

UC Santa Barbara

NCGIA Technical Reports

Title

Environmental modeling with GIS: A strategy for dealing with spatial continuity (93-3)

Permalink

<https://escholarship.org/uc/item/06n4f3rm>

Author

Kemp, Karen K.

Publication Date

1993

NCGIA

National Center for Geographic Information and Analysis

Environmental Modeling with GIS: A Strategy for Dealing with Spatial Continuity

by

Karen K. Kemp

University of California, Santa Barbara

Completed as a PhD Dissertation

Department of Geography, University of California, Santa Barbara

Copyright by Karen Kathleen Kemp 1992

Technical Report 93-3

May 1993

Simonett Center for Spatial Analysis University of California
35 10 Phelps Hall, Santa Barbara, CA 93106-4060
Office (805) 893-8224; Fax (805) 893-8617
ncgia@ncgia.ucsb.edu

State University of New York
301 Wilkeson Quad, Box 610023, Buffalo NY 14261-0001
Office (716) 645-2545; Fax (716) 645-5957
ncgia@ubvms.cc.buffalo.edu

University of Maine
348 Boardman Hall, Orono ME 04469-5711
Office (207) 581-2149; Fax (207) 581-2206
ncgia@spatial.maine.edu

ABSTRACT

Environmental modeling with GIS:
A strategy for dealing with spatial continuity

by

Karen Kathleen Kemp

Linking a GIS to a spatially distributed, physically-based environmental model offers many advantages. However, the implementation of such linkages is generally problematic. Many problems arise because the relationship between the reality being represented by the mathematical model and the data model used to organize the spatial data in the GIS has not been rigorously defined. In particular, while many environmental models are based on theories that assume continuity and incorporate physical fields as independent variables, current GISs can only represent continuous phenomena in a variety of discrete data models. This document develops and outlines a strategy in which *field variables* are used to enable modelers to work directly with the spatial data as spatially continuous phenomena. Field variables are declared like other data types in standard computing languages. Specifications of field variables include several special properties which are used to express the relationship between the physical field and the discrete data model. These properties determine how the spatial data can be manipulated. This allows the manner in which the spatial data has been discretized and the ways in which it can be manipulated to be treated independently from the conceptual modeling of physical processes.

Several outcomes from the use of this strategy are explored. Modelers can express their spatial data needs as representations of reality, rather than as elements of a GIS database, and a GIS-independent language for model development results. By providing a formal linkage between the various models of spatial phenomena, a mechanism is created for the explicit expression of transformation rules between the models of spatial data stored and manipulated by GIS. The incorporation of field variables allows several operations (such as determining integrals, slope and aspect) and reserved variables (such as latitude and longitude) which are commonly used in environmental models to be defined. While scalar fields are the focus of this document, consideration of the potential for the definition of vector fields and related operations (such as divergence and gradient) using this strategy is also included.

ACKNOWLEDGMENTS

There are many people I would like to recognize for their contributions to my successful completion of this work. Primarily, I thank Dr. Michael Goodchild, whose inspiration, guidance, perspective and patience has been ultimately responsible for the initiation and completion of this research. The other members of my committee, Drs. Tobler, Smith and Haymon, have also provided valuable assistance. I must also thank the National Science Foundation for their support of the National Center for Geographic Information and Analysis (NCGIA), through grant SES 88-10917, which supported this research and the other professional activities I undertook during my years at the University of California at Santa Barbara (UCSB). Less direct, but equally critical, help has been given to me by my many friends and colleagues at UCSB and elsewhere. While I cannot thank them all, I would like to recognize a few individuals: Mark Kumler, with whom the experience of traveling the Ph.D. road together has bonded us as life-long colleagues and friends; Kelly Cayocca, whose friendship helped me through some of the most emotionally trying parts of this effort; Luc Anselin and Emily Talen, whose loan of an office chair came at a most critical time; my mother and father, Don and Gwen Kemp, whose relief at seeing me finally complete my formal education has inspired me; and, all the faculty and graduate students at the Department of Geography, who have provided a continuing intellectual and recreational challenge. And finally, I must thank the late Dr. David Simonett who challenged a generation of Geography graduate students and whose untiring commitment to ideals was largely responsible for the creation of the superb Geography department and graduate program at UCSB and for the presence of the NCGIA there.

TABLE OF CONTENTS

| | |
|--|--|
| CHAPTER 1 - INTRODUCTION | |
| A. Environmental modeling for the 90's | |
| B. Continuity in environmental modeling | |
| C. Related research | |
| D. Overview | |
| CHAPTER 2 - MATHEMATICAL MODELING | |
| A. Models and modeling defined | |
| B. Theories of models and modeling | |
| C. Model building | |
| D. The role of data in models | |
| CHAPTER 3 - ENVIRONMENTAL MODELING | |
| A. Environmental modeling Domains | |
| B. The mathematics of environmental modeling | |
| C. Spatial data for environmental modeling | |
| D. The role of GIS in environmental modeling | |
| CHAPTER 4 - MEASURING AND REPRESENTING FIELDS | |
| A. Measuring and representing value and location | |
| B. Representing fields | |
| C. Spatial data models for fields | |
| D. Modeling with continuous variables | |
| E. Field variables | |
| CHAPTER 5 - OPERATIONS ON FIELD VARIABLES | |
| A. GIS operations | |
| B. Mathematics on fields | |
| C. Converting field data models | |
| D. Arithmetic operations | |
| E. Operations involving different spatial data models | |
| F. Descriptive statistics for fields and field variables | |
| G. Inferential statistics | |
| CHAPTER 6 - SPECIFYING FIELD VARIABLES | |
| A. Conventions and assumptions | |
| B. Declaring field variables | |
| C. Properties of field variables | |
| D. Defining special components | |
| E. Spatial equality and nesting | |
| F. Neighborhoods | |
| G. New operations and variables | |
| H. GIS operations for field variables | |
| I. Examples of the use of the strategy | |

CHAPTER 7 - VECTOR FIELDS
A. Operations on vector fields
B. Representing vector fields
C. Using vector fields in environmental models

CHAPTER 8 - IMPLEMENTING THE STRATEGY
A. An example - The soil water budget
B. Implementation issues

CHAPTER 9 - CONCLUSION
A. Summary of the research.....
B. Critique and analysis of results
C. Directions for the future

REFERENCES.....

CHAPTER 1 - INTRODUCTION

Environmental issues are among the most important facing decision-makers today. The dynamics of the hydrologic and atmospheric systems of the earth imply that all environmental systems are tightly interrelated, dynamically and spatially. Impacts in one location usually have effects in others. Spatial data, systems for managing that data and analytical techniques for converting that data into information are now vital tools in the assessment and management of a healthy natural environment.

Considerable progress is being made in integrating spatial information systems, GIS, and mathematical models of the environment (cf. Goodchild, et al, 1993). For most environmental modeling projects, GIS is seen as a convenient and well structured database for handling the large quantities of spatial data needed. Traditional GIS tools such as overlay and buffering are also important for developing derivative datasets that serve as proxies for unavailable variables. Many experts also expect that as better spatial analysis methods become incorporated into GIS, GIS will also become an important tool in all aspects of modeling, including model building, validation and operation. However, there are significant incompatibilities preventing true integration. GIS manages static and discrete data while environmental models deal with dynamic and continuous phenomena. GIS databases contain information on location, spatial distribution and spatial relationships while environmental models work on a basic currency of mass and energy transfer (Maidment, 1993). In order to fully integrate the two we need to add dynamics and continuity to our understanding of spatial data, and spatial interaction and functionality to the environmental models (Fedra, 1993).

This research seeks to address the first of these needs by considering the implications of working with continuous phenomena directly in the context of GIS. Following a consideration of how spatially distributed phenomena are characterized in environmental models, a strategy for working with these continuous phenomena so that they can be efficiently and conveniently incorporated into the discrete digital computer code of mathematical models is developed. Since most environmental models deal with phenomena which are continuous in space, this strategy provides a means by which modelers can work with the spatially continuous phenomena directly rather than being forced to work with individual discrete spatial elements. As Goodchild (1992) and Tobler (1989) have contended, we must recognize that digital spatial data are *representations* of reality, not reality itself and we should strive to achieve a "frame free analysis" (Tobler, 1989, p. 116), one which is not dependent upon the representations but upon reality. Therefore, the fundamental theses of this research are:

- 1) it is both desirable and possible to separate the mathematical operations which will be performed on data about spatially continuous phenomena from the form of spatial discretization used to represent those phenomena in the computer, and
- 2) this separation allows issues about the implementation and manipulation of these digital representations to be dealt with automatically, without external control, in such a way that they can be considered extraneous to the modeling task.

This chapter sets out the intellectual challenge of this research effort. It introduces the theme of environmental modeling and in particular that of the physically-based modeling of spatially-distributed environmental processes which is the focus of this research. We consider briefly the value and problems of these modeling efforts and discuss the possibility of a new spatial paradigm for environmental modeling. Continuity is a constant theme in this research and its role in environmental modeling is examined. Finally, related research efforts are reviewed so that this work can be placed in context.

A. ENVIRONMENTAL MODELING FOR THE 90'S

Mathematical models have been around since the development of mathematics by the Chinese, Greeks and Hindus, but modeling has progressed rapidly since the 1950's as a result of developments in computers and programming languages (Cross and Moscardini, 1985). While computer models became popular initially with scientists in the physical sciences who had well-developed formally defined theories to work with, the use of such models has now gradually shifted into biology, medicine, natural resources, urban development, and social and behavioral sciences (Jacoby and Kowalik, 1980).

While decades of less than satisfying experience with large-scale models have provided many sound arguments against dependence upon the outcome of models for decision making (Lee, 1973), governments at all levels are encouraging the use of mathematical models in research and planning. The anticipation is that such models will provide useful tools for dealing with the multitude of complex issues facing them. NASA, in a recent call for proposals, demonstrates the fervent anticipation of success from such modeling efforts:

It is the function of this research program to use the available satellite data; derive its essential information about the interacting environmental regimes, assimilate the derived quantitative information sets into initial, sequential, and boundary conditions for numerical forecast models; to run the coupled, multidisciplinary prediction models in likely scenarios that will simulate the consequences of stressing the global environment with perceived processes of global change. (NASA, 1991, Appendix A, p. 1)

Decision-makers increasingly are being asked to make legally justifiable decisions about complicated environmental issues (Breidenbach, 1976) and physically based models are particularly attractive for this purpose. Unlike earlier black-box models, these models attempt to describe real physical processes and interactions occurring across the landscape. Though he refers to the use of hydrologic models, Beven's comments are relevant to all environmental modeling areas:

Distributed [physically-based] models will be increasingly used in the future. The incentive to do so will be due to the fact that they are 'better' models in the sense of having a more rigorous theoretical basis.... There is an important danger of a subversion of hydrological reasoning by ill-conceived applications of distributed models, arising from the very fact that their theoretical rigor may lead to an uncritical belief in their predictions. (Beven, 1985, p. 432)

While models are, by definition, simplifications of reality, it is often assumed in management situations that models are precise and that all relevant factors are included. It is difficult to argue against the model conclusions without an intimate knowledge of the model itself (Freeman, 1973). Breidenbach suggests that in order to assist decision-makers, it is imperative that scientists begin working more closely with them:

Development of a consistent and understandable modeling language is becoming extremely important to the manager, the scientist, and the public... The person who developed the model probably knows precisely what assumptions are important. It is the user -- the decision-maker -- who will feel the effects of these assumptions... It is your job and responsibility as modelers to reduce the manager's risk. (Breidenbach, 1976, p. 4)

Decision-makers need to know what assumptions have been made and how sensitive the model is to changes in the variables. This will assist in the determination of what elements are important and what can be ignored. They also need confirmation of the validity of the model and information about what data were used, where they came from, their currency, which variables were measured and which were estimated. Both the model and its data must be available for scrutiny by non-specialists. A "universal language for communicating the structure and behavior of models to others" (Zeigler, 1976) is needed. This research begins to address this need.

Modelers, too, need help dealing with the new modeling opportunities. In spite of increasing popularity, there are many practical difficulties in developing spatially-distributed physically-based models. Besides the problems of creating stable numerical solutions to non-linear continuous equations and the need to create spatial discretizations for the iterative procedures, the acquisition of accurate and meaningful spatial data for these models is of prime importance in ensuring the validity of results. These models require massive quantities of physically-based data for calibration, processing and validation.

On the surface, it may appear that this problem is diminishing. Huge quantities of spatial data are currently being collected and stored. Many government agencies have taken on primary roles as data producers, compilers and archivists. For example, while the USGS mission has always been that of "collecting, analyzing and disseminating earth-science information" (USGS, 1991, p. 8), it now is responsible for coordinating all digital cartographic activities of the Federal Government (USGS, 1991). NASA and the remote sensing agencies of other countries are now accumulating formidable amounts of raw spatially-referenced data about the earth and its atmosphere. Local and state governments are compiling regional databases. And commercial organizations are preparing updated and niche collections of all types of spatial data (GIS World, 1991).

Unfortunately, quantity does not necessarily ensure that the data needs of physically-based environmental models are being met. In fact, as the models become more sophisticated and sensitive, the cry from the modelers for more data becomes louder. As Thornes points out "the demand made by these models quickly outstrips the capacity to supply them with data, even where national or international agencies are involved" (Thornes, 1989, p. 12). Clearly, from the modelers' viewpoint, the available data does not match the needs of the modelers. Modelers are being forced to invent creative ways to transform the data they can get into the data they need. And since modeling is a creative activity, it is unlikely that we will ever be able to forecast accurately the data modelers will need.

Each new data source encourages increased modeling efforts, more demands for new data collections and the possibility of misuse of available data. In the early days of the development of digital mathematical models, scientists had to deal directly with their data; collecting, verifying and interpolating it. Now copious amounts of data are available at arms length and "the cult of information presumes that all information is good" (Batty, 1989, p. 166). While there may be fewer opportunities for simple typographical mistakes, the possibility of misuse and misunderstanding of data quality and suitability is greatly increased (Rumble, 1984). Interfaces between models and spatial databases generally tend to use a brute force method of transformation in which source data is converted with little regard for

the spatial consequences of the manipulation. The development of a sound, theoretical basis for the use of spatial data in environmental models is needed.

Fortunately, the groundwork for this effort has been well started. Decades of work by geographers passing through the quantitative revolution have provided us with many concepts and tools for handling spatial data in numerical operations (cf. Getis and Boots, 1978). Since the quantitative revolution, geographers have tended to study not the phenomenon itself, but its spatial characteristics and abstract properties related to topology and geometry (Haggett and Chorley, 1967). The topological-geometrical framework of reality relates directly to the data models now available in GIS. The soundness of these principles can be seen in their frequent rediscovery by numerous environmental modeling teams as they construct individual interfaces between reality and spatial data for their specific projects.

A paradigm change for environmental modelers

Proper integration of geographic knowledge into these modeling efforts requires a form of paradigm change for environmental modelers. We can see Harvey's 1969 statement (not made in reference to geographic information systems as a computer technology) in this newest sense:

The current issue is not one of revolutionizing geographic information systems - it is, rather, one of examining the underlying logic of any geographic information system, and formalizing the method whereby geographers [and now others] may spatially order the mass of information available to them. (Harvey, 1969, p. 214)

The spatial paradigm for environmental modeling is just beginning. Opportunities offered by spatial data and functionality through GIS will lead the sciences into new realizations, new "patterns of searching the real world" (Haggett and Chorley, 1967, p. 26). On paradigm change, Harvey comments that "Experiences that seemed irrelevant now seem surprising and demand explanation. This general change in expectations involves a shift in the scientist's perceptions of the world around him" (Harvey, 1969, p. 18). Insight will come from the consideration of spatial effects, new data will be collected or generated and theories will be revised and extended.

Although a paradigm shift does amount to "a matter of judgment, an act of subjective choice, an act of faith", it must be "backed up by substantive evidence from logic and experiment" (Harvey, 1969, p. 17). Scientists in many different fields are investigating the significance of the accuracy of spatial databases, effects of scale change, relative efficiencies of data models, etc. They need formal methods for incorporating spatial effects with known confidence into their models. This research draws on basic geographic principles to provide a sound theoretical structure for the integration of environmental models and GIS.

There is another perspective to this paradigm shift taking place in the environmental sciences. As the models move gradually away from direct descriptions of physical laws to complex models with uncertain spatial parameters, effects and inputs, environmental modelers will find themselves working with models more similar to those of the soft-sciences like economics and behavioral science in which relationships are ill-defined and full system description impossible. The difficulty of validating these new spatially distributed models often means that "truth" must be found by means other than by reference to universal physical laws (Beven, 1987). New tools to help environmental modelers examine and confirm the validity of their models are needed.

Paraphrasing Casti (1989), when he spoke of issues to be addressed in developing a theory of modeling, there are several interesting questions that can be explored as the role of spatial data in models is formalized. How many different ways can we measure space? How can we represent space in a formal system? What is the relationship between space and its formal representation? How can we compare two representations of the same space? How do these models relate? What is the role of space in the mathematical model? How do the models of space relate to the mathematical models? Can two models of space be interchanged in mathematical models? What procedures can we invoke that can be used to choose the appropriate model of space? Such questions provide the intellectual challenge to this effort.

B. CONTINUITY IN ENVIRONMENTAL MODELING

Since physically based environmental models depend upon physical principles, the mathematics of these models is often in the form of differential equations. These equations implicitly recognize the continuity of space and the constantly changing values of the independent variables. The challenge for scientists and GIS analysts working with environmental models is to transform this continuity into the discrete world of the computer.

Discrete representations for both continuous equations and the continuity of space have been devised and are widely used. Finite difference numerical solutions to differential equations discretize time and space into small units. These stepped algebraic solutions for the governing differential equations are calculated for each time and space unit and a final solution is achieved by "integrating" (generally through simple addition) the results across the entire study area and time period. Finite element solutions divide the study area into units which are homogeneous in ways that allow some terms of the governing equations to be simplified. Analytical solutions can then be determined for each element and the total solution is determined through simultaneous solution of a set of equations. Alternatively, some of the global climate models achieve discretization by spectral analysis. In this case, instead of discretizing space, the response spectrum itself is dissected into a set of ordinary differential equations for which solutions can be found (Bourke, 1988).

Just as the equations themselves in these mathematical models are continuous, so are most of the phenomena being described. For example, air temperature, soil infiltration rate and solar radiation are continuous physical fields. Since we cannot measure continuous phenomena everywhere, it is necessary to develop techniques for gathering information about fields by collecting data at a finite number of points. Likewise, we need techniques for representing continuity with this finite collection of data. A few environmental variables, particularly those in the biological sciences which deal with individuals like trees or animals, are not continuous in the strict sense. Conceptually, such phenomena can be converted to fields by taking the limit of the value of the phenomenon (in this case the count or the frequency) divided by the area as the area tends to zero, "stopping short in the usual way before molecular lumpiness manifests itself" (Shercliff, 1977, pp. 11-12). This process results in a continuous density surface. Models using such variables calculate the rate of change of the density. In the urban environment, Angel and Hyman (1976) used this continuous conceptualization of discrete phenomena to develop a continuous model of transportation systems. However, since molecular lumpiness sets in at the scale of the phenomenon itself (i.e. the individual being counted), the measurement (count) must be made over a defined area, otherwise the phenomenon would yield a binary presence/absence value only. Density estimation cannot be taken to or even near this binary limit. Thus the value of density is dependent upon the unit over which it is measured (e.g. 100 over each 10 km² versus 10 over each km²) and all density measurements have an implied scale. Since the

value of such a field changes as the area over which observation occurs changes, density fields are not true physical fields and cannot be modeled in an entirely similar fashion.

Continuity, of course, also exists in time. Like space, time is difficult to discretize. It is common in many different areas of mathematical modeling to discretize time as either a series of instantaneous snapshots, such as daily noon temperature, or as a series of averages over a time slice, such as average monthly temperature. Since continuity through time can be as important as continuity through space, many researchers are working on new temporal data models for computation (cf. Langran, 1989). However, for the purposes of this particular investigation, it is assumed that the continuity of time is adequately represented by either of these traditional temporal discretizations.

Finally, continuity arises in measurement. Many phenomena in the natural environment are measured on continuous scales. Temperature, solar radiation and precipitation can be measured to as many decimal places as the measuring instrument allows. Continuity in many dimensions is a fundamental characteristic of all natural systems and it must be scrupulously and explicitly addressed whenever the natural environment is transformed into a digital representation. This research considers continuity in the spatial dimension.

C. RELATED RESEARCH

This section reviews some of the related research being carried out by others who are seeking to find solutions for the integration of GIS and the modeling of natural processes.

Map algebra and cartographic modeling

Map algebra is perhaps the first significant attempt at achieving this separation of operations from consideration of the form of the spatial discretization. Here this is accomplished by allowing only a single discretization. Beginning with the requirement that all spatial data to be manipulated is stored in co-registered sets of gridded data, Tomlin's map algebra provides an organizing framework and a vocabulary for expressing many standard GIS operations (Tomlin, 1990).

Founded in the map overlay tradition of landscape architecture, map algebra provides a computer language which automates many of the concepts of cartographic modeling, a technique for land planning made popular by McHarg (1969). Tomlin maintains that "cartographic modeling is a general methodology for the analysis and synthesis of geographical data" (Tomlin, 1993, p. 361). In cartographic modeling each different theme in the landscape is expressed as a separate map or layer. Manually, cartographic modeling was accomplished by physically overlaying several transparent maps of different themes in order that composite characteristics at individual locations could be observed. Map analysis provides a means to implement and extend these manual overlay techniques with the computer. Chan and Tomlin note that "At the heart of cartographic modelling is the idea of decomposing data sets, data processing-capabilities and data-processing control specifications into elementary components that can then be recombined with ease and flexibility". (Chan and Tomlin, 1991, p. 351). Tomlin recognizes that the cartographic modeling approach is "function-oriented" since the language is concerned only with the functions that are performed on the data rather than with the objects that are represented by it. The form in which data is stored and the functions specified do not purport to represent reality, but rather to facilitate the manipulation of digital data.

This reduction to elementary components is achieved by requiring that each element of the landscape which is to be examined using map algebra must be discretized as a continuous layer of rectangular cells. This greatly facilitates the implementation of the analysis functions since mention of the spatial discretization can be ignored and data processing can be specified for this particular form of discretization only. However, as a result, all operations are conceptualized on the grid rather than on the fields themselves. For example, distance from a point must be conceptualized in terms of "rook's case" or "queen's case" adjacencies or an nxn window rather than by the linear distance determined by the physical process being modeled. Thus the modeler must work within a gridded universe rather than with the continuous one described by the physical laws.

Is a gridded universe the only one an environmental modeler needs? It may be possible to convert all datasets to grids, but is this the best approach? Certainly not if one considers the reality that is represented. Certain phenomena may be better represented by other forms. Unnecessary conversion from one form of representation to another leads to an unacceptable loss of information if the mathematics does not demand a gridded structure. All movement must be discretized to 4 or 8 directions, paths may become unnecessarily distorted. By reducing all fields to a single gridded data model in which values are constant across cells but change suddenly at cell boundaries, it is difficult to conceptualize and work with continuous functions such as differentiation and integration.

Map algebra requires us to enforce a structure on reality rather than allowing reality to suggest a more appropriate structure for our analysis. The modeler is forced to work within this single discretization and to be constantly aware of its limitations. Unless the grid size can be allowed to shrink towards 0, it is impossible to conceptualize continuous natural processes operating across continuously varying surfaces within this gridded world. Rather than being a true modeling algebra, it is instead a unified nomenclature for formally defined GIS operations on gridded data. The weakness of the link with reality is critical.

In spite of its limitations, the map algebra approach has achieved wide popularity. Most notably, it has been chosen as a framework for the modeling language included in ESRI's raster-based product GRID (ESRI, 1991) and is the basis for the GIS software, MapBox (Decision Images, undated). Once the common data model is accepted and achieved, it is possible to express many mathematical operations between different datasets representing environmental variables in a structured syntax. For example, if we wish to add two themes, call them FIRST and SECOND, together, Tomlin's 1990 syntax would be (Tomlin, 1990):

RESULT = LocalSUM of FIRST and SECOND

though in the implementation of map algebra in ESRI's GRID software the more direct syntax:

RESULT = FIRST + SECOND

is employed. Using such a common syntax for basic algebraic operations, it is possible to conceptualize the operations distinct from the implementation details.

Model/data interfaces

An altogether different approach to the problem of linking digital data about spatially distributed phenomena to mathematical models of environmental processes drives the many attempts that have been made to develop generic interfaces between spatial information systems and applications programs (cf. Abel and Wilson, 1990, Evans et al, 1992). In

general, these interface systems allow data to be stored in several different files and formats in information systems that are logically and often even physically remote from the mathematical model. When the mathematical model is executed, required data is requested automatically from the interface. The interface then accesses the appropriate database(s), extracts the required dataset(s) and converts it (them) automatically to whatever format is required by the model. Once in a form suitable for input, the interface forwards the data to the model. Output from the model is handled similarly but in reverse.

Breunig and Perkhoff describe the development of one such interface system designed to mediate between several heterogeneous databases and different task-oriented applications programs. In this case, "the goal of logical data integration is the transparent use of data managed in different autonomous database systems" (Breunig and Perkhoff, 1992, p. 273). Using a client server structure, the Data Integration System (DIS) sits between the applications programs and the databases. Users define and use application dependent views of the databases in such a way that "the structure and location of the data and the different data handling languages of the external systems" (Breunig and Perkhoff, 1992, p. 273) are irrelevant to the user.

As it is currently implemented the DIS provides a Functional Global Data model (FGD) that enables schema integration between INGRESS and the DASDBS Geokernal databases and provides a high level of data abstraction. The FGD is a functional language (as opposed to an object oriented language) which is strongly typed, containing type variables and constructors for defining new types, and provides a small set of primitive functions" (Breunig and Perkhoff, 1992, p. 275). A distinct advantage of this system is that standard algorithms which are specific to certain representations of reality, such as distance or aspect, can be retained within the databases rather than written specifically for each application.

These data integration systems are designed to serve the data exchange needs of specific software. Rather than allowing system independent manipulation of spatial data, they provide tools for the conversion of datasets from one format to another. They do not provide a means by which modelers can express and manipulate the representations of spatially distributed phenomena directly outside of the constraints of specific software.

Database oriented solutions

Other approaches to dealing with the problems of representing the spatial complexity of reality work from the perspective of the data rather than from formalization of operations as provided by map algebra or from the development of interface programs between the database and the model or application. These approaches seek to find a means for rationalizing and organizing the various ways spatial data can be conceptualized and represented. Here, the data itself becomes the focus of scrutiny rather than the operations to be performed.

In the 70's, IBM developed a research prototype for a comprehensive database/application system which was designed to manipulate both standard and geographic data within the same framework.

Geographic data support allows the system to handle information extracted from maps in the same way as the more conventional string and numeric data. This is achieved by the provision of a geographic data type and an appropriate set of geographic operators and functions. (Aldred and Smedley, 1974).

Using a relational database design and representing geographic data as points, lines and areas, this prototype was intended to provide a comprehensive urban management

system. While the prototype is described in several internal IBM documents, it was never fully implemented. It seems likely that advances in technology and computing science overtook this ambitious project. Though incomplete, this project did attempt to address the issue of whether geographic data as a fundamental and unique data type could be handled concurrently with more traditional data types.

More recently, the concepts of object orientation have been seen to provide a possible means by which related but individually unique classes of complex spatial objects, with specialized operations inherited and encapsulated, can be defined for incorporation into computational models. Smith outlines the elements of a logic based modeling and database language which has as a long term goal the unification of computational modeling and database systems of many areas of scientific research (Smith, 1992). By developing a formal language for defining and manipulating complex spatio-temporal entities at all levels of abstraction, "from simple datasets to large, mathematical models", Smith hopes to be able to derive tools that will enable scientists to "focus on scientific rather than on computational issues" (Smith, 1992, p. 592). At the core of his research is the "Term Definition Language", a metalanguage that will allow complex spatio-temporal entities, their properties and relationships to be defined and manipulated independent of specific programming and spatial database implementations. Data objects can be conceptualized and structured according to the requirements of any single scientist or modeling task.

When Smith's work is complete and scientists have access to the multitude of distributed large heterogeneous spatial databases via this logic-based database and modeling language, modeling will become transformed and the full benefit of object orientation in mathematical modeling may be realized. In the meantime, the research described in this document seeks to work in the middle ground between function oriented and data oriented approaches to integrating mathematical modeling with GIS. Rather than seeking ways to restructure the way spatial data is modeled or ways to redefine the operations that modelers seek to apply to the data, this research is focused on tools for implementing the current models of natural phenomena using existing spatial data models and GISs. The result is a new way of looking at old tools.

Formal languages and algebras

Formal languages and algebras are important tools for developing and expressing abstract concepts such as those which are developed here. Formal languages consist of sets of symbols, operations and rules about how to combine the operations and symbols. Algebras are a subset of formal languages in which properties of the operations, such as symmetric, transitive and reflexive, are important (Egenhofer, 1992, pers. comm.). Such properties allow the identification of minimal sets of operations and permit the simplification of complex tasks.

Efforts to develop formal languages for working with geographic data have been underway for several decades. In the late '60's/early 70's Dacey outlined some concepts for the development of such a formal language (Dacey, 1967, 1970 and 1971). Though little of this fundamental work is easily accessible, much of it published only in technical reports, many of his contemporaries did recognize the significance of this work. More recently, the call for the development of a formal language has been taken up by Molenaar. He outlines his vision of a Theory of Geoinformation whose purpose is "to structure the whole field of geoinformation systems in the sense that the common aspects and the differences of existing GIS systems can be made more clear... [and] the common aspects and differences of the applications of GIS can be made more transparent" (Molenaar, 1991, p. 98). He notes that the definition of entities within a geographic database are very dependent upon the reason they were collected. For example, the definition of a soil unit depends on whether the data is

to be used for landuse planning, erosion studies or hydrological modeling. He argues that the theory should provide a grammar for specifying the context of the information handled and methods for transforming data from one context to another (Molenaar, 1991).

Working over a period of more than two decades, Serra and others at the Centre de Morphologie Mathematique in France have developed a very sophisticated algebra for the analysis of digital images through image analysis and mathematical morphology (Serra, 1982 and 1988, and Serra and Lay, 1988). Mathematical morphology uses a transformation algebra to convert bounded sets of points into significant numbers. The technique uses "structuring elements" which "interact with the object under study [on the image], modifying its shape and reducing it to a sort of caricature which is more expressive" than the original image object (Serra, 1982, p. v). This approach is particularly useful for medical imagery where the objects being studied (e.g. blood vessels or connective tissue) are depicted on single cross-sections that may not be oriented in a direction useful for understanding the characteristics of the object being examined. Applications of mathematical morphology range from medical imagery to plant physiology and satellite imagery.

While their data models are limited to regular (hexagonal) point grids and their applications are concerned only with extracting measurements of form, this effort demonstrates how a transformational algebra for analyzing spatially distributed data can be formally developed. Beginning with a number of "criteria" which limit the types of phenomena that can be considered under their algebra, they gradually build a set of "tools" and principles that can be used to extract spatial structure from point based data. Their most recent efforts provide the algorithmic basis for the automatic analysis of medical imagery. While this research provides useful direction, it is important to note that the work of Serra and his group seeks to understand the relationships between patterns in the data and the morphological characteristics of the real phenomena rather than considering how relevant characteristics of reality are captured in the data set.

Harvey gives an extended review of the role of artificial languages (e.g. mathematics) in the development of theories. Such artificial languages provide "an abstract system of signs and relata which have no empirical content or substantive meaning" (Harvey, 1969, p. 21). He notes that since abstract symbols and relationships must be defined precisely within the language, a clarity and lack of ambiguity that is impossible in natural languages can be expressed. Once concepts are defined in abstract languages and theories they can be linked to the real world through empirical definitions of the abstract concepts. For example, a gravity model of interaction between cities is linked to reality by stating that cities form the bodies of the formula and their masses are represented by their populations.

The basic fundamentals of algebras are well developed in any elementary algebra text. Algebraic tradition provides simple definitions of sets, mappings, relations and functions that offer an established structure for the development of an algebra. Selby and Sweet note that it is possible to "construct many different mathematical systems, depending on the choice of different sets of elements, relations, operations, and postulates (Selby and Sweet, 1969, p. 270). While many algebras are developed for specific purposes, each one has its own merits and range of interpretations. They add that "Mathematical systems may be looked upon as games that involve certain objects (elements) and are played according to specified rules (postulates)" (p. 271).

While the development of a formal language or algebra is a useful way to express theoretical concepts, it is presupposed here that environmental modelers are generally conversant with one or more formal languages appropriate for the expression of the particular mathematics of their models. Frequently these are universal computer languages such as FORTRAN or C though they may be more specialized modeling languages such as STELLA.

Furthermore, it is assumed that to develop an additional formal language to handle the spatial data would be counterproductive as modelers would be required to integrate two separate languages. Therefore, rather than developing a complete formal language, this research focuses on the development of concepts which can be implemented within existing mathematical and computer languages, or in other words a *strategy* for dealing with spatial continuity.

Object orientation

Finally, it is useful to introduce the broad concept of object orientation as it forms a central, though often unremarked, theme in this research. Object orientation means different things to different people. and the appellation object orientation has been attached to techniques, languages, databases and interfaces (Khoshafian and Abnous, 1990). STELLA, described above, is an example of an object oriented programming language while Smith's logic based database language (Smith, 1992) described above is an example of an object oriented database language. Object orientation is particularly relevant to this effort as it provides a new paradigm for modeling "the real world as close to a user's perspective as possible" and for allowing the user to interact "easily with a computational environment, using familiar metaphors" (Khoshafian and Abnous, 1990, p. 1). It allows "the normal radical transformation from system requirements (defined in the user's terms) to system specification (defined in computer terms)" (Ledbetter and Cox, 1985, p. 314) to be greatly reduced.

Central to object orientation is the concept of abstract data typing. Abstract data types are defined according to the needs of the user and the operations which are to be performed upon data stored in them:

The actual data structures chosen to store the representation of an abstract data type are invisible to its users or clients. The algorithms used to implement each of an ADT's [abstract data type] operations are also encapsulated within the ADT (Khoshafian and Abnous, 1990, p. 8).

Individual objects, which are, in many ways, the object orientation equivalent of variables in procedural languages, are associated with particular abstract data types. Thus, rather than the traditional approach where the flow of operations is determined by the statements of a procedure and performed on the passive variables, object oriented operations are initiated and defined by the objects themselves.

The relevance of object orientation to this particular effort is in the notion that the objects, or variables, themselves determine how operations are to be performed on them. As demonstrated at length later, specific implementation details controlling the manipulation of different digital representations of spatially continuous phenomena can be determined by the representations themselves and can, therefore, be encapsulated within them. Encapsulation and the definition of new data types are important themes here.

D. OVERVIEW

The fundamental theses of this research were stated early in this chapter. These theses contend that it is possible to separate considerations about the form in which spatial data about spatially continuous phenomena are stored in the computer from the operations which are performed on them. If this separation can be achieved, it should be possible to perform mathematical manipulations of the data independently of its representation. This

document seeks to confirm these theses by first exploring a number of related issues and then integrating these issues in a generic strategy.

Chapters 2 and 3 provide a background on mathematical modeling in general and the mathematical modeling of environmental processes in particular. These chapters introduce and discuss the basic characteristics of environmental models and highlight the great heterogeneity of the environmental modeling field. Thus, while this research seeks to provide a high level, generic strategy for the separation of representation from manipulation, it is abundantly clear that individual modeling domains and even kinds of data require unique treatment. The need for a generic strategy which allows for the incorporation of unique, individual solutions is demonstrated.

Chapter 4 turns to a consideration of the measurement and representation of continuous phenomena in the discrete world of the computer. The characteristics of different representations and the relationships between these spatial data models and the reality they represent are examined. This chapter develops the fundamental concept of the field variable which forms the basis of the strategy developed later. The field variable concept allows both the continuous physical field and its discrete representation to be handled separately but within a single entity.

Chapter 5 examines the typical operations that are performed on spatial data in mathematical models of environmental processes. When these operations involve discrete spatial data models of physical fields, simple mathematical operations are confounded by the need to manipulate different forms of the representations of the fields. This manipulation generally requires the conversion of one discrete field representation to another. There are many issues related to these spatial data model conversions. Some of these are generic to all physical fields and others are specific to certain modeling domains and kinds of data. How these phenomena are represented determines how their representations can be implemented in mathematical models. This chapter lays out the aspects of the conversion processes which need to be considered, either automatically by the computer or individually by the modeler, when performing various mathematical manipulations.

In Chapter 6, the themes examined in earlier chapters are brought together in a strategy for dealing with spatial continuity. Using the critical aspects of spatial data manipulation described earlier, a plan for the declaration of field variables is outlined. This declaration allows the important characteristics of each particular spatial dataset to be clearly expressed in such a way that they can be used to determine how that dataset should be manipulated. Then, in most cases, the implementation of these manipulations can be controlled automatically by the computer without the direct involvement of the modeler. This approach provides a generic solution while at the same time allows the special characteristics of unique datasets to be respected.

Once field variables can be declared, the specification of new mathematical operations using fields becomes possible. Chapter 6 outlines several such operations for scalar fields. On a more speculative theme, Chapter 7 considers some of the untapped potential of the concept of field variables by examining the role of vector fields in mathematical modeling. The declaration of vector fields follows directly from the principles outlined in earlier chapters. Finally, returning to practical matters, Chapter 8 outlines how this strategy for dealing with spatial continuity through the declaration of field variables can be implemented. Chapter 9 provides a brief summary and a look forward.

In summary, this document seeks to lay out the foundation for a generic, high level strategy which allows spatial data about continuous phenomena to be handled both as continuous physical fields and as discrete digital representations of these fields. It is

recognized that little of the data and few of the processes handled across the various domains of environmental modeling can be generalized and that much of the important work done in these fields requires an unique approach to the modeling of the physical world. However, the strategy outlined here seeks to achieve a generic solution to these diverse problems by providing a means by which the uniqueness of each modeling problem can be incorporated. Throughout this document, exceptions to the general case are discussed to illustrate this diversity and to demonstrate how exceptions can be included within this high level strategy. While the true test of the value of the strategy outlined in this document can only come through practical application, discussion of issues of a practical nature which would require consideration of specific implementations or applications are avoided in order that generic issues can be discerned before implementation is attempted. In this way, it is hoped that this effort will lead to the establishment of a solid theoretical basis for future work.

We begin, then, with a consideration of the field of mathematical modeling in general.

CHAPTER 2 - MATHEMATICAL MODELING

This chapter provides some background on the art and science of mathematical modeling in general. We confine the discussion largely to the characteristics of mathematical models of physical processes though much of what is described below can be applied equally to models of physical form and of human activities. Following a definition and discussion of models and their general purposes, some of the theories and logical constructs that have been developed to offer organization to a complex process are reviewed. A consideration of the stages that are followed during model building provide additional organizational structure. Finally, since this research is concerned with the link between data and models, the role of data in models is examined.

A. MODELS AND MODELING DEFINED

A model is, quite simply, any representation of reality. Casti has suggested that models are "toy versions of real-world situations" (Casti, 1989, p. vii). However, because of:

- the complexity of the real world,
- inaccuracies in measurement,
- imperfect hypotheses,
- modeling approximations, and
- necessity of scaling, up or down, in space and time,

the representation is a simplification of reality, elements are omitted and complexity reduced (Jacoby and Kowalik, 1980).

Many definitions of models assume a mathematical form. Cross and Moscardini (1985) define models simply as mathematical descriptions of hypotheses about physical processes. Casti defines a model as

an encapsulation of some slice of the real world within the confines of the relationships constituting a formal mathematical system... A model is a mathematical representation of the modeler's reality, a way of capturing some aspects of a given reality within the framework of a mathematical apparatus that provides us with a means for exploring the properties of that reality mirrored in the model. (Casti, 1989, p. 1)

More formally, a model can be defined as a formal representation of the relationships between defined quantities or qualities (Jeffers, 1982). "Defined" quantities or qualities describe objects in the real world. This contrasts with the abstract entities that are the hallmark of mathematics. In this sense modeling is always applied and defined within the

context of a natural system. Formal expression provides tools which the modeler can use to confirm that the model is valid with respect to reality, that it can be tested against reality and that predictions about reality may be made.

Simulation is often equated with modeling. Although a number of authors would hotly debate this equation, it appears that much of the debate is based on definitions. Zeigler (1976) and Bekey (1985) both define simulation as the implementation of mathematical models of physical systems on computers. Zeigler states "*modelling* deals primarily with relationships between real systems and models; *simulation* refers primarily to the relationships between computers and models" (Zeigler, 1976, p. 3). Thus they distinguish between the mathematical model and its implementation on the computer - the simulation. This is a useful distinction since there is a significant step between conceptualizing a natural system as a continuous model and its implementation in the necessarily discrete world of the computer. Here, numeric approximations and enforced finite word length make the conceptualization of the model somewhat remote from the realized implementation. In a similar vein, there is a big step between the continuous natural world and its discretization in a GIS.

Casti (1989), on the other hand, distinguishes between models and simulations on the basis of structure. In his view, models which provide a mathematically created black box linkage between observed inputs and outputs represent simulations rather than proper models of reality. Such models do not contain the inherent functional relationships of reality. By his definition, if A simulates B and B simulates C, A does not necessarily simulate C; "in fact, A is only a simulation of a simulation of C" (Casti, 1989, p. 468). On the other hand, if A models B and B models C, then A models C. In essence, however, the distinction between model and simulation is one of degree of correspondence between reality and the model.

Model purpose

Models can have a number of purposes. From a management perspective, models can be used to (Trakowski, 1976):

- evaluate the consequences of various environmental management decisions and regulatory strategies,
- assist in the planning process,
- develop optimal control systems and technology,
- project future environmental phenomena and variables,
- aid in interpretation and analysis of monitoring data intended to depict the state of the environment,
- estimate the risks of adverse effects of environmental hazards, and
- assess the economic and social costs arising from environmental hazards.

From the scientific point of view, models are also constructed to improve understanding of natural systems. The process of building a model forces the modeler to "rationalize one's conceptual view of a process or system and quantify the influence of each main factor" (Cross and Moscardini, 1985, p. 18). If a model is built based on theoretical principles and the model performs in a manner similar to the natural system being modeled, modelers conclude that the theory may provide some explanation of the behavior of the real

world. If the results do not match reality, deficiencies in the hypotheses may become apparent. Of course, the measurement of "similarity" is not simple and suitable tools are limited.

In many cases prediction is the objective of the modeling task. Here similarity between the model and reality must be achieved. While it is important for models to represent reality, Casti notes that "models should not be judged by their truth or falsity but rather by their usefulness as algorithms for correlating observations and making predictions". As such, a model can be used for "reality organization, i.e. a tool for ordering experiences rather than a description of reality" (Casti, 1989, p. 458). In fact, Chorley and Haggett note that models are "highly subjective approximations in that they are valuable in obscuring incidental detail and in allowing fundamental aspects of reality to appear" (Haggett and Chorley, 1967, p. 22). This is the approach to model building that must be used in the less physically-based social sciences, but it must become more common in the environmental sciences as their models become more complex and inclusive.

Models provide a means for conducting carefully controlled experiments repeatedly and under impossible natural conditions such as ecosystem destruction. Jeffers (1982) cautions that experiments on models should not be regarded as a total substitute for experiments on actual physical systems; however, models can often be used to identify the most critical field experiments which should be done. As well, model output can be used to produce comprehensive data sets that would otherwise be unobtainable, an approach used in meteorological and oceanographic research.

Types of models

Given the broad definition of "model" that is generally applied, the term has been appropriated for many different uses. With regard to GIS alone, models are commonly discussed in the following senses: geographic models, data models, GIS models, ecological models, mathematical models, decision models, environmental models, scientific models and map models. In a formal sense, it may be argued that there are only two basic types of models: physical and logical (Chorley and Haggett, 1967). Physical models are the scale-models physically constructed of materials. Logical or symbolic models are "a collection of symbolic entities satisfying a particular set of axioms and theorems" (Casti, 1989, p. 458). Mathematical models are a major subset of logical models. They incorporate variables, parameters, equations and inequalities and have no physical resemblance to the phenomena being modeled. In order to construct and use a mathematical model it is necessary to define each of the components and solve a mathematical problem (Jacoby and Kowalik, 1980). Mathematical models are the focus of this research.

Mathematical models can be classified in many different ways. The dichotomous classification developed by Jorgensen (1990) shown in Table 1 provides a useful summary of some of these. The dimensions by which models can be classified include: time and space related behavior; type of data, parameters and expressions used; model structure; and, type of mathematics. **Time related behavior** determines whether results depend on conditions existing at the moment of calculation or on conditions at previous time(s). The latter situation is typical of recursive models. Another aspect of this is whether the rules of interaction and/or the parameters change over time, *time varying*, or are constant over time, *time invariant*. **Space related behavior** likewise considers whether other locations affect the results of calculations at a particular location (Jacoby and Kowalik, 1980). *Dynamic* and *static* are the opposite ends of the spectrum in terms of time and space related behavior. Dynamic systems are those which change over time and, possibly, over space. They may have initial, transient and steady states. Transient states can only be modeled using differential or difference equations. While, by definition, steady states correspond to

conditions where derivatives are zero and thus can be modeled with algebraic equations, oscillations around a steady state can be modeled using continuous mathematical forms. Static models normally assume variables and parameters are independent of time and space (Jorgensen, 1990). Changes over space and time may be *continuous* or *discrete* and the model may be *autonomous* if it is completely isolated from the environment or *nonautonomous* if it receives inputs from outside the model (Zeigler, 1976).

Table 1 - Classification of Models (from Jorgensen, 1990)

| | |
|-----------------------|---|
| Research models | Used as a research tool |
| Management models | Used as a management tool |
| Deterministic models | Predicted values are unique and exact |
| Stochastic models | Predicted values depend on probability distribution, vary from one realization to another |
| Differential models | Mathematics based on differential equations |
| Matrix models | Mathematics based on matrix algebra |
| Reductionistic models | Include as many relevant details as possible |
| Holistic models | Use general principles |
| Static models | Variables not dependent on time or space |
| Dynamic models | Variables are a function of time and/or space |
| Distributed models | Parameters considered function of time and space |
| Lumped models | Parameters considered as constants |
| Linear models | First degree equations used consecutively |
| Nonlinear models | One or more equations not first degree |
| Causal models | Input, states, output related by physical principles (causal relation) |
| Black box models | Input and output only statistically related |

When examined in terms of the **type of data, parameters and expressions** used, models can be classified in a number of other ways: *deterministic* versus *stochastic*, *fixed-parameter* versus *adaptive-parameter* and, again but in a different context, *continuous* versus *discrete* (Jacoby and Kowalik, 1980).

Additional descriptive terms for models arise when **model structure** is considered. *Distributed* models have one or more independent variables which vary according to spatial position. Such models are frequently used for continuum or field problems and generally are described in the form of partial differential equations. On the other hand, *lumped models* are zero-dimensional; average or otherwise representative values are used irrespective of spatial location. Such models may be expressed in difference, ordinary differential or simple algebraic equations. In actual implementation, however, all distributed models are actually lumped models at some level of detail since it is impossible to solve continuous equations over continuous space in the computer. All spatial distributions must be discretized for computation. Hence, distributed models are defined only by the degree to which spatial position affects the output of a large number of lumped models distributed across the spatial dimension. Finally, the within model structure category, the complexity of the relationships between model elements also leads to descriptions such as *tree*, *network* and *feedback* models.

When the type of **mathematics** used is considered, models can be classified in yet another manner. Jeffers (1982) uses this approach to classification when he considers the vast range of mathematical models. *Dynamic* models are very flexible and allow considerable freedom from constraints and assumptions. Differential equations, non-linearity and feedback are often hallmarks of these models. However, it is often difficult to predict the behavior of such models or to estimate the values of basic parameters. *Matrix* models (input/output models) sacrifice reality and replace it with the logic of pure mathematics. Such models are easy to use and the determination of parameters results directly from mathematical manipulation. *Stochastic* models come in many varieties. Many of these models incorporate very stringent assumptions, those relating to independence of observations and error are particularly difficult when dealing with spatial data. Distribution models can be used to describe spatial distributions and to test predicted distributions against observed to confirm the adequacy of the model. Other stochastic models include those based on regression techniques. *Multivariate* models include principal component, cluster and discriminant analysis. *Optimization* models locate the maximum(s) or minimum(s) of a given mathematical expression. *Game theory* models and *catastrophe* models are other, less frequently used techniques (Jeffers, 1982).

Finally, a geographic and time dependence classification of models has been provided by Tobler (1979). This system considers the relationship between the value at one location at one time and that at other locations and times. *Independent* models show no dependence between values at different locations. *Dependent* models incorporate dependence upon the immediately previous condition at a location (e.g. the previous land use type). Historical models extend temporal dependence to the sequence of previous conditions at that location. *Multivariate* models make the value at a single location dependent upon several variables at that location. *Geographic* models incorporate dependence between different locations so that the value at a single location depends upon that at other locations.

B. THEORIES OF MODELS AND MODELING

While "the field of modelling and simulation is as diverse as the concerns of man" (Zeigler, 1976, p. vii), it can be argued that there are several fundamental principles. Zeigler and Casti have both attempted to develop formal systems.

Zeigler's theory of modeling

Zeigler incorporates aspects from general systems theory, mathematical systems theory and automata theory into his theory of modeling (Zeigler, 1976). General systems theory provides an underlying structure which assumes the unity and interdependence of natural systems, mathematical systems theory provides the formal structure while automata theory provides a logical and algebraic analysis of computer models. Beginning with a recognition of five elements in modeling - the real system, the experimental frame, the base model, the lumped model and the computer - Zeigler sets out to link the elements together in a formal structure.

The *real system* is the natural system, the one which supplies data on which the model is based. Some of the inputs and outputs to this natural system can be measured, while others are unobservable. These inputs and outputs can be linked into pairs and, according to Zeigler, this I/O behavior is "all that can be directly known about the real system" (Zeigler, 1976, p. 30). The *experimental frame* describes the range of conditions under which the real system is observed. This frame may be determined by time or space scales or by the range of measurements made. It determines the set of variables that can be measured. Identification of this frame is important as it determines the range of validity of the model constructed

based on the observed I/O behavior. The *base model* is a conceptual representation of the natural system. This model accounts for all the interactions which produce the observed I/O behavior. The definition assumes complete correspondence between the model and the natural system. However, due to the incompleteness with which we can observe the natural system, it is impossible to fully describe this model. The *lumped model* is the simplification of the base model. This is the only model that can actually be constructed and includes a number of components representing aggregations of components of the base model. Since it can be constructed, the structure of the lumped model is known completely. Although the *computer* is simply the computational device used to implement the lumped model, computational methods add another level of complexity to the relationship between the final model and the natural system. Computers and computing programs add additional structure to the model that is unrelated to the true structure of the natural system.

Using these five elements, Zeigler develops a hierarchy of models from models based on simple input/output analysis to fully functional network models. This hierarchy allows an examination using mathematical set theory of the homomorphism between the real system, the base model and the lumped model.

Casti's theory of models

In the book *Alternate Realities: Mathematical Models of Nature and Man*, Casti (1989) also attempts the monumental task of developing a theory of models. His work is commendable for the clarity it provides on a number of issues that are normally obscured by application specific aspects. According to Casti, in order to create a good model, one needs to know the properties of models, the techniques for encoding specific realities into formal systems and the procedures for interpreting the properties of the formal system in terms of the real world (Casti, 1989). In an annotation he notes that the role of the environmental modeler is to "understand the ways and means of how to encode the natural world into 'good' formal structures, and then to see how to use these structures to interpret the mathematics in terms of the questions of interest to the experimental scientist" (Casti, 1989, p.43).

Casti begins by identifying a number of elements and relationships. The infinite variety of nature is seen as an innumerable set of **states** which fully describe the total condition of the natural system under study. While these states may not in reality be distinct from one another and may be infinite in variety, any modeling effort must proceed by identifying a minuscule subset of this total set, the set of **abstract states**, which can be observed and measured. The specific abstract states that are chosen for observation and measurement depend upon the measuring apparatus and the observer's aims and knowledge.

An **observable** is "a rule associating a real number with each abstract state" (Casti, 1989, p.5). For example, "temperature" is an observable in that it is a rule for associating a real value (e.g. degrees Celsius) with a state of the system under study (thermal energy). We can measure the temperature of the system for each abstract state we have identified. Or in a simple mathematical example, consider a straight line. If we view this line in Cartesian space there are several observables: location in space given by distances along x and y axes, the slope, the y-intercept, the x-intercept, the quadrant in which x has the greatest value, the curvature, and so on. Note that we can also observe and measure the line using other origins and different coordinate systems. Casti notes that in order to "see" an entire natural system, we would need an infinite number of observables and the entire set of states. However, for practical modeling purposes we must "boldly just throw most of them away and focus our attention on a proper subset" (Casti, 1989, p. 5). As well, Casti notes that

the quantities that remain invariant under coordinate changes are the only aspects of the system that have any right to be termed intrinsic system-

theoretic properties, and we should ideally only make use of the invariants in answering questions about the system. (Casti, 1989, p. 38)

Other observables are "coordinate dependent" and are simply artifacts of the observation method. This comment is particularly relevant to spatially-based models. This implies that conclusions based solely on how we have discretized a continuous world do not truly describe the system, but rather describe aspects of our construction of it.

Equations of state are mathematical statements expressing the dependency relationships among observables. For the example of the straight line, the equation $y = ax + b$ represents such an equation of state. Using the equations of state, a number of different types of observables can be identified. The first are those which remain constant for every state, in other words the *parameters*. In the straight line model the slope and intercept are the parameters. We can also identify those observables which are functions of others. Again with the line and the slope/intercept equation of state, y is a function of x . This allows us to identify other observables as *inputs* (x) and *outputs* (y). Using these concepts, Casti develops formal descriptions of complexity, error and many specific modeling tasks in the social, behavioral and life sciences.

Herring has discussed Casti's and other theories of modeling within the context of GIS. While he is using modeling in the much broader sense that a GIS is a model of reality (a model of its structure rather than of its function), he notes that Casti's modeling theory falls within the *category theory of models*. "Category theory is the study of structure-preserving morphisms between "algebraic" structures, especially concerned with the types of information and inferences that can be preserved under such mappings" (Herring, 1991, p. 319). A morphism between reality and the model in the GIS establishes a correspondence between entities in each system so that inferences about entities drawn in one apply to the corresponding entities in the other. Like other model theories, Casti's requires that the following be identified or constructed:

- the entities or objects included in the model,
- the operations and relationships between the entities, and
- an algebra which expresses the rules of calculation and composition.

In summary, Herring concludes that all models which seek to represent the real world embody two fundamental assumptions (Herring, 1991):

- the model and the data represented reflect the pertinent facts about reality, and
- the rules of calculation and composition correspond to processes and relationships in the real world.

C. MODEL BUILDING

There is considerable advice available on model construction, particularly in the fields of engineering and computer science where modeling and simulation form an important part of their education and work. The following section is based on material merged from several sources including Cross and Moscardini (1985), Jorgensen (1990), Marani (1988), Sklar and Costanza (1990) and Haines-Young and Petch (1986).

The procedure for developing a mathematical model of a natural system has several stages. The first is the formulation of the problem. This stage involves articulating objectives, defining the variables of interest, deciding upon the assumptions and specifying the functional relationships. In order to determine the variables to be included, the modeler must identify the specific aspects of reality which represent the abstract terms in the theory which will underlie the model. A theory on plant succession, for example, may refer to successional stages which can be expressed in nature as specific plant communities. Next, it is necessary to determine measurable surrogate variables which can be used to quantify the defined concepts. In the succession model, this may be the number of individuals of specific species or species density. These variables are generally called the *state* or *response variables* of the model. Once these variables have been established, it is necessary to identify the elements of the environment which influence their value (called *forcing functions* or *external variables*) and, finally, to devise the mathematical equations which specify the functional relationships between all the variables. Selecting the appropriate model structure is critical and it may be necessary to simplify the components, observations and relationships of the real world. Zeigler (1976) discusses a number of ways to simplify models including dropping components, variables or interactions, replacing deterministically controlled variables by random variables, coarsening the range of variables, and aggregating variables and components.

The implementation stage involves coding, parameter estimation (in some cases also known as calibration) and verification. Coding may require the choice of a numerical method for the solution of differential equations in the theoretical model. *Parameters* are the coefficients in the equations. In many cases, parameters will be physically-based and derived from experimental results (e.g. rate of growth of a specific species), though frequently these parameters function as error terms and are used to account for unknown interactions or subscale processes. Parameter estimation in this latter case is often an iterative, statistical procedure requiring, in some situations, intelligent guesswork. Software for automatic parameter estimation is available (Jorgensen, 1990). Verification is a test of the internal logic of the model. This involves ensuring the model behaves as intended and that it is stable, an important consideration when working with finite difference models.

Finally, sensitivity analysis and validation must be performed. Since, in most models, interactions are non-linear, the individual behavior of model elements is unknown. Sensitivity analysis is critical to identify those elements which must be specified with the most accuracy. If it is determined that the model is very sensitive to changes in variables that cannot be supplied with known accuracies, it will be necessary to reformulate the model. Validation is concerned with how well the model output fits the empirical data. There is considerable literature available on validation procedures (Sargent, 1980). Validation techniques include asking experts if the model is reasonable, tracing entities through the model, examination of graphic simulations, and comparing model results to other model results and to historical data.

D. THE ROLE OF DATA IN MODELS

Zeigler (1976) identifies three different levels of model validity: the degree with which the output from the model matches the output from the natural system; the correspondence between predictions generated by the model and those observed; and, the degree to which the model's structure matches the structure of the natural system. All of these conditions require experimental or observational data about the reality that is being modeled. It is clear that data plays a critical role in model development and operation.

It is important to note that the effect of data collection methods has a major impact on model design and validation (Thornes, 1989). It seems likely that some modeling efforts now arise simply because relevant data is available. Cox notes "too much research in the Brave New World will be guided by whatever data happen to be readily and copiously available. The danger will be... that preoccupation with data will limit imagination and creativity" (Cox, 1989, p. 208). If this is to be the case, it behooves data provider and information system designers to take whatever steps are possible to ensure that the data they make available is used as religiously and appropriately as possible. Providing tools which help modelers handle data intelligently is perhaps a good step in this direction.

In many ways, while models are attempts to build representations of real processes, data can be seen to represent the shape or form of reality. Linking the representations of form in the data to the representations of process in the models in such a way that meaning and understanding can be obtained is the challenge of modeling. This chapter has introduced many of the fundamental aspects of modeling and provided several organizational frameworks for considering this large and growing activity. We now move forward to a detailed consideration of environmental modeling.

CHAPTER 3 - ENVIRONMENTAL MODELING

In this chapter the theme of environmental modeling is examined broadly. Interesting differences and similarities between the different environmental modeling domains are highlighted. The common mathematical structures used to implement these models are discussed, particularly with respect to the relationship between these structures and the spatial data which is placed within them. This leads to a consideration of the manner in which GIS can be linked with such physically-based mathematical models.

A. ENVIRONMENTAL MODELING DOMAINS

While the modeling of systems in the hard sciences such as physics and chemistry may be more rigorous and precise, considerable progress has been made in the modeling of complex environmental systems. Given the functional link between hydrological modeling and engineering (e.g. the US Army Corps of Engineers), it is not surprising that hydrological systems were the first to receive concerted attention in the modeling arena. Now, however, all environmental disciplines have entered the modeling arena with an interesting variety of approaches and results. Sklar and Costanza (1990) compiled a very useful summary of models in disciplines related to ecology, including geography, hydrology, biology and ecosystem science. Their summary chart, reproduced in Table 2, is a concise review of the range of scales, variables and model structures used in these different areas. While efforts to work in interdisciplinary teams are growing, much of the modeling work is still discipline-specific. Thus, the traditional "spheres", the hydrosphere, atmosphere, biosphere and lithosphere, provide a convenient way to organize the following review.

Since many of the differences between the processes and their models described below are related to their relative *scales*, it is useful to clarify the sense in which the word scale is used below. For the following discussion, scale is used in its common, non-technical sense such that large scale processes are those which occur over large territories and large time scales are those covering long time periods. Small scale processes and time scales refer to small territories and short periods, respectively.

Hydrological models

Hydrological modeling is concerned primarily with the flow of water and various mobile constituents (salts, chemical pollution, suspended sediment) over the land surface and through the upper subsurface. Hydrological processes can be modeled quite successfully by dependence upon the fundamental conservation laws of mass, momentum and energy (Maidment, 1993). These models may include advection, dispersion and various chemical and physical transformations such as adsorption, reaction and degradation. Coupled with extensive field data and widely accepted empirical relationships, these models have considerable predictive ability.

Table 2 - Summary of Model Characteristics (from Sklar and Costanza, 1990)

| Discipline | Category | Model objective | Typical time and space scales | Typical modeling variables | Basic Characteristics | Advantages | Disadvantages |
|----------------|------------------|---|--|---|--|--|---|
| Geo- graphy | Geo- metric | Define a landscape by a system of shapes | $t_f = n/a$ $dt = n/a$ $sf = km$ $ds = meters$ | shape size distance | Shape and design can explain spatial distributions | Used to organize and categorize natural landscape features; creates an emergent geometry related to theories of spatial structure | Descriptive model with little dynamic behavior; generalized shapes don't match reality |
| | Demo- graphic | Predict the flow from point i to point j | $t_f = years$ $dt = days$ $sf = continents$ $ds = km$ | diffusion transition rates density information birth death | Spatial pattern emerges from influence of origin and destination characteristics on birth, death and migration | Applicable to many situations; focus is on diffusion; easy to develop; first to use statistical parameters such as power spectra and spatial autocorrelation | Tends to be descriptive, linear, non-dynamic; focus is on pattern of diffusion without concern for process; no feedbacks; spatial attributes are few and point-averaged |
| | Network | Find the shortest distance (or cost) when going from i to j | $t_f = n/a$ $dt = n/a$ $sf = continents$ $ds = km$ | flow rates production demand distance | Flows between nodes can be optimized based upon each node's capacity to produce and consume | Widely used and mathematically well understood; flows are function of distance; finds the best solution; used in operations research and transportation problems | All intersections are linear, feedbacks are ignored; flows are assumed to be infinitely divisible in an attempt to optimize; nondynamic behavior |

Table 2 (cont)

| Discipline | Category | Model objective | Typical time and space scales | Typical modeling variables | Basic Characteristics | Advantages | Disadvantages |
|----------------|-----------------------|--|---|--|---|---|--|
| Fluid Dynamics | Hydro-dynamics | Predict the velocity and mass of a spatial fluid field | tf = days dt = seconds sf = 100's km ds = km | momentum acceleration depth friction | Flows between nodes are based upon the equations of motion and momentum bounded by link characteristics | Accurate estimates of flow for short time scales; has some feedback; mathematically well understood | Requires extensive computer processing; data intensive; network structure is fixed; only links and nodes are simulated |
| | GACM's | Predict the velocity, mass and direction of flows in the atmosphere/ocean system | tf = 10's years dt = days sf = global ds = 5° lat and long | momentum turbulence temperature moisture salinity density pressure | Space is divided into three dimensional grid cells and flows are based upon equations of motion, momentum and mass balance | Spatially explicit; simulates flows for large regions; incorporates many temporal and spatial feedbacks | Very complex; requires a super-computer; little interaction with the biosphere; coarse spatial scales |
| Ecology | Growth and population | Predict size of populations | tf = years dt = days sf = meters ds = cm | migration density information birth death resources | Spatial pattern is result of populations having different rates of growth, birth and death as a function of competition for resources | Widely used; has simple linear mathematics; focus on temporal density-dependent interactions | Spatial feedbacks are ignored; landscapes are independent, exogenous variables |
| | Eco-system | predict resource distribution and size of populations | tf = years dt = hours sf = n/a ds = n/a | migration density information birth death growth resources | Spatial pattern is result of communities having different rates of energy and material flows as a function of biological and environmental interactions | Versatile and widely used; focus on whole system components and flow rates; easily incorporates spatial variables | Mathematics can get complex; functional aggregations don't explicitly address spatial dynamics |

Table 2 (cont)

| Discipline | Category | Model objective | Typical time and space scales | Typical modeling variables | Basic Characteristics | Advantages | Disadvantages |
|-------------------|---------------|---|---|--|--|---|--|
| Landscape Ecology | Stochastic | Predict changes in a spatial pattern | $t_f = 10$'s years $dt = \text{years}$ $sf = \text{counties}$ $ds = 10$'s of m | density transition rates distance habitat | Probability of transition of a cell in a landscape is a function of intrinsic change and neighbor characteristics | Mathematics is simple and well understood; combines geometric, demographic and population modes; modifies Markovian models to include spatial resolution; finds emergent landscape structure | Transition probabilities are constant; environmental feedbacks don't exist; cannot relate cause and effect; still in development stage |
| | Process-based | Predict flows from i to j and mechanisms of change in a spatial pattern | $t_f = 10$'s years $dt = \text{weeks}$ $sf = \text{counties}$ $ds = 100$'s m | diffusion mass density information growth death birth momentum turbulence resources | Space is compartmentalized and flows are based upon equations of motion and mass balance, modified by the information, energy and material flows within each compartment | Spatially explicit; combines all types of spatial models; incorporates many temporal and spatial feedbacks; has realistic physical and biological processes; can link cause and effect; useful for evaluating specific alternatives | Uses complex non-linear mathematics; requires extensive CPU time; coarse space and time scales; still in the development stage |

As early as the '60's, fieldwork had revealed enough about the physical relationships in watersheds that the outline for physically-based hydrologic response models could be described (Freeze and Harlan, 1969). The recognition of the value of these models for water resources planning encouraged considerable research interest in the early '70's (see Bowers et al, 1972, Brown et al, 1974, and World Meteorological Organization, 1975). Research activities in this area are now focused on moving away from traditional spatially lumped models to fully distributed models. The availability of spatial databases coupled with GISs has helped intensify this activity (see for example Goulter and Forrest, 1987, Lindhult et al, 1988, Stuebe and Johnston, 1990, and any recent GIS conference proceedings for descriptions of several GIS/hydrological model interfaces). Recently, however, concern has been voiced regarding the ability of modelers to develop truly valid distributed models given the quality of data available (Beven, 1989).

Hydrologic models often incorporate two distinct geographic models. Water originates as a distributed phenomena over the watershed; therefore, modeling runoff from the entire watershed as a distributed process is an important component. However, water quickly concentrates in stream channels, hence these models also incorporate network models of the channel system. Terrain data is extremely important as input to hydrologic models. Several topographic components are commonly used as forcing functions, including elevation, slope angle, aspect and slope length (distance downslope before encountering a major change in slope) (Corbett and Gersmehl, 1987). As well, in channelized flow, channel depth, roughness, slope and width are important (Brown et al, 1987).

Atmosphere models

Like hydrologic models, atmospheric models are also firmly grounded on the basic laws of conservation, with motion fueled by energy from the sun and phase changes of water (Lee et al, 1993). Originally devised as lumped models where components were passed from land surface to the atmosphere, they can now be extremely complex models predicting the state of the global atmosphere.

The range of spatial and temporal scales at which different processes operate is an important problem in atmospheric models. Convective storms are significant energy converters and are very small in area compared to the air masses in which they are imbedded. Some processes become critical only when their size reaches a threshold. The influence of the land surface is now recognized as an important determining factor in circulation and energy fluxes in the boundary layer (Hay et al, 1993), but the heterogeneity of slope aspects and land cover makes it very difficult to combine these small scale variations with others operating at much larger scales. Parameterization of the land surface is one way in which these multiple scale factors can be combined (Henderson-Sellers, 1989) and many hope that GIS will provide a means for automatic derivation of these parameters.

Mathematically and physically speaking, ocean models have more in common with atmosphere models than they do with other models of the hydrosphere. In fact, many large ocean circulation models are fundamental components of climate models with links through the physics of energy storage, momentum and thermal convection (Han, 1988). While often modeled as 3D grids, they are generally limited to horizontal grids with spacings of a few hundred kilometers and a very few vertical layers with vertical processes introduced through parameterization.

Ecological models

Unlike hydrological and atmosphere models, ecological models are not so easily developed from strictly physical principles. Many of the fundamental theories in biology are qualitative and cannot be stated completely in mathematical terms. For example, Darwin's theory of survival of the fittest with genetic mutation occurring at random cannot be specified in a deterministic equation (Vansteenkiste and Spriet, 1980). Many biological principles depend on the identification of species and vegetation types that often cannot be unambiguously defined. As well, biological systems and organisms have tremendous variability.

In addition to the difficulty of discriminating the entities on which the models must be based, the structure of ecological models are extremely complex due to (Jeffers, 1982):

- non-linearity,
- interaction of factors,
- feedback,
- discontinuities in which a large change in behavior is often associated with a relatively small change in other variables
- bimodality in which there are a countable number of distinct states with only a few individuals between these
- hysteresis marking delayed responses to changing inputs, and
- markedly different behaviors related to increasing versus decreasing stimuli

Like atmosphere models, hierarchies are important in ecological models since processes are dissimilar at different time and space scales (Nemani, 1993). However, the magnitude of differences between micro- and macro-scale processes in biology (e.g. cellular vs. plant scale) is much smaller than in other natural sciences (e.g. local vs. general circulation) while critical time scales for processes across the range of spatial scales may be similar (Vansteenkiste and Spriet, 1980). Such problems led Vansteenkiste and Spriet to conclude in the late 70's that "the theoretical foundations in biology have not yet reached the level of abstraction of certain other fields" (Vansteenkiste and Spriet, 1980, p. 16) necessary for true physically-based modeling.

Nevertheless, many important ecological models do exist (see reviews in Baker, 1989, Costanza and Sklar, 1985, Johnson, 1990, Dale et al, 1985). Landscape ecology, in particular, is contributing a great deal to the current modeling efforts of biological systems. Ecological models normally include representation of one or more of the hierarchy of system structure elements (e.g. cell, leaf, canopy and community) and simple biological functionality such as photosynthesis, transpiration and nutrient cycling (Nemani, 1993). While until recently many of these models concentrated on temporal density variation and assumed that the future state of a landscape unit is independent of adjacent units, it is clear from evidence gathered during field experiments that this is an invalid assumption (Hunsaker et al, 1993). In the larger ecosystem models, spatial components have generally been aggregated into single variables via black-box or other non-physically-based models or by providing for interaction through flows of energy, minerals or individuals between submodel components (Sklar and Costanza, 1990).

An area of particular concern in landscape ecology is how to identify natural landscape units and how to recognize critical sized "patches" (Baker, 1989). Clearly traditional geographical research linked to the analytical capabilities in GIS can contribute considerably to this work. Perhaps the integration of these models with GIS can help overcome some of the problems that arise from the ill-defined and qualitative nature of the discipline.

Land surface and subsurface models

This category of models covers a broad range of themes. Some of these, specifically soil erosion and groundwater models, might be also considered within the hydrological category. However, groundwater models, though concerned with the flow of water, differ from hydrological models since flow is constrained more by characteristics of the medium through which the water flows than by characteristics of the flow itself. Likewise, soil erosion models include modeling of overland flow, but the focus is on interaction between the water and the landsurface. This group also includes geomorphological models of landforming processes which, of course, are related to many of the soil erosion models. While all of these models, like hydrological and atmospheric models, depend upon basic physical laws, they are more frequently spatially distributed in structure. Thus these models tend to have the highest demand for accurate spatial data, including topographical characteristics, soils, geology, land cover, land use and hydrography (Moore et al, 1993).

Since these models depend very heavily upon spatial data, problems with this data have received particular attention from researchers in this field. Many of the spatial variables on which these models are based are very poorly defined (e.g. slope, surface roughness) and depend largely on the units selected for discretization of surfaces (Evans, 1972). Considerable effort is currently being expended on developing an understanding of the effect of various discretizations and on investigating the existence of universal landscape characteristics (see for example Moore et al, 1991, Franklin, 1987, Pike, 1988 and Frank et al, 1986). As well, researchers are devising methods for dealing with a lack of appropriate data, including interpolation techniques which use data sparse models common in geological modeling and the development of reliable empirical indices to act as proxies for unknown spatial characteristics (Moore et al, 1991, 1993).

Surprisingly, very few fully developed physically-based models of land surface processes exist. This is likely due to the fact that very little substantive quantitative theoretical development has occurred. Smith and Bretherton's paper on drainage basin evolution is one of the few successful attempts to develop theoretical treatment of land surface change (Smith and Bretherton, 1972). Though Smith and Bretherton's analysis has been incorporated into several recent models (see for example Willgoose et al, 1990), its use in spatially distributed models is still limited. As a result of this lack of theoretical development, many land surface models widely used in environmental management are based on a few empirically based equations. In soil erosion modeling, ANSWERS, AGNPS, CREAMS and EPIC models are all based on the empirical Universal Soil Loss Equation (Beach, 1987), despite frequent criticisms (Moore and Burch, 1986).

General issues in environmental modeling

In spite of numerous efforts to formulate, implement, verify and validate physically-based environmental models, significant problems with these mathematical models are widely recognized. Many arise from the fact that it is often not possible to define the precise mathematical relationships due to an inadequate theoretical understanding of reality. As a result, many models include fundamental empirical relationships as part of their apparently physical basis. Processes that cannot be treated directly are included through parameters.

Modeling also is hampered by the fact that critical processes operate on many different scales in time and space and there may be scale thresholds at which critical processes change. Processes occurring over time scales which are longer than those being modeled are often considered invariant, and those over shorter scales are neglected as random statistical fluctuations (Henderson-Sellers, 1989). Statistical techniques are important elements in deterministic modeling efforts. They are required in the calibration, parameterization and validation stages and stochastic noise may arise in the model from numerical procedures, boundary conditions and data and coding errors (Unwin, 1989).

Many environmental models have grown immense over years of improvement and enhancement. Large models present particularly difficult validation problems since they are so complex that the interactions within the model may create obscure errors. These models often have hidden, hardwired constraints designed to ensure the model works that may unknowingly become invalid when new modules are added (Thornes, 1989). Unfortunately,

There is a great deal of political, emotional and organizational capital and prestige bound up in the largest of these models, and original objectives have a tendency to be lost in the momentum developed in the operation and the established wisdom resulting from it. (Thornes, 1989, p. 14)

Space in environmental models

Modelers in many fields are gradually recognizing the value of incorporating a consideration of spatial structure or processes in their models. "Spatial structure arises from the operation of processes in which spatial relationships enter explicitly into the way the process behaves.... A spatial process is a process where changes of state are due to spatial properties of the attribute" (Haining, 1990, p. 24). At present, spatial functionality in environmental models is limited primarily to simple flows between adjacent neighbors. Models for these flows are well developed for advection, convection and dispersion. There are certainly many additional spatial factors that affect natural systems, such as connectivity and patchiness. Sklar and Constanza suggest there are three ways to incorporate spatial characteristics or processes into models: 1) geometrically through system element interactions, 2) statistically, and 3) mechanically through physically-based simulation of the processes (Sklar and Costanza, 1990). However, it will be some time before the experimental data and theory development arising out of a new spatial paradigm in environmental modeling provide the basis for mathematical models incorporating spatial effects.

B. THE MATHEMATICS OF ENVIRONMENTAL MODELING

It has been noted above that the majority of physically-based environmental models depend largely upon variations in a few fundamental laws of physics. For example, following experiments on soil water movement conducted in 1856, Darcy concluded that the flow rate in porous materials is directly proportional to the hydraulic gradient. This can be expressed by what is known as Darcy's law (Skaggs, 1982):

$$q_s = -K \frac{\partial H}{\partial s}$$

where q_s is the volume of water moving through the soil in the s direction per unit area per unit time, H is the hydraulic head, $\partial H / \partial s$ is the hydraulic gradient in the s direction and K is

the hydraulic conductivity. As another example, Laplace's equation for steady-state flow in two dimensions is:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

where u is the quantity being measured (i.e. heat, water). Such relationships form the basis of many modern soil and ground water models where they are used as expressions of true physical processes or as analogies to others.

In order to solve such equations, for example to find the value of u at any place, it is necessary to find a method for isolating this variable on one side of the equation. Ideally, a solution can be found by analytical means. However,

In general, however, closed form solutions are only attainable for models which are very simple, or which have some special form. Because of the complexity of most ecological systems, models of dynamic systems must be solved numerically. (Jeffers, 1982)

Numerical solutions for differential equations in two dimensions

Computers are discrete and finite machines. Numerical methods are required to solve continuous equations and data can be stored only to finite precision. Thus while many natural processes occur continuously over continuous surfaces, it is necessary to discretize the processes and/or surfaces for implementation in computers. The search for techniques for the solution of differential equations has produced two general numerical methods suitable for implementation on computers - the method of finite elements and the method of finite differences. These numerical solutions rely on techniques which allow complex partial differential equations to be simplified in such a way that they can be solved by analytical means. As well, spectral methods are commonly used in global models of atmospheric flow.

Finite difference method

Finite difference solutions are commonly used in all of the environmental modeling areas. This solution discretizes both time and space into small steps and solves the equations as simple algebra. For example, Laplace's equation above can be replaced with the following difference equation (Gerald and Wheatly, 1989):

$$\frac{1}{h^2} [u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{ij}] = 0$$

where h is the step size in both x and y directions.

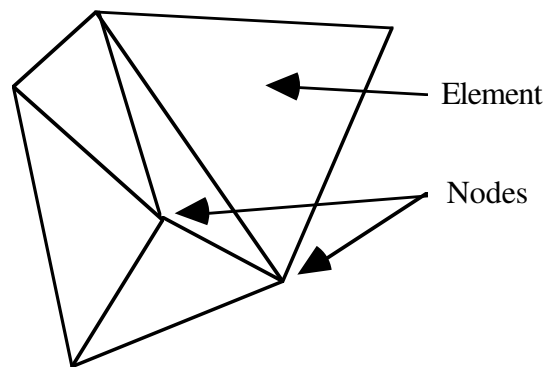
Once a finite difference equation has been devised for the differential equation, the space over which the solution is required is divided into an appropriate number of equal intervals and the difference equation is written for each of these intervals where the value is unknown. Finally, the system of equations is solved simultaneously. If the differential equation is dependent upon time, the simultaneous solution is iterated over the required number of time steps. These solutions have significant problems with compounded errors but there are many different methods for dealing with of which the Runge-Kutta method is perhaps the most widely used (Gerald and Wheatley, 1989).

It is useful to note the similarity between the manner in which space is divided in finite difference methods and traditional raster or grid-based GIS databases. This has led to the widespread use of raster GISs as database managers for such modeling efforts.

Finite elements

Developed in the 50's and 60's for use in the aircraft industry, finite element methods did not become widely popular as a solution technique for differential equations until sufficiently powerful computers were available (Barron, 1988). Finite element solutions are particularly well suited for problems where the study area cannot be easily broken into simple rectangular units or are better treated as a set of homogeneous but irregularly-shaped areas. Each element has certain known physical properties which may be the same or different than those of other elements.

Briefly, the finite element method proceeds through four stages (Gerald and Wheatley, 1989). First, the region over which the value of a given continuous phenomenon is to be predicted is divided into elements. The boundaries of each element are defined by a set of nodes and a set of lines and surfaces between these nodes:



The values at some nodes may be known from initial values and/or they may be constrained in some way (e.g. steady-state conditions). However, for the majority of the nodes, the value is unknown.

The second stage requires the identification of a set of functions which can be used to interpolate the values of the phenomenon at any location within each element. These interpolating functions are based on the values at the nodes, though these values may initially be unknown. While these functions can take many forms, they are often linear. Therefore, triangular elements are particularly useful in finite element techniques since it is easy to express linear relationships for each element based on values at the three triangle nodes. Interpolating functions can also be simplified by rotating coordinates within each element. Thus, two dimensional flow problems can be simplified by expressing the flow within each finite element in one dimension along the axis of the element instead of as a vector based on two Cartesian coordinates.

The set of these interpolating functions over all elements (where the value of any function outside its corresponding element is 0) is used in the next step as an approximation of the solution. The final result of the finite element method provides the coefficients for these interpolating functions.

The next stage requires the derivation of a functional corresponding to the general differential equations which govern the behavior of the phenomenon being studied. A functional is a function whose domain is a set of functions (Daintith and Nelson, 1989). For

the finite element method, the set of functions comprising the domain is the set of interpolating functions. Thus the functional describing the physics of the process incorporates the nodal values which are functions of the interpolating functions which are functions of space (in 1 or more dimensions) and, possibly, of time. When the matrix of interpolating functions is substituted into the functional, a large but sparse set of linear equations is obtained. In final step the linear equations are solved for the vector of unknowns (the nodal values) using well-known Gaussian elimination or iterative methods (Desai, 1979). The solution is then used to determine nodal values which can be used to describe the resulting continuous phenomenon.

Like raster GISs for finite difference methods, vector-based GISs are particularly well suited for the implementation of finite element solutions for distributed physically-based environmental models (Vieux, 1991).

Spectral methods

While confined mainly to applications in global atmospheric models, spectral methods provide a different type of solution for models based on complex differential equations. In these models, while vertical variation is discretized into layers in a manner similar to three dimensional finite difference methods, variation horizontally over the surface of the earth is expressed in terms of a series of smooth spherical harmonic functions of latitude and longitude. Note that latitude and longitude are used as spherical coordinates rather than as the rectangular coordinates commonly used in finite difference methods (Bourke, 1988). Thus, instead of discretizing space, the dynamic processes themselves are discretized. Bourke lists the advantages of the spectral method:

- (i) the intrinsic accuracy of evaluation of horizontal advection, (ii) the elimination of aliasing arising from quadratic nonlinearity, (iii) the ease of modelling flow over the entire globe, and (iv) the ease of incorporating semi-implicit time integration.... These characteristics of the spectral method afford in practice both highly accurate and stable numerics and efficient and simple computer coding. (Bourke, 1988, p. 169)

The data stored in GISs are by necessity spatially discretized, a situation which is quite compatible with both finite difference and finite element methods. However, such spatial discretization is incompatible with spectral methods which discretize across the response spectrum instead of across space. Thus the integration of spectral methods with current GISs is very limited and will not be considered further in this research.

C. SPATIAL DATA FOR ENVIRONMENTAL MODELING

The modeling of spatial processes and structures requires empirical data about the natural system. Spatial data included in environmental models are derived from available data on topography, climate and weather, soil properties, geological properties, landcover, landuse, hydrography and water quality. An excellent review of these different data sources is given in Moore et al (1993). Each of these data sources has unique and problematic qualities when it is used in environmental models. Some of these phenomena vary continuously across space (elevation, soil properties), others can be seen to be discrete (geological faults, river networks) while others may fall in either category depending upon the level of detail considered (shore- and coastlines). As well, these phenomena may be measured on continuous scales of measurement (elevation, infiltration rate) or on discrete scales of measurement (rock type, soil color). Since data cannot be stored in computers in continuous form, one of the major problems in using spatial data in environmental models is

the mismatch between reality, the forms of discretization used to collect and store data about continuous phenomena and the form in which it must be used in the model. The connection between the reality being represented and the data in the database is often very tenuous.

Roles for spatial data in environmental modeling

Spatial data may play many roles in environmental modeling. These roles can be roughly classified into three categories: visualization of model output, proxy or surrogate building, and arithmetical calculations during model operation and calibration. GIS may be integrated with environmental models for any or all of these activities.

Visualization of model output is important during both model development and model operation. Visualization of spatial data requires techniques for translating discrete data into continuous surfaces and for displaying the results in effective and meaningful ways. While visualization of spatial data has traditionally been the sole domain of cartographers, research in visualization is now also being conducted in the fields of computer graphics and spatial cognition.

In addition to its role as a spatial database management system, GIS's other primary role in environmental modeling has been as a means for deriving *proxy* data. Mathematical models of environmental processes frequently call for dense datasets about phenomena which are normally only sparsely measured, if they are measured at all. Often, spatial datasets provide only data on potentials (e.g. elevation) while the models require forcing functions (e.g. slope). Often *characterization* of the landscape, a process which combines data about various phenomena which vary at scales less than that considered in the model, is required for many regional and global models. *Resampling*, in which data must be transformed from one spatial representation into another, is also frequently required.

GIS has become a very important tool for generating such proxy datasets. In order to transform measured spatial data about one phenomenon into estimates of other phenomena, it may be necessary to use a combination of traditional GIS tools along with mathematical transformations. For example, to estimate an input value "soil moisture deficit", spatial data on cropping practices, vegetation cover, slope angle and rainfall may be transformed through the construction of Thiessen polygons, the use of buffers, polygon overlay and regression analysis, and, finally, the conversion of the result to a gridded data model for use in the finite difference solution.

Research into the construction of valid and defensible proxies predates the appearance of GIS as a scientific modeling tool by decades. Water resources modelers have been concerned with the accuracy of their predictions and the development of sufficient datasets for operating their models since the 60's. Clarke (1973) provides a review of some of the work on such issues in hydrology research during the early 1970's and Wood et al (1988) examine some of the issues being studied more recently. Laslett et al (1987) review related issues in soil science. The Long-Term Ecological Research Program of the National Science Foundation (Franklin et al, 1990) and the International Satellite Land Surface Climatology Project of UNESCO's Man and the Biosphere Program (Sellers et al, 1990) are two major programs addressing the problems of land characterization. As a result, considerable information is available in many disciplines on techniques for the derivation of these proxies and their accuracy. Burrough suggests that some formalization of this knowledge is needed so that it can be included in intelligent geographic information systems (Burrough, 1992). For the moment, such transformations depend upon the skill and knowledge of GIS analysts trained in specific disciplines.

In terms of the mathematical manipulations in models, spatial data play an important role during model calibration, verification and operation. In all of these activities, the spatial data are used as simple non-spatial numeric values in algebraic equations. For example, when we calculate $c=a+b$ in the computer, it uses two finite real numbers to produce a third finite real number. At the very core of all computer operations, each single process can involve only one or two real finite numbers (this applies even with parallel processing). Use of spatial data about continuous phenomena in this manner raises an important contradiction. While the data stored in the computer are discrete, they are modeled to represent continuous phenomena and, while the mathematical equations themselves are discrete, they are frequently also designed to represent continuous processes. Thus when designing and implementing mathematical equations using spatial data, we must think in terms of continuous phenomena while at the same time manipulating the values of discrete entities. In many cases, the mechanism for and timing of the switch between the continuous conceptual model and the discrete implemented model is critical but uncertain.

While mathematical calculation and proxy building are not mutually exclusive, the distinction does serve a useful purpose here. Mathematical calculations operate on discrete values for specific locations using traditional scalar, vector or matrix algebra, trigonometry, geometry and statistics. All of these processes can be broken down to simple arithmetic algorithms which operate on individual numbers. Location, if it does play a role in these operations, is restricted to consideration of the position of a value in a vector or a matrix. Proxy building, on the other hand, often requires direct manipulation of the spatial data through the techniques of spatial analysis. Spatial analysis requires explicit consideration of the geographical location of spatial elements. Thus again we have the contradictory situation of operations which require discrete values about individual spatial entities being used in conjunction with operations which require geographically extensive datasets. While it is useful to maintain this conceptual distinction, it is clear that mathematical and spatial operations become thoroughly entwined in mathematical modeling with spatial data.

We must also consider how data about phenomena measured on discontinuous scales are incorporated into environmental models. Discontinuous scales of measurement are typical of soil, geological, land cover and land use data. To be used in non-Boolean algebraic equations classified data must be converted to numerical values. This is accomplished in a number of ways. As is frequently the case with soils data, the functional relationships of the model may will use values that are obtained from tables relating the classified soil types given in the spatial database to specific characteristics such as permeability or erosiveness. In Boolean statements, data classes may be used to select a specific path at a branch in the computer program. Classified data may also be used in the model development process to select parameter values. For example, an examination of the study area may show that there are three vegetation types. Based on previous field research on these vegetation types, parameters related to physical characteristics such as rate of transpiration or conversion rate of carbon might be determined for use in the model.

D. The role of GIS in environmental modeling

Linking GIS and environmental modeling

The value of GIS in environmental modeling efforts is widely acknowledged. Links between environmental models and GIS are becoming common and interest in merging the technologies is growing (cf. Goodchild et al, 1993). Nyerges (1992) has reviewed the issues related to the coupling of GIS and spatial analytic models (which he defines as all those models which implement a set of numerical expressions in stand-alone software). He notes that

coupling can occur through any combination of GIS and model subsystems, implementing graphic, analysis and data management functions. However, the most common coupling occurs through the data management subsystem. Consequently, the data management subsystems form the basis of the coupling architecture.... Since the data models provide the key to the coupling, an interface which supports data model translation or conversion plays a key role in the coupling effort. (Nyerges, 1992, p. 538)

Fedra has suggested that there are three levels of integration that can be achieved between GIS and environmental models (Fedra, 1993). The lowest level involves a simple exchange of files. Since file formats may differ, it may be necessary to modify output files manually before they can be used as input to other components. The next higher level adds a special interface program which manages the file format conversions so that file sharing is transparent to the user. At the highest level, the model becomes one of the analytical functions inside the GIS or the GIS is an option in the file management and output components of the model.

Independently, several researchers have developed conceptual frameworks for the data integration that must take place if GIS and environmental models are to become fully integrated. Generally these frameworks call for interface programs which can handle all of the issues related to the conversion of different file formats (cf. Breunig and Perkhoff, 1992). Others have suggested that the solution lies in defining unified data modeling languages that can manipulate and integrate data from many different formats (cf. Smith, 1992). Whatever integration approach is the best, a conceptual framework which can guide the way in which environmental modelers think about and handle data about spatially continuous phenomena is still needed.

This chapter has examined many aspects of environmental modeling. In particular, we considered how these models deal with the problem of discretizing continuous space and continuous processes and touched upon how this may relate to the issue of integrating environmental models with GIS. With some understanding of how discretization takes place on the modeling side of the modeling/GIS enterprise, we turn now to the problem of managing and working with data about spatially continuous phenomena.

CHAPTER 4 - MEASURING AND REPRESENTING FIELDS

Just as models of natural processes are "toy versions of real world situations" (Casti, 1989, p. vii), geographic databases are also real world models, but they are descriptive rather than functional models. While there has been considerable discussion about how humans perceive the world (Couclelis, 1992), how we develop conceptual spatial data models of it (Burrough, 1992) and eventually how we construct abstractions of it in the form of spatial databases (Nyerges, 1991), for the most part the question of how to construct a database model of reality is out of the hands of environmental modelers. Databases often exist before the modeling task is conceived and the model is designed to take advantage of what is available. Thus the question is not how to represent reality but rather how to understand and work with the database's representation of reality.

The majority of the phenomena and processes which are modeled in environmental models are continuous in space as well as in value and variation. Techniques for discretizing the processes have been discussed in a previous chapter. This chapter examines how we discretize spatially continuous phenomena conceptually and represent them in computers. We begin, however, with a brief consideration of the related questions of how to measure, record and represent the value of phenomena. We also briefly consider the issue of measuring and representing location.

A. MEASURING AND REPRESENTING VALUE AND LOCATION

The "scales of measurement"

In an earlier chapter, we discussed Casti's approach to modeling the infinite variety of nature (Casti, 1989) as an infinite set of "states". Since we cannot enumerate this infinity, we are forced to concentrate on only a tiny subset of these states when we model a natural system. The states that we choose to include depend upon our goals, interests and, importantly, the measurement tools we have available. Stevens, in his landmark 1946 article "On the Theory of Scales of Measurements" notes that measurement, "in the broadest sense, is defined as the assignment of numerals to objects or events according to rules" (Stevens, 1946, p. 677). Our measurement tools allow us to develop such rules for assigning numbers to the abstract states (Casti calls these "observables"). These numbers in turn can be used "as a model to represent aspects of the empirical world" (Stevens, 1946, p. 677). In any situation there is a multitude of ways by which we can measure these abstract states and, as a result, we have many different measurement systems. How we measure reality affects how we can model it. The following identifies the different types of measurement systems available to the environmental modeler. Each spatial dataset incorporated into a model must embody one of these systems.

Steven's "scales of measurement" divide the ways we can measure phenomena into four systems. The first two of these are categorical since each observation is assigned to one of a finite, often small, number of categories or classes. These values serve as a form of identification which is used to assign a name or a class to the phenomena. Although

mathematical symbols may be used to identify these classes (e.g. 1 = rocky, 2 = loam), by definition, these class values cannot be used in mathematical expressions. They have no explicit mathematical value. If the values have no inherent order, if they serve simply as names as a means of distinguishing one entity or its characteristics from another, the measurement is *nominal*. Binary measurement systems are a special case of nominal in which the number of classes is equal to 2. Binary systems are generally used to indicate opposite conditions such as high/low, exists/does not exist; inside area of interest/outside area of interest. Values in *ordinal* systems identify a set of ordered classes (e.g. 1 = good, 2 = medium, 3 = poor). Examples of phenomena measured in nominal and ordinal systems are:

- nominal: land cover, soil type, soil texture, rock type
(binary: existence of hard pan, high concentration of salts)
- ordinal: drainage class, erosion potential

As we have noted earlier, although values based on a categorical measuring system cannot be used in mathematical equations, they may be used in computer programs in one of the following ways:

- to select a subroutine,
- to select a substitute real number from a lookup table, or
- (rarely) as a parameter (ordered sets only).

Observations using measuring systems based on real numbers are the fundamental fuel for mathematical models. There are two types of real number measurement systems: *interval* systems which have a feasible range of $[-\infty, \infty]$ and *ratio* systems with the more limited range of $[0, \infty]$. Frequently, ratio measurements are seen simply as a subset of interval numbers. However, there are some critical differences between these measurement systems when it comes to performing arithmetic. For example, while it is possible to subtract 15 m from an elevation of 10 m to get -5 m (interval system), it is not possible to subtract 15 m from 10 m of water (ratio system). Examples of spatial phenomena which are measured using the real number system are:

- interval: temperature, elevation, head
- ratio: rainfall, wind speed, infiltration rate, pH, NDVI

There are a number of specialized real number measurement systems that have additional constraints on their use in mathematical equations. Two of these which are particularly important for environmental modeling are radial systems and vectors.

Radial measuring systems produce values stated in degrees or radians and are the common units of measurement in geographic coordinates based on latitude and longitude as well as in measures of aspect and turn angle. Since these systems are circular (or semicircular as in the case of latitude) they are usually limited in range and may cycle back to initial values. For example, aspect and wind direction are measured from 0° to 360° with 360° equal to 0° , while longitude ranges from -180° to 180° which is equal to -180° . It is important to note that within the set of radial systems, the values may be either interval (latitude, turning angle) or ratio (slope, compass direction).

Vector measuring systems produce values with two components which together determine direction and magnitude. There are two ways to express these elements:

- direction (in degrees) and magnitude, and
- the x and y components of a geometric vector based on Cartesian coordinates.

Each of these can be determined geometrically from the other. Generally for mathematical manipulation, the (x,y) form is preferred. Examples of phenomena which can be expressed as vectors are horizontal groundwater flow, wind (observed as direction and magnitude) and ground slope (observed as aspect and amount of inclination).

While there are important differences in how each of these measurement systems can be manipulated, for our purposes here it is convenient to collapse these measurement systems into *categorical* and *numerical*. As noted above, while numbers can be used freely in arithmetic expressions and thus can be incorporated freely into mathematical models, categorical values are restricted to the uses expressed above. To be sure there are certain situations where operations on interval numbers create invalid values. For example, multiplication and division of numbers representing physical phenomena are only valid for ratio values -- $20^{\circ}\text{C}/2$ does not make physical sense, since 10° is not $1/2$ of 20° . Practically speaking, however, interval and ratio values are often combined in mathematical equations based on physical principles. In this case, one or more empirically based parameters will be included which may be assumed to convert the interval value to a ratio value in appropriate units. An example of this parameterized conversion is Turc's empirical formula for potential evapotranspiration (Turc, 1961):

$$PET = \frac{0.40T(R_s + 50)}{T + 15}$$

where PET = potential evapotranspiration (mm/month) (ratio)
 T = mean air temperature ($^{\circ}\text{C}$) (interval)
 R_s = solar radiation (langleys = cal/cm^2) (ratio)

Radial (numerical) data also has complex restrictions on arithmetic operations. Results must conform to the range and circularity characteristics of the resulting variable. For example, if a and c measure declination from north (feasible range is $[0,360)$) and b measures the turning angle (feasible range is $(-\infty, \infty)$) then if $a = 330^{\circ}$ and $b = 85^{\circ}$,

$$c = a + b = (330 + 85 - 360) = 55^{\circ}$$

$$c = b \times 10 = (850 - 360)^{\circ} = (490 - 360)^{\circ} = 130^{\circ}$$

Or, if a and c measure longitude (feasible range $(-180,180]$) and b measures movement in the east-west direction (feasible range is $(-\infty, \infty)$), and negative values indicate angles measured in a westward direction, then if $a = -150^{\circ}$ and $b = 70^{\circ}$

$$c = a + b = -150 + 70 = -80^{\circ}$$

$$c = a - b = -150 - 70 = (-220 + 360) = 140^{\circ}$$

So while only categorical and numerical measurement systems are distinguished here, it is assumed that the constraints on the use of the different measurement systems, as well as normal considerations for appropriate measurement units, are considered during the construction of the mathematical equations.

Data types in computer languages

The measurement systems described by Stevens become practical issues when mathematical models are implemented on computers. Computer languages are formal languages based on abstract symbols and relationships. Their power lies in our ability to relate these abstract elements to real world phenomena (Harvey, 1969). One of the most important aspects of this abstraction is the provision of a set of standard *data types*. A data type "consists of a set of valid data values, a means of denoting those values, and a set of operations that are allowed on them" (Metcalf and Reid, 1990, p. 15). For example, addition and subtraction may be performed on integer values but not on character strings, while concatenation is valid on characters and not on integers. Computer languages require each variable to be associated with a single data type. The data type of each variable determines how it can be used in a computer program and how it will be handled in memory. While the naming conventions vary from language to language, there are several common basic data types. Table 3 shows some examples of these basic data types for different languages.

Table 3 - Examples of basic data types

| Examples | FORTRAN | C | Pascal | Mathematica |
|--------------------------|------------------|-------------------|----------|-------------|
| -3,-2,-1, 0,1,2,3 | integer | int | integer | integer |
| 23.456, -10.6E-11 | real (double) | float (double) | real | real |
| 2/3 | | | | rational |
| John | character | char | char | string |
| 3 + 4.2i | complex | | | complex |
| true, false | logical | | boolean | |
| pi, e | | | | symbol |
| Basic types called... | intrinsic | basic | required | atomic |

It is useful to note that, unlike all the other data types listed, complex data types are represented by a set of two numbers, the first element for the real part and the second for the imaginary. In this way a single variable and its related operations may be defined for a 2-component value.

Most higher level languages also allow the definition of *derived* data types based on combinations of the basic types. Derived data types are used to produce vectors, arrays and other data structures specific to a particular programming task. For example in FORTRAN 90 a new data type, employee, might be defined as follows:

```

type employee
  character(len=20) name
  real payrate
  integer ssn
end type employee

```

This high level of data abstraction allows the programmer to define and manipulate data objects without being concerned about the way they are actually represented in the computer (Metcalf and Reid, 1990). In object oriented languages, these derived data types, called *abstract data types*, also allow the programmer to associate specific operations with the data type.

Measuring location

Since space is continuous, it is possible to express a location on the earth's surface to any degree of precision required. Thus values used to express location are elements of the set of real numbers. Location is usually expressed in terms of a rectangular or spherical coordinate system imposed on the surface being examined and we often use (x,y) as a generic method of expressing the location of a point.

Of course, when a rectangular coordinate system is imposed upon the curved surface of the earth, distortions arise. Map projections are the tools we use to geometrically relate locations on the earth's surface expressed as a set of latitude and longitude (spherical coordinate system) to position in a, normally, rectangular grid, expressed as an (x,y) pair. Fortunately for the large majority of modeling efforts in which location is an integral part, there are readily available formulas which can be used to convert between rectangular coordinate and spherical coordinate systems. Snyder (1987) has provided an excellent summary of the most important coordinate transformations.

B. REPRESENTING FIELDS

We now return to the specifics of working with spatially continuous phenomena for which it is useful to establish the concept of a *field*. A physical field is traditionally defined as an entity which is distributed over space and whose properties are functions of space coordinates and, in the case of dynamic fields, of time:

$$z = f(x, y) \quad \text{or} \quad z = f(x, y, t)$$

Scalar fields are characterized by a function of position and, possibly, time, whose value at each point is a scalar, while the value at any location in a vector field is a vector (i.e. wind fields where the value at a location has both magnitude and direction). Goodchild has suggested that the fundamental element of geographic information is the tuple

$$T = \langle x, y, z_1, z_2, \dots, z_n \rangle$$

which describes the value of n spatial variables at the location (x,y) (Goodchild, 1992). Since x and y are continuous, he concludes that the number of tuples is infinite. Thus the infinite set of tuples $\langle x,y,z \rangle$ containing the values of a single spatial variable over space describes a field.

Since they are continuous, physical fields are particularly distinguished by their extremely high degree of spatial autocorrelation. Thus, while we cannot measure the value of a continuous phenomenon everywhere, we know that locations near those we can measure will have very similar values. Knowledge of spatial autocorrelation, however, gives us little information about how rapidly and erratically the values change between locations at which we know the value. In order to represent and manipulate fields for mathematical modeling, we must have some way of linking the continuous variation of the field as it is observed in nature to the individual numbers or letters stored in the computer as representations of the value of the field at certain locations. In a few special cases, values and variation in space can be represented by an equation such as:

$$z = x^2 + xy + y^2$$

where x , y are horizontal Cartesian coordinates and z is the value of the phenomena at any (x,y) location. However, since surfaces in reality are rarely this smooth, the linkage between continuous reality and its representation in the computer is achieved by:

1. dividing continuous space into discrete locations for which discrete values can be measured and recorded, and
2. establishing a rule for interpolating unknown values between these locations.

The first of these is known as *discretization*. The second is accomplished through the use of *spatial data models*. The following sections consider these two issues in greater detail.

Discretization of space

Geographers have sought to bring structure to the spatial complexity of nature for centuries. Even when geography is seen as primarily qualitative, it is nevertheless necessary to partition space so that regions and elements can be described in an analytical manner. During the quantitative revolution in geography, the focus turned to the discovery of geometric structures and mathematical techniques that could be used to explain spatial distributions. While much of the initial euphoria about spatial structures has disappeared, particularly in the human side of geography (Haggett and Chorley, 1989), it is impossible to repudiate the tremendous progress that has been made in understanding how spatial structures arise and how they can be described. This is particularly evident in the development of GIS which, while apparently not driven strongly by geographic research, has its roots in much of this early quantitative effort.

As an important tool for understanding spatial structures, maps are recognized by many geographers as a special language for spatial information (Harvey, 1969). From maps, the line work and shading created by cartographers can be reduced to the basic spatial primitives of point, line and area. Tobler calls this the cartographic paradigm of geographic phenomena (Tobler, 1990). Within the map model of geographic reality, continuous phenomena are given structure in the point/line/area model through the use of contour lines and other isolines.

The point/line/area model for geographic investigation has not been, of course, the only approach. In fact, as Golledge notes, there are "many geographies and many possible worlds" (Golledge, 1982, p.21). However, while it may be argued that this map based definition of spatial primitives is too limiting given our current technologies (Grelot, 1985, Goodchild, 1988), it has become the basis of much of our current jargon in geography and, in fact, in GIS. These concepts certainly do provide powerful analytical tools. The landmark work by Getis and Boots, *Models of Spatial Processes: an approach to the study of point, line and area patterns* (Getis and Boots, 1978), provides an excellent early summary of the value of this approach while much of the current work on error modeling (cf. Goodchild and Gopal, 1989) and spatial statistics (cf. Haining, 1990, Arbia, 1989b) also demonstrates the usefulness of this spatial structure.

With the development of satellite remote sensing and grid-based computer maps, a second formal model of space, the raster, became popular. Much debate has been generated over the value and representativeness of the raster model versus the point/line/area or vector model. Many GIS companies currently market their products as "integrated", i.e. capable of displaying both raster and vector data and in some cases translating between them. David Sinton, chief of systems engineering for a large GIS company, expounds upon this common theme,

I believe the systems that integrate these data models will continue to mature and evolve through the 90's. Thus, the great debate about geographic data models that consumed so much energy in the early days of the GIS industry has become moot as the industry has matured. (Sinton, 1992, p. 4)

Unfortunately, Sinton and others who have called the debate off, have missed the most important issue. The problem is more fundamental than simply the development of algorithms to convert raster images to vector representations of them. At the heart of this debate should be the issue of how well these models represent the reality they are intended to portray.

Peuquet has attempted to bring this question into focus (Peuquet, 1988). By reviewing different concepts used to deal with representations of spatial phenomena in several different fields, she proposes a "dual conceptual model" for representing geographic data. In this model, entities in reality are seen as either locations or objects. Location entities have attributes, some of which may point to objects (e.g. county in which the location occurs) and object entities have locations as one of their attributes. By equating rasters to the locational perspective and vectors to the object perspective, Peuquet attempts to bring raster and vector together within the same model. Harvey, on the other hand, stresses that working with locations and objects requires two different languages (Harvey, 1969). For example, he notes that the concept of similarity produces entirely different results when considering similar objects than when considering similar locations. He concludes that, "Deriving 'individuals' in one language... from 'individuals' in another language requires an adequate translation procedure. Simply mixing up two very different languages will only yield garbled results" (Harvey, 1969, p. 216).

In fact, the simple concepts of raster and vector are incomplete for working with the various representations of continuous phenomena. There is more to representing reality than just breaking it into pieces. It is useful to consider the distinct stages of increasing abstraction involved in the building and implementation of a mathematical model of a natural process. Bekey identified three levels in his *hierarchy of representations* (Bekey, 1985). These are

- the reference model, which is purely conceptual since it is a perfect model exactly equivalent to the process being modeled;
- the mathematical model, which formally expresses the process variables and the relationships between them; and
- the computer implementation of the model.

In terms of the representation of space, Peuquet recognizes three similar *levels of abstraction*: (Peuquet, 1988):

- the conceptual representation
- the functionally-oriented representation, and
- the implementational format.

Within the field of database management, the last two levels are referred to as *data models* and *data structures* respectively. When considering the discretization of space for computer representation it is useful to recognize these different kinds of abstractions and recognize the differences and relationships between them. Based on these distinctions, we propose to use the following set of abstractions:

- *Geographic models* (a term proposed by Grelot, 1985) are those conceptual models used by environmental modelers as they evolve an understanding of the phenomenon being studied and extract its salient features from the background of infinite complexity in nature. Examples of such geographic models are visualizing terrain as a continuous surface which can be measured everywhere and regarding soils as a highly variable continuous phenomena with specific, measurable physical characteristics. Like Bekey's reference model or Zeigler's base model, models at this level cannot be completely specified though they can be described in a number of ways (e.g. size of smallest unit considered or the scale on which characteristics can be measured).
- *Spatial data models* are formally defined sets of entities and relationships used to discretize the complexity of geographic reality (Goodchild, 1992). The entities in these models can be measured and the models completely specified. These models provide a vehicle for interpretation of spatial data and a formal link between the geographic models and the data structures. Spatial data models are the method by which we discretize the complex natural and man-made environment so that it can be examined within the computer.
- *Data structures* describe details of specific implementations of spatial data models.

Grelot distinguishes the three levels in this manner,

a geographic model is an abstract modelling of an object... which can have at the very same time several representations built on different and complementary data models... each of which requires a specific mode of spatial data distribution [the data structure]. (Grelot, 1985, p. 576)

This research focuses on the link between a specific geographic model, the field, and various spatial data models which can be used to represent it. We now turn to a consideration of these data models.

C. Spatial data models for fields

In a literal sense, just as a hydrological model represents hydrology and a plant growth model represents plant growth, the term "data model" suggests the idea of a formal representation of *data*, not of reality. It is important to recognize that the process of developing spatial data models of a specific reality, called *data modeling*, involves the discretization of the spatial variation of that reality. Unfortunately data modeling is often confused with issues of data structure (Goodchild, 1992) and becomes mired in questions of how points, lines and areas should be represented. In fact, this confusion of terms may be partially at fault for a lack of understanding about the fundamentally different ways these data models represent reality. Each one embodies one or more important assumptions about the form of the reality represented. These assumptions critically affect how the data model can be manipulated mathematically.

For the representation of fields, there are six different spatial data models available (Goodchild, 1992): cellgrids, polygons, TINs, contour models, pointgrids and irregular points. In the next sections, we define and discuss each of these models and describe how each one models reality. For this and later discussions, it is useful to define the concept of

spatial elements. These are the basic geometric components of spatial data models (e.g. point, cell (pixel), line) and are the individual entities which are referenced and manipulated by the computer. Each of these elements is located in space and assigned one or more specific values. The full set of spatial elements in a single spatial data model used to represent a specific instance of a phenomena is a *dataset*.

Cellgrids

A cellgrid partitions the entire study area into rectangles which are evenly aligned in two perpendicular directions. Regular tessellations which are not rectangular (e.g. hexagonal tessellations) are considered polygon models. The most common example of a cellgrid is a remotely sensed scene composed of pixels. The value of the phenomenon over the entire area covered by each cell (the spatial element) is represented by a single value even though there may be considerable variation within the cell. As a result, values change abruptly at cell edges. The geography of a cellgrid can be fully described by specifying the cell width and height, the origin of the grid, the orientation of the rows or columns to a compass direction and the projection used.

Polygons

Polygons partition the entire study area into irregularly shaped contiguous regions. Like cellgrids, the value of the phenomenon within a single polygon is defined as a constant and changes abruptly at polygon edges. The boundaries of a set of polygons may be defined either by the phenomenon (e.g. vegetation zones) or they may be independent of the phenomenon (e.g. cut blocks, watersheds when used to partition soil characteristics). In environmental databases, polygon-structured data is often categorical (e.g. soils, vegetation types, watershed). To be useful in mathematical models, these categorical datasets are usually linked to a relational table which describes various numerical and other properties of each class.

TINs (Triangulated Irregular Networks)

TINs partition the entire study area into triangular regions. The value of the phenomenon is specified only at triangle nodes. However, since the surface of each triangle is assumed to be a simple function of rectangular coordinates, values anywhere on a triangle face can be calculated directly from the values at the nodes. (Many different forms can be assumed for the variation on these triangular faces, but since planar faces are the easiest to work with and the majority of commercial implementations of TINs allow only planar faces, we consider only that form here.) While there is no abrupt change in value at planar triangle edges, there is an abrupt change in slope. The location of boundaries is defined by the location of the nodes. Thus the correspondence between the real surface and that represented by the surface of triangles is determined by the set of points (nodes) selected to define the critical points of the surface. Since TINs define continuously varying surfaces, TIN models can never be used to structure categorical, non-numerical data.

Pointgrids

Pointgrids store the value of the phenomenon at every intersection in a rectangular grid. These values represent the actual value of the phenomenon at that location. The location of each sampling point is determined by the grid, independently of the phenomenon. If a pointgrid dataset has been derived from a primary data source (e.g. elevation from stereo pairs), no assumptions are made regarding the representativeness of each value within the neighborhood of sampling points. However, if the dataset has been derived from some other spatial dataset, points may indeed be representative of the neighborhood. For example, the

values in a pointgrid derived from a cellgrid are representative of the cell neighborhoods rather than of the point values. The geography of a pointgrid may be described by specifying the x and y spacing, the origin and orientation of the grid and the projection used.

Irregular points

Irregular point models store the value of the phenomenon at irregularly scattered point locations. The location of the points may be determined by the phenomenon. In this case, values may be assumed to be representative of neighboring locations (e.g. carefully selected representative locations for the collection of rainfall data). However, irregular point data may also be collected at locations determined by concerns other than the phenomenon under study (e.g. weather stations located at airports). In this case, the value at each point is less likely to be representative of the surrounding conditions.

Contour models

Contour models are unique amongst these spatial data models used for continuous phenomena. Unlike the other spatial data models, contour models are constructed by holding the value of the phenomenon constant and determining the location. Lines are constructed to connect adjacent locations whose value matches that of the desired contour line value. This model explicitly identifies all places which exhibit a value expressed by one of the contour lines. However, the value of the surface is defined only along the contour lines. The location of the contour lines is determined by both the phenomenon and the selected values at which contour lines are drawn.

A unique characteristic of this model is that the spatial elements are lines, rather than points or areas as in the other five models. These lines are ordered by value so that neighboring contour lines are either equal in value or different by only one contour interval. Like TINs, contour models partition space into regions over which the value of the phenomenon varies. Unlike TINs, the variation between contour lines is not linear or otherwise clearly defined. The only assumptions that can be made about the variation between the lines are that the value of the phenomenon remains within the range defined by the values of the bounding contour lines and that locations close to contour lines have a value close to that of the lines. Finally, like TINs, since contour lines must be measured on a continuous measurement system, contour models can never represent categorical data.

Characteristics of the spatial data models

Having defined and briefly described the six spatial data models used to represent fields, it is now possible to consider their different characteristics in a holistic manner. Goodchild has suggested that these six models represent two distinct ways of exploiting the spatial autocorrelation of fields (Goodchild, 1992). *Piecewise* models make use of the assumption that nearby locations are similar while *sampled* models exploit the fact that if we know the value at one location we can estimate the values at nearby locations.

Piecewise models dissect the surface into contiguous regions. A value is defined at every location on the surface. The continuous variation of the value of the phenomenon within each region is described by a simple mathematical function of the coordinates. In two models, gridcell and polygons, this mathematical function is a constant while in the TIN model the function is linear. Thus if the values of the phenomena represented are drawn as a third dimension, gridcell and polygon models produce a stepped surface of horizontal regions, while the regions of the TIN model are sloping planes with the edges of each region coincident to those of its neighbors. The crucial assumption in all piecewise models is that the value or function assigned to each region is representative of the average value or general

trend of the surface in the region. While each individual point may not be represented precisely, it is assumed that the integral of the values over this surface would produce the value or linear function assigned.

Sampled models use an entirely different approach. In these models, the phenomenon is sampled precisely at a number of different points. Sampling is done either at points as in pointgrids and irregular point models, or along lines as in contour models. No values are assigned to locations that have not been sampled and, except in the limited case of contour models, no information is provided about the variation in the value of the phenomenon between sample sites. In order to represent the continuous surface between these sample locations, an assumption must be made that variation between these points can be described by a mathematical function. However, unlike piecewise models, the form of this function is not always clearly defined. Frequently, linear functions are used though other forms are also common (e.g. higher order functions to fit a surface exactly to points in a 3x3 window of a pointgrid). The interpolation function chosen may vary for a single dataset in different applications. As well, the accuracy with which the value of a given point on the surface can be predicted varies depending upon its distance from a sampled site since, in general, the value of a point very close to a sampled site can be predicted with greater accuracy than the value of a location at a great distance away.

Contour models actually display a combination of sampled and piecewise model characteristics. While they are indeed sampled in the sense described above, they are complete samples of all locations with the selected contour line values. This provides additional information about the variation between lines. Thus, from the piecewise perspective, they do imply the existence of some information about the variation within each region bounded by lines. And, like TINs, contoured surfaces drawn in three dimensions display smoothly varying surfaces between lines.

Thus we have two groups of models with widely different basic assumptions. While piecewise models provide a generalized representation of the continuous phenomenon, sampled models provide precise data at a limited number of locations. Sampling schemes may be unbiased (as in pointgrids) or biased (as in contours and some irregular point models). In terms of surface representation, it is useful to regard the 6 models in three distinct groups. Constant piecewise models depict a stepped horizontal surface with vertical breaks at cell or polygon borders. Surface models, TIN and contour models, depict a continuous surface with varying values within regions and continuity across borders (triangle edges or contour lines). Point models do not depict a continuous surface; interpolation must be used to construct one.

Spatial data models as representations of reality

How well a spatial data model represents reality is a multi-dimensional issue. Many authors have discussed the sources of error in digital data (good summaries are found in Burrough, 1986, Chrisman, 1991, Goodchild and Gopal, 1989). Error arises initially in the measurement of the phenomenon and may include mismeasurement and incorrect recording of one or both of the value and the location. This has led to the development of separate measures for locational and positional accuracy and a recognition that errors may be random or systematic (Chrisman, 1991).

If we assume that the true surface has been accurately sampled and recorded, spatial data models are still only representations of reality. Is it possible to estimate the accuracy with which a spatial data model fits reality? How can the fit of a discrete representation of a continuous surface be tested? The direct answer to these questions is simply, with difficulty. Since it is continuous, the true surface cannot be completely described. Therefore the model

cannot be directly compared to it. As well the high degree of spatial autocorrelation between values at points in a field invalidates many of the traditional statistical tools that can be used to measure the degree of correspondence between the model and reality. While there are some measures, including the US National Map Accuracy Standard (Chrisman, 1991) and kriging (Burrough, 1991), none of them are completely satisfactory. We must often resort to qualitative assessments of the veracity of the models.

But even qualitative assessment is difficult. Unlike the resolution of satellite images, the amount of detail we can discern on a continuous surface is infinite. The closer we look the more we see (Tobler, 1988). Our assessment of the relationship between the surface represented by the model and the true surface will be affected by the scale of the phenomenon we are considering and the sampling frequency with which we have measured and recorded the surface:

The interaction between sampling frequency and terrain variation is important because the type of land feature captured at a certain resolution is site specific.... Sampling theory dictates that a grid DEM can represent terrain features no smaller than twice the spacing (the Nyquist frequency) and therefore for a 30 m DEM, the minimal feature resolution is 60m. Conversely, the TIN-based DEM theoretically responds to the variation change in a terrain, with the sampling intensity increasing in areas where the terrain variability increases. (Theobald, 1989, p. 105)

Another important issue related to phenomenon variability concerns the smoothness of the spatial data model used to represent the physical field. While many phenomena may exhibit high frequency variability, digital datasets representing these fields often give produce models of only low frequency variation. This, of course, is related to the issue addressed by the Nyquist frequency since the size of the spatial elements in the spatial data model can only capture phenomenon variation at scales greater than the spatial elements. As well, different types of spatial data models represent this high frequency variability in different ways. Constant piecewise models replace local variation with a single local average while surface models depict variability by expressing the rate of change of the value of the phenomenon through measures of slope (i.e. the slope of TIN triangles or the closeness of contour lines). How these models of reality are used and interpreted by a modeler must be determined by the application in which they are to be used. For example, if small variations in the value of the phenomenon are important, it is essential that the spatial data model is capable of representing these variations. This may be done either through the use of very small spatial elements or the development of measures of variability such as those provided by the application of geostatistical techniques. In other cases, where small variations in the value of the phenomenon lead to inconclusive model results, a spatial data model which smoothes these high frequency variations while maintaining a good representation of low frequency variation will be required. If a dataset cannot depict the appropriate level of variability for the modeling effort, then some means to introduce that information must be sought. This is the responsibility of the modeler, not of the dataset.

A further concern about the relationship between reality and models of it is due to artifacts that arise in certain datasets. An excellent example of such artifacts are those which can be observed in TIN models which have been derived from contour models. Due to the algorithms which are used to create TINs from contours, a derivative model may display such processing artifacts as flat triangles (which occur when the three nodes of a triangle are obtained from a single contour line) and dams and divots (which occur when triangles cut across ridges and valleys which are not captured by the contour lines) (Kumler, 1992). Whether these artifacts create problems depends upon the application to which the derivative TIN will be put. If the TIN is to be used as a means for creating a shaded relief image of the

surface, these characteristics may be extremely misleading. Similarly they may have a great impact on the results of a hydrological model since they will cause flow directions to be modified. On the other hand, if the spatial model is to be used in a mathematical model where only the scalar value at any point is important, values interpolated on TIN triangles should be similar to those on the contour model. Thus the match between reality and the model is in some cases a matter of subjective choice - does this model match the modeler's version of reality? And finally, in the absence of expert knowledge about the phenomenon represented, we must rely upon the basic assumptions of the spatial data model which has been chosen to model reality and hope that its selection was driven by a desire to represent reality as accurately as possible.

Data structures for field spatial data models

As Goodchild pointed out, data structures often become confused with data models (Goodchild, 1992). The reason for this is simple - there is a complex mapping between data models and data structures. If we consider only two large categories of data structures - raster and vector - the mapping between data models and data structures might look like this:

| | |
|------------------|---------------------|
| Cellgrid | -> raster |
| Polygons | -> vector |
| TINs | -> vector |
| Contours | -> vector |
| Pointgrids | -> vector or raster |
| Irregular points | -> vector |

Thus, a dataset may be stored in a vector format, but it may represent one of several different spatial data models. In order to use a spatial dataset appropriately, it is necessary to know which spatial data model has been used during the data modeling stage of database development.

D. MODELING WITH CONTINUOUS VARIABLES

For environmental modelers, designing and coding a mathematical model is an entirely different task than accessing and manipulating spatial data in a GIS. On the one hand modelers can use well-known and well-structured algebraic and computer languages, following widely accepted and proven rules for substitution and solution. On the other hand, when manipulating spatial data for use in the models, the modelers have only the idiosyncratic language of a specific GIS to work with. The procedures they must follow to get at the spatial data are not codified in any common language. There are no widely accepted common rules and defaults to guide how spatial data are used in environmental models. Thus, while modelers can use a common symbolic language to express the development of their mathematics and thus prove the validity of their approach, there is no simple way to express the transformations and manipulations that are necessary to incorporate the spatial data into the model. The consequence of this is that it is very difficult to assess the validity of the data incorporated into models which have been based on spatial data and, as a result, it is difficult to evaluate the validity of the model results.

A strategy for dealing with spatial continuity

What is needed are common strategies and techniques for handling spatial data about continuous phenomena in all its forms. A common strategy for handling data about fields in mathematical models provides a framework in which many issues related to the representation of continuous phenomena can be addressed. An awareness of the basic

assumptions which are embodied in each field data model and a means for expressing exceptions to these assumptions can be provided for. Specifically this strategy should:

- allow expression and manipulation of variables and data about continuous phenomena in common symbolic languages. In other words, the strategy should be capable of being incorporated into computer language implementations of environmental models. This is in direct contrast to the natural language-like structure of Tomlin's map algebra and is more amenable to the scientific environment.
- eliminate the necessity to consider the form of the spatial discretization (the data model) whenever possible. While we believe it is desirable and possible to achieve this objective for most operations, it is necessary to provide for input of additional information for some operations.
- provide a syntax for incorporating primitive operations appropriate for environmental modeling with fields but which are not yet available in GIS or common programming languages. These include operations to perform discrete versions of "differentiation" and "integration" on variables representing fields and the incorporation of the concept of vector fields.
- guide and enable the rapid development of direct linkages between environmental models and any GIS.

The remainder of this chapter establishes the fundamentals of this proposed strategy for handling spatially continuous data in environmental modeling projects.

E. FIELD VARIABLES

The field data type and field variables

In order to manipulate data about spatially continuous phenomena, we begin by defining the *field data type* to be used in addition to the traditional data types (e.g. float, integer, character and so on). Variables declared as field data types are *field variables*. Field variables are the logical or functional representation of the concept of fields. These variables are spatially continuous and represent values of the field during a single slice or instant of time. Like other types of variables, fields are represented with symbols. In this document, uppercase letters are used to denote field variables. For example, the field of temperature may be represented by the symbol T or by TEMP.

For any field variable, it must be possible to determine a value at any location and these values may differ from location to location within the same field variable. Using a Cartesian coordinate system, we can refer to the temperature at a specific location in the field using the notation $T(x,y)$. While Cartesian coordinates are the default, it is possible to denote the value of the variable at a point using any coordinate system. The notation $T(x,y)$ reinforces the notion that the value at any point in a field is a function of its location.

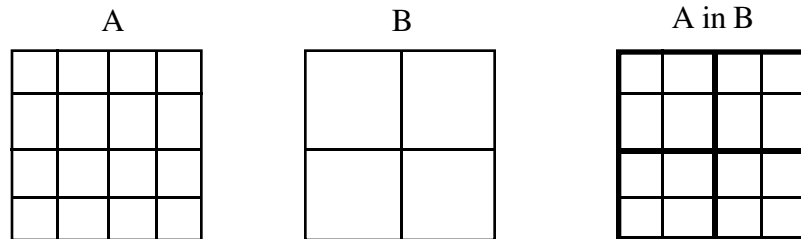
Spatial equality and nesting

It is useful to define the concepts of spatial equality and nesting as they are essential when doing mathematics on field variables. These concepts are used to compare the specific spatial discretizations of different field variables. In spatially equivalent field variables, the geography of all spatial elements correspond exactly and completely. Such a condition is

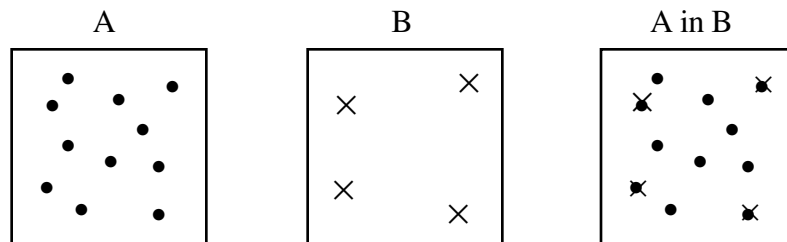
found in co-registered cellgrids of equal dimensions, i.e. if A and B are spatially equivalent cellgrids ("A as B"), they have the same cell dimensions, origin, orientation and projection. Spatially equivalent polygons are less common in environmental datasets since boundary locations are generally determined by the phenomena observed. However, when a field variable is derived from another by simple substitution of a set of classes for a set of numeric values, the resulting variable will be spatially equivalent to the original. Spatial equivalence is essential for most mathematical operations on field variables.

Spatial nesting indicates that one spatial variable nests spatially within another. The definition varies slightly for piecewise and sampled models. For piecewise spatial models, if A spatially nests within B ("A in B"):

- each element in A falls completely within one element in B, and
- the set of lines which form the boundaries of B is a subset of the set of lines which form the boundaries of A.



For sampled models, spatial nesting means simply that the spatial elements of A are a subset of the set of spatial elements of B; again A is nested in B:



Nesting arises most frequently when working with cellgrids. For example, consider two cellgrids, A and B, with the same origin, orientation and projection but different cell dimensions. If A has a cell width of 10 seconds and B has a cell width of 1 degree, A is nested in B such that there are 36 A cells in each B cell.

Declaring field variables

Earlier in this chapter, a strategy for dealing with spatial continuity was outlined. We propose that this strategy can be achieved by incorporating field variables into environmental models. The essential element of the implementation of this strategy, then, is the statement in which field variables are declared. The objective of this declaration statement is to provide the information needed to establish all the parameters for spatial data manipulation and to determine which specific operations must be used given the specific spatial data model and the characteristics of the fields represented. In most cases, sufficient information can be provided in the declaration of the field variables that no further input is required from the modeler when these variables are manipulated. In this manner, several of the strategy's objectives can be met, specifically it:

- allows the "expression and manipulation of variables and data about continuous phenomena in common symbolic languages", and
- eliminates "the necessity to consider the form of the spatial discretization (the data model)".

Like declaration statements in standard programming languages, these declarations establish certain functionality constraints and options particular to the specific data type. For example, in C, when a variable is declared as

```
char A[10,8]
```

memory is set aside to allow for a 10 by 8 array of characters. It is subsequently impossible to place a floating point or integer data value in this variable location, or to algebraically add two elements of this array. Similarly, the declaration of field variables establishes not only the type of data contained within the variable, but how it can be manipulated. Declaration of the field variable is not a difficult task. It simply provides a structure in which the essential properties can be unambiguously and completely described. These properties will be used during runs of the model to determine automatically the appropriate operations and conversions that must be performed.

In this chapter we have introduced the concept of fields and considered how fields are discretized and represented in the computer. Since the manner in which fields are represented is fundamental in determining how mathematical operations can be performed, the properties associated with field variables describe the data model used and other critical characteristics related to information density, temporality and measurement system. These properties are critical in determining how the variable can be manipulated mathematically within the computer and help determine how variables represented by different data structures may be combined in a single mathematical statement. The issue is to determine what these critical properties are and how they can be expressed clearly. The next chapter considers the operations that may be performed on field variables. These considerations lead, in Chapter 6, to a complete description of how field variables may be specified and an outline of the requirements for the specification for a mathematical model using field variables.

CHAPTER 5 - OPERATIONS ON FIELD VARIABLES

There is a wide range of operations which may be carried out using field variables in environmental models. This chapter seeks to bring some organization to this variety. We begin with a quick review of some of the different classification schemes that have been devised to organize the wide range of operations that can be performed on spatial data in GISs. This provides a useful background to and contrast for the remaining sections of this chapter in which we examine in detail the operations which may be performed on field variables within mathematical models.

A. GIS OPERATIONS

GIS research concentrates on the special characteristics and problems of manipulating spatial data. It is useful, therefore, to begin with the GIS perspective. Three contrasting approaches to organizing the wide range of operations that can be performed on spatial data and thus might be available in GIS are considered here.

Tomlin's map algebra

While Tomlin's map algebra approach (Tomlin, 1991) was designed to manipulate only spatial data in gridded spatial data models, his fundamental organizational scheme has found wide acceptance. Tomlin organized the various operations that might be applied to different data layers (which may be different attributes or different time slices) into four general classes - local, focal, incremental and zonal (Tomlin, 1990).

- Local operations operate on the values of a single location (a cell) through a number of different layers.
- Focal operations use the values of a neighborhood around a single location.
- Incremental operations allow extension of the operation to neighboring cells which exhibit a connected attribute, such as flow direction or position along a linear feature. This provides for the consideration of larger anisotropic neighborhoods.
- Zonal operations operate on all locations within the same zone (class).

This classification scheme emphasizes the difference between operations which are carried out on data for a single location and those which relate to neighborhoods defined by proximity, distance and direction.

Raper and Maguire's categories

Raper and Maguire (1992) have identified five major functional categories for GIS operations. The five categories are:

- data capture, transfer, validation and editing
- data structuring
- data manipulation - restructure, generalize and transform
- analysis and query
- presentation

This approach organizes the functional operations contained within typical GISs, classifying them according to the sequence of steps that might be made during the implementation of a single GIS project.

Burrough's classification

Burrough's approach is much more theoretical and has broader application (Burrough, 1992). He has constructed 9 classes of GIS operations:

- Class 1 operations derive new values from the exact values of discrete objects and include arithmetic and Boolean operations and classification (numerical taxonomy) methods.
- Class 2 operations produce non-exact values from exact values of discrete objects and include statistical and regression methods of value estimation.
- Class 3 operations derive new object values from the values of locations within the neighborhood of discrete objects. Included are operations of adjacency, connectivity and proximity.
- Class 4 are also neighborhood operations, but these are concerned with neighborhood operations on continuous surfaces and include filtering, indices of spatial variation, computation of derivatives (slope, aspect, drainage networks), interpolation and surface fitting.
- Class 5 operations are the inverse of 3 and 4 since they assign values based on the value at the original location to the locations in its neighborhood through the use of buffer zones and point-in-polygon operations.
- Class 6 operations create new spatial objects through overlay, buffering, centroid calculation and smoothing.
- Class 7 operations derive values based on geometrical attributes of the objects under study and include measurement of shape, size and topology.
- Class 8 operations produce summary reports including histograms, counts of occurrences and cross-sections.

- Class 9 operations are for data management and include rectification, projection change and join functions.

This classification system makes some distinction between operations performed on data representing continuous fields and those on data representing discrete objects. However, only class 4 explicitly deals with operations on continuous surfaces. All other operations, if they are to be performed on data representing continuous phenomena, require discrete data as input. Continuous operations must be approximated.

B. MATHEMATICS ON FIELDS

While Burrough's classification comes the closest to describing mathematical operations that may be performed on spatial data, even it does not address the issues from the perspective of traditional mathematical manipulation. We now turn to consideration of operations that might be performed on spatial data from the mathematical perspective, leaving the traditional GIS approach behind.

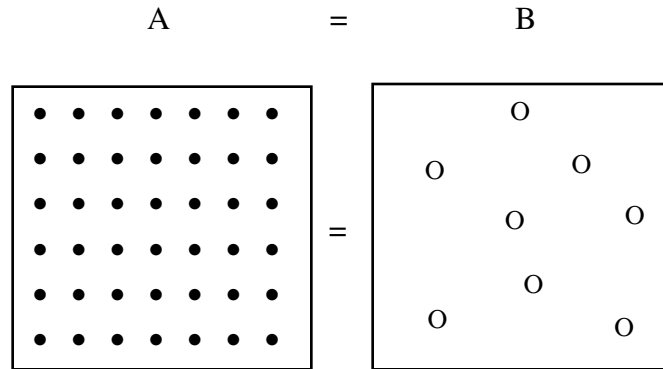
We have noted above that the computer is incapable of adding two continuous fields to produce a third continuous field. All fields must be reduced to simple finite numbers before mathematical manipulation can proceed. This is the function of spatial data models of continuous phenomena. However, there is an additional complication. In order to manipulate two fields simultaneously (as in addition or multiplication), the locations for which there are simple finite numbers representing the value of the field must correspond. To add field A to field B, one must add the value of A to the value of B at the same location. Different spatial data models express location in ways which are generally incompatible. This implies that in order to perform mathematical operations on data in various spatial data models, we must first convert all models to spatially equivalent ones, or at least to extract estimates of values for locations in one field variable for which we have data in the other field variable. This condition can be expressed most directly in the "=" or assignment operation of traditional algebra. We begin, therefore, with a detailed consideration of assignment and then examine other mathematical operations important in the construction of mathematical models.

Assignment

This operation is the most fundamental of all mathematical operations. By definition, all mathematical equations require assignment. In standard programming languages, assignment statements such as $(A = B)$ or $(A := B)$ or $(A <- B)$ replace the value of the left hand variable with the value of the right hand variable. If the type of the two variables is not the same, a conversion is performed to restate the value of the right hand variable in the data type required by the left hand variable. A similar convention must hold here.

As with simple scalar variables, the conceptual version of the assignment operation for fields is simple. If B is the temperature field and $A = B$, then A is a copy of the temperature field. Every location has the same value in A as it does in B. But there are a number of different ways to represent fields in the computer. If A is declared as a different spatial data model than B, then it is entirely possible that a value which must be specified at a given location in A is not precisely specified at the same location in B (see diagram on the next page). Thus assignment, the simple, fundamental mathematical operation, becomes a complex spatial operation when fields are involved. It requires the conversion of one spatial data model to another. Since each model provides a different representation of reality, it is important to confront these differences directly during the operation. However, it is our contention that it is possible to codify these differences in such a way that the decisions

regarding how to convert one model to another can be handled automatically, without input from the modeler. In the next section we discuss issues which determine how these conversions should be done and lay out a scheme for organizing and selecting appropriate procedures.



It should be noted that the term *conversion* is used instead of the possibly more common term *transformation* to describe this model-to-model assignment process. The term transformation is often synonymous with function and commonly used to refer to changes in coordinate systems. Such transformations generally are invertible; in other words, a transformation can be inverted to return to the original data without any change or degradation in the data or loss of information (Tobler, 1979a). In the case of spatial data models, the conversions or transformations which are performed are rarely invertible since they generally lead to a loss of information. For example, even though exact values for a given set of points can be extracted from a TIN, it is not possible to return from that set of points to the original TIN if the TIN nodes are not part of the point structure. (An exception is any conversion process which densifies the structure without changing the spatial data model.) Therefore, we prefer the term conversion to emphasize that the model as well as the information content is, in most cases, permanently changed by one of these operations.

C. CONVERTING FIELD DATA MODELS

Selection of appropriate techniques for converting field data models to other field data models requires consideration of several issues. Most important is the consideration of how each model represents reality. In an earlier chapter, the six models were reviewed and the ways in which they model continuity by taking advantage of spatial autocorrelation were described. The models differ in the assumptions that must be made to derive the continuous surface from the discrete representation but each provides some link with reality. In order to convert models, we must exploit each model's link with reality as data is extracted from one model and placed in another. This process may be conceptualized in two stages. First we must derive a continuous surface from the original discrete spatial data model, then we must use an appropriate technique to sample the continuous surface to produce the target model.

The derivation of a continuous surface from a discrete representation involves *spatial interpolation*. Goodchild has defined spatial interpolation as the task of computing a complete continuous surface from a set of sample points (Goodchild, 1992), though Tobler suggests that it also includes computation using any other spatial data model used to represent continuous phenomena (Tobler, 1988). Here we define spatial interpolation as *a set of rules for obtaining a complete field from a spatial data model*. Spatial interpolation has a long history. It has always been an important tool for geologists wishing to interpret the

limited clues they collected on the surface of the earth. One of the first things that geographers have traditionally learned is how to interpolate elevations from contour maps. With the rapid expansion of the use of computers for scientific investigation and the development of elevation matrices and digital contour maps in the late '60's, interpolation became an important and widely examined computational procedure. Many different approaches and algorithms for interpolation and the resulting conversion between different spatial data models exist. Excellent reviews of spatial interpolation methods can be found in Schut (1976), Lam (1983) and more recently in Burrough (1986).

Sampling may be similarly defined as *a set of rules for obtaining a spatial data model from a complete field*. Together these two processes, spatial interpolation and sampling, may be regarded as *resampling* (Tobler, 1988). By splitting resampling into these two stages, passing through a best guess of reality, we ensure that the link with reality is maintained and that the final representation is as close to it as possible.

When performing spatial data model conversion it is necessary to consider the type of data involved. Numerical data are measured on a continuous scale and thus permit the derivation of values which lie between those in the original dataset. Since values in categorical datasets are elements of small, finite sets of discrete classes, derived values must also be members of these sets and no new values can be created in a conversion operation. This fundamental difference suggests that different approaches to the conversion of spatial data models will be appropriate for these different types of data. As well, operations involving categorical data about continuous phenomena are restricted to a subset of the six spatial data models. Categorical data cannot be stored in the surface models - TINs and contour models. Point models may be used to store categorical data, though any practical use of this data requires the initial construction of piecewise models as a representation of the continuous surface.

The next section discusses how each spatial data model may be converted so that mathematics can be performed on data stored in different representations. This demonstrates a fundamental principle of this work - no common model is assumed for the representation of fields. Each field may be represented in a different manner, the choice of which model being dependent upon many things, including the phenomenon being represented. Thus we make an important fundamental step away from the cellgrid confines of map algebra.

The approach outlined below is prescriptive, suggesting how each of these models may be converted most precisely. The purpose of this prescription is to demonstrate how rules for conversion can be devised, not to set out definitive techniques. The derivation of definitive techniques will require implementation and widespread application of the concepts outlined here. In later chapters, we discuss how these rules can be encapsulated with the spatial data itself or added to standard programming language compiler libraries and accessed as required during mathematical model runs. The approach is generic although we recognize that specific phenomena and datasets may require specialized handling. Provision for such specialized handling is described later.

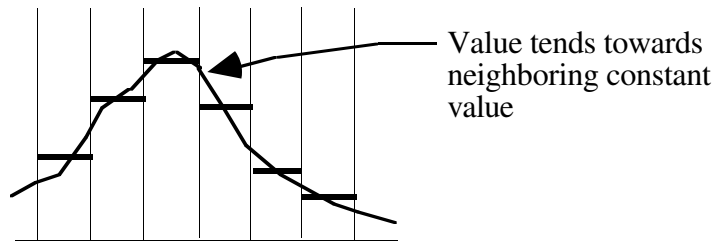
We begin by considering how each model can be used to obtain a continuous surface. This is followed by consideration of how each model can be sampled from the field. Conversion of categorical data models is treated separately from conversion of numerical data models. We define the *source* data model as the original or right hand side of the assignment statement, and the *target* data model as the destination or left side of the assignment. The term *region* is used to refer to either cells or polygons. The structure of the target data models must be predefined, since, within the context of performing mathematics on fields represented by spatial data models, models will only be converted if there is a need to make them spatially equivalent to other models. Thus while generalization or

enhancement of a TIN or contour model may be desirable for storage or display purposes, these conversions are not considered in this context.

Interpolating fields from numerical field models

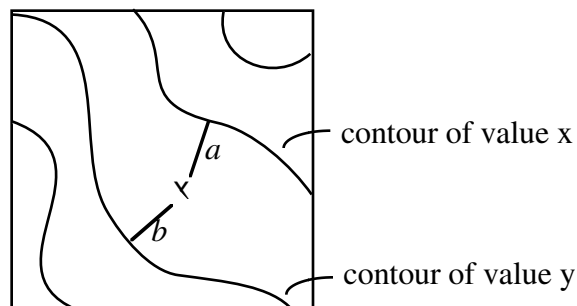
Interpolation of a field from a discrete data model can be thought of as the process of finding the value of the field at an infinite number of locations. Therefore to determine the appropriate interpolation procedure for a particular spatial data model we need to find a technique that allows us to find the value at any location. TINs provide a complete description of a surface, so no interpolation is required to construct a continuous surface. Cellgrids and polygons also provide a continuous representation in the spatial sense but the values change abruptly at region boundaries. Therefore, it is not necessary to devise a technique by which the value at any location can be found. Although it is clear that this stepped surface is not identical to the original surface, we can use the models' assumption that the value within each region is representative of all locations covered by that region to suggest that the best guess of a value at any location is that of its enclosing cell or polygon.

On the other hand, near the boundaries of source regions, it may be more appropriate to assign a value intermediate between the value of adjacent regions. Since constant piecewise models do represent continuous surfaces in which values of points near region boundaries will tend toward the value of neighboring regions:



this approach may provide more accurate estimates of the value on the true surface. Thus, in the case of cellgrid interpolation, it may be reasonable to calculate point values on the surface as a distance weighted average of the center points of each cell. Some form of averaging is certainly necessary for deriving the values of points on the surface which fall exactly upon the corners or boundaries of source regions:

For the other data models, interpolating fields involves finding the values at a number of points. Although contour models do describe a smoothly varying surface, values are only known on the contour lines. Many algorithms for finding values between contour lines are available (Schut, 1976, Weibel and Heller, 1991).



Conceptually, finding the value at any location on a contour model may be determined by calculating the distance of a point from the adjacent contour line(s) and using linear interpolation along these lines to determine the intermediate value which should be assigned to the point. The value at the point will be

$$z = \left(\frac{a}{a+b} \right) (y-x) + x = \frac{ay+bx}{a+b}$$

Interpolation of a surface from points (pointgrids or irregular points) has a long history and there are almost as many techniques for interpolation as there are applications (e.g. see Lam, 1983 and McCullagh, 1988):

Numerous algorithms for point interpolation have been developed in the past. But none of them is superior to all others for all applications, and the selection of an appropriate interpolation model depends largely on the types of data, the degree of accuracy desired, and the amount of computational effort afforded. (Lam, 1983, p. 130)

Selection will also depend upon the phenomenon being modeled, the data model, the available software and the knowledge and experience of the modeler. Techniques include distance weighting, Kriging, splines, interpolating polynomials, Fourier series and least-squares. Therefore, the generic term, point interpolation, is used here to refer to the spectrum of techniques available. The selection of a specific technique will be made either by 1) the person who compiles the data and encapsulates an appropriate technique with the dataset, 2) the modeler who will make a selection based on knowledge of the phenomenon, the data or the application, or 3) the software through limitations on available techniques.

Sampling from numerical fields

As described in Chapter 4, each of the spatial data models provides a different representation of the spatial variation of a continuous phenomenon. For the constant piecewise models, the value recorded for each cell or polygon represents the average of all values within that region in the field. Such averaging is performed automatically by remote sensors with spatial resolutions of a given pixel (cell) size. However, for fields that must be measured by taking samples on the ground, values cannot be collected everywhere. In order to determine average values for the regions in a constant piecewise model, it is necessary to determine the mean value of a sample of values falling within the region. Thus to generate a constant piecewise model, we need to begin with a dense set of point values on the surface.

Sampling a field for representation as a TIN requires selection of the critical points on the surface that upon triangulation produce planes approximating the true surface. When many points have been measured, selection of critical points can be done automatically (Kumler, 1992). Construction of a contour model requires a dense net of point values so that the selected contour lines can be threaded between the points as accurately as possible (Yoeli, 1984). Finally, creating point models from fields is the most direct, since the surface is simply sampled at any location for which a value is required by the model.

Resampling numerical field models

Having considered the two stages independently, it is important now to consider the combination of interpolation and sampling from one specific model to another since there are certain aspects that require modification as the conversion proceeds through the intermediate representation of continuity. Some of the procedures described above can or must be

modified, combined or simplified in order that the conversion proceed as directly as possible with the least loss of information. Note again that no conversions to contour or TIN models are considered.

In the following section, it is assumed that each source data model incorporates all the available information about the true field. Thus, if we resample from a contour model to a cellgrid, for example, the assumption of constant variation between contour lines allows us to assign values interpolated between lines to the target cells. This may, in fact, be seen to be creating data since the cellgrid is likely to have several spatial elements between each contour line. However, if we accept the assumptions inherent in the contour model, the target cellgrid becomes simply an alternate representation of the same field. Rather than adding data, we have lost information implicit in the precise location of the contour lines. If the basic assumptions of a particular data model cannot be assured (for example, a contour model constructed from very sparse data), then additional knowledge must be used to determine the appropriate conversion procedure.

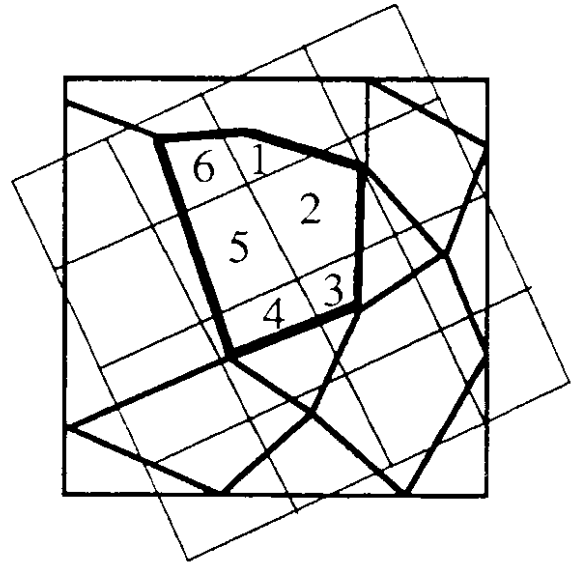
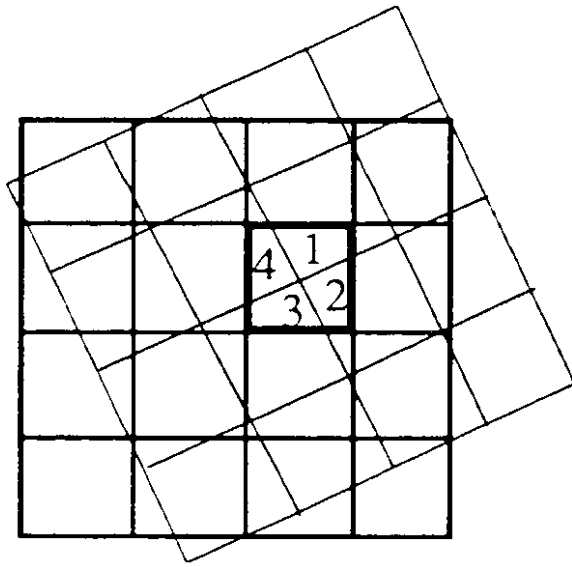
From constant piecewise models

From constant piecewise models to constant piecewise models. Since cellgrids and polygons represent a continuous surface, creation of new models from these source models requires rearrangement of the cell or polygon boundaries (see diagram below). A method for estimating a new value from the values in a set of whole or partial regions on the original (stepped) surface is required. This involves the use of *areal interpolation*. As for point interpolation, there are many different techniques for areal interpolation (Goodchild and Lam, 1980, Lam, 1983, Flowerdew and Green, 1992). However, many of these, including volume-preserving techniques such as pycnophylatic interpolation (Tobler, 1979b) or methods based on the EM algorithm (Flowerdew and Green, 1992), have been devised for categorical and count data and are unnecessary here. (Image on following Page)

In fact, simple *areal weighting* is sufficiently precise and conceptually consistent for fields. It may be expressed as (Flowerdew and Green, 1992):

$$z_t = \sum_s \frac{z_{st} a_{st}}{a_t}$$

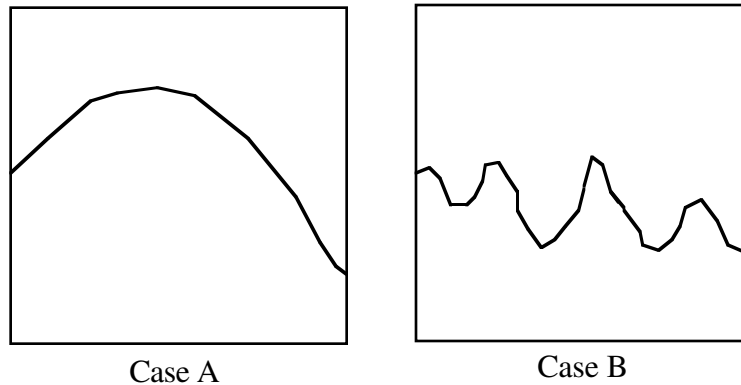
where t is the target zone, s is a source zone, st is the intersection of the target and a source zone, z is the value in the indicated zone, a is the area of the zone and



$$z_{st} = z_s$$

From constant piecewise models to point models. With the cellgrid or polygon model as the representation of the continuous surface, values for points can be sampled directly from the surface. It must be kept in mind, however, that when created from one of the constant piecewise models, point values will not be indicative of the value at the location, but instead will be the average of the neighborhood. As a result, derivative point based models are not conceptually the same as original point based models.

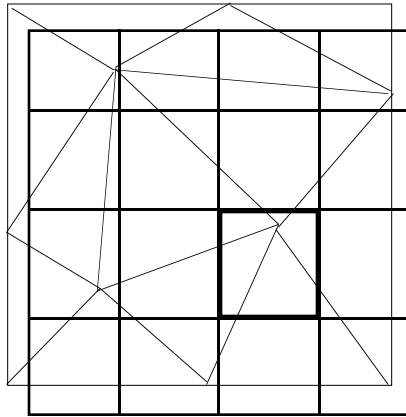
Unfortunately, there are several complications with region to point resampling. The provision for distance weighting values near the boundaries of regions was discussed above. As well, for models in which polygons or cells are large in comparison to the point spacing, sampling will produce large patches of same valued points which do little to reflect the true nature of the underlying surface. Selective use of an appropriate spatial filter may provide some smoothing across former region boundaries and produce a more realistic representation of the true surface. The decision whether to use a filter or the distance weighting approach should depend upon knowledge of the smoothness of variation in the phenomenon. For example,



if the surface is very smooth, as in case A above, filtering or weighting may be appropriate. In case B, however, irregularity of the surface invalidates the underlying assumptions of smoothness inherent in filtering and weighting techniques in such a way that it is clear they are inappropriate for this type of field. For this latter case, alternate techniques must be used. Knowledge of the surface form must be used to determine appropriate conversion procedure(s). Fortunately, as it will be demonstrated later, it is possible to encapsulate this knowledge with the source data model.

From TINs

From TINs to constant piecewise models. TINs provide a complete description of a continuous surface. Therefore, to sample to constant piecewise models it is necessary to perform areal averaging over the area covered by each target region (see diagram on the next page). Conceptually, areal averaging would be done by finding the average value within each piece of triangle in each target region and then carrying out areal weighting. To find the average value within each triangle piece it is necessary to divide the pieces into smaller triangles for which the centroid and its related value can be found. The values and areas of these smaller triangles can be used as the basis for the areal weighting.

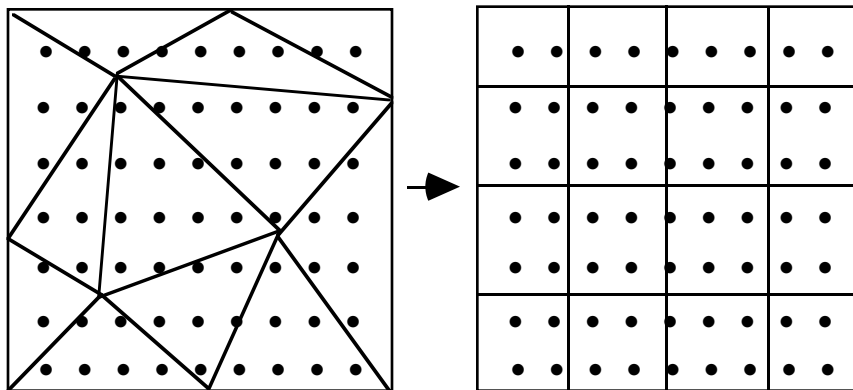


From TINs to point models. TINs to point models allows an exact TIN-to-point interpolation procedure which calculates the linear equation on each triangle face from the rectangular coordinates of the nodes and their values. The equation which results may be used to find the value at any point which falls on that triangle face.

From contour models

From contour models to constant piecewise models. Since the surface represented by a contour model cannot be simply described by a simple mathematical relationship, it is necessary here to resort to an intermediate data model, a pointgrid, though which this conversion is performed. Using a contour-to-point interpolation procedure, a dense grid of points is created. This provides information about the variation of the value of the field within each target region which can be used to estimate the required representative value. This estimation from a pointgrid requires the use of *point averaging*.

The validity of point averaging depends upon the number of points that fall within each target region though the critical minimum number of points that should be used is not clear. In many cases, the number of points from a set of interpolated points which fall within each target region will vary depending on the configuration of the interpolated points and the target data model:



It is important to recognize that averaging a different number of points for each target region may affect the statistical validity of the target model.

Given a specified minimum number of points to be interpolated for each target region, the spacing of the grid of interpolated points can be determined directly from the dimensions of the cells or from the size of the polygons. For example, let us assume that a minimum of 4 points is sufficient. Thus, for a target cellgrid, if Δx and Δy are the width of grid cells in the two perpendicular directions, the interpolated grid should have spacings of $\Delta x/3$ and $\Delta y/3$. For a target polygon model, the grid spacings in both directions should equal:

$$\sqrt{\frac{a_p}{4}}$$

where a_p = area of the smallest polygon. For polygons which are roughly rectangular, this should place at least 4 grid intersections in all polygons. However, since it is likely that polygons will not be rectangular, it will be necessary to check that at least 4 points fall in each polygon. For any polygon without the minimum number of points, a new, denser grid may be calculated for the area covered by that polygon. While this point-in-polygon checking procedure is computationally intensive, algorithms for its solution do exist (Burrough, 1986). Hence it would not be necessary to obtain additional input from the user.

From contour models to point models. Contour to point conversions use the contour-to-point interpolation procedures described above to combine interpolation and sampling from the continuous surface described by the contour model.

From point models

From point models to constant piecewise models. If the number of points in a point model is much greater than the number of cells or polygons, resampling of point models can proceed directly by calculating the mean of each set of points falling in each region. However, if this is not the case, then it will be necessary to invoke an interpolation procedure which can be used to determine a continuous surface represented by a dense net of point values which can be averaged within each region. As described above, the relationship of the density of the point model to the size of cells or polygons can be determined automatically without additional input.

From point models to point models. In this case, interpolation and sampling will be done simultaneously by interpolating point values in the source model for destination points in the target model. The appropriate interpolation procedure may be determined by the characteristics of the source model as well as the relative densities of source and target.

Interpolating fields from categorical field models

We now turn to a consideration of the conversion of fields expressed using categorical data. Much of what has been discussed above applies here although several differences arise. Recalling that not all spatial data models are appropriate for categorical data, this discussion is restricted to consideration of constant piecewise models and point models.

Whether numerical or categorical, cellgrids and polygons present a continuous surface. The blocky nature of the constant piecewise models is the only possible representation of fields measured on categorical scales since the data is discrete and cannot be continuous across boundaries. Thus the field represented by these models is the closest approximation to reality that can be provided. Near zone boundaries, it may be appropriate to assume that a transition occurs with the class at any point being determined by a binary

probability function. However, since the assumption is that the regions in a piecewise model depict the most common class within the zone, it is equally appropriate to assume that the most likely class at any location is that indicated by the class of the zone in which it falls.

Continuous fields of categorical data cannot be generated from points using any of the mathematical interpolation techniques since values cannot be "interpolated" between classes. Geometric techniques must be used. These techniques involve proximal mapping through the construction of Thiessen polygons (also known as Voronoi and Dirichlet graphs, diagrams or tessellations). By constructing Thiessen polygons each point is associated with the region of the plane closer to it than to any other point. By assigning the value of the enclosed point to the entire polygon, the assumption is made that the closest known point provides the best information about any unknown point (Burrough, 1986, Boots, 1986). (Note that this assumption may be invalid for fields which are known to be highly heterogeneous.) Many implementations of algorithms for the construction of Thiessen polygons exist (Aurenhammer, 1992). Once converted to Thiessen polygons, the characteristics of these representations of fields are those of constant piecewise models and conversions proceed accordingly.

Sampling from categorical fields

Once the continuous categorical fields exist, sampling for cellgrids and polygons must be performed by partitioning source zones into target zones, the first step in areal weighting. However, for this categorical data, it is not possible to use the full areal weighted procedures to determine a representative value for the target regions since categorical data cannot be mathematically manipulated. Thus, to determine the value of a region in the target model it is necessary to devise a set of rules which can be used to determine which class should take precedence in the new region. There are several different types of rules that may be invoked, including:

- the source class covering the largest portion of the target region becomes the new value.
- precedence rules determine which class takes precedence if it appears anywhere within a target region. This approach might be taken when the classes indicate the level of expected impact; if any part of a target region contains a portion of any source region classified as high impact, the entire target region is classified similarly.
- proportions are retained. This requires a different data structure for the target model which can retain several pairs of values with each pair indicating the class and the proportion of the region covered by that class.

Sampling categorical fields for point models is simply a matter of determining the class of the area in which the point falls. No mathematical manipulation is possible.

Resampling categorical field models

The options available for resampling of categorical models are limited by the restriction on suitable data models. Since constant piecewise models give the best representation of the surface, only point models need to be interpolated through construction of proximal regions to create complete representations as piecewise models. Sampling is performed through partitioning and rules as described above or by point sampling.

Problems with converted fields

It is worthwhile to reemphasize here a point made earlier regarding the veracity of spatial data model conversions. Each conversion creates a new model of the reality being represented. While a conversion process generally causes some loss of information, the number of spatial elements may actually increase. This may cause an apparent change in the smoothness of the variation and an apparent addition of data. We have argued that an increase in the number of spatial elements does not increase the degrees of freedom in the data set and that it may be useful to attach relevant information about the character of the source data model to the target model. However, we have also noted that these changes are artifacts of the processing, and in the absence of expert knowledge about the field, they do not substantially change what is known about the relationship between the model and reality. The effect of such transformations of the representation are important, but their effects are manageable if the relevant characteristics of the datasets are recognized.

The conversion matrices

Based on the preceding discussion, we can summarize these various procedures in a pair of *conversion matrices*, shown below in Tables 4 and 5. While organized for the purpose of conceptual clarity, the procedures outlined in these matrices could be implemented as a specific set of decision rules and operations within any computer programming language. Note that in Table 4, TINs and contours cannot be used for predefined destination models since the structure is determined by the phenomenon represented. The conversions for TINs and contour models outlined here may be performed to reduce data volume or for visualization purposes. They are included here for completeness but are shaded to indicate that they will not be used for mathematical manipulations.

Table 4 - Summary of spatial data model conversions for numerical data

| To From | Cellgrid | Polygon | TIN | Contour | Point grid | Irreg pt |
|------------|------------------------|------------------------|------------------------------|------------------------|---------------|--------------|
| Cellgrid | areal weights | areal weights | point interp & triang | point interp & contour | point sample | point sample |
| Polygon | areal weights | areal weights | point interp & triang | point interp & contour | point sample | point sample |
| TIN | areal weights | areal weights | add or remove nodes & triang | thread contours | point interp | point interp |
| Contour | point interp & average | point interp & average | select nodes & triang | remove or add contours | point interp | point interp |
| Point grid | point interp & average | point interp & average | select nodes & triang | contour | point interp | point interp |
| Irreg pt | point interp & average | point interp & average | select nodes & triang | contour | point interp | point interp |

Table 5 - Summary of spatial data model conversions for categorical data

| From \ To | Cellgrid | Polygon | Point grid | Irreg pt |
|------------|-----------------------------------|-----------------------------------|-------------------------|-------------------------|
| Cellgrid | partition & apply rules | partition & apply rules | point sample | point sample |
| Polygon | partition & apply rules | partition & apply rules | point sample | point sample |
| Point grid | Thiessen, partition & apply rules | Thiessen, partition & apply rules | Thiessen & point sample | Thiessen & point sample |
| Irreg pt | Thiessen, partition & apply rules | Thiessen, partition & apply rules | Thiessen & point sample | Thiessen & point sample |

The computational structure

As noted earlier, continuous mathematics requires some form of discretization in order for computation to be performed on a computer. Frequently differential equations are solved on the computer through the use of finite difference grids. The dimensions of these computational grids are defined with respect to the scale of the processes under study. Thus global circulation models with grids of 10° latitude and longitude are concerned with processes that operate at scales of 100's of kilometers while soil erosion models based on grids of cell width in the order of 100m are concerned with processes operating near the hectare scale.

Virtually all environmental models will be designed for implementation in a specific spatial data model with predefined dimensions. This is the *computational structure* of the mathematical model. Since the definition of the computational structure depends upon the processes being studied and not the spatial data available to the modeler, there is no expectation of any fixed relationship between the computational structure and the spatial properties of the field variables the model uses. As we shall see, since this is the structure in which all mathematical calculations of the model will be performed, it becomes the default structure, the one towards which all others will tend during mathematical and proxy building operations.

D. ARITHMETIC OPERATIONS

Having examined the fundamental assignment operator and its related conversions, we now briefly review the range of other operators and functions that may be used in mathematical equations. Categorical fields can only be used in Boolean arithmetic, so, with the exception of the Boolean section, the following applies only to numerical fields.

Binary arithmetic operators

Binary arithmetic operators combine two numbers through the simple operations of addition, subtraction, multiplication and division. If one variable is scalar and one a field, the result of the operation is to increase or decrease field values uniformly according to the specified operation. Adding and subtracting 0 and multiplying and dividing by 1 create identical fields. Multiplying by 0 creates a *null field*, one in which the value everywhere is 0. Division by 0 cannot be done. If both variables are fields, these arithmetic operations can be visualized as combining the values of the variables for each location in space:

$$\text{WATER_DEPTH} = \text{IMPERV_LAYER_DEPTH} - \text{WATER_TABLE_DEPTH}$$

In abstract mathematics, binary arithmetic operations can be performed without restriction on any real or integer numbers (arithmetic with vectors has different conventions and is treated separately below), with the two exceptions: division by 0 which is not permitted and integer division which may create a real number.

Unary operators

Unary operators operate on a single value to create a derivative value. These include negation, absolute value, log, roots and exponentiation. Exponentiation may also be seen as a special case of multiplication operating on a single value. With a field, negation creates a reflection of the surface across the 0-plane, while absolute value reflects only those values which fall below 0.

Vector arithmetic operations

Arithmetic can also be performed on vector data. Some of these operations are defined as follows. If $a = \langle a_1, a_2 \rangle$, $b = \langle b_1, b_2 \rangle$ and c is a scalar,

| | |
|--|--|
| addition: | $a + b = \langle a_1 + b_1, a_2 + b_2 \rangle$ |
| subtraction: | $a - b = \langle a_1 - b_1, a_2 - b_2 \rangle$ |
| multiplication by a scalar: | $ca = \langle ca_1, ca_2 \rangle$ |
| dot or inner product: | $a \cdot b = a_1b_1 + a_2b_2$ (a scalar) |
| cross product cannot be done in 2 dimensions | |

As with binary arithmetic operations on scalar data, these operations when performed on field data assume the combination of the corresponding vector values for each location in the field.

Trigonometry

While not strictly arithmetic operations, like arithmetic, trigonometric operations require scalar values. Radians and degrees, which can be exactly transformed one to the other, are the only permissible arguments in trigonometric operations. There are no restrictions on the execution of trigonometric operations on field variables which are given in radians or degrees. Standard language compiler operations are suitable. Trigonometric operations are common in models which consider intensity of solar radiation (using a field variable of aspect) or rate of flow (using a variable of slope).

Boolean arithmetic

Both categorical and numerical fields can be manipulated through Boolean arithmetic. The operators of Boolean arithmetic provide tools for working with categorical data and are useful for determining the flow of operations to be performed in a mathematical model.

E. OPERATIONS INVOLVING DIFFERENT SPATIAL DATA MODELS

Even simple binary arithmetic operations on field variables are not necessarily simple. However, if A and C are spatially equivalent field variables and b is a scalar, the equations:

$$\begin{aligned} C &= A + b \\ C &= A - b \\ C &= A / b \\ C &= A * b \end{aligned}$$

perform the specified operation on each value in the data set comprising the right hand variable and place the result in the data set comprising the left hand variable. If all three variables are field variables and spatially equivalent, then the arithmetic is performed directly on the values in each spatial element and the result is placed in the corresponding element of the left hand variable. Recall that a spatial element, as defined in Chapter 4, is the basic geometric component of a spatial data model (e.g. point, cell, polygon).

Difficulties arise when the field variables are not spatially equivalent. In this case, conversion must be performed so that 1) the operation on the right hand side can be performed and 2) the answer can be placed in the left hand variable. The question now is which conversion should be performed first. Consider the following case. The computational structure of the model is a cellgrid. We wish to develop a proxy variable for monthly precipitation. The only data available is a contour map of total annual precipitation and scattered weather station records detailing the percent of total annual precipitation that falls in each month. Hence we have a contour model which must be multiplied by an irregular point model to create a cellgrid. Do we convert the contour model to irregular points, multiply and then convert the result to the cellgrid, or do we convert both the contour model and the irregular points to the computational structure before the multiplication? Clearly a set of priority rules is needed.

Intuitively, it is possible to develop a set of rules for conversion. Since the most convenient structure for most mathematical and spatial operations is the grid, a simple rule might be that all variables are converted into grids before calculation is performed. However, this may lead to an unnecessary loss of information, particularly if the target variable is not a grid. Figure 1 shows a wide range of such combination operations and indicates the model to which variables should be converted before the arithmetic operation is carried out. From such an analysis, a set of rules can be devised. The decisions upon which many of these rules are based depend upon the relative size or spacing of the spatial elements. This concept is expressed here as *density* which is defined as *the number of spatial elements per unit area*. The following is an example of such a set of rules, listed in order of priority:

1. If both sources are spatially equivalent, use that structure.
2. If each source is either a TIN or a contour model, use the target structure.
3. If all variables are spatially nested grids, use the densest grid.
4. If the target is spatially identical to one source, use the target structure.
5. If one source is a TIN or contour, and the other is a grid, use the grid.
6. If all are of approximately the same density, use the target.
7. If only one is points, use points, unless the points are very sparse.
8. Use the densest structure. If there is a tie, use the target.

While there is opportunity for experimentation to devise the perfect set of priority rules, it is clear that implementation of any such set of rules can be done without input from the modeler.

FIGURE 1 - Combining spatial data models

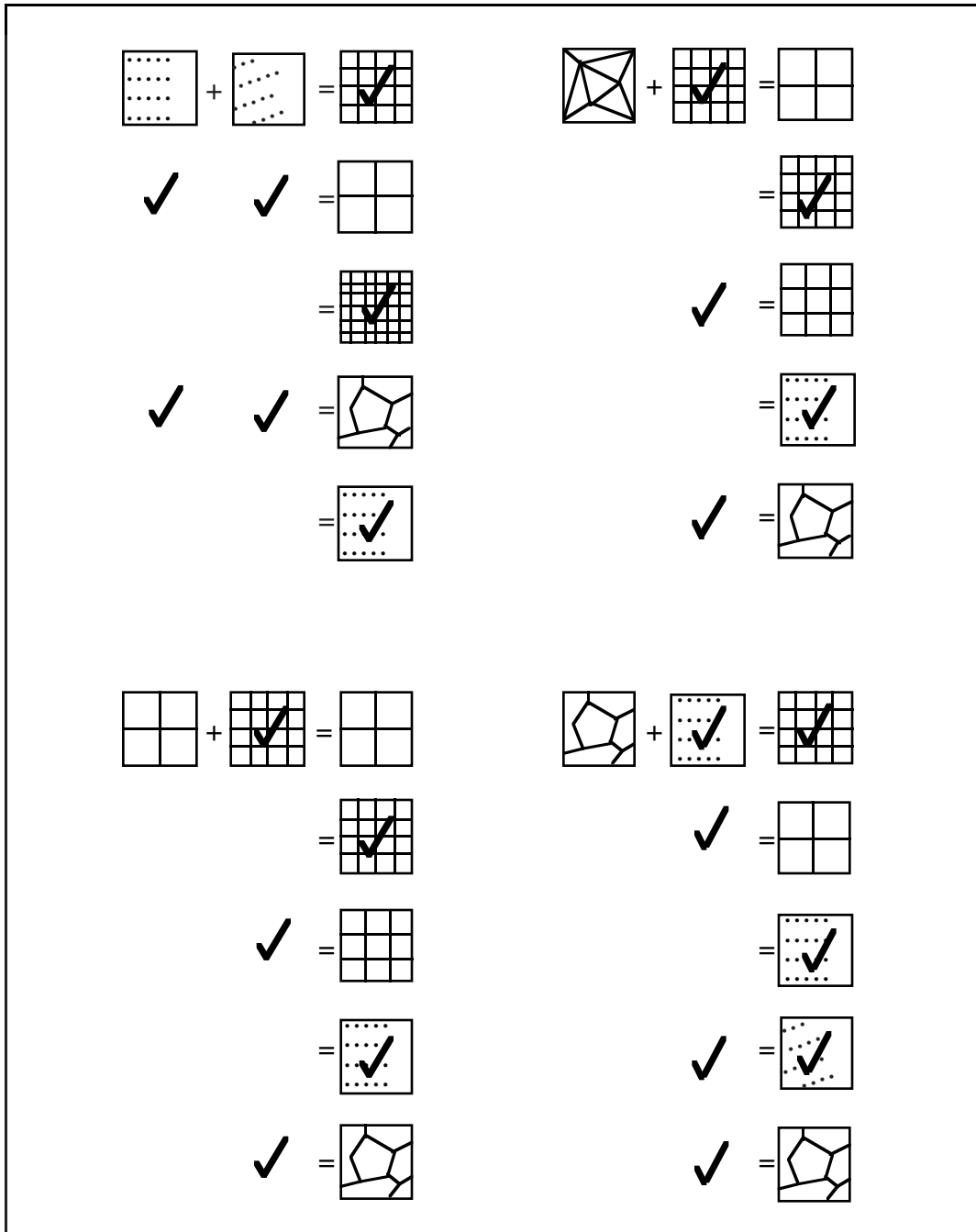
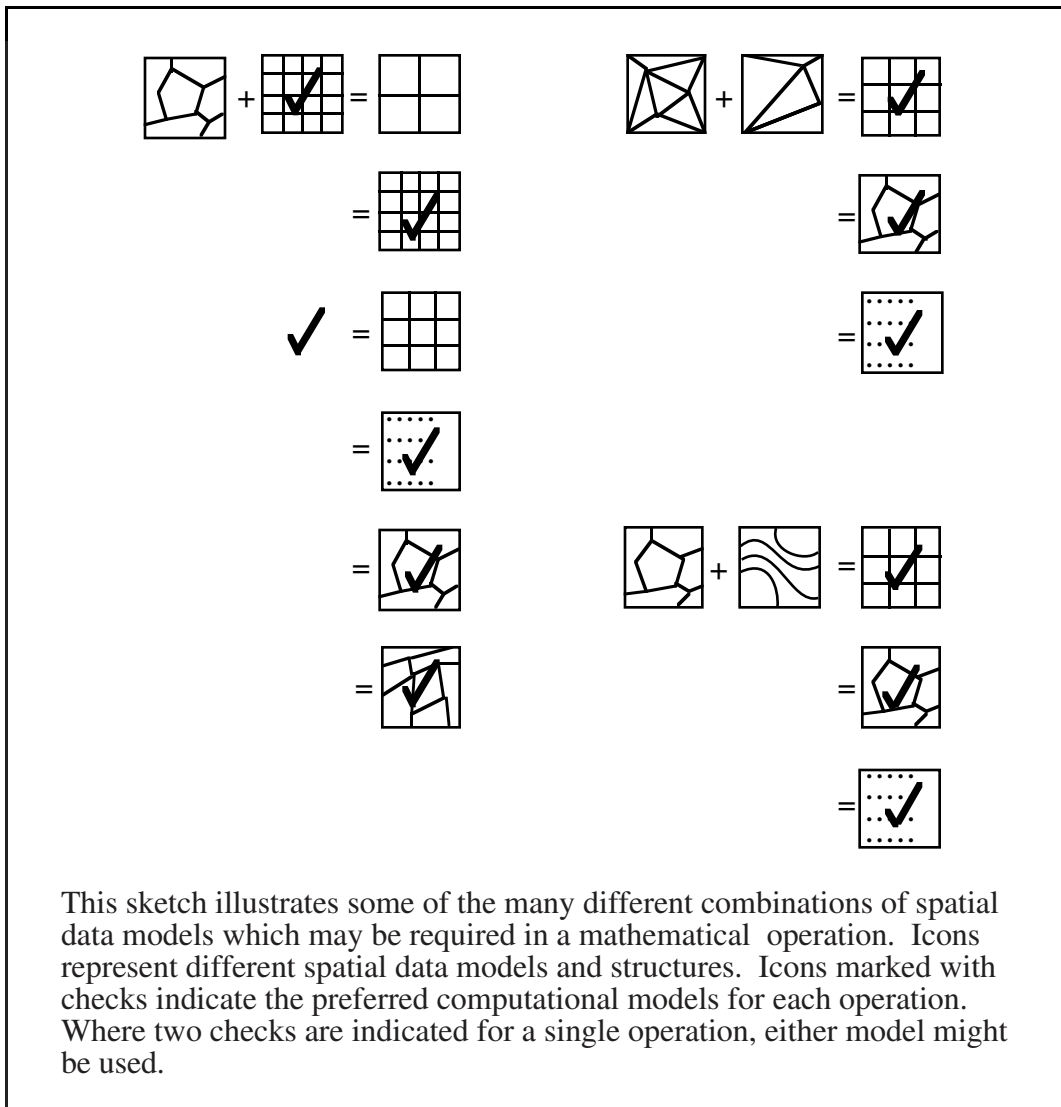


FIGURE 1 (cont)



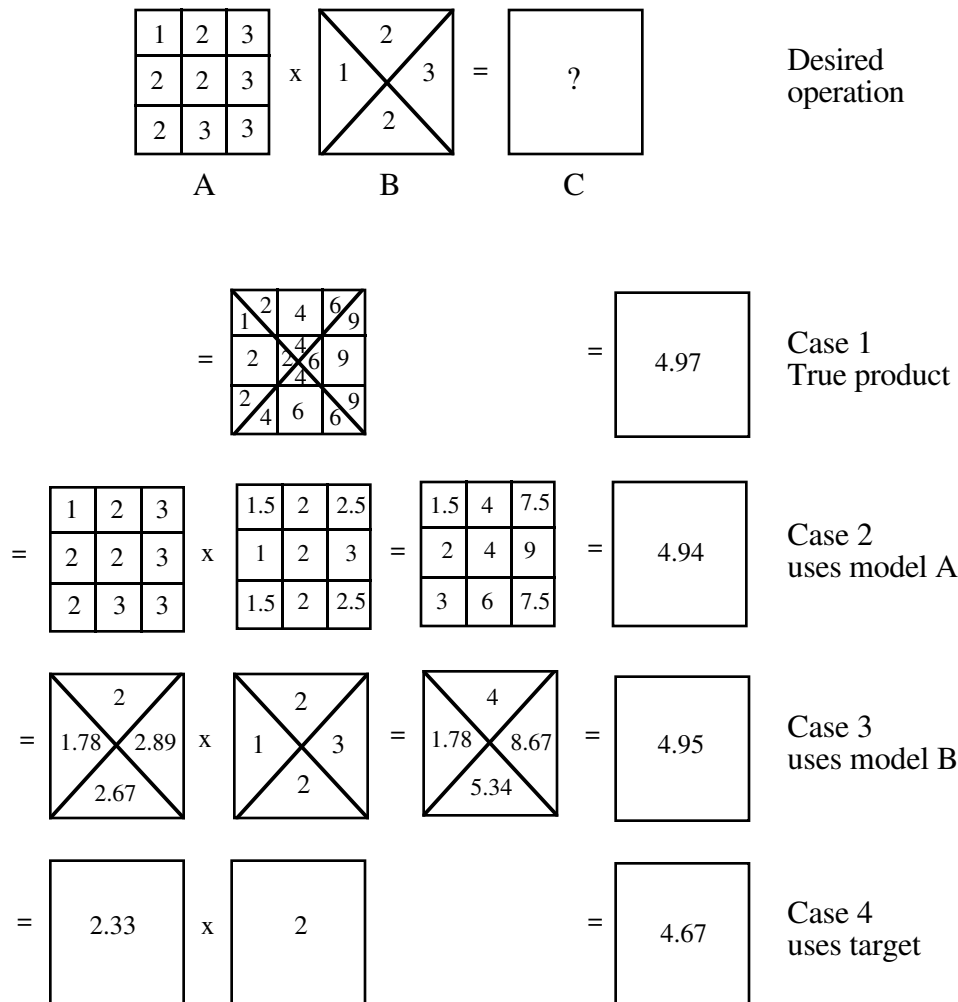
Incompatible scales

The foregoing discussion has ignored the fundamental issue of incompatible scales, a multi-faceted and complex topic. Since there are processes operating at many different scales in any natural system, the sampling frequency and the density of a spatial dataset will determine which processes are represented and can be modeled. Mixing datasets based on different sampling frequencies and spatial densities may lead to confused or misleading modeling results. This aspect of incompatible scales is a recurring theme in all the modeling sciences but, since the nature of its resolution is dependent upon specific modeling domains and processes, it is beyond the scope of this research.

If it can be assumed that datasets which are to be used together depict the appropriate level of variability of the fields being represented, there are some generic issues of scale incompatibility that may be addressed. For example, suppose a model is being developed which will use a finite difference grid of 10' latitude and longitude and one of the necessary

input variables is only available with a grid spacing of 1° latitude and longitude. For statistical purposes, it may be suggested that it is inappropriate to significantly densify the resolution of this dataset so that it can be used with another dataset. This is because the number of degrees of freedom, normally determined by the number of data elements, remains that of the original dataset, although the number of data elements increases significantly through densification. Since the new points or cells have been created from the original data, they do not represent separate independent observations (i.e. degrees of freedom). Thus, the available degrees of freedom are not determined by the number of spatial elements in the densified variable but rather by number of elements in the original variable. This implies that each variable which results from an operation in which one or more of the original variables has been densified should carry an indication of its true degrees of freedom. While this is not provided for in the syntax described later, it is a desirable addition for future versions.

However, for spatial data model conversions, it is often the densest model which determines how operations should be performed and how accurately the results can be determined. As the following figure demonstrates, when undertaking data model conversions, restricting operations to only the sparsest models can lead to unnecessary loss of information.



Here, cellgrid A is multiplied by polygon model B to produce the sparse polygon (or cellgrid) in C. There are several ways these different data models can be integrated for this

calculation. Case 1 gives the true result since the two source data models are overlaid to produce an intermediate dense model which includes all recorded variation. An area weighted average of these values produces the final result. The other three cases demonstrate each of the alternatives for selection of a computational model from the source or target model structures. In Case 2, model B is converted to the cellgrid of model A before multiplication and averaging. Case 3 uses model B as the computational structure. In Case 4, both target models are converted, by area weighted averaging to the target model. It is clear that using the sparsest model, the target (Case 4), as the computational structure produces the least accurate estimate of the product. Thus, while densification should not be used to simulate an increase in the available information, it is appropriate during model conversions.

F. DESCRIPTIVE STATISTICS FOR FIELDS AND FIELD VARIABLES

There are several statistical measures that are used to summarize or describe sets of observations. These measures include those of central tendency which give an indication of the typical values observed and those of dispersion which suggest the degree of variability in the set (Silk, 1979). For sets of observations, the mean value is the sum of all values divided by the number of values, maximum and minimum values are simply the largest and smallest values in the dataset, and the standard deviation is the square root of the sum of the squares of differences between each value and the mean divided by the number of observations.

While these measures are simply defined for a set of discrete values, it is possible to express some of these summary statistics in terms of derivatives and integrals of continuous functions. For example, a local maximum value of a function occurs at those points where the first derivative equals 0 and the second derivative is negative. The mean value of a two dimensional function $f(x,y)$ on the intervals $x=[a,b]$ and $y=[c,d]$ is defined as

$$\bar{f} = \frac{1}{(b-a)} \frac{1}{(d-c)} \int_a^b \int_c^d f(x,y) dx dy$$

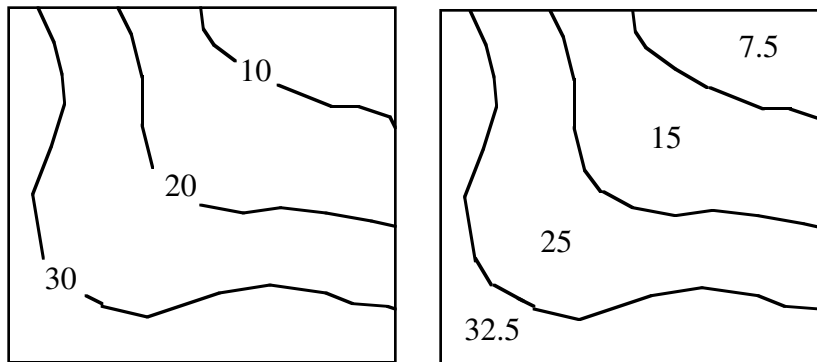
However, when the fields are physical fields that cannot be described by simple functions, the definition and determination of such descriptive statistics is extremely difficult. Fortunately, the discrete data sets provided by spatial data models provide us with a means to devise representative statistics for fields. In this section we review the methods by which simple descriptive statistics can be devised for fields represented by the various spatial data models. It must be remembered that many sophisticated statistical measures for analyzing spatial data exist (cf. Haining, 1990 and Anselin, 1989), including measures of spatial autocorrelation and indices of spatial pattern. While these may often be better descriptors of the datasets themselves, spatial statistics and for that matter, spatial effects, have yet to become widely incorporated into mathematical models. Therefore, we confine the discussion below to those statistical measures which are likely to be used in mathematical models of environmental processes. These measures can be used only with numerical fields.

Mean

While the mean is difficult or impossible to calculate for a continuous field which is not described by a simple function with a definite integral, it is possible to calculate means for fields represented by discretized field variables. Given that constant piecewise models are representative of the continuous surface and that the value for each region is representative of all values in that region, the region value can be considered the local mean. The mean of the entire dataset in a constant piecewise model is calculated as the sum of each

region value times its area divided by the total area of the study area. Similarly, fields represented by TIN models also can be averaged by finding the mean of each triangular facet (the value at the centroid) and calculating the area weighted average across the field.

Assuming that the variation between contour lines is smooth, means for fields represented by contour models can be calculated by constructing polygons between adjacent contour lines. Contour-bounded polygons are defined as contiguous areas bounded but not intersected by contour lines and, possibly, the border of the area under study:



Each polygon is assigned a value based on the values of bounding contour lines. The mean of the entire surface is calculated using the area weighted average.

Since pointgrids are sets of points sampled on a regular grid and thus independent of any phenomena being sampled, a simple arithmetic mean of all values in a pointgrid dataset provides a reasonable estimate of the true mean of the field in the same way that the mean of a sample is used to estimate the mean of a population. This approach requires an assumption that there is no systematic spatial variation with a period equal to the grid spacing. For irregular points, the mean of the values of the set of points can also provide an estimate of the surface mean if the points are assumed to be random samples of the surface value. However, spatial autocorrelation may make this approach invalid if the points are not evenly distributed over the surface since points sampled close together will weight the mean to the value in that area. Thus, if irregular points are clustered, it is more appropriate to construct Thiessen polygons around each point and continue as if the dataset were a polygon model (i.e. use the area weighted average) or to triangulate the surface and continue as for TINs. The choice of which process to use depends upon whether the points can be considered representative of their neighborhoods (in which case Thiessen polygons are appropriate) or critical points on the surface (in which case triangulation is preferred). Information which is used to make this choice may be encapsulated with the dataset or, in the case of possible spatial autocorrelation problems, a test for spatial randomness (Silk, 1979) may be invoked.

Of course it is quite possible that the mean is not stationary over the surface. In this case more sophisticated approaches to calculating means may be necessary, including trend surface and Fourier analysis.

Sum (Integration)

When working with discrete sets of count data such as cancer cases by county or votes by electoral district, if we wish to know "How much?" is represented by the complete dataset, we simply add together all the elements of the dataset. When working with fields such as rainfall depth or nitrogen content, "How much?" requires integration of the surface

function over the area under study. In two dimensions, the sum of a defined function $f(x,y)$ over the intervals $x=[a,b]$ and $y=[c,d]$ is

$$\text{sum}(f) = \int_a^b \int_c^d f(x,y) dx dy$$

As any calculus student knows, the result of the integration operation is the area under the curve. In two dimensions the result is a volume.

When working with fields which are not described by functions, integrals can be found by dividing the study area into regions for which a simple mean value can be calculated. The mean of each region is multiplied by the region area to give the region volume and the total of all region volumes is the integral over the study area. Therefore, in order to calculate the integral of any spatial data model we need simply to find the set of regions for which we can determine mean values. For cellgrids and polygons, these regions are defined by the model. The mean of a region defined by a TIN triangle is the value at the centroid of the triangle. Mean-valued polygons can be devised from contour models as described above. The point models require the initial construction of regions. This is achieved through the construction of Thiessen polygons or through triangulation with the choice based on whether points are representative or critical as described above. Once regions have been constructed from the point models, integration proceeds as outlined for polygon or TIN models.

The final output of many mathematical models is a final single scalar value which is used to represent the integration of individual contributions across the entire study area. Since most computer languages include the clearly defined explicit operation of `sum` which adds up all values in the specified set, we propose that a new operation, `integrate` replace `sum` when working with field variables. Here it is abbreviated to the form

$$a = \text{integ}(B)$$

where a is a scalar and B is a numerical field. As well, we often wish to know the individual integrals within specific regions such as counties or soil types. Therefore, it is useful to add a second, optional categorical field to the `integ` function which defines the regions over which integration is to be calculated.

$$a = \text{integ}(B, C)$$

where C is a categorical field.

Maximum and minimum

The maximum and minimum of a field described by a spatial variable can generally be considered to be the maximum and minimum observations in the dataset. Since the surface is completely described in constant piecewise models, these descriptive measures are simply the extreme values in the set of spatial element values. For TINs, extremes are extracted from the set of node values. For point models, sophisticated interpolation procedures such as splines or fitted polynomials could be used first to define the surface from which extremes will be extracted. However, since many of these techniques use the sample points as extremes, and it may be difficult to determine the extreme values of some fitted surfaces, it is also reasonable to extract the maximum and minimum values directly from the set of point values.

Contour models do not yield exact extreme values since the value of the field varies smoothly across contour lines. Maximum and minimum values must be extrapolated from the value of the highest and lowest contour lines contained within the study area. While it is certain that these extreme values lie within one contour interval of the extreme valued contour lines, there is no additional information within the spatial data model itself available to produce accurate estimates.

Standard deviation

The standard deviation of a function, $f(x,y)$ may be defined as

$$\sigma = \sqrt{\frac{\int_a^b \int_c^d (f(x,y) - \bar{f})^2 dx dy}{\int_a^b \int_c^d dx dy}}$$

where \bar{f} is the mean of function f . Since this is difficult to determine for a field not defined by a simple function, it is necessary to resort again to the use of the discrete representations.

To calculate standard deviation for a constant piecewise model, each squared difference must be weighted by the area of each region:

$$\sigma = \sqrt{\sum_i \left((x_i - \bar{x})^2 \times \left(\frac{area_i}{total_area} \right) \right)}$$

where i is an index value for each region. The standard deviations of TIN and contour models can be found in a similar manner once the mean value for each region has been determined as described above. It must be noted that standard deviations found in this manner will be less than the true value since the local means of each region have already removed some of the true variation.

Pointgrids and irregular points are the only models which can produce a statistically correct standard deviation value. In these models, values represent the true value at a point and the dataset itself can thus be treated simply as a set of observations. As noted above, problems with spatial autocorrelation may arise for irregular point models so it may be preferable to construct Thiessen polygons or to triangulate and proceed as for polygons or TINs.

G. INFERENCE STATISTICS

Finally we consider briefly the area of inferential statistics as a operation in environmental models. Strictly speaking, inferential statistics will not form a part of the models themselves, but they are important in all other aspects of mathematical modeling including model fitting, testing and evaluation, proxy building, visualization of results and importantly in interpolation and sampling.

Due to inherent characteristic of spatial autocorrelation and the resulting lack of independence in spatial datasets, researchers dealing with spatial data often find it necessary to avoid strict parametric approaches to inferential statistics in favor of nonparametric, qualitative and robust procedures. Parametric approaches require assumptions of independence and normality, which is unrealistic for many spatial data sets. However, results

from robust and nonparametric techniques are limited in their generality (i.e. results may not hold for the population from which the sample was drawn) and tend to be very conservative (Anselin, 1989).

As well, many statisticians argue that a dataset comprising a field variable is not the set of random samples required by traditional inferential statistics but rather it is a single instance of a stochastic variable. In fact, it has been suggested that a spatial dataset forms the complete population and as a result, no inferential statistics can be used (Summerfields, 1983, in Anselin, 1989), although Anselin has argued that this condition does not often hold due to "the imperfect nature of measurement, and the inherent error (or noise)" (Anselin, 1989, p. 8). If this is the case, the spatial dataset can be viewed as the result of a set of stochastic processes that have measured the underlying signal along with the stochastic error of measurement (Haining, 1990). It is thus one realization of many.

Geostatistics and the theory of regionalized variables

Fortunately, a new approach to inferential statistics which has recently gained in popularity provides some new tools for dealing with fields. Geostatistics and the theory of regionalized variables have evolved in a number of different disciplines, originally in mining and climatology, but the common driving force behind this development has been the need to develop statistical techniques which work with spatially correlated continuous phenomena.

Fields are traditionally described as functions of location, (Oliver, et al, 1989):

$$z(x, y) = f(x, y) + \varepsilon$$

where $z(x, y)$ is the value of a variable z at location (x, y) and f denotes a deterministic function, ε is a random error term. Using this concept, the use of fitted polynomials to interpolate surfaces from a limited number of sample sites implies that the majority of the variation observed is deterministic. An alternative way to look at the observed variation is to consider it as essentially random but spatially dependent.

Geostatistics is the application of the theory of regionalized variables to spatially distributed data (Journal, 1986). Although Journal suggests

From a theoretical standpoint, little is new in this theory, which borrows most of its models and tools from the concept of stationary random functions and techniques of analysis of variance and generalized least-squares prediction. (Journal, 1986, p. 120)

it provides a "concise, coherent and useful body of theory" (Oliver et al, 1989, p. 268) providing a new set of probabilistic tools which can be used to estimate and describe the spatial variation of phenomena on the earth's surface.

All of geostatistics depends upon the development of a model of the spatial dependence observed in the regionalized variable. This model is the *variogram* which relates the variance of the value of the variable between any two sites to the distance separating them. Thus, the initial stage in any geostatistical analysis is the development of the variogram.

Constructing the variogram

The variogram is constructed from sample data by calculating the variance between sample values for a range of different *lags*, or separating distances. If the spatial variation is

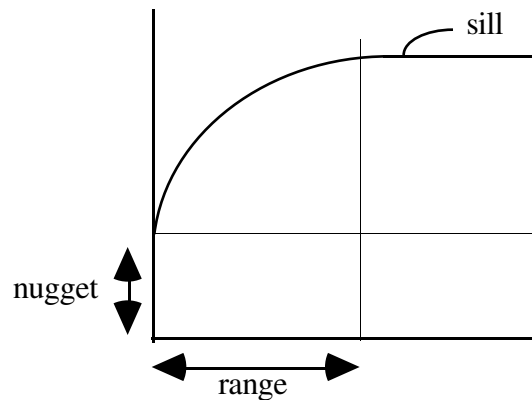
isotropic then the estimate of the semivariance, γ^* , is calculated as (Oliver and Webster, 1991):

$$\gamma^*(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} \{z(x_i) - z(x_i + h)\}^2$$

where $z(x_i)$ and $z(x_i+h)$ are the observed values at x_i and x_i+h , h is the lag, $\gamma^*(h)$ is the estimate of $\gamma(h)$ and $m(h)$ is the number of paired comparisons at that lag. Once the variance has been calculated for several different h , a *sample* or *experimental* variogram is constructed by plotting $\gamma^*(h)$ against h . If the field is anisotropic, this calculation is done individually for several different directions with each set of points treated separately in later manipulations.

The resulting sample variogram consists of a set of discrete points (or sets of points if anisotropy is considered). So that this graph can be used to estimate the semivariance at any given lag, it necessary to fit a smooth model to these points. Using weighted least squares approximation, the points of the sample variogram will be fitted to one of several different models, including spherical, linear, exponential, Gaussian, De Wijssian and Bessel functions (Burrough, 1991). Davis notes that the process of fitting models is "to a certain extent an art, requiring experience, patience and sometimes luck. The process is not altogether satisfying, because the conclusions are not unique" (Davis, 1986, p. 245).

The resulting variogram has several interesting features which can be interpreted with respect to the phenomenon being studied (see diagram below). The *range* is the distance over which points exert influence on their neighbors; beyond the range, marked by the beginning of the *sill*, no influence is observed. The *nugget* represents the wholly random, non-spatial, variation (Oliver and Webster, 1991). However, many sample variograms and their fitted models will not match this classic form.



Kriging

Kriging is "a generic name for a variety of generalized least-squares estimation algorithms" (Journel, 1986) based upon the variogram and its derivatives. In its simplest case, a kriged point estimate is a linear weighted sum of the sample data (Oliver et al, 1989):

$$z^*(x_o) = \sum_{i=1}^n \lambda_i z(x_i) \quad \text{with} \quad \sum_{i=1}^n \lambda_i = 1$$

where x_i are the n points used in the estimation, $i=1$ to n , and $z^*(x_o)$ is the estimate of the value at point x_o . The weights are chosen by solving a set of simultaneous equations:

$$\sum_{i=1}^n \lambda_i \cdot \gamma(x_i, x_j) + \psi = \gamma^*(x_i, x_o) \quad \text{for all } j$$

where $\gamma(x_i, x_j)$ is the semivariance between x_i and x_j , $\gamma(x_i, x_o)$ is the semivariance between x_i and the point being estimated (x_o) and ψ is the Lagrange multiplier added to achieve minimization. Semivariances are determined from the variogram. The variance of the estimate is

$$\sigma^2(x_o) = \sum_{i=1}^n \gamma^*(x_i, x_o) + \psi$$

Kriged estimates are unbiased and optimal (Davis, 1986).

Using geostatistics

Geostatistics is rapidly moving out of the domain of statistics researchers and into the hands of practitioners in many different fields (Oliver and Webster, 1991). Its most important application is in interpolation since kriged estimates have been found to be as good or better than those made with other techniques (Burrough, 1986, Journel, 1986) though this contention has been argued by others (Philip and Watson, 1986). The existence of an error measure in the estimate variance is unusual among the range of interpolation procedures. Techniques for devising "block" estimates from point samples were one of the early developments in the set of tools available. Block estimates are particularly important in mining geology where scattered core samples must be used to determine the grade of ore over a large mining block (Davis, 1986), though the application of this technique to the derivation of estimates for the value of phenomena modeled using continuous piecewise models is obvious. Geostatistics is also useful in planning optimal and efficient sampling designs for continuous phenomena and for carrying out spatially constrained classification (Oliver and Webster, 1991). Finally, analysis of the variogram itself provides some insight into the form of the spatial variation of the phenomenon under study.

However, there are many problems with geostatistics and Kriging. Some statisticians question the fundamental statistical bases of the theory of regionalized variables and its derivatives in geostatistics (Philip and Watson, 1986 present a particularly strong criticism). At a less fundamental level, major problems with geostatistics arise due to the extremely high computational demands of the method. Each point estimate requires the solution of a set of simultaneous equations - 17 or more are used in some contouring packages (Davis, 1986). As well, Lam has noted that "the method is unreliable unless a large number of sample values are available" (Lam, 1983, p. 134); Webster and Oliver (1992) have shown that at least 100 and preferably 200 points should be used.

Finally, as Oliver, et al, have noted, "the method is by no means universally applicable and investigators need to be quite sure that it is appropriate to their circumstances when using it" (Oliver, et al, 1989, p. 262-3). In fact, while "the kriging process is an elegant solution to a difficult problem [it] requires a clear understanding of its approach and rather more elegant data than is often available to achieve a reliable result" (McCullagh, 1988, p. 753).

Geostatistics for spatial data models

How can geostatistics be used in environmental modeling with field variables? Clearly kriging provides an excellent interpolation alternative for many of the conversions. However, if it is to be used upon demand within mathematical models, it would be absolutely essential that the variogram is encapsulated with the dataset as the derivation of it is a time consuming and skilled task. Such encapsulation seems entirely appropriate as it is expected that the data gatherer would be sufficiently familiar with the phenomenon and the data which represents it to perform this geostatistical modeling task.

Since geostatistics is becoming more widely promoted (note its inclusion in standard statistical texts like Davis' *Statistics and Data Analysis in Geology*), computer software to assist in the development of variograms and kriged estimates are now appearing. Kriging is already included in many commercial GISs, including Arc/Info (from ESRI, Inc.), GeoEAS (from the EPA, Las Vegas) and Surfer (from Golden Software).

Having laid out the fundamental aspects of the mathematical manipulation of field variables, the next chapter identifies the specifics that must be included in declarations of these variables. These specifics are intended to allow the mathematical operations discussed here to be performed automatically, without direct control from the modeler.

CHAPTER 6 - SPECIFYING FIELD VARIABLES

Having examined the types of operations that may be performed on field variables it is now possible to consider the requirements for their specification. Based on the discussion of the previous chapter it is evident that each field variable must be described by two required properties and may also be described by several additional optional ones. Required properties that must be described are:

- spatial data model
- measurement system

Optional properties include:

- field type
- measures of accuracy
- interpolation procedure
- temporal characteristics

The value of any field may be one dimensional (scalar) or multidimensional (vector). Therefore, *scalar field variables* represent those fields for which there is a scalar value at every location while *vector field variables* represent vector fields. Scalar fields may be numerical or categorical but vector fields can only be numerical. While vector fields are important in many dynamic atmosphere and ocean models, the additional complexity of representing and manipulating dynamic phenomena in static geographic databases suggests that it is useful to consider each of these types separately. We begin therefore by limiting our initial consideration only to the representation and manipulation of scalar field variables. In Chapter 7, these concepts are expanded by considering vector fields, their properties and the operations which may be performed on them.

A. CONVENTIONS AND ASSUMPTIONS

Since mathematical models which will be run on the computer must be constructed in a computer language, and many are now being run on UNIX workstations, the strategy developed here is expressed in a simple pseudo-code similar in structure to C. However, it is designed to be generic and should be implemented easily in any computer language. Implementation issues are considered in Chapter 8.

As in C, all statements end with a semi-colon. Any statement may contain a block of statements which are contained within braces. In this document, pseudo-code statements representing elements of the strategy are shown by the use of courier typeface. Field variables are indicated in capital letters. Arguments are indicated by italic characters. For

the purposes of this strategy, it is assumed that all datasets which will be combined for use in a single mathematical model adequately cover the geographic area under consideration, are co-registered and expressed in a common geographic referencing system.

While there are some mathematical models which do use polar coordinate systems, rectangular coordinate systems are the basis of finite element numerical methods and are the most common ones used in GIS. While not explicitly treated within this initial framework for the spatial language, it is assumed that many of the conditions and operations described can be readily translated into ones appropriate for other geometric coordinate systems. As well we recognize as rectangular any coordinate system in which the axes meet at a right angle. Thus the latitude and longitude coordinate system are rectangular and in fact form the most important rectangular coordinate system for most regional and global mathematical models.

B. DECLARING FIELD VARIABLES

This strategy makes use of two types of environment statements. *Declaration* statements identify variables to be used within a particular mathematical model code while *definition* statements are used to define and identify a name for specific formats or operations to be performed later in the code. Spatial variables are declared using a declaration statement of the form:

```
field NAME:property(parameter),property(parameter) ...;
```

C. PROPERTIES OF FIELD VARIABLES

Spatial data model

As described in Chapter 4, there are six spatial data models currently used to represent fields. Each of these models has different inherent assumptions about the manner in which continuity is represented and may be inferred from data stored in that form. In the field variable declaration this property is identified as:

```
model(name: parameters)
  where name may be
    cellgrid           (cg)
    polygon            (py)
    tin                (tn)
    contour            (cn)
    pointgrid          (pg)
    irregular points   (ip)
```

Two letter codes indicated in brackets after each model name may be used as abbreviations. Some of these models have additional required parameters to provide information on the resolution or density of the data in each field variable and on the geometry and location of grids.

As demonstrated in the previous chapter, some measure of the density of spatial elements is required in order that decisions about some conversion operations can be made. Considering irregular point models, Tobler has suggested that the

average influence domain of each sample point might be calculated by dividing the total area (in square units) containing the observations by the number of observations. The square root of this value is an estimate of the average resolution (in linear units), or the effective average length per observation. (Tobler, 1988, p. 134)

He goes on to add that

The definition given here for spatial resolution enables one to make the calculation for point, line, polygon or raster Geographic Information Systems, and for categorical, scalar, vector, or tensor data. (Tobler, 1988, p. 135)

Since we are concerned here more with the "average influence domain" rather than the resolution, we choose to express the "density" of data as the *spatial elements per unit area, calculated by dividing the number of spatial elements by the total area.*

Of course, the density measure does not address a number of important issues about how the spatial dataset represents reality. Of particular importance is the issue of the relationship between the size of spatial units and the rate of variation of the phenomenon. For example, polygons may be large because the phenomenon under study is very homogeneous, or they may be large because the sampling interval was large and variation at a smaller scale was not recorded. This suggests a variation of the concept of resolution which may be defined as *the minimum distance over which variation has been recorded.* Since this aspect is related to issues of accuracy, we leave discussion of this to a later section.

Declaring grids

Since cellgrids and pointgrids have simple geometric characteristics, details on the grid spacing and orientation are supplied with this property identifier.

```
model (cellgrid: c1, c2,  $\alpha$ ,  $\Delta x$ ,  $\Delta y$ , nx, ny)
```

```
model (pointgrid: c1, c2,  $\alpha$ ,  $\Delta x$ ,  $\Delta y$ , nx, ny)
```

where

$c1$, $c2$ are the *geographic* coordinates of the center point of the grid expressed in the rectangular coordinate system of the entire database (e.g. latitude and longitude, UTM).

α is the number angle between -45° and 45° , measured clockwise, from north to the first axis. This axis defines the x or row direction. The angle has a range of $(-45^\circ, 45^\circ]$.

Δx , Δy are the size of the grid spacing (i.e. cell width or height or distance between adjacent points) in the x and y (row and column) directions respectively, includes units.

nx , ny are the number of cells or points in x and y respectively.

The density of the grid is

$$\frac{1}{\Delta x \times \Delta y}$$

and, therefore, it does not need to be specified.

An example of this property declaration is:

```
model (cellgrid: 34°N, 128°W, 15°, 100m, 100m, 256, 256)
```

Declaring polygons, TINs and irregular point models

Polygons, TINs and irregular point models are declared simply with the addition of the density measure. For example:

```
model (polygons: 10-5m-2)
```

Declaring contour models

The density of information stored in contour models cannot be expressed using the measure described above. Fortunately, since the density does not affect the manner in which data stored in contour models can be manipulated, this property is not needed for contour models. While it may seem that it would be useful to include items such as contour interval and lowest and highest contour values in the specification of a contour model, the previous discussion of spatial data model conversions did not suggest a use for such information. Thus contour models require no additional parameters to be declared in the model property.

Measurement system

For implementation, it is only necessary to recognize two different measurement systems. *Numerical* measurement systems are those which yield interval or ratio numbers. *Categorical* systems include nominal and ordinal measurements. While we often think of categorical data as being linked via IDs to relational tables of numerical and other categorical data, in this strategy we define each attribute as a separate field variable, though related variables are linked through a common spatial structure. This allows each attribute in the related table to be treated as an individual component in mathematical equations while at the same time economizing on spatial data model conversion efforts. We discuss this aspect in greater detail after all the components of the strategy have been specified.

This property is expressed as:

```
measurement (name)  
    where name is either numerical or categorical, abbreviated as num or  
           cat
```

Field type (required only if vector)

If the field is scalar, field type is expressed as:

```
type (scalar)
```

If the field is vector, field type is expressed as:

```
type (scalar, n)  
    where n is the number of components in the vector
```

Since the default type is scalar, this property is only specified when the field is a vector field. Vector fields are discussed in Chapter 7.

Specific interpolation procedure (optional)

This optional property allows the user to specify a specific interpolation procedure that is to be used in place of the default procedure for the dataset representing a particular variable. For example, if the points stored in an irregular point model are critical points, it may be appropriate to specify that interpolation should use a triangulation procedure rather than a default Thiessen polygon procedure. Kriging may be specified for a point model to override a default distance weighted averaging technique.

In the declaration statement, this procedure is specified only by a name:

```
interpolation (name)
```

The specific procedure will be identified and described in a *define procedure* statement.

Accuracy (optional)

The measurement and description of the accuracy of spatial data is a area of considerable concern under active research on a number of dimensions. Questions of accuracy in field variables relate to how accurately the true value at any point is predicted. Accuracy of prediction depends upon many things including: the resolution (the minimum distance over which variation has been recorded) and its relationship to the distance over which the phenomenon varies; and, how accurately a spatial data model represents the true continuous field. There is a growing body of literature on the accuracy of spatial databases and many measures have been proposed (cf. Goodchild and Gopal, 1989); however, the selection of appropriate measures to use in particular cases depends upon the way in which the data is to be used (Goodchild, 1992). Thus, although not specifically discussed, the structure of the declaration of field variables, their properties and optional procedures described here can be easily extended to include one or more measures of accuracy which can in turn be used to determine the appropriate spatial operations to perform.

Temporal characteristics (optional)

As discussed in an earlier chapter, time, like space, is continuous. However, in order to constrain the scope of this research, we have chosen to assume time can be represented adequately by the time slice or the time instant. Since many environmental models do include analysis of time sequences, it is necessary to include syntax for dealing with temporal aspects. Thus, temporal characteristics are described by the length of the time slice, the time of the beginning of the series and the number of observations in the set:

```
time (length_of_slice, start, number_of_observations)
```

where

length_of_slice is the amount of time during which the measurement was recorded. If the recording was a single moment then length is 0.

Except in the case of 0 length, this parameter includes both a number and a unit. For example "12 hours", "1 month".

start is the time at which the first moment was recorded or the first slice began. Time is recorded using the SI format, yy/mm/dd/hh/mm/ss. This sequence must be filled from the left, but unnecessarily precise elements on the right may be omitted.

number_of_observations is the total number of datasets in the temporal sequence.

Thus, a daily average temporal sequence starting on September 15, 1954 would be recorded as:

```
time (1 day, 54/9/15, 15)
```

For those sequences which are averages of cyclical time periods, only those elements of the start date which vary need to be replaced with numbers. For example, to express a sequence whose values are representative of all values during a calendar month, starting with January, 12 months in the series, temporality would be specified as

```
time (1 month, yy/1, 12)
```

The default condition is static. Thus, temporality does not need to be declared if the variable represents a phenomenon which is constant over time. In this case, though it is unnecessary as this is the default, static variables can be declared as

```
time (i, 0, 1)
      where i represents infinity (or static conditions)
```

Examples of field declarations

```
field TEMP: model(irregular_points: 10-2 km-2),
              measurement(numerical), time(0, 91/5/1, 31) ;
```

```
field VEG_TYPE: model(polygons: 10-5 m-2),
                  measurement(categorical) ;
```

```
field CLOUD_COVER:
  model(cellgrid:34N°,10W°,0°,100m,100m,256,256),
  measurement(numerical), time(1 month,yy/1,12);
```

D. Defining special components

In order to allow for user control in the implementation of this strategy, it includes a number of different types of *definition* statements. These definition statements name and describe properties, rules, tables or spatial structures which can subsequently be referred to only by name. The standard syntax for definitions is

```
define component name: parameters ;
```

Define model

Since it is likely that several variables will share similar spatial properties, it is useful to allow for the definition and naming of specific spatial models.

```
define model NAME: (spatial_model: parameters) ;
```

For example,

```

define model AGRID:
    (cellgrid:34N°,12W°,0°,100m,100m,256,256) ;

field PRECIP, HUMIDITY, TEMP: model(AGRID),
    measurement(numerical), time(1 month,yy/1,12);

field VEG_TYPE:
    model(AGRID), measurement(categorical) ;

```

The model name is capitalized here as it can be used, as described in the next section, as a generic term in other statements to refer to the several fields which share the particular structures.

Define time

Since the temporal property can become lengthy and is likely to be repeated in the declarations of several variables, we also allow the definition of time properties:

```

define time name: (parameters) ;

```

Define computational structure

In the earlier section on the numerical solutions for continuous equations, a number of different approaches to discretizing continuity for mathematical manipulation was described. In that section it was noted that finite difference and finite element solutions appear to be the most widely used. Both of these solutions have specific spatial structures in which all model calculations are performed. Therefore, it is useful to declare the *computational* structure so that intermediate conversions will tend to terminate in datasets which are spatially identical and algebraic operations can be performed on identical spatial elements. The computational structure is defined using the same structure as the model property declaration:

```

define model COMP:(spatial_model: parameters) ;

```

or

```

define model COMP(NAME):
    (spatial_model: parameters);

```

If the computational structure is a grid, the first form may be used since the grid is fully described by the model properties and does not need to be supplied externally. Any computational structure which is not a grid must be described using the second form which specifies that the computational model will be input as a variable.. *NAME* refers either to one of the independent variables or to a dummy variable which has only spatial structure and no values.

In subsequent field declarations, this structure may be used like other defined models to establish spatially similar field variables. Variables with the computational structure are given structural precedence in many conversion operations. Since conversion to the computational model is frequently required, a special mark (") is used to signify a field variable which has been reformatted to the computational model. For example,

```

T" = T

```

indicates that the spatial model of T has been converted to that defined for the computational model.

Define table

The define table statement is used to define lookup tables used to derive numerical values for computation from categorical field variables. These tables may be one or two dimensional. A one dimensional table may be used to relate a numeric or categorical value to a class in a categorical field or to a range of numeric values in a numerical field:

| Class or upper range value | Value |
|----------------------------|-------|
| 10 | 150 |
| 20 | 200 |
| 30 | 400 |

If the field which is being used to determine the value is numerical, the values in the first column represent the upper inclusive value of a range extending from but not including the next lower range limit in the column. For example, in the table shown above, if the field referred to in the first column is numerical, then the table indicates that any field value >10 and ≤20 would be related to the table value 200.

Within the syntax outlined here, a one dimensional table would be defined as:

```
define table name: field_name ;
```

A two dimensional table would be used to determine a derivative value from the joint values of two different fields. These fields may be categorical or numerical. For example, a two dimensional table might look like this:

| | SOIL | | |
|------------|---------|---------|---------|
| VEGETATION | Class 1 | Class 2 | Class 3 |
| Type 1 | 2 | 2 | 3 |
| Type 2 | 2 | 1 | 3 |
| Type 3 | 1 | 1 | 2 |

Like one dimensional tables the fields in the rows and columns may be either categorical or numerical. If numerical, like one dimensional tables, values under the field names indicate the upper limits of ranges to which the table value is to be related. The exclusive lower limit is defined by the next lower range limit listed. A two dimensional table would be defined as:

```
define table name: row_field_name, column_field_name;
```

where

row_field_name is the field variable whose classes define the rows,

column_field_name is the field variable whose classes define the columns.

For example, in the table given above VEGETATION is the `row_field` and SOIL is the `column_field`.

While the content of most tables will be supplied as input, for some operations it is necessary to define a table in advance which can hold the output of an operation on categorical data. The use of tables is described further below in the section "New operations on field variables".

Define procedure

Metaknowledge about the data in the database can be used to determine critical characteristics about the relationship between the representation and the surface. Metaknowledge itself is difficult to codify but may be critical in choosing correct interpolation and other conversion procedures. By using metaknowledge to define procedures initially, the modeler or the database specialist can:

- specify an appropriate interpolation procedure for a particular dataset.
- specify appropriate smoothing operations for densification of cellgrids and pointgrids.
- specify the appropriate conversion rules for densifying or generalizing categorical data.

Conversion procedures can be model-to-model conversions, variable-to-variable conversions or specialized functions for performing particular algebraic operations. Procedures are specified with a symbol (often algebraic) and the field variables or spatial models to which it is to be applied. Such overloading of system defined functions is a common feature of most higher level programming languages.

```
define(VARIABLE_1 SYMBOL VARIABLE_2){algorithm};
```

For example:

```
define (A = B): {algorithm} ;

define (pointgrid = irregular_points):
  {algorithm} ;

define (VEG + SOILS): {
  overlay (VEG, SOILS)
  lookup (potential)} ;
```

The use of the terms *overlay* and *lookup* in the previous example illustrate how this syntax can be used as a window to a GIS or any other applications program. According to the modeler's familiarity with other software, these procedures can be defined as generically or specifically as desired. Procedures can be specified simply by a common name (like *overlay* or *buffer* for GIS operations) or, if necessary or desired, by pseudo-code descriptions of the algorithms involved. When implemented, procedures can be specified as functions or subroutines in procedural languages or applications software.

Once defined, the symbol can be used directly in mathematical statements. For example,

```
A = B ;  
POTENTIAL = VEG + SOILS ;
```

will perform the conversions as defined above. In the first case B is assigned to the new field A according to the algorithm. In the second case, the fields VEG and SOILS are overlaid (using some undefined GIS) to create a new field POTENTIAL. The value in each region of POTENTIAL will be determined by the table potential. (Categorical operations are discussed in greater detail below.)

Define conversion matrix

Having outlined the default conversion matrices for field variables in the previous chapter, it is useful to allow for their formal specification. The procedures specified in these matrices would be invoked if a field declaration does not specify a special procedure which is to be applied to it. Thus these procedures can be seen as the generic operations, those which can be used in the absence of additional information encapsulated with the field variable itself. Note that for conversion of categorical fields to piecewise models, the specification of rules is required. Thus categorical conversion matrices are always incomplete and require encapsulation of additional procedures with each variable or the definition of suitable procedures.

There are many different ways the matrices could be implemented in different computer languages. Here we suggest a conceptual form which may be easily translated when required. A matrix is constructed as a simple table with each cell containing one or more procedures and a set of decision rules if necessary. Tables 4 and 5 in Chapter 5 outline possible default matrices. Physically, these tables may be implemented as lists of procedures which can be selected based on the type of spatial data model of the source and target field variables. Each of the procedures included may be defined in a definition statement or provided externally through GIS or other software.

Define priority rules

Similarly, it is possible to also formally specify the priority rules as a procedural list with appropriate decision paths.

E. Spatial equality and nesting

The concepts of spatial equality and nesting were defined in Chapter 4. Unless specified, it is assumed that field variables are not spatially equivalent or nested. If these properties do exist they are specified explicitly as follows:

```
declare equivalent: (A, B) as C ;  
declare nested: (D, E) in F ;
```

However, equivalence will more often be specified by declaring several field variables in the same statement, for example:

```
field A, B, C: model(COMP), measurement(numeric);
```

This indicates that fields A, B and C are spatially equivalent. This is the statement which is used to declare individual attributes of a categorical variable as separate field variables.

To simplify data manipulations, spatially equivalent fields can be jointly referred to in other types of statements as well. For example, to input a set of variables possessing the same spatial structure, the following can be used:

```
define model (APOLYGON) : (polygon: 10 km2) ;  
field A,B,C: model (APOLYGON), measurement (num) ;  
read APOLYGON (A,B,C) ;
```

This read statement can be interpreted either as reading in the spatial structure and a related table of attributes or as reading in three separate field variables. On the other hand, the input of these variables could also be expressed as:

```
read A,B,C ;
```

in which case their individuality is stressed.

If we wish to force conversion of a set of spatially equivalent field variables to the computational structure, then using the definition and declaration above, we can write:

```
APOLYGON" = APOLYGON ;
```

or

```
A" = A ;  
B" = B ;  
C" = C ;
```

F. Neighborhoods

For many natural processes, the values of phenomena which we observe at particular locations are determined by conditions at other locations in their neighborhoods. Therefore, it is essential to establish a syntax for working with neighborhoods. This section considers the definitions of neighborhoods for fields and provides a syntax for working with neighborhoods.

Definition of neighborhood

In many cases, devising a precise definition of a neighborhood is difficult. We often identify the area surrounding a place as its neighborhood. Funk and Wagnalls' dictionary defines a neighborhood as "the quality or condition of being near; proximity". For his map algebra Tomlin defines a neighborhood as "any set of one or more locations that bear a specified distance and/or directional relationship to a particular location" (Tomlin, 1990, p. 96).

For our purposes, a neighborhood can be defined as *the area adjacent to a specific location whose condition affects the process under study at that location*. Note that the extent of the neighborhood is determined by the process. For example, in a study of plant root growth, the neighborhood of a plant might be limited to that area over which the plant roots compete for nutrients with other plants. In a study of soil erosion, the neighborhood might encompass all of the hillside upslope from a point. Thus neighborhoods may be defined in terms of proximity relationships (e.g. within a distance of) or connectedness

relationships (e.g. flows into this location). If the process is isotropic, then the neighborhood may be expressed simply as a circle surrounding the location whose radius is equal to the distance over which the process is affected by other locations. If the process is anisotropic, then the neighborhood will extend in the direction of the effect. As well, physical barriers, such as rivers or mountains may reduce the extent of the neighborhood on the affected side.

To define this concept for spatial data models, it is necessary to translate the field view of a neighborhood to one which can be expressed as a set of spatial elements. Beginning with Tomlin's definition and rephrasing it in the context of a field variable represented by a discrete spatial data model, a neighborhood is *the set of complete or partial spatial elements which bear some distance or connectedness relationship to the spatial element under consideration*. We recognize two types of neighborhoods - within a distance, and connected to.

Within neighborhoods are geometrical. They can be calculated from the geometry of the spatial elements and can be used equally well for numerical and for categorical fields. Many connected to neighborhoods, in particular watersheds, are determined by the value of the phenomenon under study. These neighborhoods also can be determined automatically but they are restricted to use with numerical fields. Other connected neighborhoods, such as those determined by adjacency to a transportation corridor or water body, are determined by factors external to the phenomenon being studied. They cannot be derived directly from the values of the field and must be input.

Using a neighborhood

However a neighborhood is defined, it may be referenced by a simple syntax:

```
FIELD [neighborhood]
```

where [*neighborhood*] has been previously defined by some function or read in as input. For example, for a hierarchically nested set of polygons depicting subwatersheds, the topological relationships will be stored in some manner with the polygons, either in the geometry or in an attribute table. Thus

```
SURPLUS [UPSTREAM]
```

refers to the set of subfields of the field SURPLUS, one subfield for each location, defined as the neighborhood UPSTREAM. Each location will have a unique subfield but subfields may overlap.

When implemented for the spatial data models, each spatial element will be associated with a set of spatial elements as its neighborhood. We assume that the spatial elements in the neighborhood will be a subset of the set of spatial elements which collectively comprise the field variable itself. Thus subfields defined as individual neighborhoods in a pointgrid will be sets of points, not regions.

G. NEW OPERATIONS AND VARIABLES

Integrals

Integral functions were introduced in the previous chapter. These may produce either ordinary scalar variables or scalar field variables. If we wish to determine the integral of a single quantity over the field then:

$$a = \text{integ}(B)$$

will produce a single scalar value for that function. If the result is assigned to a field then the parameter of the function must be associated with a neighborhood:

$$A = \text{integ}(B[\text{up}])$$

The result will be a scalar field variable A whose value at each point is the integral of the value of the phenomena in the neighborhood [up] of each point of B.

A third variation of this function produces a table of scalar values. If we wish to know the integral of the field over a set of regions, then it can be expressed as:

$$\text{table}(a) = \text{integ}(B, C)$$

where C is a categorical field which defines the regions and table(a) has been defined previously as.

```
define table a: C, null;
```

The result will be a table with one column:

| Field classes | Integ value |
|---------------|-------------|
| | |
| | |
| | |

Latitude and longitude variables

Latitude and longitude are fields. They can be determined for any location on the surface of the earth and vary continuously in space. Fortunately, this geographic coordinate system is geometric and mathematical functions can be used to determine latitude and longitude at any location that falls within a region covered by a geographically referenced dataset. For this reason, we suggest that functions for determining latitude and longitude be incorporated as the reserved field variables, LAT and LONG. These variables can be used directly in any equation in which latitude and/or longitude are independent variables.

Earlier we have noted that any mathematical operation must be performed on discrete data representing a single location. Therefore, any mathematical statement which incorporates the reserved variables LAT and LONG assumes the values of these variables are determined for each iteration of the statement by the specific location for which the equation is being calculated. Since latitude and longitude are point values, they can be determined directly for the location of any point in a point model. For piecewise models, a location within each region must be identified as the location for which latitude and longitude will be determined. The most obvious location for these points would be the centroid of each spatial

element. While centroids are easy to locate within rectangular cells or triangles, centroid calculation for irregular polygons can be quite complex. Fortunately, most polygon data structures include reference to an included point which can serve as a representative location for the location of latitude and longitude.

Slope and aspect

Slope and aspect are important field variables in many mathematical models. (While these two scalar quantities can be jointly regarded as the vector quantity *gradient*, we leave consideration of vectors to a later chapter.) Aspect determines the impact of solar radiation and the local direction of flow on the surface. Slope is important in many physically based models as it plays a critical role in determining the rate of flow and the energy content of mobile constituents. Slope can also be thought of as the physical description of the first derivative of a function describing the field. These fields can be derived from any numerical field representing a potential such as elevation or hydraulic head.

Mathematically, slope can be defined as the inclination of a line or plane tangent to surface at the point. If the surface can be described by a function then slope can be defined as

$$\text{slope} = \lim_{h \rightarrow 0} \left(\frac{f(a+h) - f(a)}{h} \right) = \frac{df}{dh}$$

where h is a small distance. (Note that this formula is the textbook definition of the derivative of the function.) In a Cartesian coordinate system, slope can be calculated as an angle (Burrough, 1986):

$$\tan(\text{slope}) = \left(\frac{\partial z}{\partial x} \right)^2 + \left(\frac{\partial z}{\partial y} \right)^2$$

where z is the value at a point. Aspect can similarly be defined in Cartesian space as (Burrough, 1986):

$$\tan(\text{aspect}) = \frac{-\partial z / \partial y}{\partial z / \partial x}$$

While slope and aspect are defined strictly at points, practically speaking they are often used in a more regional sense. For example, when considering water flowing over a hillside, we are not particularly concerned where each individual drop of rain goes, but instead need to know the integrated flow path of a large number of minute flows. Fortunately, the spatial discretization that is required for the numerical solution of mathematical models as well as for the digital representation of fields provides a basis for this spatial generalization.

In order to calculate the parameters of slope and aspect, it is first necessary to estimate a function which describes the shape of the surface at any given point. This estimation is easiest with a TIN model since the slope of each triangle can be described by a simple linear function in x and y . In fact, many implementations of TIN models automatically calculate slope and aspect as a standard attribute. However, due to the discontinuous nature of the functions at TIN boundaries, this method cannot provide slope and aspect calculations at the triangle nodes.

Point models can be used to derive functions of continuous surfaces through the use of several different interpolation procedures including fitted polynomials and splines. While global methods can be used, more frequently, piecewise polynomial surfaces are calculated. For example, for pointgrids, a polynomial can be fitted to the set of 9 points centered in the 3 x 3 window around each point in the grid (e.g. Evans, 1980, uses a six-parameter quadratic equation to estimate slope and aspect at the central point). For irregular points the determination of the appropriate surface patches to use for calculating piecewise polynomials can be difficult (Lam, 1983). Therefore, it is common to use simpler methods for calculating slope and aspect on pointgrids. For pointgrids, Burrough notes that most researchers "show most interest in gradient [slope] and aspect which are computed from finite differences" (Burrough, 1986, p. 50). For example, a simple finite difference estimate of slope in the x direction at point i,j is (Burrough, 1986):

$$\left(\frac{\delta z}{\delta x} \right)_{ij} = \frac{z_{i+1,j} - z_{i-1,j}}{2\delta x}$$

For irregular point models, it may be more efficient to triangulate the points and calculate slope and aspect as for TIN models.

For the constant piecewise models, the slope of each spatial element is 0° except at element boundaries where it is 90° (derivative of infinity). Thus, slope and aspect cannot be defined on a polygon or cellgrid model. However, for cellgrids, if the value of each cell is assigned to its center point, cellgrids can be converted to pointgrids which in turn can be used to calculate continuous surfaces from which slope and aspect can be determined.

Contour models also provide a defined surface on which it is possible to determine slope and aspect at specific points. Given that a pointgrid or set of irregular points at which values must be determined is placed over the contour model, it is possible to determine the required values. While some algorithms may be more efficient or precise than others, any application program which manipulates contour models is likely to have built-in modules for determining slope and aspect at any point.

Given the above discussion, we suggest that it is appropriate to devise slope and aspect functions:

A = slope (B) ;

A = aspect (B) ;

where each of these returns a new field of the desired parameter. Specific procedures used to derive these values are determined by the spatial data model.

It should be noted that Chapter 7 - Vector Fields provides an alternate way of operationalizing the concepts of slope and aspect.

Categorical functions (lookup)

By definition, categorically measured fields are not continuous in value since categories are discrete. However, since categorical field variables are often used conceptually for continuous phenomena (e.g. soils classes in the determination of erosion potential) it is useful to devise a categorical function using tables and the define procedure statement. Unlike algebraic functions discussed in earlier chapters, categorical functions do not return mathematically derived values. However, a large portion of the data available to

environmental modelers is categorical. It is therefore necessary to allow for the specification of functions for categorical field variables. There are several variations on this theme which need to be considered separately.

One input variable, input and output in same spatial model

This function might be thought of as a simple reclassify function. Each class in the input variable is assigned to a different class or value for the output variable. None of the geometry changes. This is easily operationalized as in the following:

```
define model AGRID: (cellgrid: ...) ;
field SOIL_IN:
  model(AGRID), measurement(cat) ;
field SOIL_OUT: model(AGRID), measurement(num) ;
define table reclass: SOIL_IN ;
define (SOIL_OUT = SOIL_IN): {
  lookup (table(reclass)) } ;
read SOIL_IN ;
SOIL_OUT = SOIL_IN ;
```

This function works equally well for any combination of numeric and categorical variables.

Two input variables, input and output in same spatial model

A two dimensional table may be used to combine the values of two variables to produce a third. Since the table produces discrete values for each combination, these are considered categorical functions even though all variables may be declared numerical. As with one dimensional tables above, if input variables are numerical, then the table defines ranges within which values may be found.

Suppose A and B are categorical fields which are to be jointly considered to determine a numerical result for field C:

```
define model AGRID:
  (cellgrid: 34°N,128°W,15°,100m,100m,256,256) ;
field A, B: model(AGRID), measurement(cat) ;
field C: model(AGRID), measurement(num) ;

define table AbyB: A,B ;

define (A * B): {lookup (table(AbyB)) } ;

C = A * B
```

Other lookup tables can be defined for the same pair of field variables and defined for other symbolic operators.

Two input variables, all variables in different spatial models

If there are several different spatial models involved in the lookup, it becomes more complex:

```
define model AGRID:
  (cellgrid: 34°N,128°W,15°,100m,100m,256,256);

field A: model(polygon: 1000 m2),
  measurement(cat) ;
field B: model(polygon: 1200 m2),
  measurement(cat) ;
field C: model(AGRID), measurement(num) ;

define table AbyB: A,B ;

define (A * B): {lookup (table (AbyB)) } ;

C = A * B
```

First the two input variables must be overlain to produce a temporary merged variable. The lookup then results in the assignment of a single value for each new polygon. This temporary numeric polygon field variable must then be converted to the target model using the a procedure either defined within the program or established in the defaults.

H. GIS OPERATIONS FOR FIELD VARIABLES

For completeness, we include reference to standard GIS operations for field variables. In the pseudo-code described here it is acceptable to include reference to standard GIS operations such as overlay and buffer. These operations are to be understood in their traditional sense. If desired, it is reasonable to incorporate more detailed specifications for these operations such as those described in map algebra.

I. EXAMPLES OF THE USE OF THE STRATEGY

Generalize a categorical cellgrid

Reduce the density of a cellgrid by generalizing 2x2 windows of cells into single cells. The cellgrid represents a categorical variable which has four classes (1 to 4). Assign a class to the new larger cells using the following rules:

1. if 3 or 4 of the set of 4 original cells are a single class, assign that class to the new cell.
2. if 2 original cells are class 4, assign class 4 to the new cell,
3. if 2 original cells are class 3, assign class 3 to the new cell,
4. if 2 original cells are class 2, assign class 2 to the new cell,

5. if all original cells are different classes, assign class 4 to the new cell.

```
define model SMALL-GRID:
  (cellgrid: 34°N,118°W,0°,100m,100m,1000,1000) ;
define model LARGE-GRID:
  (cellgrid: 34°N,118°W,0°,200m,200m,500,500) ;

field VEG_TYPE: model(SMALL-GRID), measurement(cat);
field VEG: model(LARGE-GRID), measurement(cat) ;

declare (VEG-TYPE in VEG) ;
define (VEG = VEG-TYPE): {
  {values} = {set of 4 values in VEG_TYPE in
             a single cell in VEG}
  for each {values} {
    if mode({values}) exists, return mode, end;
    if count({values}=4) = 2, return 4, end;
    if count({values}=3) = 2, return 3, end;
    if count({values}=2) = 2, return 2, end;
    else return 4 } } ;

read VEG-TYPE ;
VEG = VEG_TYPE ;
```

Polygons to grid

Given two field variables, one a cellgrid the other a polygon model with several attributes, calculate the multiple of a numerical field and one attribute from the table related to the categorical field.

```
define model AGRID:
  (cellgrid:30°N,15°S,0°,1km,1km,200,200) ;
define model VEG_TYPE: (polygon: 10-4m-2) ;

field HOURS_SUN, GROWTH:
  model(AGRID), measurement(num) ;
field GROWTH_RATE, DENSITY, HEIGHT:
  model(VEG_TYPE), measurement(num) ;

read VEG_TYPE(GROWTH_RATE,DENSITY,HEIGHT), HOURS_SUN;

GROWTH = HOURS_SUN * GROWTH_RATE ;
```

Functions on grids

Calculate photosynthetically active radiation (PAR) at the surface as a function of latitude, reduced by cloud cover to 45% to 65% of total irradiance.

```
define model AGRID:
  (cellgrid: 30°N,15°S,0°,10°,10°,200,200) ;
```



```

define time atime: 1 month, yy/1, 12 ;
field LATITUDE: model(AGRID), measurement(num) ;
field CLOUD-COVER, PAR:
    model(AGRID), measurement(num), time(atime) ;
read CLOUD-COVER;
for each month: {
    PAR = ( f(LAT) * (65 - CLOUD-COVER/5) /100 ) } ;

```

where $f(LAT)$ determines the total irradiance received at the surface under 0% cloud cover.

Neighborhood functions

Calculate the growth at a number of point locations based on a function of rain, cloud cover, soil, elevation and aspect given rain at irregular points, cloud cover as a cellgrid, soil as polygons, elevation as a pointgrid and aspect calculated from the elevation pointgrid.

```

define model COMP: (ip: 10-2 km-2) ;
define time months; 1 month, yy/1, 12 ;
field GROWTH:
    model(COMP), measurement(num), time(months) ;
field RAIN: model(ip: 10-3 km-2),
    measurement(num), time(months) ;
field CLOUD-COVER:
    model(cg: 24°N,128°W,0°,1km,1km,100,100),
    measurement(num), time(months) ;
field FERTILITY: model(pn: 10-3 km-2), measurement(num) ;
field ELEV: model(pg:24°N,128°W,0°, .5km, .5km,200,200),
    measurement(num) ;
define find-growth {equation for calculating growth from
    several parameters. Conversion of all spatial models to
    computational model is enforced before calculation} ;
read SITES, RAIN, CLOUD-COVER, FERTILITY, ELEV ;
GROWTH = find-growth(RAIN,CLOUD-COVER,FERTILITY,
    ELEV,aspect(ELEV) ) ;

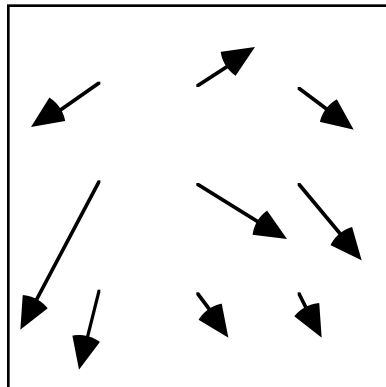
```

This chapter has outlined the essentials of a pseudo-code syntax which can be used to specify those aspects of field variables required to fulfill the requirements of the strategy for dealing with spatial continuity set out in Chapter 4. The next chapter suggests an extension to this syntax by considering the possible inclusion of vector fields.

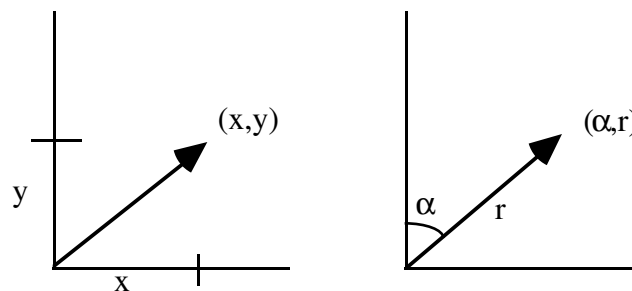
CHAPTER 7 - VECTOR FIELDS

Having demonstrated the fundamental properties and use of scalar field variables for mathematical modeling, it is now useful to expand consideration to include vector fields. While vector fields are important in some areas of environmental modeling, particularly those dealing with fluid flow in 3 dimensions (i.e. groundwater and oceanic models), the incorporation of vector fields into GIS has not yet been explored. In this chapter we consider the potential role of vector fields for all areas of environmental modeling. Vectors are denoted here by bold face type.

While scalars have only magnitude, physical vectors have both magnitude and direction. Examples of physical vectors are displacement, force and velocity, though practically speaking a vector may be thought of as any field for which there are two or more related values at any location. Two and three dimensional vectors are graphically represented by directed line segments in which length and direction are relevant:

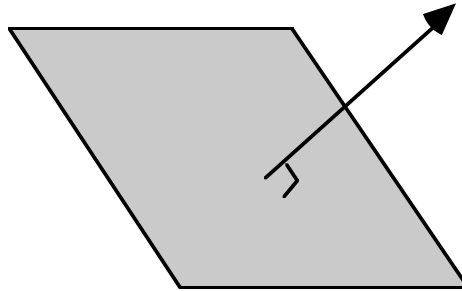


The origin of the directed line segment signifies the location of the vector while the location of the endpoint of the line is determined by the vector's value. Hence, these values can be described either by the Cartesian coordinates of the line's endpoint or by the line's magnitude and direction:



Since it is easier to perform mathematics on the Cartesian coordinate form of vectors, unless otherwise indicated below, the Cartesian coordinate form is assumed.

Vectors can be used to represent directed quantities other than those related to movement. In particular, *plane elements* can be represented by vectors such that the direction component is normal to the plane and the magnitude is the area of the element (see diagram below). As is shown later, this view of vectors is particularly useful for calculating flux.

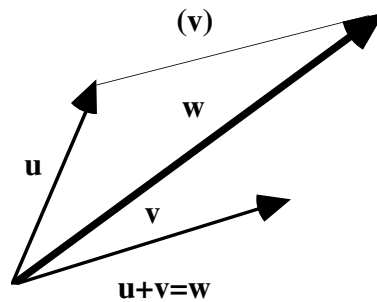


A. OPERATIONS ON VECTOR FIELDS

To illustrate the utility of vectors, this section reviews some basic operations for vectors.

Vector algebra

Adding vectors results in a vector describing total displacement. It is graphically expressed as the triangular sum:



Thus, addition of vectors is simply the addition of similar elements:

$$\text{if } \mathbf{u} = [a, b, c] \text{ and } \mathbf{v} = [d, e, f] \text{ then } \mathbf{w} = [a + d, b + e, c + f]$$

There are two types of multiplication for vectors. The scalar, or dot, product is the sum of the product of similar elements:

$$\mathbf{u} \bullet \mathbf{v} = (a \times d) + (b \times e) + (c \times f).$$

The scalar product can also be expressed as

$$\mathbf{u} \bullet \mathbf{v} = |\mathbf{u}| |\mathbf{v}| \cos \theta \quad (8-1)$$

where $\cos \theta$ is the angle between the vectors and $|\mathbf{u}|, |\mathbf{v}|$ are the magnitudes of the vectors (Stewart, 1987). Since $\cos(90^\circ) = 0$, the dot product of two orthogonal vectors is equal to 0. An excellent physical example of a scalar product is flux (flow rate) which can be expressed as the dot product of the vector describing flow and the vector describing the plane element (Shercliff, 1977) or the algebraic product of the magnitude of the flow times the area of the plane element (the magnitude of the plane element vector) times the cosine of the angle between them (equation 8-1).

The vector, or cross, product is:

$$\mathbf{u} \times \mathbf{v} = [bf - ce, cd - af, ae - bd]$$

and

$$\mathbf{u} \times \mathbf{v} = |\mathbf{u}| |\mathbf{v}| \sin \theta$$

It is important to note that the vector product produces a vector which is perpendicular to both original vectors and that two vectors are parallel if their cross product is the *null vector*, a vector whose elements are all 0 (i.e. [0,0,0]).

Vector calculus

Integrals, which may be thought of as the sum of field values over space and time, and derivatives, which describe the variation of field values in space and time, provide conceptually and mathematically useful abstractions for environmental modeling purposes. Interestingly there is a close relationship between scalars and vectors through calculus. Calculating gradient transforms a scalar field to a vector field (Tobler, 1989) and potential is a scalar field which can be derived from a vector field (Brauer, 1988). For example, the vector field heat flow is associated with the scalar field temperature (potential) through the operation of differentiation.

The differentiation operator, *del* or ∇ , is defined in three dimensions as:

$$\nabla = \frac{\partial}{\partial x} \mathbf{i} + \frac{\partial}{\partial y} \mathbf{j} + \frac{\partial}{\partial z} \mathbf{k}$$

where \mathbf{i}, \mathbf{j} and \mathbf{k} are the base vectors in the three orthogonal directions (i.e. $\mathbf{i}=[1,0,0], \mathbf{j}=[0,1,0]$ and $\mathbf{k}=[0,0,1]$). Applying *del* to scalar fields and to vector fields produces complimentary results (Shercliff, 1977). *Gradient* (grad) produces a vector field by differentiation of a scalar field:

$$\text{grad } q = \frac{\partial q}{\partial x} \mathbf{i} + \frac{\partial q}{\partial y} \mathbf{j} + \frac{\partial q}{\partial z} \mathbf{k} = \nabla q$$

Grad describes the nonuniformity of the scalar field. Vectors point in the direction in which the scalar quantity increases and their magnitude expresses the rate with which the quantity changes. This is directly analogous to the physical aspect and slope of a topographic surface. Darcy's law for seepage flow through a porous medium (e.g. aquifers) and related diffusion

laws can be expressed using the grad operator. In two dimensions, diffusion can be expressed as (Shercliff, 1977):

$$\mathbf{n} = -d \text{ grad } n = -d \left(\frac{\partial n}{\partial x} \mathbf{i} + \frac{\partial n}{\partial y} \mathbf{j} \right)$$

where \mathbf{n} is a vector describing the diffusion vector with magnitude equal to the quantity of diffusing phenomena passing per unit area and time, n is the concentration of the phenomena, and d is the diffusivity of the phenomena through the medium.

Divergence (div) produces a scalar field by differentiation of a vector field:

$$\text{div } \mathbf{v} = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} = \nabla \bullet \mathbf{v}$$

Divergence calculates the *potential* of a vector field and describes its nonuniformity by measuring the net efflux (flux out minus flux in) in a flow field.

While divergence can be used to determine the quantity that can be displaced, gradient is used to determine the field of displacement. It is useful to note that both grad and div are invariant on translation or rotation of the axes (Shercliff, 1977).

The product of the div and grad operators is the Laplace operator:

$$\text{div}(\text{grad } f) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = \nabla^2 f \quad (\text{a scalar})$$

This operator is commonly found in many mathematical models of environmental processes as it "forms the basis for solving a very wide range of field problems and a great variety of mathematical, numerical and analogue methods have been developed for extracting solutions" (Shercliff, 1977, p. 150). The Laplace operator can also be applied to vector fields (Stewart, 1987):

$$\text{grad}(\text{div } \mathbf{f}) = \frac{\partial^2 f_x}{\partial x^2} \mathbf{i} + \frac{\partial^2 f_y}{\partial y^2} \mathbf{j} + \frac{\partial^2 f_z}{\partial z^2} \mathbf{k} = \nabla^2 \mathbf{f}$$

Vector calculus can be extended in many different directions to suit many different problems of distributed flow. Since this section sought only to highlight some aspects of this large and fundamental mathematical topic, there is no need to explore these further.

B. REPRESENTING VECTOR FIELDS

The concept of vector fields has not been implemented in current commercial GISs and so there are no spatial data models specifically designed for vector fields. It is, of course, quite reasonable to construct models of vector fields using the separable components of vectors. Thus, vector fields are expressed by recording multiple values at each location in the dataset. What is missing, however, is a single common symbol or naming convention which allows these vector components to be handled as a single variable and manipulated in special vector functions.

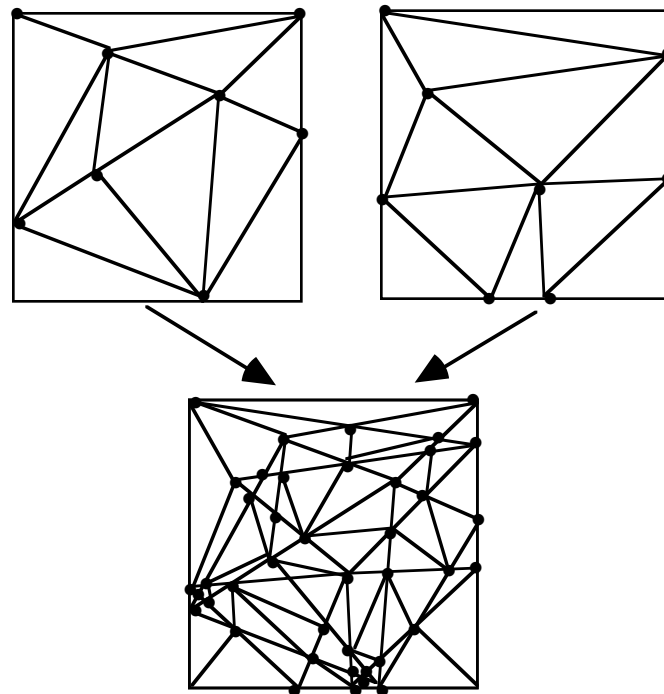
Point based spatial data models are the most appropriate models for representing multiple component vector fields since the field values can be attached directly to their specific locations. However, cellgrids and polygons can also be used to identify the value and location of vectors, though, like similar scalar fields, the vector is assumed to represent the mean vector value over the entire area included in the spatial element.

Using TINs as models of vector fields is less direct. Recall that TINs are modeled by storing the value of the phenomenon at the triangle nodes and using these values to determine the slope and aspect of the triangular planes between nodes. While it is possible to store multiple values at each node, the triangular planes between each node must be modeled individually for each vector component such that if the vector components are $[u,v]$ then the equations of the planes are

$$u = f(x, y)$$

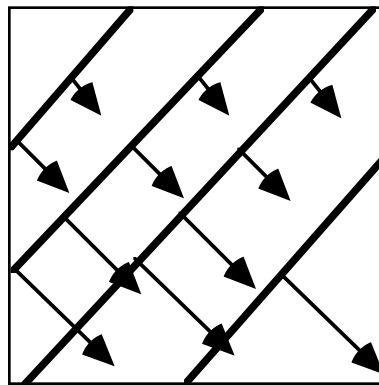
$$v = f(x, y)$$

The result is simply a set of individual TINs with common nodes. However, since the selection of locations for TIN nodes and the related planar faces is determined by the variation in the phenomenon being represented, it is inappropriate to have two TINs with common nodes. What is a "break in slope" or "ridge" for one component may be a smooth plane for another. The only correct way to use TINs in for vector fields would be to select critical points and triangulate for each component separately, then to combine the set of nodes and triangles such that each of the component TINs is a superset of the triangles of the result:

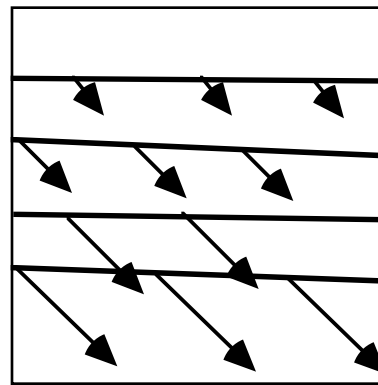


Since it is unlikely that data can be collected to support such a dense triangulation, we reject the use of TINs for vector field spatial data models.

The similarity between the appearance of contour models and diagrams of vector field lines may suggest that contour models are appropriate models for vector fields. Field lines are a common way of displaying certain aspects of vector fields. The tangent of a field line indicates the direction of the vector at that point. Tangents to contour lines also indicate the direction of the gradient of the scalar represented by the model. However, field lines are incomplete representations of vector fields. Field lines give information only about direction, not about magnitude, joining points of equal magnitude does not produce lines whose tangent indicates direction:



Field lines - tangent indicates directions



Lines of equal magnitude

The contour model cannot model more than one component, even though it can give information about that component's gradient. Thus, we find that only four of the six spatial data models are appropriate for representing vector fields.

C. USING VECTOR FIELDS IN ENVIRONMENTAL MODELS

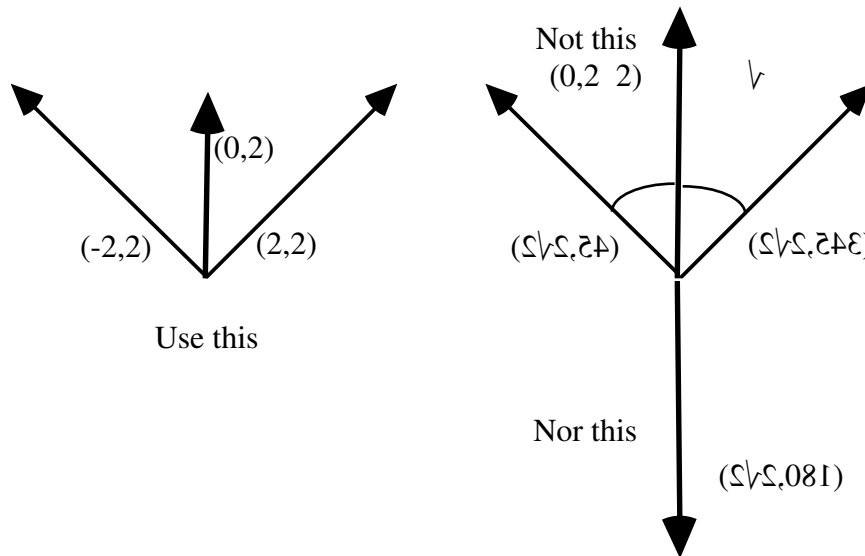
Declaring vector fields

The declaration of vector fields is identical to scalar fields with the exception of the replacement of `type(scalar)` with `type(vector, n)`. A naming convention can be established to distinguish between scalars and vectors though this would only be for convenience since the declaration statement specifies relevant characteristics. All other properties for vector field variables can be declared in a manner similar to that for scalar field variables.

Resampling vector fields

If we wish to use vector fields in mathematical operations, it is necessary, just like with scalar fields, to establish ways in which values can be found in one vector field model at locations specified in another vector field model. Thus we need to consider the issue of resampling from vector fields. In a general sense, the operations performed will be similar to those discussed for scalar fields in Chapter 5, including partitioning, point interpolation and areal averaging. The only difference is that each required arithmetic operation is performed on each vector component individually. Thus the sum of two vectors is the sum of each component and the mean of two vectors is the mean of each component. It is important to recall that vector components are expressed as Cartesian coordinates. This rule is particularly

important when calculating vector means, since averaging angle and distance components produces different and ambiguous results in the mean operation:



Using vector field variables

Vector fields have not been widely implemented in environmental models linked to spatial databases. As a result, vector operations for GIS have yet to be adequately considered by the software designers. Since this is uncharted territory, we consider only a few aspects of vector fields for GIS here. Further extension of this concept will only come with the incorporation of vector fields in conceptual data modeling and the implementation of vector functions.

Since vector operations can be performed on the separable components of vectors and it is possible to convert different spatial data models of vector field such that spatial elements can be made coincident, the use of vector field variables in algebraic statements is straightforward, so long as proper attention is paid to the dimensions of variables included in a single statement. For example, while it is possible to multiply a scalar times a vector, addition requires two vectors of equal dimensions. This is no different than similar algebraic rules, such as the prohibition against division by 0, and creates no particular difficulty. Given these restrictions, analytical, finite difference and finite element solutions of complex equations may proceed in the normal scalar fashion with solutions being found for each spatial element and integrated for the whole.

Once vector field variables have been defined, related functions can be established to create vector fields or to create derivatives of them. These include gradient, divergence, surface normals, flux and curl. The next sections briefly discusses some of these possible functions and other applications of vector fields.

Gradient

The grad operator is the most immediately practical vector concept to implement since the result of grad on a continuous surface produces the vector field of gradient which is readily realized as the two scalar fields slope and aspect, two parameters commonly found in environmental models. Although grad is conceptually a continuous operations, calculating grad on discrete spatial data models is best accomplished through the joint calculation of the

parameters of slope and aspect. Most simply, the grad operator can be implemented as two functions, one which returns slope and the other aspect. The derivation of slope and aspect from scalar fields has been discussed in Chapter 6. Results from these individual slope and aspect functions can be converted to their equivalent representation as the 3 components of a Cartesian coordinate vector.

Divergence

While gradient uses a scalar to produce a vector, divergence begins with a vector. Since vectors are not widely implemented in most databases, the need for divergence operators has not been recognized. However, if vector fields are made available to modelers, then this basic operation will be needed.

Plane element vectors

The plane element vector representation discussed earlier provides an interesting alternative technique for the storage of gradient data in cellgrid spatial data models. A plane element vector gives both the normal to the surface, an important input to many solar irradiance models, and the area of the tilted plane, also a useful quantity in radiation models. As well, plane element vectors may also provide a new form in which the slope and aspect attributes attached to TIN triangles may be stored.

This chapter has explored some basic issues related to the use of vector fields and their manipulation and representation in GISs. We have reviewed some basic principles about vectors and have suggested how vector representations and associated vector functions may be incorporated into environmental models through the field variable concept. Since a mechanism for expressing vector fields has not been previously available, the true value of these concepts cannot be known until it has been widely incorporated into modeling efforts.

CHAPTER 8 - IMPLEMENTING THE STRATEGY

Since the concept of field variables can be generically described, the implementation of field variables is relatively straightforward. This chapter demonstrates how the strategy can be used in the development of a spatially distributed environmental model and then discusses some of the implementation issues.

A. AN EXAMPLE - THE SOIL WATER BUDGET

To begin, we consider Thornthwaite's relatively simple, non-spatial soil moisture balance, or water budget, model. This model has proven a useful tool for the longterm analysis of water needs as well as a method for quantitatively describing different climates. A discussion of the soil water balance is traditional in introductory physical geography texts (cf. Strahler and Strahler, 1989; Marsh, 1987; Briggs and Smithson, 1985). We begin with a brief outline of the model and then discuss how to express the model in the syntax developed here.

Thornthwaite's model

In this model, the amount of soil moisture at any time is a function of evapotranspiration and precipitation (Thornthwaite and Mather, 1955). Soil moisture is a store of water available to plants when precipitation is insufficient. In the model's simplest form, its value may vary from 0 to a theoretical maximum. Moisture is removed from the store by evaporation and transpiration and replaced by precipitation. If more precipitation falls than is needed for evapotranspiration, it replaces any soil moisture deficit up to the maximum. Surplus beyond that needed to replenish the soil moisture is removed from the system and is often regarded as a contribution to streamflow. Since this model does not consider small fluctuations in rates, it is best applied on a monthly basis, i.e. total monthly precipitation, storage at the end of a month.

Given the following values:

PET - average monthly potential evapotranspiration

P - average monthly precipitation

FC - soil storage capacity (field capacity, a constant)

the model calculates the following :

S - soil moisture storage at the end of the month

ΔS - change in soil moisture storage during the month

D - total monthly soil moisture deficit

AET - total monthly actual evapotranspiration

R - total monthly surplus water (runoff)

All values are in water depth. The governing equations discretized to time slices of one month are:

$$\Delta S = P - PET$$

$$AET = P - \Delta S$$

except that since $0 \leq S \leq FC$ and $AET \leq PET$:

$$\text{if } \Delta S + S_{t-1} < 0, \text{ then } \Delta S = S_{t-1}$$

$$\text{if } \Delta S + S_{t-1} > FC, \text{ then } \Delta S = FC - S_{t-1} \text{ and } AET = PET$$

$$D = PET - AET$$

$$R = P - AET$$

$$S_t = S_{t-1} + \Delta S$$

Given these governing equations, the pseudo-code routine for calculating one month's values is as follows. Note that the notation [t] is used to refer to a single month's values.

```
define procedure monthly-values {
    dS[t] = P[t]-PE[t] ;
    if dS[t]+S[t-1] > FC, then dS[t] = FC-S[t-1] ;
    if dS[t]+S[t-1] < 0, then dS[t] = S[t-1] ;

    AET[t] = P[t]-dS[t] ;
    if AET[t] > PET[t], then AET[t] = PET[t] ;

    D[t] = PET[t]-AET[t] ;

    R[t] = P[t]-AET[t]-dS[t] ;
}
```

Calculations are performed on a monthly basis beginning with an arbitrary choice of starting month. A guess for the value of the previous month's storage must be made to begin the monthly calculations. A good guess is (P-PET). A maximum of two iterations through the annual cycle will produce a stable result. Table 6 illustrates the results derived from this model when monthly PET and P are given and FC is estimated to be 100 mm.

A simple point-based query version

A very simple version of this model may be implemented as a point query. Given two sets of base maps of monthly precipitation and potential evapotranspiration, we wish to design a model which allows the user to select a specific location and outputs a water balance chart or graph for that location. While this model could be designed as a query of the output field at a single point, to illustrate how the computational structure can be used, we conceptualize the query location as a separate field, POINT, which contains the location of the single query point as the computational structure. Using the pseudo code developed here we could describe this model as follows:

Table 6 - Water Balance for Vancouver, B.C.

| | J | F | M | A | M | J | J | A | S | O | N | D | Yr. |
|------------|-----|-----|-----|-----|----|-----|-----|-----|----|-----|-----|-----|------|
| PET | 6 | 13 | 27 | 48 | 77 | 103 | 119 | 104 | 70 | 44 | 21 | 11 | 643 |
| P | 218 | 147 | 128 | 85 | 72 | 62 | 31 | 43 | 92 | 147 | 210 | 222 | 1457 |
| P-PET | 212 | 134 | 101 | 37 | -5 | -41 | -88 | -61 | 22 | 103 | 189 | 211 | |
| ΔS | 0 | 0 | 0 | 0 | -5 | -41 | -54 | 0 | 22 | 78 | 0 | 0 | |
| S | 100 | 100 | 100 | 100 | 95 | 54 | 0 | 0 | 22 | 100 | 100 | 100 | |
| AET | 6 | 13 | 27 | 48 | 77 | 103 | 85 | 43 | 70 | 44 | 21 | 11 | 548 |
| D | 0 | 0 | 0 | 0 | 0 | 0 | -34 | -61 | 0 | 0 | 0 | 0 | 95 |
| R | 212 | 134 | 101 | 37 | 0 | 0 | 0 | 0 | 0 | 25 | 189 | 211 | 909 |

All values shown are in mm.

```

define model COMP: (irregular points: 8x107) ;
field PRECIP: model(irregular_points: 8x104),
    measurement(num), time (1 month, yy/1, 12) ;
field PET:
    model(cellgrid:0°N,0°W,0°,10°,10°,36,18),
    measurement(num), time (1 month, yy/1, 12) ;
field S,D,AET,R,dS: model(COMP), measurement(num),
    time(1 month, yy/1, 12) ;

read PRECIP, PET ;
input POINT ;

for each month, cycling from t=12 to t=1 {
    do procedure (monthly-values)
    repeat until old S[t] = new S[t] } ;

output P,PET,D,R,S,AET ;

```

Since all but two of the variables are in the single point structure, all calculation will be done in this computational structure. Although it would be much more efficient to convert P and PET to single scalar values and then calculate the result as simple scalar variables, this example shows the simplicity of the approach. The next example makes better use of field variables.

A grid-based map version

In this version, the complexity is increased by adding additional field variables to determine PET and calculating the result across the region under study (continental or subcontinental in scope) displaying the results as a monthly series of cellgrid based maps.

Based on field studies (Christiansen, 1968), the relationship between potential evapotranspiration and the evaporation measured from a shallow pan of water has been found to have an annual average relationship of

$$PET = 0.75 PAN$$

where *PAN* is pan evaporation (in cm) which itself is calculated from a combination of climatic and other variables:

$$PAN = 0.324 C_T C_W C_H C_S C_E R_{et} \quad (9-1)$$

$$C_T = 0.463 + 0.425 \left(\frac{T}{20} \right) + 0.112 \left(\frac{T}{20} \right)^2 \quad (9-2)$$

$$C_W = 0.672 + 0.406 \left(\frac{W}{6.7} \right) - 0.078 \left(\frac{W}{6.7} \right)^2 \quad (9-3)$$

$$C_H = 1.035 + 0.240 \left(\frac{H}{0.6} \right)^2 - 0.275 \left(\frac{H}{0.6} \right)^3 \quad (9-4)$$

$$C_S = 0.340 + 0.856\left(\frac{S}{0.8}\right) - 0.196\left(\frac{S}{0.8}\right)^2 \quad (9-5)$$

$$C_E = 0.970 + 0.030\left(\frac{E}{305}\right) \quad (9-6)$$

and

$$R_{et} = f(\text{latitude, date})$$

where T = mean air temperature, °C
 W = mean wind velocity 2 m above ground, km/hr
 H = mean relative humidity, decimal
 S = mean % sunshine, decimal
 E = elevation, m

R_{et} = extraterrestrial radiation, expressed as equivalent depth of vaporized water, cm

Under standard climatic conditions, each of the coefficients calculated in equations (9-2) to (9-6) equals 1 so that the total effect in equation (9-1) is to produce a 68% attenuation of the incoming solar radiation.

Pseudo-code for this revised model might look like this:

```
define model COMP:
    (cellgrid: 10°S, 50°W, 0°, 100km, 100km, 100, 100);

define model CLIMATE:
    model(irregular points: 105 km2) ;

field T, W, H, S, PRECIP, CT, CW, CH, CS: model(CLIMATE),
    measurement(numeric),
    time(1 month, yy/1, 12) ;

field ELEV: model(cellgrid:
    10°S, 50°W, 0°, 50km, 50km, 200, 200), measurement(numeric)
;

field RET, PET, AET, D, S, R, dS: model(COMP),
    measurement(numeric), time(1 month, yy/1, 12);

read CLIMATE(T, W, H, S, PRECIP), ELEV ;

define procedure (CLIMATE"=CLIMATE): {
    thiessen and partition } ;
define procedure (ELEV"=ELEV): {
    3x3 window mean } ;
define procedure (LAT(cellgrid)): {
    calculate latitude at center point of cell};
```

```

define procedure (RADIATION(X,y)): {
    determine radiation from on latitude X,
    date y};

int date ;
CLIMATE" = CLIMATE;
ELEV" = ELEV ;

for each month[t] {
    date = t*12+15 ;
    RET[t] = RADIATION(LAT(GRID), date) ;
    CT[t] = 0.463 + 0.425*(T"[t]/20) + 0.112*(T"[t]/20)^2
        ;
    CW[t] = 0.672 + 0.406*(W"[t]/6.7) -
        0.078*(W"[t]/6.7)^2 ;
    CH[t]= 1.035 + 0.240*(H"[t]/0.6)^2 -
        0.275*(H"[t]/0.6)^3 ;
    CS[t] = 0.340 + 0.856*(S"[t]/0.8) -
        0.196*(S"[t]/0.8)^2 ;
    PET =0.75*0.324 * RET[t] * CT[t] * CW[t] * CH[t] *
        CS[t] * (0.97 + 0.083*(ELEV"/305)) } ;

for each month, cycling from t=12 to t=1 {
    do procedure monthly-values ;
    repeat until old S[t] = new S[t] }

output P,PET,D,R,S,AET ;

```

A spatial interaction version

Finally, we consider a version of the water budget which includes spatial interaction. Here we wish to determine how much surplus flows past given locations during each time step assuming that all surplus water passes through the system within a single time step. Calculation of surplus water may proceed as outlined above. For this version it is necessary to add a statement which describes how the surplus can be integrated over the region. This may be expressed as

$$Q(x,y) = R(x,y) + \int R_{up} dt$$

where $Q(x,y)$ is the total quantity of water passing (x,y) , $R(x,y)$ is the runoff generated at point (x,y) and R_{up} is the runoff generated at all points upstream of (x,y) (the watershed).

We have discussed previously how integrals and watersheds might be calculated for each of the spatial data models. If we have defined both the watershed for each spatial element ([up]) and the integral function (integ) for the specific combination of spatial data models (both source and target), then the combined function would be simply:

$$Q = R + \text{integ}(R[\text{up}])$$

which should be inserted in the previous code after the monthly values procedure is iterated.

B. IMPLEMENTATION ISSUES

There are a number of ways this strategy can be implemented and several different reasons for doing so. In this section we consider the motivations for using this strategy and consider how implementation might be achieved.

At its most fundamental level, this strategy can be used simply as a thinking tool, as a means for conceptualizing the relationships of the model in a procedural fashion. By isolating issues related to the manipulation of the spatial data models, the difficulty of dealing with data model conversions is eliminated and the modeler can work directly with the variables of the model.

Pedagogically speaking this strategy emphasizes the fundamental differences between the representation and the reality that is being modeled. While reality can be modeled as abstract scalar and field variables, the variables themselves must be declared as specific spatial data models and the conversion procedures and other special operations defined in these terms. This separation of spatial operations from modeling operations helps clarify the limitations and restrictions inherent in working with discrete representations of a continuous world.

At a more practical level, this strategy maps out how a linkage between a model and a GIS may be implemented. There are several different ways this linkage could be achieved. These are related to the levels of GIS/model integration proposed by Fedra (1993) and discussed in Chapter 3. The lowest level of integration, file exchange, is not addressed by this strategy since it can be achieved by map algebra. In that case, all data must be in the same format before modeling is begun.

The next level of integration is more appropriate for this strategy. This level involves a special interface program which manages the file format conversions so that file sharing is transparent to the user. At this level the strategy could be implemented as a compiler which converts the syntax of the strategy into calls to various data management systems and data model conversion subroutines and integrates these with the standard mathematical operations of the program. This solution suggests that implementation using a traditional procedural language is suitable.

The highest level of integration, in which the model becomes one of the analytical functions inside a GIS or the GIS becomes an option in the file management and output components of the model suggests a final approach to the implementation of this strategy. Here, the various functions and conversion procedures can be "unbundled" and made available as individual tools in a GIS or model toolbox. The model itself uses these functions directly. This particular implementation approach requires direct implementation of the declaration of field variable statements as an integral part of the model. This solution may be most appropriate for implementation using an object oriented environment.

Implementation in scientific programming languages

In standard procedure-oriented languages such as C or FORTRAN, field variables can be specified as user defined data types. Properties can be stored as flags or values which can be accessed when necessary to signal the appropriate subroutine or to insert the appropriate value in a routine. Specialized functions and subroutines using these user defined data types can be created. With procedural languages, details of the implementation of field data model conversions and related procedures do not need to be handled within the program itself, but may be developed as standard subroutine libraries accessed through operations on field variables. As well, these standard libraries may be written to interface directly between

various programming languages and specific GISs. It is possible to visualize a full range of subroutines contained in a set of GIS libraries for the standard programming languages.

With object-oriented programming languages, the six field data models may be defined as abstract data types with associated methods determining appropriate conversion routines. Specific field variables can then be defined as instances of these object classes. Appropriate conversion procedures will be inherited. Instance-specific overloading of operators can be enabled by creating subclasses of the 6 object classes.

Thus it seems apparent that existing programming facilities which allow us to define abstract or user-defined data types and to establish standard procedures either as subroutines or encapsulated operations provide an immediate means by which field variables can be incorporated into mathematical models of environmental phenomena. We hope that this strategy will provide the impetus to take advantage of such opportunities.

GIS implementation issues

Linking a mathematical model to a GIS using this strategy provides explicit details about how the spatial data is to be manipulated. By separating the manipulation of the spatial data from the mathematics of the model, the spatial operations which need to be linked become readily apparent. This detail can be used directly to design the interface between model and GIS. The conversion procedures and functions defined by the modeler must be translated into functions in the available GIS and other software.

An advantage of using this strategy lies in the fact that the modeler can specify spatial operations to any level of detail desired. If he or she is unfamiliar with the types of operations that may or should be performed on spatial data, information about spatial data model conversion procedures and even field variable declarations can be omitted from the pseudo-code. These elements can be specified as required when the model is implemented with the GIS.

This chapter has considered some of the issues related to the implementation of this strategy for dealing with spatial continuity. The development of a simple model of water balance demonstrated how the syntax outlined in Chapter 7 can be used. We have also considered the ways in which this strategy can be used and some of the tools available today for its implementation.

CHAPTER 9 - CONCLUSION

In this final chapter, we return to the beginning, review the progress made towards the original goals and consider the value and implications of the strategy developed here. The premise of this research is that environmental modeling is based on an essentially continuous world. Processes, phenomena and time are all continuous. While there are many advantages to the integration of the spatial data management and analysis tools of GIS with environmental modeling, the antithesis between the continuity of the natural world and the discrete nature of the computer makes it very difficult to obtain perfect integration. To assist in easing this integration, this research considers how continuous phenomena can be handled directly within the context of GIS. The result is the outline of a strategy for dealing with spatial continuity through the use of field variables.

The fundamental theses of this research were laid out in Chapter 1. They are:

- 1) it is both desirable and possible to separate the mathematical operations which will be performed on data about spatially continuous phenomena from the form of spatial discretization used to represent those phenomena in the computer, and
- 2) this separation allows issues about the implementation and manipulation of these digital representations to be dealt with automatically, without external control, in such a way that they can be considered extraneous to the modeling task.

By examining aspects of the discretization of space and processes and the manner in which discrete representations of continuous phenomena can be manipulated, we sought to affirm these theses. The strategy developed provides a means by which to achieve the separation that is proposed.

A. SUMMARY OF THE RESEARCH

Discretizing processes and phenomena

Environmental modeling deals with continuous processes in continuous space. Numerical solutions for the discretization of continuous processes in continuous space exist and are widely used. Of course, the discretizations necessary for such modeling efforts have direct links to the form of discretization that can be used to represent data about continuous phenomena, or fields.

A field is defined as an entity which is distributed over space and whose properties are functions of space coordinates. There are six spatial data models which can be used to represent fields. They are the constant piecewise models, cellgrids and polygons, the surface models, TINs and contour models, and the point models, pointgrids and irregular point models. By taking advantage of the spatial autocorrelation between values at different locations in a field, each of these spatial data models provides an inherent set of rules or assumptions that can be made about the form of the variation over space. Constant piecewise

models provide a value at every location, but these values are averages of the value over a region. Surface models provide a means for estimating the variation between locations at which values are known. Point models provide exact data at specific locations, but do not provide any information about how to determine the value at locations between known points. These different assumptions inherent in each model are crucial in determining how data in these models can be manipulated.

Mathematical modeling with spatial data models

Having recognized that data about continuous phenomena must be stored and manipulated as discrete spatial data models, we next considered the implications of performing mathematical operations upon these representations of reality. It is fundamental to recognize that in order to do mathematics on fields, it is necessary to determine the value of every variable being used in a single operation at exactly the same location. This implies that any representation of a continuous phenomenon must permit the determination of the value of that phenomenon at any location. This is consistent with the definition of fields and with the assumptions inherent in the various spatial data models. However, to use discrete representations of fields in these models, it is essential that consideration be given to the means by which values can be determined in one spatial data model for locations which are specified in others. This, it turns out, is the crux of the problem in the clear separation between mathematical operations and the representation of spatial data.

Fortunately, algorithms which can be used to convert one spatial data model to others are widely available and implemented in many different forms (e.g. standalone software modules, contour modeling packages, full-scale GISs). However, in some cases there are several different ways one representation of reality can be converted to another representation. The choice of which procedure to use should be determined by the reality being represented rather than by the form of the representation itself. We need to consider both the interpolation procedure which can be used to conceptually create a continuous surface from the discrete model and the sampling procedure which is used to sample the continuous surface for representation in the discrete target model. This implies that the appropriate conversion procedure should be determined by the characteristics of datasets themselves rather than by the modeling environment. This, in turn, supports the contention that the mathematical operations of a mathematical model should be treated separately from issues related to the form of spatial discretization.

Most of the operations which must be performed to convert spatial data from one representation to another can be determined simply by a consideration of the characteristics of each spatial data model. These conversions were considered in depth in Chapter 5 and summarized in two conversion matrices, one for numerical data and the other for categorical data. However, since similar spatial data models can be used to represent very different phenomena, in a few cases it is desirable that information about how a spatial dataset should be manipulated be attached to specific datasets. In particular, since point models do not explicitly include rules or assumptions about how values vary between the points where values are known, knowledge about the reality being represented should be used to determine appropriate interpolation procedures. For example, in irregular point data models, the values at the points included in the dataset may be representative of the neighborhood of each point or they may be the critical points in the field (i.e. if the variable value of the field is visualized as a topographic surface these would be the pits, peaks, ridges and other places where breaks in slope occur). In the first case a kriging approach may be appropriate whereas in the second case a triangulation procedure might be used. This information cannot be deduced from the representation itself but must be supplied externally. Since the appropriate conversion is determined by the representations of the phenomena rather than the operations to be performed on them, we suggest that such information should be attached to

the datasets at the time of their creation since this is the only time when reality and its representation can be directly compared. Such expert knowledge about the phenomenon can be used to encapsulate appropriate procedures with the data about each field before they are incorporated into the model. Since encapsulation is currently the exception rather than the norm, in most cases this information must be deduced and appended by the modeler him or herself who, it is hoped, understands at least a little about the nature and sampling of the phenomenon being represented. In any case, it seems clear that the addition of any relevant knowledge must improve the modeling results.

The conversion of categorical fields also requires special attention since the manipulation of such fields can rarely be handled in a simple mathematical context. Rules about how to perform model conversions can only be determined by consideration of the phenomena represented by the datasets. However, such rules can be specified independently of any mathematical operations that will be performed.

A strategy for dealing with spatial continuity

In order to promote the desired separation of operations from representation that is sought in the theses of this research, we have developed a strategy for dealing with spatial continuity in mathematical models. This development has been driven by a realization that designing and coding a mathematical model is an entirely different task than accessing and manipulating spatial data in a GIS. On the one hand modelers can use well-known and well-structured algebraic and computer languages, following widely accepted and proven rules for substitution and solution. On the other hand, when manipulating spatial data for use in the models, modelers have only the idiosyncratic language of a specific GIS to work with. There are no widely accepted common rules and defaults to guide how spatial data are used in environmental models and no simple way to express the transformations and manipulations that are necessary to incorporate the spatial data into the model. The strategy we have developed for handling field variables addresses part of this problem. Specifically this strategy seeks to:

- allow expression and manipulation of variables and data about continuous phenomena in common symbolic languages. In other words, the strategy should be capable of being incorporated into computer language implementations of environmental models. This is in direct contrast to the natural language-like structure of Tomlin's map algebra and is more amenable to the scientific environment.
- eliminate the necessity to consider the form of the spatial discretization (the data model) whenever possible. While we believe it is desirable and possible to achieve this objective for most operations, it is necessary to provide for input of additional information for some operations.
- provide a syntax for incorporating primitive operations appropriate for environmental modeling with fields which are not yet available in GIS or common programming languages. These include operations to perform discrete versions of "differentiation" and "integration" on variables representing fields and the incorporation of the concept of vector fields.
- guide and enable the rapid development of direct linkages between environmental models and any GIS.

The core of this strategy is the definition of the *field variable type* and the development, in Chapter 6, of a syntax for declaring the characteristics of a specific field

variable. The earlier examinations of the characteristics of the various spatial data models and the operations which are performed on spatial data provided the means by which to determine the critical characteristics which need to be specified. Required characteristics include information about the spatial data model and its density and the measurement system used (i.e. numerical or categorical). These specifications allow the spatial operations required by the mathematics of the model to be performed without external control, such that the implementation and manipulation of the spatial data can be considered extraneous to the modeling task

In addition to the declaration of field variables, the strategy outlines several other parameters that may be defined for use in the mathematical operations of the model. These include establishing the specific algorithms which will be used for specific model-to-model conversions and defining the spatial structure which will be used for the computation of the model. The definition of a computational structure allows a finite difference grid or a set of finite elements to be specified and used as the default target structure during execution of the mathematical operations in the model. All of these additional parameters are designed to establish everything needed to conceptualize and manage the manipulation of the spatial operations separately from the mathematical operations.

Working with field variables

Once field variables have been declared they can be used directly in mathematical statements. Whenever a field variable is invoked, its declared characteristics will determine how operations will be performed upon it. As well, given the existence of field variables, it is possible to formulate new functions designed specifically for fields. In particular, *integrate* is an operation that has wide application in models working with continuous phenomena. Its implementation can be relatively easily specified for each of the field spatial data models. *Slope* and *aspect*, variables which are important in many environmental models, can be seen as field functions. While these are currently implemented in traditional forms in many GISs, their expression as field functions would be useful in the context of this strategy. Certain statistical operations like *mean* and *standard deviation* have special meanings when performed on fields, but their implementation is simply a matter of defining appropriate algorithms for the various data models.

As well, we suggest the identification of two reserved field variables, *lat* and *long*. Like other fields, latitude and longitude can be determined for any location. Since mathematical statements using field variables must be performed on values about individual locations, it is easy to evaluate the values of these reserved variables for use in mathematical operations.

The problems of working with categorical data which represent continuous phenomena have also been explored. We have already reviewed the approach which has been devised to specify rules about how to perform model conversions on categorical data. The strategy also outlines a technique for defining new operations for categorical data (i.e. the "overloading" of standard operators for certain fields) so that this data can be used directly with numerical fields.

In addition to relatively easily implemented scalar field variables, the concept of field variables also allows the consideration of vector fields. Since vector fields have not been available in GISs, their use has been restricted to a small group of very specialized, highly technical mathematical models that have little in common with current GISs. Vector fields and related operations such as divergence and gradient do provide some useful new ways of conceptualizing interactions in continuous phenomena. However, before vector fields and the related operations can become widely used by environmental modelers using spatial data

stored in GISs, it is necessary that vector fields become accessible and manageable within available software. Vector field variables provide a first step in this direction.

B. CRITIQUE AND ANALYSIS OF RESULTS

How successfully have the objectives of this research been achieved? Let us consider, first, the main theses outlined above. The first thesis which states that separation of mathematical operations from the spatial discretization is possible has been clearly supported by the development of the strategy which allows details about the manipulation of spatial data to be handled independently from the mathematics. That this is desirable has been demonstrated in the clarity that the strategy brings to the manner in which data about continuous phenomena can now be handled during the development stage of mathematical models. The second thesis which suggests that manipulation of the spatial data can be handled extraneously to the modeling task has been the focus of much of the latter part of this document. Field variables, their declarations and related aspects of the strategy are designed specifically to demonstrate that this thesis can be achieved.

The objectives which the strategy has been designed to meet are also listed above in this chapter. The first objective seeks to ensure that the strategy can be immediately useful in the current modeling environment. Chapter 8 discusses some of the implementation issues for this strategy and points out that there are many ways in which it can be used. Each of these uses allows incorporation of the strategy at a different level into existing modeling and programming procedures. The second objective, elimination of consideration of the form of spatial discretization, has been achieved through the separation of mathematics from spatial data manipulation. Several primitive operations devised for field variables, the theme of the third objective, have been described. Finally, as Chapter 8 has outlined, the clarity brought to the distinction between mathematics and spatial data should encourage and aid the development of direct linkages between GIS and environmental models.

Value of the strategy

Chapter 8 outlined a number of ways this strategy may be used. It was suggested that it may be used as a thinking tool to aid in the conceptualization of the mathematical model. Alternatively, it has value as a pedagogic tool through its emphasis on the distinction between the spatial data and the reality that it represents. Perhaps most importantly, the strategy can be used to help map out how a linkage between a specific mathematical model and a GIS may be implemented.

The strategy presented here provides a flexible approach to handling variables representing continuous phenomena. Modelers can express as much or as little as they prefer about how these variables are to be represented and manipulated. The use of field variables allows the spatial data to be incorporated into the models in a form which is similar to their conceptual form in mathematical equations. This strategy also provides a means by which the manipulations which are to be performed on spatial data can be explicitly described.

A significant feature of this approach is that it does not assume a homogeneous database. Data can be stored in the form closest to that in which they have been collected. Conversion to the spatial data model required by a particular mathematical model is performed only when necessary. This helps to keep to a minimum the number of model conversions that are performed on a particular spatial dataset and thus reduces the inevitable information loss that occurs with each conversion.

Weaknesses in the strategy

The weaknesses in this strategy may be seen to stem from its generic form. Since the intent was to map out a strategy which could be used in the broad range of environmental modeling domains, it has been difficult to provide specific solutions to individual problems or to suggest tools that may be particularly useful in some cases. During the development of this strategy, specific details of the implementation of the strategy, such as how to implement specific interpolation procedures for specific spatial data models, often obscured the larger picture which was being sought. It was necessary to constantly pull back, to remember that implementation is not the difficult issue since many of these specific problems have already been addressed by other researchers. It is hoped that this strategy will allow a framework within which these specific solutions can be organized and made accessible to the widest audience.

These generic aspects may also lend an air of indecisiveness to the conclusions generated. Many issues were assumed away in order to prepare the groundwork for the development of this strategy. This is not to imply that accuracy, generalization, data modeling and so on are not important themes. What we have sought to do here is to provide a framework within which these diverse themes can be joined and dealt with explicitly and comprehensively. However, future research should explore these issues in the same detailed manner as mathematical operations have been explored here so that commonalities between procedures and operations can be uncovered and methods for codifying critical properties can be devised.

C. DIRECTIONS FOR THE FUTURE

Implementation of the strategy is the next step which must be made. The obvious relationship between the conceptual framework outlined here and the concepts of object orientation must be explored in detail. While many mathematical models will continue to be developed in traditional procedural languages, object orientation does provide a conceptual structure in which field objects can be defined and related procedures encapsulated.

There are many directions that implementation of this strategy should pursue. Implementation in a number of different environmental modeling domains would provide considerable opportunity for refinement of the generic concepts outlined here. The development of interfaces between various programming languages and GISs which interpret statements about field variables and translate them into computable elements is also highly desirable. Finally, the creation of libraries of field variable functions would be useful extensions to existing programming languages.

There are also several opportunities for extensions to the strategy outlined here. Certainly three dimensional fields merit examination within the framework developed in this document. Consideration of other types of continuous variables such as tensors and statistical fields could be an interesting and useful extension. Many of the issues raised and solutions provided are applicable in areas beyond the limited range of applications in the environmental modeling domains considered. Models of human activities frequently make use of density variables modeled as fields and thus may benefit from treatment similar to that given to models of physical processes. As well, the continuity of time was not considered in this analysis but it is apparent that a similar approach to the development of a general theoretical framework for handling time data may be useful.

Environmental modeling is an important and growing enterprise. The expanding availability of spatial databases and the rapid evolution of GIS software are encouraging many modelers to turn to GIS as a means of handling their spatial data needs. GIS developers must work to understand the needs of these modelers and to incorporate facilities which meet their current and future needs. This strategy is one step in this direction.

REFERENCES

- Abel, D.J., and M.A. Wilson, 1990. A systems approach to the integration of raster and vector data and operations, *Proceedings of the 4th International Symposium on Spatial Data Handling*, vol. 2, pp. 559-565.
- Aldred, B.K., and B.S. Smedley, 1974. *Urban Management System Technical Review*, internal technical paper UK-SC.0050, UK Scientific Centre, IBM United Kingdom Limited, Peterlee, UK.
- Anselin, L., 1989. *What is Special about Spatial Data? Alternative perspectives on spatial data analysis*, NCGIA Technical Paper 89-4, National Center for Geographic Information and Analysis, University of California Santa Barbara, CA.
- Arbia, G., 1989a. Statistical effect of spatial data transformations: a proposed general framework, in M.F. Goodchild and S. Gopal, eds, *Accuracy of Spatial Databases*, Taylor and Francis, pp. 249-259.
- Arbia, G., 1989b. *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*, Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Baker, W.L. 1989. A review of models of landscape change, *Landscape Ecology* 2(2):111-133.
- Baron, G.E., 1988. History and development, in J.R. Brauer, ed., *What Every Engineer Should Know About Finite Element Analysis*, Marcel Dekker, Inc., New York, pp. 1-6.
- Batty, M., 1989. Urban modelling and planning: reflections, retrodictions and prescriptions, in B. Macmillan, ed, *Remodelling Geography*, Basil Blackwell, Oxford, pp. 147-169.
- Beach, T., 1987. A review of soil erosion modeling, in D.A. Brown and P.J. Gersmehl, *File Structure Design and Data Specifications for Water Resources Geographic Information Systems*, Special Report No. 10, Water Resources Research Center, University of Minnesota, St. Paul, MN, Chapter 9.
- Bekey, George A., 1985. Simulability of dynamical systems, *Transactions of the Society for Computer Simulation* 2(1):57-71.
- Beven, K., 1987. Towards a new paradigm in hydrology, *Water for the Future: Hydrology in Perspective, Proceedings of the Rome Symposium*, April 1987, IAHS publication no. 164, pp. 393-403.
- Beven, K., 1989. Changing ideas in hydrology - the case of physically-based models, *Journal of Hydrology* 105:157-172.

- Boots, B.N., 1986. *Voronoi (Thiessen) polygons*, Concepts and techniques in modern geography series no. 45, Geo Books, Norwich, UK.
- Bourke, W., 1988. Spectral methods in global climate and weater prediction, in M.E. Schlesinger, ed., *Physically-Based Modelling and Simulation of Climate and Climatic Change - Part I*, Kluwer Academic Publishers, pp. 169-220.
- Bowers, C.E., A.F. Pabst, and S.P. Larson, 1972. *Computer Programs in Hydrology*, Bulletin 44, Water Resources Center, University of Minnesota, MN.
- Brauer, J.R., 1988. *What Every Engineer Should Know About Finite Element Analysis*, Marcel Dekker, Inc., New York.
- Breidenback, A.W., 1976. Keynote address: Toward a common language, in Ott, W.R., ed, *Proceedings of the EPA Conference on Environmental Modeling and Simulation*, April 19-22, 1976, Cincinnati, Ohio, US Environmental Protection Agency, Report EPA 600-9-76-016, Washington DC.
- Breunig, M., and A. Perkhoff, 1992. Data and system integration for geoscientific data, *Proceedings of the 5th International Symposium on Spatial Data Handling*, Charleston, SC, International Geographical Union, Commission on GIS, Humanities and Social Sciences Computing Lab, University of South Florida, Columbia, SC, pp. 272-281.
- Briggs, D., and P. Smithson, 1985. *Fundamentals of Physical Geography*, Hutchinson and Co. (Publishers) Ltd., Dover, NH.
- Brown, D.A., K.L. Anderson, P.J. Gershmehl, 1987. Hydrographic data for a water resources GIS, in D.A. Brown and P.J. Gersmehl, *File Structure Design and Data Specifications for Water Resources Geographic Information Systems*, Special Report No. 10, Water Resources Research Center, University of Minnesota, St. Paul, MN, Chapter 12
- Brown, J.W., M.R. Walsh, R.W. McCarley, A.J. Green, Jr., and H.W. West, 1974, *Models and Methods Applicable to Corps of Engineers Urban Studies*, Misc. Paper H-74-8, U.S. Army Engineers Waterways Experiment Station, Vicksburg, MI.
- Burrough, P.A., 1983. *Principles of Geographic Information Systems and Land Resources Assessment*, Oxford University Press.
- Burrough, P.A., 1991. Soil information systems, in D.J. Maguire, M.F. Goodchild and D.W. Rhind, eds, *Geographical Information Systems: Principles and Applications*, Longman, London, pp. 153-169.
- Burrough, P.A., 1992. Are GIS data structures too simple minded?, *Computers & Geosciences* 18(4):395-400.
- Casti, John L, 1989. *Alternate Realities: Mathematical Models of Nature and Man*, John Wiley & Sons.
- Chan, K.K.L., and C.D. Tomlin, 1991. Map algebra as a spatial language, in *Cognitive and Linguistic Aspects of Geographic Space*, D.M. Mark and A.U. Frank, eds, NATO ASI Series D: Behavioural and Social Sciences, Vol. 63, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 351-360.

- Chorley, R.J., and P. Haggett, 1967. *Models in Geography*, Methuen.
- Chorley, R.J., ed., 1972. *Spatial Analysis in Geomorphology*, Harper and Row Publishers, London.
- Chrisman, N.R., 1991. The error component in spatial data, in D.J. Maguire, M.F. Goodchild and D.W. Rhind, eds, *Geographical Information Systems: Principles and Applications*, Longman, London, pp. 165-174.
- Christiansen, J.E., 1968. Pan evaporation and evapotranspiration from climatic data, *ASCE Journal of the Irrigation and Drainage Division* 94:243-265.
- Corbett, J.D., and P.J. Gersmehl, 1987. Terrain data for water resources GIS, in D.A. Brown and P.J. Gersmehl, *File Structure Design and Data Specifications for Water Resources Geographic Information Systems*, Special Report No. 10, Water Resources Research Center, University of Minnesota, St. Paul, MN, Chapter 11.
- Costanza, R. and F.H. Sklar, 1985. Articulation, accuracy and effectiveness of mathematical models: a review of freshwater wetland applications, *Ecological Modelling* 27:45-68.
- Couclelis, H., 1992. People manipulate objects (but cultivate fields): beyond the raster-vector debate in GIS, in A.U. Frank, I. Campari, U. Formentini, eds, *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, Lecture Notes in Computer Science 639, Springer-Verlag, pp. 65-77.
- Cox, N.J., 1989. Modelling data analysis and Pygmalions problem, in B. Macmillan, ed, *Remodelling Geography*, Basil Blackwell, Oxford, pp. 204-208.
- Cross, M., and A.O. Moscardini, 1985. *Learning the Art of Mathematical Modeling*, Halsted Press, New York.
- Daintith, J., and R.D. Nelson, 1989. *Dictionary of Mathematics*, Penguin Books, London.
- Dacey, M.F., 1967. Description of line segments, *Northwestern Studies in Geography* 13:277-287.
- Dacey, M.F., 1970. The syntax of a triangle and some other figures, *Pattern Recognition* 2:11-31.
- Dacey, M.F., 1971. Poly: a two dimensional language for a class of polygons, *Pattern Recognition* 3:197-208.
- Dale, V.H., T.W. Doyle and H.H. Shugart, 1985. A comparison of tree growth models, *Ecological Modelling* 29:145-169.
- Davis J.C., 1986. *Statistics and Data Analysis in Geology*, 2nd edition, John Wiley and Sons.
- Decision Images, undated. Various publicity materials on MapBox from Decision Images, Inc, Skillman, NJ.
- Desai, C.S., 1979. *Elementary Finite Element Method*, Prentice-Hall, Inc, Englewood Cliffs, NJ.

- EPA, 1990. *Environmental Monitoring and Assessment Program: Overview*, EPA/600/9-90/001, US Environmental Protection Agency, Office of Research and Development, Office of Modeling, Monitoring Systems and Quality Assurance, Washington, DC.
- ESRI, 1991. *Cell-based Modeling with GRID*, Environmental Systems Research Institute, Inc., Redlands, CA.
- Evans, I.S., 1972. General geomorphology, deviations of altitude and descriptive statistics, in R.J. Chorley, ed., *Spatial Analysis in Geomorphology*, Harper and Row Publishers, London, pp. 17-90.
- Evans, I.S., 1980. An integrated system of terrain analysis and slope mapping, in H. Hagedorn and M. Thomas, eds, *Perspectives in Geomorphology, Zeitschrift für Geomorphologie*, Supplementband 36:274-95.
- Evans, J.D., J. Ferreira, Jr., and P.R. Thompson, 1992. A visual interface to heterogeneous spatial databases based on spatial metadata, *Proceedings of the 5th International Symposium on Spatial Data Handling*, Charleston, SC, International Geographical Union, Commission on GIS, Humanities and Social Sciences Computing Lab, University of South Florida, Columbia, SC, pp. 282-293.
- Fedra, 1993. GIS and environmental modeling, in M.F. Goodchild, B.O. Parks and L.T. Steyaert, eds, *Environmental Modeling and GIS*, Oxford University Press, Chapter 5.
- Flowerdew, R., and M. Green, 1992. Developments in areal interpolation methods and GIS, *Annals of Regional Science* 26:67-78.
- Frank, A.U., B. Palmer and V.B. Robinson, 1986. Formal methods for the accurate definition of some fundamental terms in physical geography, *Proceedings of the Second International Symposium on Spatial Data Handling*, July 1986, Seattle, WA.
- Franklin, J.F., C.S. Bledsoe and J.T. Callahan, 1990. Contributions of the Long-Term Ecological Research Program, *BioScience* 40(7):509-523.
- Franklin, S.E., 1987. Geomorphometric processing of digital elevation models, *Computers and Geosciences* 13(6):603-609.
- Freeman, C., 1973. Malthus with a computer, in Cole, H.S.D., C. Freeman, M. Jahoda and K.I.R. Pavitt, eds, *Thinking about the future: A critique on the limits to growth*, Chatto and Windus, London.
- Freeze, R.A., and R.L. Harlan, 1969. Blueprint for a physically-based digitally-simulated hydrologic response model, *Journal of Hydrology* 9:237-258.
- Gerald, C.F., and P.O. Wheatley, 1989. *Applied Numerical Analysis*, 4th edition, Addison Wesley.
- Getis, A., and B. Boots, 1978. *Models of Spatial Processes*, Cambridge University Press, Cambridge.
- GIS World, 1991. *1991-92 International GIS Sourcebook: Geographic Information System Technology in 1991*, GIS World Inc., Fort Collins, CO.

- Golledge, R., 1982. Fundamental conflicts and the search for geographical knowledge, in P. Gould and G. Olsson, eds, *A Search for Common Ground*, Pion, London, pp. 11-23.
- Goodchild, M.F., 1988. Stepping over the line: technological constraints and the new cartography, *American Cartographer* 15:311-320.
- Goodchild, M.F., 1989. Modeling error in objects and fields, in M.F. Goodchild and S. Gopal, eds, *Accuracy of Spatial Databases*, Taylor and Francis, pp. 107-113.
- Goodchild, M.F., 1990a. Geographic Information Systems and environmental modeling, *Proceedings of the Seventh International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology*, Boston, American Meteorological Society, pp. 5-10.
- Goodchild, M.F., 1990b. Modeling error in spatial databases, *Proceedings of GIS/LIS '90* 1:154-162.
- Goodchild, M.F., 1991. Integrating GIS and environmental modeling at global scales, *Proceedings of GIS/LIS '91* 1:117-127.
- Goodchild, M.F., 1992. Geographical data modeling, *Computers and Geosciences* 18(4):401-408.
- Goodchild, M.F., B.O. Parks and L.T. Steyaert, eds, 1993. *Environmental Modeling with GIS*, Oxford University Press, New York.
- Goodchild, M.F., and N.S-N. Lam, 1980. Areal interpolation: A variant of the traditional spatial problem, *Geo-Processing* 1:297-312.
- Goodchild, M.F., and S. Gopal, eds, 1989. *Accuracy of Spatial Databases*, Taylor and Francis, London.
- Goulter, I.C., and D. Forrest, 1987. Use of Geographic Information Systems (GIS) in River Basin Management, *Water Science and Technology* 19(9):81-86.
- Grelot, J-P, 1985. Archaic data models or hardware as a concept killer, in M. Blakemore, ed., *Proceedings - Auto Carto London : International Conference on the acquisition, management and presentation of spatial data*, London, September 1986, Auto Carto London., pp. 572-577.
- Haggett, P., and R.J. Chorley, 1967. Models, paradigms and the new geography, in R.J. Chorley and P. Haggett, eds, *Models in Geography*, Methuen, pp. 19-41.
- Haggett, P., and R.J. Chorley, 1989. From Madingley to Oxford: A foreword to Remodelling Geography, in B. Macmillan, ed, *Remodelling Geography*, Basil Blackwell, Oxford, pp. xv-xx.
- Haines-Young, R.H., and J.R. Petch, 1986. *Physical Geography: Its nature and methods*, Harper and Row.
- Haining, R., 1990. *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge University Press, Cambridge.

- Han, Y.-J., 1989. Modeling and simulation of the general circulation of the ocean, in Schlesinger, M.E., ed, *Physically-Based Modeling and Simulation of Climate and Climatic Change - Part I*, Kluwer Academic Publishers, pp. 465-508.
- Harvey, D., 1969. *Explanation in Geography*, Edward Arnold, London.
- Hay, L.E., W.A. Battaglin, R.S. Parker and G.H. Leavesley, 1993. Modeling the effects of climate change on water resources in the Gunnison River Basin, Colorado, in M.F. Goodchild, B.O. Parks and L.T. Steyaert, eds, *Environmental Modeling and GIS*, Oxford University Press, Chapter 16.
- Henderson-Sellers, A., 1989. Climate, models and geography, in B. Macmillan, ed, *Remodelling Geography*, Basil Blackwell, Oxford, pp. 117-146.
- Herring, J.R., 1991. The mathematical modeling of spatial and non-spatial information in Geographic Information Systems, in D.M. Mark and A.U. Frank, eds, *Cognitive and Linguistic Aspects of Geographic Space*, NATO ASI Series D: Behavioural and Social Sciences, Vol. 63, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 313-350.
- Hunsaker, C.T., R.A. Nisbet, D.C.L. Lam, J.A. Browder, W.L. Baker, M.G. Turner and D.B. Botkin, 1993. Spatial models of ecological systems and processes: the role of GIS, in M.F. Goodchild, B.O. Parks and L.T. Steyaert, eds, *Environmental Modeling and GIS*, Oxford University Press, Chapter 22.
- Jacoby, S.L.S., and J.K. Kowalik, 1980. *Mathematical Modeling with Computers*, Prentice-Hall, Inc., Englewood Cliffs, N.J.
- James, L.D., and S.J. Burges, 1982. Selection, calibration and testing of hydrologic models, in C.T. Haan, H.P. Johnson and D.L. Brakensiek, *Hydrologic Modeling of Small Watersheds*, American Society of Agricultural Engineers, Monograph 5, pp. 437-472.
- Jeffers, J.N.R., 1982. *Modeling*, Chapman and Hall, New York.
- Jensen, M.E., and R.H. Haise, 1963. Estimating evapotranspiration from solar radiation, *ASCE Proceedings* 89(IR4):15-41.
- Jenson, S.K., 1991. Application of hydrologic information automatically extracted from digital elevation models, *Hydrological Processes* 5(1):31-44.
- Jenson, S.K., and J.O. Domingue, 1988. Extracting topographic structure from digital elevation model data for geographic information system analysis, *Photogrammetric Engineering and Remote Sensing* 54:1593-1600.
- Johnson, L.B., 1990. Analyzing spatial and temporal phenomenon using geographical information systems: a review of ecological applications, *Landscape Ecology* 4(1):31-43.
- Jorgensen, S.E., 1990. Modelling concepts, in S.E. Jorgensen, ed., *Modelling in Ecotoxicology*, *Developments in Environmental Modelling* 16, Elsevier.
- Journel, A.G., 1986. Geostatistics: Models and tools for the earth sciences, *Mathematical Geology* 18(1):119-140.

- Kelmelis, J.A., and R.D. Watts, 1991. Terrestrial process modeling - A challenge for cartographers, *ACSM Proceedings*, Vol 2, ACSM, Baltimore, MD, pp. 141-149.
- Khoshafian, S., and R. Abnous, 1990. *Object Orientation: Concepts, Languages, Databases, User Interfaces*, John Wiley and Sons.
- Kirkby, M., 1989. The future of modelling, in B. Macmillan, ed, *Remodelling Geography*, Basil Blackwell, Oxford, pp. 255-272.
- Kumler, M.P., 1992. *An Intensive Comparison of TINs and DEMs*, unpublished PhD dissertation, Department of Geography, University of California, Santa Barbara, CA.
- Lam, N.S-N., 1983. Spatial interpolation methods: a review, *The American Cartographer* 10(2):129-149.
- Langran, G., 1989. A review of temporal database research and its use in GIS applications, *International Journal of Geographical Information Systems* 3:215-232.
- Ledbetter, L., and B. Cox, 1985. Software-ICs, *Byte*, June 1985, pp. 307-316.
- Lee, D.B., Jr., 1973. Requiem for large-scale models, *American Institute of Planners Journal* 39(3):163-178.
- Lee, T.J., R.A. Pielke, T.G.F. Kittel and J.F. Weaver, 1993. Atmospheric modeling and its spatial representation of land surface characteristics, in M.F. Goodchild, B.O. Parks and L.T. Steyaert, eds, *Environmental Modeling and GIS*, Oxford University Press, Chapter 10.
- Lindhult, M.S., P.J. Godfrey and J.G. Fabos, 1988. *Geographic Information Systems in Water Resources Research*, Completion Report Project No. G1023-04, Water Resources Research Center, University of Massachusetts, Amherst, MA.
- Macmillan, B., ed, 1989. *Remodelling Geography*, Basil Blackwell, Oxford.
- Maidment, D.R., 1993. GIS and hydrologic modeling, in M.F. Goodchild, B.O. Parks and L.T. Steyaert, eds, *Environmental Modeling and GIS*, Oxford University Press, Chapter 14.
- Marani, A., 1988. On mathematical models of natural systems, in A. Marani, ed., *Advances in Environmental Modelling, Developments in Environmental Modelling* 13, Elsevier, pp. 5-26.
- Marsh, W.M., 1987. *Earthscape: A physical geography*, John Wiley & Sons.
- McCullagh, 1988. Terrain and surface modelling systems: theory and practice, *Photogrammetric Record* 12(72):747-779.
- McHarg, I.L., 1969. *Design with nature*, Natural History Press, Garden City, N.Y.
- Metcalf, M., and J. Reid, 1990. *Fortran 90 Explained*, Oxford University Press.
- Molenaar, M., 1991. Status and problems of geographical information systems: the necessity of a geoinformation theory, *ISPRS Journal of Photogrammetry and Remote Sensing* 46:85-103.

- Moore, I.D., and G.J. Burch, 1986. Physical basis of the length-slope factor in the Universal Soil Loss Equation, *Journal of the Soil Science Society of America* 50:1294-1298.
- Moore, I.D., J.C. Panuska, R.B. Grayson and K.P. Srivastava, 1988. Application of digital topographic modeling in hydrology, in American Society of Agricultural Engineers, *Modeling Agricultural, Forest and Rangeland Hydrology, Proceedings of the 1988 International Symposium*, Chicago, IL, ASAE Publication 07-88, ASAE, St. Joseph, MI., pp. 447-461.
- Moore, I.D., R.B. Grayson and A.R. Ladson, 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications, *Hydrological Processes* 5(1):3-30.
- Moore, I.D., A.K. Turner, J.P. Wilson, S.K. Jenson, L.E. Band, 1993. GIS and land surface-subsurface modeling, in M.F. Goodchild, B.O. Parks and L.T. Steyaert, eds, *Environmental Modeling and GIS*, Oxford University Press, Chapter 19.
- NASA, 1991. Earth Systems Science Modeling, Data and Information Systems, Research Announcement, NRA 91-OSSA-07, Office of Space Science and Applications, National Aeronautics and Space Administration, Washington, DC.
- Nemani, R., S.W. Running, L.E. Band and D.L. Peterson, 1993. Regional hydro-ecological simulation system (RHESSys): an illustration of the integration of ecosystem models in a GIS, in M.F. Goodchild, B.O. Parks and L.T. Steyaert, eds, *Environmental Modeling and GIS*, Oxford University Press, Chapter 28.
- Nyerges, T.L., 1991. Geographic information abstractions: conceptual clarity for geographic modeling, *Environment and Planning A* 23:1483-1499.
- Nyerges, T.L., 1992. Coupling GIS and spatial analytic models, *Proceedings of the 5th International Symposium on Spatial Data Handling*, Charleston, SC, International Geographical Union, Commission on GIS, Humanities and Social Sciences Computing Lab, University of South Carolina, Columbia, SC, 29208, pp. 534-543.
- Oliver, M.A., R. Webster and J. Gerrard, 1989. Geostatistics in physical geography. Part I: theory, *Transactions of the Institute of British Geographers* 14:259-269.
- Oliver, M.A., and R. Webster, 1991. How geostatistics can help you, *Soil Use and Management* 7(4):206-217.
- Peuquet, D.J., 1988. Representations of geographic space: toward a conceptual synthesis, *Annals of the Association of American Geographers* 78(3):375-394.
- Philip, G.M., and D.F. Watson, 1986. Matheronian geostatistics - quo vadis?, *Mathematical Geology* 18(1):93-117.
- Pike, R.J., 1988. The geometric signature: quantifying landslide-terrain types from digital elevation models, *Mathematical Geology* 20(5):491-510.
- Raich, J.W., E.B. Rastetter, J.M. Melillo, D.W. Kicklighter, P.A. Steudler, B.J. Peterson, A.L. Grace, B. Moore III, C.J. Vorosmarty, 1991. Potential net primary productivity in South America: Application of a global model, *Ecological Applications* 1(4):399-429.

- Rasmus, D.J., 1988. A STELLA performance, *MACUSER* June 1988, p. 195.
- Rumble, J. R., Jr., 1984. Data, computers and database management systems, in J.R., Rumble, Jr., and V.E. Hampel, eds, *Database Management in Science and Technology: A CODATA Sourcebook on the Use of Computers in Data Activities*, Elsevier Science Publishers B.V., Amsterdam.
- Sargent, R.G., 1980. Verification and validation of simulation models, in F.E. Cellier, ed., *Progress in Modelling and Simulation*, Academic Press, pp. 159-169.
- Saxton, K.E., and J.L. McGuinness, 1982. Evapotranspiration, in C.T. Haan, H.P. Johnson and D.L. Brakensiek, eds, *Hydrologic Modeling of Small Watersheds*, American Society of Agricultural Engineers, Monograph Number 5, pp. 229-273.
- Schut, C.H., 1976. Review of interpolatin methods for digital terrain models, *The Canadian Surveyor* 30(5):389-412.
- Selby, S., and L. Sweet, 1969. *Sets-Relations-Functions*, McGraw Hill.
- Sellers, P.J., F.G. Hall, G. Asrar, D.E. Strebel and R.E. Murphy, 1988. The First ISLSCP Field Experiment (FIFE), *Bulletin of the American Meteorological Society* 69(1):22-27.
- Serra, J., 1982. *Image Analysis and Mathematical Morphology*, Academic Press, London.
- Serra, J., and B. Lay, 1988. *Algorithms in Mathematical Morphology*, Academic Press, London.
- Serra, J., ed., 1988. *Image Analysis and Mathematical Morphology, Volume 2: Theoretical Advances*, Academic Press, London.
- Shercliff, J.A., 1977. *Vector Fields: Vector analysis developed through its application to engineering and physics*, Cambridge University Press.
- Silk, J., 1979. *Statistical Concepts in Geography*, George Allen and Unwin, London.
- Sinton, D.F., 1992. Reflections on 25 years of GIS, supplement to *GIS World* 5(1).
- Skaggs, R.W., 1982. Infiltration, in C.T. Haan, H.P. Johnson and D.L. Brakensiek, *Hydrologic Modeling of Small Watersheds*, American Society of Agricultural Engineers, Monograph 5, pp. 121-166.
- Sklar, F.H., and R. Costanza, 1990. The development of dynamic geographic models for landscape ecology: A review and prognosis, in M.G. Turner and R.H. Gardner, eds, *Quantitative Methods in Landscape Ecology: The Analysis and Interpretation of Landscape Heterogeneity*, Ecological Studies, Vol 82, Springer-Verlag, pp. 239-288.
- Smith, T.R., 1992. Towards a logic-based language for modeling and database support in spatio-temporal domains, *Proceedings of the 5th International Symposium on Spatial Data Handling*, Charleston, SC, International Geographical Union, Commission on GIS, Humanities and Social Sciences Computing Lab, University of South Florida, Columbia, SC, p. 592-601.

- Smith, T.R., and F.P. Bretherton, 1972. Stability and conservation of mass in drainage basin evolution, *Water Resources Research* 8(6):1506-1529.
- Snyder, J.P., 1987. *Map projections: A working manual*, U.S. Geological Survey Professional Paper No. 1395, General Printing Office, Washington, DC.
- Stevens, S.S., 1946. On the theory of scales of measurement, *Science*, 103(2684):677-680.
- Stewart, J., 1987. *Calculus*, Brook/Cole Publishing Company, Pacific Grove, California.
- Strahler, A.N., and A.H. Strahler, 1989. *Elements of Physical Geography*, 4th ed., John Wiley & Sons.
- Stuebe, M.M., and D.J. Johnston, 1990. Runoff volume estimation using GIS techniques, *Water Resources Bulletin* 26(4):611-620.
- Summerfield, M., 1983. Populations, samples and statistical inference in geography, *Professional Geographer* 35:143-149.
- Theobald, D.M., 1989. Accuracy and bias issues in surface representation, in M.F. Goodchild and S. Gopal, eds, *Accuracy of Spatial Databases*, Taylor and Francis, pp. 99-106.
- Thornes, J., 1989. Geomorphology and grass roots models, in B. Macmillan, ed, *Remodelling Geography*, Basil Blackwell, Oxford, pp. 3-21.
- Thorntwaite, C.W., and J.R. Mather, 1955. The water balance, *Publications in Climatology* 8:1, Drexel Institute of Technology, Philadelphia Laboratory of Climatology, Centerton, NJ.
- Tobler, W.R., 1969. Geographical filters and their inverses, *Geographical Analysis* 1(3):234-253.
- Tobler, W.R., 1979a. Transformational view of geography, *American Cartographer* 6(2):101-106.
- Tobler, W.R., 1979b. Smooth pycnophylactic interpolation for geographic regions, *Journal of the American Statistical Association* 74(367):519-536.
- Tobler, W.R., 1988. Resolution, resampling, and all that, in H. Mounsey and R. Tomlinson, eds, *Building Databases for Global Science*, Taylor and Francis, pp. 129-137.
- Tobler, W.R., 1989. Frame independent spatial analysis, in M.F. Goodchild and S. Gopal, eds, *Accuracy of Spatial Databases*, Taylor and Francis, pp. 115-122.
- Tobler, W.R., 1990. GIS transformations, *Proceedings of GIS/LIS '90* 1:163-166.
- Tomlin, C.D., 1983. *Digital Cartographic Modeling Techniques in Environmental Planning*, unpublished PhD dissertation, Yale University.
- Tomlin, C.D., 1990. *Cartographic Information Systems and Cartographic Modeling*, Prentice-Hall, Englewood Cliffs, NJ.

- Tomlin, C.D., 1991. Cartographic modelling, in D.J. Maguire, M.F. Goodchild and D.W. Rhind, eds, *Geographical Information Systems: Principles and Applications*, Longman, London, pp. 361-374.
- Trakowski, A.C., 1976. Conference Goals, in W.R. Ott, ed, *Proceedings of the EPA Conference on Environmental Modeling and Simulation*, April 19-22, 1976, Cincinnati, Ohio, US Environmental Protection Agency, Report EPA 600-9-76-016, Washington DC.
- Turc, L., 1961. Evaluation des Besoins en Eau d'Irrigation, Evapotranspiration Potentielle, *Ann. Agron.* 12(1):13-49, referenced in K.E. Saxton and J.L. McGuinness, Evapotranspiration, in C.T. Haan, H.P. Johnson and D.L. Brakensiek, *Hydrologic Modeling of Small Watersheds*, American Society of Agricultural Engineers, Monograph 5, pp. 229-273.
- Unwin, D., 1981. *Introductory Spatial Analysis*, Methuen.
- Unwin, D., 1989. Three questions about modelling in physical geography, in B. Macmillan, ed, *Remodelling Geography*, Basil Blackwell, Oxford, pp. 53-57.
- USGS, 1991. *U.S. Geological Survey: Earth Science in the Public Service*, U.S. Geological Survey, Government Printing Office, Washington, DC.
- Vansteenkiste, G.C., and J.A. Spriet, 1980. Modelling ill-defined systems, in F.E. Cellier, ed., *Progress in Modelling and Simulation*, Academic Press, pp. 11-38.
- Vieux, B.E., 1991. Geographic information systems and non-point source water quality and quantity modelling, *Hydrological Processes* 5:101-113.
- Vieux, B.E., V.R. Bralts, L.J. Segerlind, 1988. Finite element analysis of hydrologic response areas, in American Society of Agricultural Engineers, *Modeling Agricultural, Forest and Rangeland Hydrology, Proceedings of the 1988 International Symposium*, Chicago, IL, ASAE Publication 07-88, ASAE, St. Joseph, MI., pp. 437-446.
- Vorosmarty, C.J., B. Moore III, A.L. Grace, M.P. Gildea, J.M. Melillo, B.J. Peterson, E.B. Rastetter, P.A. Steudler, 1989. Continental scale models of water balance and fluvial transport: an application to South America, *Global Biogeochemical Cycles* 3(3):241-265.
- Waters, Nigel, 1989. Geographic modelling on microcomputers: Quo vadis?, *The Operational Geographer* 8(1):41-44.
- Webster, R., and M.A. Oliver, 1992. Sample adequately to estimate variograms of soil properties, *Journal of Soil Science* 43(1):177-192.
- Weibel, R., and M. Heller, 1991. Digital terrain modelling, in D.J. Maguire, M.F. Goodchild and D.W. Rhind, eds, *Geographical Information Systems: Principles and Applications*, Longman, London, pp. 269-297.
- Whittow, J., 1984. *The Penguin Dictionary of Physical Geography*, Penguin Books.
- Willgoose, G., R.L. Bras and I. Rodriguez-Iturbe, 1989. Results from a new model of river basin erosion, *Earth Surface Processes and Landforms* 16:237-254.

- Willgoose, G., R.L. Bras and I. Rodriguez-Iturbe, 1990. A model of river basin evolution, *EOS*, November 20, 1990, pp. 1806-1807.
- Wischmeier, W.H., and D.D. Smith, 1978. *Predicting rainfall erosion losses - a guide to conservation planning*, Agricultural Handbook 537, Science and Education Administration, US Department of Agriculture, Washington, DC.
- World Meteorological Organization, 1975. *Intercomparison of Conceptual Models Used in Operational Hydrological Forecasting*, WMO Operational Hydrology Report No. 7, Geneva, Switzerland.
- Yoeli, P., 1984. Cartographic contouring with computer and plotter, *American Cartographer* 11:139-55.
- Zeigler, Bernard P., 1976. *Theory of Modelling and Simulation*, John Wiley & Sons, New York.