A Geographer Looks at Spatial Information Theory

Michael F. Goodchild[1]

[1] National Center for Geographic Information and Analysis, and Department of Geography, University of California, Santa Barbara, California 93106-4060, USA
good@geog.ucsb.edu

Abstract. Geographic information is defined as a subset of spatial information, specific to the spatiotemporal frame of the Earth's surface. Thus geographic information theory inherits the results of spatial information theory, but adds results that reflect the specific properties of geographic information. I describe six general properties of geographic information, and show that in some cases specialization has assumed other properties that are less generally observed. A recognition of the distinction between geographic and spatial would allow geographic information theory to achieve greater depth and utility.

1 Introduction

The term *geographic* might be said to refer to features and phenomena at or near the surface of the Earth, and if so, geographic information is information about such features and phenomena. More formally, geographic information might be defined as consisting of atomic pairs of the form <**x**,**z**> where **x** is a location in space-time, and **z** is a set of properties of that location [10]; or of information that is reducible to such pairs. Thus *geographic* refers to a spatial domain consisting of the Earth's surface and near-surface, and times extending forwards and backwards from the present. The term also implies a certain range of spatial resolution, from perhaps 1cm to 10km, that excludes any quantum or relativistic effects and is thus rigidly Newtonian.

In this sense geographic is a subset or specialization of *spatial*, which by extension refers to any spatiotemporal frame, and any spatial resolution, and also includes non-Cartesian spaces. The spaces defined by the human body, or an automobile, or the universe are instances of *spatial*. A spatial frame may contain the geographic frame, as in the case of the universe, but the geographic frame may also contain spatial frames that may move within it. Thus a human sees the geographic frame as a rigid and fixed structure, and other spaces as variously embedded within it. From this perspective the term *geospatial* is essentially identical to and redundant with *geographic*.

While *geographic* inherits many of its properties from *spatial*, it also adds new ones, and thus specializes the definition. If "spatial is special", as many have suggested [1], [13], then geographic should be even more special, and a theory of geographic information should be distinct from a theory of spatial information, inheriting all of the generality of the latter, but adding its own specifics. Thus when a geographer looks at spatial information theory, he or she logically asks not whether

D.R. Montello (Ed.): COSIT 2001, LNCS 2205, pp. 1−13, 2001.

the conclusions of spatial information theory are useful, as they must necessarily be in any subclass, but whether the conclusions could be more useful if the specifics of the subclass were exposed.

In this paper I examine the specific nature of geographic information, by discussing six principles that appear to be generally true of geographic information but not necessarily true of spatial information. By doing so I hope to demonstrate that while its generalities are undoubtedly useful, a theory of spatial information can be made even more useful and effective for geographic information if it recognizes and exploits those specifics. The specific nature of geographic information imposes constraints, narrowing the options that must be considered in the general case. It also suggests underlying structures and causal mechanisms that may further narrow the options, and allow theory development to proceed to deeper levels.

2 General Properties of Geographic Information

2.1 Positions in the Geographic Frame Are Uncertain

First, consider the determination of position on the Earth's surface. All measuring instruments are subject to error, and many of the instruments used to measure position are subject to substantial errors. For example, routine measurements using the Global Positioning System are subject to errors on the order of 10m. Monuments in supposedly fixed positions move as a result of tectonic activity and the movement of crustal plates. More fundamentally, the frame as defined by the Earth's axis moves as the Earth wobbles, and along with it the Poles and Equator; and the ellipsoids and other mathematical functions used to approximate the shape of the Earth are defined only to limited precision. For all of these reasons, it is impossible to measure location on the Earth's surface exactly, or to determine equality of position based purely on information about location.

As a consequence the geometry underlying all geographic information technologies is approximate. Moreover, errors are normally much greater than the uncertainty inherent in using discrete numerical methods in computing systems, although these sometimes contribute significantly. For example, single-precision arithmetic normally offers 7 decimal digits; but 1 part in 10^7 of the Earth's radius is less than a meter, and thus substantially more precise than the accuracy of most global databases. Double precision offers 14 decimal digits, which supports accurate positioning on the Earth's surface at sub-micron levels, an absurd level of precision given the typical accuracy of geographic data. In practice, positional accuracy seems to fall within a fairly narrow range of 10^{-3} to 10^{-4} of the extent of a project for a variety of reasons [8]. Thus only when coordinates are represented by short integers is there a need to be concerned about machine precision.

The limited precision of positional representations has motivated a number of projects, such as ROSE [11], that have developed algorithms that are consistent with a discrete rather than continuous space. In effect, these algorithms assume that numerical methods are implemented in discrete form in a space that is fundamentally continuous, and that position is knowable to an accuracy that is greater than the precision of the methods. But inaccuracy requires a somewhat different approach, because each point's position must be conceptualized as located at the center of a circle of possibility in continuous space—one might term this an *object-centered* approach, to distinguish it from the *space-centered* approach of a discrete-space geometry (Figure 1). Because inaccuracies are likely to be many orders of magnitude greater than imprecisions, the object-centered approach seems to be much more strongly motivated for geographic information than the space-centered approach, but to have received much less attention.

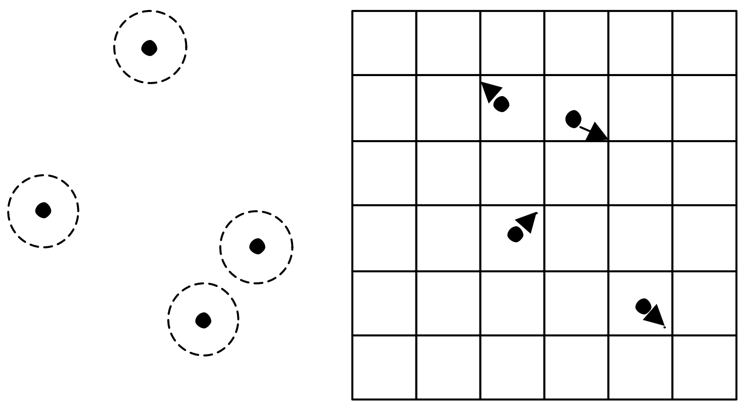

Fig. 1. (A) In an object-centered approach, limited accuracy requires that the possible true locations of a point are located within a given distance of the point's apparent position. (B) In a space-centered approach, limited precision requires that points appear to be at the intersections of a fixed grid.

Thus while it is interesting to theorize about spatial information in ways that include the possibility of equality of position, in practice for geographic information it is almost never possible to determine equality. We cannot determine whether a point lies exactly on a line, or whether two lines are exactly equal, based on position alone. Thus point-in-polygon routines designed to determine enclosure normally offer only a binary response (in or out), and polygon overlay routines infer equality of position using user-defined positional tolerances, not by exact comparison. It is generally unwise to compute topology from geometry, and better to allow independently determined topology to over-ride geometry when the two conflict, as they often will. Because the distance between a house and its street centerline is often less than the accuracy of positioning of either, many databases code the house's side of the street directly (*e.g.*, TIGER and its derivatives). It is generally unsafe to rely on point-in-polygon operations to determine the parcel containing a point, such as a utility pole, and better to code the containment directly, and to allow this topological information to over-ride any information obtained from geometry.

In summary, a theory of geographic information can often afford to drop the equality option, because it implies an unrealistic level of accuracy in positioning.

Moreover, the nature of inaccuracy suggests that an object-centered approach to imprecise geometry will be more productive than a space-centered approach.

2.2 Spatial Dependence Is Endemic in Geographic Information

A variable is said to possess spatial dependence if correlations exist between its values at distinct points. Frequently the degree of similarity between the values at two points increases as the two points approach each other, implying a degree of continuity and smoothness. Geographic information is observed to possess this type of spatial dependence, and this observation is sufficiently general to warrant the status of a law, often identified with Waldo Tobler [18] and stated thus: "All things are similar, but nearby things are more similar than distant things." The effect expressed in the law is easily measured by the Geary and Moran statistics of spatial autocorrelation, and by the *variogram*, and the field of geostatistics is founded on what Matheron [15] termed *regionalized* variables, or variables possessing strong spatial dependence in accordance with Tobler's law.

It is possible to distinguish between positive and negative spatial autocorrelation; in the positive case nearby pairs of points are more similar than distant pairs, while in the negative case nearby pairs are more different than distant pairs. But such measures are scale-specific, and it is generally impossible for a variable to be negatively autocorrelated at all scales. Thus the familiar chessboard shows strong negative autocorrelation between adjacent squares, but strong positive autocorrelation within squares.

Zero spatial autocorrelation results when values at distinct points are uncorrelated, or statistically independent. This is a reasonable condition when the points are very far apart, or separated by what geostatistics terms the *range* of the variable. But consider a world in which spatial autocorrelation is zero at *all* scales. In such a world, an infinitesimal movement would be sufficient to encounter the entire range of the variable, and it would be impossible to construct descriptions or representations of the world that were less than infinitely large. In effect, spatial dependence is essential for description, mapping, and the very existence of geographic information as a useful commodity. A world without spatial dependence would be an impossible world to describe or inhabit.

Many statistical methods assume independence of observations, and thus are problematic when applied to geographic information. Inferential tests associated with the Geary and Moran coefficients [5] invoke a null hypothesis of zero spatial dependence, which is virtually untenable with respect to geographic information. Thus any experiment which results in acceptance of this null hypothesis suggests a Type II statistical error—acceptance of the null hypothesis when in fact it is false.

Tobler's law is an observation about geographic space, and thus clearly not true of all spaces, although it seems that much theorizing about spatial information has assumed it implicitly. For example, Tobler's law is clearly implicit in any discussion of uniform regions or polygons.

2.3 Geographic Space Is Heterogeneous

In the discipline of geography there is an ancient debate that is still annually rehearsed in seminars on geographic thought, concerning whether the purpose of research should be to discover general truths, or to document specific facts; the two positions are termed *nomothetic* and *idiographic* respectively. While the former is often presented as more *scientific*, it is also possible for idiographic description to follow scientific principles of replicability. In a geographic context the two are expressed as distinct strategies with respect to our understanding of the Earth's surface; in the nomothetic strategy, research is successful if it uncovers principles that are true everywhere in the domain, while the idiographic strategy supports detailed study of the unique characteristics of places, that may or may not lead to generalizations about the entire domain. Clearly the nomothetic strategy requires some degree of homogeneity of the domain, not perhaps in its form, but probably in the processes that modify and shape it; and the search for such processes dominates the nomothetic approach. On the other hand the idiographic strategy requires no homogeneity at any level.

One possible compromise between these two positions exploits the potential of geographic characteristics to repeat themselves. For example, all of the world is not like Bloomfield Hills, Michigan; but market researchers are well aware that the characteristics of Bloomfield Hills relevant to marketing are very much like those of Scottsdale, Arizona. Thus it may not be possible to generalize from one region to the entire planet, but it is often possible to generalize from one region to several similar regions; geography may not be uniform, but it may be repetitive. The strategy relies on our ability to define and measure suitable metrics of similarity.

Recently another compromise strategy appears to be emerging, and to be gaining in popularity. This strategy argues that any general laws relevant to the geographic world are likely to be of limited predictive power, unlike, say, the general laws of physics. The unexplained variation in any law is likely to be geographically variable, because the Earth's surface is essentially heterogeneous. Thus it is appropriate to define a law to the level of its inputs and outputs, but to regard one or more of the parameters of the law as geographic variables. For example, consider a law $z = f(y)$, and assume that f is a linear function. We might expect the law to apply everywhere, but we might expect its constants a and b (as in $z = a + bx$) to vary geographically. Such variation can be readily exposed as shown in Figure 2. Geographically Weighted Regression (GWR; [7]) is one of a number of *place-based* analytic techniques that adopt this compromise between the nomothetic and idiographic strategies.

The Earth's surface exhibits enormous variation, and because of Tobler's law it is often necessary to scan a large fraction of the surface to encounter all of its variability; a small area of the surface typically encompasses only a small fraction of any variable's total variation. It follows that the results of an analysis almost always depend explicitly on the geographic bounds of the study region, and that a shift of boundaries will produce different results. A small region does not produce a representative sample of the Earth's surface. As with spatial dependence, spatial heterogeneity appears to be a defining characteristic of geographic space.

2.4 The Geographic World Is Dynamic

As noted in the previous section, the heterogeneous nature of the geographic world, coupled with the nomothetic need to generalize, leads inevitably to an emphasis on the study of process in preference to form. Geomorphologists, for example, have long argued that study of process is of greater significance and value than study of form; that understanding how the world *works* is more important than understanding how it *looks*. The processes of interest to geomorphologists are natural, but the argument applies as well to the human processes that modify the landscape, such as settlement and migration, as to the physical processes such as erosion and tectonic activity. The world of geographic information is also concerned with *design*, or the study of deliberate, normative modifications of the landscape by human action (*e.g.*, [16]).

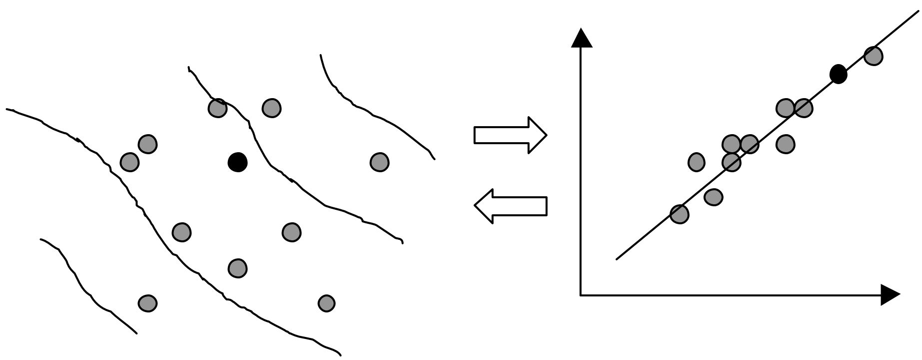

Fig. 2. Geographically Weighted Regression is conducted as follows: (1) Select one observation as reference point, and weight all other observations according to a decreasing function of distance from the reference (*e.g.*, by a negative exponential function of distance); (2) Fit the constants *a* and *b* using points weighted in this way, and assign the derived values to the reference point; (3) Repeat for all observations, and interpolate complete surfaces for *a* and *b* (only one surface is shown).

By contrast, our perspective on the geographic world is relatively static, and most of our information comes in the form of snapshots at specific instants of time. The lack of attention to time in geographic information systems, which draw heavily from cartographic roots, is often recognized, as is the relative importance of information about change to the development of public policy and the making of decisions. Escaping a static view of the world remains one of the most important challenges of GIS.

There are many kinds of geographic information. One normally cites maps and images as the most familiar examples, but geographic information can also take the form of text description, spoken narrative, and even music (the *songlines* of the Australian aborigine are a form of geographic information; [4]), since all of these meet the definition of geographic information given above. Information about dynamic processes is expressed in many different forms: as mathematical models, such as partial differential equations (PDEs; *e.g.*, [17]); as conceptual models expressed in text or diagrams; and as computational models expressed in computer code. But none of these meet the definition, since none is reducible to the atomic form <**x**,**z**>. Yet they are certainly expressible in binary form, given appropriate methods of coding.

Dynamic process models are analogous to the transformations familiar to users of GIS, because they map the geographic world from one state to another. For example, a PDE expressed in numeric form as a finite difference computer code takes the initial state of the system, and computes future states based on appropriate functions and parameters. In that sense dynamic process models are similar to GIS operations such as buffering, which similarly accept input geographic information and produce new geographic information as output. From an object-oriented perspective, dynamic process models are akin to the *methods* that can be *encapsulated* with object classes.

The field of *geocomputation* has emerged in recent years as an intellectual home for research on dynamic process models and their implementation. The relationship between geocomputation and geographic information theory has been discussed by several authors and in several presentations (*e.g.*, [2], [6]), but remains controversial. If study of process trumps study of form, as it clearly does in many areas of science, and if it motivates much acquisition and analysis of geographic information, then an understanding of process is clearly important to effective theorizing about geographic information. I believe therefore that links to the study of process can enrich geographic information theory, and that dialog is essential between the geocomputation and geographic information theory communities.

2.5 Much Geographic Information Is Derivative

The raw data of science often consist of original measurements made with instruments. The terms *accuracy* and *precision* refer to the fit between measurements and truth, and repeated measurements respectively. For many instruments these parameters are well known, and can be used to analyze the impacts of errors on subsequent analyses.

The geographic information that is presented on maps and in databases is rarely composed of original measurements, however. A user of a soil database sees polygons with uniform classes, rather than the original measurements that were obtained by analyzing soils collected in pits, or the aerial photographs that were used to extrapolate the information obtained from pits to create a complete coverage. Much geographic information is similarly the result of compilation, interpretation, analysis, and calculation, almost all of which remains hidden from the user. The visible form of representation (classified polygons in this example) may have little relationship to the forms of representation used at earlier stages (*e.g.*, point samples, rasterized aerial photographs, digital elevation models).

Consider the soil database example in the context of uncertainty, and the impact of uncertainty on its polygons and homogeneous classes. Let uncertainty be interpreted as meaning that other databases might equally well have arisen through the process of derivation, and that in the absence of other information all such alternative databases should be taken to be equally likely. For example, the errors inherent in the measurement of properties in the field will eventually result in alternative databases.

Without any knowledge of the process of derivation, we have no guidance about the form that such alternative databases might take, and must therefore consider every possibility. Thus we have no reason to assume that alternative databases will have boundaries in the same positions, or even the same numbers of polygons, edges, and nodes. For example, different compilers will most likely have produced databases that are topologically as well as geometrically and thematically distinct, despite working from the same original data.

To pursue this example a little further, as an instance of any representation of a nominal field (including databases of soils, land cover, land use, vegetation type, land ownership, *etc.*), it is clear that many of the models of uncertainty studied in spatial information theory represent somewhat arbitrary choices. The egg-yolk model, for example, focuses on individual polygons, and implies that alternative databases will have the same topology (the same boundary network). Moreover, it proposes that the region of uncertainty in each polygon will be adjacent to the boundary. Although this is in a sense consistent with Tobler's law, there are several reasons why it and other similar models such as the epsilon band may be inappropriate for many types of geographic information.

First, in the case of a database derived from remote sensing, the definition of a class is statistically based, and influenced by the relative abundances of pixels in different parts of the spectrum. If spectral responses vary systematically with distance inwards from the polygon boundary, then the responses typical of the periphery will be more common than the responses typical of the center. Thus the choice of a class for a polygon may depend more on the peripheral areas than the central area, just as the suburban class is more typical of a city than the central core. In this sense the periphery may be more certain than the core.

Second, an important stage in the compilation of any soil map is cartographic—a cartographer ultimately determines the positions of polygon boundaries, and decides which small patches should be separate polygons, and which should be merged with their surroundings. Any mapped polygon will likely contain many small inclusions, or patches of some other class, that have been deleted by the cartographer. Now consider such an inclusion near the polygon edge, and assume it is similar to the class of the neighboring polygon across the edge (Figure 3). When the line is drawn, it may be able to accommodate the inclusion by modifying the polygon boundary. But inclusions near the core of the polygon must be ignored. In summary, the cartographic process of map compilation may lead to greater lack of homogeneity in the core of polygons than on the periphery.

Finally, a distance-based epsilon band or egg-yolk model raises awkward issues of process, since it is difficult to think of real processes that might lead to a zone of uncertainty of uniform width inside a polygon. In a botanical example, it is possible that dispersion of seeds into an area from outside its boundary might produce a uniform gradient of uncertainty, but it is hard to imagine a similar process operating in the case of soils. Thus from a geographic perspective, there seem to be good reasons not to believe in epsilon bands or egg-yolk models, but to take a broader view of the alternatives that result from uncertainty. In such cases spatial information theory seems *more* restrictive than geographic information theory, which is counter to the arguments presented earlier.

What is clearly lacking in all such discussions of the uncertainty associated with polygons is a clearly defined and reasonable model of how one specific set of polygons resulted from the process of compilation—such a model would also provide a basis for theorizing about uncertainty, by modeling the generation of alternative databases. What is missing, then, is a set of comprehensive models of uncertainty in nominal fields, that serve to frame the methods used to compile databases. Although many such models have been discussed in the literature, I will focus only on one [9] as an example to demonstrate the efficacy of this approach to conceptualizing uncertainty in one class of geographic information. I do not want to suggest that this model is in any way unique, or even the most appropriate in many instances. However it seems to provide one conceptual *framework* for the process by which the polygons of a soil database came into being.

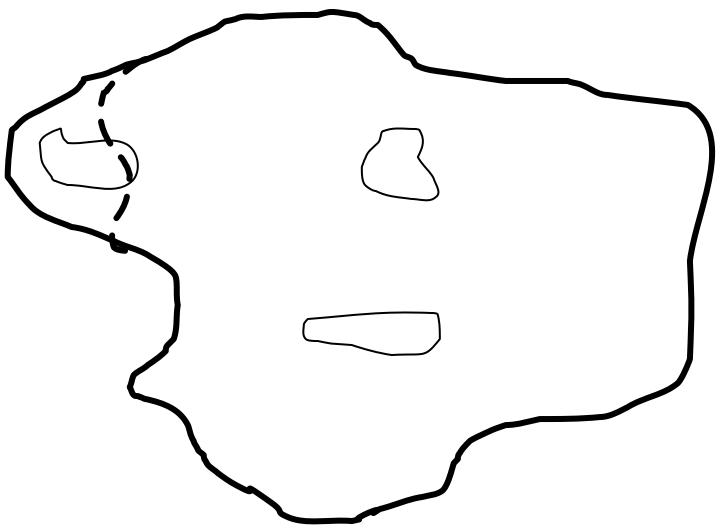

Fig. 3. In drawing a polygon (heavy solid line) a cartographer will ignore inclusions of a different class that fall below the size of the minimum mapping unit. But an inclusion near the boundary may result in a modification of the boundary's position. Thus there may be greater uncertainty in the center of the polygon than in the periphery.

Consider a set of fields $\{z_1(\mathbf{x}), z_2(\mathbf{x}), \ldots\}$ measured on continuous scales. Each field represents the spatial variation of one measurable quantity relevant to soil mapping, such as soil pH, depth to water table, or organic carbon content. Now consider a space defined by these variables (Figure 4 shows an example in the case of only two variables). Define $c(\mathbf{z})$ as a function over this space, the discrete values of c defining a set of classes. The space and its function are analogous to the classifiers used in remote sensing (where the defining variables are spectral responses in the various bands of a sensor; for example, a *parallelepiped* classifier is named for the geometric form of the domains formed in $\mathbf{z}$ by values of c). For the purposes of this paper I term this a *phase space* by analogy to the physical states of a substance. Finally, map any geographic location **x** to a class $c(\mathbf{x})$ by determining its measurable quantities $\{z_1(\mathbf{x}), z_2(\mathbf{x}), \ldots\}$, and identifying the class associated with those quantities.

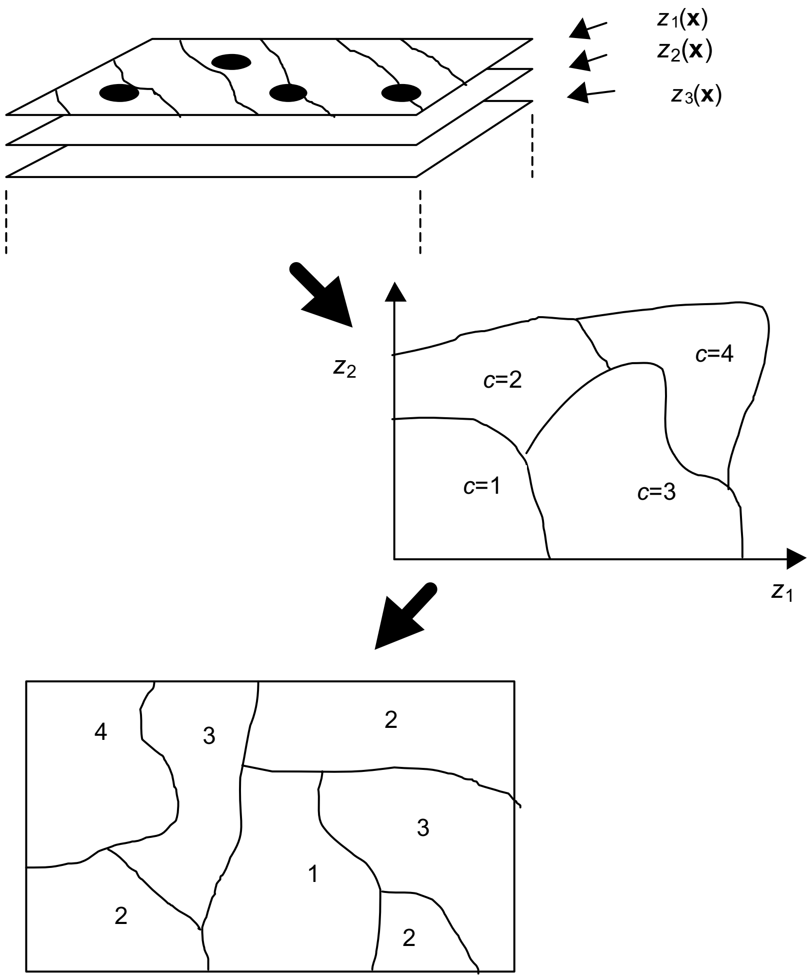

Fig. 4. A possible model for the derivation of polygons in a soil map. Variables are measured at sample points, and interpolated to form continuous fields. A phase space assigns every vector of field values to a class. Finally, the interpolated fields and phase space are combined to form a nominal field.

Now consider the implications of this model. First, successive determinations of the underlying variables $z_1, z_2, \ldots$ will be subject to the measurement errors inherent in the relevant instruments. In practice the variables will not have been measured everywhere, but will have been interpolated from point measurements, so interpolation errors will need to be included, perhaps using the techniques of geostatistics [12]. Thus it will be possible to simulate alternative measurements and interpolations (see specifically *conditional simulation*), and consequently alternative databases. Second, the implications of scale change can be examined by coarsening the underlying variables $\{z_1(\mathbf{x}), z_2(\mathbf{x}), \ldots\}$, which is readily done using simple convolution filters. Third, the implications of coarsening or refining the classification scheme can be examined by making appropriate changes to the phase space (*e.g.*, a class can be subdivided into two or more classes by subdividing its domain).

The model provides an easy way of conceptualizing the implications of Tobler's law. Since all of the underlying variables are geographic, we expect them to exhibit strong spatial dependence, and this of course is the basis for all techniques of spatial interpolation. It follows that two classes can be adjacent in geographic space if and only if they are adjacent in phase space.

In summary, much geographic information is derivative, in the sense that it is the result of compilation, interpretation, analysis, and calculation from original measurements that are not normally exposed to the user; these processes can involve many stages and many individuals. A model such as that presented above provides a way of conceptualizing the process of creation of a nominal field (and the collection of polygons used to represent it). Moreover, uncertainty is represented explicitly in the model, in this case as measurement error in the original point observations, and errors in the process of interpolation used to create continuous fields. Thus the alternatives to be expected due to uncertainty can be modeled explicitly, as a comparatively narrow range of options. Models such as the epsilon band or egg-yolk, which assume no such background conceptual framework, can be examined to see if they are feasible within the framework, and to determine the degree of generality of their assumptions with respect to the framework.

2.6 Many Geographic Attributes Are Scale-Specific

Consider the field defined by the elevation of the Earth's surface. Overhanging cliffs, or locations $\mathbf{x}$ where the field is many-valued, are sufficiently rare to be ignored in most circumstances. Elevations are discontinuous at cliffs, where $z(\mathbf{x}+\delta\mathbf{x})$ does not tend to $z(\mathbf{x})$ as $\delta\mathbf{x}$ tends to zero. More importantly gradients are discontinuous at ridges and sharp valleys, where the surface lacks well-defined tangents or derivatives. Such properties are characteristic of fractal surfaces, and Mandelbrot [14] has shown how fractal behavior is typical of many geographic phenomena.

One of the commonest GIS functions applied to digital elevation models is the determination of slope. Since the elevation surface is already represented in such models as a finite-difference approximation, or a regular grid of point measurements, it is convenient to estimate slope by comparing elevations over a neighborhood of such points, typically a 3 by 3 neighborhood. Burrough and McDonnell [3] and others review the alternative estimating equations. Implicit in this approach is the dependence of the resulting estimates of slope on the grid spacing. But if the elevation surface lacks tangents, these slope values are not estimates of the derivatives of the surface, but explicitly scale-specific. In essence, there is no such thing as the slope of a geographic surface, only slope *at a specific scale* or grid spacing.

This property of scale specificity is very general for geographic data, and extends well beyond the case of interval fields such as elevation. The derivation process for nominal fields discussed in the previous section is also scale specific, as is the definition of many of the classes used in geographic databases. Consider the example of the land cover class *urban*. Scale is not often specific in its definition, but is clearly important. The pixels covering New York City may be roughly homogeneous in spectral response when seen from the AVHRR satellite, with a ground resolution of approximately 1.1km, but at the 4m resolution of the multispectral IKONOS sensor the homogeneity breaks down into grass, concrete, asphalt, roof materials, *etc.*

Scale specificity has obvious implications for any theory of the effects of spatial and thematic resolution on geographic data. Rather than breaking down at finer scales, the domain *urban* in the phase space discussed in the previous section disappears completely, and its replacement classes of grass, concrete, asphalt, *etc*. may share none of its boundaries. Thus we cannot assume a hierarchical relationship between coarse and refined classes. Instead, it seems likely that new classes will be needed *on the boundaries of coarse classes* in phase space as well as within their domains, since it is probably here that the greatest heterogeneity exists.

3 Conclusion

I have focused in this paper on the differences between *spatial* and *geographic*, defining those terms such that geographic is a specialization of spatial. The six general properties discussed above are clearly only a sample, and there may well be others that are equally or more important in specializing spatial. Each of these specializing properties provides a basis for extending spatial information theory, by narrowing the set of possibilities that it must consider, and thus allowing theory to be extended and deepened. In the case of the framework model, the geographic case provides a basis for additional theorizing through the formulation of a background framework, or model of the process by which geographic information was compiled. Finally, I have identified ways in which the specialization of spatial appears to have proceeded in a direction that is inconsistent with the general properties of geographic.

The properties discussed in this paper are generalizations from empirical observations, and as such fall into a classic tradition of observations that serve to drive theory. Although there is value in theorizing in the absence of such general observations, there is clearly much greater practical value in theory that is grounded in empiricism. In this case, the domain of spatial is far greater than the domain of geographic, and many more subclasses exist, each of which can be expected to exhibit general properties that may or may not be similar to those exhibited by the geographic domain. Thus theorizing about spatial information results in conclusions that apply in all domains, whereas theories about geographic information may apply only to the geographic domain, and this potential disadvantage must be weighed against the advantages of domain-specific theory driven by empiricism.

Acknowledgment

The Center for Spatially Integrated Social Science is supported by the National Science Foundation.

References

1. Anselin, L.: What is Special About Spatial Data? Alternative Perspectives on Spatial Data Analysis. Technical Report 89-4. National Center for Geographic Information and Analysis, Santa Barbara, Calif. (1989)
2. Berry, B.J.L.: Towards Computational Geography. Paper presented at GIScience 2000, Savannah, Georgia, October 28 (2000)
3. Burrough, P.A., McDonnell, R.A.: Principles of Geographical Information Systems. Oxford University Press, New York (1998)
4. Chatwin, B.: The Songlines. Viking, New York (1987)
5. Cliff, A.D., Ord, J.K.: Spatial Processes: Models and Applications. Pion, London (1981)
6. Couclelis, H.: Geocomputation in Context. In Longley, P.A., Brooks, S.M., McDonnell, R., MacMillan, W., Geocomputation, A Primer. Wiley, New York (1998) 17–30
7. Fotheringham, A.S., Brunsdon, C., Charlton, M.E.: Quantitative Geography. Sage, London (2000)
8. Goodchild, M.F.: Scale in Remote Sensing and GIS. In McMaster, R., Sheppard, E.S., editors, Scale and Geographic Inquiry. Blackwell, Oxford (in press)
9. Goodchild, M.F., Dubuc, O: A Model of Error for Choropleth Maps with Applications to Geographic Information Systems. Proceedings, Auto Carto 8. ASPRS/ACSM, Falls Church, Virginia (1987) 165–174
10. Goodchild, M.F., Egenhofer, M.J., Kemp, K.K., Mark, D.M., Sheppard, E.: Introduction to the Varenius Project. Int. J. Geogr. Info. Sci. 13 (1999) 731–745
11. Güting, R.M., Schneider, M.: Realm-Based Spatial Data Types: The ROSE Algebra. VLDB J. 4 (1995) 100–143
12. Isaaks, E.H., Srivastava, R.M.: Applied Geostatistics. Oxford University Press, New York (1989)
13. Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W.: Geographic Information Systems and Science. Wiley, New York (2001)
14. Mandelbrot, B.B.: The Fractal Geometry of Nature. Freeman, San Francisco (1983)
15. Matheron, G.: The Theory of Regionalized Variables and Its Applications. Ecole National Superieure des Mines, Paris (1971)
16. McHarg, I.L.: Design with Nature. The Natural History Press, Garden City, N.Y. (1969)
17. Smith, T.R., Birnir, B., Merchant, G.E.: Towards an Elementary Theory of Drainage Basin Evolution I: The Theoretical Basis. Comp. Geosci. 23 (1997) 811–822
18. Tobler, W.R.: A Computer Movie: Simulation of Population Change in the Detroit Region. Economic Geography 46 (1970) 234–240