Fields as a framework for integrating GIS and environmental process models. Part 1: Representing spatial continuity

IMPORTANT!
See errata at end.

KAREN K KEMP

National Center for Geographic Information and Analysis, Department of Geography, University of California, Santa Barbara, California 93106–4060, USA. e-mail: kemp@ncgia.ucsb.edu

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 paper outlines a strategy in which *field variables* are used to enable modellers to work directly with the spatial data as spatially continuous phenomena. 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 modelling of physical processes. Modellers 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 may result. 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 different spatial data models stored and manipulated by a GIS.

Introduction

The value of GIS in environmental modelling efforts is widely acknowledged. Links between environmental models and GIS are becoming common and interest in merging the technologies is growing (e.g. the proceedings of three international conferences Goodchild et al 1993; Goodchild et al 1996; NCGIA 1996). There is now general agreement that there are several different model/GIS coupling strategies ranging from simple, operator-initiated exchange of files and special interface programs which manage the file format conversions to a level at which 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. As Nyerges noted:

> *The compatibility of data models for the GIS and the analytic model determine how easy or difficult it is to couple the two. [...] 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: 538)

There are a number of different 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 sophisticated interface programs which can handle all of the issues related to the conversion of different file formats (e.g. Breunig and Perkhoff 1992). Others have suggested that the solution lies in defining unified data modelling languages that can manipulate and integrate data from many different formats (e.g. Smith et al 1995). Whatever integration approach is used, the linkage will produce meaningful model results only if the relationship between the real world as it is represented by both the model and the geographic

1361-1682/97/0003-0219

database is understood and accounted for. One of the most difficult aspects of this relationship is the need to force continuous real world phenomena into the discrete environment of the computer.

The problem of continuity

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. 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. Air temperature, soil infiltration rate, solar radiation, and many other natural phenomena are continuous physical fields. A physical field is traditionally defined as an entity which is distributed over space and whose properties are functions of space co-ordinates and, in the case of dynamic fields, of time:

$$z = f(x,y) \text{ or } z = f(x,y,t) \qquad [1]$$

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 (e.g. wind fields where the value at a location has both magnitude and direction). 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. And in turn, 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 such as 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: 11–12). This process results in a continuous density surface. Models using such variables may 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 km^2 versus 10 over each km^2) 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 modelled 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 modelling 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 (e.g. Langran 1992; Newell et al 1992; Kemp and Groom 1994).

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. However, many continuous phenomena, such as soil, vegetation, and geology, are often measured using discrete scales of measurement. Thus we often have a measurement incompatibility within spatially compatible data sets.

Continuity in many dimensions is a fundamental characteristic of all natural systems. It must be scrupulously and explicitly addressed whenever the natural environment is transformed into a digital representation. We continue with a more detailed consideration of continuity in the spatial dimension.

Working with spatial continuity

Working with continuous phenomena has always been a challenge for the sciences. Until calculus, the mathematics of continuity, was developed concurrently by Newton and Leibniz, many problems in classical physics defied solution. Outside physics, the dichotomy between discrete and continuous concepts of the world have challenged researchers in many fields, including mathematics, logic, semantics, geography, and GIS.

Couclelis (1992) provides a broad perspective on the continuous/discrete dichotomy by comparing the object versus field debate in GIS to the opposition between the atom and the plenum ontologies in the philosophy of physics. She notes that there are traditionally two different ways to view the world: one sees the world as a collection of discrete, well-defined objects (the discrete or object view) while the other sees the world as a set of attributes distributed through space (the continuous or field view). In this second view, there is some 'thing' everywhere since a value for any attribute can be determined at any place. She concludes that:

> *the geographic world is best compatible with the field perspective. As the supporters of the plenum ontology in physics have pointed out, quantitative descriptions can only deal with relations between properties, not between things and it is properties, not things, that mathematical fields are about. Also, a field-based framework is much better suited to modelling change, and therefore time, because it is much easier both for our minds and for our formal tools to deal with volatile variables than with volatile objects.*
>
> (Couclelis 1992: 73)

In GIS, much debate has been generated over the value and representativeness of the 'field' (= raster) model versus the 'object' (= 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. Like many of his private sector colleagues, Sinton, at the time chief of systems engineering for a large GIS company, was willing to announce publicly in 1992 that '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: 4). Unfortunately, Sinton, and others who have called the debate off, has 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 and vice versa. At the heart of this debate should be the issue of how well these models represent the reality they are intended to portray.

In the context of digital spatial information, Peuquet was one of the first to examine comprehensively the different concepts used to deal with representations of spatial phenomena in several fields of study (Peuquet 1988). She proposes a dual conceptual model which can be used to represent 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. the 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 (1969), on the other hand, stresses that working with locations and objects requires two different languages. He notes, for example, 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: 216). No matter what the philosophy, in order to do mathematics on continuous phenomena in the computer, we must have a means by which continuity can be discretized.

The field of semantics may provide some direction for resolving this dichotomy. In some of the earlier work on designing computing systems for artificial intelligence, the concept of *mass nouns* caused considerable discussion (Bunt 1985). Mass nouns describe quantities which are not defined as discrete objects (i.e. jewellery, money, time, furniture, water) and are in a sense conceptually continuous. The opposite to mass nouns are *count nouns* which are used to refer to individual elements (i.e. necklace, dollar, hour, table, water drop). Mass nouns cannot be handled through traditional logical formalisms since while set theory does support the concept of a member of a set, it does not support the concept of a part of a whole (e.g. a portion of the whole quantity of water). Thus as Bunt suggests: 'This means that we have to face two problems. One is the design of a representational formalism for continuous notions, the second is the "interfacing" of this formalism with a set-based formalism for representing "discrete" notions' (Bunt 1985: 38). According to Bunt, this 'interface problem' can be resolved in one of several ways:

- embedding a 'continuous' representational formalism within a set-based one;
- a combination of the 'continuous' and the 'discrete' representational formalism;
- having two alternative representations of the same or loosely-related knowledge, and the possibility of relating and transforming the one into the other.

The latter solution is the one most commonly employed in GIS. Bunt concludes that:

> *To satisfy all these interface requirements, there really seems to be only one possibility: we must have one general representational formalism that can accommodate alternative views, allowing both 'continuous' and 'discrete' representations and meaningful mixtures of the two.*
>
> (Bunt 1985: 39)

Thus, given all these arguments, it seems clear that the simple concepts of raster and vector may be incomplete for working with representations of continuous phenomena. There is more to representing reality than just breaking it into pieces that can be fit into the computer.

Levels of abstraction for modelling reality

It is useful to consider the distinct stages of increasing abstraction involved in the building and implementation of a model of a natural process. For mathematical modelling, Bekey (1985) identified three levels in his *hierarchy of representations*: (1) the *reference model*, which is purely conceptual since it is a perfect model exactly equivalent to the process being modelled; (2) the *mathematical model*, which formally expresses the process variables and the relationships between them; and (3) the *computer implementation* of the model. Similarly, in terms of the representation of space, Peuquet (1988) recognizes three similar *levels of abstraction*: (1) the *conceptual representation*; (2) the *functionally-oriented representation*; and (3) 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 necessary to recognize these different kinds of abstractions and to understand the differences and relationships between them. Based on these distinctions, the following set of abstractions for process modelling of spatially-distributed phenomena are used here:

- *Geographic models* (a term proposed by Grelot 1985) are those conceptual models used by various environmental modellers as they evolve an understanding of the phenomenon being studied and extract its salient features from the background of nature's infinite complexity. For example, this might be visualizing terrain as a continuous surface which can be measured everywhere or regarding soils as highly variable continuous phenomena with specific, measurable physical characteristics. 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 at 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 examined within the computer.
- *Data structures* describe details of specific implementations of spatial data models.

The link between specific geographic models of fields and various spatial data models which can be used to represent them is of vital importance to environmental modellers. We now turn to a consideration of spatial data models for fields.

Representing fields

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 modelling, 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:

$$z = ax^2 + bxy + cy^2 + d \qquad [2]$$

where x,y are horizontal Cartesian co-ordinates 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 set of rules for interpolating unknown values between these locations.

The first of these steps is known as *discretization*. The second step is accomplished through the use of *spatial data models*.

Discretization of space

As discussed earlier, the discretization of space is an essential step in numerical methods used to solve complex equations in the computer. Finite elements and the cells used in finite difference solutions require boundaries to be drawn on the continuity of the phenomena being modelled. This drawing of boundaries is not a simple task. As Couclelis (1992) points out, the continuum of the natural world may contain things like hills, valleys, and clouds:

> *but these are not 'objects' to pick up and move about: they are salient features, breaks in a plenum that is otherwise continuous, not by the mathematical criterion of infinite sub-divisibility, but because of the indefinite number of different ways one could draw boundaries around these features.*
>
> (Couclelis 1992: 72–3)

In the geographic sphere, and subsequently in the spatial analysis tradition which influenced the original development of GIS, it is clearly 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 in the human and natural landscapes and the related mathematical techniques that could be used to explain spatial distributions. As the original tool for understanding these spatial structures, maps have provided a means for expressing them. From maps, the line work and shading created by cartographers can be reduced to the basic spatial primitives of point, line, and area. Tobler (1990) calls this the cartographic paradigm of geographic phenomena. Within the map model of geographic reality, continuous phenomena are given structure in the point/line/area model. These concepts certainly do provide powerful analytical tools. The landmark work by Getis and Boots (1978), *Models of Spatial Processes: An Approach to the Study of Point, Line, and Area Patterns*, provides an excellent early summary of the value of this approach, while much of the more recent work on error modelling (e.g. Goodchild and Gopal 1989; Caspary and Scheuring 1992; Monckton 1994) and spatial statistics and analysis (e.g. Arbia 1989; Haining 1990; Fotheringham and Rogerson 1994) also demonstrates the usefulness of such spatial structure.

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: 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 the current jargon in GIS.

Spatial data models

There are two general classes of data models used to represent fields (Goodchild 1992). One class explicitly assigns a value everywhere in space by using a simple function to estimate the character of the spatial variation. Fields represented by functions are modelled by the use of constant or linearly-varying piecewise functions. *Polygons* partition the entire area into irregularly-shaped constant-valued contiguous regions while *cellgrids* partition the area into rectangular constant valued regions. Values do not change within each region but may change abruptly at region boundaries. *Triangulated Irregular Networks* (TINs) also completely partition the area but the regions are triangular and within each region values are assumed to vary linearly between the known values at the nodes. These piecewise models produce values that are only representative of the value of the true surface, there is no assumption of exact measurement for any specific location (with the general exception of TIN nodes).

The other group of field data models samples the field precisely at selected locations and provides no information about the values between these locations. Values between sampled locations must be interpolated. Sampling the surface at locations determined by the nodes of a rectangular grid produces a *pointgrid*. In this case, no knowledge of the phenomenon is assumed in the design of the sampling procedure, although the grid spacing may (should!) be determined by the rate of variability of the phenomenon. For data models composed of *irregular points*, the selection of locations to sample may be based upon rules that adapt to prior knowledge of the field (e.g. denser sampling in areas of more rugged terrain) or that are independent and often irrelevant to the phenomenon (e.g. weather data collected at airports). In *contour models*, sampling locations are controlled by the value of the phenomenon, locations are recorded wherever the surface has one of the selected contour values. The result is a set of non-intersecting lines of constant value. For any of these sampled field data models, assumptions that may be made about the form of the surface between sampled points and techniques to interpolate specific values vary depending on the specific data model and the method used to select sample points.

It is also important to note that similarly-structured data models can be used to represent phenomena which are discontinuous in two dimensions. Cities and well sites can be represented as irregular points while road and river networks and contours all consist of linear features (although intersection is not allowed for contours). Therefore, to conceptually distinguish data models used for discrete objects from those used to represent continuous phenomena, we must recognize a *field* sub-class of spatial data models. Any of the six *field data models* described can be used to represent any field. The specific model chosen will depend on the sampling scheme used, the functionality of the software available to the data gatherer and modeller, and, possibly, the proposed analysis.

A strategy for modelling with data about fields

For environmental modellers, designing and coding a mathematical model is an entirely different task from accessing and manipulating spatial data in a GIS. On the one hand, modellers 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 modellers 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 the way in which spatial data are used in environmental models. Thus, while modellers 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. Wilson (1996) provides a good discussion of this problem in the context of land-surface and sub-surface models.

Common strategies and techniques for handling spatial data about continuous phenomena in all its forms are needed. A common strategy for handling data about fields in mathematical models would provide 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 (1990) original 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 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 modelling 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 paper establishes the fundamentals of this proposed strategy for handling spatially continuous data in environmental modelling projects.

The field data type and field variables

In order to manipulate data about spatially continuous phenomena, we begin by specifying 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. Within the computer, field variables are stored as a set of values corresponding to each *spatial element*, the basic geometric components of spatial data models (e.g. point, cell (pixel), line). Spatial elements 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.

Like other types of variables, fields are represented with symbols. 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. By considering the operations performed on field variables, we can determine the critical properties which must be specified if we are to remove the modeller from unnecessary involvement with the discretization.

Since the manner in which fields are represented within the computer is fundamental in determining how mathematical operations can be performed, field variables must have associated properties describing the data model used and other critical characteristics related to resolution, temporality, and sampling scheme, each of which is defined as follows:

Resolution. The concept of resolution is used in place of the oft-misused term scale. While scale is properly used and defined in cartography where it describes the relationship between distances on the map and distances on the earth, the difficulty of measuring the scale of a TIN or of an irregularly distributed set of points illustrates how the concept is unworkable for digital data. Resolution is defined here as the minimum distance over which variation is recorded. It describes the density of information contained in the field variable and provides critical parameters necessary for its manipulation.

Temporality. While time, like space, is continuous, given the current state of the technology, temporal characteristics are limited to one of three temporal data models:

- a single static view of the phenomenon, in which case temporal changes are not considered;
- a time series of periodic averages – analogous to piecewise spatial models;
- equally spaced instantaneous values – analogous to sampled spatial models.

Sampling scheme. The sampling scheme is particularly important for understanding how irregular point data models represent particular fields. While all irregular point data models will be structured similarly, if we wish to determine the value at a location between points in the data set, we need to know if the point values are representative of their neighbourhoods or if they are the extreme values of the surface. Given this information, the selection of an interpolation method based on, for example, Thiessen polygons in the first case and triangulation in the second might be made automatically.

Other properties. The concepts of *spatial equality and spatial nesting* are used to compare the specific spatial discretizations of different representations of fields. 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, they have the same cell dimensions, origin, orientation, and projection. Spatially equivalent polygons are unlikely to occur in environmental data sets since boundary locations are generally determined by the phenomenon 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. For reasons described below, 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 (Figure 1). For piecewise spatial models, if A spatially nests within B:

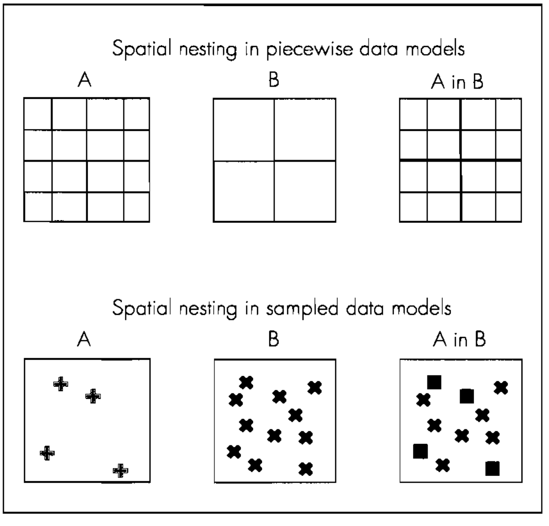
- each element in A falls completely within one element in B; and
- the set of lines which form the boundaries of B is a sub-set of the set of lines which form the boundaries of A.

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

Mathematics with field variables

Since digital computers are discrete machines, they are 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

Figure 1. Spatial nesting in piecewise and sampled data models.

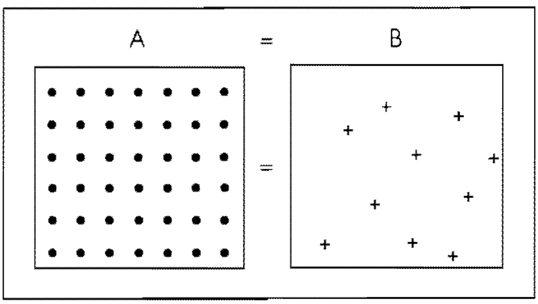


other field variable. This condition can be expressed most directly in the '=' or assignment operation of traditional algebra.

Assignment 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 spatial data model different 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 (Figure 2). 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 this author's contention that it is possible to codify these differences in such a way

Figure 2. Determining values in one spatial data model from another. It is not possible directly to equate individual values in one spatial data model to those in another. For example, how can values at locations in this grid of points (A) be directly determined from this set of irregular points (B)?



that the decisions regarding how to convert one model to another can be handled automatically, without input from the modeller but without compromising the relationship between the data model and reality. In the next section we discuss issues which determine how these conversions should be accomplished and lay out a scheme for organizing and selecting appropriate procedures.

From data models through reality and back

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. Earlier, 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 (1988) suggests that it also includes computation using any other spatial data model used to represent continuous phenomena. Here spatial interpolation is defined as *a set of rules for obtaining a complete field from a spatial data model.* 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 Lam (1983), Burrough (1986) and Myers (1994).

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.

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 – 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 and the raster-based field model.

Converting spatial data models

Based on a detailed review of the procedures used to interpolate and sample spatial data models (Kemp 1993), it is possible to summarize the procedures to convert spatial data models in a pair of *conversion matrices*, shown and discussed below as Tables 1 and 2. While organized this way 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. The reader is directed to the original publication for complete details of the conversion procedures generalized here.

Table 1. Summary of spatial data model conversions for numerical data.

To / From	Cellgrid	Polygon	Pointgrid	Irregular points	TIN	Contour
Cellgrid	areal weighting	areal weighting	point sample	point sample	point interpolate & triangulate	point interpolate & contour
Polygon	areal weighting	areal weighting	point sample	point sample	point interpolate & triangulate	point interpolate & contour
TIN	areal weighting	areal weighting	point interpolate	point interpolate	add or remove nodes & triangulate	thread contours
Contour	point interpolate & average	point interpolate & average	point interpolate	point interpolate	select nodes & triangulate	remove or add contours
Pointgrid	point interpolate & average	point interpolate & average	point interpolate	point interpolate	select nodes & triangulate	contour
Irregular points	point interpolate & average	point interpolate & average	point interpolate	point interpolate	select nodes & triangulate	contour

Table 1 lists conversion procedures for numerical data. Thus, for example, if source data on rainfall amounts are stored in a Thiessen polygon data model and we wish to use these values in a differential equation which has been solved using finite elements (i.e. a cellgrid model), it is necessary to use an area-weighed procedure to rearrange the region boundaries from the original representation of a continuous surface to produce a new representation of one. Likewise, if the rainfall were stored as a set of irregular points, to convert to the cellgrid model it is necessary first to perform an interpolation routine which estimates the rainfall amounts at points between those for which data is stored in the source model. Assuming that the locations of the interpolated points have been appropriately chosen for the target model (i.e. in a density greater than one per cell), the interpolated values falling within each individual cell can then be averaged to produce a representative value for each cell.

Note that TINs and contours cannot be used as predefined target models since the structure is determined by the phenomenon represented. Thus, it would be inappropriate to convert a cellgrid of soil moisture to a TIN whose structure is determined by elevation as the critical points in the elevation TIN which determine where the facets should lie would not match those of the soil moisture field. However, the conversions for TINs and contour models listed in Table 1 may be performed to reduce data volume or for visualization purposes. They are included here for completeness but they have been separated from the main table columns to indicate that they will not be used for mathematical manipulations.

Table 2 outlines the conversion procedures for categorical data. This type of data presents many limitations when used in mathematical applications. During conversion procedures, resulting data values are limited to the small, finite set of discrete classes in the source data, no new values can be created in a conversion operation. Thus, if a source data set has four classes of vegetation types, the target data set will also have at most four vegetation types. Mathematical transformations of source values such as those produced by interpolation procedures cannot be performed. Thus conversion of a vegetation cellgrid to a vegetation polygon data model requires first the repositioning of the region boundaries and then the use of rules to determine which vegetation class will be assigned to the new region. Such rules might be:

- the source class covering the largest portion of the target region becomes the new value; or
- an order of precedence which determines which class is assigned first if it occurs anywhere within a target region.

The blocky nature of the constant piecewise models is the only possible representation of fields measured on categorical scales since the data are discrete and cannot be continuous across boundaries. Thus, conversions involving categorical data about continuous phenomena are restricted to a sub-set of four spatial data models since categorical data cannot be stored in the surface models – TINs and contour models. Point models may be used to store categorical data, though any conversion of this data requires the initial construction of piecewise models as a representation of the continuous surface. Therefore, conversion from a data model of irregular points representing soil types found at sample sites to a cellgrid requires first that the soil types be spread outward towards other points so that Thiessen polygons are formed, this being the only logical means to create the continuous representation of reality. Conversion would then proceed by either partitioning the space into a new piecewise model and applying rules such as those described above, or simply by observing the polygon values of points at new locations for target point models.

Clearly these tables are only an initial formulation of the conversion tables. Within each table cell there are many different possible algorithms. Which specific algorithm is chosen will depend upon the selection available to the modeller through the GIS or supplementary programs and upon the characteristics of the specific field variable. Fortunately, much of this decision work can be performed by an intelligent interface capable of comparing algorithm capabilities to the resolution and other properties specified in the declaration of the field variable. However, since the reality represented by the six field data models cannot be completely expressed, the specification of field variables should permit the definition of alternate procedures which override defaults provided in the program itself.

Arithmetic with field variables

Unary arithmetic operators operate on a single value to create a derivative value. These operators include negation, absolute value, log, roots, and

Table 2. Summary of spatial data model conversions for categorical data.

To / From	Cellgrid	Polygon	Pointgrid	Irregular points
Cellgrid	partition & apply rules	partition & apply rules	point sample	point sample
Polygon	partition & apply rules	partition & apply rules	point sample	point sample
Pointgrid	Thiessen, partition & apply rules	Thiessen, partition & apply rules	Thiessen & point sample	Thiessen & point sample
Irregular points	Thiessen, partition & apply rules	Thiessen, partition & apply rules	Thiessen & point sample	Thiessen & point sample

exponentiation. Exponentiation may also be seen as a special case of multiplication operating on a single value. As these operators use only one variable, their application to fields represented by digital data models is straightforward.

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, of course, cannot be performed. If both variables are fields, these arithmetic operations can be visualized as combining the values of the variables for each location in space such that:

```
WATER_TABLE_ELEV =
   SURFACE_ELEV_WATER_DEPTH                [3]
```

If A and C are spatially equivalent field variables and b is a scalar, the equations:

```
C = A + b                                  [4a]
C = A - b                                  [4b]
C = A / b                                  [4c]
C = A * b                                  [4d]
```

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, A, B, and C, are field variables, the equations are:

```
C = A + B                                  [5a]
C = A - B                                  [5b]
C = A / B                                  [5c]
C = A * B                                  [5d]
```

If all three field variables are 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. 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. 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. The target model is a cellgrid. 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 target model before the multiplication? Clearly a set of priority rules is needed.

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 3 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. For example, in the top right of the figure, the icons illustrate the combination of a TIN with a medium dense cellgrid to produce a coarse grid. This might be the case where elevation (TIN) is being combined with rainfall (denser cellgrid). The result might be, say, an elevation–rainfall index in a coarse cellgrid to be used later in a global modelling exercise. For this first stage, it would be best to do the calculation using the denser of the cellgrids as the computational model. In the figure, where there are two checks shown for one operation, either of the checked data models might be used as the computational model.

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

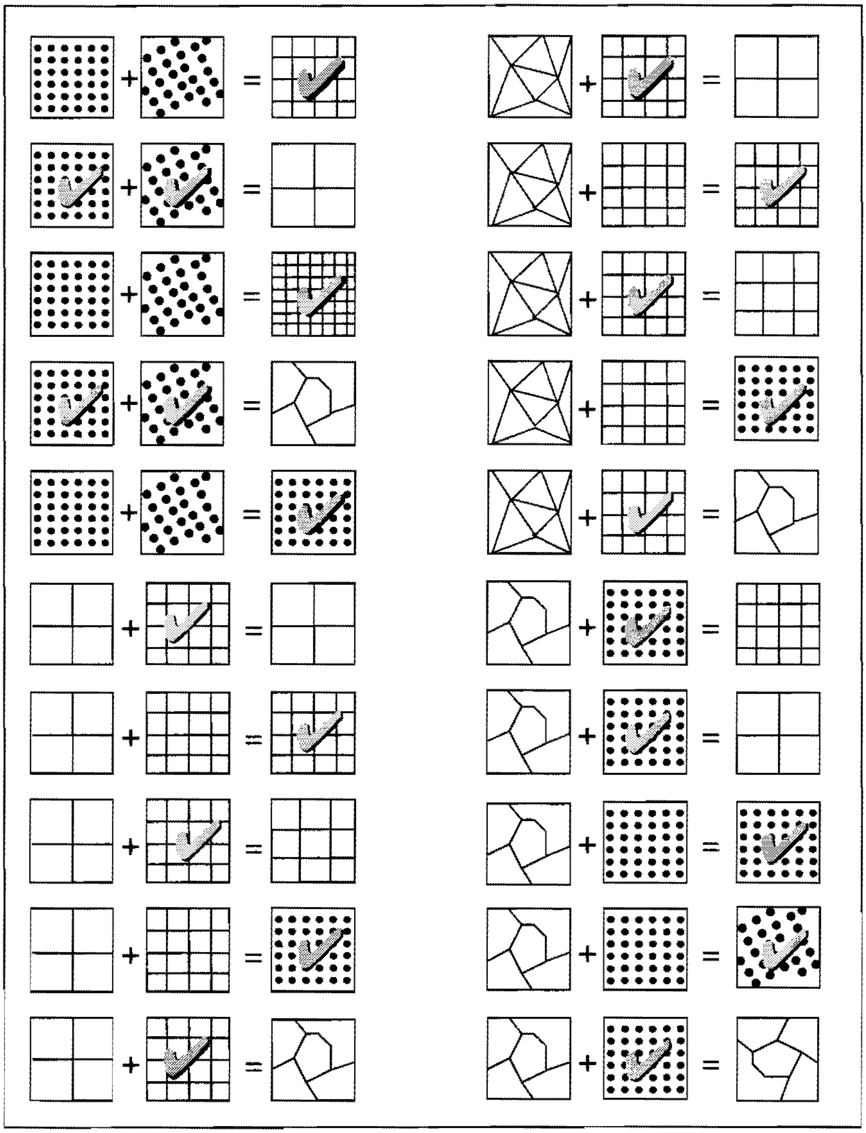


Figure 3. Combining spatial data models. This sketch illustrates some of the many different combinations of spatial data models which may be required in a mathematical operation. Icons represent different spatial data models and structures. Icons marked with checks indicate the preferred computational models for each operation. Where two checks are indicated for a single operation, either model might be used.

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 priority rules, listed in decreasing order of priority:

1 If both sources are spatially equivalent, use that model;
2 If each source is either a TIN or a contour model, use the target;
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 should be clear that implementation of any such set of rules can be accomplished without input from the modeller.

Once field variables have been specified and conversion procedures established, it is possible to implement easily all traditional mathematical computer operations and to consider several new functions specific to continuous data. Perhaps the most readily apparent of such functions is integration. Integration is clearly defined for continuous functions and therefore is easy to conceive of for continuous phenomena. Implementation of an integration function for discrete representations of fields is relatively straightforward once the properties of the field variable are specified. Similarly, differentiation (i.e. calculation of slope from elevation data) and smoothing can be conceptualized for fields and implemented for specific field data models. Part 2 of this paper explores the implementation of field variables in detail.

Conclusions

The value of GIS for environmental modelling projects is clear. For most current environmental modelling 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 important for developing derivative data sets that serve as proxies for unavailable variables. As better spatial analysis methods become incorporated into GIS and programming languages, GIS will also become an important tool in 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. 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. It is hoped that implementation of field variable types as introduced here will lead to full integration of GIS and spatial analysis with environmental modelling.

Acknowledgements

This work was completed at the National Center for Geographic Information and Analysis which is funded by a grant from the US National Science Foundation (Grant No. SBR 88–10917). The author wishes to acknowledge the considerable advice and direction given during the original development of this paper from Professor Michael Goodchild and also the continued encouragement from the Editor of this journal to finally publish formally material which languished as grey literature for some years.

References

Angel S and Hyman G M 1976 *Urban Fields: A Geometry of Movement for Regional Science.* London, Pion

Arbia G 1989 *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems.* Dordecht, Kluwer Academic Publishers

Bekey G A 1985 Simulability of dynamical systems. *Transactions of the Society for Computer Simulation,* 2: 57–71

Bourke W 1988 Spectral methods in global climate and weather prediction. In Schlesinger M E (ed) *Physically-based Modelling and Simulation of Climate and Climatic Change.* Dordecht, Kluwer Academic Publishers: 169–220

Breunig M and Perkhoff A 1992 Data and system integration for geoscientific data. In *Proceedings Fifth International Symposium on Spatial Data Handling, 3–7 August 1992.* Columbia, Humanities and Social Sciences Computing Lab, University of South Carolina: 272–81

Bunt H 1985 The formal representation of (quasi-) continuous concepts. In Hobbs J R and Moore R C (eds) *Formal Theories of the Commonsense World.* Norwood, Ablex Publishing Corporation: 37–70

Burrough P A 1986 *Principles of Geographical Information Systems for Land Resources Assessment.* New York, Oxford University Press

Caspary W and Scheuring R 1992 Error-bands as a measure of geometrical accuracy. In *Proceedings EGIS '92, 23–6 March 1992.* Munich, EGIS Foundation: 226–33

Couclelis H 1992 People manipulate objects (but cultivate fields): Beyond the raster-vector debate in GIS. In Frank A U, Campari I, and Formentini U (eds) *Theories and Methods of Spatio-temporal Reasoning in Geographic Space.* Berlin, Springer-Verlag: 65–77

Fotheringham A S and Rogerson P 1994 *Spatial Analysis and GIS.* London, Taylor & Francis

Getis A and Boots B 1978 *Models of Spatial Processes: An Approach to the Study of Point, Line, and Area Patterns.* Cambridge, Cambridge University Press

Golledge R 1982 Fundamental conflicts and the search for geographical knowledge. In Gould P and Olsson G (eds) *A Search for Common Ground.* London, Pion: 11–23

Goodchild M F 1988 Stepping over the line: Technological constraints and the new cartography. *American Cartographer,* 15: 311–20

Goodchild M F 1992 Geographical data modeling. *Computers and Geosciences,* 18: 401–08

Goodchild M F and Gopal S 1989 *Accuracy of Spatial Databases.* London, Taylor & Francis

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

Goodchild M F, Parks B O, Steyaert L T, Johnston C, Maidment D, Crane M, and Glendinning S (eds) 1996 *GIS and Environmental Modeling: Progress and Research Issues.* Fort Collins, GIS World, Inc.

Grelot J-P 1985 Archaic data models or hardware as a concept killer. In Blakemore M (ed) *Proceedings AutoCarto London: International Conference on the Acquisition, Management and Presentation of Spatial Data, London, 14–19 September 1986.* AutoCarto London Ltd: 572–7

Haining R 1990 *Spatial Data Analysis in the Social and Environmental Sciences.* Cambridge, Cambridge University Press

Harvey D 1969 *Explanation in Geography.* London, Edward Arnold

Kemp K K 1993 *Environmental Modeling with GIS: A Strategy for Dealing with Spatial Continuity.* Santa Barbara, National Center for Geographic Information and Analysis, University of California, Technical Report 93–3

Kemp Z and Groom J 1994 Incorporating generic temporal capabilities in a Geographical Information System. In *Proceedings, 29 March – 1 April 1994, EGIS '94.* Utrecht, EGIS Foundation: 86–95

Lam N S-N 1983 Spatial interpolation methods: A review. *American Cartographer,* 10: 129–49

Langran G 1992 *Time in Geographic Information Systems.* London, Taylor & Francis

Monckton C 1994 An investigation into the spatial structure of error in digital elevation data. In Worboys M F (ed) *Innovations in GIS.* London, Taylor & Francis: 201–14

Myers D E 1994 Spatial interpolation: An overview. *Geoderma,* 62: 17–28

NCGIA (ed) 1996 *Proceedings Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, 21–6 January 1996.* Santa Barbara, National Center for Geographic Information and Analysis, University of California. WWW and CD-ROM

Newell R G, Theriault D, and Easterfield M 1992 Temporal GIS – Modeling the evolution of spatial data in time. *Computers and Geosciences,* 18: 427–33

Nyerges T L 1992 Coupling GIS and spatial analytic models. In *Proceedings Fifth International Symposium on Spatial Data Handling, 3–7 August 1992.* Columbia, Humanities and Social Sciences Computing Lab, University of South Carolina: 534–43

Peuquet D J 1988 Representations of geographic space: Toward a conceptual synthesis. *Annals of the Association of American Geographers,* 78: 375–94

Shercliff J A 1977 *Vector Fields: Vector Analysis Developed through its Application to Engineering and Physics.* Cambridge, Cambridge University Press

Sinton D F 1992 Reflections on 25 years of GIS. *GIS World,* 5: supplement

Smith T R, Su J, El Abbadi A, Agrawal D, Alonso G, and Saran A 1995 Computation modeling systems. *Information Systems,* 20: 127–53

Tobler W R 1988 Resolution, resampling, and all that. In Mounsey H and Tomlinson R (eds) *Building Databases for Global Science.* London, Taylor & Francis: 129–37

Tobler W R 1990 GIS transformations. In *Proceedings GIS/LIS '90*. Anaheim, American Congress of Surveying and Mapping: 163–6

Tomlin C D 1990 *Cartographic Information Systems and Cartographic Modeling*. Englewood Cliffs, Prentice Hall

Wilson J P 1996 GIS-based land-surface/sub-surface modeling: Potential for new models? In *Proceedings Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, 21–6 January 1996*. Santa Barbara, National Center for Geographic Information and Analysis, University of California. WWW and CD-ROM

Errata

'Economic, legal, and public policy issues influencing the creation, accessibility, and use of GIS databases' by D W Rhind. TG 1(1): 3–12

The original printing of this paper contained one typographical error in Figure 3 so that the tones in the key did not match the tones in the diagram. This figure should read as follows:

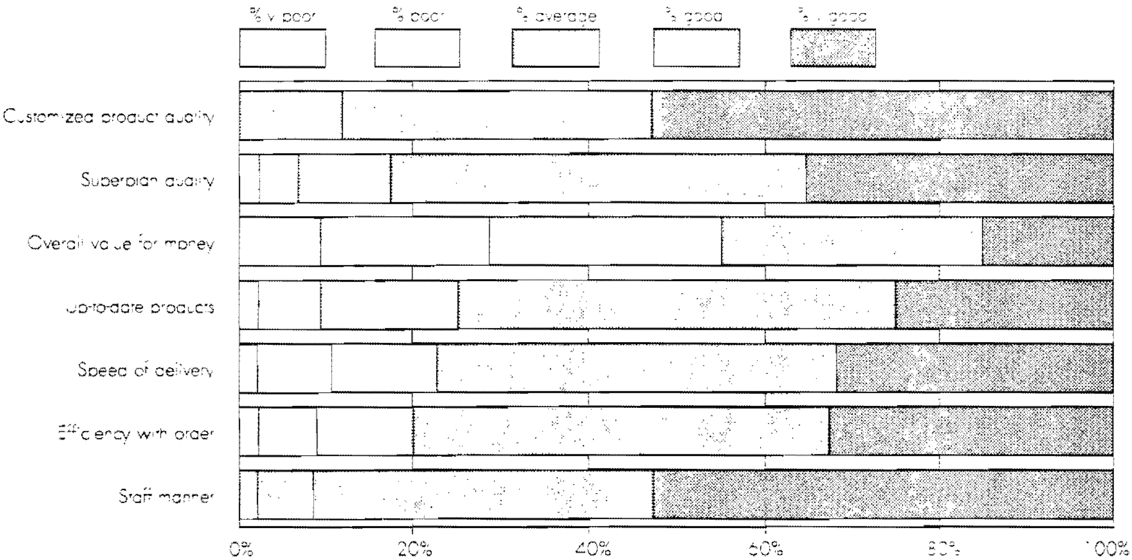


Figure 3. Results of survey of OS customer satisfaction in 1994.

'Fields as a framework for integrating GIS and environmental process models. Part 1: Representing spatial continuity' by K K Kemp. TG 1(3): 219–34

The original printing of this paper contained one typographical error in equation (3). This equation should read as follows:

```
WATER_TABLE-ELEV =
   SURFACE-ELEV - WATER_DEPTH
```

'Fields as a framework for integrating GIS and environmental process models. Part 2: Specifying field variables' by K K Kemp. TG 1(3): 235–46

The original printing of this paper contained a series of typographical errors in the equations reproduced on pp. 237–40 and p. 244. These equations (with page numbers and positions in parentheses) should read as follows:

(10 lines up from bottom of right-hand column of p. 237)

```
model(pointgrid: c1, c2, a, Δx, Δy, nx, ny);
```

(14 lines up from bottom of right-hand column of p. 238)

```
interpolation(name);
```

(18 and 23 lines down from top of right-hand column of p. 239 respectively)

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

(4, 21, and 24 lines down from top of left-hand column of p. 240 respectively)

```
define time name:(parameters);
define model COMP:
  (spatial_model: parameters);
define model COMP(NAME):
  (spatial_model: parameters);
```

(20–1 and 12–18 lines up from bottom of right-hand column of p. 240 respectively)

```
define(VARIABLE_1 SYMBOL
  VARIABLE_2):{algorithm};
define(A = B):{algorithm};
define(pointgrid =
  irregular_points):
  {algorithm};
define(VEG - SOILS):
  {overlay(VEG, SOILS)
  lookup(potential)};
```

(the example that was originally printed in the left-hand column of p. 244)

```
define model COMP:(irregular
  points:10^4 km^2);
define time months:
  1 month, yy 1, 12;
field GROWTH: model(COMP),
  measurement(numeric),
  time(months);
```

defines the computational structure, its parameters are input as a null data set

```
field RAIN:model
  (irregular points: 10^4 km^2),
  measurement(numeric),
  time(months);
field CLOUD_COVER:model(cellgrid:
  24°N, 128°W, 0°, 1km, 1km, 100, 100),
  measurement(numeric),
  time(months);
field FERTILITY:
  model(polygon): 10^4 km^2),
  measurement(numeric);
field ELEV:model(pointgrid:
  24°N, 128°W, 0°, .5km, .5km,
  200, 200),
  measurement(numeric);
```

these statements may be within the mathematical model's code or the information may be included as metadata with each data set

```
define find_growth:
  {equation(s) for calculating growth
  from several parameters};
read GROWTH,SITES,RAIN,
  CLOUD_COVER,FERTILITY,ELEV;
GROWTH = find_growth(RAIN,
  CLOUD_COVER,FERTILITY,ELEV,
  aspect(ELEV));
```

conversion of all spatial models to the computational model is enforced before calculation