Can. J. Remote Sensing, Vol. 36, No. 3, pp. 155–165, 2010

Fine-scale population estimation: how Landsat ETM+ imagery can improve population distribution mapping

Guiying Li and Qihao Weng

Abstract. The spectral response based and land use based methods were compared for population estimation with Landsat Enhanced Thematic Mapper Plus (ETM+) imagery for Marion County, Indiana, USA. With the spectral response based method, impervious surface and vegetation fractions and land surface temperature derived from the Landsat ETM+ thermal band along with spectral bands were tested for estimating population density at the census block group level. With the land use based method, land use and land cover images were first extracted from Landsat ETM+ reflected bands. The population count for the block groups was estimated based on the area of residential land use. Population was then dasymetrically redistributed onto a 30 m grid based on a land cover and land use map of the study area. A comparative analysis shows that the spectral response method produced a larger mean relative error of 237% for population density at the block group, whereas the land use based method yielded a much better estimation, with a mean relative error of 21.4% for population counts for the block group. It is suggested that the dasymetric map provides a more accurate portrayal of population distribution than a choropleth map at small geographical scales such as that of a US county.

Résumé. Les méthodes basées sur la réponse spectrale et l'utilisation du sol ont été comparées pour l'estimation de la population à l'aide des données ETM+ (« Enhanced Thematic Mapper Plus ») de Landsat dans le comté de Marion, en Indiana, É.-U. Dans le cas de la méthode basée sur la réponse spectrale, les fractions de surfaces imperméables et de végétation ainsi que la température de la surface du sol dérivée de la bande thermique de ETM+ de même que des bandes spectrales ont été testées pour estimer la densité de population au niveau du groupe d'îlots de recensement (« census block group »). Dans le cas de la méthode basée sur l'utilisation du sol, l'image de l'utilisation du sol et du couvert a été préalablement extraite des bandes réfléchies de ETM+ de Landsat. Le décompte de la population pour les groupes d'îlots a été estimé à partir de la surface d'utilisation du sol de type résidentiel. La population a ensuite été redistribuée par méthode dasymétrique sur une grille de 30 m basée sur la carte du couvert et de l'utilisation du sol de la zone d'étude. Une analyse comparative montre que la méthode basée sur la réponse spectrale a donné une erreur relative moyenne plus grande de 237 % pour la densité de population au niveau du groupe d'îlots, alors que la méthode basée sur l'utilisation du sol a donné une bien meilleure estimation, avec une erreur relative moyenne de 21,4 % pour le décompte de la population au niveau du groupe d'îlots. Les résultats suggèrent que la carte dasymétrique donne un portrait plus précis de la répartition de la population que la carte choroplèthe à des petites échelles géographiques comme celle d'un comté américain.
[Traduit par la Rédaction]

Introduction

Population growth and urban expansion in the past decades have occurred at magnitudes unprecedented in human history (UNFPA, 2008), producing considerable impacts on socioeconomic development, availability of resources, and environmental protection at the local, regional, and global scales. Obtaining timely and accurate information on urban characteristics and associated demographic information is valuable for both urban management and decision-making (Fabos and Petrasovits, 1984) and for socioeconomic and environment related studies. The traditional census survey method cannot meet the needs of contemporary socioeconomic development and urban planning and management. It is important to develop suitable approaches for estimating population and mapping its distribution in a timely and cost-effective manner. Remote sensing data with their advantages with respect to spectral, spatial, and temporal resolutions have demonstrated to be effective in providing information on physical characteristics of urban areas, including size, shape, and rates of change, and have been widely used for mapping and monitoring urban biophysical features (Haack

Received 2 October 2009. Accepted 9 June 2010. Published on the Web at http://pubservices.nrc-cnrc.ca/cjrs on 28 October 2010.

G. Li. Anthropological Center for Training and Research on Global Environmental Change, Indiana University, Bloomington, IN 47405, USA.

Q. Weng.[1] Center for Urban and Environmental Change, Department of Earth and Environmental Systems, Indiana State University, Terre Haute, IN 47809, USA.

[1]Corresponding author (e-mail: qweng@indstate.edu).

et al., 1997; Jensen and Cowen, 1999; Weng et al. 2006; 2008).

Population estimation with remote sensing data has long been explored, and many approaches have been developed based on aerial photographs and satellite images, including Landsat, Satellite pour l'Observation de la Terre (SPOT), night light of the Defense Meteorological Satellite Program (DMSP) operational line scanner, spaceborne imaging radar A (SIR-A), and radar QuickSCAT (Hsu, 1971; Lindgren, 1971; Lo and Welch, 1977; Lo, 1986a; 1986b; 1995; 2001; 2008; Iisaka and Hegedus, 1982; Meliá and Sobrino, 1987; Langford et al., 1991; Elvidge et al., 1995; 1997; Sutton, 1997; Sutton et al., 1997; 2001; Harvey, 2000; 2002a; 2002b; Li and Weng, 2005; Wu and Murray, 2005; 2007; Lu et al., 2006; Briggs et al., 2007; Nghiem et al., 2009; Viel and Tran, 2009). Wu et al. (2005) grouped population estimation methods into two categories, namely statistical modeling and areal interpolation. The statistical modeling method established an estimation model based on the relationship between population and remote sensing derived variables, and the areal interpolation method generated a refined population surface by redistributing census population data into specific pixel sizes. The statistical modeling approach can be further categorized into four types depending on the information extracted from the remote sensing data: (*i*) measurement of built-up areas, (*ii*) counting of dwelling units, (*iii*) measurements of land use areas, and (*iv*) image pixel characteristics. Each method has its own advantages and disadvantages depending on the geographical scale of study and the remote sensing data used (Lo, 1986a; Li and Weng, 2005). Statistical models are usually used to estimate population for a specific areal unit such as a census unit, which has arbitrary boundaries. However, some researchers have attempted to integrate census data with data from different spatial divisions, which are rarely consistent with each other. This process is called areal interpolation or cross-area estimation (Goodchild and Lam, 1980; Langford et al., 1991; Goodchild et al., 1993). Different methods of areal interpolation have been developed (Goodchild and Lam, 1980; Goodchild et al., 1993), including the dasymetric method. The dasymetric method is defined as utilizing ancillary data to redistribute population data from arbitrarily delineated enumeration districts like a census unit into zones of increased homogeneity with the purpose of better representation of the underlying statistical surface (Eicher and Brewer, 2001). Previous studies show that with ancillary data such as remote sensing derived products, the dasymetric mapping technique can provide a visual and statistically more accurate representation of population distribution than that represented by aggregated units, such as a census block group or tract (Eicher and Brewer, 2001; Mennis, 2003; Briggs et al., 2007; Maantay et al, 2008).

Although many methods for population estimation have been developed, rarely has research combined the statistical-based and dasymetric-based methods (Langford et al., 1991; Yuan et al., 1997; Harvey, 2002b). Except for the study by Harvey (2002b), previous studies have redistributed the census population to all land use and land cover (LULC) classes instead of to populated or residential areas, leading to unoccupied areas being populated. The objectives of this research are three-fold: (*i*) examine the utility of land surface temperature and fraction images derived from a subpixel classifier for population estimation; (*ii*) compare the effectiveness of the spectral response based and the land use based methods for population estimation of US census block groups; and (*iii*) produce a more accurate presentation of population distribution by combining dasymetric mapping with the land use based method.

Study area and datasets

Indianapolis (Marion County), Indiana, USA (**Figure 1**), was chosen as the study area. According to the US Census Bureau, Marion County has a total area of 1044 km^2, a population of 860 454 with a density of 838 people/km^2, and 387 183 housing units. With its large population, Indianapolis ranked as the twelfth largest city in the US in 2000 (US Census Bureau, Census 2000 summary file). In recent decades, the city has been expanding by encroaching on agricultural land and other nonurban lands due to population increases and economic growth. From 1990 to 2000, the population of Marion County increased by approximately 7.9%.

The primary data sources used in this research include Census 2000 and Landsat ETM+ imagery. The Landsat ETM+ image (L1G product, path 21, row 32) was acquired on 22 June 2000 under clear sky conditions. Although the L1G Landsat ETM+ data were geometrically corrected, their geometrical accuracy was not high enough for combining them with other high-resolution datasets. Hence, the image was further rectified to the Universal Transverse Mercator (UTM) coordinate system based on 1 : 24 000 scale topographic maps. A root mean square error of less than 0.5 pixels was obtained in the rectification.

The Census 2000 data at the block and block group level were downloaded from the US Bureau of the Census. Ancillary data used in this research included aerial photographs taken in the summer of 2003 and zoning data provided by the Indianapolis City Metropolitan Planning Department. All data were co-registered to UTM coordinates before integration for subsequent analysis.

Methodology

Two methods for population estimation were explored, namely (*i*) the combination of spectral bands, land surface temperature, vegetation abundance, and impervious surface; and (*ii*) residential land use and dasymetric mapping based on land use. Four steps were involved in population estimation: (*i*) extraction of population from the census data, (*ii*) development of remote sensing variables,

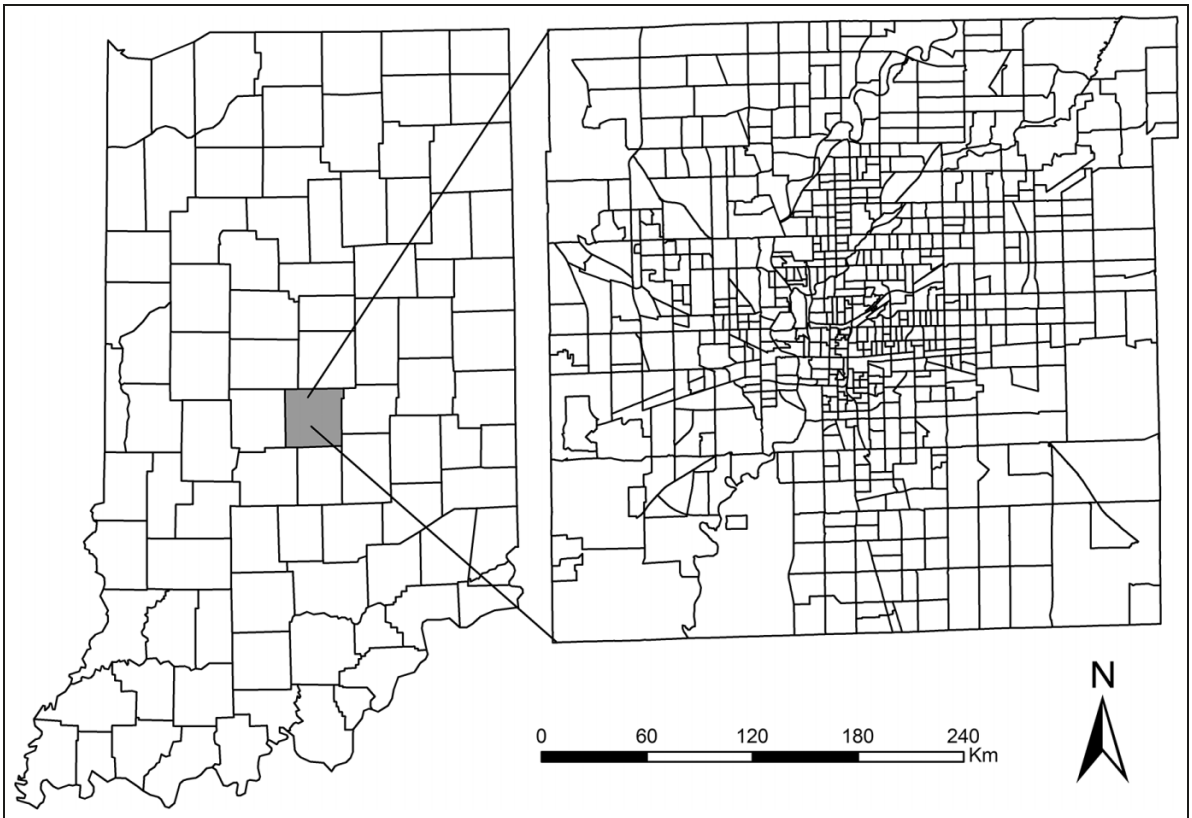

Figure 1. Study area of Marion County, Indiana.

(*iii*) development of population estimation models by the integration of population and remote sensing derived variables, and (*iv*) accuracy assessment. **Figure 2** illustrates the analytical procedures for population estimation.

Extraction of population parameters

The population data for Marion County at the block group level were extracted from the 2000 census data. Two population parameters were used in this research, namely total population and population density. Population density was calculated by dividing the total population in a block group by the corresponding area. Marion County has 658 block groups, with an average area of 1.59 km^2 and an average population density of 1587 people/km^2.

Extraction of remote sensing variables

Development of fraction images

The linear spectral mixture analysis (LSMA) approach has been widely used for deriving biophysical parameters from medium-resolution satellite images (Weng et al., 2004; Weng and Hu, 2008; Weng and Lu, 2009). LSMA assumes that the spectral signature of the mixed pixel measured by a sensor is a linear combination of the spectral signatures of all pure components, called endmembers, within the pixel (Adams et al., 1995; Roberts et al., 1998). In this study, LSMA was used to derive green vegetation and impervious surface fraction images. Details of the procedure are described by Lu and Weng (2006).

Computation of land surface temperature (LST)

The LST image was derived from the Landsat ETM+ thermal band (TIR) (10.44–12.42 μm), with a spatial resolution of 60 m. A detailed description of the LST calculation can be found in Weng et al. (2004).

Urban LULC classification

Maximum likelihood classifier was used to classify the Landsat ETM+ image. Housing density at the block level, zoning data, and high spatial resolution aerial photographs were used to assist the selection of training samples. Eleven LULC classes (i.e., commercial areas, transportation, industrial areas, water, low-density residential, medium-density residential, high-density residential, grass, cropland, fallow, and forest) were initially classified. In the final classification result, six LULC classes were produced by combining commercial, transportation, and industrial as urban and crop, fallow, forest, and grass as vegetation (**Figure 3**). Since no universal standards exist for separating residential density into high, medium, and low classes, this study established its own criteria based on housing density and zoning data. The residential lands having housing units with densities of more than 1300 people/km^2 were assigned as high-density residential areas, those having housing units with densities

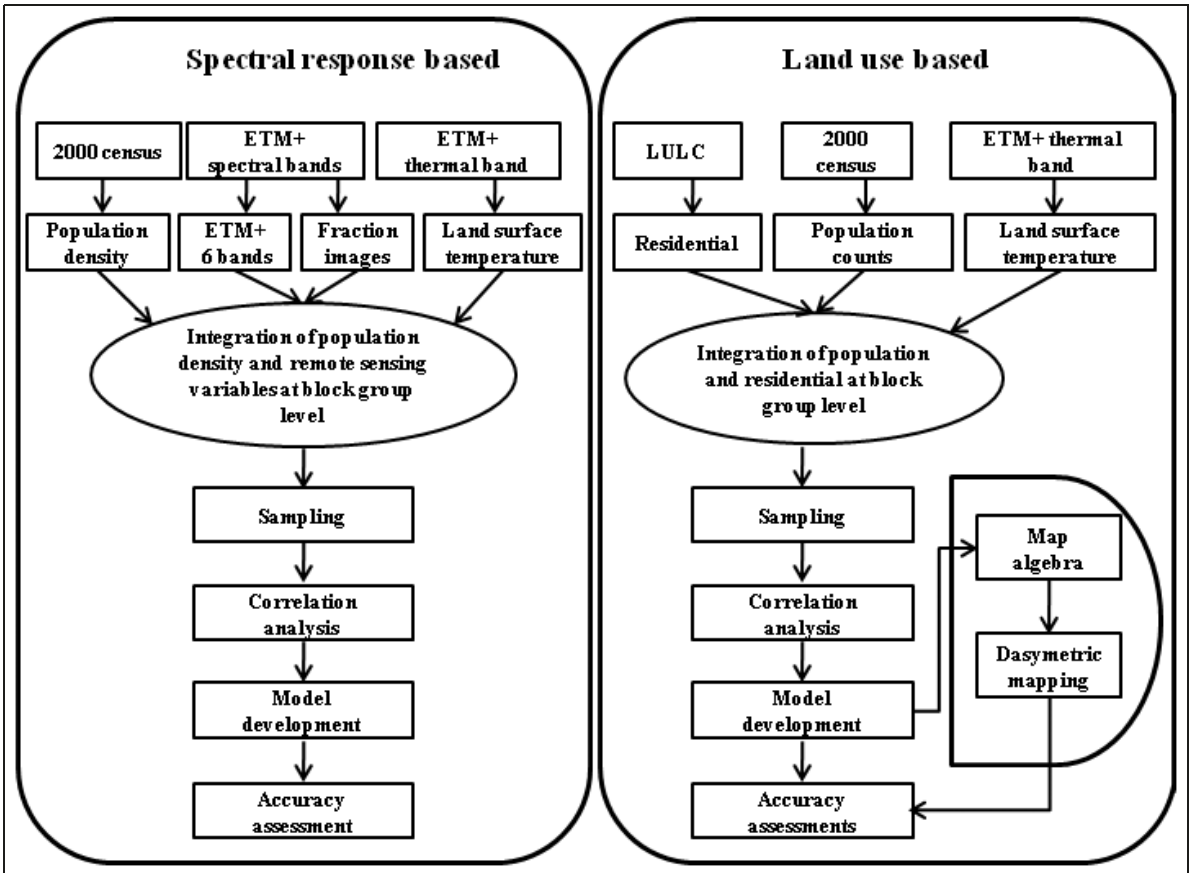

Figure 2. Procedures for population estimation.

of less than 400 people/km^2 as low-density residential areas, and those with densities in between as medium-density residential areas. The resulting image was refined by developing a decision tree based on housing density at the block level. Fifty samples for each LULC type were randomly selected and compared to references collected from high spatial resolution aerial photographs. The overall accuracy, producer's accuracy, user's accuracy, and Kappa statistic were calculated based on the error matrices (**Table 1**). The overall classification accuracy reached 86%.

Data integration and model development

Since population data at the block group have a different format and spatial resolution than remote sensing data, it is necessary to conduct data conversion before implementing further analyses. Remote sensing derived variables (e.g., spectral bands, fraction images, and land surface temperature) were aggregated to the block group level, and the mean values of these variables for the block group were then computed. LULC types were summarized as the count of pixels falling within a block group. In addition, the sum of temperature values for residential land use types within a block group was also calculated. All these remote sensing derived variables, as well as population data, were stored as the attributes of block groups and used for statistical analyses.

A random sampling technique was used to extract samples based on 25% of the total number of block groups. Initially, a total of 174 from 658 block groups were selected. A threshold of 2.5 standard deviations was used to identify the outliers for the samples during regression processing. Pearson's correlation coefficient was used to explore the correlation between population parameters and remote sensing variables. Furthermore, stepwise regression analysis was conducted to identify suitable variables for development of population estimation models.

There are two classes of statistical regression models depending on population parameters. The first class is designed to indicate the relationship between average population density in each census unit and scale-invariant indicators derived from remote sensing imagery (Harvey, 2002a; 2002b; Wu and Murray, 2005). The spectral response based method belongs to this category. Spectral response based population estimation was carried out with independent variable of six Landsat ETM+ spectral bands, fraction images, and land surface temperature. Population density was used as the dependent variable. Six Landsat ETM+ spectral bands were tested; vegetation fraction, impervious surface, and land surface temperature were then incorporated; and a series of regression models were developed. Another class of regression models aims at showing the relationship between population counts in each census unit and a number

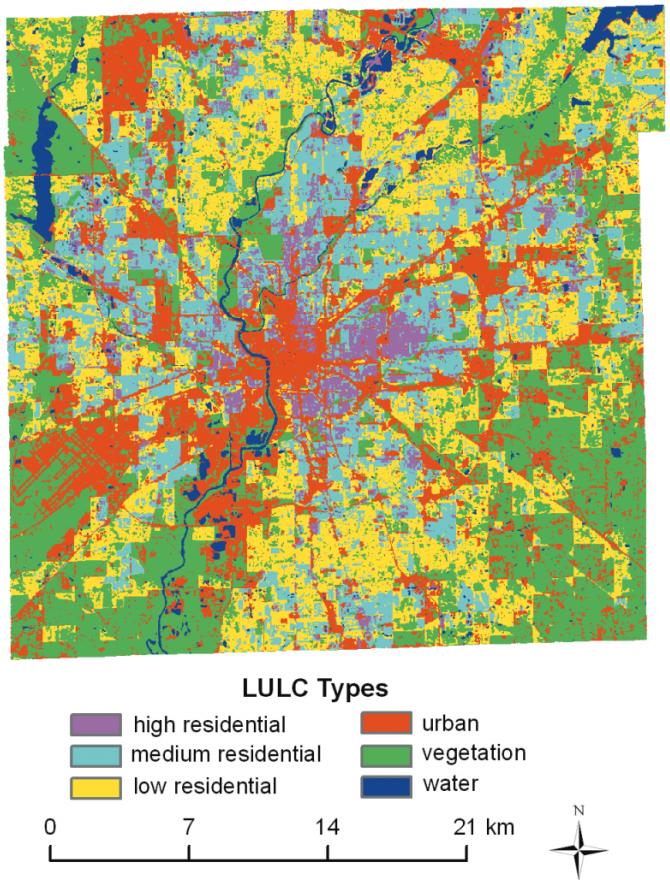

Figure 3. LULC map developed from Landsat ETM+ imagery with the maximum likelihood classifier.

of scale-dependent indicators (e.g., pixel counts) from remote sensing data for the corresponding census unit (Langford et al., 1991; Sutton et al., 1997). Land use based population estimation in this study belongs to this category and was conducted based on the urban LULC classification image. Population count was used as the dependent variable, and the areas of each type of residential land were used as independent variables. Three types of residential land use were examined using the following equation:

$$P_i = D_{\mathrm{l}}A_{\mathrm{l}i} + D_{\mathrm{m}}A_{\mathrm{m}i} + D_{\mathrm{h}}A_{\mathrm{h}i} \quad (1)$$

Table 1. Accuracy assessment of land use and land cover classification results.

LULC type	Producer's accuracy (%)	User's accuracy (%)
Water	97.9	95.9
Urban	86.6	80.6
High-density residential	51.6	72.7
Medium-density residential	81.4	74.5
Low-density residential	76.3	63.0
Vegetation	91.9	97.0
Overall accuracy (%)	86.0	
Overall kappa	0.81	

where P_i is the total population in census block group i; $A_{\mathrm{l}i}$, $A_{\mathrm{m}i}$, and $A_{\mathrm{h}i}$ are pixel counts of low, medium, and high residential land falling within block group i, respectively; and D_{l}, D_{m}, and D_{h} are coefficients of $A_{\mathrm{l}i}$, $A_{\mathrm{m}i}$, and $A_{\mathrm{h}i}$, respectively. Because Equation (1) did not take into account the variation within a particular residential land use type, land surface temperature was then incorporated into the regression model with the assumption that population had a strong correlation with land surface temperature:

$$P_i = D_{\mathrm{l}}A_{\mathrm{l}i} + D_{\mathrm{m}}A_{\mathrm{m}i} + D_{\mathrm{h}}A_{\mathrm{h}i} + \alpha T_i \quad (2)$$

where T_i is the sum of the temperatures of residential pixels in block group i, and α is a coefficient of T.

The population estimation for each block group was conducted using the models described previously. The dasymetric mapping method was used to redistribute population onto the LULC map. Different types of residential land were first extracted from the LULC map as binary images. For each type of residential land, the population for each pixel was calculated by multiplying a binary image by the corresponding coefficients obtained from the regression models. When land surface temperature was incorporated, the temperature of residential pixels was also used as a coefficient. A dasymetric population map was obtained by adding these images together. On the dasymetric population map, each pixel value was a population count. Because population was redistributed only onto residential lands, other LULC types had a value of zero. In addition, due to constant pixel size, the population count can also be considered as "population density," except that the total area was 900 m^2, not 1 km^2.

Accuracy assessment

When a developed model is applied for prediction, there are some discrepancies between estimated values and true values, and the differences are called residuals. It is necessary to examine whether the model fits with training datasets (the process of internal validation) or with other datasets that are not used as training sets (the process of external validation) (Harvey, 2002a). Mean relative error (RE) is often calculated to indicate estimation accuracy as follows:

$$\mathrm{RE} = \frac{\sum_{k=1}^{n} |(\mathrm{Pg} - \mathrm{Pe})/\mathrm{Pg}|}{n} \times 100 \quad (3)$$

where Pg and Pe are the referenced and estimated values of either the population or population density, respectively; and n is the number of block groups used for accuracy assessment. The smaller the value of RE, the better the model. In this study, all block groups were used to assess population estimation models. In addition, a residual distribution map was generated for the selected model and is used to show the geographical distribution of residuals. Higher positive residuals indicate larger underestimation, and higher negative residuals imply larger overestimation.

Table 2. Correlation between population density and the mean of Landsat ETM+ spectral bands and derived variables.

	B1	B2	B3	B4	B5	B7	GV	IMP	TEMP
PD	0.144	0.083	0.088	−0.253*	−0.281*	0.006	−0.178	0.458*	0.513*

Note: B1–B5 and B7, Landsat ETM+ bands 1–5 and 7; GV, vegetation fraction; IMP, impervious fraction; PD, population density; TEMP, land surface temperature.
*Significant at the 0.01 level (two-tailed test).

Table 3. Best models based on spectral data.

Model	Potential variables	Explanatory variables	R^2	RE (%)
1	Landsat ETM+ bands	B1 mean; B5 mean	0.18	
2	Landsat ETM+ bands; fractions; land surface temperature	TEMP mean; B7 mean	0.48	237

Note: For model 1, PD = 4083.082 − 61.875(B5 mean) + 36.677(B1 mean); for model 2, PD = −112 582.438 + 388.179(TEMP mean) − 61.511(B7 mean).

Results

Analysis of population estimation with spectral response based methods

Table 2 indicates that Landsat ETM+ bands 4 (near infrared) and 5 (middle infrared) were significantly correlated with population density. Impervious surface was better correlated with population density than vegetation abundance. Of the nine selected variables, land surface temperature was the most strongly correlated with population density. The stepwise regression analysis based on Landsat ETM+ reflective bands shows that the coefficient of determination (R^2) was only 0.18 with bands 1 and 5 as explanatory variables (**Table 3**). Obviously, Landsat ETM+ bands alone were not sufficient to accurately estimate population density. **Table 3** further indicates that the incorporation of land surface temperature in the regression model can significantly increase the R^2 value (from 0.18 to 0.48). The following is the best estimation model based on these variables:

$$\mathrm{PD} = -112\,582.438 + 388.179(\mathrm{TEMP}) - 61.511(\mathrm{B7}) \quad (4)$$

where TEMP is the mean temperature of the block group, and B7 is the mean value of band 7 (middle infrared) of the block group. This model was then used to calculate population density for each block group and the RE was computed.

The distribution of residuals from this model is illustrated in **Figure 4**. The block groups with extremely high population density (especially greater than 3000 people/km^2) were highly underestimated, and those with extremely low density (less than 400 people/km^2) were largely overestimated, leading to high relative error.

Analysis of population estimation with the land use based method

Table 4 shows correlation coefficients between population counts and residential land use and indicates that the correlation became stronger from low-density to high-density residential land use. The sum of land surface temperature had the strongest correlation with population counts. Two models based on these variables were then developed and are summarized in **Table 5**. These models were forced through the origin point based on the assumption that no people resided in block groups that had no residential land use. Model 3 was generated using only residential land uses as explanatory variables, and model 4 was developed by adding land surface temperature as an additional independent variable. The RE values show that the two models yielded a similar accuracy.

A comparative analysis indicates that the land use based method provided better estimation than the spectral response based method in terms of RE. Model 3, which was based on residential land uses, provided the best result for this study. Therefore, it was chosen to estimate population for the study area.

A residual distribution map based on model 3 was created (**Figure 5a)**. Blue and cyan indicate the overestimated population, and red and orange the underestimated population. Comparing the residual map to the population density distribution from census data (**Figure 5b**), the high-density block groups located in the center of the city tended to be underestimated. However, there was a tendency for overestimation of the low-density block groups, especially in rural areas such as the northwestern, southwestern, and southeastern corners of the city. In such areas, residential dwellings were scattered and highly mixed with vegetation, and thus were difficult to separate from surrounding vegetation.

The estimated population based on model 3 for each block group was redistributed within residential classes using the dasymetric method. The dasymetric map shows the real population distribution pattern, on which LULC types such as agriculture, forest, and water had no people (**Figure 6a**), and the presentation of population distribution by a choropleth map (**Figure 6b**) did not conform to the population distribution patterns in the real world.

Conclusions and discussions

This research compared two types of population estimation methods, namely spectral response based and land use based methods. In the spectral response method, impervious surface and vegetation image fractions and land surface tem-

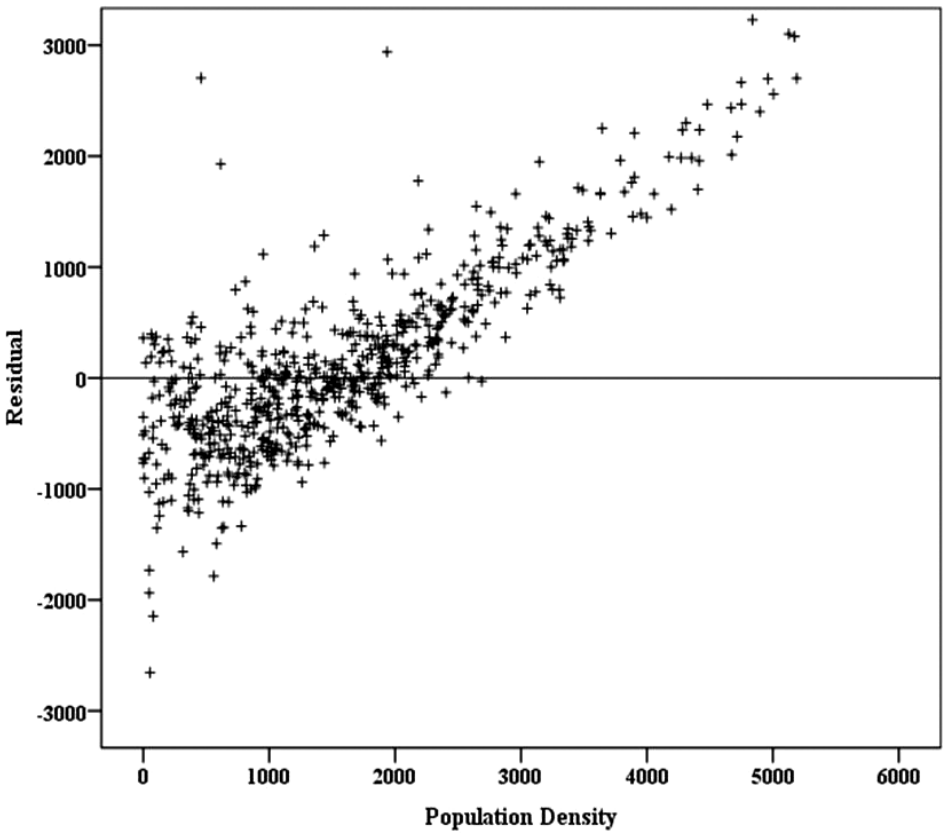

Figure 4. Residual distribution with population density (number of people per square kilometre) based on model 2.

Table 4. Correlation between population counts and residential area and temperature.

	Low-density residential	Medium-density residential	High-density residential	TEMP sum
Population	0.392*	0.758*	0.824*	0.905*

*Significant at the 0.01 level (two-tailed test).

Table 5. Models developed based on land use data.

Model	Explanatory variables	R^2	RE (%)
3	Low-, medium-, and high-density residential	0.98	21.4
4	High-density residential; land surface temperature	0.98	21.4

Note: For model 3, $P = 0.654445$(residential low) + 0.711886(residential medium) + 2.319887(residential high); for model 4, $P = 0.002334$(TEMP sum) + 1.524954(residential high), where TEMP sum is the sum of temperatures of residential pixels within a block group. For regression through the origin (the no-intercept model), R^2 measures the proportion of the variability in the dependent variable about the origin explained by regression. This cannot be compared with R^2 for models that include an intercept.

perature were used as explanatory variables in addition to traditional variables using spectral radiance. The results indicate that, when applied to the census block group, Landsat Enhanced Thematic Mapper Plus (ETM+) spectral bands alone cannot provide satisfactory estimation results, but the combination of land surface temperature and spectral bands can improve the estimation performance. The land use based method provided better estimation performance than the spectral-based method. The land use based prediction either by stratifying the density of residential land uses or by the combination of residential land use and land surface temperature can produce reasonable estimation. Dasymetric mapping provided a vivid representation of the spatial distribution of population.

Using remote sensing to estimate population is still a challenging task both in theory and methodology due to remote sensing data per se, the complexity of the urban landscape, and the complexity of population distribution. An important

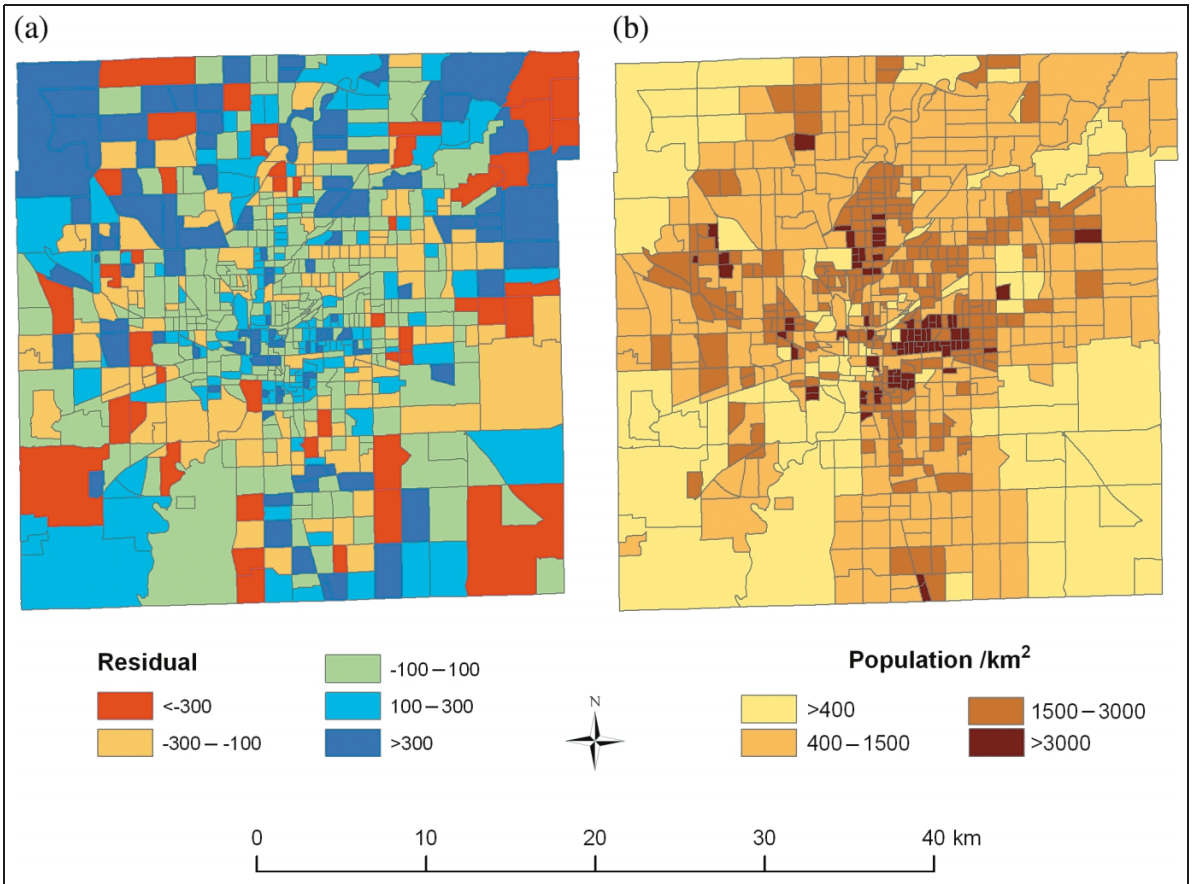

Figure 5. (a) Residuals resulting from population estimation based on model 3. (b) Population density from census.

concern is the selection of population indicators from remote sensing data. The most commonly applied indicators in previous studies are spectral radiance and their transforms. Since Iisaka and Hegedus (1982) first used spectral radiance based on Satellite pour l'Observation de la Terre (SPOT) images to estimate the population of Tokyo, there have been several attempts to explore the spectral radiance and its transforms as population predictors in zonal statistical models (Lo, 1995; Harvey, 2000; 2002a; Li and Weng, 2005; Wu and Murray, 2005). However, spectral radiance is not directly related to population. For example, the areas with low population density could be located in industrial or commercial areas or in areas dominated by forest or agriculture, but the spectral characteristics of these landscapes are significantly different. In addition, the aggregation of spectral radiance to the zonal unit level on which the population information is collected altered the inherent relationships between population and spectral radiance due to the effect of the modifiable areal unit problem (MAUP). Studies by Lo (1995) and Harvey (2002a) and this study showed that the direct use of mean values of spectral radiance could produce significant errors in population and population density. Using the transforms of spectral radiance in population estimation models (Harvey, 2002a; Wu and Murray, 2005) significantly increased the correlations between population density and remote sensing variables. However, the selection of transforms seemed to be arbitrary, and the variables used in the regression model may be highly correlated with one another, thus multicollinearity may exist in multiple regression. Incorporation of the impervious surface fraction, vegetation surface fraction, and land surface temperature into population estimation models improved the model performance, which may also strengthen the rationale for indicator selection and eliminated collinear variables from the model. It appears that the residential impervious surface image fraction is an excellent population indicator because of its stability and underlying relationship with population (Wu and Murray, 2005; Lu et al., 2006). However, the development of high-quality impervious surface data and the differentiation of nonresidential (e.g., commercial, industrial, transportation) from residential impervious surfaces warrant further studies. Moreover, under certain circumstances, the same amounts of impervious surface may have significantly different population densities because of different patterns of residential use. Thus, LULC data, especially those with high categorical resolution, have irreplaceable advantages for population estimation. In this research, different classes of residential use (high, medium, and low) were distinguished with the aid of ancillary data (i.e., housing density and zoning data), and regression models were then

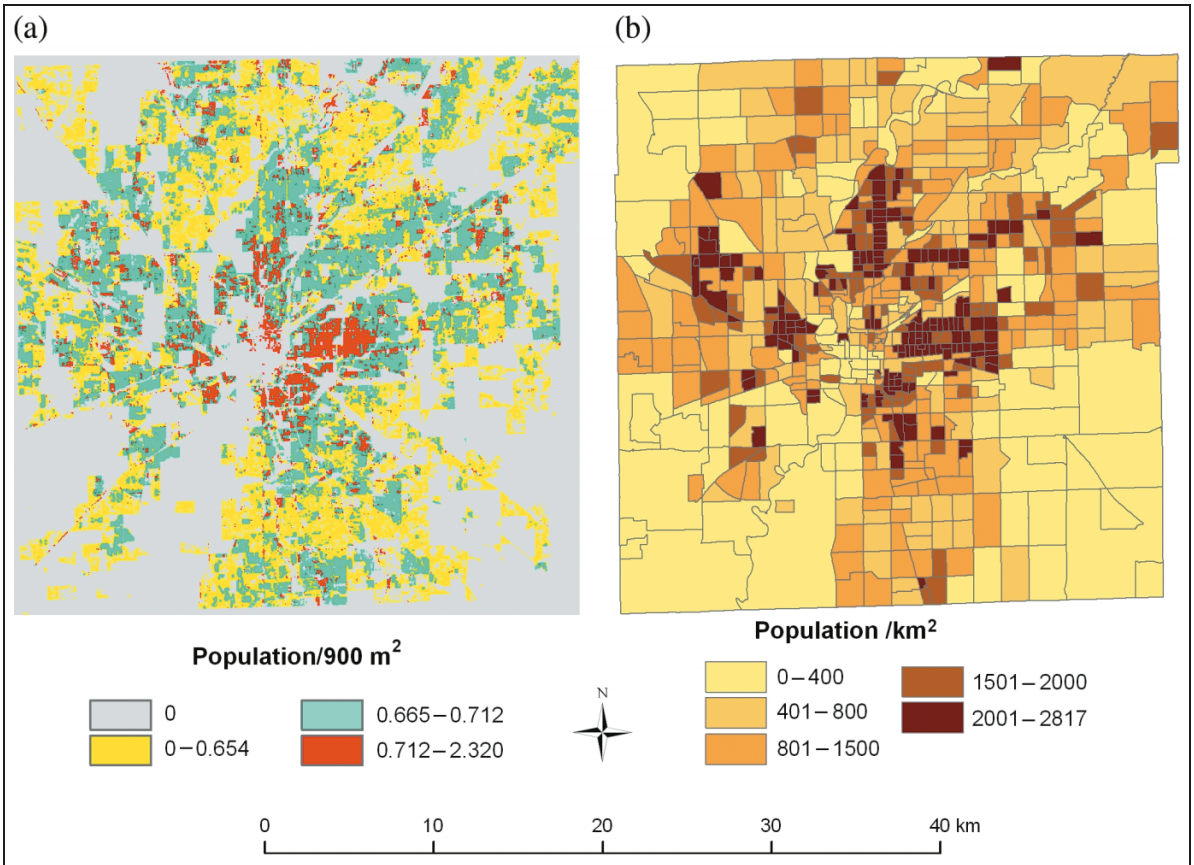

Figure 6. (a) Dasymetric distribution of population based on residential land uses. (b) Estimated population density based on model 3.

developed using these residential classes as explanatory variables. This procedure has been demonstrated to be capable of providing a good estimation result. This finding is in agreement with the studies of Lo (1995) and Wu and Murray (2005). Both studies applied high and low residential land uses in the development of population estimation models. Another advantage of using LULC data is that they are more stable predictors compared with spectral responses and can be obtained from different remote sensing data sources. Therefore, the land use based method exemplified in this study may be easily transferred to other datasets or other study areas for population estimation. However, the definitions of different LULC types, especially different residential land uses, are somewhat subjective. Moreover, extraction of accurate urban LULC information from moderate spatial resolution satellite images presents another challenge. The areas with high population density are mainly located in urban regions with multistory buildings, with which optical remote sensing data may not be capable of identifying vertical information such as the height and structure of multistory buildings. In addition, low-density residential areas scattered in forest and agricultural areas can have very similar spectral responses, making them difficult to separate.

One major uncertainty with population estimation comes from low-density residential areas with overestimation and from high-density residential areas with underestimation. This study has shown that the use of residential impervious surface, different densities of residential LULC classes, and land surface temperature can partially solve this problem. The incorporation of building height information seems to be another useful approach. Since light detection and ranging (lidar) data can provide building height information, incorporation of lidar data with impervious surface or with residential classes may provide new insights for population estimation, especially for high-density residential areas.

Acknowledgments

This research was conducted with support from the Center for Urban and Environmental Change and the University Research Committee (grant UNR10-18), Indiana State University. We appreciate the constructive comments and suggestions provided by anonymous reviewers which help to improve this manuscript.

References

Adams, J.B., Sabol, D.E., Kapos, V., Filho, R.A., Roberts, D.A., Smith, M.O, and Gillespie, A.R. 1995. Classification of multispectral images based on fractions of endmembers: Application to land cover change in

the Brazilian Amazon. *Remote Sensing of Environment*, Vol. 52, pp. 137–154. doi:10.1016/0034-4257(94)00098-8.

Briggs, D., Gulliver, J., Fecht, D., and Vienneau, D. 2007. Dasymetric modeling of small-area population distribution using land cover and light emission data. *Remote Sensing of Environment*, Vol. 108, pp. 451–466. doi:10.1016/j.rse.2006.11.020.

Eicher, C.L., and Brewer, C.A. 2001. Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science*, Vol. 28, pp. 125–138. doi:10.1559/152304001782173727.

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

Elvidge, C.D., Baugh, K.E., Hobson, V.R., Kihn, E.A., Kroehl, H.W., Davis, E.R., and Cocero, D. 1997. Satellite inventory of human settlements using nocturnal radiation emissions: A contribution for the global toolchest. *Global Change Biology*, Vol. 3, pp. 387–395. doi:10.1046/j.1365-2486.1997.00115.x.

Fabos, J.G., and Petrasovits, I. 1984. *Computer-aided land use planning and management*. Massachusetts Agricultural Experiment Station, Amherst, Mass. Research Bulletin 693, pp. 25–45.

Goodchild, M.F., and Lam, N.S. 1980. Areal interpolation: A variant of the traditional spatial problem. *Geoprocessing*, Vol. 1, pp. 297–312.

Goodchild, M.F., Anselin, L., and Deichmann, U. 1993. A framework for the areal interpolation of socioeconomic data. *Environment and Planning A*, Vol. 25, pp. 383–397. doi:10.1068/a250383.

Haack, B.N., Guptill, S.C., Holz, R.K., Jampoler, S.M., Jensen, J.R., and Welch, R.A. 1997. Urban analysis and planning. In *Manual of photographic interpretation*. 2nd ed. Edited by W. Philipson. American Society for Photogrammetry and Remote Sensing, Bethesda, Md. pp. 517–554.

Harvey, J. 2000. Small area population estimation using satellite imagery. *Statistics in Transition*, Vol. 4, pp. 611–633.

Harvey, J. 2002a. Estimation census district population from satellite imagery: Some approaches and limitations. *International Journal of Remote Sensing*, Vol. 23, pp. 2071–2095. doi:10.1080/01431160110075901.

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

Hsu, S.Y. 1971. Population estimation. *Photogrammetric Engineering*, Vol. 37, pp. 449–454.

Iisaka, J., and Hegedus, E. 1982. Population estimation from Landsat imagery. *Remote Sensing of Environment*, Vol. 12, pp. 259–272. doi:10.1016/0034-4257(82)90039-6.

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

Langford, M., Maguire, D.J., and Unwin, D.J. 1991. The areal interpolation problem: Estimating population using remote sensing in a GIS framework. In *Handling geographical information: methodology and potential applications.* Edited by L. Masser and M. Blakemore. Longman Scientific and Technical and John Wiley and Sons Inc., New York. pp. 55–77.

Li, G., and Weng, Q. 2005. Using Landsat ETM+ imagery to measure population density in Indianapolis, Indiana, USA. *Photogrammetric Engineering & Remote Sensing*, Vol. 71, pp. 947–958.

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

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*, Vol. 52, pp. 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*, Vol. 16, pp. 17–34. doi:10.1080/01431169508954369.

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

Lo, C.P. 2008. Population estimation using geographically weighted regression. *GIScience and Remote Sensing*, Vol. 45, pp. 131–148. doi:10.2747/1548-1603.45.2.131.

Lo, C.P., and Welch, R. 1977. Chinese urban population estimation. *Annals of the Association of American Geographers*, Vol. 67, pp. 246–253. doi:10.1111/j.1467-8306.1977.tb01137.x.

Lu, D., and Weng, Q. 2006. Use of impervious surface in urban land use classification. *Remote Sensing of Environment*, Vol. 102, pp. 146–160. doi:10.1016/j.rse.2006.02.010.

Lu, D., Weng, Q., and Li, G. 2006. Residential population estimation using a remote sensing derived impervious surface approach. *International Journal of Remote Sensing*, Vol. 27, pp. 3553–3570. doi:10.1080/01431160600617202.

Maantay, J.A., Maroko, A., and Porter-Morgan, H. 2008. A new method for population mapping and understanding the spatial dynamics of disease in urban areas. *Urban Geography*, Vol. 29, pp. 724–738. doi:10.2747/0272-3638.29.7.724.

Meliá, J., and Sobrino, J.A. 1987. Study on the utilization of SIR-A data for population estimation in the eastern part of Spain. *Geocarto International*, Vol. 2, pp. 33–37. doi:10.1080/10106048709354118.

Mennis, J. 2003. Generating surface models of population using dasymetric mapping. *Professional Geographer*, Vol. 55, pp. 31–42.

Nghiem, S.V., Balk, D., Rodriguez, E., Neumann, G., Sorichetta, A., Small, C., and Elvidge, C.D. 2009. Observations of urban and suburban environments with global satellite scatterometer data. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 64, pp. 367–380. doi:10.1016/j.isprsjprs.2009.01.004.

Roberts, D.A., Batista, G.T., Pereira, J.L.G., Waller, E.K., and Nelson, B.W. 1998. Change identification using multitemporal spectral mixture analysis: applications in eastern Amazônia. In *Remote sensing change detection: environmental monitoring methods and applications.* Edited by R.S. Lunetta and C.D. Elvidge. Ann Arbor Press, Ann Arbor, Mich. pp. 137–161.

Sutton, P. 1997. Modeling population density with nighttime satellite imagery and GIS. *Computers, Environment, and Urban System*, Vol. 21, pp. 227–244. doi:10.1016/S0198-9715(97)01005-34.

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

Sutton, P., Roberts, D., Elvidge, C.D., and Baugh, K. 2001. Census from heaven: An estimate of the global human population using night-time

satellite imagery. *International Journal of Remote Sensing*, Vol. 22, pp. 3061–3076. doi:10.1080/01431160010007015.

UNFPA. 2008. *The state of world population 2008*. United Nations Population Fund (UNFPA), New York.

Viel, J.F., and Tran, A. 2009. Estimating denominators: satellite-based population estimates at a fine spatial resolution in a European urban area. *Epidemiology*, Vol. 20, pp. 214–222. doi:10.1097/EDE.0b013e31819670dc.

Weng, Q., and Hu, X. 2008. Medium spatial resolution satellite imagery for estimating and mapping urban impervious surfaces using LSMA and ANN. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 46, pp. 2397–2406. doi:10.1109/TGRS.2008.917601.

Weng, Q., and Lu, D. 2009. Landscape as a continuum: an examination of the urban landscape structures and dynamics of Indianapolis city, 1991–2000. *International Journal of Remote Sensing*, Vol. 30, pp. 2547–2577. doi:10.1080/01431160802552777.

Weng, Q., Lu, D., and Schubring, J. 2004. Estimation of land surface temperature – vegetation abundance relationship for urban heat island studies. *Remote Sensing of Environment*, Vol. 89, pp. 467–483. doi:10.1016/j.rse.2003.11.005.

Weng, Q., Lu, D., and Liang, B. 2006. Urban surface biophysical descriptors and land surface temperature variations. *Photogrammetric Engineering & Remote Sensing*, Vol. 72, pp. 1275–1286.

Weng, Q., Hu, X., and Lu, D. 2008. Extracting impervious surface from medium spatial resolution multispectral and hyperspectral imagery: A comparison. *International Journal of Remote Sensing*, Vol. 29, pp. 3209–3232. doi:10.1080/01431160701469024.

Wu, C., and Murray, A. 2005. A cokriging method for estimating population density in urban areas. *Computers, Environment and Urban Systems*, Vol. 29, pp. 558–579. doi:10.1016/j.compenvurbsys.2005.01.006.

Wu, C., and Murray, A. 2007. Population estimation using Landsat Enhanced Thematic Mapper imagery. *Geographical Analysis*, Vol. 39, pp. 26–43. doi:10.1111/j.1538-4632.2006.00694.x.

Wu, S., Qiu, X., and Wang, L. 2005. Population estimation methods in GIS and remote sensing: a review. *GIScience and Remote Sensing*, Vol. 42, pp. 58–74.

Yuan, Y., Smith, R.M., and Limp, W.F. 1997. Remodeling census population with spatial information from Landsat imagery. *Computers, Environment, and Urban Systems*, Vol. 21, pp. 245–258. doi:10.1016/S0198-9715(97)01003-X.