Using Landsat ETM+ Imagery to Measure Population Density in Indianapolis, Indiana, USA

Guiying Li and Qihao Weng

Abstract

Remote sensing techniques have been previously used in urban analysis, settlement detection, and population estimation. This research explores the potentials of integration of Landsat ETM+ data with census data for estimation of population density in City of Indianapolis, Indiana. Spectral signatures, principal components, vegetation indices, fraction images, textures, and temperature were used as predictive indicators. Correlation analysis was used to explore the relationships between remote sensing variables and population, and stepwise regression analysis was then used to develop models for estimating population quantities. Two sampling schemes (non-stratified versus stratified sampling) were compared. It was found that the integration of textures, temperatures, and spectral responses substantially improved the accuracy of estimation. Stratification of the population into three categories of low-, medium-, and high-densities and development of different models for individual population density category provided better estimation results than a non-stratified scheme. The total population for City of Indianapolis was estimated to be 832,792 in 2000 yielding an accuracy of 96.8 percent.

Introduction

World population has experienced a high growth rate since the Industrial Revolution in the 18th century, especially in recent decades. The large world population has produced great pressures on global resources, environment, and sustainable development (Lo, 1986a; Sutton *et al.*, 1997). The pressure from population increase often results in urban expansion at the expense of decreased non-urban lands, such as agricultural land and forest. Timely and accurate population estimation, its spatial distribution, and dynamics become considerably significant for understanding the effect of population increase on the social, economic, and environmental problems. Moreover, population information at different levels, such as national, regional, and local, are very important for many purposes such as urban planning, resource management, and service allocation. Conventional census methods of population estimation are found to be time-consuming, costly, and difficult to update. Besides, the census interval is often too long for many types of application; for example, the census is conducted every ten years in the United States. Due to large migrations, population distributions can change quickly, thus the census data are frequently found to be obsolete. Therefore, it is necessary to develop suitable techniques for estimating population in an accurate and timely manner at different spatial scales.

Geographic research of population estimation started early in the 1930s. John K. Wright, a geographer working at American Geographical Society, pioneered in population estimation study by producing a map of population distribution in Cape Cod, Massachusetts (Wright, 1936). Wright termed his method *dasymetric mapping*, in which the breaks in the population distribution map were related to types of land use. With the maturing of GIS technology, some applications suggest that Wright's seminal work could be applied to areal interpolation (Flowerdew and Green, 1992) suitable for statistically modeling a wide range of phenomena including population. Following the idea of dasymetric mapping and implemented by using the pycnophylactic interpolation method, the National Center for Geographic Information and Analysis created global raster images of population distribution from a set of 15,000 administrative units in the Global Demography project (Tobler *et al.*, 1995). Remote sensing techniques have been used for population estimation since the 1950s, when Porter (1956) estimated population in a settlement of Liberia by counting the number of huts on aerial photography and by multiplying it by mean occupants per hut derived from ground sampling survey. With advances of remote sensing and GIS technology, remotely sensed data have become an important resource in population estimation due to their strengths in data coverage, reasonable accuracy, and low cost (Lo, 1995; Jensen and Cowen, 1999). Different methods have been developed to estimate population based on aerial photography and satellite imagery (Lo and Welch, 1977; Watkins and Morrow-Jones, 1985; Lo, 1986a, 1986b, 1995, 2001; Langford *et al.*, 1991; Sutton, 1997; Sutton *et al.*, 1997, 2001; Weeks *et al.*, 2000; Harvey, 2002a, 2002b; Qiu *et al.*, 2003). Satellite-based variables have been recently combined with other geographic variables to produce a comparatively high-resolution (30 arc-sec) population database for the entire globe (Dobson *et al.*, 2000; Dobson, 2003; Dobson *et al.*, 2003).

Lo (1986a) reviewed the strengths and limitations of remote sensing methods of population estimation and classified them into four major categories: (a) based on

Department of Geography, Geology, and Anthropology, Indiana State University, Terre Haute, IN 47809 (geweng@isugw.indstate.edu; gli@mymail.indstate.edu).

Photogrammetric Engineering & Remote Sensing
Vol. 71, No. 8, August 2005, pp. 947–958.

0099-1112/05/7108–0947/$3.00/0

counts of dwelling units; (b) based on measurement of built-up areas; (c) based on measurement of different land use areas; and (d) based on spectral radiance of individual pixels. Each of these approaches has different requirements in terms of the scale and resolution of photography or imagery (Lo, 1986a).

The first approach, based on counts of dwelling units, is regarded as the most accurate remote sensing method (Forester, 1985; Lindgren, 1985; Lo, 1986a, 1995; Holz, 1988; Haack *et al.*, 1997; Jensen and Cowen, 1999). Photography or imagery used should have sufficiently high spatial resolution in order to identify the type of individual buildings, and the average number of persons per dwelling unit should be available. It also assumes all dwellings are occupied. Aerial photography is more commonly used than satellite imagery due to its high resolution. Hsu (1971) and Lindgren (1971) used this method to estimate and map the population distribution in the Atlanta and Boston metropolitan areas. The main difficulty with this approach was how to distinguish high-rise apartment buildings from multistory office buildings.

The second approach, based on measurements of built-up areas, employs an allometric growth model that describes the relationship between built–up areas and population size. Different versions of the following formula are used: $r = aP^b$, where r is the radius measured from the center of a settlement, a is a coefficient, P is the population size of that settlement, and b is an exponent. This method was employed by Wellar (1969) for a study in Houston and San Antonio, Texas with Gemini XII photographs. Similarly, Lo and Welch (1977) studied Chinese cities, and Ogrosky (1975) studied the Puget Sound Region in Washington using infrared aerial photography. In addition to aerial photography and Landsat imagery, low-resolution nighttime images (spatial resolution: 2.7 km) from the Defense Meteorological Satellite Program had also been used to map human settlements (Elvidge *et al.*, 1995, 1997) and urban extent (Imhoff *et al.*, 1997), and to estimate population nationally and globally (Welch and Zupko, 1980; Sutton, 1997; Sutton *et al.*, 1997, 2001; Lo, 2001). Radar images were also used in population estimation (Henderson and Xia, 1997).

The third method, based on measurement of the area of different land use types, involves classification of remotely sensed imagery into discrete land use categories. Langford *et al.* (1991) used a TM image to estimate the population of the 49 wards of Leicestershire, United Kingdom. The TM image was classified into five land use classes and the pixel count of each category within each ward was correlated with population. It was found that ward population had a relatively high positive correlation with the pixel counts in industry, commerce, dense residential and ordinary residential categories, respectively, and had low negative correlations with those areas of no population and agricultural use. Webster (1996) developed models to estimate dwelling densities in the 47 suburbs of Harare, Zimbabwe based on measures of tones (six TM bands), texture (three measures derived from classification of pixels into urban and non-urban: urban pixel density, homogeneity, and entropy), and context (distance from the city center) using SPOT and TM images, and found that R^2 values ranged from 0.69 to 0.81. Chen (2002) studied the relationship between areal census dwelling data and residential densities, which were derived from Landsat TM image covering thirteen census collection districts of Hornsby Heights, Sydney, Australia.

The fourth method utilizes spectral radiance or transformed remote sensing variables to estimate population. Hsu (1973) used Landsat MSS multispectral radiance data, cell by cell (1×1 km), to estimate population through implementation of a multiple regression model. Iisaka and Hegedus (1982) studied population distribution in residential sections of suburban Tokyo, Japan using MSS data. Two multiple linear regression models were developed, in which population was the dependent variable and mean reflectance values of 10×10 pixel grid of four MSS bands were independent variables. Correlation coefficients of 0.77 and 0.899 for 1972 and 1979 were obtained, respectively.

These population estimation approaches have also been combined for use. Lo (1995) used a mixed approach including spectral radiance of image pixels and pixel counts in residential classes to estimate population and dwelling densities in 44 tertiary planning units (TPUs) in Kowloon, Hong Kong using multispectral SPOT imagery. Four linear regression models were developed using the following independent variables: means of SPOT bands 1, 2, and 3; mean of SPOT band 3 alone; percentages of pixels classified as high and low density residential use in each TPU; proportion of high density residential use pixels in each TPU. In addition, an allometric model was built with the number of pixels in the high-density residential class as the independent variable. The models were validated by applying them to 44 TPUs, and found that the allometric growth model was the best one at a macro scale. At a micro scale, its estimation accuracy was not satisfactory due to highly mixed land use and difficulty in distinguishing residential from non-residential uses. Harvey (2002a) refined the methods developed by Iisaka and Hegedus (1982) and Lo (1995), and introduced a variety of standard spectral transformations of Landsat TM Imagery (squares of six basic band means, 15 band-mean to band-mean cross-product, 15 pairwise band-to-band ratios, and 15 pairwise difference-to-sum ratio of the TM data) into regression models for population estimation in Ballarat, Sydney, Australia. The population densities of 132 collect districts (CDs), or its logarithmic and square root transformation, were regressed against the transformed remote sensing variables. The results showed that spectral transformations and application of both the square root and the logarithmic forms improved correlations. Among the six models validated, three produced a median proportional error of 17 to 21 percent for the population of individual CDs, while the median proportional error for the total population of Ballarat was within 3 percent. Similar to other population estimation studies, all the models overestimated population in low-density rural sections and underestimated them in high-density urban sections.

Rarely has research explored the integration of spectral, textural, temperature data, and advanced transformed remote sensing variables to estimate population. Such incorporation may provide new insights for population density estimation. Although previous research has indicated that population with high density were often underestimated and those with low density were often overestimated (Harvey, 2000a), no suitable solution has been proposed to correct these errors. The main objective of this research is to develop techniques for population estimation using Landsat ETM+ data for city of Indianapolis, Indiana. Spectral bands, principal components, vegetation indices, fraction images, textures, and land surface temperature derived from Landsat imagery were examined as explanatory variables. Moreover, two sampling schemes (i.e., stratifying population density into low-, medium-, and high-, and then developing estimation model separately, and non-stratified) were compared. This paper is organized into five sections. The following section introduces the study area; then, the methodology used is described including data processing, and model development; finally the results are presented, followed by a section of discussion and conclusions.

Study Area

The City of Indianapolis, Indiana, (Figure 1) has been chosen to implement this study. According to the 2000

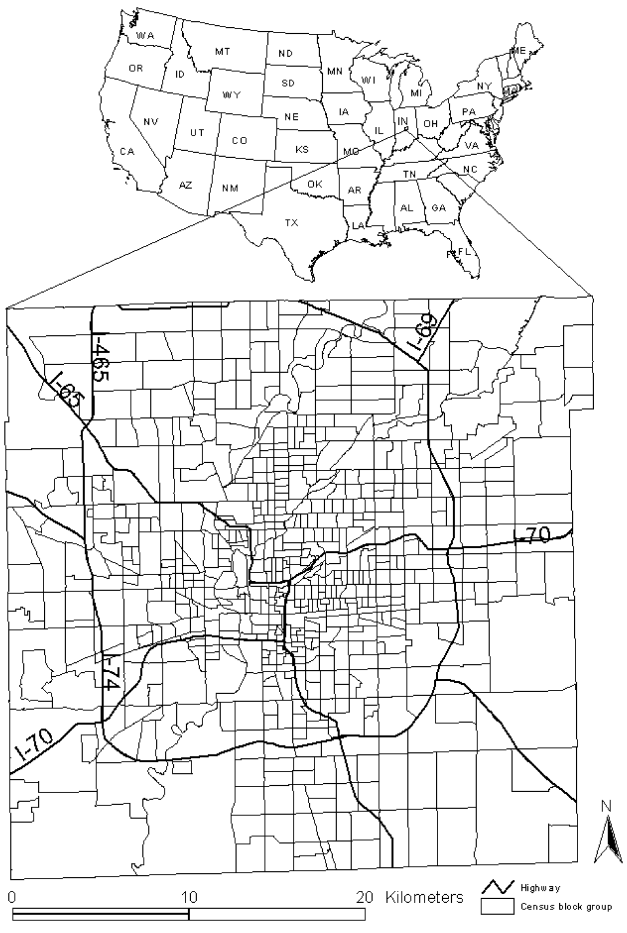

Figure 1. Study area: City of Indianapolis, Indiana.

census results, the city has 860,454 people, increasing by 7.9 percent over 1990. With this population, Indianapolis ranks as the twelfth largest city in the nation. It has the highest concentration of major empoyers and manufacturing, professional, technical, and educational services in the state. With its moderate climate, rich history, excellent education, social services, arts, leisure, and recreation, Indianapolis was named one of "America's Best Places to Live & Work" (Employment Reviews, August 1996). Its flatness and relatively symmetrical allocation provide the possibilities for expansion in all directions. Like most American cities, Indianapolis has been experiencing areal expansion through encroachment on agricultural land and other non-urban land as population increases and economic growth. The timely and accurate population information is significant for urban planning and civic applications.

Data and Methods

Data

Landsat 7 Enhanced Thematic Mapper Plus (ETM+) image (Row/Path: 32/21) dated on 22 June 2000 was used in this research. Atmospheric conditions were clear at the time of image acquisition, and the image was acquired through the USGS Earth Resource Observation Systems Data Center, which had corrected the radiometric and geometrical distortions of the image to a quality level of 1G before delivery.

Population data at block group level were obtained from an ESRI Data and Maps CD, which was provided by ESRI, based on the combination of TIGER files and 2000 population census data. Because of different coordinate systems used for the census data and the ETM+ image, the geographic coordinates of census data were converted to the UTM to match with those of the ETM+ image (Figure 2).

Image Processing

Principal Component Analysis

Remotely sensed data, such as visible bands in Landsat TM/ETM+ images, are highly correlated between the adjacent spectral bands (Barnsley, 1999). Several techniques have been developed to transform highly correlated bands into an orthogonal subset. The principal component analysis (PCA) is the most commonly used one. After performing PCA, the original correlated bands are transformed into independent principal components (PC), of which the first PC contains the largest portion of data variance and the second PC contains the second largest data variance, and so on. The higher numbered PCs often appear noisy, since they contain very little variance of information (Richards, 1994). In this study, six ETM+ multispectral bands (i.e., 1, 2, 3, 4, 5, and 7) were used to perform PCA. The first three PCs were used in population estimation analysis because they accounted for 99 percent of total variance.

Vegetation Indices

Many vegetation indices have been developed based on the fact that plants reflect less in visible red light, but more in near infrared radiation compared with non-vegetated surface (Bannari *et al.*, 1995; Jensen, 2000). Thus, vegetation indices can enhance or extract some specific features that single spectral bands cannot. In this research, six vegetation indices, namely, normalized difference vegetation index (NDVI), soil adjusted vegetation index (SAVI), renormalized difference vegetation index (RDVI), transformed NDVI (TNDVI), simple vegetation index (SVI), and simple ratio (RVI), were examined to use for population estimation (Table 1).

Fraction Images

Spectral mixture analysis (SMA) is regarded as a physically-based image processing tool that supports repeatable and accurate extraction of quantitative sub-pixel information (Smith *et al.*, 1990; Roberts *et al.*, 1998; Mustard and Sunshine, 1999). It assumes that the spectrum measured by a sensor is a linear combination of the spectra of all components within the pixel (Adams *et al.*, 1995; Roberts *et al.*, 1998). Because of its effectiveness in handling spectral mixture problems, SMA has been widely used in estimation of vegetation cover (Smith *et al.*, 1990; Asner and Lobell, 2000; McGwire *et al.*, 2000; Small, 2001), in vegetation or land cover classification and change detection (Adams *et al.*, 1995; Roberts *et al.*, 1998; Cochrane and Souza, 1998; Aguiar *et al.*, 1999; Lu *et al.*, 2003), and in urban studies (Rashed *et al.*, 2001; Small, 2001; Phinn *et al.*, 2002; Wu and Murray, 2003). In this study, SMA was used to develop green vegetation and impervious surface fraction images. Endmembers were initially identified from the ETM+ image based on high-resolution aerial photographs. The shade endmember was identified from the areas of clear and deep water, while green vegetation was selected from the areas of dense grass and cover crops. Different types of impervious surfaces were selected from building roofs and highway intersections. An unconstrained least-squares solution was used to decompose the six ETM+ bands (1, 2, 3, 4, 5, and 7) into three fraction

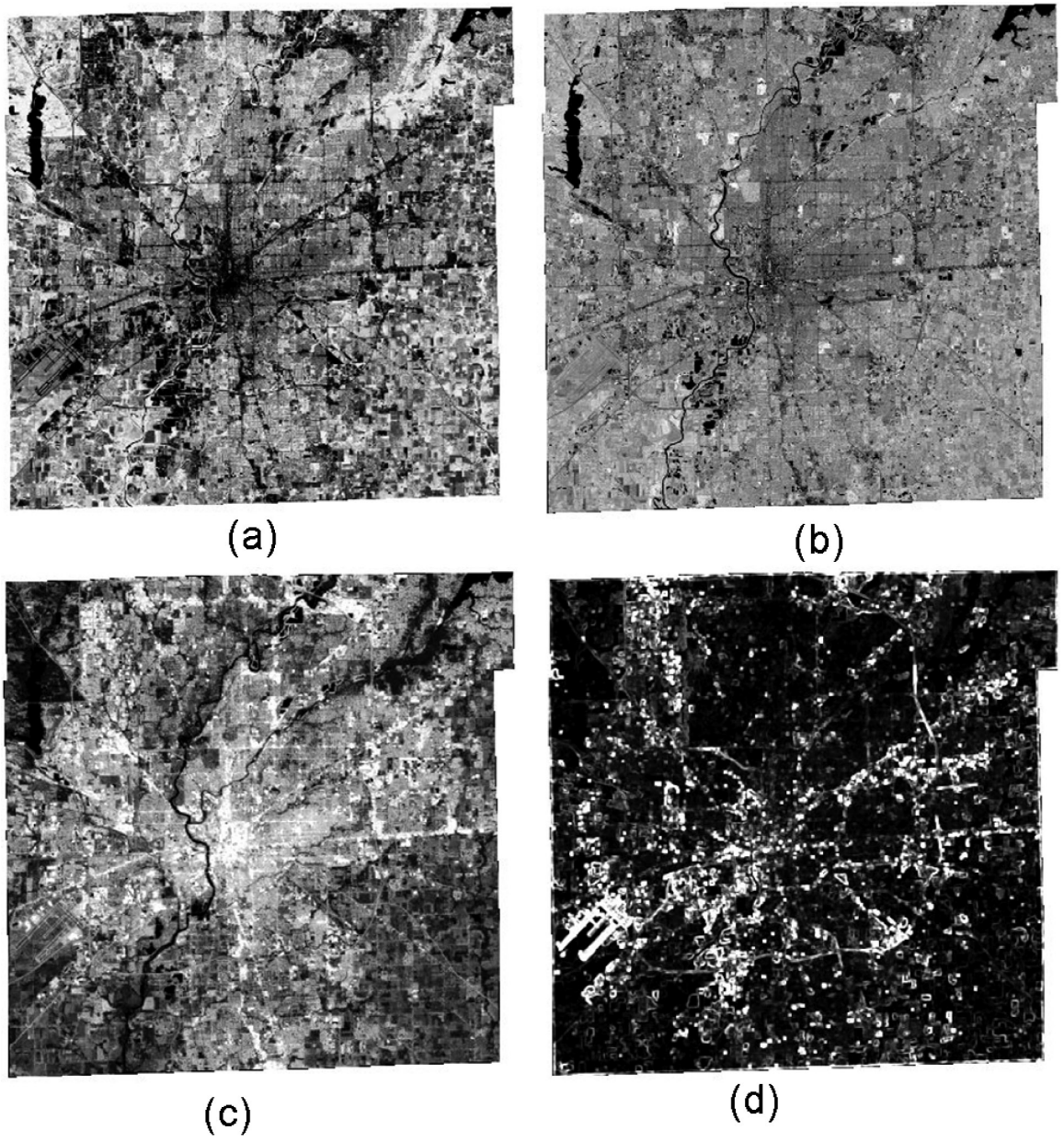

Figure 2. Some images used as remote sensing variables for population estimation. (a) NDVI, (b) Principal Component 2, (c) Temperature, and (d) Texture image derived from Landsat ETM+ band 3 with 7 × 7 window size.

images (vegetation, impervious surface, and shade). The fractions represent the areal proportions of the endmembers within a pixel. The shade fraction was not used due to its irrelevance to the population distribution. A detailed description of the above procedure can be found in Lu and Weng (2004).

Texture

Texture often refers to the pattern of intensity of variations in an image. Many texture measures have been developed (Haralick *et al.*, 1973; Haralick, 1979; He and Wang, 1990), and used for land cover classification (Marceau *et al.*, 1990; Gong and Howarth, 1992; Shaban and Dikshit, 2001; Narasimha Rao *et al.*, 2002). A common texture measure, variance, has shown to be useful in improving land cover classification (Shaban and Dikshit, 2001). In this study, variance was developed and used to examine its relationship with population. The ETM+ bands 3 and 7, which correlate strongly with the urban features, were used for deriving texture images with window sizes of 3 × 3, 5 × 5 and 7 × 7.

Temperature

The surface temperature image was extracted from the ETM+ thermal infrared band (band 6). The procedure to develop the surface temperature involves three steps: (a) converting the digital number of Landsat ETM+ band 6 into spectral

TABLE 1. DEFINITION OF VEGETATION INDICES USED IN THIS STUDY

Vegetation Index	Abbr.	Formula	Reference
Normalized difference vegetation index	NDVI	$\frac{NIR - RED}{NIR + RED}$	Rouse *et al.* (1974)
Soil adjusted vegetation index	SAVI	$\frac{(1 + L)(NIR - RED)}{NIR + RED + L}$, L = 0.5	Huete (1988)
Renormalized difference vegetation index	RDVI	$\frac{NIR - RED}{\sqrt{NIR + RED}}$	Roujean and Breon (1995)
Transformed NDVI	TNDVI	$\sqrt{(NDVI + 0.5)}$	Deering *et al.* (1975)
Simple vegetation index	SVI	NIR-RED	
Simple ratio	RVI	NIR/RED	Birth and McVey (1968)

Note: NIR – Near infrared wavelength, ETM+ band 4;
RED – Red wavelength, ETM+ band 3.

radiance; (b) converting the spectral radiance to at-satellite brightness temperature, which is also called blackbody temperature; and (c) converting the blackbody temperature to land surface temperature. A detailed description for developing the temperature image can be found in Weng *et al.* (2004).

Model Development

Since census data and ETM+ data have different formats and spatial resolutions, they need to be integrated. With the use of ERDAS Imagine® software, remotely sensed data were aggregated to block group level. The mean values of selected remote sensing variables at the block group level were computed. The variables include radiances of ETM+ bands, principal components, vegetation indices, green vegetation and impervious surface fractions, temperatures, and texture indicators. All these data were then exported into SPSS software for correlation and regression analysis.

Twenty-five percent of the total block groups (658) in the city were randomly selected, and a 2.5 standard deviation was used to identify the outliers. A total of 162 samples were used for developing models with a non-stratified sampling scheme. The population density in Indianapolis was calculated to range from 0 to 7253 persons per km^2, while most BGs had a population density ranged from 400 to 3000 persons per km^2 (Figure 3).

Previous research has indicated that extremely high or low population density is difficult to estimate using remotely sensed data (Lo, 1995; Harvey, 2002a, 2000b), hence, the population densities of the city were divided into three categories: low (less than 400 person/km^2), medium (401 to 3000), and high (greater than 3000) according to the data distribution. All block groups in the low- and high-density categories were used for sampling owing to their limited number. For the medium density category, samples were chosen using a random sampling technique. Table 2 summarizes the statistical characteristics of selected samples for different categories.

Pearson's correlation coefficients were computed between population densities and the remote sensing variables. Stepwise regression analysis was further applied to identify suitable variables for developing population estimation models. The coefficient of determination (R^2) was used as an indicator to determine the robustness of a regression model. To improve model performance, various combinations of the remote sensing variables were explored, as well as the transformation of population densities into natural logarithm (LPD) and square root (SPD) forms.

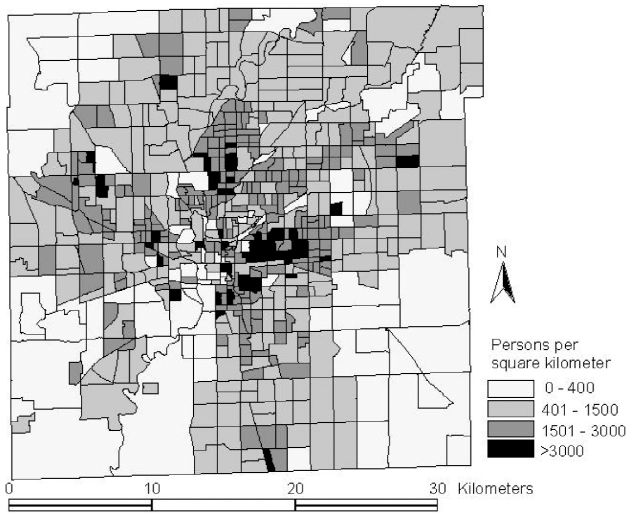

Figure 3. Population density distribution by block groups, Indianapolis, Indiana.

TABLE 2. STATISTICAL DESCRIPTIONS OF SAMPLES OF POPULATION DENSITIES (PERSONS/KM^2)

Categories	Samples	Min.	Max.	Mean	Std. Dev.
Non-stratified	162[a] (175[b]) (658[c])	8	4479	1470.71	948.62
Low	77[a] (82[c])	1	393	208.94	123.11
Medium	114[a] (125[b]) (499[c])	402	2824	1417.31	676.97
High	70[a] (77[c])	3015	5189	3707.66	579.08

Note: [a]is the samples that removed outliers and finally used for data analysis;
[b]is the samples selected based on a random sampling technique;
[c]is the total number of block groups corresponding population.

Accuracy Assessment

Whenever a model developed is applied for prediction, there are always discrepancies between true and estimated values, which are called *residuals*. It is necessary to validate the model whether it fits training set data, which is called internal validation, or to test its fitness with other data sets that are not used as training sets, which is called *external validation* (Harvey, 2002a). Relative and absolute error can be computed. For an individual case, the relative error (RE) can be expressed as:

$$\text{RE} = (P_g - P_e)/P_g \times 100 \quad (1)$$

where P_g and P_e are the reference and estimated values respectively. The residual ($P_g - P_e$) for individual cases may be negative or positive, so absolute values of the residuals were used to assess the overall performance of a developed model, i.e.,

$$\text{Overall relative error (RE)} = \frac{\sum_{k=1}^{n} |REn|}{n} \quad (2)$$

$$\text{Overall absolute error (AE)} = \frac{\sum_{k=1}^{n} |P_g - P_e|}{n} \quad (3)$$

where n is the number of block groups used for accuracy assessment. The smaller the RE and AE, the better the models

would be. A total of 483 un-sampled BGs were used to assess the performance of models in the non-stratified sampling scheme. For the stratified sampling scheme, a total of 521 samples were used for accuracy assessment. A residual map was created based on the best estimation model for geographical analysis of predicted errors.

Results

Population Estimation Based on Non-stratified Sampling Scheme

Six groups of remote sensing variables were used to explore their relationships with population parameters, and their correlation coefficients are presented in Table 3. Table 3 indicates that among the ETM+ spectral bands, band 4 was the most strongly correlated with population density, the transforms of population density into natural logarithm or square-root forms did not improve correlation coefficients of single ETM+ bands except for band 5. The principle components, especially PC2, improved the correlation with population parameters when compared with single ETM+ bands. All selected vegetation indices had a significant correlation with population density. The GV fraction had a better correlation with population density than impervious surface fraction. Selected textures, especially band 7 associated with a window size of 7 by 7, were strongly correlated with population density. Among all selected remote sensing variables, temperature was the most correlated variable with population density. Moreover, it is found that vegetation related variables such as band 4, PC2, vegetation indices and green vegetation fraction all had a negative correlation with population parameters. This is because for a given area, more vegetation is often related to less built-up area, and thus to less population.

The strong correlation between population parameters and several remote sensing variables imply that a combination of temperature, textures, and spectral responses could be used to improve the models of population estimation. A series of estimation models were developed by performing stepwise regression analysis, based on different combinations of remote sensing variables. The predictors and R^2 of regression models developed are presented in Table 4.

Table 4 indicates that any single group of remote sensing variables did not produce a satisfactory R^2 except for vegetation indices. Incorporation of vegetation-related variables or use of all variables provided better modeling results. The square root form of population density improved the regression models, while the natural logarithm form degraded the regression performance, with an exception in the textures. Table 5 summarizes the best performing regression models and associated estimation errors.

Overall, larger R^2 values resulted in less estimation errors. The regression models using a combination of spectral, textures, and temperature provided the best estimation results. The R^2 value for the best model reached 0.83, but the estimation errors were still high. The overall relative errors were larger than 123 percent, and the overall absolute errors were greater than 439 persons/km^2 (the mean population density is 1,470 persons/km^2). The extreme low and high population density BGs were the main sources of error. Low population density BGs had more severe impacts on relative errors, while high population density BGs had more impacts on absolute errors. These impacts can be clearly illustrated in the scatter plot of the residuals. Figure 4 shows the residual distributions of the best model (Model 4). It indicates that population in very low-density BGs was overestimated, while population in high-density area was greatly underestimated. The high estimation errors imply that no single model worked well for all levels of population density. In order to improve population estimation results, separating the population density into sub-categories such as low, medium, and high densities, and developing models for each category becomes necessary.

TABLE 3. RELATIONSHIPS BETWEEN POPULATION PARAMETERS AND REMOTE SENSING VARIABLES BASED ON NON-STRATIFIED SAMPLES

Variables		PD	SPD	LPD
Bands	B1	0.226[a]	0.160[b]	0.019
	B2	0.163[b]	0.096	−0.039
	B3	0.164[b]	0.096	−0.039
	B4	−0.255[a]	−0.209[a]	−0.108
	B5	−0.155[b]	−0.196[b]	−0.251[a]
	B7	0.068	0.003	−0.115
PCs	PC1	0.123	0.056	−0.073
	PC2	−0.319[a]	−0.283[a]	−0.190[b]
	PC3	−0.248[a]	−0.239[a]	−0.178[b]
VIs	NDVI	−0.244[a]	−0.182[b]	−0.052
	RDVI	−0.242[a]	−0.178[b]	−0.040
	SAVI	−0.245[a]	−0.182[b]	−0.053
	SVI	−0.221[a]	−0.156[b]	−0.023
	RVI	−0.385[a]	−0.337[a]	−0.206[a]
	TNDVI	−0.164[b]	−0.098	0.026
Frac.	GV	−0.231[a]	−0.171[b]	−0.045
	IMP	0.109	0.043	−0.082
Text.	B3_3 × 3	−0.196[b]	−0.223[a]	−0.267[a]
	B7_3 × 3	−0.295[a]	−0.326[a]	−0.347[a]
	B3_5 × 5	−0.280[a]	−0.317[a]	−0.360[a]
	B7_5 × 5	−0.368[a]	−0.406[a]	−0.427[a]
	B3_7 × 7	−0.322[a]	−0.364[a]	−0.407[a]
	B7_7 × 7	−0.402[a]	−0.444[a]	−0.463[a]
Temp.	TEMP	0.519[a]	0.513[a]	0.411[a]

Notes: [a]Correlation at 99 percent confidence level (2-tailed);
[b]Correlation at 95 percent confidence level (2-tailed);
Bn—band *n*;
PD—population density;
SPD—square root of population density;
LPD—natural logarithm of population;
PCs—principal components;
VIs—vegetation indices;
Frac.—fraction images;
GV—green vegetation fraction;
IMP—impervious surface fraction;
Text.—texture;
Temp.—temperature.

Population Estimation Based on Stratified Sampling Scheme

Table 6 shows correlation coefficients between population parameters and remote sensing variables in the low, medium, and high population density categories. It is clear that in the low-density category, correlations were not as strong as those in medium and high density. Similarly, as with the non-stratified scheme, in the medium and high-density categories, temperature had the strongest positive correlation with population, while vegetation related variables had negative correlations with population. The low correlation between remote sensing variables and population in the low-density category imply that population estimation for these areas were more complicated, and the issue warrants further studies.

The R^2 values for individual regression models are summarized in Table 7. In the low- density and high-density categories, the highest R^2 were only 0.13 and 0.180, respectively. This indicates that Landsat ETM+ data may not be suitable for population estimation in these categories. In the medium density category, the combination of vegetation indices or vegetation related variables, and the incorporation of spectral response, textures and temperature can provide good estimations, especially the later when R^2 reached as

TABLE 4. COMPARISON OF REGRESSION RESULTS FOR POPULATION DENSITY ESTIMATION BASED ON NON-STRATIFIED SAMPLES

Potential Variables	PD		SPD		LPD	
	Selected Var.	R^2	Selected Var.	R^2	Selected Var.	R^2
Bands	B4	0.065	B5, B1, B2	0.212	B1, B2, B5	0.160
PCs	PC2, PC3	0.159	PC2, PC3	0.134	PC2, PC3, PC1	0.107
VIs	RVI, TNDVI, SAVI, RDVI	0.622	TNDVI, SAVI, RDVI, RVI	0.645	TNDVI, SAVI, RDVI	0.548
Frac.	GV, IMP	0.079	GV, IMP	0.065	non*	
Text.	B7_7 × 7, B7_3 × 3, B3_3 × 3	0.369	B7_7 × 7, B3_3 × 3, B3_5 × 5	0.465	B7_3 × 3, B7_5 × 5	0.448
Temp.	TEMP	0.269	TEMP	0.263	TEMP	0.169
VRV	RVI, TNDVI, SAVI, PC2, B4	0.768	RVI, TNDVI, SAVI, PC2, B4	0.797	RVI, TNDVI, SAVI, PC2, B4	0.678
B-temp	Temp, B5	0.351	Temp, B7	0.376	Temp, B7	0.338
Mixture	B7_7 × 7, RVI, B2, TNDVI, SAVI, B5	0.785	TEMP, RVI, TNDVI, SAVI, B5, RDVI, SVI	0.828	TNDVI, SAVI, B5, TEMP, RVI	0.698

Notes: VRV—vegetation related variables, including band 4, PC2, VIs, and GV;
B-temp—combination bands and temperature;
Mixture—combination of all variables.

TABLE 5. SUMMARY OF SELECTED ESTIMATION MODELS FOR POPULATION DENSITY ESTIMATION BASED ON NON-STRATIFIED SAMPLES

Model		Var.	Regression Equation	R^2	RE	AE
PD	1	Mixture	−83613.428 − 58.830*B7_7 × 7 + 5914.817*RVI + 117300.115*TNDVI − 65068.691*SAVI − 65.723*B5 + 64.369*B2	0.785	204.3	505
	2	VRV	−95394.477 + 6378.881*RVI + 132709.023*TNDVI − 73728.142*SAVI − 137.526*PC2 + 129.704*B4	0.768	204.4	523
SPD	3	Mixture	−1293.678 + 1.318*TEMP + 57.79*RVI + 1347.089*TNDVI − 789.683*SAVI − 1.124*B5 − 11.674*RDVI + 1.325*SVI	0.828	123.1	439
	4	VRV	−1226.463 + 72.752*RVI + 1754.789*TNDVI − 1.915*PC2 − 945.565*SAVI+ 1.742*B4	0.797	142.1	452

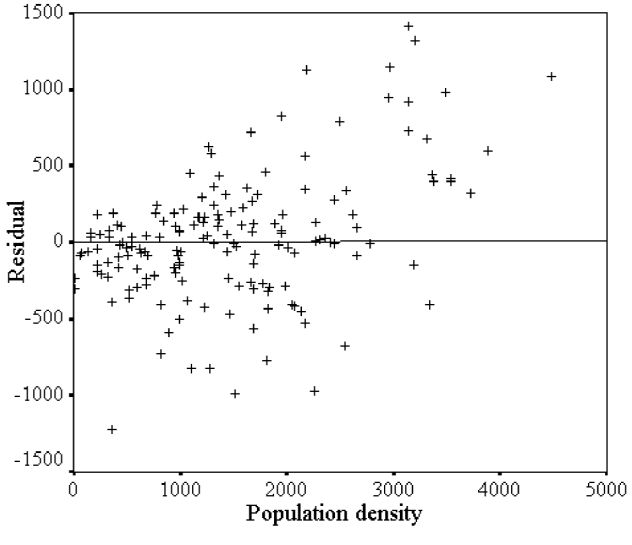

Figure 4. Residual distribution from model 4, negative indicates overestimated and positive indicates underestimated.

high as 0.87, 0.86, and 0.83 for different forms of dependent variables. Overall, the transforms of population density did not significantly improve the estimation in these categories.

Table 8 displays the four best models, selected based on R^2 and estimation errors. It shows that the results of population estimation for the low-density category using remote sensing variables were not satisfactory due to its high estimation errors. In the medium density category, both models provided very good estimations, using vegetation related independent variables, or using a combination of spectral, texture, and temperature variables. The R^2 values reached 0.83 and 0.86, respectively with a relative error of less than 29 percent, and an absolute error of less than 384 in both models (compared with the mean values of 1,417). For the high-density category, the model using temperature as the only independent variable yielded the best estimation result. The relative error was only 11.4 percent, but absolute error reached 429 (compared with mean value of 3,707). Overall, the performance of estimation models was much improved after stratification of population density into three categories. This finding implies that stratification based on population density is necessary for developing population estimation models using remotely sensed data.

Figure 5 shows the distribution of residuals when combination of Model 5, 7, and 8 was applied to predict the population of Indianapolis in 2000. Most underestimations and overestimations were located in the central part of the city. For example, BGs 1, 2, 3, 4, and 5 (marked on Figure 5) with very high population densities, were greatly underestimated. These BGs usually had several multi-story apartments for residential use (Figure 6). On the other hand, most overestimated BGs were found in the downtown area, where commercial or service uses dominated. For example, the most overestimated BG, BG3 had a population density of 678/ km^2 according to the census data, but the estimated population density reached up to 1,884/ km^2. The second-most overestimated area was observed in BG7 (Figure 7), where university and residential uses shared the land.

Based on the models developed for estimating population densities, the population of individual block groups can be calculated, and the total population of the whole city can be summed up. The total population estimates were 832,792 with relative error of 3.2 percent using stratified scheme (i.e., combination of models 5, 7, and 8)

TABLE 6. RELATIONSHIPS BETWEEN POPULATION PARAMETERS AND REMOTE SENSING VARIABLES BASED ON STRATIFIED SAMPLES

Remote Sensing Variables		Low Density			Medium Density			High Density		
		PD	SPD	LPD	PD	SPD	LPD	PD	SPD	LPD
ETM	B1	−0.231[b]	−0.237[b]	−0.232[b]	0.398[a]	0.398[a]	0.391[a]	0.274[b]	0.269[b]	0.264[b]
	B2	−0.232[b]	−0.234[b]	−0.230[b]	0.340[a]	0.342[a]	0.338[a]	0.248[b]	0.243[b]	0.237[b]
	B3	−0.244[b]	−0.245[b]	−0.237[b]	0.349[a]	0.351[a]	0.346[a]	0.267[b]	0.263[b]	0.259[b]
	B4	0.223	0.231[b]	0.207	−0.354[a]	−0.335[a]	−0.304[a]	−0.371[a]	−0.371[a]	−0.371[a]
	B5	−0.164	−0.141	−0.132	−0.060	−0.047	−0.029	−0.085	−0.087	−0.089
	B7	−0.256[b]	−0.248[b]	−0.234[b]	0.243[a]	0.247[a]	0.246[a]	0.194	0.191	0.188
PCs	PC1	−0.249[b]	−0.247[b]	−0.237[b]	0.302[a]	0.304[a]	0.302[a]	0.231	0.227	0.223
	PC2	0.164	0.181	0.168	−0.391[a]	−0.373[a]	−0.347[a]	−0.379[a]	−0.378[a]	−0.377[a]
	PC3	−0.001	0.027	0.068	−0.269[a]	−0.291[a]	−0.316[a]	−0.019	−0.009	0.001
VIs	NDVI	0.253[b]	0.257[b]	0.237[b]	−0.388[a]	−0.376[a]	−0.354[a]	−0.346[a]	−0.344[a]	−0.342[a]
	RDVI	0.255[b]	0.257[b]	0.242[b]	−0.411[a]	−0.409[a]	−0.398[a]	−0.318[a]	−0.315[a]	−0.312[a]
	SAVI	0.253[b]	0.257[b]	0.237[b]	−0.388[a]	−0.377[a]	−0.354[a]	−0.347[a]	−0.345[a]	−0.342[a]
	SVI	0.252[b]	0.256[b]	0.241[b]	−0.392[a]	−0.384[a]	−0.365[a]	−0.340[a]	−0.337[a]	−0.335[a]
	RVI	0.211	0.210	0.191	−0.514[a]	−0.506[a]	−0.485[a]	−0.354[a]	−0.353[a]	−0.351[a]
	TNDVI	0.260[b]	0.266[b]	0.246[b]	−0.320[a]	−0.308[a]	−0.286[a]	−0.335[a]	−0.333[a]	−0.330[a]
Frac.	GV	0.254[b]	0.258[b]	0.238[b]	−0.388[a]	−0.376[a]	−0.353[a]	−0.357[a]	−0.356[a]	−0.355[a]
	IMP	−0.273[b]	−0.266[b]	−0.249[b]	0.291[a]	0.290[a]	0.283[a]	0.264[b]	0.262[b]	0.260[b]
Text.	B3_3 × 3	−0.119	−0.124	−0.129	−0.047	−0.025	0.001	−0.123	−0.125	−0.127
	B7_3 × 3	−0.149	−0.143	−0.133	−0.150	−0.138	−0.123	−0.130	−0.132	−0.133
	B3_5 × 5	−0.104	−0.111	−0.125	−0.133	−0.114	−0.090	−0.103	−0.105	−0.107
	B7_5 × 5	−0.142	−0.137	−0.132	−0.218[b]	−0.210[b]	−0.198[b]	−0.106	−0.108	−0.109
	B3_7 × 7	−0.093	−0.102	−0.122	−0.177	−0.159	−0.136	−0.080	−0.083	−0.085
	B7_7 × 7	−0.136	−0.132	−0.130	−0.251[a]	−0.244[a]	−0.234[b]	−0.076	−0.078	−0.080
Temp	TEMP	−0.234[b]	−0.231[b]	−0.212	0.622[a]	0.635[a]	0.637[a]	0.425[a]	0.423[a]	0.420[a]

[a]Correlation at 99 percent confidence level (2-tailed);
[b]Correlation at 95 percent confidence level (2-tailed).

TABLE 7. COMPARISON OF REGRESSION RESULTS FOR DIFFERENT POPULATION DENSITY CATEGORIES

Scale	Var.	PD		SPD		LPD	
		Selected Var.	R^2	Selected Var.	R^2	Selected Var.	R^2
Low	Bands	B7	0.066	B7	0.062	B3	0.056
	PCs	PC1	0.062	PC1	0.061	PC1	0.056
	VIs or VRV	TNDVI	0.068	TNDVI	0.071	TNDVI	0.061
	Frac.	IMP	0.075	IMP	0.071	IMP	0.062
	Text.	non*		non*		non*	
	Temp.	TEMP	0.055	TEMP	0.053	TEMP	0.045
	Mixture	IMP, B3_7 × 7	0.130	IMP	0.071	IMP	0.062
Medium	Bands	B1, B2	0.290	B1, B2	0.283	B1, B2, B4	0.290
	PCs	PC2, PC3	0.277	PC2, PC3	0.277	PC2, PC3	0.269
	VIs	TNDVI, SAVI, RDVI, SVI	0.641	TNDVI, SAVI	0.611	TNDVI, SAVI	0.597
	Frac.	GV	0.151	GV	0.141	GV	0.124
	Text.	B7_7 × 7, B7_3 × 3	0.215	B7_7 × 7, B7_3 × 3	0.230	B3_3 × 3, B3_5 × 5	0.296
	Temp.	TEMP	0.387	TEMP	0.404	TEMP	0.407
	VRV	RVI, TNDVI, PC2, SAVI, RDVI, SVI	0.825	RVI, TNDVI, PC2, SAVI, RDVI, SVI, B4	0.829	TNDVI, PC2, RDVI, NDVI, SVI	0.794
	B-temp Mixture	TEMP, B7	0.442	TEMP, B7	0.461	TEMP, B7	0.466
		B3_7 × 7, RVI, TNDVI, SAVI, B5, PC1	0.869	TEMP, B3_7 × 7, RVI, TNDVI, SAVI, B5, PC1	0.863	TEMP, B7_7 × 7, RVI TNDVI, AVI, B5, RDVI	0.831
High	Bands	B4	0.138	B4	0.138	B4	0.138
	PCs/VRV	PC2	0.144	PC2	0.143	PC2	0.142
	VIs	RVI	0.125	RVI	0.124	RVI	0.123
	Frac.	GV	0.128	GV	0.127	GV	0.126
	Text.	non*		non*		non*	
	B-temp Mixture	TEMP	0.180	TEMP	0.179	TEMP	0.177

and 789,756 with relative error of 8.2 percent using the non-stratified model 3 (in Table 5). Remote sensing techniques can provide reasonably accurate estimation results for the total population, and dividing population density into different categories was more effective than conventional, non-stratified methods.

TABLE 8. BEST ESTIMATION MODELS FOR DIFFERENT DENSITY CATEGORIES BASED ON THE STRATIFIED SAMPLING SCHEME

Model		Potential Var.	Dep. Var.	Regression Equation	R^2	RE	AE
Low	5	Mixture	PD	296.190 – 9.136*IMP + 7.746*B3_7 × 7	0.130	315.3	97
Medium	6	VRV	SPD	−649.707 + 37.317*RVI + 1027.726*TNDVI – 1.579*PC2 –538.330*SAVI + 0.608*B4 –12.092*RDVI + 1.183*SVI	0.829	28.9	357
	7	Mixture	SPD	−966.765 + 0.684*TEMP – 0.632*B3_7 × 7 + 46.507 *RVI + 1110.756*TNDVI – 581.426*SAVI –1.003*B5 + 0.514*PC1	0.863	28.4	384
High	8	B-temp/Mixture	SPD	−565.332 + 2.047*TEMP	0.179	11.4	429

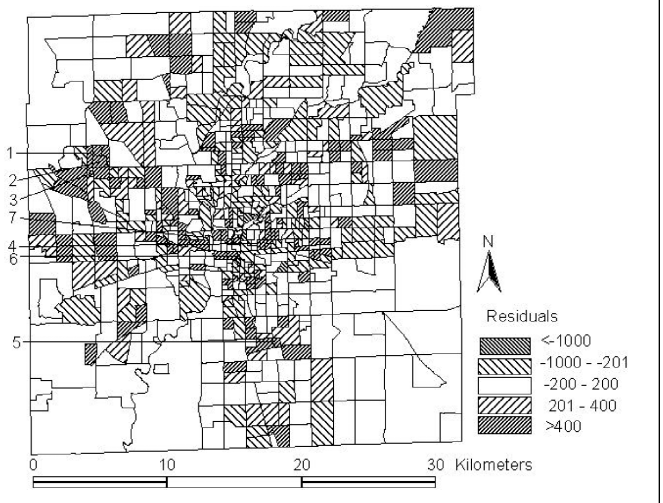

Figure 5. Residual map of estimated population densities based on the best model of the stratified sampling scheme. Negative value indicates overestimated and positive value indicates underestimated.

Figure 6. The most underestimated block group. The solid line shows the boundary of block groups. The background image is a digital orthophoto with a pixel resolution of 0.5 foot (international survey feet), which was acquired in March and April of 2002 and was provided for use by the Indianapolis Mapping and Geographic Infrastructure System.

Discussion and Conclusions

Discussion

Using remote sensing techniques to estimate population density is still a challenging task both in terms of theory and methodology, due to remotely sensed data, the complexity of urban landscapes, and the complexity of population distribution. Remotely sensed data are associated with the characteristics of surface features, but not directly related to population counts or population densities. For example, the areas with low population density may be located in industrial / commercial areas, or in forest / agriculture dominated areas, but the spectral characteristics of these landscapes are fundamentally different. The areas with high population density largely consist of multi-story apartments. The distribution of these high-density areas varies greatly with urban spatial structure which has identified three classical types: concentric zone, sector, and the multiple-nuclei city (Wheeler and Muller, 1981). In developed countries, substantial urban growth frequently occurs in suburban areas due to the re-distribution of population and decentralization of metropolitan urban functions; while in developing countries, urban expansion is more related to rapid population growth and industrialization. In the Eastern cities, such as Hong Kong, Singapore, and Shanghai, high population density is often observed in the central part of a city, where high-rise residential buildings coalesce. Optical remote sensing data, however, have not directly been linked to the vertical or internal features of such buildings for digital analysis of population estimation.

Remotely sensed data and census data are often collected with different formats and stored with different data structures. The values of any census variable are aggregated totals or mean values for the entire spatial extent of a census unit. In other words, census units are assumed to be homogenous, no matter what types of variation in land use are within them. A problem is created when unoccupied areas, such as water, airports, and forests, are given a population in the census. Remote sensing data have finer resolution than the census data. When integrating these two types of data, a common method is to aggregate remotely sensed data to an appropriate census level. For example, this research aggregated various remote sensing variables at a block group

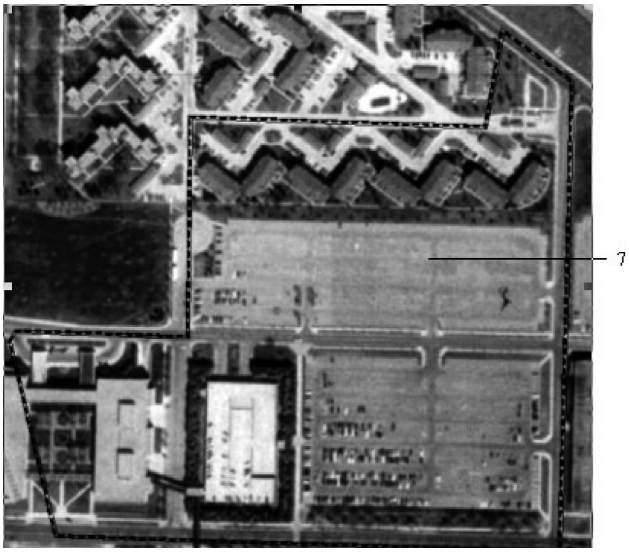

Figure 7. The most overestimated block group. The solid line shows the boundary of block groups. The background image is a digital orthophoto with a pixel resolution of 0.5 foot (international survey feet), which was acquired in March and April of 2002 and was provided for use by the Indianapolis Mapping and Geographic Infrastructure System.

level. This aggregation has the potential to result in the same values for different BGs in spite of their differences in land use and land cover types, causing errors for population estimation.

Because of the complexity of population distribution, a single model is often difficult to fit all the data. Stratification of population density in this study has proved to be effective in improving estimation results. However, the method of stratifying population based on density may result in spatial discontinuity of the data, and may be difficult to find suitable thresholds for stratification. The population categories identified here correspond, to some extent, to land cover categories in urban residential areas in the National Land Cover Database (NLCD) (Vogelman *et al.*, 1998; Vogelman *et al.*, 2001). Low intensity residential areas in NLCD, mostly single-family housing units, relate to the medium population density category, while high intensity residential, such as apartment complexes and row houses, relate to high population density category. Moreover, due to various factors impacting remotely sensed data quality, it is usually difficult to directly transfer the model developed in one site to other sites. Many factors need to be considered, including image acquisition date and time, the atmospheric condition when the image data were acquired, and the characteristics of the urban landscapes under investigation. Model transfer to nearby areas with similar socio-economic conditions is favorable, if these areas are within the same image scene or adjacent scenes in the same path of image acquisition.

Conclusions

Population estimation models developed based on the integration of satellite imagery and census data have numerous applications. They can be used to provide information on intra-urban population distribution, which is essential in urban planning, natural hazard risk assessment, disaster prevention and response, environmental impact assessment, transportation planning, economic decision-making, and evaluation of quality of life. They can also be applied to validate urban growth models if the time series image data become available.

This study demonstrates that Landsat ETM+ imagery could be used to provide reasonably accurate population density estimation by combining various remote sensing derived variables, such as original ETM+ bands, principal components, vegetation indices, fraction images, temperature, and textures as explanatory variables. Vegetation related variables were especially effective. Remote sensing based models were more suitable for estimation of population with medium density than with low- and high-densities. The stratification of population density into some categories and development of estimation models for individual categories improved the model performance. Although population estimation using remotely sensed data are not straightforward, especially for low- and high- density regions, a major advantage of this approach is that it can provide a timely update of a population database and its spatial distribution, which is impossible by conventional census approaches.

More research is needed to improve the population estimation through development of suitable models and use of multi-source data, such as high spatial resolution imagery and lidar, which are capable to identify individual buildings and measure the heights of the buildings. Comparative analyses of different methods are further suggested, for example, the method based on dwelling counts could be used to estimate population in low-density area using high spatial resolution remote sensing images. Further studies are warranted by incorporating Elvidge's radiance calibrated nighttime lights imagery with original frequency data from satellite imagery, since the former has been proved to be effective in estimating low and high population extremes (Elvidge *et al.*, 1995; 1997). In addition, the dynamic changes of population can be examined if multi-temporal remote sensing and census data become available.

Acknowledgments

The authors wish to thank three anonymous reviewers for their constructive comments and suggestions. We further acknowledge the financial supports of University Research Committee at Indiana State University through Grant #UNR184 awarded to Dr. Qihao Weng. Last but not least, suggestions, comments, and help provided by Dr. Dengsheng Lu, Dr. Paul Mausel, and Dr. Brian Ceh at various stages of this research and manuscript writing are greatly acknowledged.

References

Adams, J.B., D.E. Sabol, V. Kapos, R.A. Filho, D.A. Roberts, M.O. Smith, and A.R. Gillespie, 1995. Classification of multispectral images based on fractions of endmembers: application to land cover change in the Brazilian Amazon, *Remote Sensing of Environment*, 52:137–154.

Aguiar, A.P.D., Y.E. Shimabukuro, and N.D.A. Mascarenhas, 1999. Use of synthetic bands derived from mixing models in the multispectral classification of remote sensing images, *International Journal of Remote Sensing*, 20:647–657.

Asner, G.P., and D.B. Lobell, 2000. A biogeophysical approach for automated SWIR unmixing of soils and vegetation, *Remote Sensing of Environment*, 74:99–112.

Bannari, A., D. Morin, F. Bonn, and A.R. Huete, 1995. A review of vegetation indices, *Remote Sensing Reviews*, 13:95–120.

Barnsley, M.J., 1999. Digital remote sensing data and their characteristics, *Geographical Information Systems: Principles, Techniques, Applications, and Management* (Second Edition)

(P. Longley, M. Goodchild, D.J. Maguire, and D.W. Rhind, editors), New York: John Wiley & Sons, pp. 451–466.

Birth, G.S., and G. McVey, 1968. Measuring the color of growing turf with a reflectance spectrophotometer, *Agronomy Journal*, 60:640–643.

Chen, K., 2002. An approach to linking remotely sensed data and areal census data, *International Journal of Remote Sensing*, 23:37–48.

Cochrane, M.A., and C.M. Souza, Jr., 1998. Linear mixture model classification of burned forests in the eastern Amazon, *International Journal of Remote Sensing*, 19:3433–3440.

Deering, D.W., J.W. Rouse, R.H. Haas, and J.A. Schell, 1975. Measuring forage production of grazing units from Landsat MSS data, *Proceedings of Tenth International Symposium on Remote Sensing of Environment*, Ann Arbor, ERIM, 2, pp. 1169–1178.

Dobson, J.E., E.A. Bright, P.R. Coleman, R.C. Durfee, and B.A. Worley, 2000. LandScan: A global population database for estimating populations at risk, *Photogrammetric Engineering & Remote Sensing*, 66(7):849–857.

Dobson, J.E., 2003. Estimating population at risk, *Geographical Dimensions of Terrorism* (S.L. Cutter, D.B. Richardson, and T.J. Wilbanks, editors), Routledge, New York and London, pp. 161–167.

Dobson, J.E., E.A. Bright, P.R. Coleman, and B.L. Bhaduri, 2003. LandScan2000: A new global population geography, *Remotely-Sensed Cities* (V. Mesev, editor), Taylor and Francis, London, pp. 267–279.

Elvidge, C.D., K.E. Baugh, E.A. Kihn, H.W. Kroehl, and E.R. Davis, 1995. Mapping city lights with nighttime data from the DMSP operational linescan system, *Photogrammetric Engineering & Remote Sensing*, 63:727–734.

Elvidge, C.D., K.E. Baugh, V.R. Hobson, E.A. Kihn, H.W. Kroehl, E.R. Davis, and D. Cocero, 1997. Satellite inventory of human settlements using nocturnal radiation emissions: A contribution for the global toolchest, *Global Change Biology*, 3:387–395.

Flowerdew, R., and M. Green, 1992. Developments in areal interpolation methods and GIS, *Annals of Regional Science*, 26:67–78.

Forester, B.C., 1985. An examination of some problems and solutions in monitoring urban areas from satellite platforms, *International Journal of Remote Sensing*, 6:39–151.

Gong, P., and P.J. Howarth, 1992. Frequency-based contextual classification and gray-level vector reduction for land-use identification, *Photogrammetric Engineering & Remote Sensing*, 58:423–437.

Haack, B.N., S.C. Guptill, R.K. Holz, S.M. Jampoler, J.R. Jensen, and R.A. Welch, 1997. Urban analysis and planning, *Manual of Photographic Interpretation*, American Society for Photogrammtry and Remote Sensing, Bethesda, Maryland, pp. 517–553.

Haralick, R.M., K. Shanmugam, and I. Dinstein, 1973. Texture features for image classification, *IEEE Transactions on Systems, Man and Cybernetics*, 3:610–621.

Haralick, R.M. 1979. Statistical and structural approaches to texture, *Proceedings of the IEEE*, 67, Seattle, Washington, pp. 786–804.

Harvey, J.T., 2002a. Estimation census district population from satellite imagery: Some approaches and limitations, *International Journal of Remote sensing*, 23:2071–2095.

Harvey, J.T., 2002b. Population estimation models based on individual TM pixels, *Photogrammetric Engineering & Remote Sensing*, 68:1181–1192.

He, D.C., and L. Wang, 1990. Texture unit, textural spectrum and texture analysis, *IEEE Transactions on Geoscience and Remote Sensing*, 28:509–512.

Henderson, F.M., and Z. Xia, 1998. Radar Applications in Urban Analysis, Settlement Detection and Population Estimation, *Principal & Applications of Imaging Radars*, *Manual of Remote Sensing*, 3rd Edition, John Wiley & Sons, New York.

Holz, R., 1988. Population estimation of Colonias in the Low Rio Grande Valley using remote sensing techniques, *Paper Presented at the Annual Meeting of the Association of American Geographers*, Phoenix, Arizona.

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*, pp. 583–591.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI), *Remote Sensing of Environment*, 25:295–309.

Iisaka, J., and E. Hegedus, 1982. Population estimation from Landsat imagery, *Remote Sensing of Environment*, 12:259–272.

Imhoff, M.L., W.T. Lowrence, D.C. Stutzer, and C.D. Elvidge, 1997. A technique for using composite DMSP/OLS city lights satellite data to map urban area, *Remote Sensing of Environment*, 61: 361–370.

Jensen, J.R., 2000. *Remote Sensing of the Environment*: *An Earth Resource Perspective*. Prentice Hall, Upper Saddle River, New Jersey, 544 p.

Jensen, J.R., and D.C. Cowen, 1999. Remote sensing of urban/suburban infrastructure and socio-economic attributes, *Photogrammetric Engineering & Remote Sensing*, 65:611–622.

Langford, M., D.J. Maguire, and D.J. Unwin, 1991. The areal interpolation problem: estimating population using remote sensing in a GIS framework, *Handing Geographical Information: Methodology and Potential Applications* (L. Masser and M. Blakemore, editors), Longman Scientific & Technical and John Wiley & Sons, Inc., New York.

Lindgren, D.T, 1971. Dwelling unit estimation with color-IR photos, *Photogrammetric Engineering*, 37:373–378.

Lindgren, D.T., 1985. *Land Use Planning and Remote Sensing*, Martinus Nijhoff Inc., Boston, 230 p.

Lo, C.P., and R. Welch, 1977. Chinese urban population estimation, *Annals of the Association of American Geographers*, 67:246–53.

Lo, C.P., 1986a. *Applied Remote Sensing*, Longman, New York.

Lo, C.P., 1986b. Accuracy of population estimation from medium-scale aerial photography, *Photogrammetric Engineering & Remote Sensing*, 52:1859–1869.

Lo, C.P., 1995. Automated population and dwelling unit estimation from high resolution satellite images: A GIS approach, *International Journal of Remote Sensing*, 16:17–34.

Lo, C.P, 2001. Modeling the population of China using DMSP operational linescan system nighttime data, *Photogrammetric Engineering & Remote Sensing*, 67:1037–1047.

Lu, D., Moran, E., and M. Batistella, 2003. Linear mixture model applied to Amazonian vegetation classification, *Remote Sensing of Environment*, 87(4):456–469.

Lu, D., and Q. Weng. 2004. Spectral mixture analysis of the urban landscape in Indianapolis with Landsat ETM+ imagery, *Photogrammetric Engineering & Remote Sensing*, 70(9): 1053–1062.

Marceau, D.J., P.J. Howarth, J.M. Dubois, and D.J. Gratton, 1990. Evaluation of the grey-level co-occurrence matrix method for land-cover classification using SPOT imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 28:513–519.

McGwire, K., T. Minor, and L. Fenstermaker, 2000. Hyperspectral mixture modeling for quantifying sparse vegetation cover in arid environments, *Remote Sensing of Environment*, 72: 360–374.

Mustard, J.F., and J.M. Sunshine, 1999. Spectral analysis for earth science: investigations using remote sensing data, *Remote Sensing for the Earth Sciences: Manual of Remote Sensing*, Volume 3, 3rd Edition, (A.N. Rencz, editor), John Wiley & Sons, Inc., New York, pp. 251–307.

Narasimha-Rao, P.V., M.V.R. Sesha Sai, K. Sreenivas, M.V. Krishna-Rao, B.R.M. Rao, R.S. Dwivedi, and L. Venkataratnam, 2002. Textural analysis of IRS-1D panchromatic data for land cover classification, *International Journal of Remote Sensing*, 23:3327–3345.

Ogrosky, C.E, 1975. Population estimation from satellite imagery, *Photogrammetric Engineering & Remote Sensing*, 41:707–712.

Phinn, S., M. Stanford, P. Scarth, A.T. Murray, and P.T. Shyy, 2002. Monitoring the composition of urban environments based on the vegetation-impervious surface-soil (VIS) model by subpixel analysis techniques, *International Journal of Remote Sensing*, 23:4131–4153.

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

Qiu, F., K.L. Woller, and R. Briggs, 2003. Modeling urban population growth from remotely sensed imagery and TIGER GIS road data, *Photogrammetric Engineering & Remote Sensing*, 69(9):1031–1042.

Rashed, T., J.R. Weeks, M.S. Gadalla, and A.G. Hill, 2001. Revealing the anatomy of cities through spectral mixture analysis of multispectral satellite imagery: a case study of the Greater Cairo region, Egypt, *Geocarto International*, 16:5–15.

Richards, J.A., 1994. *Remote Sensing Digital Image Analysis: An Introduction*, Springer-Verlag, Berlin, Germany, 340 p.

Roberts, D.A., G.T. Batista, J.L.G. Pereira, E.K. Waller, and B.W. Nelson, 1998. Change identification using multitemporal spectral mixture analysis: applications in eastern Amazônia, *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications* (R.S. Lunetta and C.D. Elvidge, editors), Ann Arbor Press, Ann Arbor, Michigan, pp. 137–161.

Roujean, J.L., and F.M. Breon, 1995. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements, *Remote Sensing of Environment*, 51:375–384.

Rouse, J.W., R.H. Haas, J.A. Schell, and D.W. Deering, 1974. Monitoring vegetation systems in the Great Plains with ERTS, *Proceedings of Third Earth Resources Technology Satellite-1 Symposium*, Greenbelt: NASA SP-351, pp. 310–317.

Shaban, M.A., and O. Dikshit, 2001. Improvement of classification in urban areas by the use of textural features: the case study of Lucknow city, Uttar Pradesh, *International Journal of Remote Sensing*, 22:565–593.

Small, C., 2001. Estimation of urban vegetation abundance by spectral mixture analysis, *International Journal of Remote Sensing*, 22:1305–1334.

Smith, M.O., S.L. Ustin, J.B. Adams, and A.R. Gillespie, 1990. Vegetation in deserts: I. A regional measure of abundance from multispectral images, *Remote Sensing of Environment*, 31:1–26.

Sutton, P., 1997. Modeling population density with nighttime satellite imagery and GIS, *Computers, Environment, and Urban System*, 21:227–244

Sutton, P., D. Roberts, C.D. Elvidge, and H. Meij, 1997. A comparison of nighttime satellite imagery and population density for the continental United States, *Photogrammetric Engineering & Remote Sensing*, 63:1303–1313.

Sutton, P., D. Roberts, C.D. Elvidge, and K. Baugh, 2001. Census from heaven: An estimate of the global human population using night-time satellite imagery, *International Journal of Remote Sensing*, 22:3061–3076.

Tobler, W.R., U. Deichmann, J. Gottsegen, and K. Maloy, 1995. *The Global Demography Project. Technical Report No. 95–6*, National Center for Geographic Information and Analysis. UCSB. Santa Barbara, California, 75 p.

Vogelman, J.E., T.L. Sohl, and S.M. Howard, 1998. Regional characterization of land cover using multiple sources of date, *Photogrammetric Engineering & Remote Sensing*, 64:45–57.

Vogelman, J.E., S.M. Howard, L. Yang, C.R. Larson, B.K. Wylie and N.V. Driel, 2001. Completion of the 1990s national land cover data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources, *Photogrammetric Engineering & Remote Sensing*, 67:650–662.

Watkins, J.F., and H.A. Morrow-Jones, 1985. Small area population estimates using aerial photography, *Photogrammetric Engineering & Remote Sensing*, 51:1933–1935.

Webster, C.J., 1996. Population and dwelling unit estimates from space, *Third World Planning Review*, 18:155–176.

Weeks, J.R., M.S. Gadalla, T. Rashed, J. Stanforth, and A.G. Hill, 2000. Spatial variability in fertility in Menoufia, Egypt, assessed through the application of remote sensing and GIS technologies, *Environment and Planning A*, 32:695–714

Welch, R., and S. Zupko, 1980. Urbanized area energy utilization patterns from DMSP data, *Photogrammetric Engineering & Remote Sensing*, 46:1107–1121.

Wellar, B.S., 1969. The role of space photography in urban and transportation data series, *Proceedings of the Sixth International Symposium on Remote sensing of Environment, II*, pp. s831–854.

Weng, Q., D. Lu, and J. Schubring. 2004. Estimation of land surface temperature-vegetation abundance relationship for urban heat island studies, *Remote Sensing of Environment*, 89(4):467–483.

Wheeler, J.O., and P.O. Muller, 1981. *Economic Geography*, New York: John Wiley & Sons, Inc., pp. 133–137.

Wright, J.K., 1936. A method of mapping densities of population with Cape Cod as an example, *Geographical Review*, 26: 103–110.

Wu, C., and Murray, A.T., 2003. Estimating impervious surface distribution by spectral mixture analysis, *Remote Sensing of Environment*, 84:493–505.

(Received 04 February 2004; accepted 05 May 2004; revised 15 May 2004)