Population Estimation Methods in GIS and Remote Sensing: A Review

Shuo-sheng Wu, Xiaomin Qiu, and Le Wang
Department of Geography, Texas State University–San Marcos, 601 University Drive, San Marcos, Texas 78666-4616

Abstract: This paper reviews existing population estimation methods in the GIS and remote sensing literatures. The methods can be grouped into two categories: areal interpolation methods and statistical modeling methods. Areal interpolation methods can be further separated into two categories depending on whether ancillary information is used. Statistical modeling methods can be further grouped into five categories based on the relationship between populations and urban areas, land uses, dwelling units, image pixel characteristics, or other physical or socioeconomic characteristics.

INTRODUCTION

Many methods for population estimation have been reported in the GIS and remote sensing literatures. Depending on the intended goal and the required information, these methods can be grouped into two categories: areal interpolation and statistical modeling. Areal interpolation methods are primarily designed for the zone transformation problem that involves transforming data from one set of spatial units to another. This approach uses census population data as the input and applies interpolation or disaggregation techniques to obtain a refined population surface. In contrast, the statistical modeling approach is more interested in inferring the relationship between population and other variables for the purpose of estimating the total population for an area. The statistical modeling approach does not directly use census data as the input. Rather, it makes use of socioeconomic variables and applies theories in urban geography for population estimation; census population data only participate in the model training process. This approach is originally designed to estimate the intercensal population or population of an area difficult to enumerate, though it can also be incorporated into the process of interpolating census population. Before reviewing these two approaches of population estimation, we would like to review the early population density models from urban geography.

POPULATION DENSITY MODELS FROM URBAN GEOGRAPHY

The simple gravitational population density model from urban geography is the heart of what has been called social physics (Stewart and Warntz, 1958). Although many people have noticed the decrease of population density from inner city to outer

GIScience and Remote Sensing, 2005, **42**, No. 1, p. 58-74.

suburbs, it was Clark (1951) who first associated this observation with specific mathematical functions (Liu, 2003) in the following negative exponential function:

$$d\,(r) = K * e^{-\lambda r}, \tag{1}$$

where $d(r)$ is the population density at distance r from the center of the city ($r = 0$); K is a constant that equals the central density $d(0)$; M describes the rate of decline of density. This relationship has been demonstrated to exist for many cities of the United States (Weiss, 1961), as well as for many cities outside the U.S. (Newling, 1965). Although the goodness of fit varies, the model always holds statistically significantly in every place studied. Kramer (1958) also incorporated a sectoral model of a city and showed how various models of urban forms interact with each other.

Some studies have explored other mathematical forms to describe the relationship between population density and location. For example, Sutton et al. (1997) examined the Gaussian and the parabolic forms and found that both are statistically significant. Some studies have criticized the use of the negative exponential function. For example, Batty and Longley (1994) stated that the exponential population density function has been used solely for its convenience and elegance, rather than its appropriateness to empirical data. Parr (1985) suggested that an inverse power function is more appropriate to the urban fringe and hinterland, and the negative exponential function is more appropriate for describing density in the urban area. A similar suggestion to make a modification at the fringe of the urban area is also found in Tobler's (1999) comment on Martin's (1996) population interpolation algorithm. Tobler pointed out that "the exponential distance decay function is a relevant approximation for the whole of an urban area, yet its repeated use farther out from the urban center hardly seems reasonable. In the periphery, far from the center, the density gradient is much more nearly linear" (1999, p. 85). To correct this problem, Tobler proposed a "tent function" by first decomposing each census unit into triangles, with one vertex being the geometric centroid of the unit; populations inside triangles are then assigned based on the coordinates and population of the vertices.

AREAL INTERPOLATION METHODS

The negative exponential function can only be regarded as the empirical results showing how population is distributed in urban areas. It has not been used, in existing literatures, for practical population estimation that concerns accuracy. Instead, most studies used areal interpolation or statistical modeling methods for population estimation. Areal interpolation, as mentioned above, is primarily designed for zone transformation that involves transforming data from one set of spatial units to another. The two sets of spatial units could be referred as the source zone and the target zone (Lam, 1983). The general strategy for zone transformation is to apply certain areal interpolation operations to transform source zone data to finer-scale raster data and then aggregate them for target zones. In the context of population interpolation, census data are the vector-based source zone data and are interpolated to finer-scale raster data by a certain interpolation method. Areal interpolation is subject to errors from original areal aggregation. The quality of the interpolation estimates depends largely on how source zones and target zones are defined, the degree of generalization in interpolation process, and the characteristics of the partitioned surface (Lam, 1983). Areal interpolation methods can be further separated into two categories depending on whether ancillary information is used.

Areal Interpolation without Ancillary Information

For areal interpolation methods without ancillary information, there are point-based methods and areal-based methods (Lam, 1983). In point-based interpolation, a control point is assigned to represent each source zone and a grid map is generated with grid point values estimated from control points. In contrast, area-based interpolation uses the source zone itself as the unit of operation instead of arbitrarily assigned control points. Also, area-based interpolation is more concerned with volume preservation; i.e., the summation of population data to the original set of areal units is preserved in the transformation to a new set of areal units. Based on theoretical and limited empirical evidence, volume preservation is an essential requirement for accurate interpolation estimates (Lam, 1983).

Point-Based Methods. There have been many point-based interpolation methods developed in the past. Some researchers put such methods into two groups, global and local, depending on whether they consider all of the data values at once or the values within a pre-defined neighborhood of each point. Here we adopt Lam's (1983) approach to group point-based methods into exact methods and approximate methods, depending on whether they are concerned with preserving the original sample point values or with determining an overall surface function $f(x, y)$. The reason for this categorization is that whether interpolation methods preserve original data values on the inferred surface is fundamental in analyzing their accuracy (Lam, 1983).

The exact methods include interpolating polynomials, most distance-weighting methods, kriging, spline functions, and finite difference methods, while the approximate methods include power-series trend models, Fourier series models, distance-weighted least squares, and least-squares fitting with splines. Each of these methods has its own advantages and disadvantages, and none of them is superior to all others for all applications (Lam, 1983). Furthermore, the results from all the methods are seriously affected by the quality of the original data, especially the density and the spatial arrangement of data points, and the complexity of the surface. The choice of an appropriate interpolation method depends largely on the type of data, the degree of accuracy desired, and the amount of computational effort afforded. In general, exact methods are more reliable than approximate methods because of the former's simplicity, flexibility, and reliability (Lam, 1983).

One of the point-based methods widely used in the UK census is a kernel-based interpolation proposed by Martin (1989) (Bracken, 1991; Martin and Bracken, 1991; Bracken and Martin, 1989). This method uses a source zone centroid as the control point. A window is positioned over each control point in turn and the source zone population is allocated to grid cells falling inside the window using a unique weighting based on the distance decay function between the source zone centroid and the grid cell.

Point-based areal interpolation methods experience a few problems (Lam, 1983; Liu, 2003). First, the use of a control point, usually the centroid or the center of an area, to represent the source zone often introduces errors. The calculation of the centroid or center of an area depends on the coordinates of the points defining the boundary of a source zone. If the source zone is symmetrical and relatively simple, the center or centroid would be a convenient control point, and the estimated value for each grid cell would be reliable. However, if the boundaries are not symmetrical or well generalized, the location of the centroid can be significantly effected, and the interpolation results may be biased. In reality, census units are rarely symmetrical, and the non-uniform distribution of population within a census unit further complicates this issue.

Another problem associated with point-based interpolation methods is that all have some kind of *a priori* assumption about the surface involved. This rather arbitrary assumption rarely fits the complex geographical phenomena in the real world. Nevertheless, it is worth noting that the task of areal interpolation is to search for the best method, whose output is as close to the ground truth as possible. Violation of the assumption of a method only implies that the results obtained may not be optimal, but does not mean that this method is necessarily inferior to others. For example, in the context of kriging interpolation, if the source zones can be reduced to control points and the population distribution can be described by the semi-variogram, kriging is the best linear unbiased estimation. If the assumption is not satisfied, results from kriging interpolation are not necessarily inferior to those from others. Comparative studies using empirical datasets are needed to further research this issue.

The most important problem of point-based methods is that they mostly do not conserve the total value within each source zone. Volume preservation is important in that it gives reliability to the approximation of grid values for source zone, and thus the subsequent estimation for target zone is less subject to error. Besides, in the context of population interpolation, people should not be "destroyed" or "manufactured" during the redistribution process (Langford and Unwin, 1994). To correct this problem, Martin (1996) modified the original kernel-based interpolation algorithm (which is a point-based method) to ensure that the populations reported for target zones are constrained to match the overall sum of the source units.

Area-Based Methods. In contrast to point-based interpolation, area-based methods are volume-preserving methods. The simplest method in this category is the overlay operation based on the geometric properties of the source and target zones. It superimposes the target zone on the source zone to obtain the proportion of each source zone in each target zone. The proportion then serves as a weight and the values of target zones become a weighted linear function of source zones.

The major problem with the overlay method is that it assumes homogeneity within each source zone. Source zones having homogeneous distributions, unfortunately, seldom occur in the real world. This may well be true of some phenomena such as rainfall or agricultural productivity, but is harder to justify for human phenomena such as population. In addition, very often the source zones were originally delineated for other purposes and may not show the important distribution information for the target zones. For these reasons, the reliability of target zone estimates is controlled mainly by the nature and degree of the homogeneity of the source zone and by the size of the target zone in relation to the source zone (Lam, 1983).

Tobler's (1979) pycnophylactic interpolation is probably the most widely quoted area-based interpolation method. This method assumes a smooth density function that takes into account the effect of adjacent source zones while preserving its volume.

The smooth condition intends to minimize the curvature of estimated surface by requiring the value of any grid point to approach the averages of its four neighbors. Other smoothing conditions may be used depending on the type of application. The interpolation process begins by assigning the mean density to each grid cell superimposed on the source zones, and then modifies this by a slight amount to bring the density closer to the value required by the smoothing condition. The volume-preserving requirement is then enforced by either incrementing or decrementing grid densities within each source zone after each computation. The result is a smooth population density surface.

The original pycnophylactic interpolation uses regular lattice grids as its spatial configuration. Rase (2001) extended it to a surface representation based on an triangular irregular network (TIN). The basic step is to generate a TIN from the boundary network first, and then to interpolate a smooth surface by an iterative procedure, in which the two steps of smoothing and difference distribution are repeated until the threshold for the overall smoothness measured by the relative variance is reached or the maximum number of iterations is exceeded. Compared to the original grid-based method, the TIN-based pycnophylactic interpolation is argued to have several advantages (Rase, 2001), including that the error resulting from converting source zone polygons to a regular grid is avoided, and that the TIN-based method is more suitable for fast display in real-time applications. On the other hand, the TIN model is more difficult to implement because it requires more effort and support for the data and program structures.

Lam (1983) stated that the overlay methods will yield better estimates if the surface is discontinuous, whereas the pycnophylactic method gives better results when smoothness is a real property of the surface. In cases where the surface is intermediate between discontinuous and maximally smooth, different target equations and side conditions should be imposed for reliable results, but such methods are yet to be developed.

Areal Interpolation with Ancillary Information

Population is related to other information, e.g., land use and transportation networks, that can be used to assist population interpolation. This section will review the interpolation methods with ancillary information, particularly those can be extracted from remotely sensed data.

The dasymetric method is the most well-known method in this category. It was originally developed by Wright (1936) out of a concern that choropleth maps do not give a valid representation of population distribution within enumeration units. Wright's idea was to use knowledge of the locality to identify areas within zones that have different population densities, thus allowing refinement of the assumption of an even distribution (Fisher and Langford, 1995). In his population density mapping of Cape Cod, Wright made binary partitions iteratively to disaggregate general zones to detailed zones of population density while making certain that the original zone population was preserved.

In the past, Wright's dasymetric mapping of binary partition was difficult to implement. With the development of digital data and GIS technology, the dasymetric method became easier through use of the GIS overlay process, which also provides the convenience of integrating various types of ancillary spatial data. For example, Monmonier and Schnell (1984) demonstrated the integration of classified residential land use classes from Landsat satellite imagery as the ancillary information in the dasymetric method.

Wright's dasymetric method relies on knowledge of the local areas to determine subzone population densities. Flowerdew and Green (1989) proposed using statistical regression analysis to estimate subzone population densities; yet Langford et al. (1991) first applied multivariable regression techniques to estimate dasymetric subzone population densities. Their approach is based on the following function:

$$P_i = \Sigma P_{ij} = \Sigma A_{ij} * D_j, \quad (2)$$

where P_i is the total population of source zone i; P_{ij} is the total population of land use j within source zone i (subzone ij); A_{ij} is the total area of land use j within source zone I; and D_j is the average population density of land use j. A_{ij} can be obtained by a GIS overlay operation of a land use map and a source zone map. Since there are multiple source zones, multivariable regression can be applied to estimate D_j of multiple land use types. Volume preservation is further maintained by scaling up or down derived density measures to fit the original total population for each source unit.

Despite the ease of implementation, the dasymetric method is still subject to the problem of an even distribution assumption within subzones. In other words, while the difference between subzones is recognized, differences within subzones are ignored. For example, for single-family land use, there is the difference between low-density, medium-density, and high-density zones. To incorporate such a consideration, one may conduct a more detailed land use classification, and associate each land use class with a certain population density. Although this approach could improve population interpolation accuracy, it requires effective ways to classify detailed land use types and to estimate their population densities.

The easiest dasymetric mapping approach with remote sensing–derived land use data is a binary division approach in which land use is classified to "populated" and "unpopulated" and census populations are simply redistributed to those populated areas; some example studies included Holt et al. (2004), Fisher and Langford (1996), and Langford and Unwin (1994). Furthermore, a more specific dasymetric mapping approach would classify a number of land use classes and redistributed census populations to these classes; some example studies include Mennis (2003), Eicher and Brewer (2001), Yuan et al. (1997), and Langford et al. (1991). For the latter group of studies, some ways of determining the population density ratio between land use classes must be applied. Some studies used an empirical sampling approach (e.g, Mennis, 2003), some used pre-defined population density statistics (e.g., Eicher and Brewer, 2001), whereas some used regression analysis to derive population density estimates (e.g., Yuan et al., 1997; Langford et al., 1991). The regression analysis seems to provide a preferred approach because of its objectivity in testing model accuracies through statistical significance tests.

Harvey (2002b; 2000) adopted an extreme approach to deal with the homogeneity assumption within subzones by estimating population density in the spatial unit of pixels. His method first assigned all residential pixels within a source zone with an equal share of the total population in the following equation:

$$P_{ij} = P_i/n,\ i = 1, 2, \ldots, n \tag{3}$$

where P_{ij} is the population initially assigned to the jth pixel in source zone i whose total population is P_i, and n is the number of pre-classified residential pixels in source zone i. Since there were many source zones, each of which had some residential pixels with different digital values, an ordinary least-squares regression could be conducted between the population and the digital value of the pixels. With the regression coefficients obtained, the population of each pixel was adjusted by the following equation:

$$P_{ij}(\text{adj}) = \hat{P}_{ij} + \hat{r}, \tag{4}$$

where $\hat{P}_{ij}$ is the regression estimate, and

$$\hat{r} = \frac{\sum_{j=1}^{n} (P_{ij} - \hat{P}_{ij})}{n}. \tag{5}$$

The result was that the adjusted reference population lay closer to the regression line than the initially assigned population. If the iteration was run again with the adjusted value, the R^2 would be improved. The process was repeated iteratively, and R^2 continued to increase monotonically with decreasing increments, and stopped when a predefined threshold was reached. Harvey proved that this iterated regression procedure is a least-square approximation to the Expectation Maximization (EM) algorithm that was originally presented by Dempster et al. (1977) and applied by Flowerdew and Green (1989, 1991) for combining data from two incompatible sets of spatial zones.

Harvey (2000, 2002b) argued that the pixel-level dasymetric method has several advantages over zone-level dasymetric methods, including: human habitation by individual residences should be better delineated by pixels; the mathematical form of pixel-based model is simple and relatively robust; and implementation and refinement of routine pixel-based classifiers are easier.

We separated areal interpolation methods into two categories depending on whether ancillary information is utilized. The methods using ancillary information, particularly the dasymetric method, usually yield more accurate results than those without ancillary information, assuming the ancillary information reflects the spatial distribution of the variables being mapped. It is worth noting that methods in these two categories can be incorporated with each other for certain purposes. For example, Langford and Unwin (1994) applied a kernel-based smoothing function to the result of a dasymetric method in order to create a cartographically pleasing and informative map, so that the readers won't see too many pixel-level details. Also, most methods in the first category (without ancillary information) can still make use of ancillary information when it is available. For example, in the case of Tobler's pycnophylactic interpolation, if information about residential areas is available, one can first allocate the population of a census unit to residential polygons within it, assuming non-residential polygons have no population, and then perform the smoothing interpolation operation.

STATISTICAL MODELING METHODS

We will review the second category of population estimation methods, the statistical modeling methods, in this section. As reviewed previously, theories in urban geography have demonstrated that population distribution in an urban area is affected by morphological factors such as distance to the central business district (CBD), distance to roads, etc. Many of the morphological factors can be extracted from remotely sensed data. Consequently, remote sensing has been actively explored as a means to study population distribution. Strictly speaking, models from this group are mainly designed to estimate an overall population count rather than population density that is relevant to population distribution. However, since population count and population density can be derived from each other through the size of the area of interest, the method designed to estimate population counts can also be used to estimate population distribution.

Statistical modeling approaches for population estimation started in the 1950s. The initial motivation was to remedy the shortcomings of the decennial population census, such as high cost, low frequency, labor intensity, etc. (e.g., Kraus et al., 1974), but these approaches also have been applied to check the reliability of the census enumeration (e.g., Clayton and Estes, 1980), and the inference of socioeconomic characteristics such as housing value and residential quality (e.g., Forster, 1983). The use of remote sensing in statistical modeling approaches started in the mid-1950s, particularly with the goal of searching for an alternative to a population census. Researchers have conducted various statistical modeling methods for population estimation on different scales with different types of remotely sensed imagery. In general, there are five categories of approaches, based on the relationship between population and (1) urban areas, (2) land use, (3) dwelling units, (4) image pixel characteristics, and (5) other physical or socioeconomic characteristics (Lo, 1986; Liu, 2003).

Correlation with Urban Areas

This category of methods is a general approach based on a functional relationship between urban areas and population size. Inspired by the biological law of allometric growth (Huxley, 1932), Nordbeck (1965) studied the relationship between urban areas and population size of many U.S. cities and concluded that the built-up area (A) of a settlement is proportional to its population (P) raised to some power:

$$A = a * P^b. \tag{6}$$

Tobler (1969) was the first to use satellite imagery to study the relationship between population and urban areas. He used images from the Gemini manned space flight program to study populations of many cities in the world. Assuming that if cities can be considered circular in shape, and if shape varies insignificantly with time, Tobler found the correlation coefficients between radii and populations of 0.87 or higher in the following function:

$$r = a * P^b. \tag{7}$$

The results of his study also indicated that the coefficient (a) and exponent (b) for cities in the United States was comparable to those for cities in Sweden and Canada; cities in Japan and in the Nile Delta, however, had coefficients and exponents reflecting the dense and compact structure of settlements in Asia and the Middle East.

The availability of Landsat satellite imagery and advancement in image processing techniques allowed researchers to efficiently study the relationship between population and urban areas, although one of the major difficulties involved differentiating rural/urban boundaries (Lo and Welch, 1977). Using 1972 to 1974 Landsat MSS imagery of 10 large cities in China with 500,000 to 2,000,000 populations, Lo and Welch (1977) found correlation coefficients of 0.82 or higher between populations and classified urban areas in a modified function from (6):

$$P = a * A^b. \tag{8}$$

This function can be referred as the allometric growth model (Lee, 1989; Lo, 2003), which describes that the relative growth rate of population is proportional to the relative growth rate of the residential land area.

Researchers also used urban light as an indicator for population size. Prosperie and Eyton (2000) found a quite high R^2 of 0.974 between light volumes and populations of 254 Texas counties using DMSP (Defense Meteorological Satellite Program) imagery. Adopting a similar approach but at a smaller scale using cities, Lo (2002) found a correlation coefficient of 0.91 between the light volumes of 35 Chinese cities and their non-agricultural populations. He further evaluated derived population models using data from other 18 Chinese cities and obtained acceptable accuracies.

Correlation with Land Use

The second approach for population estimation is based on correlating population counts with different types of land use areas, which should achieve higher precision than the first approach. The total population for an area can be calculated according to the following function:

$$P = \sum_j A_j * D_j, \tag{9}$$

where P is the total estimated population; A_j is the area of land use j; and D_j is the population density for land use j, which is to be determined through regression analysis. This basic function is similar to that used in the dasymetric method reviewed previously, only that the former intends to disaggregate census population by maintaining the original census unit population count, whereas the latter intends to estimate the total (intercensal) population for an area.

Areas for different types of land use could be extracted from remote sensing images. The accuracy of population estimation would largely rely on how accurate different types of land use are classified. In Weber's (1994) study of population estimation for the City of Strasbourg, France, he classified six types of urban land use from SPOT HRV XS images with a Kappa coefficient of 0.915. Then he ran a regression analysis between population counts and land use areas of 2,812 census units based on function (9) and obtained an R^2 of 0.91. After applying his regression model to estimate the total population for the city, his model estimate was 7.91% below the census population of the city.

In Lo's (2003) study of Project ATLANTA, he classified six types of land use from Landsat TM images and obtained a Kappa coefficient of 0.878. The area of low-density urban use class was then regressed with population counts of 418 census tracts using a logarithmic transformed allometric growth model. The result had an R^2 of 0.68. He then applied the regression model to estimate populations of 373 census tracts and the results had a relative error of 14.80% and a overall underestimate of 8.07%.

Population densities for different types of land use could also be determined from sample surveys or census statistics, in addition to the regression analysis. For example, in Kraus et al.'s (1974) study of population estimation for four California cities, four types of urban land use were first classified. Then the authors calculated the characteristic population densities for each type of land use from sampled census block-level population data. Finally they estimated city populations based on function (9). The results ranged from an underestimate of 9.17% to an overestimate of 7.00% when compared with census populations of the four cities.

Correlation with Dwelling Units

The total population of an area can be estimated by multiplying the total number of dwelling units with the number of persons normally living in a dwelling unit. It is also possible to categorize dwelling units and apply a different persons-per-dwelling unit ratio to each category. This ratio can be derived from sample surveys or calculated from census data with the assumption that a single household occupies one dwelling unit. The total number of dwelling units in an area may be estimated from remote sensing images.

Green (1956) was probably the first researcher to propose using individual dwelling unit counts observed from aerial photographs for population estimation. Porter (1956), however, was the first to actually apply this methodology (Kraus et al., 1974), with the persons-per-dwelling unit ratio established from ground observation in their study of Liberia. Hsu (1971) applied the same methodology for intercensal population estimation of the Atlanta area, but he derived his persons-per-dwelling unit ratio from U.S. census tract data. Collins and El-Beik (1971) and Dueker and Horton (1971) further identified different types of residential buildings from aerial photographs for population estimation, with their population density statistics calculated from census data. To obtain a more accurate persons-per-dwelling unit ratio, Lo and Chan (1980) used a field survey methodology to calculate the average population density for various types of housing. Furthermore, in an effort to automate the time-consuming procedure of counting dwelling units, Lo (1989) used a raster approach to extract residential building density, on a grid cell by grid cell basis, from high-altitude aerial and space photographs. He first calculated the maximum possible occurrence of dwelling units in each grid cell with reference to the dwelling unit size. Then the percentage of occurrence of residential buildings in each grid cell was able to be estimated.

In the past, no effective ways of automatically extracting residential buildings existed. Researchers relied on manually identifying and counting dwelling units from high-spatial-resolution aerial photographs, even though visual interpretation is laborious and time consuming. With the advance of very high spatial resolution satellite images, such as IKONOS and QuickBirds, and the improvement of feature extraction techniques (Haverkamp, 2004), automatic extraction of dwelling units from satellite images has become possible. Another prospect for automatic building extraction is the advancement of 3D object extraction techniques from LIDAR data (Rottensteiner, 2003). With these new remote sensing data and building extraction techniques, population estimation by dwelling unit counts may become a viable approach.

Correlation with Image Pixel Characteristics

Other than the physical characteristics extractable from remotely sensed imagery, population density can also be directly correlated to the spectral reflectance value of image pixels. Hsu (1973) was probably the first to suggest the idea of using imagery pixel values to develop a multiple regression model for population estimation (Lo, 1986). His idea, however, was not implemented until Iisaka and Hegedus's (1982) pioneering work in estimating population distribution of Tokyo, Japan. They reported that the mean spectral values of Landsat MSS bands 4, 6, and 7 were strongly correlated to population density. Lo (1995) adopted a similar approach while using higher resolution imagery of SPOT, and reported a correlation coefficient of –0.91 between population density and the mean spectral values of SPOT band 3 for the Hong Kong area. Webster (1996) argued that the spectral values alone cannot discriminate areas of different population densities effectively. Alternatively, he combined numerous spectral and textural measures from Landsat TM imagery in a regression model and found textural measures have more significant predictive power for housing densities than spectral measures. Harvey (2002a, 2002b) also incorporated a variety of spectral transformation measures, such as the band-to-band ratio and difference-to-sum ratio, in addition to textural measures, in a series of stepwise regression models for population estimation. Further, Harvey (2000; 2002b) developed an innovative iterated regression procedure (reviewed previously as a dasymetric method) to improve the predictive power of a regression model based on pixel spectral values. There are also studies using imagery texture analysis to categorize pixels first and then correlate pixel counts in different categories with population density, which is similar to the approach of inferring population through land use. For example, Chen (2002) used a homogeneity texture measure to categorize pixels of different levels of homogeneity and correlate the number of pixels in each category to housing densities.

Correlation with Other Physical and Socioeconomic Characteristics

Other than the mentioned physical and pixel characteristics extractable from remotely sensed imagery, numerous other physical and socioeconomic variables can also be incorporated for population estimation. A notable example is the LandScan Global Population Project (Dobson et al., 2000), in which light volume from nighttime imagery, land cover derived from various types of remotely sensed imagery, and other information about demography, topography, and transportation networks have all been combined in a model to estimate population at a 30 × 30 second (approximately 1 × 1 km) resolution. Many of the variables can be extracted from remotely sensed imagery. Similar approaches have also been applied in a smaller scale. For example, Liu and Clarke (2002) found that the total population in urban areas is correlated with distance to the CBD, accessibility to the transportation system, slope, and the time when the residential community was first built. Overall, the accuracy and robustness are improved with increasing model complexity. It is worth noting that although multivariable approaches for population estimation tend to improve the overall accuracy compared to methods using a single variable, the selection of variables in the model requires guidance from theories in urban geography.

Many physical and socioeconomic variables can assist the estimation of population, yet only those attributes that can be directly or indirectly observed and extracted from remotely sensed imagery are applicable in the remote sensing context. Residential areas constitute a major component of such analysis. The data are usually of two types: (1) the structural conditions of individual residential units; and (2) attributes reflecting the residential or neighborhood environment. Green (1957) and Green and Monier (1959) pioneered research using aerial photograph to study socioeconomic and demographic variables. They cited an extensive literature to demonstrate that social values are attached to housing and residential communities and, by extension, that observable physical data have meaningful sociological correlations.

Regardless of whether socioeconomic or pixel characteristics are used in statistical modeling for population estimation, all studies inferring population from remotely sensed data have reported a consistent finding, i.e., that small-area population estimation is often not as accurate as large-area estimation. It may be explained that overestimation and underestimation are canceled out for large-area population estimation and thus the overall accuracy is high (Lo, 1995). Nonetheless, more studies are needed before remote sensing can be applied to population estimation on an operational basis.

SUMMARY

Of all the population estimation methods, the dasymetric method is commonly regarded as a more accurate approach, provided that the used ancillary information gives a truthful description of where people actually live. Furthermore, the dasymetric method is not only more accurate, but also relatively stable. It is robust to the variation of population density associated with a certain type of land use, as well as the anomaly of highly urbanized but sparsely inhabited areas (Fisher and Langford, 1996). The reason is because the volume-preserving property preserves the population of the source unit in the transformation to raster representation, and thus all associated errors are inherently limited to variation within each individual source unit.

The dasymetric method used with remote sensing is also robust to imagery classification error. Fisher and Langford (1996) reported that errors of up to 40% in the classified TM image still yield better estimates of the interpolated populations than other regression or surface methods they tested. The reason for the relative robustness of the dasymetric method under classification error is due to the aggregated error within zones. Specifically, even if the classification error is high, the frequency of pixels in different land classes may not vary significantly within a zone. As observed by Donnay and Unwin (2001, p. 220), "even though individual pixels may have a weak probability of being correctly assigned to a land use category, when aggregated into a target zone for density estimation, the relative frequencies within these target zones do not degrade substantially." Relevant empirical studies by Lo (1995), Webster (1996), and Harvey (2002b) also indicated that classification errors at the pixel level can be high without impacting the accuracy of areal population estimates.

In this review, we separated population estimation methods into areal interpolation and statistical modeling. It is worth noting that the statistical modeling approach can also be incorporated into the dasymetric method. For example, Langford et al. (1991) described a dasymetric procedure based on five types of land use classified from TM multispectral imagery, with their average population densities derived from regression analysis. Yuan et al. (1997) also applied a multivariate regression model to correlate census block-group populations with different land use areas classified from Landsat TM images in their dasymetric study.

Studies on population issues generally use census data as the primary data source. The census, however, may not be applicable to the intended purpose of these studies. This is because in many counties, including the United States, the census population figure is actually a de jure population, in contrast to a de facto population. A de jure population reports all usual residents of the given area, whether or not they are physically present there at the reference date. A de facto population, in contrast, reports all persons physically present in the area at the reference date. The U.S. census is a de jure census because it is based on people's home address, rather than where they work or travel during the day, or if people are out of town. The U.S. census is mainly concerned with residential populations and the daytime population distribution can be very different from that described by the census. For example, Las Vegas has a much higher daytime population than that reported by the census because of its high proportion of tourists. Since some applications (e.g., emergency response) require knowledge of daytime population whereas others (e.g., urban growth) require residential population, it is desirable that both types of population be estimated. Unfortunately, to date little research has attempted to model daytime population. Theoretically, urban land use information is more related to daytime than nighttime population because land uses such as industrial, commercial, and recreational provide information about where people are during the day. The methodology of relating nonresidential land use to daytime population is yet to be explored; research on people's traveling behavior may provide some guidance in this area.

This paper reviewed past GIS and remote sensing literatures on population estimation, particularly those making use of remotely sensed data. It is clear from the review that remote sensing provides valuable resources for useful ancillary information. Past studies of population estimation mainly relied on images of relatively coarse spatial resolution. With the availability of high-spatial-resolution commercial images, such as QuickBird and IKONOS, as well as the advancement of image processing techniques, improvement in population estimation accuracies is expected.

REFERENCES

Batty, M. and P. Longley, 1994, *Fractal Cities: A Geometry of Form and Function*, London, UK/San Diego, CA: Academic Press, 394 p.

Baudot, Y., 2001, "Geographical Analysis of the Population of Fast-Growing Cities in the Third World," in *Remote Sensing and Urban Analysis*, Donnay, J. P., Barnsley, M. J., and P. A. Longley (Eds.), New York, NY: Taylor and Francis, 225-241. **[Author, this reference is not cited in the text; please add text citation of delete reference from the references list]**

Bracken, I., 1991, "A Surface Model Approach to Small Area Population Estimation," *Town Planning Review*, 62(2):225-237.

Bracken, I. and D. Martin, 1989, "The Generation of Spatial Population Distributions from Census Centroid Data Source," *Environment and Planning A*, 21(4):537-543.

Chen, K., 2002, "An Approach to Linking Remotely Sensed Data and Areal Census Data," *International Journal of Remote Sensing*, 23(1):37-48.

Clark, C., 1951, "Urban Population Densities," *Journal of the Royal Statistical Society*, 114:490-496.

Clayton, C. and J. Estes, 1980, "Image Analysis as a Check on Census Enumeration Accuracy," *Photogrammetric Engineering and Remote Sensing*, 46:757-764.

Collins, W. G. and A. H. A. El-Beik, 1971, "Population Census with the Aid of Aerial Photographs: An Experiment in the City of Leeds," *Photogrammetric Record*, 7:16-26.

Dempster, A. P., Laird, N. M., and D. B. Rubin, 1977, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of Royal Statistics Society Series B–Methodological*, 39(1):1-38.

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., and B.A. Worley, 2000, "LandScan: A Global Population Database for Estimating Populations at Risk," *Photogrammetric Engineering and Remote Sensing*, 66(7):849-857.

Donnay, J. P. and D. Unwin, 2001, "Modeling Geographical Distributions in Urban Areas," in *Remote Sensing and Urban Analysis*, Donnay, J. P., Barnsley, M. J., and P. A. Longley (Eds.), New York, NY: Taylor and Francis, 205-224.

Dueker, K. and F. Horton, 1971, "Toward Geographic Urban Change Detection Systems with Remote Sensing Inputs," *Technical Papers*, 37th Annual Meeting, American Society of Photogrammetry, 204-218.

Eicher, C. L. and C. A. Brewer, 2001, "Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation," *Cartography and Geographic Information Science*, 28(2): 125-138.

Fisher, P. F. and M. Langford, 1995, "Modeling the Errors in Areal Interpolation between Zonal Systems by Monte Carlo Simulation," *Environment and Planning A*, 27:211-224.

Fisher, P. F. and M. Langford, 1996, "Modeling Sensitivity to Accuracy in Classified Imagery: A Study of Areal Interpolation by Dasymetric Mapping," *Professional Geographer*, 48(3):299-309.

Flowerdew, R. and M. Green, 1989, "Statistical Methods for Inference between Incompatible Zonal Systems," in *Accuracy of Spatial Databases*, Goodchild, M. and S. Gopal (Eds.), London, UK: Taylor and Francis, 239-247.

Flowerdew, R. and M. Green, 1991, "Data Integration: Statistical Methods for Transferring Data between Zonal Systems," in *Handling Geographical Information: Methodology and Potential Applications*, Masser, I. and M. Blakemore (Eds.), Harlow, UK: Longman, 38-54.

Forster, B.C., 1983, "Some Urban Measurements from Landsat Data," *Photogrammetric Engineering and Remote Sensing*, 49:1693-1707.

Green, N. E., 1956, "Aerial Photographic Analysis of Residential Neighborhoods: An Evaluation of Data Accuracy," *Social Forces*, 35:142-147.

Green, N. E., 1957, "Aerial Photographic Interpretation and the Social Structure of the City," *Photogrammetric Engineering*, 23:89-96.

Green, N. E. and R. B. Monier, 1959, "Aerial Photographic Interpretation of the Human Ecology of the City," *Photogrammetric Engineering*, 25:770-773.

Harvey, J. T., 2000, "Small Area Population Estimation Using Satellite Imagery," *Statistics in Transition*, 4(4):611-633.

Harvey, J. T., 2002a, "Estimating Census District Populations from Satellite Imagery: Some Approaches and Limitations," *International Journal of Remote sensing*, 23(10):2071–2095.

Harvey, J. T., 2002b, "Population Estimation Models Based on Individual TM Pixels," *Photogrammetric Engineering and Remote Sensing*, 68(11):1181-1192.

Haverkamp, D., 2004, "Automatic Building Extraction from IKONOS Imagery," *Proceedings of ASPRS 2004 Conference*, Denver, Colorado, May 23-28, 2004 [available online at http://www.spaceimaging.com/techpapers/default.htm].

Holloway, S. R., Schumachey, J., and R. Redmond, 1996, *People and Place: Dasymetric Mapping Using ArcInfo*, Missoula, MT: Wildlife Spatial Analysis Lab, University of Montana. **[Author, this reference is not cited in the text. Please add text citation or indicate that this reference should be deleted from the references list]**

Holt, J. B., Lo, C. P., and T. W. Hodler, 2004, "Dasymetric Estimation of Population Density and Areal Interpolation of Census Data," *Cartography and Geographic Information Science*, 31(2):103-121.

Hsu, S. Y., 1971, "Population Estimation," *Photogrammetric Engineering*, 37:449-454.

Hsu, S. Y., 1973, "Population Estimation from ERTS Imagery: Methodology and Evaluation," *Proceedings of the American Society of Photogrammetry 39th Annual Meeting*, 583-591.

Huxley, J., 1932, *Problems of Relative Growth*, London, UK: Methuen, 356 p.

Iisaka, J. and E. Hegedus, 1982, "Population Estimation from Landsat Imagery," *Remote Sensing of the Environment*, 12:259-272.

Kramer, C., 1958, "Population Density Patterns," *Chicago Area Transportation Study (CATS) Research News*, 2:3-10.

Kraus, S. P., Senger, L. W. and J. M. Ryerson, 1974, "Estimating Population from Photographically Determined Residential Land Use Types," *Remote Sensing of Environment*, 3(1):35-42.

Lam, N. S., 1983, "Spatial Interpolation Methods: A Review," *The American Cartographer*, 10(2):129-149.

Langford, M., Maguire, D. J., and D. J. Unwin, 1991, "The Areal Interpolation Problem: Estimating Population Using Remote Sensing in a GIS Framework," in

Handling Geographical Information: Methodology and Potential Applications, Masser, I. and M. Blakemore (Eds.), New York, NY: Wiley, 55-77.

Langford, M. and D. J. Unwin, 1994, "Generating and Mapping Population-Density Surfaces within a Geographical Information-System," *Cartographic Journal*, 31(1):21-26.

Lee, Y., 1989, "An Allometric Analysis of the United States Urban System: 1960-80," *Environment and Planning A*, 21(4):463-476.

Liu, X., 2003, *Estimation of the Spatial Distribution of Urban Population Using High Spatial Resolution Satellite Imagery*, Ph.D. thesis, University of California, Santa Barbara, 175 p.

Liu, X. and K. C. Clarke, 2002, "Estimation of Residential Population Using High Resolution Satellite Imagery," *Proceedings of the 3rd Symposium in Remote Sensing of Urban Areas*, Maktav, D., Juergens, C., and F. Sunar-Erbek, June 11–13, 2002, Istanbul, Turkey: Istanbul Technical University, 153-160.

Lo, C. P., 1986, *Applied Remote Sensing*, New York, NY: Longman, 393 p.

Lo, C. P., 1989, "A Raster Approach to Population Estimation Using High-Altitude Aerial and Space Photographs," *Remote Sensing of Environment*, 27(1):59-71.

Lo, C. P., 1995, "Automated Population and Dwelling Unit Estimation from High-Resolution Satellite Images—a GIS Approach," *International Journal of Remote Sensing*, 16(1):17-34.

Lo, C. P., 2002, "Urban Indicators of China from Radiance-Calibrated Digital DMSP-OLS Nighttime Images," *Annals of The Association Of American Geographers*, 92(2):225-240.

Lo, C. P., 2003, "Zone-Based Estimation of Population and Housing Units from Satellite-Generated Land Use/Land Cover Maps," in *Remotely Sensed Cities, Mesev, V. (Ed.), Lon*don, UK/New York, NY: Taylor & Francis, 157-180.

Lo, C. P. and H. F. Chan, 1980, "Rural Population Estimation from Aerial Photographs," *Photogrammetric Engineering and Remote Sensing*, 46:337-345.

Lo, C. P. and R. Welch, 1977, "Chinese Urban Population Estimates," *Annals of the Association of American Geographers*, 67(2):246-253.

Martin, D., 1989, "Mapping Population Data from Zone Centroid Locations," *Transactions of the Institute of British Geographers*, 14(1):90-97.

Martin, D., 1996, "An Assessment of Surface and Zonal Models of Population," *International Journal of Geographical Information Systems*, 10(8):973-989.

Martin, D. and I. Bracken, 1991, "Techniques for Modeling Population-Related Raster Databases," *Environment and Planning A*, 23(7):1069-1075.

Mennis, J., 2003, "Generating Surface Models of Population Using Dasymetric Mapping," *Professional Geographer*, 55(1): 31-42.

Monmonier, M. S. and G. A. Schnell, 1984, "Land-Use and Land-Cover Data and the Mapping of Population Density," *The International Yearbook of Cartography*, 24:115-121.

Newling, B. E., 1965, "Urban Growth and Spatial Structure—Mathematical Models and Empirical Evidence," *Annals of the Association of American Geographers*, 55(4):637-637.

Nordbeck, S., 1965, "The Law of Allometric Growth," *Michigan Inter-University Community of Mathematical Geographers*, Discussion Paper No. 7.

Parr, J. B., 1985, "A Population-Density Approach to Regional Spatial Structure," *Urban Studies*, 22(4):289-303.

Porter, P. W., 1956, *Population Distribution and Land Use in Liberia*, Ph.D. thesis, London School of Economics and Political Science, London, UK, 213 p.

Prosperie, L. and R. Eyton, 2000, "The Relationship between Brightness Values from a Nighttime Satellite Image and Texas County Population," *Southwestern Geographer*, 4:16-29.

Rase, W., 2001, "Volume-Preserving Interpolation of a Smooth Surface from Polygon-Related Data," *Journal of Geographical Systems*, 3(2):199-213.

Rottensteiner, F., 2003, "Automatic Generation of High-Quality Building Models from Lidar Data," *IEEE Computer Graphics and Applications*, 23(6):42-50.

Stewart, J. Q. and W. Wartnz, 1958, "Physics of Population Distribution," *Journal of Regional Science*, 1:99-113.

Sutton, P., Roberts, C., Elvidge, C., and H. Meij, 1997, "A Comparison of Nighttime Satellite Imagery and Population Density for the Continental United States," *Photogrammetric Engineering and Remote Sensing*, 63(11):1303-1313.

Thompson, D., 1975, "Small Area Population Estimation Using Land Use Data Derived from High Altitude Aircraft Photography," *Proceedings of the American Society of Photogrammetry 41st Annual Meeting*, 673-696.

Tobler, W. R., 1969, "Satellite Confirmation of Settlement Size Coefficients," *Area*, 1(3):30-34.

Tobler, W. R., 1979, "Smooth Pyncophylactic Interpolation for Geographical Regions," *Journal of the American Statistical Association*, 74(367):519-530.

Tobler, W. R., 1999, "Linear Pycnophylactic Reallocation—Comment on a paper by D. Martin," *International Journal of Geographical Information Science*, 13(1):85-90.

Weber, C., 1994, "Per-zone Classification of Urban Land Use Cover for Urban Population Estimation," in *Environmental Remote Sensing from Regional to Global Scales*, in Foody, G. M. and P. J. Curran (Eds.), Chichester, UK: New York, NY: Wiley, 142-148.

Webster, C. J., 1996, "Population and Dwelling Unit Estimation from Space," *Third World Planning Review*, 18(2):155-176.

Weiss, H. K., 1961, "The Distribution of Urban Population and an Application to a Servicing Problem," *Operations Research*, 9(6):860-874.

Wright, J. K., 1936, "A Method of Mapping Densities of Population," *The Geographical Review*, 26(1):103-110.

Yuan, Y., Smith, R. M., and W. F. Limp, 1997, "Remodeling Census Population with Spatial Information from LandSat TM imagery," *Computers, Environment and Urban Systems*, 21(3-4):245-258.