Research Article

Singly – and Doubly – Constrained Methods of Areal Interpolation for Vector-based GIS

Richard D Mrozinski, Jr  Robert G Cromley
Department of Geography,  Department of Geography,  
University of Connecticut, USA  University of Connecticut, USA

Abstract
Traditionally, areal interpolation has referred to techniques for transferring attribute values from one partitioning of space to a different partition of space but this is only one of several situations that create the need for estimating unknown data values for areal units. This paper presents a categorization of four areal interpolation problems that includes the “missing” data problem, the traditional “alternative geography” problem, the overlay of a choropleth and an area-class data layer, and the overlay of two choropleth data layers and demonstrates the relationship between the last three problems and general spatial interaction modelling. The “alternative geography” and overlay of choropleth and area-class data layers mirrors a singly constrained spatial interaction model while the overlay of two choropleth layers is analogous to a doubly constrained interaction model. Iterative proportional fitting techniques with and without ancillary data are developed to solve these three classes of problems.

1 Introduction
There are numerous ways that the earth’s surface can be divided into non-overlapping areal units for the purpose of data measurement and tabulation. Given the different methods for partitioning space into regions and the number of agencies that report spatially aggregated data, it is possible that units used for data collection and reporting may not be the units most appropriate for subsequent spatial analysis. In these instances, data values must be estimated for the appropriate areal units from the available information. Areal interpolation refers to a set of techniques for transferring

Address for correspondence: Richard D Mrozinski, Jr, Department of Geography, U-2148, 354 Mansfield Road, Room 438, University of Connecticut, Storrs, CT 06269-2148, USA
attribute values from one partitioning of space to a different division of space (Goodchild and Lam 1980, Lam 1983, Flowerdew and Green 1989, Green 1990, Flowerdew et al 1991, Goodchild et al 1993). In this case, attribute values are known for a set of “source” zones and the interpolation procedure estimates this attribute value for the set of “target” zones.

Although areal interpolation methods were initially developed in response to the need in spatial analysis to estimate data for target zones, the use of GIS has also increased the range of situations that utilize areal interpolation procedures. Through operations such as polygon overlay and buffering, the creation of new polygons is inherent in many GIS applications. In overlay analysis, values for a given attribute may be present in two polygon layers, but not for the polygons that are formed by the overlay of the two layers. Areal interpolation is performed to calculate the non-spatial attribute values for these newly created regions. This form of interpolation differs from the data transfer type because the object of the interpolation is a newly created polygon layer having two sets of source zones. This paper reviews the spatial context for estimating the attribute values of areal units and presents several new methods for performing areal interpolation. The proposed methods are then tested against existing techniques for different spatial problems that require areal interpolation.

2 Background

The need for areal interpolation arises during the compilation stage in data collection, wherein individual objects are aggregated into areal units for further analysis. Areal interpolation would be unnecessary if all subsequent analyses used data for individual observational units rather than spatially aggregated data. Openshaw and Taylor (1981) listed several reasons why this option is not used more often. First, in the case of data collected for individual people, the widespread use of reporting data aggregated by geographical area avoids the problem of disclosing confidential information. Second, there are computational advantages in the analysis of data in an aggregated form. Aggregation greatly reduces, for example, the size of matrices used in origin-destination analysis, such as in transportation modelling. Finally, geography has a long tradition of studying associations among variables in regions for which zonal approximations already exist.

The modifiable areal unit problem in geography arises because there is no single zoning system appropriate to the study of all spatial processes (Openshaw and Taylor 1981). Because the reporting zones are not fixed and unique, individual data observations are frequently aggregated differently. Comparing zonal data sets is a problem because zone delineations change over time and different agencies use alternative geographies to collect their data. If the disaggregated data underlying these different geographies no longer exist or are not available because of confidentiality restrictions, then areal interpolation is used to redress the incompatibility of zonal systems by estimating zonal values for target zones based on the values for the source zones.

Additionally, areal interpolation arises from the need to estimate new data values for areal units formed by the overlay of map layers containing different sets of areal units. The overlay of two map layers requires a spatial intersection operator to
determine which areal units in one layer intersect with particular units in the other layer. The intersection of the two layers can be expressed as a matrix (Figure 1) in which the objects from Layer A (the source zones) are presented along the vertical axis and the objects from Layer B (the target zones) are presented along the horizontal axis. The intersection matrix will contain a cell value of one if the corresponding objects from the two layers intersect and a cell value of zero if they do not intersect. The new map layer (Layer C) contains a new set of areal objects with a spatial resolution different from before. The reconstruction of the least common geographic units – the objects defined as the areas left “uncut by any further partitioning” (Peucker and Chrisman 1975, 63) – requires an estimation of attribute values for the new units given the values of the former layers. These new units are also called resels (resolution elements), and for vector data are the equivalent to the pixel in the raster world (Tobler 1984).

The spatial attributes of the new objects, such as area and perimeter, are easily calculated within a GIS as measurement of spatial objects is a basic function of these systems (Tomlinson and Boyle 1981). Computation of the non-spatial characteristics of the new objects is not as straightforward. Because the overlay function inherently disaggregates the original data (Chrisman 1987), the assignment of non-spatial attributes can pose major problems. The initial values for the characteristics of each polygon are themselves a function of the aggregation process that was used to delineate the polygons in the original layers; any areal interpolation must take into account differences between the types of polygon delineations found in the original layers.

Geographic data can be dichotomized into field-based data and entity-based data (Worboys 1994). Field-based data view space as continuous and space-filling whereas entity-based data view space as being punctiform, empty except where entities are located. These two types of data can be represented using either the raster or vector data model. In vector representations of areal features as polygons, the map layers commonly found in geographic databases – choropleth and area-class – correspond to entity- and field-based data respectively.

For choropleth layers, the polygons are entities determined independently of the phenomenon being measured (Sinton 1978, Mark and Csillag 1989). This type of map layer is used for the display of socio-economic data, and the areal units serve as discrete reporting zones for such data. The type of non-spatial attribute data associated with choropleth mapping are usually interval or ratio data that are either spatially extensive (counts) or intensive (averages, ratios, and other normalized data) in nature (Goodchild and Lam 1980).

<table>
<thead>
<tr>
<th>Source Zones</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 1  The intersection matrix of source and target zones.
Area-class maps, on the other hand, have boundaries between classes of categories which occur over connected regions of geographic space (Mark and Csillag 1989). The polygons that comprise an area-class map are delineated by boundary lines that are estimates of where the character of the continuous phenomenon being measured changes. The primary use of these maps is for portraying naturally occurring phenomena. The most common type of non-spatial attribute data in these maps is categorical. In an area-class layer, there is only one initial non-spatial attribute – the characteristic used to determine the boundaries of the layer – whereas a choropleth layer may have many non-spatial attributes associated with it.

Whenever two area-class maps are overlaid, vector-based GIS software packages such as ARC/INFO attach the non-spatial attribute value of each source layer polygon directly to the polygon in the new layer formed by their intersection. These values are reasonable estimates of the non-spatial attributes for the new units if the original layers were homogenous; there is a related issue that delineations in the naturally occurring phenomenon resulting in crisp boundary lines do not exist and result in attribute errors (Chrisman 1987). In raster-based systems, these can be represented as fuzzy boundaries and fuzzy set rules can be applied in the overlay process (see Burrough and O’Donnell 1998, 280–2). In an overlay involving at least one choropleth layer, a GIS software package such as ARC/INFO also attaches the non-spatial attributes of the choropleth source layer(s) to the polygons formed by its(their) intersection(s). For count data, however, these estimates are completely spurious and should be estimated using areal interpolation rather than assignments.

Areal interpolation is used then to estimate unknown values associated with areal units for four different situations that can arise in the creation and manipulation of various map layers. First, the “missing data” problem (AI1) arises whenever one or more unknown values exist for areal units within the same source layer. AI1 differs from other areal interpolation problems in that it does not involve an intersection of two layers and data disaggregation. In this case, the total count for the full set of units is not known due to the missing data. This problem has been solved using cartographic methods based on spatial proximity (see Tobler and Kennedy 1985) and statistical methods using other attribute information to generate maximum likelihood estimates (Griffith et al 1989).

Next, the “alternative geography” problem (AI2) is the traditional form of areal interpolation discussed at the outset of this paper involving the transfer of data values collected for one geography of areal units to another. One desired property of AI2 is that it should preserve the volume of the discrete surface. The first volume-preserving method for solving AI2 was pycnophylactic interpolation (Tobler 1979), developed to create isopleth maps of population of the United States. Although this smoothing technique was originally proposed to fit a rasterized continuous surface to data for the purpose of isopleth mapping, it is also used to reaggregate areas of the interpolated surface into new target zones.

Another raster-based approach is to use the weighted centroid of an enumeration unit as the location of highest density within the enumeration unit. Using the cell containing that centroid location as the central cell, the total count associated with the unit is distributed over all cells within a window limit of the central cell based on a distance-decay probability (see Bracken and Martin 1989, Martin 1989, Bracken 1991, Martin and Bracken 1991). A given raster cell would receive partial count values from
all centroids within whose window it is located. While the purpose of this approach is to generate more local estimates of population distributions, the cells can be reaggregated into new target zones. This technique is really a point interpolation method and represents the attribute as a continuously distributed surface; however, spatially extensive counts are aggregations of discretely located individuals. Additionally, for vector-based GIS, these raster approaches have the drawback of not only a vector-to-raster conversion to implement the technique but then a raster-to-vector conversion of the final results.

A polygon-based method, developed for AI2, is areal weighting in which values are estimated as averages proportionally weighted by area of each resel (Goodchild and Lam 1980). The area weighting procedure’s one major weakness is it is founded on the assumption that the attribute being interpolated is uniformly distributed within each source zone. To overcome this drawback, Fisher and Langford (1995) suggest using a dasymetric mapping approach in which additional knowledge of the local environment is used to identify zones having different population densities although in a raster format.

Other techniques for solving AI2 use statistical techniques involving ancillary data collected for target zones (Flowerdew and Green 1989, 1994, Green 1990, Flowerdew et al 1991). Extending this research, Goodchild et al (1993) defined areas represented by ancillary data as “control zones.” One problem with using ancillary data for target zones is that these data may not exist for target zones associated with area-class map layers. A second problem is that within current GIS software packages, there is no direct method to incorporate the statistical analysis necessary for areal interpolation methods based on ancillary data. To incorporate statistics into areal interpolation, researchers have been forced to link the attribute files of an overlay to a computer statistical package (Flowerdew and Green 1989, Flowerdew et al 1991). Difficulties can be encountered during this process if the necessary links between GIS software and the appropriate statistical software are impractical or infeasible on certain hardware platforms (Flowerdew et al 1991, 314). Using Monte Carlo simulation, Fisher and Langford (1995) report that the statistical methods perform less satisfactorily than dasymetric-based areal interpolation.

The third form of areal interpolation (AI3) arises from the need to estimate the attribute values for a new layer of resels resulting from the overlay of a choropleth map layer and an area-class layer; this case is very similar to AI2 but the focus of attention is the resel formed by the intersection rather than the target zones of the second initial layer. Most approaches for solving the alternative geography problem can be applied to this problem with the restriction that ancillary non-spatial metric attributes, such as income and housing, probably do not exist for the area-class zones although categorical attributes such as land use do.

Finally, the fourth form of areal interpolation (AI4) is the estimation of the attribute values for the layer of polygon intersections resulting from the overlay of two choropleth layers (Table 1); in this case, the values may be known with respect to a non-spatial characteristic for the polygons of both source layers but not for the new polygons of the intersection layer. Volume must now be preserved with respect to two layers rather than one. No existing areal interpolation method has addressed this problem.
3 Methodology

AI2, AI3, and AI4 also share many characteristics with flow and transition models such as spatial interaction modelling (Wilson 1971, Senior 1979), multidimensional demography (Willekens 1982, 198), and voting patterns (Johnston and Hay 1983) that are extensions of multidimensional contingency table analysis. For example, in a doubly-constrained spatial interaction model, the count for each origin and destination zone is known but the count for each origin/destination pair is unknown and must be estimated; this is analogous to the areal interpolation problem in which two choropleth layers are overlaid and the counts for each zone in the ‘origin’ layer and ‘destination’ layer are known but the common portion between each ‘origin’/‘destination’ pair is unknown.

The doubly-constrained, two-dimensional log-linear model used in contingency table estimation in its multiplicative form is:

\[ X_{ij} = W_0 W_i W_j W_{ij} \]  

(1)

where \( X_{ij} \) is the estimate for cell \((i,j)\), \( W_0 \) is a proportionality constant (the total count), \( W_i \) and \( W_j \) are the first order effects (the row and column totals), and \( W_{ij} \) is a second-order iteration term. In a singly constrained model either the row or column effect is eliminated. Whereas distance between origin/destination zones is the basic attribute used for estimating the interaction term in spatial interaction modelling, area shared by origin/destination zones is the basic attribute used for estimating the interaction term in areal interpolation. Using the terminology of spatial interaction models, singly-constrained areal interpolation corresponds either to the alternative geography problem (AI2) or the overlay of a choropleth and an area-class layer problem (AI3) discussed in the previous section because only one source zone layer is constrained by a given set of data values. Doubly-constrained areal interpolation refers to the problem

<table>
<thead>
<tr>
<th>Category</th>
<th>Problem Definition</th>
<th>Types of Data Layers Involved</th>
<th>Focus of Inquiry</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI1</td>
<td>Missing Data</td>
<td>One Choropleth Layer</td>
<td>Estimating attribute value missing for certain Areal Objects in the layer</td>
</tr>
<tr>
<td>AI2</td>
<td>Alternate Geography</td>
<td>One Choropleth Layer and either a second Choropleth or an Area-Class Layer</td>
<td>Estimating attribute values for all Areal Objects in one layer given attribute values for different Areal Objects layer</td>
</tr>
<tr>
<td>AI3</td>
<td>Polygon Overlay I</td>
<td>One Choropleth Layer and one Area-Class Layer</td>
<td>Estimating attribute values for new areas formed by the intersection of the two layers</td>
</tr>
<tr>
<td>AI4</td>
<td>Polygon Overlay II</td>
<td>Two Choropleth Layers</td>
<td>Estimating attribute values for new areas formed by the intersection of the two layers</td>
</tr>
</tbody>
</table>
of overlaying two choropleth layers (AI4) because both source zones have respective sets of given values.

Spatial interaction and areal interpolation models can be formulated as a spatial relation matrix and be solved using contingency table analysis. However, unlike most forms of flow and transition analysis, areal interpolation involves estimations based on incomplete tables. Incomplete tables arise when there is a priori knowledge that a cell value should be zero as a naturally occurring feature of the data (Bishop et al 1975). In areal interpolation, the intersection matrix contains many zero elements that denote the infeasibility of two polygons intersecting (Figure 1). The presence of “structural zero” cells increases the informational content of the matrix but also necessitates adjustments be made in the maximum likelihood estimation procedures because the logarithm of zero is not defined.

Solution methodologies for making adjustments to a doubly-constrained matrix, used in procedures such as entropy-maximizing and spatial interaction modeling, are rooted in a procedure known by several names, such as maximum likelihood estimation, iterative proportional fitting, and biproportional matrix adjustments. In this paper, iterative proportional fitting (IPF) (Fienberg 1970) is the term referring to the procedure for making biproportional adjustments to a matrix. In the IPF procedure, a matrix is adjusted using the following set of equations adopted from Bishop et al (1975):

\[
m_{ij}^{\nu+1} = m_{ij}^{\nu} \frac{Q_i}{\sum_j m_{ij}^{\nu}}
\]

\[
m_{ij}^{\nu+2} = m_{ij}^{\nu+1} \frac{Q_j}{\sum_i m_{ij}^{\nu+1}}
\]

where \(m_{ij}^{\nu}\) is the matrix element in row \(i\), column \(j\), and iteration \(\nu\), \(Q_i\) is the predefined row sum, and, \(Q_j\) is the predefined column sum. This procedure will theoretically converge in iteration \(k\) when the following two conditions are met:

\[
\sum_j m_{ij}^k = Q_i
\]

\[
\sum_i m_{ij}^k = Q_j
\]

The properties of the IPF procedure have been studied by many researchers (Deming and Stephan 1940, Mosteller 1968, Bacharach 1970, Feinberg 1970, Bishop et al 1975, Macgill 1977). Macgill (1977) has shown that if there is a solution for the biproportional adjustments, this solution will be unique. Also, if a biproportional adjustment is a consistent adjustment, then a solution will exist. A matrix adjustment is consistent if the prespecified row and column totals can be satisfied by some matrix, with a pattern of zeros that at least cover the pattern of zeros in the previous matrix (Macgill 1977, 690). The final quality addressed by Macgill is convergence. An iterative procedure will converge if the adjustments are consistent. Convergence has been proven for several types of matrices, including the non-square and non-negative types of matrices which will be addressed in this research (see Macgill 1977 for proof of
convergence references). Iterative proportional fitting is easily adapted to doubly-constrained areal interpolation.

In adapting singly-constrained models for AI2 and AI3, a different fitting technique is necessary because only one set of marginal totals is available. The first singly-constrained method for solving AI2 and AI3 proposed here is based on polygon operations by extending pycnophylactic interpolation to create a smooth density function at the polygon (or resel) level instead of at the grid cell level, as was the case in Lam’s (1983) adaptation of Tobler’s original pycnophylactic method.

In the basic polygon smoothing approach, each intersection polygon is assigned a population density value that is equal to the population density value for the source zone within which the intersection polygon is located. Next, using the topological data structure of the source layer, a smoothing function is performed by averaging the density value of the intersection polygon with the density values of any intersection polygon that shares a common boundary with the one being smoothed. The volume-preserving constraint is then placed on the process to proportionally increase or decrease each intersection polygon value so the total population value within each source zone remains unchanged. This process is then repeated until convergence is reached. Next, for AI3 the data value for each intersection polygon is calculated by multiplying the interpolated density value of the intersection polygon by its area. Lastly, for AI2 the values for the intersection polygons that comprise a target zone will be summed to estimate the value for that particular target zone.

For the dasymetric polygon smoothing approach, the existing road network within each region is used as ancillary data. Road networks are readily available in the United States from the USGS Digital Line Graph files. Using the road network of the study area, control zones can be constructed, relying on the assumption that individuals in a human population will live within a certain set distance of a roadway. A set of new area zones are created using the buffering operation of a GIS within each source zone. All population values are transferred to the areas within the buffer polygons and the polygons outside the buffer polygons are assigned a value of zero. The iterative smoothing technique is then applied to the densities of these polygons.

In the doubly-constrained areal interpolation necessary for solving AI4, the first step is to assign the interaction term for each non-zero term in the intersection matrix as the area of the associated intersection polygon. Using these values, iterative proportional fitting is applied until convergence is obtained while the row and column constraints are maintained. Again, the dasymetric version is implemented using a buffered road network and making all intersection polygons representing areas outside the buffer polygons structural zeros in the matrix. The initial interaction term for each non-zero element is the area of the associated buffer polygon.

The main problem in the application of spatial interaction models to the problem of estimating values is the specification of the initial interaction terms. As discussed earlier, for flow models the interaction term is specified as a form of distance-decay which is also used in point interpolation problems (Bracken and Martin 1989 and Martin 1989 have also used distance decay for areal interpolation). The concept of interaction in an overlay of two polygon layers is most commonly modelled as the area of the common intersection. Fisher and Langford (1995) suggest that this interaction term can be improved by adding additional knowledge of the local environment as is done in converting a choropleth map to a dasymetric map. The choice of which data to use in the dasymetric approach is context-specific but the utility of GIS is the ability to
integrate many data layers for the same geographic domain; the GIS analyst can choose which data layers to use for the specific areal interpolation application. Buffered roads are suggested here for population estimates because of the availability of this information in digital form in the United States and the fact that most people dwell within a certain distance of roads.

4 Data and Analysis

These different singly and doubly constrained methods for solving AI2, AI3, and AI4 were used to estimate population values for a database geo-referenced to the state of Connecticut. For evaluation of these methods when solving AI2 and AI3, the results are compared against similar results from existing pycnophylactic, areal weighting and dasymetric areal weighting methods for the same database; no additional comparisons are made for AI4 because no existing method exists. The first source layer is a choropleth layer of the political geography of the 169 Connecticut towns (Figure 2); the towns had a combined population of 3,288,547 in 1990 (U.S. Bureau of the Census 1995). The second source layer is an area-class layer of 52 regional drainage basins in Connecticut (Figure 3a). However, for testing purposes, this layer was modified so that the boundaries were made congruent to the census geography at the block group level. By building the drainage basins as aggregates of block groups (Figure 3b), the population for this source level is also known. This aggregation would not be necessary if the purpose was to estimate the values for the 52 drainage basins and is done here purely for test purposes. The overlay of the two source layers resulted in 303 intersection polygons (Figure 4).

Figure 2  The choropleth layer of town boundaries in Connecticut.
A third map layer containing ancillary spatial information was the Roads and Trails Data Layer for the state of Connecticut. This map layer is comprised largely of primary highways, secondary highways, local roads, streets, and trails that appear on the 1:24,000-scale USGS quadrangle maps (DEP 1994). Since these data have their origin in the quadrangle maps, the temporal quality of each quadrangle is based on the update year, ranging from 1969 to 1984. This causes a temporal discrepancy between this layer (having, at best, a date of 1984) and the population data (dated 1990). The Roads and Trails Data Layer was modified from its original format to include only certain features, because certain features present in the original layer such as trails and footbridges are inappropriate for this analysis (Figure 5). The selected features were then buffered 100 feet on each side to create the control data layer for the dasymetric approaches.

For solving AI2, Connecticut towns serve as source zones and the approximation of the drainage basins serve as target zones. Summary statistics including minimum and maximum deviations and the root mean standard error for each method of singly-
**Figure 3b**  Block group approximation of regional drainage basins.

**Figure 4**  Overlay of source and target layers.
constrained areal interpolation are presented in Table 2. The root mean standard error ($E^{RMS}$) is calculated using the following equation:

$$E^{RMS} = \sqrt{\frac{\sum(\text{Interpolated Value} - \text{Actual Value})^2}{N}}$$

Each of the techniques, except the pycnophylactic, interpolated the value of at least one target zone perfectly. This is because source zone 102 and target zone 34 represent the same area, North Stonington, and thus, have the same population. The error in the pycnophylactic interpolation for this polygon is due to the vector-to-raster and raster-to-vector conversions that must be applied to source and target zones for this method. The largest absolute deviation for any target zone is associated with areal weighting (59,206) (Table 2); the dasymetric areal weighting and dasymetric polygon smoothing methods had much smaller maximum absolute deviations (42,325 and 39,758 respectively) than the other approaches. Similarly, the root mean square error for the dasymetric areal weighting and polygon smoothing, 1103 and 1096 respectively, was much smaller than that for the non-dasymetric methods (Table 2).

While the dasymetric areal weighting method was slightly worse than the dasymetric polygon smoothing method with respect to the AI2 problem, its was slightly better than the latter at the level of intersection polygons – the task for AI3 (see Tables 2 and 3). This outcome occurred because over- and under-predicted values were in close proximity to one another for the dasymetric polygon smoothing method (Figure 6a) and averaged each other out with respect to aggregation at the target zone scale. The other outcomes were similar to the trends for the AI2 estimates: the RMS errors for the dasymetric approaches were much less than the non-dasymetric ones.

For the AI4 problem, both the town and drainage basin layers were used as source layers and values were estimated for the individual intersection polygons. Similar
results occurred with respect to the doubly-constrained methods as with the singly-constrained methods. The dasymetric approach using buffer polygons had a maximum absolute deviation that was 3,000 less than that for the simple iterative proportional fitting of intersection polygons (Table 4). The overall RMS error for the dasymetric method was also 30 less (66 versus 96) than that for simple IPF.

Another measure of how well doubly-constrained areal interpolation performed is to compare the values estimated for the intersection polygons as an AI4 problem to the values estimated for the same intersection polygons as an AI3 problem. This gives a measure of how much extra value is added by knowing the totals for both layers. As might be expected, no AI3 method performed as well as either AI4 method. Even the dasymetric methods for AI3 had RMS values almost twice as high as the worst AI4 method. A comparison of the spatial distribution of value deviations showed similar spatial arrangements (Figures 6a, 6b, and 6c) but fewer large deviations as one progressed from the singly-constrained dasymetric polygon smoothing algorithm (Figure 6a) to the doubly-constrained interpolation (Figure 6b) to the doubly-constrained interpolation using dasymetric data (Figure 6c).

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Summary statistics of the AI2 estimates for target zones.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Area Weighting</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>59,206</td>
</tr>
<tr>
<td>Standard Error</td>
<td>1,595</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Summary statistics of the AI3 estimates for polygon intersections.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Area Weighting</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>45,253</td>
</tr>
<tr>
<td>Standard Error</td>
<td>249</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Summary statistics of the AI4 estimates for polygon intersections.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Doubly-constrained interpolation</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>12,520</td>
</tr>
<tr>
<td>Standard Error</td>
<td>96</td>
</tr>
</tbody>
</table>
**Figure 6a** Percent deviation associated with the singly-constrained interpolation of polygon intersections using dasymetric data.

**Figure 6b** Percent deviation associated with the doubly-constrained interpolation of polygon intersections.
5 Conclusions

There is an important need for effective areal interpolation techniques, given the number of different spatial aggregations for which data are available. With the growing use of geographic information systems, data can be known for two data layers that are overlaid, but not for the resulting intersection polygons or data is known for a source layer but not for alternative geographies of the same region. This paper has linked the problem of areal interpolation to the different circumstances (AI1, AI2, AI3, and AI4) under which estimates for unknown values are needed. It has also put areal interpolation into the context of other spatial estimating procedures such as spatial interaction modelling and attempted to transfer solution methodologies based for singly- and doubly-constrained spatial interaction to that of singly- and doubly-constrained areal interpolation.

The singly-constrained methods developed here performed as well as existing methods for the AI2 problem. Using the buffered roads as dasymetric mapping information for both singly- and doubly-constrained models improved the estimates across all forms of areal interpolation (i.e. AI2, AI3, and AI4). Finally, the same techniques produced better results for estimating AI2 target zone values than estimating AI3 values for intersection polygons. This seems to suggest that the under-estimations and over-estimations of polygons are averaged when added to obtain values for target zones.
References


Bracken I and Martin D 1989 The generation of spatial population distributions from census centroid data. Environment and Planning A 21: 537–43


Deming W E and Stephan F F 1940 On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Annals of Mathematical Statistics 6: 427–44

DEP 1994 Data Dictionary for Roads and Trails Data Layer. Hartford, CT, Department of Environmental Protection


Mark D and Csillag F 1989 The nature of boundaries on area-class maps. Cartographica 26: 65–78

Martin D 1989 Mapping population data from zone centroid locations. Transactions of the Institute of British Geographers 14: 90–7


© Blackwell Publishers Ltd. 1999


Tobler W R 1984 Application of image processing techniques to map processing. In *Proceedings of the International Symposium on Spatial Data Handling*. Zurich, Geographisches Institut: 140–4


Tomlinson R F and Boyle A R 1981 The state of development of systems for handling natural resources inventory data. *Cartographica* 18: 65–95


