

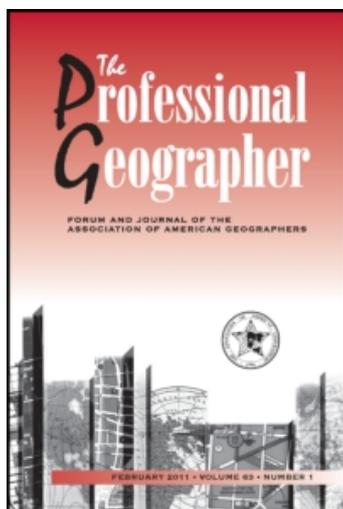
This article was downloaded by: [Florida Atlantic University]

On: 23 March 2011

Access details: Access Details: [subscription number 784176984]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



The Professional Geographer

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t788352615>

A Point-Based Intelligent Approach to Areal Interpolation

Caiyun Zhang^a; Fang Qiu^b

^a Florida Atlantic University, ^b University of Texas, Dallas

First published on: 02 March 2011

To cite this Article Zhang, Caiyun and Qiu, Fang(2011) 'A Point-Based Intelligent Approach to Areal Interpolation', The Professional Geographer, 63: 2, 262 – 276, First published on: 02 March 2011 (iFirst)

To link to this Article: DOI: 10.1080/00330124.2010.547792

URL: <http://dx.doi.org/10.1080/00330124.2010.547792>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

A Point-Based Intelligent Approach to Areal Interpolation*

Caiyun Zhang

Florida Atlantic University

Fang Qiu

University of Texas at Dallas

Areal interpolation is the data transfer from one zonal system to another. A survey of previous literature on this subject points out that the most effective methods for areal interpolation are the intelligent approaches, which often take two-dimensional (2-D) land use or one-dimensional (1-D) road network information as ancillary data to give insight on the underlying distribution of a variable. However, the 2-D or 1-D ancillary information is not always applicable for the variable of interest in a specific study area. This article introduces a point-based intelligent approach to the areal interpolation problem by using zero-dimensional (0-D) points as ancillary data that are locationally associated with the variable of interest. The connection between zonal variables and point locations can be modeled with a linear or a nonlinear exponential function, which incorporates the distribution of the variables in the transferring of the information from the source zone to the target zone. An experimental study interpolating the population data at a suburbanized area suggests that the proposed method is an attractive alternative to other areal interpolation solutions based on the evaluation of its resulting accuracy and efficiency. **Key Words:** areal interpolation, models, population estimation.

面插值是把一个分区系统的数据传输到另一个分区系统的技术。对于这个技术的文献研究指出，面插值最有效的方法是智能模式：往往采取二维（2-D）的土地用途或一维（1-D）的路由网络信息，作为辅助数据对一个变量的潜在分布进行深入分析。然而，对于一个具体的研究领域所感兴趣的变量，2-D 或 1-D 辅助信息并不总是适用。本文介绍了一种基于点数据的智能方法以解决面插值问题，即利用零维（0-D）点作为附属数据，这些附属数据与感兴趣的变量具有区域相关性。分区变量和位置点之间的联系可以通过线性或非线性的指数函数而加以建模，这些函数综合表述了从源区到目标区信息传输的变量分布。我们对郊区化地区的人口数据进行了插值方法的实验研究，研究表明，鉴于对其结果的准确性和效率的评估，上述方法是一个有吸引力的替代性面插值方法。关键词：区域插值，模型，人口估计。

La interpolación espacial es la transferencia de datos de un sistema zonal a otro. La revisión de la literatura existente sobre esta materia indica que los métodos más efectivos de interpolación espacial son los enfoques inteligentes, que a menudo adoptan la red de información bidimensional de uso del suelo (2-D) o la vía unidimensional (1-D) como datos subsidiarios para dar comprensión a la distribución subyacente de una variable. Sin embargo, la información subsidiaria 2-D o 1-D no siempre es aplicable a la variable de interés en un área específica de estudio. En este artículo se introduce un enfoque inteligente de base puntual al problema de la interpolación espacial, utilizando puntos cero-dimensionales (0-D) como datos subsidiarios que están asociados locacionalmente con la variable de interés. La conexión entre las variables zonales y las localizaciones puntuales pueden modelarse con una función exponencial lineal o no lineal, que incorpore la distribución de las variables en la transferencia de información desde una zona fuente a la zona de destino. Un

*We would like to extend our appreciation to the anonymous reviewers. Their constructive comments were very valuable in improving the quality of this article. Caiyun Zhang would also like to acknowledge the support of the Natural Science Foundation of China through projects 40730530 and 40706056.

estudio experimental en el que se interpolan los datos de población en un área suburbanizada sugiere que el método propuesto es una alternativa atractiva para buscar soluciones a otras interpolaciones espaciales basadas en la evaluación de los resultantes niveles de exactitud y eficiencia. **Palabras clave:** interpolación espacial, modelos, cálculos de población.

Areal interpolation involving the transformation of attribute data from one zonal system to another is frequently used in economic, social, and urban studies. The zonal units for which data are available are known as *source zones* and those for which data are required are called *target zones* (Markoff and Shapiro 1973). The need for areal interpolation arises when spatially aggregated data are available for one set of geographic areal units (or zones) but not for the areal units of current interest. In the United States, population data have often been collected and aggregated every ten years in area units such as census blocks, block groups, and census tracts, which are arbitrarily defined by the U.S. Census Bureau for survey purposes. Population data, however, are often needed in socioeconomic, urban, and interdisciplinary studies, where the study units are usually different from those used by the Census Bureau. For example, many businesses often need demographic data for marketing analysis zones or service or trade areas. Similarly, in the natural sciences, the geographic units of analysis are usually areas defined by land use, land cover, soil type, watershed boundaries, and a variety of other biophysical and geophysical features. Given that census geography and its concomitant demographic data seldom correspond exactly to these areas, it is crucial that population counts and other related information can be transferred into new area units so that the data from different disciplines and disparate units of analysis can be integrated.

A variety of algorithms for areal interpolation have been proposed based on different assumptions for the underlying distributions of variables. These algorithms can be grouped into two categories: simple methods that do not employ ancillary data and intelligent methods that make use of ancillary data (Hawley and Moelering 2005).

The best known simple method is the area-weighting approach, which interpolates a variable based on the area of intersection between the source and target zones without using ancillary data (Lam 1983). This method has been used in various applica-

tions, for example, to generate a consistent global georeferenced population data set known as Gridded Population of the World (GPW) in 1995 (Tobler et al. 1995), which is now updated by the Center for International Earth Science Information Network of Columbia University with improved methods (see <http://sedac.ciesin.columbia.edu/gpw/>). A fundamental problem with this approach is that it assumes spatial homogeneity of the variable of interest within each source zone, an assumption that is rarely true in the real world. The pycnophylactic method (Tobler 1979) is also a simple method that provides a solution to this problem. It fits a continuous surface to the source zone data and then uses that surface to aggregate values for the target zones. Tobler applied this method to generate population density maps of 1970 at the state level and the census tract level. A centroid-based method proposed by Martin (1989) and Bracken and Martin (1989) is another simple solution and was used for census mapping in the United Kingdom. This method places a kernel density window over a source zone centroid to allocate the population into each grid cell with a distance-based weighting strategy. This simple approach does not conserve the total value of each source zone, a property known as *volume preserving* in the literature and regarded as an important characteristic in areal interpolation. To solve this problem, Martin (1996) modified the original centroid-based algorithm to ensure that the populations reported for target zones are constrained to match the overall sum of the source units. Similarly, Harris and Chen (2005) employed UK postcodes to estimate population density using the population surface modeling technique proposed by Martin (1989) and Bracken and Martin (1989).

With the inclusion of additional data, various algorithms have been proposed to provide better solutions to the areal interpolation problems. These methods are known as *intelligent methods*, because they utilize the ancillary data to inform the distribution of the variable of interest in the source zone. These intelligent methods have been extensively applied to

population interpolations (Mennis 2003; Holt, Lo, and Hodler 2004; Langford 2006; Reibel and Agrawal 2007), socioeconomic variable estimations (Goodchild, Anselin, and Deichmann 1993; Eicher and Brewer 2001; Mennis and Hultgren 2006), and time-variant population analysis with changing historical administrative boundaries (Gregory 2002a, 2002b). Among the intelligent methods within the literature, the so-called dasymetric approach is the most cited. The dasymetric method was originally designed for mapping purposes by Wright (1936) to improve the depiction of population distribution in choropleth areal maps. It was then adopted as an areal interpolation approach that divides source zones into smaller constituents possessing different but internally consistent density of the variable (Langford, Magnire, and Unwin 1991; Langford 2006). Dasymetric approaches commonly employ two-dimensional (2-D) ancillary data, such as land use, to provide insight on the underlying distribution of the variable in the source zone (e.g., Fisher and Langford 1995, 1996; Eicher and Brewer 2001; Mennis 2003; Holt, Lo, and Hodler 2004). With the availability of one-dimensional (1-D) Topologically Integrated Geographic Encoding and Referencing (TIGER) line geographic information system (GIS) data, Xie (1995) was the first to attempt to solve the areal interpolation problem by using these types of data as ancillary information using a similar idea. Reibel and Bufalino (2005) investigated the errors in Xie's algorithm.

The ancillary data are often called *control zones* if they are 2-D (Goodchild, Anselin, and Deichmann 1993). Similarly we can refer to the 1-D ancillary data as control lines if they are 1-D. Many variations of these intelligent methods, such as smart dasymetric methods (Deichmann 1996; Turner and Openshaw 2001), intelligent dasymetric methods (Mennis and Hultgren 2006), and regression dasymetric methods (Langford, Magnire, and Unwin 1991; Yuan, Smith, and Limp 1997; Langford 2006), have also been proposed.

Statistical methods have also been employed in areal interpolation. Those methods model the relationship between the distribution of the variable of interest and the ancillary data, as in the regression dasymetric method. The regression analysis is usually conducted at the aggre-

gate level, for example, by using the total count of the variable in each source zone and the aggregate area of the control zones falling within the source zone. The resulting regression function is then applied to each target zone to estimate its total count by using the aggregate area of the control zones falling into the target zone. A statistical method called the EM (expectation/maximum) algorithm that was proposed by Flowerdew and Green (Flowerdew 1988; Flowerdew and Green 1991) is innovative. It models the relationship between estimated population density and socioeconomic variables such as the number of people or car ownership per household at the disaggregate level (i.e., control unit level). Similarly, Harvey (2002a, 2002b) used an iterated regression procedure as a least-squares approximation of the EM algorithm to correlate population with the digital number of each resident pixel of a satellite image. These methods often derive the regression formula from individual estimations with assumed statistical rules, and then apply those rules to the whole units being studied (Xie 1995). The success of these methods depends heavily on the availability of detailed control variables at the disaggregate level and the degree that the variable of interest precisely follows a specific standard statistical distribution, a demand that might limit its wide application.

Geostatistical approaches that were originally designed for point interpolation have also recently been adopted in areal interpolation by incorporating spatial autocorrelation into the modeling process. Kyriakidis (2004) established a theoretical framework of area-to-point interpolation based on spatial cross-correlation between areal and point variables through cokriging procedures. Kyriakidis and Yoo (2005) illustrated the realization of this theoretical work through the application of a synthetic image data set. Liu, Kyriakidis, and Goodchild (2008) further extended the model to disaggregate the residuals that resulted from a regression between population density and built-up and vegetation compositions obtained from an IKONOS image. In a separate study conducted independently by Wu and Murray (2005), the imperviousness fraction derived from an Enhanced Thematic Mapper Plus (ETM+) image was adopted as a secondary variable in a cokriging prediction of population density. As an emerging approach, geostatistical-based areal

interpolation, with its theoretical framework being recently developed and tested by applications using simulated grid and satellite imagery, is anticipated to be a promising method if verified by more real-world applications using other ancillary data.

Hawley and Moelling (2005) conducted a detailed comparative study of the four most widely used methods: the area-weighting method (Lam 1983), the pycnophylactic method (Tobler 1979), the binary dasymetric method using 2-D remotely sensed data (Wright 1936; Fisher and Langford 1995), and the road network method using 1-D TIGER/Line data (Xie 1995). It was found that intelligent methods usually achieved better accuracy than simple methods, with the 1-D overlaid network as ancillary data resulting in the best accuracy for population interpolation in their study area.

Intelligent methods using 2-D ancillary data or 1-D road network data are not always applicable to all studies. Sadahiro (2000) pointed out that 2-D land use information employed in many dasymetric methods is often derived from remote sensing images, which are not always available or might be expensive, especially for developing countries. In addition, the computational cost tends to be problematic because great quantities of detailed polygons, raster cells, or both have to be processed. The 1-D road network-based intelligent approach often employs road information such as TIGER/Line data to partition the attribute data along the street segments, which also increases computation time due to the large volume of segments to be processed. Such applications rely on the theoretic assumption that human residents are mainly located along the sides of roads. Consequently, the applicability of the 1-D network-based approach for estimating variables other than population has not yet been explored (Xie 1995). Further, the implementation of both 2-D and 1-D intelligent methods relies on the overlay operation, a complex topological process that further increases the computational expense of these methods, especially given the large quantity of data usually involved.

In this article, we propose a point-based approach that uses 0-D points as ancillary data to help transfer information from the source zone to the target zone. These point data therefore

serve as control points, which help characterize the underlying distribution of the variable of interest. This proposed approach has several advantages, as follows. First, 0-D ancillary data have a much simpler data structure compared with 2-D and 1-D spatial data. Second, point data are widely available from various databases such as school sites, business centers, supermarkets, fire stations, and hazardous sites. They allow for the interpolation of other variables besides human population. Third, the proposed approach is based on a simpler geographic information systems (GIS) operation, straight-line distance analysis, rather than the overlay process, which makes the technique more computationally efficient than 2-D and 1-D intelligent methods. Finally, the new technique makes an intelligent areal interpolation possible when 2-D or 1-D ancillary data relevant to the variables to be interpolated are not available or applicable.

Previous Methods

To evaluate the proposed approach, five areal interpolation methods that have been widely cited in the past are employed as benchmarks. The basic algorithms underlying these methods can be described as follows.

Area-Weighting Method

Lam (1983) reported the area-weighting method as the simplest algorithm for performing areal interpolation. Albeit very straightforward, it is a volume-preserving algorithm that conserves the total value within each zone. This method can be formulated in the following steps. First, the density \hat{D}_s (e.g., population density) of a variable (e.g., population) in each source zone is calculated as:

$$\hat{D}_s = \frac{U_s}{A_s} \quad (1)$$

where U_s and A_s refer to the value (e.g., the number of population) and area of source zone s (e.g., the area of a census tract), respectively. Then the total number for a target zone t (V_t) (e.g., a ZIP code tabulation area) is estimated by performing a weighted summation of the density values of all the source zones falling in the target zone using the overlapping area

between the source zone s and target zone t (A_{ts}) as the weight:

$$V_t = \sum A_{ts} \hat{D}_s \tag{2}$$

where A_{ts} is the overlapping area between the source zone s and target zone t . The area-weighting method can be implemented in both vector and raster GIS environments. From Equation 1, we can see that this algorithm is based on the assumption that the density within each source zone is a constant. This method is probably the only choice when additional information is unavailable for the interpolation area, even though the assumption of homogeneity is rarely met in the real world (Xie 1995).

Pycnophylactic Method

The pycnophylactic method was proposed by Tobler (1979) for isopleth mapping and was applied to areal interpolation by Goodchild and Lam (1980). This method was developed in a raster GIS calculation environment. It first computes the density of the variable within each source zone. The computation can be formulated as:

$$\hat{D}_{si} = \frac{U_s}{N_s} \tag{3}$$

where \hat{D}_{si} is the value for cell i , and U_s and N_s are the value and the number of cells within source zone s , respectively. This process is the same as the area-weighting approach. The pycnophylactic approach, however, then smooths the density for each cell to combine the impacts of the adjacent neighbors on its grid value by using the following equation:

$$\hat{D}'_{si} = \frac{\sum_{i=1}^{N_n} \hat{D}_{si}}{N_n} \tag{4}$$

where N_n is the number of neighbors of cell i . The smoothed density \hat{D}'_{si} for each cell is often computed as the average of its four orthogonal neighbors or all of its immediate eight neighbors in the 3-by-3 smoothing window. Theoretically, however, the size of the smoothing window can be customized with any odd or even number based on the characteristics of the underlying data. To keep the volume-preserving

property or to meet the pycnophylactic condition, each \hat{D}'_{si} is adjusted with an iterative procedure that continues until there is either no significant difference between the estimated values and actual values within the source zones or there have been no significant changes of cell values from the previous iteration. To save computation time, Mennis (2003) and Langford (2006) suggested the following one-step approach to avoid multiple iterations of calculation:

$$\hat{D}''_{si} = \hat{D}'_{si} \frac{U_s}{\sum_{i=1}^{N_t} \hat{D}'_{si}} \tag{5}$$

The interpolated value for each target zone then can be expressed as:

$$V_t = \sum_{i=1}^{N_t} \hat{D}''_{si} \tag{6}$$

where N_t is the number of cells within target zone t .

By creating a smoothing surface, this method does not confine itself to the homogeneity assumption of the area-weighting approach, which offers some compromise between homogeneity and heterogeneity. It also conserves the original value of each source zone and is an improvement over the area-weighting method but makes no attempt to use the abundant information available in ancillary data.

Dasymetric Methods Using 2-D Ancillary Data

Wright (1936), possibly the first to propose the dasymetric method, used topological maps as ancillary data for estimating the population of Cape Cod to address the uneven distribution problem in areal interpolation. According to Hawley and Moellering (2005), Fisher and Langford (1995) were the first to publish a dasymetric areal interpolation method using 2-D land use data as control zones. If the land use information is simply the binary divide between related (e.g., residential) and unrelated (e.g., nonresidential) information to the variable of interest (e.g., population), the method is called the binary dasymetric model. It is formulated

by the following equations:

$$\hat{D}_c = \begin{cases} \frac{U_s}{\sum A_{sc}} \\ 0 \end{cases} \quad (if \sum A_{sc} = 0) \quad (7)$$

$$V_t = \sum B_{tc} \hat{D}_c \quad (8)$$

where \hat{D}_c is the density of the variable within the control zone, U_s is the value of source zone s , and A_{sc} is the overlapping area between the source zone s and control zone c . V_t is the interpolated value for target zone t , and B_{tc} is the overlapping area between target zone t and control zone c . This approach assumes that the variable might not be evenly distributed in the source zone but is evenly distributed in the control zones within a source zone.

The binary dasymetric method considers only two land use types (i.e., one related to the variable of interest, and the other not), but if more than two land use types are used, a regression model is required to derive the density for each land use type. The following equation gives the regression-based dasymetric method using land use control zone c with M classes (Langford, Magnire, and Unwin 1991; Langford and Unwin 1994; Yuan, Smith, and Limp 1997):

$$U_s = \sum_{m=1}^M A_{scm} \hat{D}_{cm} \quad (9)$$

where A_{scm} and \hat{D}_{cm} are the area and the density of land use class m within source zone s , respectively. D_{cm} is derived by using the ordinary least-squares fitting method with the corresponding matrix algebra expressed by Equations 10 and 11:

$$[U_{sn}]_{N \times 1} = [A_{snm}]_{N \times M} \times [\hat{D}_m]_{M \times 1} \quad (10)$$

$$[\hat{D}_m]_{M \times 1} = \left([A_{snm}]_{M \times N} [A_{snm}]_{N \times M}^T \right)^{-1} \times [A_{snm}]_{M \times N} \times [U_{sn}]_{N \times 1} \quad (N \geq M) \quad (11)$$

All of the source zones will be involved in the derivation of the density of each class in the control zone when the global regression method is used. N is the total number of source

zones, and it must be greater than or equal to M . The derived matrix \hat{D}_c in the left side of Equation 11 cannot be used directly for the target zones' estimations if the volume of each source zone needs to be conserved. To achieve volume preservation for each source zone, each \hat{D}_c needs to be adjusted. This preservation can be implemented with a rescaling operation similar to that in Mennis (2003) and Langford (2006):

$$\hat{D}'_{cm} = \hat{D}_{cm} \times \frac{U_s}{U'_s} \quad (12)$$

where \hat{D}'_{cm} is the new density estimation for class c in source zone s ; U'_s is the estimated value of source zone s using Equation 10. \hat{D}'_{cm} is then used to aggregate the value of the target zone, in a similar manner to that of the binary dasymetric method.

Road Network Method Using 1-D Ancillary Data

The road network-based intelligent method using 1-D road network data as ancillary information was developed by Xie (1995), who proposed three algorithms: the network length method, the network hierarchical weighting method, and the network house-bearing method. The network length method is the most straightforward and widely used approach because no detailed information on road classes and housing distributions along the roads is needed. The network length method is given as follows:

$$\hat{D}_s = \begin{cases} \frac{U_s}{L_s} \\ 0 \end{cases} \quad (if L_s = 0) \quad (13)$$

$$V_t = \sum L_{ts} \hat{D}_s \quad (14)$$

where \hat{D}_s is the value per unit length, L_s is the total length of the segments within a source zone, and L_{ts} is the total length of the network segments within the overlapping area between the source and target zone.

The notion of incorporating relevant ancillary information for characterizing the heterogeneous nature of variable distributions in areal interpolation provides the rationale for the dasymetric method that uses 2-D land uses

or 1-D road networks as control zones or lines. So far, most of the dasymmetric approaches in the literature are used to interpolate only human population. For other types of variables, such as toxic gas concentration caused by point sources in pollution studies, these 2-D land uses or 1-D road networks are not always applicable, because they might not be directly relevant, or may even be completely unrelated, to the variable being interpolated. As discussed previously, spatial data with 2-D or 1-D ancillary information also often involve complex topological vector data structures, which demands greater computation time for areal interpolation calculation.

Point-Based Intelligent Method

We propose an alternative approach to areal interpolation based on 0-D point data. Compared with 2-D polygon and 1-D line data, point data have a much simpler data structure and do not require extra storage of topological information to represent spatial relationships. The algorithms developed in this approach employ primarily a simple data processing operation, namely, straight-line distance analysis, to estimate the density distribution. Similar to the pycnophylactic approach, the density thus derived can be smoothed if necessary. A unique feature of our point-based intelligent approach to areal interpolation is that it utilizes discrete ancillary point data. It appears that the methodology of the proposed approach is similar to the centroid-based (Martin 1989) and geostatistics-based (Kyriakidis 2004) areal interpolation in that point data are employed, but it is not exactly same as those in nature. Unlike the simple centroid-based approach that utilizes the population information of source zone as weight, the proposed approach is an intelligent method that integrates the existing point ancillary data with the population information, while preserving the total population count in the source zone. The geostatistics-based approach makes use of the systematic point data derived from the grid cells or image pixels to perform area-to-point interpolation using a cokringing procedure as suggested by the publications currently available, whereas our proposed approach employs discrete point data obtained from a GIS data set to conduct a point-to-area interpolation us-

ing a simple Euclidean distance function. In our intelligent point-based method, a linear or exponential equation was used to model the relationship between density of the variable of interests and the location of the control points. The adoption of the linear or exponential functions is based on techniques used by the urban population density modeling widely cited in the literature. According to Martori and Suriñach (2002), two classical models were often used to model population density in urban areas. One is based on the linear function proposed by Stewart (1947) and another is an exponential model proposed by Clark (1951). The detailed algorithms of this approach are presented here, first without a smoothing scheme and then with a smoothing scheme.

Without Smoothing Scheme

The proposed point-based approach assumes that control points are locationally related to the concentration of a variable of interest. For example, schools, supermarkets, and business centers are often close to population centers to best serve the residents around them. However, toxic release inventory sites, landfills, and hazardous material processing plants are often situated away from neighborhoods with high population densities so that the pollution impacts on people can be minimized. Our approach aims to model these locational relationships to characterize the underlying distribution of a variable of interest. The model can be implemented in a vector or a raster environment. For simplicity, we used a raster environment only as an example to explain our approach in this article, but the vector implementation is very similar in principle. The density of a variable can be estimated by this linear function:

$$\hat{D}_{si} = a_s W_{si} \quad (W_{si} \in [0, 1]) \quad (15)$$

The density can also be characterized by a nonlinear exponential function in the form:

$$\hat{D}_{si} = a_s e^{W_{si}} \quad (W_{si} \in [0, 1]) \quad (16)$$

where \hat{D}_{si} is the estimated density value for cell i within source zone s , a_s is a constant parameter for source zone s , and W_{si} is the weight of cell i within source zone s . W_{si} is calculated

based on the inverse distance weighting:

$$W_{si} = \left(1 - \frac{\lambda_{si}}{\lambda_{s \max}}\right)^q \quad (q \geq 1) \quad (17)$$

where λ_{si} is the distance from cell i within source zone s to the closest control point, which is assumed to have the largest influence, although it might not be always absolute in dominance. This control point does not have to be inside the source zone within which cell i is located. $\lambda_{s \max}$ is the maximum values of λ_{si} for all the cells within source zone s ; q is the power parameter that controls the degree of local influence. Higher q values suggest that the rate of change in density values is higher near a cell. The smaller the distance λ_{si} , the larger the influence this control point has. If the cell is at the furthest distance ($\lambda_{s \max}$), the weight (W_{si}) becomes 0, with no influence in determining the density value for this cell. The calculation of distance from each cell to its closest control point can be easily derived using simple GIS techniques such as the straight-line distance function in a raster environment or the point distance function in a vector environment. If the distribution of the variable is just distance weighted instead of inverse distance weighted, where faraway control points have larger influence on the density than near ones, a different weighting strategy can be employed. For example, garbage processing plants are often far away from residential areas. Therefore, the population interpolation using garbage processing plants as control points will use the following formula for the weight parameter:

$$W_{si} = \left(\frac{\lambda_{si}}{\lambda_{s \max}}\right)^q \quad (18)$$

The constant parameter a_s in Equations 15 and 16 can be derived respectively with the following formulas:

$$\left. \begin{aligned} \hat{D}_{s,1} &= a_s W_{s1} \\ \hat{D}_{s,2} &= a_s W_{s2} \\ &\dots \\ \hat{D}_{s,N_s} &= a_s W_{sN_s} \end{aligned} \right\} \Rightarrow a_s$$

$$= \frac{\sum_{i=1}^{N_s} \hat{D}_{si}}{\sum_{i=1}^{N_s} W_{si}} = \frac{U_s}{\sum_{i=1}^{N_s} W_{si}} \quad (19)$$

$$\left. \begin{aligned} \hat{D}_{s,1} &= a_s e^{W_{s1}} \\ \hat{D}_{s,2} &= a_s e^{W_{s2}} \\ &\dots \\ \hat{D}_{s,N_s} &= a_s e^{W_{sN_s}} \end{aligned} \right\} \Rightarrow a_s$$

$$= \frac{\sum_{i=1}^{N_s} \hat{D}_{si}}{\sum_{i=1}^{N_s} e^{W_{si}}} = \frac{U_s}{\sum_{i=1}^{N_s} e^{W_{si}}} \quad (20)$$

where U_s and N_s are the value of source zone s and the number of cells within it, respectively. Finally, the value for a target zone is computed as the sum of \hat{D}_{si} , similar to Equation 6. This approach is volume preserving, a property that most users generally prefer.

With Smoothing Scheme

The pycnophylactic areal interpolation proposed by Tobler (1979) used a smoothing density function to incorporate the effects of adjacent source zones. To make it possible to compare with Tobler's method, we also implemented a similar smoothing scheme using a moving window. The smoothed density value for cell i within source zone s (\hat{D}'_{si}) can be computed as the average of the original density values (\hat{D}_{si}) derived from Equation 15 or 16 for all the neighboring cells in the smooth window:

$$\hat{D}'_{si} = \frac{\sum_{i=1}^{N_n} \hat{D}_{si}}{N_n} \quad (21)$$

where N_n is the total number of the smoothing window (e.g., $N_n = 9$ for a 3-by-3 window). To meet the pycnophylactic condition, a rescaling procedure similar to Equation 5 is used to derive the adjusted density \hat{D}''_{si} . The final interpolated value for any target zone can be computed as the sum of \hat{D}''_{si} , which is similar to Equation 6. By using the smoothing scheme, the impacts of its adjacent neighbors to each cell are taken into consideration, so that outliers resulting from the interpolation process can be reduced. It is also a volume-preserving method when the scale operation is applied.

Case Study

The typical application of areal interpolation is often the residential population estimation from one areal unit to another. The U.S.

Table 1 Descriptive statistics of population in the source zone

Minimum	1,832
Maximum	11,868
<i>M</i>	5,784
<i>SD</i>	2,058
Total	491,675

Census Bureau provides population data for a wide range of areal units, an ideal data set to test and evaluate alternative methods.

We performed areal interpolation on the population data for a suburbanized county, Collin County, located in the state of Texas as our experimental area. As the part of Dallas/Fort Worth metroplex, this area has been experiencing explosive growth in recent years. The U.S. Census Bureau estimated that the population has increased about 48.6 percent from 2000 to 2007 in this region, which includes eighty-five census tracts, thirty-two ZIP code tabulation areas, 196 schools, and 31,951 road network segments (Figure 1). In the experiment, census tracts were used as source zones, and ZIP code tabulation areas were used as target zones. This is a suitable framework for an interpolation experiment, because the units are not nested but rather the source zones and target zones intersect, a common practice often required in applied demographic work. The surveyed population values were available for both source zones and target zones, which were used to derive areal interpolation results and to evaluate model accuracy. The descriptive statistics of the population for the source zone are listed in Table 1. In 2000, there were 491,675 people living in this county. The minimum population among all census tracts was 1,832 and the maximum was 5,784. Land use data from the North Central Texas Council of Governments were employed to evaluate the dasymetric method using 2-D ancillary information. To evaluate the binary dasymetric model, these land use data were grouped into two classes: residential regions and nonresidential regions. To test the dasymetric regression model, it was regrouped into three classes: high residential regions, low residential regions, and nonresidential regions. For the network-based intelligent approach, TIGER/Line network data were used as 1-D ancillary data. To examine our proposed

point-based intelligent method, we chose to use schools as the control points. The site of a school is often selected based on a variety of factors, such as location, environment, safety, and zoning, among which the location and environment factors are regarded as the most important. School sites are always located to minimize the travel distance for most students, and they must be away from sources of noise, air, water, and soil pollutions, where population concentration is also sparse. Because schools are usually located close to the concentration centers of the population and serve as community centers, their locations provide ancillary information for residential population distribution (California Department of Education 1998; Georgia Department of Education Facilities Services Unit 2003; Ewing, Schroeder, and Greene 2004; Springer 2007).

To evaluate the results of these methods, five error measures of global fit were used: mean absolute error (MAE), mean absolute percentage error (MAPE; Goodchild and Lam 1980), population-weighted mean absolute percentage error (PWARPE; Qiu, Woller, and Briggs 2003), root mean squared error (RMSE), and adjusted RMSE (Adj-RMSE; Gregory 2000). To assess the accuracy at the individual zone level, we also calculated the absolute percentage error (APE) and reported the range between the minimum and maximum APE for each method, so that the stability of the interpolation models can be evaluated.

In the literature, researchers usually used only one or two of these error measures. For example, Fisher and Langford (1995) used RMSE to quantify the error introduced by various areal interpolation methods and Gregory (2000, 2002a, 2002b) used Adj-RMSE to examine the accuracy of several areal interpolation techniques suitable for use with historical data. The use of different error measures in these studies makes comparison of their results difficult. We therefore calculated all error measures for this experimental area, with the hope that future studies can use this as a benchmark for comparison of different areal interpolation approaches. These measures are calculated as follows:

$$MAE = \frac{\sum_{i=1}^{N_t} |V'_i - V_i|}{N_t} \quad (22)$$

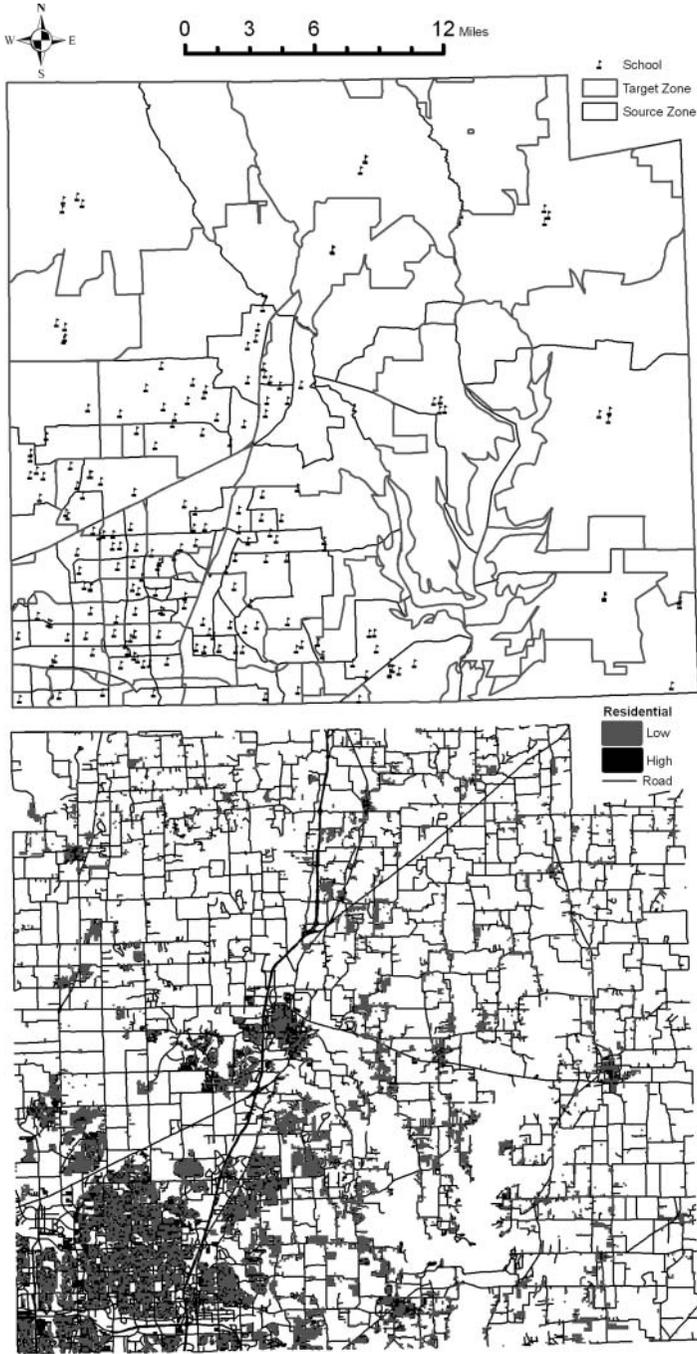


Figure 1 The experimental area, Collin County in Texas, and related data. The top map is the census tract (source zone) boundary and ZIP code tabulation area (target zone) boundary, and the bottom map is the residential area and TIGER/Line data.

$$MAPE = 100 \times \frac{\sum_{t=1}^{N_t} \frac{|V'_t - V_t|}{V_t}}{N_t} \quad (23)$$

$$PWMAPE = 100 \times \frac{\sum_{t=1}^{N_t} \frac{V'_t |V'_t - V_t|}{V_t}}{\sum_{t=1}^{N_t} V'_t} \quad (24)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N_t} (V'_t - V_t)^2}{N_t}} \quad (25)$$

$$Adj_RMSE = \sqrt{\frac{\sum_{t=1}^{N_t} \left(\frac{V'_t - V_t}{V_t}\right)^2}{N_t}} \quad (26)$$

$$APE = 100 \times \frac{|V'_t - V_t|}{V_t} \quad (27)$$

where V_t and V'_t are the original and interpolated value of each target zone, respectively, and N_t is the number of target zones in the study area.

Some of these measures might be affected by the extremes of the population base of the target zones. For example, MAE and RMSE are in the unit of population count and therefore are heavily impacted by the zones with a large population base, because they tend to have a large prediction error in count (Gregory 2000). On the contrary, MAPE and adj_RMSE, in the unit of percentage, are often disproportionately affected by the zones with small population base, because the same amount of error in count will lead to a larger percentage error when the population base is small. PWMAPE uses estimated

population as a weight to normalize the absolute percentage error, having the advantage to eliminate the effects by extreme population bases. This is believed to be the most objective error measure that is not heavily impacted by either high or low population base (Qiu, Woller, and Briggs 2003).

We tested our point-based intelligent method using 0-D school ancillary data, first with linear function without smoothing scheme (AI6), and then with smoothing scheme by a 3-by-3 window (AI7), followed by with exponential function without smoothing (AI8) and with smoothing (AI9). The power parameter q controls the degree of local influence. Like the power parameter in inverse distance-weighted (IDW) point interpolation, this parameter is often set empirically on a trial-and-error basis. Based on several experiments using values of 0.5, 1, 2, 3, and 4, we found that a q value of 2 achieved the best result and reported the results with only q equals 2 in this article.

The overall global accuracy of each method is shown in Table 2. In general, the intelligent methods using ancillary data outperformed area-weighting and pycnophylactic methods without using ancillary data, which was expected. The pycnophylactic method (AI2) had a similar result to the area-weighting approach (AI1). In terms of MAPE, the binary dasymetric method (AI3) generated the best result. The regression dasymetric method with three-class land use data (AI5) produced a poor result, which agrees with the

Table 2 Population areal interpolation errors of different methods

	AI1	AI2	AI3	AI4	AI5	AI6	AI7	AI8	AI9
MAE	1,936	1,998	1,781	1,359	2,398	1,388	1,423	1,655	1,747
MAPE	13.5%	13.8%	9.8%	10.5%	11.6%	10.1%	10.2%	11.9%	12.3%
PWMAPE	7.2%	7.5%	7.4%	5.5%	10.0%	5.7%	5.8%	6.4%	6.8%
RMSE	2,760	2,718	2,436	1,916	2,907	1,639	1,685	2,253	2,300
Adj_RMSE	0.18	0.18	0.13	0.15	0.15	0.13	0.13	0.16	0.16
APE Range	43.98%	44.71%	34.21%	34.43%	34.20%	27.97%	25.74%	32.95%	34.42%

Note: AI1 = area-weighting method; AI2 = pycnophylactic method using a 3-by-3 smoothing window; AI3 = binary dasymetric method using 2-D land use ancillary data; AI4 = network length method using 1-D TIGER/Line ancillary data; AI5 = three-class regression-based dasymetric method using 2-D land use ancillary data; AI6 = point-based intelligent method using 0-D school ancillary data, linear function, without smoothing scheme, and $q = 2$; AI7 = point-based intelligent method using 0-D school ancillary data, linear function, smoothing scheme with a 3-by-3 window, and $q = 2$; AI8 = point-based intelligent method using 0-D school ancillary data, exponential function, without smoothing scheme, and $q = 2$; AI9 = point-based intelligent method using 0-D school ancillary data, exponential function, smoothing scheme with a 3-by-3 window, and $q = 2$; MAE = mean absolute error; MAPE = mean absolute percentage error; PWMAPE = population-weighted mean absolute percentage error; RMSE = root mean squared error; Adj_RMSE = adjusted root mean squared error; APE = absolute percentage error.

conclusions drawn by Langford (2006), who suggested that the benefits of the three-class dasymetric model over a binary model are inconclusive. The network length method (AI4) provided the best result in terms of MAE and PWMAPE. These results are consistent with those obtained by other researchers. Hawley and Moellering (2005) conducted a systematic comparison of the four approaches, including the area-weighting method, the pycnophylactic method, the binary dasymetric method, and the road network approach. Their results obtained from the population interpolation for three counties showed that the road network was the best among these four approaches. Reibel and Bufalino (2005) also suggested that the road network method is a promising areal interpolation technique because it produces more consistent errors when applied to variables at different areal units in a given study area.

When comparing the point-based methods with the preceding algorithms, we can see that our methods (AI6 and AI7) outperformed area-weighting (AI1) and pycnophylactic methods (AI2), suggesting that the incorporation of schools as control points did help to reflect the population distribution information for each source zone. In terms of MAE, PWMAPE, and RMSE, our methods outperformed the dasymetric method (AI3 and AI5) using land use as ancillary data and achieved a comparable accuracy with that of the network length method (AI4). Better accuracy might be achieved by using fire stations or emergency centers that are closer to the population concentration centers. Some schools might be located at the boundary or away from the residential area, which can explain some errors in the interpolation. We did not find much difference between the results of our methods using the smoothing (AI7) and nonsmoothing scheme (AI6) based on the total error measures. Population was also interpolated using the exponential function with the two schemes (AI8 and AI9). The results illustrated that they are comparable with that of the binary dasymetric method, but performed more poorly than the linear function model, suggesting the linear function model was more suitable in this case. By examining the range of APE of each method, we can see the proposed approach had a smaller range of APE, especially the smoothed method with lin-

ear model, implying its stability in performance across all individual zones. The computational cost in seconds for each method is obtained based on a PC with a 3.4 GHz Pentium CPU and 1 GB of RAM (Table 3). We can see that the point-based models (AI6–AI9) are much more efficient than the 1-D and 2-D dasymetric methods (AI3–AI5) and are comparable to simple approaches (AI1–AI2), which makes the 0-D intelligent approach a much more attractive and effective alternative when the areal interpolation is needed for large areas. During our experiments, we also noticed that the 2-D dasymetric model and 1-D network-based method were unstable when executed using ArcGIS 9.3, resulting in many crashes of the system. The crashes are not always consistent, in that the same operation might work sometimes but not at another time or works fine with one computer but not with another one. Because map overlay operations are now standard procedures for GIS, the crashes might reflect the fact that the system implementing the procedures might not be stable when a large quantity of segments, polygons, or raster cells are involved, suggesting a demand for high computer performance for the execution of the overlay procedures and therefore the same for overlay-based 2-D dasymetric and 1-D network methods.

Summary and Discussions

In this article, a point-based method was proposed for the areal interpolation problem using 0-D point data as ancillary information. Compared with 2-D dasymetric methods and 1-D road network-based intelligent methods that use complex data structure and rely on time-consuming GIS overlay operations, the proposed approach was theoretically more efficient because it employed 0-D data without the requirement of topological structure and was implemented with a simple straight-line distance GIS operation. An experimental study in a suburban area illustrated that the proposed areal interpolation approaches took a similar amount of computation time but generated much better accuracy than the simple approaches that do not make use of ancillary data. It also produced comparable results with those using 1-D and 2-D ancillary data but with tremendous saving in computational

Table 3 Computational cost of each areal interpolation method

Algorithm	CPU time (seconds)
A11	18
A12	36
A13	148
A14	158
A15	226
A16	30
A17	63
A18	49
A19	83

Note: The computer used had a 3.4 GHz Pentium processor with 1 GB RAM. A11 = area-weighting method; A12 = pycnophylactic method using a 3-by-3 smoothing window; A13 = binary dasymetric method using 2-D land use ancillary data; A14 = network length method using 1-D TIGER/Line ancillary data; A15 = three-class regression-based dasymetric method using 2-D land use ancillary data; A16 = point-based intelligent method using 0-D school ancillary data, linear function, without smoothing scheme, and $q = 2$; A17 = point-based intelligent method using 0-D school ancillary data, linear function, smoothing scheme with a 3-by-3 window, and $q = 2$; A18 = point-based intelligent method using 0-D school ancillary data, exponential function, without smoothing scheme, and $q = 2$; A19 = point-based intelligent method using 0-D school ancillary data, exponential function, smoothing scheme with a 3-by-3 window, and $q = 2$.

cost. This suggests that the point-based intelligent method might serve as an efficient and effective alternative to the current simple and intelligent areal interpolation techniques, especially in applications with a large study area. The point-based method also has the potential to be used for other types of data; for example, pollutant interpolation using 0-D hazardous sites. Lam (1983) stated that the choice of an appropriate areal interpolation method largely should depend on three factors: the type of variable being interpolated, the degree of accuracy desired, and the amount of computational effort afforded. Given the availability of a variety of digital point data sources, the simplicity of the methodology, as well as the comparable accuracy with the intelligent approaches using 1-D and 2-D ancillary data and the amount of computational time saved in our experiment, we concluded that the point-based methods can provide additional solutions to the areal interpolation problems when the choices are limited by various factors.

The encouraging results from our case study do not imply the ultimate superiority of the point-based intelligent method over other ap-

proaches being compared under all circumstances, however. The fact that different types of ancillary data are required for different approaches did not really make them comparable in a strict sense. Similar to other intelligent approaches, the accuracy of the point-based methods heavily depends on the relative sizes of the source and target zones; the quality, suitability, and strength of relevance of the control points to the variable of interest; as well as the model employed and the associated parameters specified. Sadahiro (1999, 2000) investigated the accuracy of various areal interpolation approaches and asserted that estimation accuracy is improved when source zones are relatively small compared with the target zones; circular and square zones can achieve better results than rectangle zones and the estimates become more inaccurate as the shape of cells and target zones become elongated. Inappropriate ancillary information that is not strongly associated with the variable of interest might actually reduce the accuracy of the areal interpolation. More rigorous studies are needed to further investigate these issues in the future. ■

Literature Cited

- ArcGIS, Version 9.3. Redlands, CA: ESRI.
- Bracken, I., and D. Martin. 1989. The generation of spatial population distribution from census centroid data. *Environment and Planning A* 21:537-43.
- California Department of Education. 1998. *Site selection criteria*. Sacramento, CA: California Department of Education. <http://www.cde.ca.gov/ls/fa/sf/documents/sitecrit.pdf> (last accessed 8 January 2011).
- Clark, C. 1951. Urban population densities. *Journal of the Royal Statistical Society* 114:490-96.
- Deichmann, U. 1996. *A review of spatial population database design and modelling*. Technical Report 96-3. Santa Barbara, CA: National Center for Geographic Information and Analysis.
- Eicher, C., and C. A. Brewer. 2001. Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science* 28:125-38.
- Ewing, R., W. Schroerer, and W. Greene. 2004. School location and student travel: Analysis of factors affecting mode choice. *Transportation Research Record* 1895:55-63.
- Fisher, P. F., and M. Langford. 1995. Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A* 27:11-24.

- . 1996. Modeling sensitivity to accuracy in classified imagery: A study of areal interpolation by dasymetric mapping. *The Professional Geographer* 48:299–309.
- Flowerdew, R. 1988. Statistical methods for areal interpolation: Predicting count data from a binary variable. Research Report No. 15, Northern Regional Research Laboratory, Newcastle upon Tyne, UK.
- Flowerdew, R., and M. Green. 1991. Data integration: Statistical methods for transferring data between zonal systems. In *Handling geographic information: Methodology and potential applications*, ed. I. Masser and M. B. Blakemore, 38–53. London: Longman Scientific and Technical.
- Georgia Department of Education Facilities Services Unit. 2003. *A guide to school site selection*. Atlanta, GA: Georgia Department of Education. http://www.doe.k12.ga.us/documents/schools/facilities/site2_a.pdf (last accessed 8 January 2011).
- Goodchild, M. F., L. Anselin, and U. Deichmann. 1993. A framework for the areal interpolation of socioeconomic data. *Environment and Planning A* 25:383–97.
- Goodchild, M. F., and N. S.–N. Lam. 1980. Areal interpolation: A variant of the traditional spatial problem. *Geo-Processing* 1:297–312.
- Gregory, I. N. 2000. An evaluation of the accuracy of the areal interpolation of data for the analysis of long-term change in England and Wales. Paper presented at the Fifth International Conference on GeoComputation, Greenwich, UK.
- . 2002a. The accuracy of areal interpolation techniques: Standardizing 19th and 20th century census data to allow long-term comparisons. *Computers, Environment and Urban Systems* 26:293–314.
- . 2002b. Time-variant GIS databases of changing historical administrative boundaries: A European comparison. *Transactions in GIS* 6:161–78.
- Harris, R., and Z. Chen. 2005. Giving dimension to point locations: Urban density profiling using population surface models. *Computers Environment and Urban Systems* 29:115–32.
- Harvey, J. T. 2002a. Population estimation models based on individual TM pixels. *Photogrammetric Engineering and Remote Sensing* 68:1181–92.
- . 2002b. Estimating census district populations from satellite imagery: Some approaches and limitation. *International Journal of Remote Sensing* 2: 2071–95.
- Hawley, K., and H. Moellering. 2005. A comparative analysis of areal interpolation methods. *Cartography and Geographic Information Science* 32: 411–23.
- Holt, J. B., C. P. Lo, and T. W. Hodler. 2004. Dasymetric estimation of population density and areal interpolation of census data. *Cartography and Geographic Information Science* 31:103–21.
- Kyriakidis, P. C. 2004. A geostatistical framework for area-to-point spatial interpolation. *Geographical Analysis* 36:259–89.
- Kyriakidis, P. C., and E.-H. Yoo. 2005. Geostatistical prediction and simulation of point values from areal data. *Geographical Analysis* 37:124–51.
- Lam, N. S.–N. 1983. Spatial interpolation methods: A review. *The American Cartography* 10:129–49.
- Langford, M. 2006. Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems* 30:161–80.
- Langford, M., D. J. Magnire, and D. J. Unwin. 1991. The areal interpolation problem: Estimating population using remote sensing in a GIS framework. In *Handling geographical information*, ed. I. Masser and M. Blakemore, 55–77. London: Longman.
- Langford, M., and D. J. Unwin. 1994. Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal* 31:21–26.
- Liu, X. H., P. C. Kyriakidis, and M. F. Goodchild. 2008. Population-density estimation using regression and area-to-point residual kriging. *International Journal of Geographical Information Science* 22: 431–47.
- Markoff, J., and G. Shapiro. 1973. The linkage of data describing overlapping geographical units. *Historical Methods Newsletter* 7:34–46.
- Martin, D. 1989. Mapping population data from zone centroid locations. *Transactions of the Institute of British Geographers* 14:90–97.
- . 1996. An assessment of surface and zonal models of population. *International Journal of Geographical Information Systems* 10:973–89.
- Martori, J. C., and J. Suriñach. 2002. Classical models of urban population density. The case of the Barcelona metropolitan area. In *Urban regions: Governing interacting, economic, housing and transport systems*, ed. J. Van Dick, P. Elhorst, J. Oosterhaven, and J. E. Wever, 109–23. Utrecht, The Netherlands: Koninklijk Nederlands Aardrijkskundig Genootschap, Universiteit Utrecht.
- Mennis, J. 2003. Generating surface models of population using dasymetric mapping. *The Professional Geographer* 55:31–42.
- Mennis, J., and T. Hultgren. 2006. Intelligent dasymetric mapping and its application to areal interpolation. *Cartography and Geographic Information Science* 33:179–94.
- Qiu, F., K. L. Woller, and R. Briggs. 2003. Modeling urban population growth from remote sensed imagery and TIGER GIS road data. *Photogrammetric Engineering & Remote Sensing* 69: 1031–42.
- Reibel, M., and A. Agrawal. 2007. Areal interpolation of population counts using pre-classified land cover data. *Population Research and Policy Review* 26:619–33.

- Reibel, M., and M. E. Bufalino. 2005. Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environment and Planning A* 37:127–39.
- Sadahiho, Y. 1999. Accuracy of areal interpolation: A comparison of alternative methods. *Journal of Geographical Systems* 1:323–46.
- . 2000. Accuracy of count data estimated by the point in polygon method. *Geographical Analysis* 32:64–89.
- Springer, D. 2007. *Integrating schools into healthy community design*. Issue Brief. Washington, DC: National Governor's Association. <http://www.nga.org/Files/pdf/0705schoolshealthydesign.pdf> (last accessed 8 January 2011).
- Stewart, J. Q. 1947. Empirical mathematical rules concerning the distribution and equilibrium of population. *Geographical Review* 24:461–85.
- Turner, A., and S. Openshaw. 2001. Dissaggregative spatial interpolation. Paper presented at GISRUK 2001, Glamorgan, Wales.
- Tobler, W. 1979. Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association* 74:519–30.
- Tobler, W., U. Deichmann, J. Gottsegen, and K. Maloy. 1995. The Global Demography Project. Technical Report TR–95–6, National Center for Geographic Information and Analysis, Santa Barbara, CA.
- Wright, J. 1936. A method of mapping densities of population: With Cape Cod as an example. *Geographical Review* 26:103–10.
- Wu, C., and A. Murray. 2005. A cokriging method for estimating population density in urban areas. *Computers, Environment and Urban Systems* 29:558–79.
- Xie, Y. 1995. The overlaid network algorithms for areal interpolation problem. *Computers, Environment and Urban Systems* 19:287–306.
- Yuan, Y., R. M. Smith, and W. F. Limp. 1997. Remodeling census population with spatial information from Landsat TM imagery. *Computers, Environment and Urban Systems* 21:245–58.
- CAIYUN ZHANG is an Assistant Professor in the Department of Geosciences at the Florida Atlantic University, Boca Raton, FL 33431. E-mail: czhang3@fau.edu. Her research interests include spatial analysis and modeling, GIS application software development, remote sensing digital image processing, LiDAR, and ocean remote sensing.
- FANG QIU is an Associate Professor in the Program of GIS at the University of Texas at Dallas, Richardson, TX 75083. E-mail: ffqiu@utdallas.edu. His research interests include remote sensing digital image processing, spatial analysis and modeling, GIS application software development, and Web-based mapping and information processing.