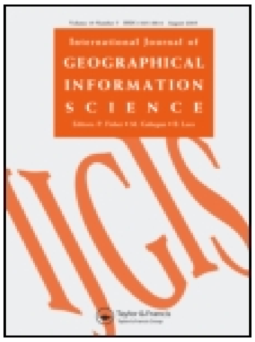


International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: http://www.tandfonline.com/loi/tgis20

# Evaluating representation and scale error in the maximal covering location problem using GIS and intelligent areal interpolation

Robert G. Cromley , Jie Lin & David A. Merwin

**To cite this article:** Robert G. Cromley , Jie Lin & David A. Merwin (2012) Evaluating representation and scale error in the maximal covering location problem using GIS and intelligent areal interpolation, International Journal of Geographical Information Science, 26:3, 495-517, DOI: 10.1080/13658816.2011.596840

**To link to this article:** https://doi.org/10.1080/13658816.2011.596840

Published online: 24 Oct 2011.

*International Journal of Geographical Information Science*
Vol. 26, No. 3, March 2012, 495–517

Taylor & Francis
Taylor & Francis Group

# Evaluating representation and scale error in the maximal covering location problem using GIS and intelligent areal interpolation

Robert G. Cromley[a]*, Jie Lin[a] and David A. Merwin[b]

[a]*Department of Geography, University of Connecticut, Storrs, CT, USA;* [b]*Department of Geography, Framingham State University, Framingham, MA, USA*

(*Received 30 September 2010; final version received 10 June 2011*)

A common problem in location-allocation modeling is the error associated with the representation and scale of demand. Numerous researchers have investigated aggregation errors associated with using different scaled data, and more recently, error associated with the geographic representation of model objects has also been studied. For covering problems, the validity of using polygon centroid representations of demand has been questioned by researchers, but the alternative has been to assume that demand is uniformly distributed within areal units. The spatial heterogeneity of demand within areal units thus has been modeled using one of two extremes – demand is completely concentrated at one location or demand is uniformly distributed. This article proposes using intelligent areal interpolation and geographic information systems to model the spatial heterogeneity of demand within spatial units when solving the maximal covering location problem. The results are compared against representations that assume demand is either concentrated at centroids or uniformly distributed. Using measures of scale and representation error, preliminary results from the test study indicate that for smaller scale data, representation has a substantial impact on model error whereas at larger scales, model error is not that different for the alternative representations of the distribution of demand within areal units.

**Keywords:** representation error; aggregation error; areal interpolation; maximal covering location problem

## Introduction

Over time, there has been a continuing confluence between location-allocation modeling and geographic information systems (GIS). From a GIS perspective, this integration has occurred because optimization models form the basis of many decision support systems (Malczewski 1999). GIS has been used in three broad application areas: (1) as an information database, (2) as an analytical tool, and (3) as a decision support system (Eastman *et al*. 1995). The use of GIS as a decision support system involves not only integrating spatially referenced data in a problem-solving environment but also integrating different types of spatial analysis functions with the GIS (Armstrong and Densham 1990). From a location-allocation perspective, the information database capabilities and spatial analytical capabilities of GIS can be used to collect and organize spatial data for the application of a location model, and GIS can be used to visualize results (Church 2002). An important

*Corresponding author. Email: robert.cromley@uconn.edu

ISSN 1365-8816 print/ISSN 1362-3087 online

http://dx.doi.org/10.1080/13658816.2011.596840
http://www.tandfonline.com

aspect of the impact of GIS on location-allocation modeling has also been the improved representation of the spatial objects under investigation (Miller 1996). The purpose here is to investigate the significance of improved demand representation within the context of the maximal covering location problem (MCLP) using GIS and intelligent areal interpolation procedures. The MCLP, developed by Church and ReVelle (1974), is used to locate a given number of facilities or service centers such that the maximum amount of demand is contained within a fixed coverage distance of an open facility.

The effect of scale and representation in location modeling has been an important area of investigation in spatial analysis (see Hillsman and Rhoda 1978, Goodchild 1979, Casillas 1987, Current and Schilling 1987, Daskin *et al*. 1989, 1990, Fotheringham *et al*. 1995, Murray and Gottsegen 1997, Erkut and Bozkaya 1999, Francis *et al.* 1999, Francis, Lowe and Tamir 2000, Plastria 2001, Murray and O'Kelly 2002). Much of the research on aggregation in location-allocation modeling has focused on methods for reducing errors associated with the aggregation process (Current and Schilling 1987, Hodgson and Neuman 1993, Francis *et al*. 1996, Bowerman *et al*. 1999, Zhao and Batta 1999, 2000, Plastria 2001) but other researchers have specifically addressed representation error especially with respect to covering problems (Murray and O'Kelly 2002, Murray and Tong 2007, Murray *et al.* 2008, Tong and Murray 2009, Alexandris and Giannikos 2010).

Although GIS technology permits a greater realism in data representations and a greater range in the scale of analysis, some assumptions regarding the spatial distribution of demand for services may still be limiting. Early models assumed that demand aggregated over areas was instead concentrated at points whereas recent models assume that demand is uniformly distributed within local regions. Intelligent areal interpolation methods are used here to provide better estimates of the spatial heterogeneity demand within areal units, and in the process evaluate errors associated with the scale and representation of demand data. Sections 'Representation and Aggregation Error in Covering Models' and 'Using Areal Interpolation in Covering Models' review representation and aggregation errors in covering models and how areal interpolation methods can be integrated with these models. Then a test design and data processing methods for implementing intelligent areal interpolation are discussed, and the results from this approach are evaluated against alternative representation methods.

## Representation and aggregation error in covering models

The fact that spatial analyses conducted at different scales (the *scaling effect*) or using different areal aggregation units (the *zonal effect*) produce different results is known as the modifiable areal unit problem (Openshaw and Taylor 1981). In the context of location-allocation modeling, the modifiable areal unit problem is manifested in aggregation and representation errors. Aggregation error occurs in any analysis conducted above the level of the individual or whenever a scale change occurs – the scaling effect. Whenever individuals are aggregated together, the aggregated demand has been spatially represented by either a demand centroid (a point) or a demand polygon within which the demand is located. Using a certain type of representation for the spatial units under investigation can also result in representation error. Murray and O'Kelly (2002) have noted that although they are related aggregation error differs from representation error. Francis *et al.* (2009) also note that the size of aggregation effects depends on model structure such that at the same aggregation level, some models can have more error than others. In the case of a centroid representation, all individuals are assumed to be concentrated at the centroid; in the case of a polygon the individuals are assumed to be either uniformly distributed within the polygon or not. Initially, the p-median problem (PMP) (ReVelle and Swain (1970), the location set-covering problem (LSCP) (Toregas *et al.* 1971), and the MCLP (Church and Revelle 1974) were developed using a centroid (or point) representation.

The effect of error due to aggregation and the geographic representation of objects was first analyzed by Hillsman and Rhoda (1978) in the context of the PMP. Using a point representation of demand, they identified three basic types of error. Source A error occurs whenever demand is not located at the centroid and results in either an overestimation or an underestimation of travel. Source B error is a special case of Source A error that occurs whenever a service is located at a centroid resulting in a zero average distance and an underestimation of travel cost for the individuals assumed to be located at the centroid. Source B error occurs frequently because service sites often are chosen from the same set of points as the centroids. Source C error occurs when the centroid is allocated to its nearest service site but more disaggregated demand points are closer to other service sites.

Current and Schilling (1990) extended these error concepts based on point representations to an analysis of the LSCP and the MCLP. Because covering models do not allocate demand to a specific service location, Source C error does not exist for covering models. For covering models Source A and B error corresponds to either demand not being covered by a service site when it actually is (underestimation) or demand being covered by a service when it is not (overestimation). In Figure 1, demand within the polygon represented by centroid A is considered covered because A lies within the coverage area of a service site denoted by a star; however, a portion of the polygon lies outside the shaded coverage area. Conversely in Figure 1, the demand polygons represented by centroids B and C are

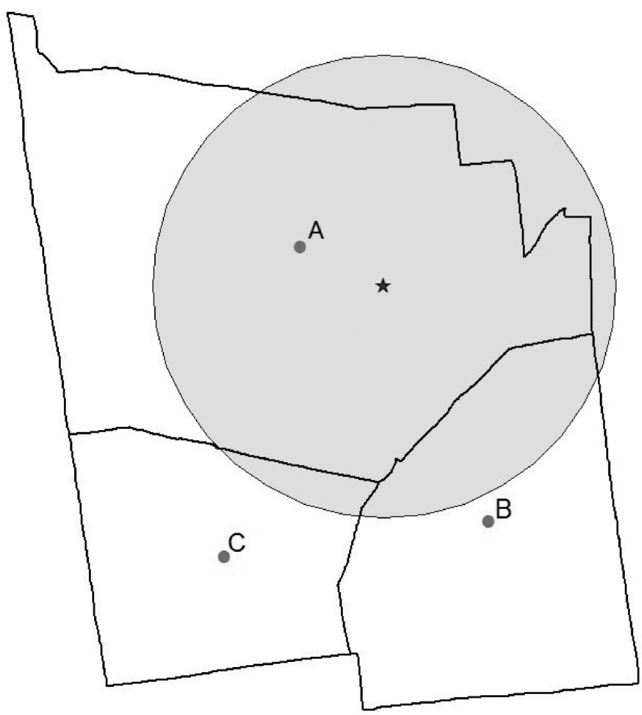


Figure 1. Source a error in the maximal covering problem. Polygon A is considered covered even though a portion of it is not and polygons B and C are considered not covered even though a portion of them are covered.

considered not covered because these centroids lie outside the shaded coverage area but a portion of each polygon does lie within the coverage area.

With respect to covering models, Source A errors can result in inaccurate estimations of the objective function due to an incorrect estimation of the total coverage as well as an incorrect location of facilities. For covering models, Casillas (1987) classified the difference in objective functions as *cost estimate error* (or simply *cost error*) – the difference in function values due to inaccurate estimation of demand arising from its aggregation and *optimality error* – the difference in function values due to incorrect locations. Later *location error* was defined as the geographic difference between service locations based on aggregate demand versus those based on more disaggregated demand (Erkut and Bozkaya 1999). Without fully disaggregated demand, these error measures are only approximated by comparing results for demand at one scale of analysis with results for the same study area with demand at a larger scale of analysis.

The same concern for the validity of results due to aggregation error also exists with respect to representation error. Initially the centroid representation for demand was the standard but the advent of GIS technology and the increased availability of geo-referenced data have enabled better model representations (Miller 1996, Church 2002). With respect to the LSCP, Murray and O'Kelly (2002) have found that point representations tend to overestimate the actual coverage of regional demand generally associated with a service configuration. This occurs because all centroids must be covered in the LSCP and some portions of the associated polygon(s) lie outside the coverage zone. Murray *et al*. (2008) later used square tessellations and triangle tessellations of regional demand to find the optimal configuration of sites when siting occurs over continuous space.

These regional demand representations have also been used to solve the MCLP for continuous space siting (Murray and Tong 2007). Later, Tong and Murray (2009) reformulated the MCLP as the PMP-multi-facility coverage, an extension of PMP to take into account partial area coverage rather than using the all or nothing approach associated with a centroid representation. In both of these papers, regional demand is assumed to be uniformly distributed throughout the entire study region (Murray and Tong 2007) or within individual demand polygons (Tong and Murray 2009). Whenever an areal representation of demand is used in coverage problems, the original demand layer will be partitioned into areas covered and/or not covered by a service configuration. In this situation, the level of demand must be estimated for the covered portion but the assumption of uniformly distributed demand is not realistic for this estimation procedure. Erkut and Bozkaya (1999) consider the assumption of uniform demand to be an analyst-induced error in evaluating the impact of aggregation errors on the p-median model. They warn that the numerical results in aggregation studies may be inflated due to the choices and/or assumptions made by the researcher. Their criticism also applies to evaluating the impact of representation errors on covering models. Spaulding and Cromley (2007) addressed representation issues in a related covering problem – the maximal capture problem – by allocating demand to road buffer zones. It is proposed here that intelligent areal interpolation be used to model the spatial heterogeneity of demand within areal units for the MCLP when estimating the amount of demand actually covered by facilities within a study region.

## Using areal interpolation in covering models

Areal interpolation refers to a set of techniques used to estimate unknown data values associated with areal units – most commonly population estimates (Goodchild and Lam 1980). Most frequently, it is used to solve the 'alternative geography' problem which involves the transfer of data values that are collected for areal units in one partition of geographic space (referred to as the source layer) to those of areal units that partition the same geographic space in a different manner (referred to as the target layer). In the case of covering problems, the source layer consists of the original demand polygons at some geographic scale (Figure 2a). The target layer corresponds to the intersection areas of the overlay of coverage zones associated with potential service sites (Figure 2b). Once demand is transferred to these coverage areas, they form the basic demand units. These units are referred to here as the least common demand coverage units (LCDCUs) using the terminology of Peucker and Chrisman (1975) with respect to units created from the overlay of successive layers. Each LCDCU is covered by a unique set of service sites. The transfer of demand from the source layer to the target layer also requires the intersection of the source and target layers (Figure 2c). Demand is estimated at the level of these smaller units and then reaggregated to the larger LCDCUs of the target layer. In Figure 2c, any portion of an LCDCU that lies outside a demand region is eliminated because it would contain no demand. If any portion of an LCDCU is covered by a potential site, then all of the LCDCU are covered. Source A error can still occur if the estimated demand associated with an LCDCU is incorrect. If LCDCU demand is overestimated, then more demand is said to be covered than actually is, and if the demand is underestimated, the reverse holds. However, some of the Source A error cancel out in the final solution because the LCDCUs themselves are aggregated to estimate the total demand covered by all open service sites, though the need is to estimate demand as accurately as possible for each LCDCU.

One of the first areal interpolation methods was Tobler's (1979) pycnophylactic (or volume-preserving) method that transformed a polygonal prism surface to a smooth, rasterized, continuous surface. Because the volume of the isoplethic surface had the same volume as the polygonal prism, the raster units could be reaggregated into target zones different from the original source polygons (Tobler 1979). Another early method is areal weighting (Goodchild and Lam 1980) in which estimates are determined by proportionally weighting source polygon values by the area of each intersection polygon to the area of the source polygon that contains it. This method is used implicitly by Tong and Murray (2009) to estimate the demand in their PMP-multi-facility coverage reformulation of the MCLP. The major weakness of areal weighting is the assumption that the attribute being interpolated is uniformly distributed within each source zone and it performs poorly in most

(a) (b) (c)

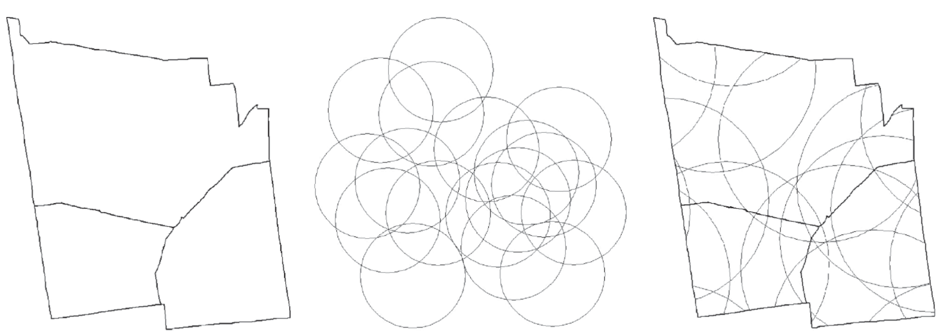

Figure 2. Construction of least common demand coverage units. (a) Source demand areas, (b) target coverage areas, and (c) intersection forming least common demand coverage units.

evaluations of areal interpolation techniques used to estimate population density (Fisher and Langford 1995, Langford 2006).

Methods such as Tobler's pycnophylactic method and areal weighting are referred to as *simple* areal interpolation methods because they do not use any ancillary data whereas *intelligent* areal interpolation methods use some form of ancillary data that correlate to the attribute being estimated (Langford *et al*. 1991, Langford 2006). The latter category of interpolators is based on the principles of dasymetric mapping (Wright 1936) in which ancillary data are used to identify zones having different population densities. Within intelligent methods, some are logical extensions of Wright's (1936) original work (see Langford *et al.* 1991, Fisher and Langford 1996, Eicher and Brewer 2001, Mennis 2003, Holt *et al.* 2004, Langford 2006, Mennis and Hultgren 2007). Other more computationally intensive techniques are based on some type of statistical approach (see Flowerdew and Green 1989, Flowerdew *et al*. 1991, Goodchild *et al.* 1993, Flowerdew and Green 1994, Bloom *et al*. 1996, Mugglin and Carlin 1998, Mugglin *et al*. 1999, Reibel and Agrawal 2007). Among potential ancillary data sets, remotely sensed land cover categories are probably the most commonly used correlates (Langford *et al.* 1991, Eicher and Brewer 2001, Mennis 2003, Holt *et al.* 2004), although cadastre databases have been used in urban areas (Maantay *et al*. 2007).

Intelligent areal interpolation is used most often to improve cartographic displays of aggregated data (see Eicher and Brewer 2001, Mennis 2003, Holt *et al*. 2004). The dasymetric map is recommended as an alternative to the choropleth map because one of the choropleth map's limiting features is the assumption that data are uniformly distributed within each areal unit whereas dasymetric mapping portrays the spatial variation within units. This logic of including the variation within spatial areal units is applied now to evaluate the representation and solution of MCLPs.

## Model and test design and data processing

The model formulation for the maximal covering problem used in this analysis is the version that minimizes demand not covered (Church and ReVelle 1974):

$$\text{Minimize} \sum_{i \in I} a_i y_i \tag{1}$$

subject to:

$$\sum_{j \in N_i} x_i + y_i \geq 1 \text{ for all } i \in I \tag{2}$$

$$\sum x_j = p \tag{3}$$

$$x_j = \ (0,\ 1) \ \text{ for all } \ j \in J \tag{4}$$

$$y_j = \ (0,\ 1) \ \text{ for all } \ i \in I \tag{5}$$

where $I$ is the set of demand units; $J$ is the set of potential service sites; $y_i$ is a decision variable equal to 1 if the $i$th demand unit is not covered, and 0 if the $i$th demand unit is covered within the maximal service distance $S_i$; $a_i$ is the estimated population to be served in the $i$th demand unit; $x_j$ is a decision variable equal to 1 if a service is located at the $j$th site, and 0 otherwise; $N_i$ is the set of service sites for the $i$th demand unit that lie within its critical distance $S_i$; and $p$ is the number of service sites to be located.

The objective is to minimize the amount of uncovered demand given the location of $p$ service sites. In this formulation, the portion of the constraint set (Equation (2)) associated with $N_i$ corresponds to the coverage portion of the MCLP matrix. For each constraint either a demand unit is covered by an open service site or the value of $y_i$ is equal to 1.

The study area for the test design is the 29 town region of Hartford County, Connecticut. The source demand is the 2000 census population of the county. To examine the effects of both aggregation and representation on results, three different data scales, three different coverage distances, and three different representations are used. Population data were acquired at the town, census tract, and block groups levels from the US Census website (www.census.gov). The study region had 222 census tracts and 666 block groups in the 2000 census. The boundary files for towns, tracts, and block groups were also downloaded from the census website. The demand at each scale is represented either by area centroids, by LCDCUs in which demand is assumed to be uniformly distributed, or by LCDCUs in which the spatial distribution of demand correlates to ancillary data. Remotely sensed land cover data for the year 2002 at a $30 \times 30$ m spatial resolution for the state of Connecticut were used as the ancillary data. These data were downloaded from the Center for Land Use Education and Research website (clear.uconn.edu). Temporally, these were the closest remotely sensed data for the study region with respect to the date of the population data.

For the test runs, 57 potential service sites were selected along major transportation arteries rather than from the set of demand centroids because as previously discussed choosing them from centroids is another form of analyst-induced error (Erkut and Bozkaya 1999). The number of service sites was arbitrary but enough sites were included to generate a multitude of different combinations of $p$ sites as well as providing complete coverage for certain representations. Three different coverage distances are used to evaluate the sensitivity of results with respect to coverage area. A 2-mile, 3-mile, and 4-mile Euclidean distance from each service defines the limit of coverage for each site. It is expected that as individual coverage areas increase in size less error would occur because for the same number of facilities being located, more demand units would be completely covered within any single coverage. Okabe and Sadahiro (1997) have shown that even for simple areal interpolation methods, smaller estimation errors are found whenever the diameters of source zones are much less than the size of target zones.

For the point representation of demand, demand units are centroids and the coverage portion of the constraint matrix can be found by calculating first a distance matrix between each service site and the demand centroids and then converting these distances into a zero/one matrix based on the critical distance. For this representation and method, the number of constraints in the MCLP is a function of the number of demand centroids that potentially could be covered. For the 2-mile coverage area at the town scale, only 24 demand centroids at most could be covered while at the tract level, only 195 centroids potentially could be covered and at the block group level, only 597 centroids could be covered (see Figure 3a). For the 3-mile coverage area at the town scale, all 29 demand centroids could be covered by at least one potential service site while at the tract level, only 219 centroids potentially could be covered and at the block group level, only 656 centroids could be covered (see Figure 3b). Finally, for the 4-mile coverage area all centroids at the town and tract scales could be potentially covered but at the block group level only 665

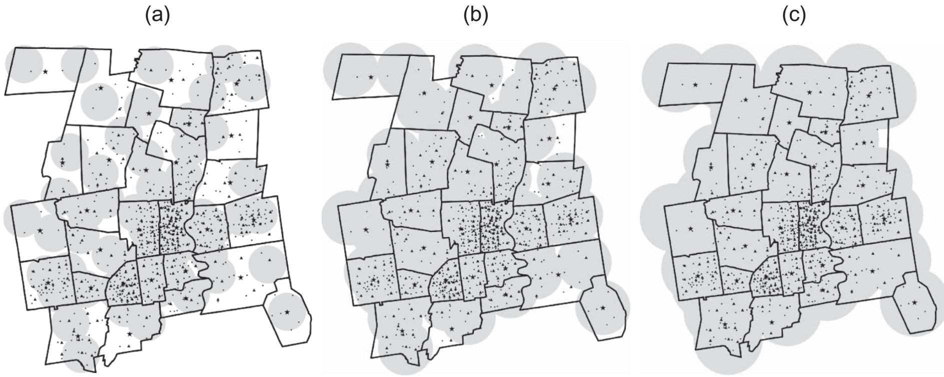


Figure 3. Location of different scale centroids with respect to the coverage distance. Town centroids are stars; tract centroids are triangles; block group centroids are dots; and the potential coverage area is shaded gray: (a) 2-mile coverage distance. (b) 3-mile coverage distance. (c) 4-mile coverage distance.

centroids could be covered (see Figure 3c). The $a_i$ values for each scale are the respective town, tract, and block group population associated with each demand centroid that could potentially be covered.

For the demand polygon representation, the demand units are the LCDCUs created by the overlay of the 57 coverage areas intersected against the demand layer of Hartford County. Not all of Hartford County could be covered by the LCDCUs for any of the three coverage distances (see Figure 4). Before the overlay of service coverage areas for a given distance is performed, each individual coverage area is given an attribute with a value of 1 designating the service site associated with the coverage area. The overlaying of the 57 coverage polygons using the UNION operation in ARCGIS 9.3 (ESRI, Redlands, California, USA) for a given distance will then build a zero/one attribute table for the LCDCUs, associating each LCDCU with the service sites that can cover it at the given distance. This resulted in 205 LCDCUs for a 2-mile coverage, 461 LCDCUs for a 3-mile coverage, and 787 LCDCUs for a 4-mile coverage. The coverage portion of the constraint matrix for the polygon representation is then extracted from the attribute table of the LCDCU layer.

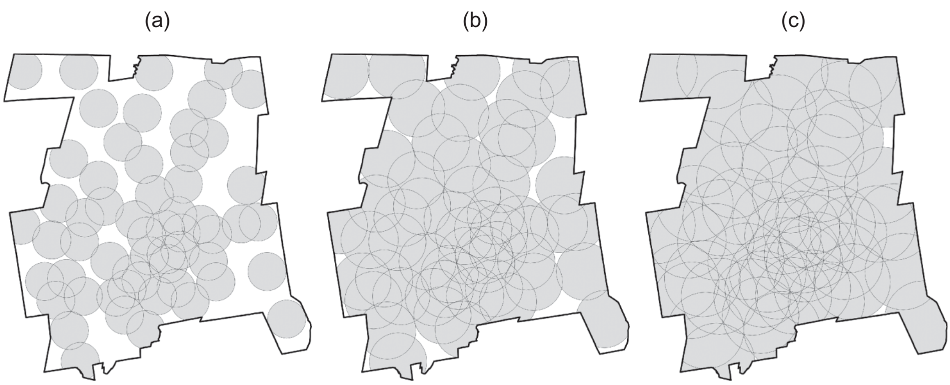


Figure 4. Location of the LCDCUs assuming uniform population density: (a) 2-mile coverage distance. (b) 3-mile coverage distance. c) 4-mile coverage distance.

Under the assumption of uniformly distributed demand, simple areal weighting is used to estimate each $a_i$ value and a binary dasymetric interpolation is used to estimate the $a_i$ values when assuming that the spatial heterogeneity of demand correlates to the ancillary data. Binary dasymetric interpolation has been found to produce estimates more accurately than more sophisticated statistical methods, such as regression analysis that uses global parameters in population estimation, because it is a local interpolator in which population densities differ by source zones (Fisher and Langford 1995, Langford 2006) and it produces more accurate estimates than areal weighting. It can also be easily implemented in a GIS environment. Given that the spatial heterogeneity of the dasymetric demand representation lies between the extremes of the centroid and the uniformly distributed demand representations, it is expected that the coverage values for the dasymetric representation will be lower than the centroid coverage values and higher than the uniformly distributed demand coverage values on average.

The land cover for Hartford County was first extracted from the state land cover image using ARCGIS 9.3 spatial analyst's masking tool. The 560,512 cells of the developed land category were extracted next as the final ancillary data layer (Figure 5). Of the 12 land cover categories in the original Center for Land Use Education and Research image, this category is most closely associated with population (the other categories were turf and grass, other grasses, agricultural field, deciduous forest, coniferous forest, water, nonforested wetland, forested wetland, tidal wetland, barren, and utility rights-of-way). The spatial distribution of the developed land category refines the distribution of demand from being located anywhere to more probable locations. Next, the raster image was converted into a point data layer. The number of points contained within each demand polygon was determined by performing a point-in-polygon spatial join in ARCGIS 9.3 between the point layer and

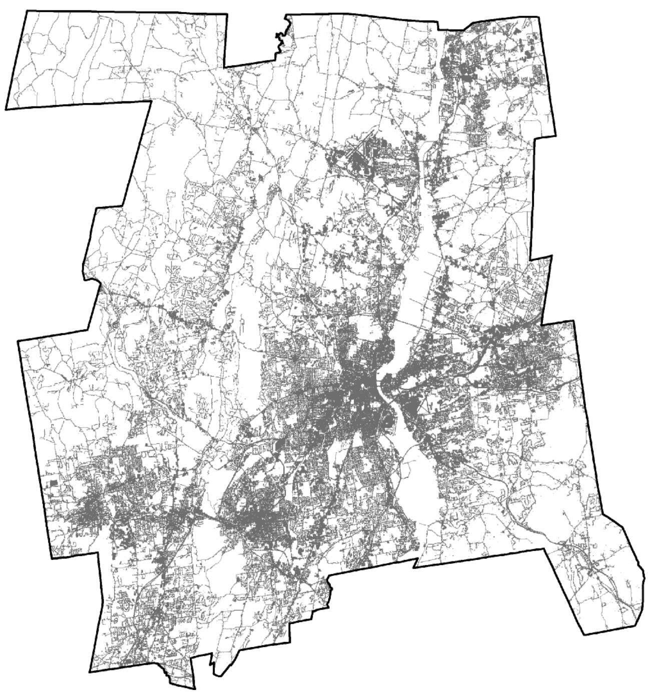

Figure 5. Spatial distribution of developed land.

the demand source layer to attach the source polygon identification to each point. Then a summary count was determined on the source polygon field. The spatial join and summary count steps were performed for the three scales of town, tract, and block group. The population density of each point for each scale was computed by dividing the population of the demand source unit at that scale containing it by the number of points in that demand unit. Next the point-in-polygon spatial join was performed between the point layer and the LCDCU target layer to attach the LCDCU identification to each point. Finally, the $a_i$ values associated with each LCDCP at each scale were computed by summarizing the population density for each scale by LCDCU. Whenever an $a_i$ value equals 0 for an LCDCU, that unit is dropped from the model lowering the number of demand constraints in the final model. Under the assumption of uniformly distributed demand no LCDCU could be dropped. Thus, although there were 205, 461, and 787 demand constraints for the 2-mile, 3-mile, and 4-mile coverage distance, respectively, for a uniform demand representation, there were 196, 439, and 727 demand constraints, respectively, for the same coverage distances with respect to the dasymetric demand representation.

Comparing representations, the number of constraints for the demand polygon representation is invariant with respect to the scale of the source units but differs with the size of the coverage area whereas the reverse is true for the centroid representation. However, the number of constraints for the centroid representation can be greatly reduced by selecting and retrieving only the LCDCUs that contain at least one centroid. This step geometrically performs the same row reductions for the MCLP that Toregas and ReVelle (1973) had earlier outlined as algebraic operations. The $a_i$ value for each retained LCDCU is the sum of the $a_i$ values for the centroids contained in each LCDCU. This selection did not reduce the number of constraints at the town level because there was at most only one centroid in any LCDCU but the centroid representation at the tract level would only need 98 constraints for a 2-mile coverage, 128 constraints for a 3-mile coverage, and 149 constraints for a 4-mile coverage. Similarly, the centroid representation at the block group level would only need 141 constraints for a 2-mile coverage, 225 constraints for a 3-mile coverage, and 292 constraints for a 4-mile coverage. Regardless of the coverage distance, a centroid representation can be solved with no more and potentially far less constraints than either the uniform or dasymetric demand representations. The remaining question is what level of accuracy does the centroid representation have in comparison with the alternative representations.

It was next determined how many sites are necessary for maximum coverage with respect to each representation, scale, and coverage distance because (1) not every combination of representation and scale can achieve 100% coverage of total demand and (2) this number of facilities is the upper bound on $p$ in the MCLP for each scale and representation. The maximum coverage results are given in Table 1. The number of facilities necessary for maximum coverage was invariant with respect to the scale of the source demand units for the LCDCU representations whereas the number increased for the centroid representation as the source scale became larger (see Table 1). As Murray and O'Kelly (2002) also found, the demand centroid representation needed far fewer sites for demand to be 'covered' than the areal demand representations although the number increased as the scale became larger. Among the two areal representations, though, more sites are necessary for the assumption of uniform demand than are necessary for dasymetric demand because some sites are covering LCDCUs with zero population.

With respect to the size of individual coverage zones, the number of facilities necessary for maximum coverage decreased as the size of the coverage area increased. At the same time, the percent of total demand covered increased as the size of the coverage area

Table 1. Maximum coverage for different representation models.

| Representation/source scale/buffer distance | Number of demand constraint | Number of sites necessary for maximum coverage | Maximum coverage* as a percent of total demand (%) |
|---|---|---|---|
| Centroid/town level/2 miles | 24 | 24 | 91.2 |
| Centroid/town level/3 miles | 29 | 22 | 100 |
| Centroid/town level/4 miles | 29 | 17 | 100 |
| Centroid/tract level/2 miles | 98 | 49 | 86.9 |
| Centroid/tract level/3 miles | 128 | 32 | 98.5 |
| Centroid/tract level/4 miles | 149 | 22 | 100 |
| Centroid/BG level/2 miles | 141 | 53 | 87.2 |
| Centroid/BG level/3 miles | 225 | 38 | 98.4 |
| Centroid/BG Level/4 miles | 292 | 27 | 99.7 |
| Uniform LCDCU/town level/2 miles | 205 | 55 | 81.7 |
| Uniform LCDCU/town level/3 miles | 461 | 45 | 97.1 |
| Uniform LCDCU/town level/4 miles | 787 | 32 | 99.8 |
| Uniform LCDCU/tract level/2 miles | 205 | 55 | 87.8 |
| Uniform LCDCU/tract level/3 miles | 404 | 45 | 97.9 |
| Uniform LCDCU/tract level/4 miles | 787 | 32 | 99.8 |
| Uniform LCDCU/BG level/2 miles | 205 | 55 | 88.1 |
| Uniform LCDCU/BG level/3 miles | 404 | 45 | 98.0 |
| Uniform LCDCU/BG level/4 miles | 787 | 32 | 99.8 |
| Dasymetric LCDCU/town level/2 miles | 196 | 55 | 86.7 |
| Dasymetric LCDCU/town level/3 miles | 439 | 44 | 98.0 |
| Dasymetric LCDCU/town level/4 miles | 727 | 31 | 99.9 |
| Dasymetric LCDCU/tract level/2 miles | 196 | 55 | 87.8 |
| Dasymetric LCDCU/tract level/3 miles | 439 | 44 | 98.2 |
| Dasymetric LCDCU/tract level/4 miles | 727 | 31 | 99.9 |
| Dasymetric LCDCU/BG level/2 miles | 196 | 55 | 88.1 |
| Dasymetric LCDCU/BG level/3 miles | 439 | 44 | 98.2 |
| Dasymetric LCDCU/BG level/4 miles | 727 | 31 | 99.9 |

Notes: LCDCU, least common demand coverage unit; BG, block group.
*This is the percent reported covered by the model for a given scale and representation although the 'true' coverage may be different.

increased. For a 2-mile coverage, the number of sites needed ranged from 24 to 55; for a 3-mile coverage, the number of sites ranged from 22 to 45; and for a 4-mile coverage, the number of sites ranged from 17 to 32. The MCLP was run then for each size coverage area, source demand scale, and representation combination for $p = 1, 2, 3, 5, 7, 10, 15, 20, 25$, and 30.

The results of these runs are then evaluated in terms of cost error and optimality error for both different levels of aggregation and types of representation. Cost error is associated with having different objective function values for different scales and representations so that the optimal set of facilities will have different objective function values. Optimality error is the difference in objective function values at the same scale that results when a difference set of optimal sites is chosen by different scale and representation combinations. Modifying terminology developed by Erkut and Bozkaya (1999), cost and optimality errors associated with both scale (aggregation) and representation are defined as follows:

$f_a$ is a larger scale objective function using representation $a$;
$f_b$ is a larger scale objective function using representation $b$;

$g_a$ is a smaller scale objective function using representation $a$;
$g_b$ is a smaller scale objective function using representation $b$;
$x_a$ is the optimal solution to the larger scale problem using representation $a$;
$x_b$ is the optimal solution to the larger scale problem using representation $b$;
$y_a$ is the optimal solution to the smaller scale problem using representation $a$;
$y_b$ is the optimal solution to the smaller scale problem using representation $b$.

With this notation, total error, cost error, and optimality error across scale and representation are defined as follows:

$$\text{Total error} = \frac{[g_a(y_a) - f_b(x_b)]}{f_b(x_b)} \tag{6}$$

$$\text{Cost error} = \frac{[g_a(y_a) - f_b(y_a)]}{f_b(x_b)} \tag{7}$$

$$\text{Optimality error} = \frac{[f_b(y_a) - f_b(x_b)]}{f_b(x_b)} \tag{8}$$

Total error can also be decomposed by first holding either representation or scale changes constant. Holding representation constant first:

$$\text{Scale (or Aggregation) error} = \frac{[g_a(y_a) - f_a(x_a)]}{f_b(x_b)} \tag{9}$$

$$\text{Representation error} = \frac{[f_a(x_a) - f_b(x_b)]}{f_b(x_b)} \tag{10}$$

Cost error and optimality error are defined as follows:

$$\text{Cost error (scale)} = \frac{[g_a(y_a) - f_a(y_a)]}{f_b(x_b)} \tag{11}$$

$$\text{Optimality error (scale)} = \frac{[f_a(y_a) - f_a(x_a)]}{f_b(x_b)} \tag{12}$$

$$\text{Cost error (representation)} = \frac{[f_a(x_a) - f_b(x_a)]}{f_b(x_b)} \tag{13}$$

$$\text{Optimality error (representation)} = \frac{[f_b(x_a) - f_b(x_b)]}{f_b(x_b)} \tag{14}$$

On the other hand, holding scale constant first:

$$\text{Representation error} = \frac{[g_a(y_a) - g_b(y_b)]}{f_b(x_b)} \tag{15}$$

$$\text{Scale (or Aggregation) error} = \frac{[g_b(y_b) - f_b(x_b)]}{f_b(x_b)} \tag{16}$$

Cost error and optimality error are defined as follows:

$$\text{Cost error (representation)} = \frac{[g_a(y_a) - g_b(y_a)]}{f_b(x_b)} \tag{17}$$

$$\text{Optimality error (representation)} = \frac{[g_b(y_a) - g_b(y_b)]}{f_b(x_b)} \tag{18}$$

$$\text{Cost error (scale)} = \frac{[g_b(y_b) - f_b(y_b)]}{f_b(x_b)} \tag{19}$$

$$\text{Optimality error (scale)} = \frac{[f_b(y_b) - f_b(x_b)]}{f_b(x_b)} \tag{20}$$

Comparing Equations (14) and (18), we can see that optimality error with respect to representation occurs at both scales and comparing Equations (12) and (20), optimality error with respect to scale occurs for both representations. If only scale issues are to be investigated, total error would be Equation (9) and Equations (11) and (12) are the only cost and optimality errors, respectively. If only representation issues are studied then total error would be Equation (15) and Equations (17) and (18) are the only cost and optimality errors, respectively.

## Results and discussion

In terms of total population covered, the results using the dasymetric LCDCU representation were mainly between the results of the centroid representation and the uniform LCDCU representation for all source scales and coverage sizes with a few exceptions (Figures 6–8). This was expected because the centroid and uniform LCDCU representations are the two extreme assumptions for a spatial distribution and the results for these two representations should bound that of the dasymetric LCDCU. The bounds for the dasymetric LCDCU results also became tighter as the source scale changed from smaller to larger scales. Comparing the results across the different scales, the dasymetric LCDCU values were closer to those of the uniform LCDCU than to the centroid values especially at the town level (Figures 6a, 7a, and 8a). However, at the block group level the results were very much the same for all coverage sizes (Figures 6c, 7c, and 8c).

Total error, cost error, and representation error were calculated next using the dasymetric LCDCU at the block group level as the base comparison. As expected, the total error for the centroid representation at the town level was the highest across all numbers of located sites and coverage sizes and the amount of error steadily declined as coverage size increased (Table 2). The centroid error at the town level peaked between 7 and 10 sites for the 2-mile coverage distance; it peaked between 10 and 15 sites for the 3-mile coverage distance; and for the 4-mile coverage distance, it peaked at one site and declined steadily to a minimum at 20 sites. In contrast both LCDCU representations at the town level peaked at one site and declined steadily for all three coverage distances. Across all coverage distances, these differences at the other two source scales were not as pronounced; the highest total error at the tract were only 4.5%, 3.6%, and 2.6%, respectively, for the three coverage distances (all associated with the centroid representation). No uniform LCDCU results differed more than 2.9% from the base.

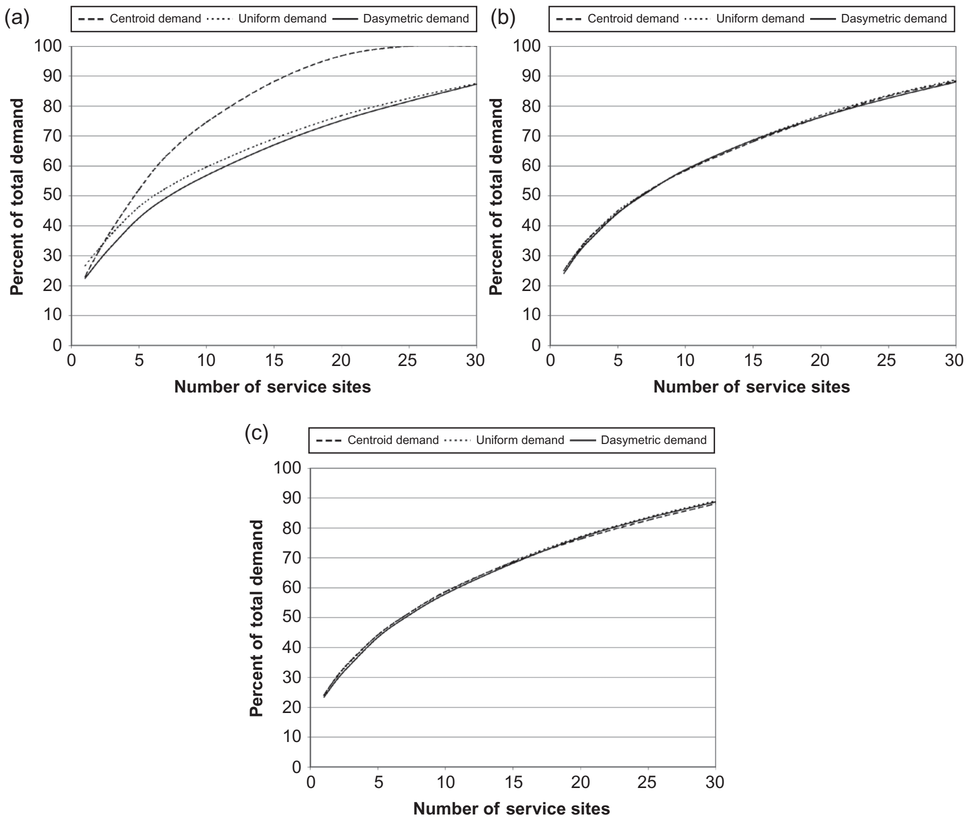


Figure 6. Percent of total demand covered at different scales for a 2-mile coverage distance. (a) Percent of total demand covered by each representation at the town level. (b) Percent of total demand covered by each representation at the tract level. (c) Percent of total demand covered by each representation at the block group level.

These results suggest that beyond a certain scale for the source demand, the type of representation has little impact especially as coverage distances increase. As noted in the areal interpolation literature (Okabe and Sadahiro 1997), the accuracy of interpolation results improves as the size of the target zones increases relative to the size of the source zones. As more source zones are completely contained within a target zone, the uncertainty of the estimates is reduced; if the source zones are nested perfectly within a target zone then the interpolation process would simply be an aggregation of population values. A percent certainty level for any target zone could be established by dividing the population of the source zones completely contained within the target zone by the total population of all source zones that intersect with the target zone and then converted into percentage terms. Thus, if no source zones are completely contained within a target zone it would have a 0% certainty level whereas if the source zones are nested perfectly within a target zone, the target zone would have a 100% certainty level.

A similar situation exists for the MCLP in which, for a given size coverage area (target zone), more source zones at a larger scale are completely contained within the coverage area. Figure 9 presents the 2-mile coverage area for one of the 57 potential sites. For town level demand, none of the four towns that intersect with it are completely contained within

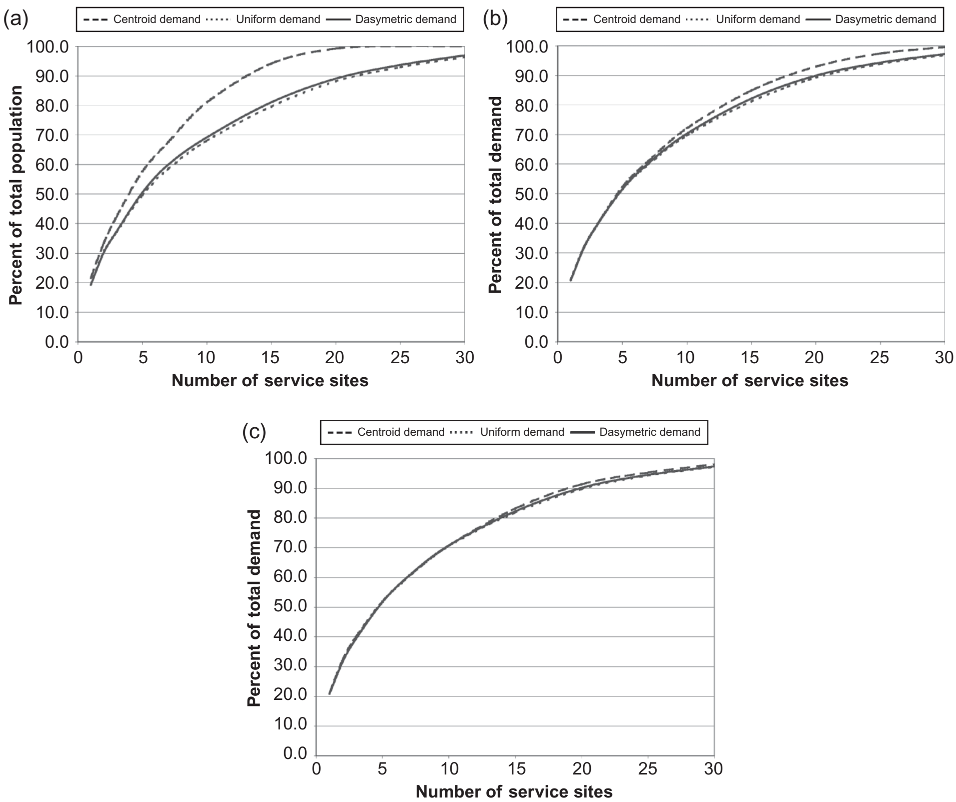


Figure 7. Percent of total demand covered at different scales for a 3-mile coverage distance. (a)Percent of total demand covered by each representation at the town level. (b) Percent of total demand covered by each representation at the tract level. (c) Percent of total demand covered by each representation at the block group level.

it (see Figure 9a). The four towns have a total population of 229,300 persons and the population of the coverage area could range from 0 to 229,300. There is a 0% certainty level regarding the population contained within it. At the tract level, 14 of the 34 tracts that intersect with the coverage area are completely contained within it (see Figure 9b). The 14 tracts completely contained within it have a population of 47,261 and the remaining 20 tracts that just intersect have a population of 69,448. For this scale, the population of the coverage area could range from 47,261 to 116,709 and the certainty level is 40.5%. Finally at the block group level, 38 of the 73 block groups that intersect with the coverage area are completely contained within it (see Figure 9c). The 38 tracts completely contained within it have a population of 53,942 and the remaining 45 tracts that just intersect with it have a population of 39,459. For this scale, the population of the coverage area could range from 53,942 to 93,401 and the certainty level is now 57.8%.

In addition, for any given source scale as the size of a coverage area increases, then it is more likely that more source zones will be completely contained within it than for a smaller coverage area. Figure 10 depicts the 4-mile coverage area for the same potential service site given in Figure 9. The source scale for Figure 10 is same as for Figure 9c, the block group level. Now, 152 of the 209 block groups that intersect with this coverage

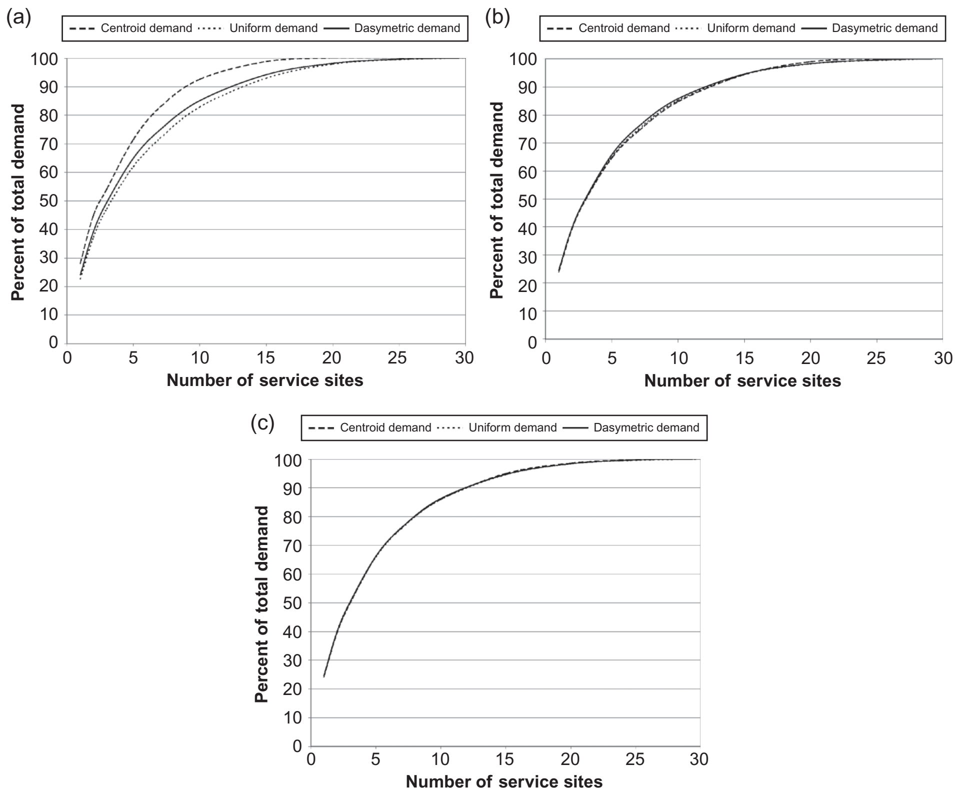


Figure 8. Percent of total demand covered at different scales for a 4-mile coverage distance. (a)Percent of total demand covered by each representation at the town level. (b) Percent of total demand covered by each representation at the tract level. (c)Percent of total demand covered by each representation at the block group level.

area are completely contained within it. These 152 tracts have a population of 178,388 and the remaining 57 intersecting block groups have a population of 69,867. The population within the total coverage area could range from 178,388 to 248,255 and the certainty level has increased to 71.9%.

Decomposing total error into cost error and optimality error, cost error was mostly higher than optimality error. The results for $P = 10$ sites are illustrative of the overall trends (Table 3). The cost error for centroid representation at the town level ranged from 25.5% to 68.9% whereas the optimality error ranged only from −8% to –26.0%. Cost errors were mostly positive for the centroid representation because as previously discussed the assumption of perfect concentration usually inflates total coverage (the only negative cost values were –0.1% at the block group level for the 2-mile and 4-mile coverage distances). With one exception, for all LCDCU representations the cost error is negative although there is no obvious reason to necessarily expect this. The more important issue is optimality error because this error is associated with incorrect locations being chosen; optimality errors are always nonpositive as expected. The optimality errors for the centroid representation at the tract and block group levels for all coverage distances are also less in magnitude than the optimality error for the uniform LCDCU representation at the town level, and the

Table 2. Percent total error for different representations and scale using (a) 2-mile buffer, (b) 3-mile buffer, and (c) 4-mile buffer.

| | Number of sites | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Representation/scale | $P=1$ | $P=2$ | $P=3$ | $P=5$ | $P=7$ | $P=10$ | $P=15$ | $P=20$ | $P=25$ | $P=30$ |
| 2-mile buffer | | | | | | | | | | |
| Centroid/town level | 24.3 | 28.5 | 32.9 | 37.3 | 43.0 | 42.9 | 41.0 | 35.4 | 27.7 | 18.7 |
| Centroid/tract level | 4.5 | 3.7 | 4.2 | −0.8 | −0.4 | −1.5 | −2.3 | −2.8 | −1.6 | −1.9 |
| Centroid/BG level | −1.5 | 1.2 | 1.4 | −0.5 | −0.5 | −0.1 | −0.9 | −2.2 | −2.2 | −2.0 |
| Uniform LCDCU/town level | −26.5 | 1.8 | 5.5 | 1.5 | 10.1 | 0.3 | .8 | 0.0 | 0.0 | .9 |
| Uniform LCDCU/tract level | −0.3 | −0.9 | 0.3 | −1.0 | −2.6 | −2.2 | −2.4 | −2.9 | −2.2 | −2.3 |
| Uniform LCDCU/BG level | 0.3 | 0.1 | 0.3 | −0.4 | −0.6 | −0.6 | −0.5 | −0.9 | −0.8 | −0.7 |
| Dasymetric LCDCU/town level | −20.3 | −15.6 | −11.2 | −7.1 | −5.2 | −5.3 | −4.6 | −4.7 | −4.3 | −3.7 |
| Dasymetric LCDCU/tract level | −1.5 | −1.2 | −0.5 | −0.9 | −1.4 | −1.1 | −0.8 | −0.9 | −0.6 | −0.7 |
| 3-mile buffer | | | | | | | | | | |
| Centroid/town level | 13.8 | 11.5 | 12.9 | 15.4 | 14.9 | 17.5 | 17.1 | 12.3 | 7.8 | 4.7 |
| Centroid/tract level | 1.8 | 0.6 | 0.2 | 1.6 | 1.0 | 2.3 | 3.6 | 3.5 | 3.3 | 2.6 |
| Centroid/BG level | 2.3 | 2.8 | 2.2 | 0.7 | −0.1 | 0.1 | 1.4 | 1.5 | 1.0 | 0.9 |
| Uniform LCDCU/town level | −13.0 | −8.0 | −8.8 | −6.4 | −5.5 | −5.6 | −4.7 | −3.5 | −2.9 | −2.2 |
| Uniform LCDCU/tract level | −0.8 | −0.9 | −1.2 | −1.4 | −2.0 | −2.0 | −1.8 | −1.4 | −1.1 | −0.8 |
| Uniform LCDCU/BG level | 0.1 | 0.1 | −0.2 | −0.3 | −0.4 | −0.6 | −0.7 | −0.6 | −0.5 | −0.3 |
| Dasymetric LCDCU/town level | −9.3 | −6.0 | −6.0 | −3.0 | −1.6 | −2.7 | −1.8 | −1.5 | −1.2 | −0.7 |
| Dasymetric LCDCU/tract level | −0.5 | −0.8 | −0.5 | −0.4 | −0.7 | −0.8 | −0.1 | −0.3 | −0.2 | −0.1 |
| 4-mile buffer | | | | | | | | | | |
| Centroid/town level | 14.1 | 13.2 | 8.3 | 8.1 | 8.7 | 7.3 | 4.3 | 1.7 | 0.3 | 0.0 |
| Centroid/tract level | −2.4 | −1.1 | −1.1 | −2.4 | −2.6 | −1.9 | −0.4 | 0.6 | 0.3 | 0.0 |
| Centroid/BG level | −1.7 | −0.7 | −0.7 | −0.1 | 0.1 | −0.4 | 0.3 | 0.2 | 0.1 | 0.0 |
| Uniform LCDCU/town level | −8.7 | −6.5 | −5.9 | −6.4 | −5.9 | −4.2 | −2.0 | −0.7 | −0.4 | −0.2 |
| Uniform LCDCU/tract level | −0.6 | −0.3 | −1.1 | −2.0 | −2.1 | −1.4 | −0.5 | −0.3 | −0.4 | −0.1 |
| Uniform LCDCU/BG level | −0.1 | 0.0 | −0.3 | −0.5 | −0.5 | −0.5 | −0.3 | −0.2 | −0.3 | −0.2 |
| Dasymetric LCDCU/town level | −1.9 | −1.1 | −1.1 | −1.8 | −1.8 | −1.4 | −0.6 | −0.2 | −0.1 | 0.0 |
| Dasymetric LCDCU/tract level | 0.0 | 0.0 | −0.2 | −0.6 | −0.6 | −0.5 | −0.1 | −0.1 | −0.1 | 0.0 |

Note: LCDCU, least common demand coverage unit; BG, block group.

optimality errors for the centroid representation at the block group level are less in magnitude than the same errors for the dasymetric LCDCU representation at the town level. With respect to optimality error, scale difference is as important as representation difference. A

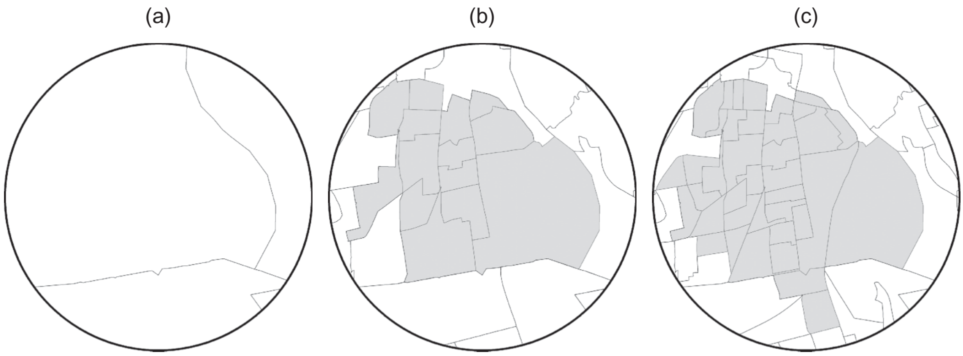


Figure 9. Source zones completely contained within a 2-mile coverage distance for an individual potential service site. Zones completely contained are shaded gray. (a) Source zones at the town scale. (b) Source zones at the tract scale. (c) Source zones at the block group scale.

Figure 10. Block group source zones completely contained within a 4-mile coverage distance for an individual potential service site. Zones completely contained are shaded gray.

further check of the importance of scale versus representation was the decomposition of the error by holding either representation or scale constant first (Table 4). The results are again inconclusive as the optimality error for scale holding representation constant first is almost the same as the optimality error for representation holding scale constant first for each coverage distance (–27.8% vs. –24.2% for 2 miles, –7.9% vs. –8.1% for 3 miles, and –22.0% vs. –22.5% for 4 miles).

Table 3. Percent cost and optimality errors when $P = 10$ for different representations, scale, and distance.

| Representation/scale/distance | Cost error | Optimality error |
|---|---|---|
| Centroid/town level/2 miles | 68.9 | −26.0 |
| Centroid/town level/3 miles | 25.5 | −8.0 |
| Centroid/town level/4 miles | 30.0 | −22.7 |
| Centroid/tract level/2 miles | 0.3 | −1.7 |
| Centroid/tract level/3 miles | 4.0 | −1.7 |
| Centroid/tract level/4 miles | 0.9 | −2.8 |
| Centroid/BG level/2 miles | −0.1 | 0.0 |
| Centroid/BG level/3 miles | 0.7 | −0.4 |
| Centroid/BG level/4 miles | −0.1 | −0.3 |
| Uniform LCDCU/town level/2 miles | −5.3 | −5.1 |
| Uniform LCDCU/town level/3 miles | −2.0 | −3.6 |
| Uniform LCDCU/town level/4 miles | −0.2 | −4.0 |
| Uniform LCDCU/tract level/2 miles | −1.4 | −0.8 |
| Uniform LCDCU/tract level/3 miles | −1.9 | −0.1 |
| Uniform LCDCU/tract level/4 miles | −0.8 | −0.6 |
| Uniform LCDCU/BG level/2 miles | −0.6 | 0.0 |
| Uniform LCDCU/BG level/3 miles | −0.6 | 0.0 |
| Uniform LCDCU/BG level/4 miles | −0.5 | 0.0 |
| Dasymetric LCDCU/town level/2 miles | −4.4 | −0.9 |
| Dasymetric LCDCU/town level/3 miles | −0.5 | −2.2 |
| Dasymetric LCDCU/town level/4 miles | −0.8 | −0.6 |
| Dasymetric LCDCU/tract level/2 miles | 0.6 | −1.7 |
| Dasymetric LCDCU/tract level/3 miles | −0.8 | 0.0 |
| Dasymetric LCDCU/tract level/4 miles | −0.2 | −0.3 |

Note: LCDCU, least common demand coverage unit; BG, block group.

## Conclusions

How geographic objects are represented at a particular scale is a fundamental issue in spatial analysis. As one moves from larger to smaller scales, the spatial dimensionality of an object often changes. When dealing with demand, individual scale demand is aggregated at smaller scales into group demand covering an area. For a given scale, the representation of this aggregated demand could have an impact on model results. The assumption that demand within an area is either concentrated at a centroid or uniformly distributed represents the two polar extremes of a continuum for a spatial distribution – perfectly concentrated versus perfectly uniform. With respect to the MCLP in which the number of sites is fixed, the two extreme assumptions can either overestimate or underestimate the amount of demand covered depending on how demand is actually distributed. These two extremes do perform bounds on the level of representation error. The objective function results using intelligent areal interpolation to model the spatial heterogeneity of demand for the most part lie somewhere between the objective function values for the centroid and uniform distributions.

It was found that the choice of representation depends as much on the scale of the source demand data as its representation. Comparing across representations, the centroid results differed from the uniform LCDCU and dasymetric LCDCU results at the town scale but were very similar to them at the tract and block group scales. The similarity of results increased also as the size of the coverage distance increased. Comparing within a representation, the dasymetric LCDCU and uniform LCDCU methods produced similar

Table 4. Percent cost and optimality errors when $P = 10$ for different decompositions using (a) a 2-mile buffer, (b) a 3-mile buffer, and (c) a 4-mile buffer.

| | | Holding representation constant first | | Holding scale constant first | |
|---|---|---|---|---|---|
| Representation/scale | Source | Cost | Optimality | Cost | Optimality |
| 2-mile buffer | | | | | |
| Centroid/town level | Scale | 70.8 | −27.8 | −4.4 | −0.9 |
| | Representation | −0.1 | 0.0 | 72.5 | −24.2 |
| Centroid/tract level | Scale | 0.4 | −1.8 | −1.3 | −1.7 |
| | Representation | −0.1 | 0.0 | 1.2 | −0.3 |
| Centroid/BG level | Scale | – | – | – | – |
| | Representation | –0.1 | 0.0 | −0.1 | 0.0 |
| Uniform LCDCU/town level | Scale | −4.4 | −5.3 | −4.4 | −0.9 |
| | Representation | −0.6 | 0.0 | −2.2 | −2.7 |
| Uniform LCDCU/tract level | Scale | −0.7 | −0.9 | −1.3 | −1.7 |
| | Representation | −0.6 | 0.0 | −0.9 | 1.7 |
| Uniform LCDCU/BG level | Scale | – | – | – | – |
| | Representation | −0.6 | 0.0 | −0.6 | 0.0 |
| Dasymetric LCDCU/town level | Scale | −4.4 | −0.9 | −4.4 | −0.9 |
| | Representation | – | – | – | – |
| Dasymetric LCDCU/tract level | Scale | 0.6 | −1.7 | 0.6 | −1.7 |
| | Representation | – | – | – | – |
| 3-mile buffer | | | | | |
| Centroid/town level | Scale | 25.3 | −7.9 | −0.5 | −2.2 |
| | Representation | 0.5 | −0.4 | 28.3 | −8.1 |
| Centroid/tract level | Scale | 3.9 | −1.7 | −0.8 | 0.0 |
| | Representation | 0.5 | −0.4 | 4.3 | −1.2 |
| Centroid/BG level | Scale | – | – | – | – |
| | Representation | 0.5 | −0.4 | 0.5 | −0.4 |
| Uniform LCDCU/town level | Scale | −1.5 | −3.5 | −2.1 | −0.8 |
| | Representation | −0.6 | 0.0 | −0.5 | −2.2 |
| Uniform LCDCU/tract level | Scale | −1.5 | 0.1 | −0.8 | 0.0 |
| | Representation | −0.6 | 0.0 | −1.3 | 0.1 |
| Uniform LCDCU/BG level | Scale | – | – | – | – |
| | Representation | −0.6 | 0.0 | −0.6 | 0.0 |
| Dasymetric LCDCU/town level | Scale | −0.5 | −2.2 | −0.5 | −2.2 |
| | Representation | – | – | – | – |
| Dasymetric LCDCU/tract level | Scale | −0.8 | 0.0 | −0.8 | 0.0 |
| | Representation | – | – | – | – |
| 4-mile buffer | | | | | |
| Centroid/town level | Scale | 29.7 | −22.0 | −0.8 | −0.6 |
| | Representation | −0.1 | −0.3 | 31.1 | −22.5 |
| Centroid/tract level | Scale | 1.5 | −3.0 | −0.2 | −0.3 |
| | Representation | −0.1 | −0.3 | 1.6 | −3.0 |
| Centroid/BG level | Scale | – | – | – | – |
| | Representation | −0.1 | −0.3 | −0.1 | −0.3 |
| Uniform LCDCU/town level | Scale | −0.2 | −3.5 | −0.8 | −0.6 |
| | Representation | −0.5 | 0.0 | −1.0 | −1.9 |
| Uniform LCDCU/tract level | Scale | −0.5 | −0.4 | −0.2 | −0.3 |
| | Representation | −0.5 | 0.0 | −0.6 | −0.2 |
| Uniform LCDCU/BG level | Scale | – | – | – | – |
| | Representation | −0.5 | 0.0 | −0.5 | 0.0 |
| Dasymetric LCDCU/town level | Scale | −0.8 | −0.6 | −0.8 | −0.6 |
| | Representation | – | – | – | – |
| Dasymetric LCDCU/tract level | Scale | –0.2 | –0.3 | –0.2 | –0.3 |
| | Representation | – | – | – | – |

Note: LCDCU, least common demand coverage unit; BG, block group.

results across the three scales whereas the centroid results were more scale dependent. The conclusion is that at smaller scales, the centroid representation should be avoided but at larger scales it probably is as good as the uniform LCDCU approach.

Another aspect of the choice of representation besides the quality of the results is its impact on problem size for computational purposes. It was discussed that the number of constraints in the traditional MCLP is a function of demand scale for the centroid approach but is a function of the number of potential service sites in the LCDCU approach. However, the two approaches are not mutually independent. Using the centroid approach will initially increase the number of constraints for a larger scale problem. However, constraint (or row) reductions are possible by performing selecting and retaining only those LCDCUs that contain centroids. In the case study, this step reduced the number of constraints for the tract and block group scale centroid representations to less than half of the number of LCDCUs assuming either uniform demand or dasymetric demand for results that were not substantially different. The intelligent areal interpolation method also used this technique to aggregate over 500,000 points expected to have demand down to a much smaller number of LCDCUs. The major difference between the reduced centroid representation and the dasymetric representation was the number of points that was aggregated into the LCDCUs. Further research will also need to examine these scale consequences from the perspective of service sites on demand representation.

Finally, this research presents an alternative representation and processing procedures for solving the MCLP in a GIS environment. The location analyst can use other readily available data layers such as land cover and intelligent areal interpolation procedures to improve assumptions regarding the spatial distribution of demand. These techniques reduce the problem size as well as improve the quality of results for the MCLP.

**Acknowledgments**

The authors would like to thank the anonymous reviewers for their valuable comments and criticisms.

**References**

Alexandris, G. and Giannikos, I., 2010. A new model for maximal coverage exploiting GIS capabilities. *European Journal of Operational Research*, 202, 328–338.

Armstrong, M. and Densham, P., 1990. Database organization strategies for spatial decision support systems. *International Journal of Geographical Information Systems*, 4, 3–20.

Bloom, L., Pedler, P., and Wragg, G., 1996. Implementation of enhanced areal interpolation using Mapinfo. *Computers and Geosciences*, 22, 459–466.

Bowerman, R., Calamai, P., and Hall, G.B., 1999. The demand partitioning method for reducing aggregation errors in p-median problems. *Computers & Operations Research*, 26, 1097–1111.

Casillas, P., 1987. Data aggregation and the p-median problem in continuous space. *In*: A. Ghosh and G. Rushton, eds. *Spatial analysis and location-allocation models*. New York: Van Nostrand Reinhold, 327–344.

Church, R., 2002. Geographical information systems and location science. *Computers & Operations Research*, 29, 541–562.

Church, R. and Revelle, C., 1974. The maximal covering location problem. *Papers of the Regional Science Association*, 32, 101–118.

Current, J. and Schilling, D., 1987. Elimination of source A and B errors in p-median location problem. *Geographical Analysis*, 19, 95–110.

Current, J. and Schilling, D., 1990. Analysis of errors due to demand data aggregation in the set covering and maximal covering location problems. *Geographical Analysis*, 22, 116–126.

Daskin, M., *et al*., 1989. Aggregation effects in maximal covering models. *Annals of Operation Research*, 18, 115–140.

Eastman, J.R., *et al*., 1995. Raster procedures for multiobjective land-use planning. *Photogrammetric Engineering and Remote Sensing*, 61, 539–547.

Eicher, C. and Brewer, C., 2001. Dasymetric mapping and areal interpolation: implementation and evaluation. *Cartography and Geographic Information Science*, 28, 125–138.
Erkut, E. and Bozkaya, B., 1999. Analysis of aggregation errors for the p-median problem. *Computers & Operations Research*, 26, 1075–1096.
Fisher, P. and Langford, M., 1995. Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A*, 27, 211–224.
Fisher, P. and Langford, M., 1996. Modeling sensitivity to accuracy in classified imagery: a study of areal interpolation by dasymetric mapping. *Professional Geographer*, 48, 299–309.
Flowerdew, R. and Green, M., 1989. Statistical methods for transferring data between zonal systems. *In*: M. Goodchild and S. Gopal, eds. *The accuracy of spatial database*. London: Taylor & Francis, 239–247.
Flowerdew, R. and Green, M., 1994. Areal interpolation and types of data. *In*: S. Fotheringham and P. Rogerson, eds. *Spatial analysis and GIS*. London: Taylor & Francis, 121–145.
Flowerdew, R., Green, M., and Kehris, E., 1991. Using areal interpolation methods in geographic information systems. *Papers in Regional Science*, 70, 303–315.
Fotheringham, A.S., Densham, P., and Curtis, A., 1995. The zone definition problem in location-allocation modeling. *Geographical Analysis*, 27, 60–77.
Francis, R.L., Lowe, T., and Rayco, M., 1996. Row-column aggregation for rectilinear distance p-median problems. *Transportation Science*, 30, 160–174.
Francis, R.L., Lowe, T., and Tamir, A., 2000. Aggregation error bounds for a class of location models. *Operations Research*, 48, 294–307.
Francis, R.L., *et al*., 1999. A synthesis of aggregation methods for multi-facility location problems: strategies for containing error. *Geographical Analysis*, 31, 67–87.
Francis, R.L., *et al*., 2009. Aggregation error for location models: survey and analysis. *Annals of Operations Research*, 167, 171–208.
Goodchild, M., 1979. The aggregation problem in location-allocation. *Geographical Analysis*, 11, 240–255.
Goodchild, M., Anselin, L., and Deichmann, U., 1993. A framework for the areal interpolation of socioeconomic data. *Environment and Planning A*, 25, 383–397.
Goodchild, M. and Lam, N., 1980. Areal interpolation: variant of the traditional spatial problem. *Geo-Processing*, 1, 297–312.
Hillsman, E. and Rhoda, R., 1978. Errors in measuring distances from populations to service centers. *Annals of Regional Science*, 12, 74–88.
Hodgson, M.J. and Neuman, S., 1993. A GIS approach to eliminating source C aggregation error in p-median models. *Location Science*, 1, 155–170.
Holt, J., Lo, C.P., and Hodler, T., 2004. Dasymetric estimation of population density and areal interpolation of census data. *Cartography and Geographic Information Science*, 31, 103–121.
Langford, M., 2006. Obtaining population estimates in non-census reporting zones: an evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems*, 30, 161–180.
Langford, M., Maguire, D., and Unwin, D., 1991. The areal interpolation problem: estimating population using remote sensing in a GIS framework. *In*: I. Masser and M. Blakemore, eds. *Handling geographical information*. London: Longman, 55–77.
Maantay, J.A., Maroko, A.R., and Herrmann, C., 2007. Mapping population distribution in the urban environment: the cadastral-based expert dasymetric system. *Cartography and Geographic Information Science*, 34, 77–102.
Malczewski, J., 1999. *GIS and multicriteria decision analysis*. New York: John Wiley & Sons.
Mennis, J., 2003. Generating surface models of population using dasymetric mapping. *The Professional Geographer*, 55, 31–42.
Mennis, J. and Hultgren, T., 2007. Intelligent dasymetric mapping and its application to areal interpolation. *Cartography and Geographic Information Science*, 33, 179–194.
Miller, H., 1996. GIS and geometric representation in facility location problems. *International Journal of Geographical Information Systems*, 10, 791–816.
Mugglin, A. and Carlin, B., 1998. Hierarchical modeling in geographic information systems: population interpolation over incompatible zones. *Journal of Agricultural, Biological, and Environmental Statistics*, 3, 111–130.
Mugglin, A., *et al*., 1999. Bayesian areal interpolation, estimation, and smoothing: an inferential approach for geographic information systems. *Environment and Planning A*, 31, 1337–1352.

Murray, A. and Gottsegen, J., 1997. The influence of data aggregation on the stability of p-median location model solutions. *Geographical Analysis*, 29, 200–213.

Murray, A. and O'Kelly, M., 2002. Assessing representation error in point-based coverage modeling. *Journal of Geographical Systems*, 4, 171–191.

Murray, A., O'Kelly, M., and Church, R., 2008. Regional service coverage modeling. *Computers & Operations Research*, 35, 339–355.

Murray, A. and Tong, D., 2007. Coverage optimization in continuous space facility siting. *International Journal of Geographical Information Science*, 21, 757–776.

Okabe, A. and Sadahiro, Y., 1997. Variation in count data transferred from a set of irregular zones to a set of regular zones through the point-in-polygon method. *International Journal of Geographical Information Science*, 11, 93–106.

Openshaw, S. and Taylor, P., 1981. The modifiable areal unit problem. *In*: N. Wrigley and R. Bennett, eds. *Quantitative geography: a British view*. London: Routledge, 60–69.

Peucker, T. and Chrisman, N., 1975. Cartographic data structures. *The American Cartographer*, 2, 55–69.

Plastria, F., 2001. On the choice of aggregation point for continuous p-median problems: a case for the gravity centre. *TOP*, 9, 217–242.

Reibel, M. and Agrawal, A., 2007. Areal interpolation of population counts using pre-classified land cover data. *Population Research and Policy Review*, 26, 619–633.

Revelle, C. and Swain, R., 1970. Central facilities location. *Geographical Analysis*, 2, 30–42.

Spaulding, B. and Cromley, R., 2007. Integrating the maximum capture problem into a GIS framework. *Journal of Geographical Systems*, 9, 267–288.

Tobler, W., 1979. Smooth pycnophylactic interpolation of geographical regions. *Journal of the American Statistical Association*, 74, 519–530.

Tong, D. and Murray, A., 2009. Maximising coverage of spatial demand for service. *Papers in Regional Science*, 88, 85–97.

Toregas, C. and Revelle, C., 1973. Binary logic solutions to a class of location problem. *Geographical Analysis*, 5, 145–155.

Toregas, C., *et al*., 1971. The location of emergency service facilities. *Operations Research*, 19, 1363–1373.

Wright, J., 1936. A method of mapping densities of population. *The Geographical Review*, 26, 103–110.

Zhao, P. and Batta, R., 1999. Analysis of centroid aggregation for the Euclidean distance p-median problem. *European Journal of Operational Research*, 113, 147–168.

Zhao, P. and Batta, R., 2000. An aggregation approach to solving the network p-median problem with link demands. *Networks*, 36, 233–241.