of repertory grid methodologies in acquiring knowledge about the relationships between thematic map representations of spatial data and the functional operators. What other knowledge acquisition techniques are available for this same task? What are the advantages and disadvantages of each?
4. Consider the following adjectives that can be applied easily to different types of GIS models and discuss how they would fit within the three basic classifications employed in this chapter:
   a. Predictive
   b. Simulation
   c. Spatiotemporal (dynamic)
   d. Land capability
   e. Land suitability
5. For at least three different knowledge domains, provide some appropriate visualization techniques, statistical approaches, and literature sources to provide the basis for spatial understanding. Example knowledge domains might include criminal justice, defense, land planning, atmospheric modeling, habitat evaluation, site selection, health care provision, real estate, and insurance.

Learning Activities

1. In your nearby community, go for an undirected road trip, carrying with you a camera (even one of the disposable ones will work), with at least 24 frames of print film available. Your mission is to photograph your environment as you observe it firsthand, in particular to observe and document patterns. After developing your film, provide a 3″ × 5″ index card for each photograph describing the nature of the environment you have photographed and the patterns you observe. Create a poster of your photographs and index cards that you will share with your other classmates.
2. From your photographs and those of your classmates, create a table similar to Table 5.1 that shows the spatial dimension, object, character, measure, possible cause, and effect for each of the objects you identified.
3. Collect 10 to 20 GIS modeling articles at random from a range of journals in the literature (especially the professional journals) without selecting them by type (descriptive versus prescriptive), by methodology, or by logic. Create a table with five columns. The first column should be an abbreviation of the article title, the

second column should be the name of the journal, and the three remaining columns should be model categories (purpose, methodology, logic). Assign each article to the first and second columns. For the model types columns, classify each article and place that information in the appropriate column. From this, suggest which journals seem to specialize in particular GIS model types. Compare your results with those of other students in the class. Does this suggest where to obtain examples of particular types of GIS models?