

Uncertainty and Interoperability: The Areal Interpolation Problem

Michael F. Goodchild, Phaedon C. Kyriakidis, Philipp Schneider, and Jorge Sifuentes

National Center for Geographic Information and Analysis, and Department of Geography

University of California, Santa Barbara, CA 93106-4060, USA

Tel: +1 805 893 8049, Fax: +1 805 893 3146

E-mail: good@geog.ucsb.edu, phaedon@geog.ucsb.edu

Abstract

Discussions of geographic interoperability normally address two sets of issues: syntactic interoperability, having to do with compatibility of data formats, and semantic interoperability, having to do with the meaning ascribed to data by the sender and the receiver. In this paper we propose two additional dimensions, both related to data quality: issues associated with data accuracy, and affecting the ability to work with data from sources of different accuracies; and spatial support, having to do with differences of spatial sampling. The case of areal interpolation is explored, and it is shown that all known methods of areal interpolation can be brought within a single, unified framework based on geostatistics, and defined by the assumptions the user is willing to make about a hypothesized, underlying continuous field. Examples are used to illustrate the approach. The nominal case is discussed, and an empirical method presented that also appears to provide some promise in overcoming differences of spatial support.

Key words: Interoperability, areal interpolation, accuracy, spatial support.

1 Introduction

Much has been written over the past ten years on the topic of interoperability (see, for example, Goodchild *et al.*, 1999), which is defined generally as the ability of systems to share data—two GISs are said to be interoperable if data from one can be read and used by the other. Issues of interoperability arise in various domains of the computing industry because software developers adopt different formats for data, and in many cases treat these formats as proprietary, reserving access to technical details of formats to their own programmers. In such situations users are forced to employ special import and export formats to exchange data, since the details of internal formats are unknown. Much progress has been made in this area through the development of open standards, notably by the Open Geospatial Consortium (www.opengeospatial.org), which advocates an open rather than a proprietary approach to geographic data formats, and has developed a series of specifications, including GML, WMS, WFS, and WCS, all of which facilitate syntactic interoperability across the Internet.

In addition to issues of format, or so-called *syntactic* issues, interoperability is jeopardized when the creators and users of data attach different *meaning* to them, or fail to understand meaning. We call these issues *semantics*, and recognize both syntax and semantics as major dimensions of the interoperability issue. Semantic issues arise whenever two agencies use different classification schemes in their data, for example when Class A has different meaning to two agencies both mapping land use (Cruz *et al.*, 2002, 2003), or when agencies adopt different definitions of key terms such as *wetland*. Semantic issues also arise when data are not described completely, allowing the creator and user to differ in the meaning they attach to the data. For example, omitting the specification of a datum from the metadata of a geographic data set can introduce a substantial positional uncertainty of as much as hundreds of meters to positions specified in latitude and longitude.

In this paper we argue that two additional dimensions of interoperability are important in GIS, both related to data quality. We first outline those two additional dimensions, and review methods that are already known for addressing them. We focus on one, and use the cases of areal interpolation and reclassification to illustrate our approach. We present a new perspective on the areal interpolation problem that integrates all known methods into a single framework, embedded in the formal structure of geostatistics.

First, two data sets may differ in their *accuracy*. For example, consider a digital elevation model of an

area, and a satellite image, and suppose that a user wishes to drape the image over the DEM to create a three-dimensional rendering (for a realistic example see Google Earth, earth.google.com). Since the image and the DEM are unlikely to have a common lineage, their relative positions will be distorted by the absolute positional errors in both, with the result that they will not fit perfectly. In many cases the degree of misfit will be insufficient to cause a problem, but inevitably there will be other cases where misfit is unacceptable, perhaps in areas of comparatively poor data registration, or perhaps in particularly detailed applications.

Goodchild and co-workers (Goodchild, 2002, 2004) have called for a distinctly different approach to GIS to address this problem, by contrasting conventional GIS, which they term *coordinate-based GIS*, with a new approach that is in many ways reminiscent of longstanding practice in surveying and geodesy, which they term *measurement-based GIS*. A measurement-based GIS stores the measurements and procedures by which positions are determined, rather than the positions themselves, thus allowing models of positional error to be constructed, and allowing the effects of improved measurements to be propagated automatically.

Our fourth dimension is concerned with differences of *spatial support*, a term that is commonly used in spatial statistics but rarely encountered in GIS. Consider a continuous field $z(\mathbf{x})$, mapping location \mathbf{x} to some function z . For example, z might represent ground surface elevation, population density, or surface temperature. In principle \mathbf{x} may have any number of dimensions, but for most purposes will be limited to the two horizontal spatial dimensions. Measurement of the field, and its representation in a digital system, require some kind of sampling or discretization, and six such methods are commonly encountered in GIS (Longley *et al.*, 2005). First, $z(\mathbf{x})$ may be measured at a finite number of *irregularly distributed sample points*, such as we find in the characterization of meteorological variables. Second, the points may be regularly distributed, normally in a *rectangular grid*, as we find in many DEMs, which record spot elevations over such a grid (note that in principle a rectangular grid cannot exist on the curved surface of the Earth; this approach requires that the Earth first be projected to a flat surface). Third, the field may be represented by assigning a value to each of the cells in a *regular tessellation*, most often a rectangular grid, such as we find in remote sensing, where the assigned value of each cell consists of the field convolved with a function characteristic of the sensor. Fourth, an *irregular tessellation* may be used, such as we find in the characterization of such variables as household income by statistical agencies; in such cases the value assigned to each cell of the tessellation will likely be either the integral of the field over the cell in the case of fields of density, or its mean value. Fifth, the field may be characterized by assigning linear functions to each cell of a triangular mesh, such that values are zero- and first-order continuous across cell boundaries—this approach is commonly known as the *triangulated irregular network*, or TIN, in GIS. Finally, the field may be characterized by creating digital representations of its *isolines*, as is often done in the case of topographic or bathymetric maps.

Many other methods are possible in addition to these six, though they are rarely supported by GIS software. The field may be represented by a collection of *splines*, by a *Fourier transform*, by *wavelets*, by the mesh of triangles and quadrilaterals that is commonly encountered in *finite-element modeling*, and by many other approaches.

Collectively, we refer to the set of features, whether points, lines, or areas, that is used to measure or represent the field as its spatial support. In some cases, such as a DEM, spatial support is unique to a field z , while in other cases, such as the summary tables provided by census agencies, the same spatial support (*e.g.*, an irregular tessellation of census tracts) may be used for many variables. Meteorological agencies also typically use the same network of weather stations to report many variables. Nevertheless GIS users frequently find themselves faced with overcoming differences of spatial support when analyzing or modeling spatial data. In the next section we present three use cases that typify such problems.

1.1 Sample use cases

First, consider the problem of estimating the value of a field from multiple sources with differing spatial support. For example, suppose that $z(\mathbf{x})$ denotes ground surface elevation, and that an estimate of elevation must be made at some point \mathbf{x}_0 . Three data sources are available, all of different lineage. A set of spot heights has been measured using GPS, and the statistical distributions and correlations of the errors associated with these measurements (of \mathbf{x} as well as z) are known. A second data set has

been obtained by digitizing the isolines of a topographic map, and information on data quality is available in the form of the quality standards of the mapping agency. A third data set consists of a DEM, again of independent lineage (perhaps derived by interferometry). We assume that \mathbf{x}_0 is not coincident with a spot height, an isoline, or a DEM point. The three data sets have incompatible spatial support, providing examples of three of the six approaches discussed above. An estimate of elevation should combine the information available from the three data sets in some optimal fashion. One approach would be to reduce all three data sets to a common spatial support, for example by using a standard method of point interpolation to estimate elevations at the DEM points from the GPS measurements, and by using a method of contour-to-grid interpolation to similarly interpolate from the isolines to the DEM. But such a method would be entirely lacking in theoretical basis, using essentially arbitrary techniques of interpolation. No estimates of estimation variance would be available, and no basis would be apparent for weighting the three sets of values. For example, intuitively one would want the weight assigned to the isoline-sourced data to vary depending on distance from the nearest isoline, but we lack a theoretical basis for assigning such weights.

Second, consider the problem of estimating the relationship between public perceptions of ambient noise, and actual measurements of noise. Assume that public perceptions are characterized through a series of interviews with households, and that measurements are made at a series of points within the study area. The locations of the interviews and the locations of the measuring instruments do not coincide. Before any relationship can be analyzed, therefore, some way must be found to overcome the incompatibility of the two sets of spatial support. As before, we might interpolate from Data Set A's locations to those of Data Set B, or vice versa, or interpolate both data sets to a new set of points, perhaps a regular grid. But again we would have no theoretical basis for doing so. Moreover one could in principle interpolate to a grid of any density, raising the possibility of increasing the apparent degrees of freedom in the ultimate analysis without limit.

Finally, consider a case of two incompatible areal supports—this third use case provides the basis for much of the subsequent discussion of the paper. *Areal interpolation* has been defined (Goodchild and Lam, 1980) as a technique for addressing such problems. For example, suppose that our task is to investigate the relationship between two variables, but that measurements of one are available for one irregular tessellation, and measurements of the other are available for a different irregular tessellation—and assume that there is no simple geometrical relationship, such as hierarchical nesting, between the two tessellations. The problem is normally addressed by estimating the values of the first variable for the second tessellation. Call the first tessellation the *source* zones and the second tessellation the *target* zones. Then such techniques allow an analysis of the relationship between both variables for the second, target set of zones. As before, note that in principle both data sets could be interpolated to a third tessellation, perhaps a regular one, and that the spatial resolution of this new tessellation could be arbitrarily fine.

2 Approaches to Areal Interpolation

As noted earlier, the focus of this paper is on the areal interpolation problem, as an issue of interoperability across incompatible spatial support; we plan to address the other use cases in future research. In this section we review several known approaches to the areal interpolation problem.

First, however, we need to introduce two terms (Longley *et al.*, 2005). Consider a field $z(\mathbf{x})$ representing population density, or the density of some other comparable property. We noted earlier that two distinct types of estimate could be made of the properties of some irregular area within a tessellation. First, the field could be integrated over the area to obtain an estimate of total population. We term this an instance of a *spatially extensive* variable because its value *extends* with extension of the area, and cannot be true of every part of the area—its value can be true only of the whole area.

Second, the field could be *averaged* within the area. We term this a *spatially intensive* variable because in principle the value could be true of the entire area (it represents an *intensity*), if the value of the field was constant within the area. Spatially extensive and spatially intensive variables behave differently when areas are split or merged; under merging, spatially extensive variables sum while spatially intensive variables average.

2.1 Area weighting

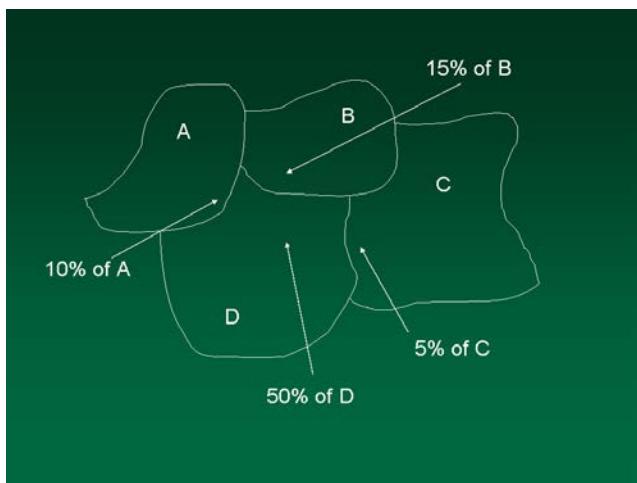


Figure 1: A simple illustration of area-weighted interpolation using four source zones and one target zone

principle integrated to obtain the source zone populations, is constant within each source zone and changes abruptly across source zone boundaries.

Several variations on this simple approach are known. A straightforward modification allows spatially intensive variables to be interpolated (Goodchild and Lam, 1980). The assumption of uniform density within *target* zones can be handled through a least-squares approach, and Goodchild, Anselin, and Deichmann (1993) extend the approach to the case of a third set of zones, termed *control zones*, that are assumed uniform when neither of the previous assumptions are tenable.

2.2 Pycnophylactic interpolation

Tobler (1979) described an approach to areal interpolation that explicitly estimates the underlying field, ensuring that the integral of the field over the source zones matches exactly the known source data. The underlying field is assumed to be maximally smooth, unlike the underlying field of the area-weighting case. The field is estimated over a fine raster of cells, and is constrained by boundary conditions that force the field outside the study area to either a value of zero or a gradient of zero.

We begin by setting the values of the field in each cell to equal shares of the value of the containing source zone. For example, if a zone has a population of 1000 and contains 100 cells then each cell will receive a value of 10. A filter is then run over the entire raster, replacing each cell's value with the mean of its neighbors, thus increasing the smoothness of the estimated field. But as a result the cells within each source zone may no longer sum to the known source value, so a proportional adjustment is made. The filter is then reapplied, and the process iterated until changes fall below some user-defined threshold. The field can then be reintegrated to any set of target zones by summing the contents of appropriate cells.

2.3 Point-based interpolation

Third, we review an intuitively reasonable approach that makes use of known methods of point-based interpolation. First we compute a centroid for every source zone, and assign a spatially intensive version of the variable of interest to it. As in Section 2.2 we then overlay a fine raster, and interpolate values of the field for every raster cell, using one of the standard methods of interpolation such as inverse-distance weighting or Kriging (Goovaerts, 1997). Finally, as in Section 2.2 we sum the cells within each target zone to obtain appropriate estimates. The characteristics of the underlying field are defined in this case by the interpolation procedure, and are explicit in the case of Kriging. However the pycnophylactic constraint is no longer satisfied—in general, the integral of the field over source zones will not match the input values.

By far the simplest method of areal interpolation provides estimates of target-zone statistics based on a simple scheme of area weighting. Figure 1 shows a simple example, in which population statistics (spatially extensive) are available for each of four source zones, and an estimate is required for one target zone that overlaps all of them. The simplest version of area weighting apportions the source zone populations to the target zone based on area of overlap. For example, since 10% of the area of Source Zone A lies in the target zone we assign 10% of Zone A's population to the target zone. Underlying this approach is the assumption that the imaginary underlying field of population density, which is in principle integrated to obtain the source zone populations, is constant within each source zone and changes abruptly across source zone boundaries.

3 A Geostatistical Framework

As noted earlier, our intent in this paper is to introduce a geostatistical framework for the areal interpolation problem that integrates known methods, and that provides an illustration of the importance of this fourth dimension of interoperability—and the closely related third dimension of accuracy. In this section we first define the basic terms of our approach, and then sketch its details. Only a brief sketch of the approach will be given here; for a more detailed exposition and analysis see Kyriakidis, Schneider, and Goodchild (2005).

As before, let the underlying, unobserved field be denoted by $z(\mathbf{x})$. We also assume that the pycnophylactic constraint is operative: that the field should integrate or average to give the known source-zone values.

Let the source-zone values be denoted by $\{z(s_k), k=1, \dots, K\}$ where K denotes the number of source zones and s_k denotes the k th source zone. Define a convolution or sampling function $g_k(\mathbf{x})$ that defines how the field is integrated to obtain the values for each source zone. For example, $g_k(\mathbf{x})$ could be the convolution function that integrates a continuous field of spectral response to the measured values of each cell of a remotely sensed image. $g_k(\mathbf{x})$ could also represent the geometry of the source zone, and the aggregation process that produces source-zone statistics from the underlying field. Then:

$$z(s_k) = \int_{\mathbf{u}} g_k(\mathbf{u}) z(\mathbf{u}) d\mathbf{u}, k = 1, \dots, K$$

and similarly target-zone values can be estimated by using a set of similarly defined target-zone convolution functions.

We now invoke a geostatistical framework, conceptualizing the underlying field as the outcome of a stochastic process characterized by a constant mean and a semivariogram $\gamma(h)$ that is a function only of distance h . Since the model can produce an infinite number of realizations, it is possible within this framework to determine the uncertainty associated with any estimate. This represents a major advance on the known methods, none of which provides any information on the uncertainty associated with its estimates.

Within this framework, the structure of the underlying field is characterized by its semivariogram, whose functional form follows one of a number of possible models. Each semivariogram is characterized by a *nugget* $\gamma(0)$, or the variance of the field at points whose distance apart tends to zero. The smooth fields of Tobler's pycnophylactic approach clearly have no nugget, but on the other hand the area-weighting method includes infinitely rapid change in population density across zone boundaries, and therefore exhibits a strong nugget. The semivariograms of spatial variables are also generally observed to rise monotonically and asymptotically to a *sill*, at a distance termed the *range*, beyond which there is little further increase.

Figure 2 shows a simple example which we use to illustrate the approach. The first data set has spatial support in the form of a tessellation of four source zones, while the second data set covers the same area, again with a tessellation of four target zones. The variable is shown in the figure in spatially intensive form.

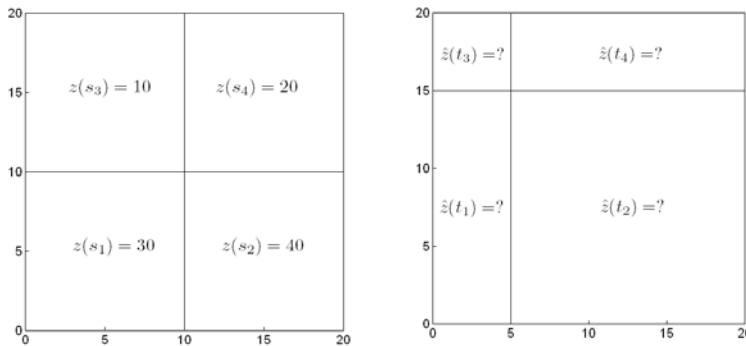


Figure 2: An example of four source zones (left) with associated spatially intensive values, and four target zones (right).

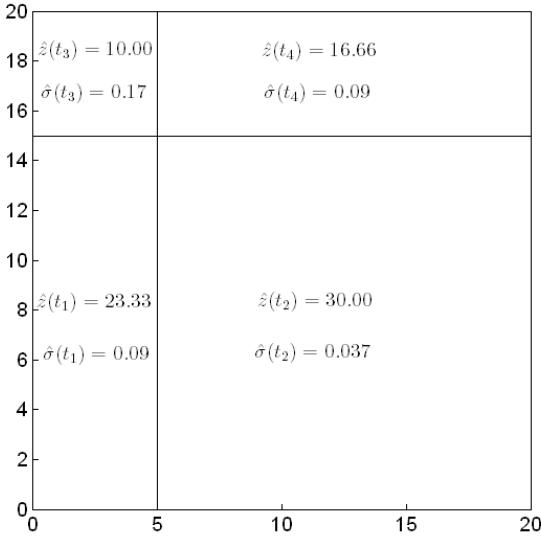


Figure 3: Results for the case of area weighting (nugget equal to sill, zero range)

of the range increases spatial effects also increase, strengthening the general trend across the area, and reducing the estimation variance. In all cases one can easily verify that the pycnophylactic constraint is satisfied: the total of target-zone values (in spatially extensive form) matches exactly the total of source-zone values.

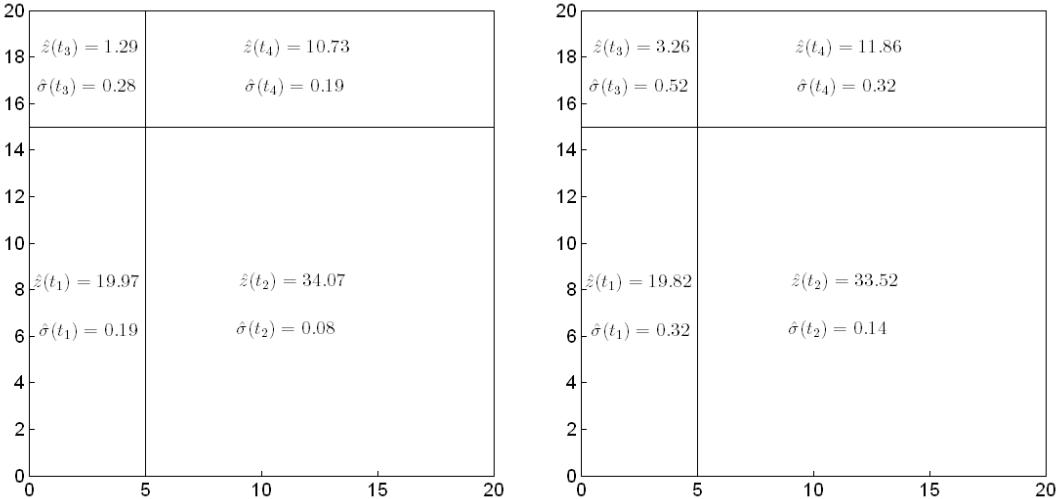


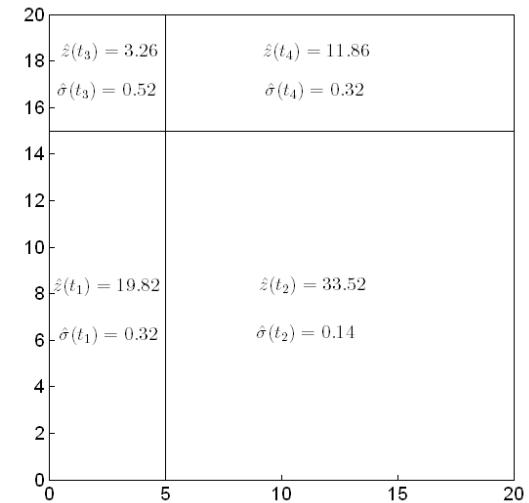
Figure 4: Estimates using different semivariograms: (a) spherical with a range of 10; (b) Gaussian with a range of 20.

4 Interoperability in the Nominal Case

In this section we consider the possibility that z represents a nominal variable, such as soil class or land-use class—in other words, that location \mathbf{u} maps to a nominal or categorical variable. Mark and Csillag (1989) term the results of such mappings *area-class* maps. Spatial support is a somewhat different concept in this case; while the boundaries of an irregular tessellation used by the census are designed somewhat independently of the spatial variation of the variables of interest, in the nominal case the boundaries are located where changes in class are observed. Smith and Varzi (2000) have termed these *bona fide* boundaries to distinguish them from the more arbitrary *fiat* boundaries of cases such as the census. Thus two agencies independently producing area-class maps of the same area but using different classification schemes will almost certainly produce different tessellations, in other

The area-weighting case corresponds to an absence of spatial covariance, in other words to a semivariogram with a nugget that is also its sill, and a range of zero. Figure 3 shows this case, and the estimated target-zone values are identical to those that would be produced by area weighting. But in addition to the estimated values the method also provides standard errors for each estimate, based on the assumption that the underlying field is only one realization of a random process. As one might expect the standard errors are higher for smaller target zones, and lowest for the largest target zone, reflecting the averaging process that occurs over each target zone.

Figures 4a and 4b show two further examples, with different ranges and semivariogram models. Figure 4a uses a spherical semivariogram and a range of 10, while Figure 4b uses a Gaussian semivariogram and a range



words different spatial support. How can classes be converted from one spatial support to another, analogous to the transfer of attributes from source zones to target zones in the interval/ratio case just considered? Achieving interoperability across such differences of spatial support is therefore a significant issue.

In this section we present an approach to the problem that is empirical in nature, in order to elicit the transformation rules that would provide a solution. Two situations appear to yield the necessary information for this approach: an overlap of the two mappings, allowing one mapping to be compared directly to another, and an adjacency between the two mappings, in other words a meeting along a common boundary. In what follows we illustrate the latter approach, using boundary length as a measure, while recognizing that the approach can be readily extended to the overlap case using area as the measure.

Figure 5 shows the example data from the Yucatan Peninsula. The peninsula is divided between three national jurisdictions: Mexico, Guatemala, and Belize, and Figure 5 shows the mappings of soil type produced by the respective agencies. Each agency used a different soil classification scheme, reflecting to some degree the different objectives of the agencies, but also reflecting the general lack of interoperability in this domain. In the Petén area of Guatemala a scheme originating with the work of Simmons (1959) and the U.S. Department of Agriculture was used; in Mexico the mapping agency INEGI (2002, 2004) adapted the classification of the U.N. Food and Agriculture Organization; and in Belize the British Honduras Land Survey Team adopted the classification of Wright, Romney, *et al.* (1959).

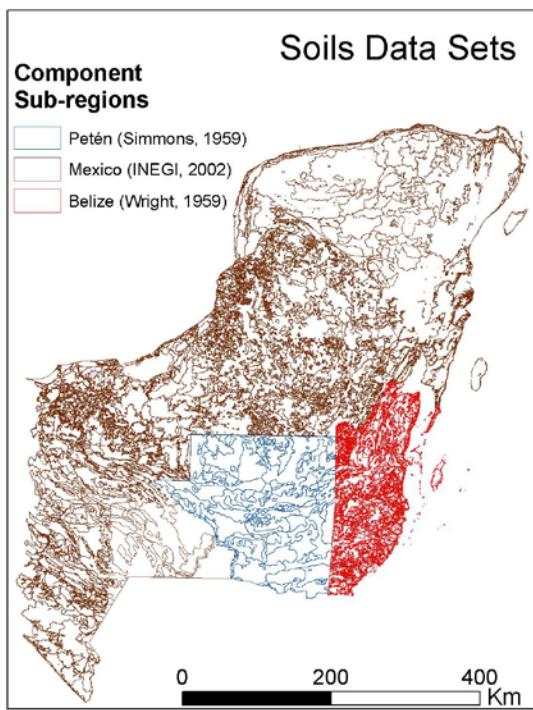


Figure 5: Soil map of the Yucatan peninsula, illustrating the three national jurisdictions.

will appear as blocks of a single entry, 1:n mappings will appear as blocks of dimensions 1 by n, and n:m mappings will appear as blocks of dimensions n by m.

In reality it is unlikely that the matrix will be structured in such a simple way, with blocks of non-zero entries and zero entries, and any simplification will involve a substantial degree of approximation. In this case we chose to organize the mappings based on two ordinal simplifications and careful examination of the definitions for each classification: classes that indicated a decreasing scale of quality of drainage, and classes that indicated a decreasing scale of fertility. Figure 7 shows the

Let the classes of Map A be denoted as i and let the classes of Map B be denoted as j . In the adjacency case we are interested in the common boundary between the two maps, and in the length L_{ij} of boundary that is classified as i on Map A and j on Map B. We assume that the boundary bears no relationship to the process of soil formation, a reasonable assumption for most but not all of the national boundaries on the peninsula, and reject those parts of the boundaries where there is reason to expect a change in the nature of soil, for example where the boundary follows a major river or sharp ridge. In the overlap case we are interested in the area A_{ij} that is classified as i on Map A and j on Map B.

Figure 5 illustrates a simple case of the matrices that will result from this process. Because we have no way of knowing which classes on Map A correspond to which classes on Map B, the matrix will appear as in the top left of Figure 6, with apparently random entries. By reordering the rows and columns of the matrix we can seek a solution of the form shown in the lower right, a type of *block matrix*, in which non-zero entries are clustered in blocks. Any 1:1 mappings between classes

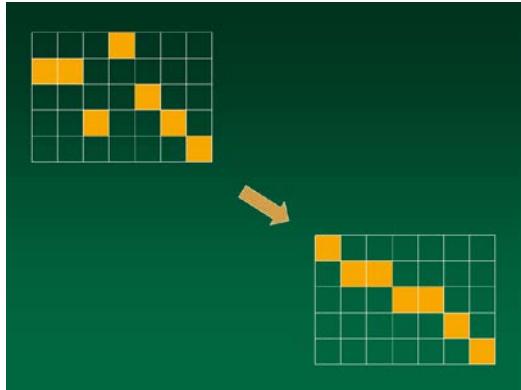


Figure 6: A matrix of overlap areas or lengths of common boundary (upper left), and its reordering (lower right). Filled cells have non-zero entries.

results. The degree to which national boundaries still emerge after reclassification is one test of the effectiveness of this method, and clearly we have been less successful in the case of fertility, since the Guatemala/Mexico border is still obvious.

5 Conclusions

The purpose of this paper has been to define two additional dimensions to the geographic interoperability problem, in addition to the customary dimensions of syntax and semantics. We have argued that accuracy and spatial support are important dimensions that need to be addressed in any comprehensive approach to interoperability. We defined spatial support, showing that six different types of spatial support are common in GIS, and that several other methods are also known, and argued that it is common for data sets to differ on this dimension, leading to problems in

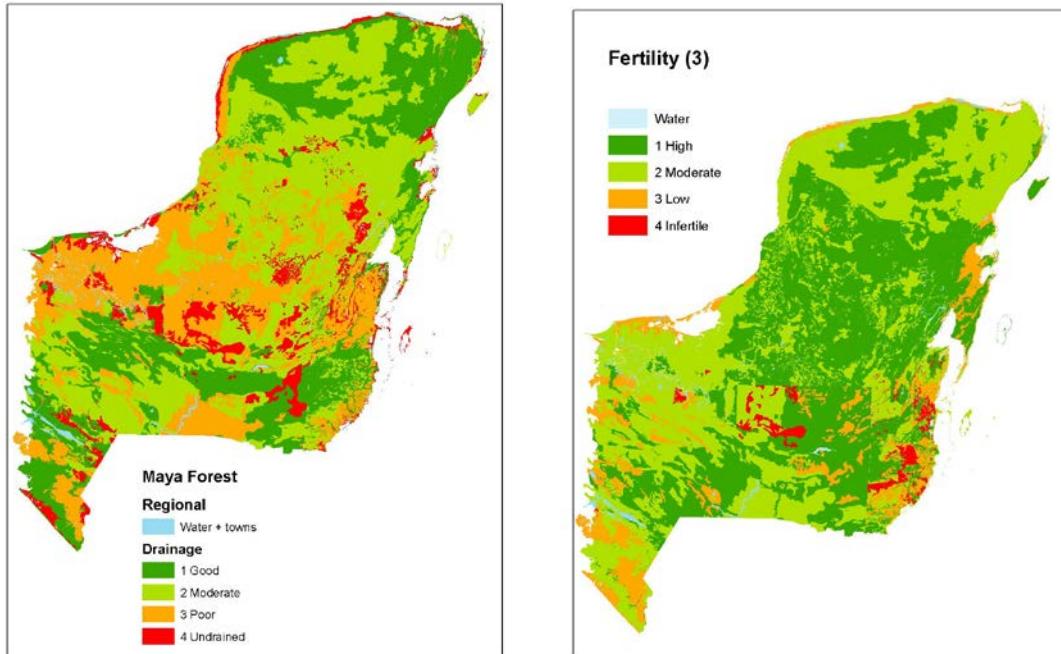


Figure 7: Integrated maps of drainage (left) and fertility (right) after reclassification across national boundaries based on prevalence of adjacent classes.

any subsequent analysis or modeling.

Areal interpolation provides a useful and convenient example of a technique for overcoming differences of spatial support. We showed that all known methods, including area weighting, Tobler's pycnophylactic method, and point interpolation can be brought under a single unifying framework based on the assumptions and concepts of geostatistics, and provided an illustration. The method requires the estimation of a point-level semivariogram, and while some information can be gleaned from the block-level analysis of input data, it seems more likely that knowledge of the characteristics of the point-level semivariogram will have to come from more intuitive observations of the spatial structure of human population distributions, and perhaps from a series of structured fine-scale studies under a rigorous experimental design. Such studies are clearly needed, since the method we have

presented in this paper provides a very powerful approach to the spatial support problem.

We also discussed the nominal case of the area-class map, which differs both in the nature of the boundaries that define the spatial support, and in the nature of the variable being mapped. We presented a method for analyzing adjacent or overlapping mappings to elicit information that can provide the basis for reclassification.

Given the always-uncertain nature of geographic information, it is perhaps surprising that so little attention has been directed to these two additional dimensions of interoperability. In the case of spatial support, we can speculate that the cause lies in the fact that the problem is much more apparent for continuous fields than for discrete objects. We hope that this paper will draw attention to the problem, and stimulate work on its other aspects, including its relationship to metadata, and to processes of data search.

Acknowledgment

The authors wish to acknowledge the support received for this research from the National Geospatial-Intelligence Agency.

References

- Cruz, I., Wiegand, N., Patterson, D., Zhou, N., and Ventura, S., 2002, Querying heterogeneous land use data: problems and potential, *Proceedings, dg.o 2002, Los Angeles, California*.
- Cruz, I., Wiegand, N., and Zhou, N., 2003, A Web query system for heterogeneous geospatial data, *Proceedings, SSDBM, July 2003*, pp. 262-265.
- Goodchild, M. F., 2002, Measurement-based GIS, in W. Shi, P.F. Fisher, and M.F. Goodchild, editors, *Spatial Data Quality*, Taylor and Francis, New York, pp. 5-17.
- Goodchild, M. F., 2004, A general framework for error analysis in measurement-based GIS, *Journal of Geographical Systems* 6(4): 323-324.
- Goodchild, M. F., Anselin, L., and Deichmann, U., 1993, A framework for the areal interpolation of socioeconomic data, *Environment and Planning A* 25: 383-397.
- Goodchild, M. F., Egenhofer, M. J., Fegeas, R., and Kottman, C., 1999, *Interoperating Geographic Information Systems*, Kluwer, Boston.
- Goodchild, M. F. and Lam, N.-S., 1980, Areal interpolation: a variant of the traditional spatial problem, *Geoprocessing* 1: 297-312.
- Goovaerts, P., 1997, *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York.
- INEGI, 2002, *Carta Edafológica Escala 1:250,000*, Instituto Nacional de Estadística, Geografía e Informática, Aguascalientes, México.
- INEGI, 2004, *Guía Para La Interpretación de Cartografía: Edafología*, Instituto Nacional de Estadística, Geografía e Informática, Aguascalientes, México.
- Kyriakidis, P. C., P. Schneider, and M. F. Goodchild, 2005, Fast geostatistical areal interpolation, *Proceedings, Geocomputation 2005, Ann Arbor, Michigan, August*.
- Longley, P. A., M. F. Goodchild, D. J. Maguire, and D. W. Rhind, 2005, *Geographic Information Systems and Science*, Second Edition, Wiley, New York.
- Mark, D. M. and Csillag F., 1989, The nature of boundaries on “area-class” maps, *Cartographica* 26(1): 65-78.
- Simmons, C. S., 1959, *Clasificación de Reconocimiento de los Suelos de la República de Guatemala*, Instituto Agropecuario Nacional, Guatemala.
- Smith, B. and Varzi, A. C., 2000, Fiat and *bona fide* boundaries, *Philosophy and Phenomenological Research* 60(2): 401-420.
- Tobler, W. R., 1979, Smooth pycnophylactic interpolation for geographic regions, *Journal of the American Statistical Association* 74: 519-536.
- Wright, A. C. S., Romney, D. H., et al., 1959, *Land in British Honduras: Report of the British Honduras Land Use Survey Team*, Her Majesty's Stationery Office, London.