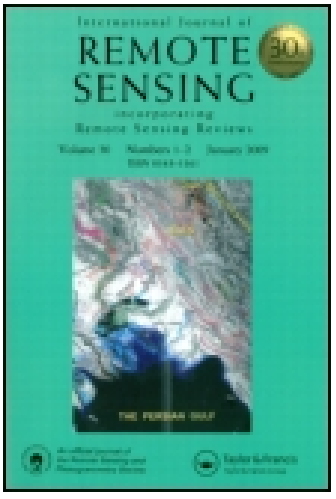

International Journal of Remote Sensing

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

Residential population estimation using a remote sensing derived impervious surface approach

Dengsheng Lu [a] , Qihao Weng [b] & Guiying Li [b]

[a] Center for the Study of Institutions , Population , and Environmental Change , Indiana University , Bloomington, IN 47408, USA

[b] Department of Geography , Geology , and Anthropology , Indiana State University , Terre Haute, IN 47809, USA

Published online: 22 Feb 2007.

To cite this article: Dengsheng Lu , Qihao Weng & Guiying Li (2006) Residential population estimation using a remote sensing derived impervious surface approach, International Journal of Remote Sensing, 27:16, 3553-3570, DOI: 10.1080/01431160600617202

To link to this article: http://dx.doi.org/10.1080/01431160600617202

International Journal of Remote Sensing
Vol. 27, No. 16, 20 August 2006, 3553–3570

Taylor & Francis
Taylor & Francis Group

Residential population estimation using a remote sensing derived impervious surface approach

DENGSHENG LU*†, QIHAO WENG‡ and GUIYING LI‡

†Center for the Study of Institutions, Population, and Environmental Change, Indiana University, Bloomington, IN 47408, USA

‡Department of Geography, Geology, and Anthropology, Indiana State University, Terre Haute, IN 47809, USA

(*Received 1 October 2004; in final form 22 January 2006*)

Residential population estimation was explored based on impervious surface coverage in Marion County, Indiana, USA. The impervious surface was developed by spectral unmixing of a Landsat Enhanced Thematic Mapper (ETM+) multispectral image. The residential impervious surface was then derived by geographic information system (GIS) overlay of residential land class and impervious surface. Regression analysis was conducted to develop population density estimation models. We found that the residential impervious surface-based approach provided the best population density estimation result, with mean and median relative errors of 38% and 23%, respectively. An overall population estimation error of −0.97% was achieved.

1. Introduction

Increasing population has generated great pressure on the sustainability of natural resources and the environmental conditions. Population increase is often associated with urban sprawl, resulting in a decrease of agricultural land and forested areas and producing problems such as increased pressure on food security, loss of biodiversity, and deterioration of environmental conditions. Timely and accurate estimation of population distribution is of considerable significance for decision makers in urban land-use planning and for a better understanding of the interactions between population growth and social, economic and environmental conditions. The traditional approach to population estimation is mainly based on a census, which is labour-intensive, time-consuming and costly, and also encounters difficulties in updating the database. Since the 1970s, remote sensing estimation of residential population has been applied more frequently, as an increasing amount of space-borne satellite data have become available (Lo 1986*a*, 2001, 2003, Langford *et al.* 1991, Sutton *et al.* 1997, 2001, Harvey 2002*a*,*b*, 2003, Li and Weng 2005). Lo (1986*b*) summarizes four categories of population estimation approaches using remotely sensed data. These methods are based on (1) counts of individual dwelling units using high spatial resolution data such as aerial photographs (Lo 1995), (2) measurements of urbanized land areas (Sutton *et al.* 1997, Lo 2001), (3) estimates derived from land use classification (Langford *et al.* 1991, Lo 2003), and (4) automatic image analysis based on spectral features (Harvey 2002*b*, 2003). These

*Corresponding author. Email: dlu@indiana.edu

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160600617202

approaches have been reviewed in previous literature (Harvey 2002*a*, Li and Weng 2005).

In previous research, population estimation has been conducted with different sensors of data that have various spatial resolutions; for example, high spatial resolution aerial photographs (Lo and Welch 1977, Lo 1986*b*), medium spatial resolution Landsat Thematic Mapper (TM) images (Harvey 2002*b*, 2003, Lo 2003, Li and Weng 2005), and low spatial resolution data from the Defense Meteorological Satellite Program Operational Linescan System (DMSP OLS) (Welch and Zupko 1980, Sutton *et al*. 1997, 2001). Very high spatial resolution data, such as aerial photographs and IKONOS, often prove difficult with regard to data processing and analysis because of their huge data content and geometric distortions in a large area, while low spatial resolution data, such as DMSP OLS, cannot provide sufficiently detailed information for population estimation at the regional and local levels. Hence, in practice, medium spatial resolution data, such as Landsat TM/Enhanced TM (ETM+) imagery, have become the primary source of data for population estimation in recent years (Lo 1995, Harvey 2002*b*, 2003, Lo 2003, Li and Weng 2005).

Previous population research using medium spatial resolution data frequently used multispectral radiance or reflectance (Harvey 2002*b*, 2003, Li and Weng 2005). The spectral signature-based method for population estimation has a common problem, in that estimations are affected by the external conditions of satellite imaging, such as the complexity of urban landscape, atmospheric conditions, phonological conditions, and sun elevation angles on land cover reflectance. Because population is not directly related to land cover surface reflectance captured by remote sensors, population estimation is still a challenging task based purely on remote sensing spectral signatures. It is important to derive reliable and stable variables from remotely sensed data for population estimation. One approach is to estimate population using land-use and land-cover (LULC) data, which has been explored in previous research (Langford *et al*. 1991, Lo 1995, 2003). Another possible approach is to use impervious surface coverage data (Ridd 1995, Wu and Murray 2003, Wu 2004), which has not been examined in previous literature. The latter is the focus of this study.

Impervious surfaces refer to any surfaces that water cannot infiltrate, and are primarily associated with transportation and buildings (Bauer *et al*. 2004). Impervious surfaces are relatively stable, and their proportions in an administrative unit (such as census units) depend on the nature of land use and the pattern of land use distribution. For residential areas, a large amount of impervious surface suggests more buildings and/or roads, and is often associated with a large population. The information on impervious surfaces can be developed using spectral mixture analysis (SMA) of remotely sensed multispectral data (Wu and Murray 2003, Lu and Weng 2004, Wu 2004). This study aimed to estimate the residential population in Marion County, Indiana, USA using a remote sensing derived impervious surface method, and to compare its effectiveness with that of the residential class-based approach.

2. Study area

The study area selected for this research is located between 39°36′ and 39°56′ N and 85°56′ and 86°19′ W, Marion County, Indiana (figure 1). Indianapolis, the capital of Indiana, is centred in Marion County. According to the US Census Bureau, the county has a total area of 1044 km^2, including 1026 km^2 of land and 18 km^2 of water.

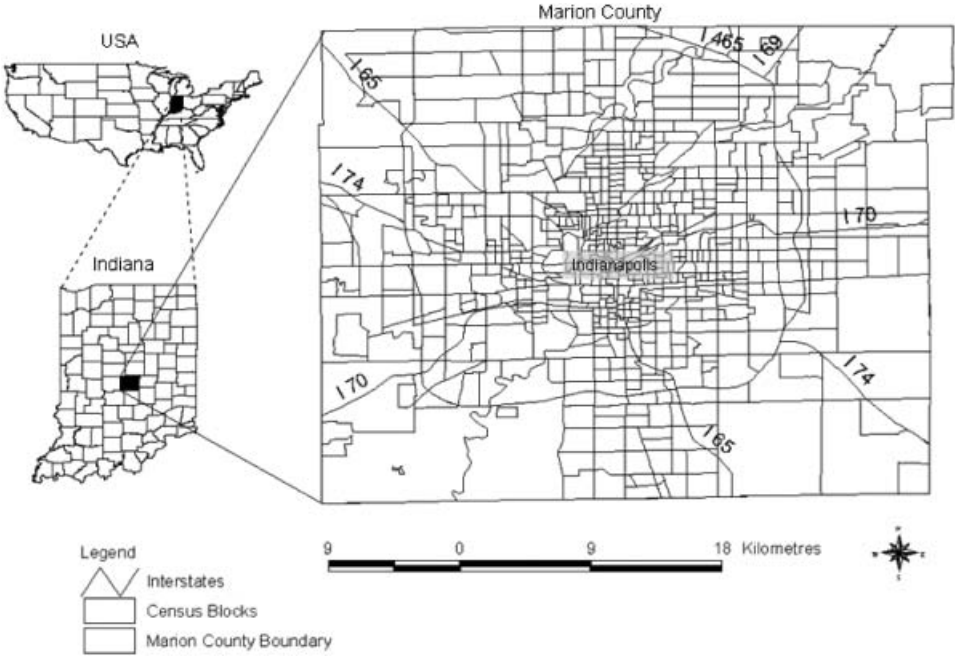

Figure 1. The study area – Marion County, Indiana, USA.

Indianapolis was called 'plain city' because of its flat topography (elevation ranges from 218 to 276 m above mean sea level). Its flatness and relatively symmetrical allocation provide the possibilities of 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. Thus, a timely estimation of the population distribution is valuable for urban land use planning.

3. Datasets

A Landsat 7 ETM + image, acquired on 22 June 2000, was used in this study. This image was rectified to a common Universal Transverse Mercator (UTM) coordinate system based on 1 : 24 000 scale topographic maps. A nearest-neighbour resampling algorithm was used during image rectification and a root mean square error (RMSE) of less than 0.5 pixels was obtained. An image-based dark object subtraction method was used to convert the digital number (DN) to surface reflectance (Lu *et al.* 2002). The ETM + data have eight bands, covering six reflective bands with 30 m spatial resolution, one panchromatic band with 15 m and one thermal band with 60 m spatial resolution. In this study, the six ETM + reflective bands were used. A land use/cover classification image was used directly, which was classified using a combination of maximum likelihood and decision tree classifiers on the fraction images from the same ETM + image (Lu and Weng 2004). Because residential land is the only relevant class in this study, the LULC map was recorded as a binary image to develop an image of residential areas.

The 2000 census data in a shapefile format were used in this study. The population data were organized as tract, block group, and block in the census

database. There are 212 tracts, 658 block groups, and 13 989 blocks in Marion County. Because of differences in coordinate systems between the census data and the ETM+ image, the geographic coordinates of the census data were converted into UTM to be consistent with the coordinate system of the ETM+ image.

4. Methods

4.1 *Spectral mixture analysis (SMA) of the Landsat ETM+ image*

SMA is regarded as a physically based image processing tool that assumes that the spectrum measured by a sensor is a linear combination of the spectra of all components (endmembers) within the pixel (Smith *et al.* 1990, Adams *et al.* 1995). The mathematical model of SMA can be expressed as:

$$R_{il} = \sum_{k=1}^{n} f_{kl} R_{ik} + \varepsilon_{il} \tag{1}$$

where $i=1, \ldots, m$ (number of spectral bands); $k=1, \ldots, n$ (number of endmembers); $l=1, \ldots, p$ (number of pixels); R_{il} is the spectral reflectance of band i of a pixel, which contains one or more endmembers; f_{kl} is the proportion of endmembers k within the pixel; R_{ik} is known as the spectral reflectance of endmember k within the pixel on band i, and ε_{il} is the error for band i in pixel l. For a constrained unmixing solution, f_{kl} is subject to the following restrictions:

$$\sum_{k=1}^{n} f_{kl} = 1 \text{ and } 0 \leq f_{kl} \leq 1 \tag{2}$$

The RMSE is often used to assess the fit of the model, and is computed based on errors and number of spectral bands used:

$$\text{RMSE} = \sqrt{\left(\sum_{i=1}^{p} \sum_{i=1}^{m} \hat{\varepsilon}_{il}^2 \right) \Big/ mp} \tag{3}$$

where $\hat{\varepsilon}_{il}$ is the estimated error or residual for band i in pixel l. With the SMA approach, selecting high-quality endmembers is the key to successfully developing high-quality fraction images. Many methods for endmember selection have been developed (Smith *et al.* 1990, Quarmby *et al.* 1992, Settle and Drake 1993, Bateson and Curtiss 1996, Tompkins *et al.* 1997, Garcia-Haro *et al.* 1999, Van der Meer 1999, Maselli 2001, Theseira *et al.* 2003). Image-based endmember selection approaches are often used, because image endmembers can be easily obtained and they represent the spectra measured on the same scale as the image data. Image endmembers can be derived from the extremes of the image feature space, assuming that they represent the purest pixels in the images (Mustard and Sunshine 1999). In addition to the endmember selection, another important step in the SMA approach is to select a suitable solution to unmix the spectral image. The least-squares solution is the most commonly used method in solving a linear mixture model (Smith *et al.* 1990, Shimabukuro and Smith 1991, Garcia-Haro *et al.* 1996) because of its simplicity and ease of implementation. The SMA has been recognized as an effective approach for many applications, such as improving urban LULC

classification (Rashed *et al.* 2001, Lu and Weng 2004) and mapping impervious surfaces (Wu and Murray 2003, Wu 2004).

In this study, principal component analysis (PCA) was used to transform the geometrically and atmospherically corrected ETM+ multispectral image into principal components (PCs). The scatterplots of PC1 *vs.* PC2 and PC2 *vs.* PC3 (figure 2(*a*) and 2(*b*)) were constructed to select four endmembers: high albedo, low albedo, soil, and vegetation. The characteristics of spectral reflectance of the selected endmembers are illustrated in figure 3. After selection of the image endmembers, a constrained least-squares solution was used to unmix the six ETM+ reflective bands into fraction images (figure 4). Visual checking of the fraction images and comparison with the ETM+ colour composite ensure that the derived fraction images are satisfactory.

4.2 *Development of the impervious surface image*

Previous research has indicated that the impervious surface can be developed by adding the high- and low-albedo fractions (Wu and Murray 2003). In general, the high-albedo fraction image mainly represents impervious surface information in the urban region and some dry soils in agricultural areas because of their similar reflectance. The low-albedo fraction image is more complex than other fraction images because it contains different features, such as water, building shadows in the central business district, vegetation canopy shadows in forested areas, and dark impervious surface materials (Lu and Weng 2004). It is important to remove the non-impervious surface from the high- and low-albedo fraction images before they are used to derive impervious surface images. The dense vegetation, bare soils in agricultural areas and water areas do not have impervious surfaces and they have highest fraction values in vegetation, soil and low-albedo fraction images, respectively. Therefore, the pixels of these classes can be first masked out from the fraction images with expert rules developed from the sample plots, which were identified from very high spatial resolution aerial photographs (Lu and Weng 2004). An impervious surface image is then produced by adding low- and high-albedo endmembers. The impervious surface image has values ranging from 0 and 1 (corresponding 0–100% coverage of impervious surface). Figure 5(*a*) shows that commercial/industrial/transportation (CIT) areas have the highest impervious surface values, appearing white in the image. Residential areas appear bright grey to grey, depending on the density. A quick comparison between the population distribution and the impervious surface image indicates that CIT areas have high impervious surface values but a very low population density. The analysis of training sample plots for the CIT and agricultural areas indicates that the majority of CIT areas (such as the downtown and airport) have an impervious surface value greater than 75% and the majority of rural areas have less than 25%. The impervious surface value in the majority of residential areas is in the range 25–75%, in agreement with the study by Wu and Marray (2003). An adjusted impervious surface image is thus developed by masking out the pixels with impervious surface values greater than 75% or less than 25% (figure 5(*b*)), assuming that the population is located in the residential areas. Another approach used to extract the residential impervious surface is by geographic information system (GIS) overlay of the residential classification image and impervious surface image (figure 5(*c*)).

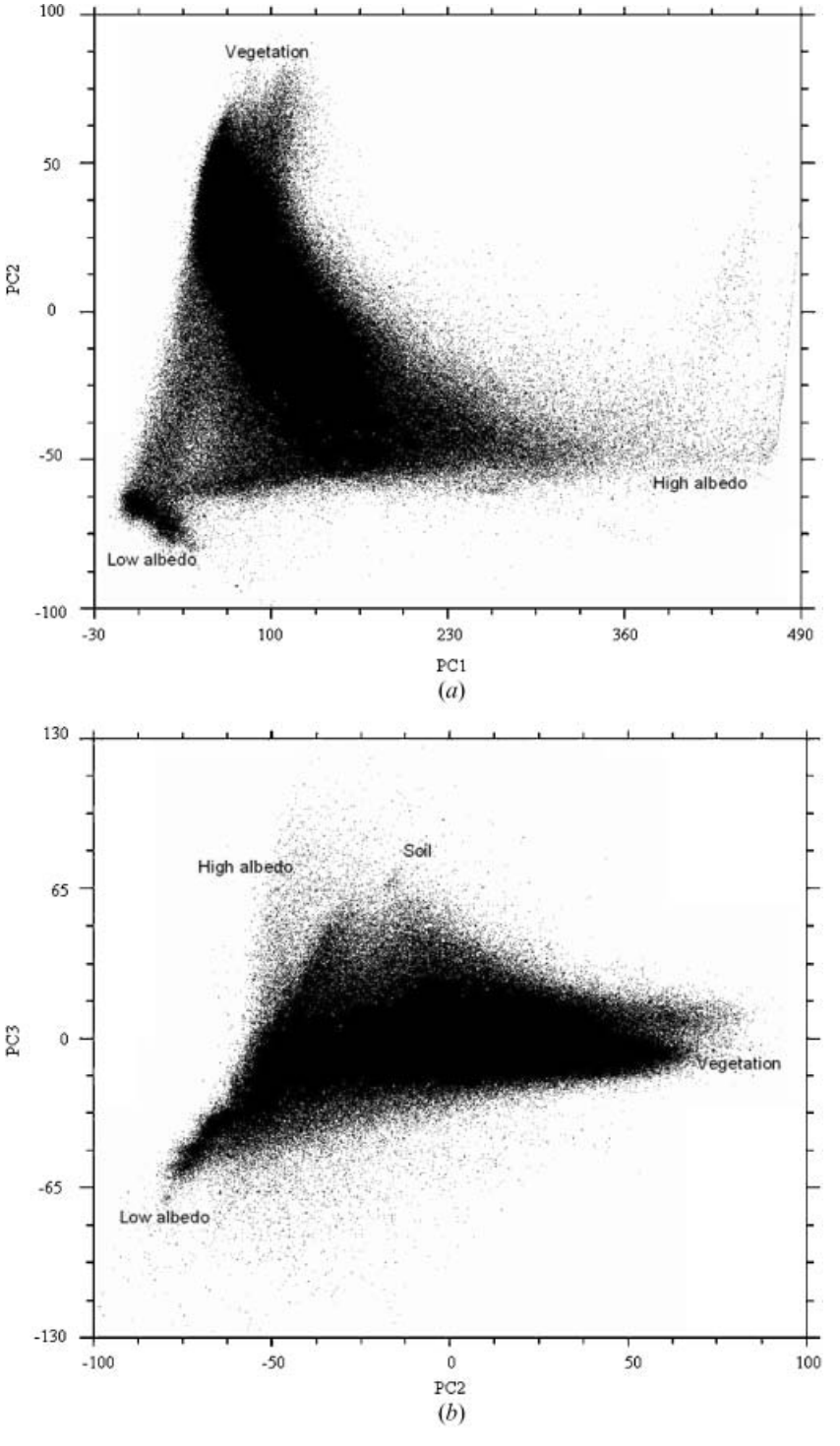

Figure 2. Scatterplots of two principal components illustrating the potential endmembers: (*a*) PC1 *vs*. PC2; (*b*) PC2 *vs*. PC3.

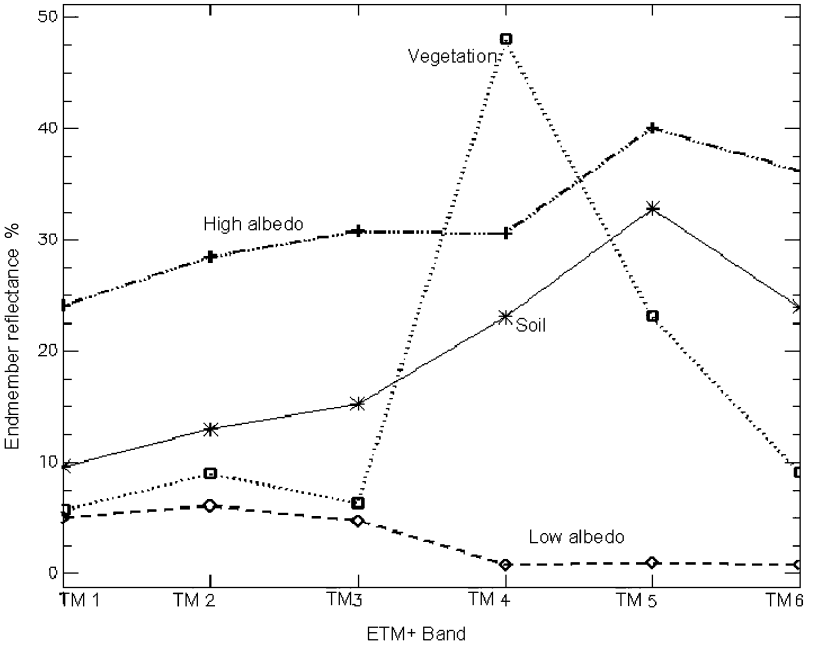

Figure 3. Reflectance characteristics of the four selected endmembers.

4.3 *Development of population estimation models*

The block group (BG) was used as the analytical unit. There are 658 BGs in this study area. Two sample datasets were developed using a random sampling technique. Half of the sample data (329 BGs) were used as a modelling data set for the development of the population density estimation models, and the other half (329 BGs) were used as a validation data set for accuracy assessment. The population density was calculated for each BG (persons/km^2). The percentage of residential class for each BG was also calculated by the division of the number of pixels of the residential class by the total number of pixels in a BG. The mean values of original, adjusted and residential impervious surface for each BG were further computed based on the overlay of the derived impervious surface images and the population density map. Figure 6 illustrates these analytical procedures for extraction of statistical variables. Linear and nonlinear regression analyses were tested to identify suitable models for population density estimation. Population density and its transformed variables using square roots and natural logarithms were selected as dependent variables, and the percentage of residential area (Resid_P), adjusted impervious surface (adjIMP), and residential impervious surface (Resid_IMP) were used as independent variables. A casewise diagnosis with 2.5 standard deviations and scattergrams of population density and relevant impervious surface variables were used to identify the outliers in the sample data. Twelve observations (BGs) were then removed from the sample of 329 BGs during the regression analysis. Thus, the population density estimation models were developed based on the 317 BGs. A coefficient of determination (R^2) was calculated to evaluate the effectiveness of the regression models for the population density estimation.

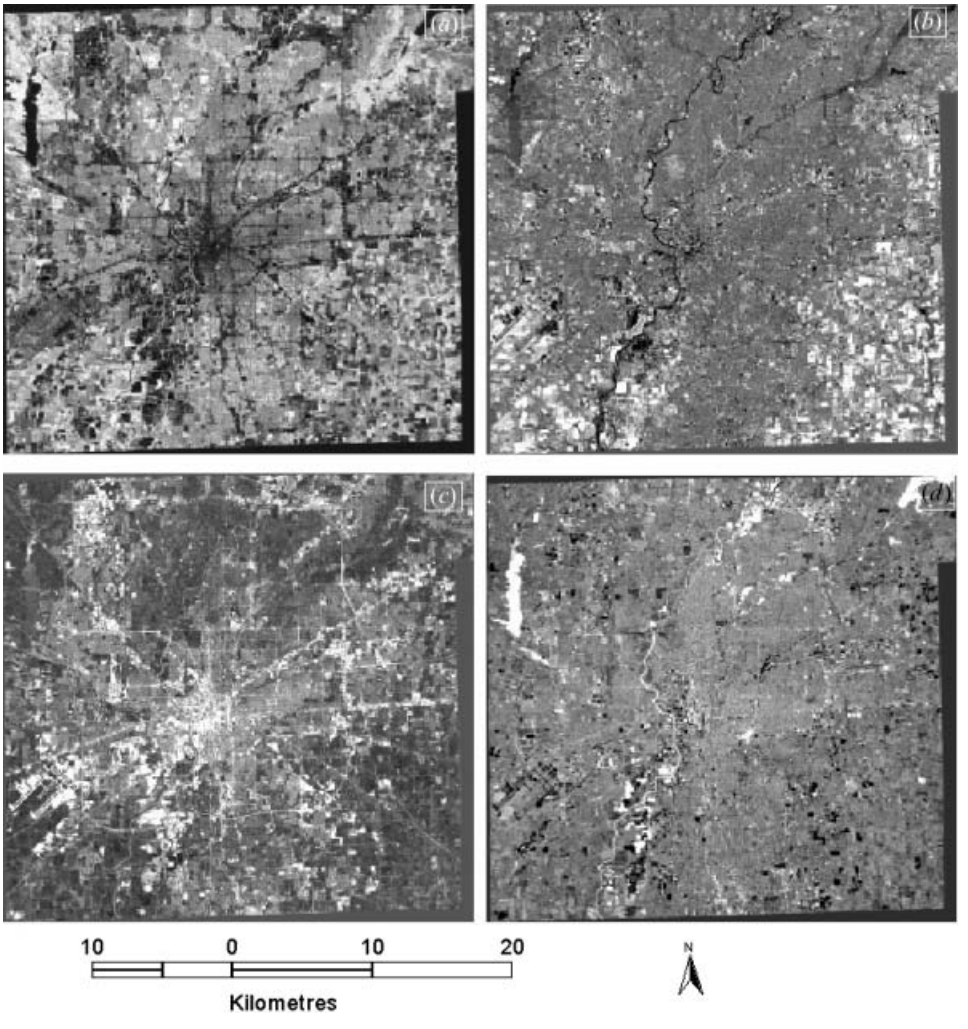

Figure 4. Four fraction images developed using spectral mixture analysis of Landsat ETM+ data: (*a*) vegetation; (*b*) soil; (*c*) high albedo; (*d*) low albedo.

4.4 *Accuracy assessment*

Accuracy assessment is an important part of the process of developing population density estimation models. In addition to R^2, which is often used to evaluate the performance of a model based on the modelling data set, a relative error was used to assess the model performance based on the validation data set. For an individual case, the relative error is defined as:

$$\mathrm{RE} = \left(P_e - P_g\right)/P_g \times 100 \tag{4}$$

where P_e and P_g are the estimated and reference values, respectively. The residual $(P_e - P_g)$ for individual cases may be negative or positive, so the absolute value of the residuals was further calculated by using the following formula:

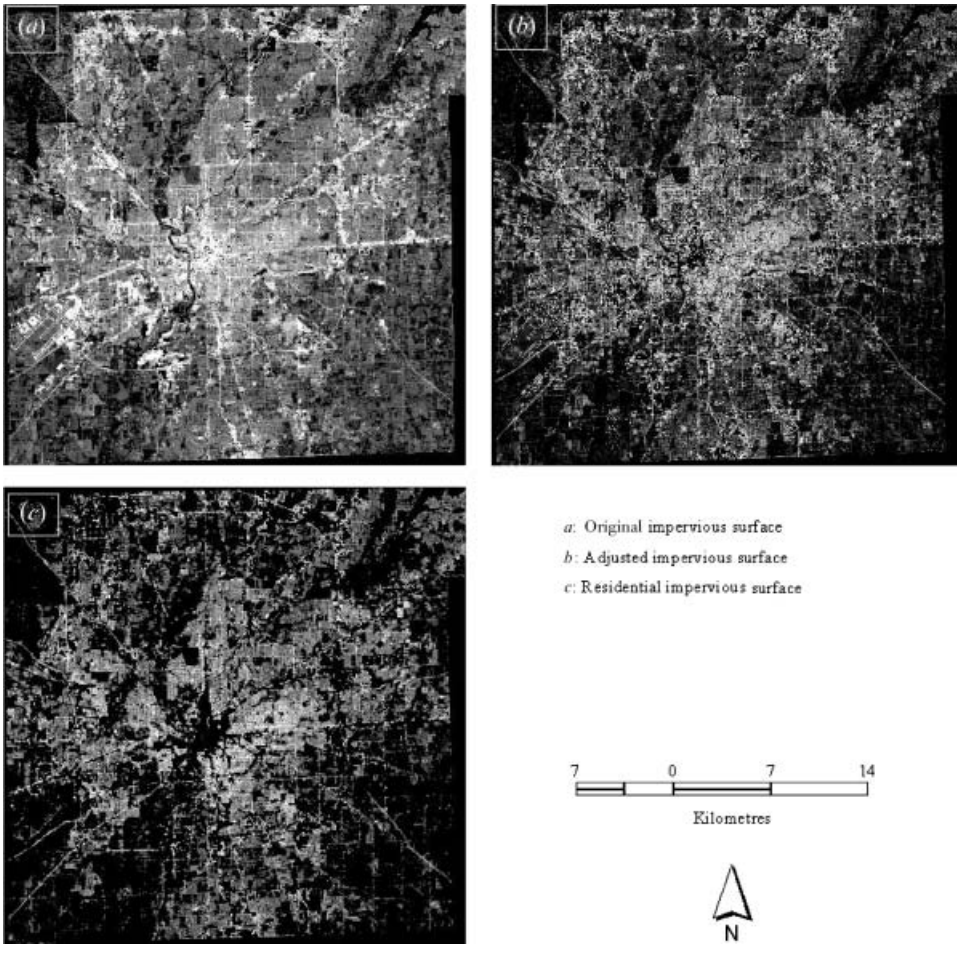

Figure 5. Impervious surface images developed from high- and low-albedo fractions: (*a*) impervious surface; (*b*) adjusted impervious surface which those pixel values greater than 75% or less than 25% were masked out; (*c*) residential impervious surface which those pixel of non-residential area were masked out.

$$\text{Mean relative error (MRE)} = \frac{\sum_{k=1}^{n} |RE_k|}{n} \tag{5}$$

MRE is often influenced by extreme values, and cannot always effectively assess the model performance. To avoid this problem, a median relative error (MdRE) was also used in this research. The MRE and MdRE were calculated for each regression model. The distribution of residuals for each model was also examined to understand the error distribution in relation to population density.

5. Results

A scatterplot can reveal the relationship between two data variables. Figure 7(*a*) shows the scatterplot of population density and percentage of the residential area. Overall, a good linear relationship exists. However, when the residential area is less

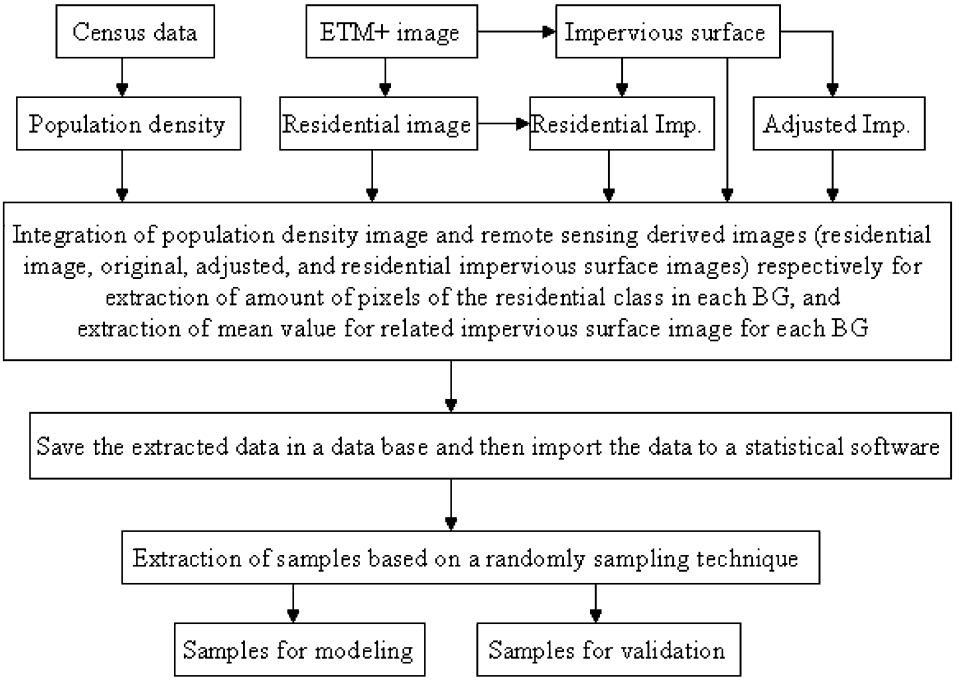

Figure 6. Framework for extraction of remote sensing variables based on integration of population density and remote sensing-derived variables.

than 25% or greater than 85% in a BG, the linear relationship does not continue to hold. This implies that population estimation with a very low or very high density would be difficult or might produce large estimation errors if the models were not optimal. Figures 7(*b*) and 7(*c*) illustrate the relationships between impervious surface and population density, and indicate that samples with the highest impervious surface fractions tend to have medium to low population density (Figure 7(*b*)). This is because the CIT areas have large proportions of impervious surface but a very low population density. However, adjusted impervious surface values show a good linear relationship with population density (Figure 7(*c*)). Residential impervious surfaces shows an even better linear relationship with population density (Figure 7(*d*)).

This study indicates that the regression model using the residential impervious surface approach appears to be the best model, with an R^2 value of 0.82 and an overall population estimation error of −0.97%. The significantly large MRE values compared with corresponding MdRE values suggest that extreme values of population density, especially the large relative errors from the low population density, have a significant impact on the models (table 1). The best models for population density estimation based on the R^2 values are summarized in table 1, where the square root of the population density was used as the dependent variable and the percentage of the residential area, adjusted impervious surface, and residential impervious surface were used as independent variables. The R^2 values range from 0.77 to 0.82, based on the modelling data set. Median relative errors range from 23.1% to 26.3%, based on the validation data set, which represent the population density estimation errors.

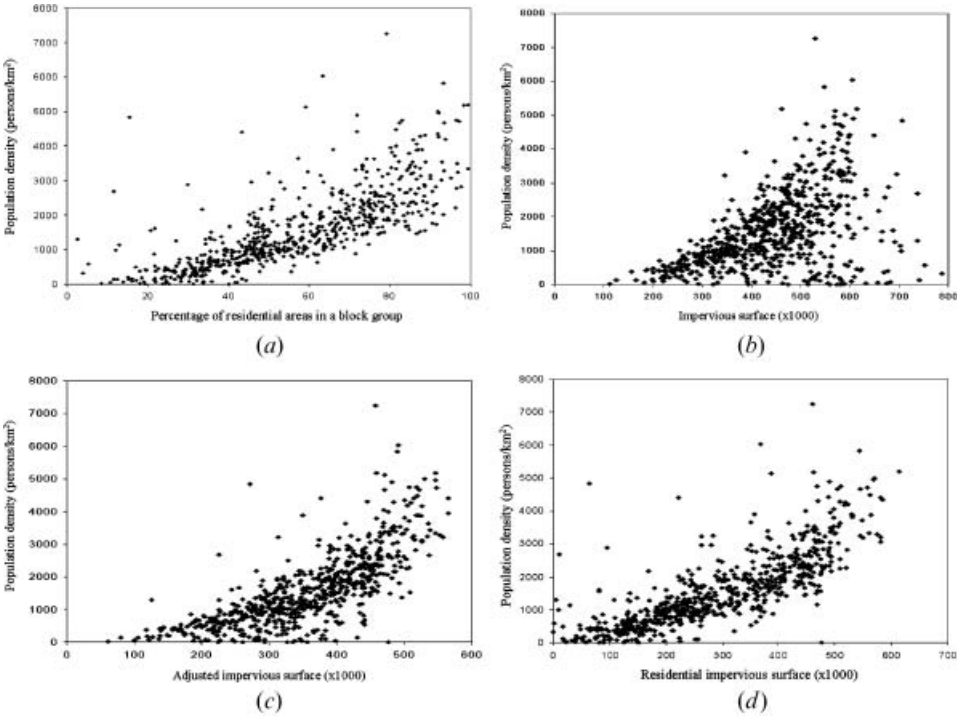

Figure 7. Scatterplots between population density and: (*a*) percentage of residential areas in a block group (BG); (*b*) impervious surface; (*c*) adjusted impervious surface; and (*d*) residential impervious surface. For the data analysis, the impervious surface images are rescaled to 0–1000 by multiplying by 1000 for each pixel, so the ranges of impervious surface values are between 0 and 1000.

Analysing error distributions is another way to understand the performance of population estimation models. Figure 8 illustrates the distribution of relative errors for the three models. All the models seem to have a common problem. Lower population density areas appear to associate with higher relative errors, and tend to

Table 1. A summary of the best models for population density estimation.

Method	Independent variable	Regression model	R^2	MRE	MdRE	Error
Resid	Resid_P	Sqrt_PD=4.865+0.557 × Resid_P	0.776	39.16	26.25	−1.97
Imp_surf	adjIMP	Sqrt_PD=−3.201+0.115 × adjIMP	0.782	58.89	24.49	1.87
Resid_Imp	Resid_IMP	Sqrt_PD=10.674+0.088 × Resid_IMP	0.821	38.04	23.13	−0.97

Resid_P, percentage of residential area in a block group; adjIMP, mean value of the adjusted impervious surface in a block group; Resid_IMP, mean value of the residential impervious surface in a block group; Sqrt_PD, square root of population density.
For the data analysis, the impervious surface images are rescaled to 0–1000 by multiplying by 1000 for each pixel, so the ranges of adjIMP and Resid_IMP are between 0 and 1000.
R^2 is the coefficient of determination, which is based on the modelling data set.
MRE and MdRE are mean and median relative errors (%), respectively, of the population density based on the validation data set. Error means total population estimation error (%) based on the overall data set in this study area.

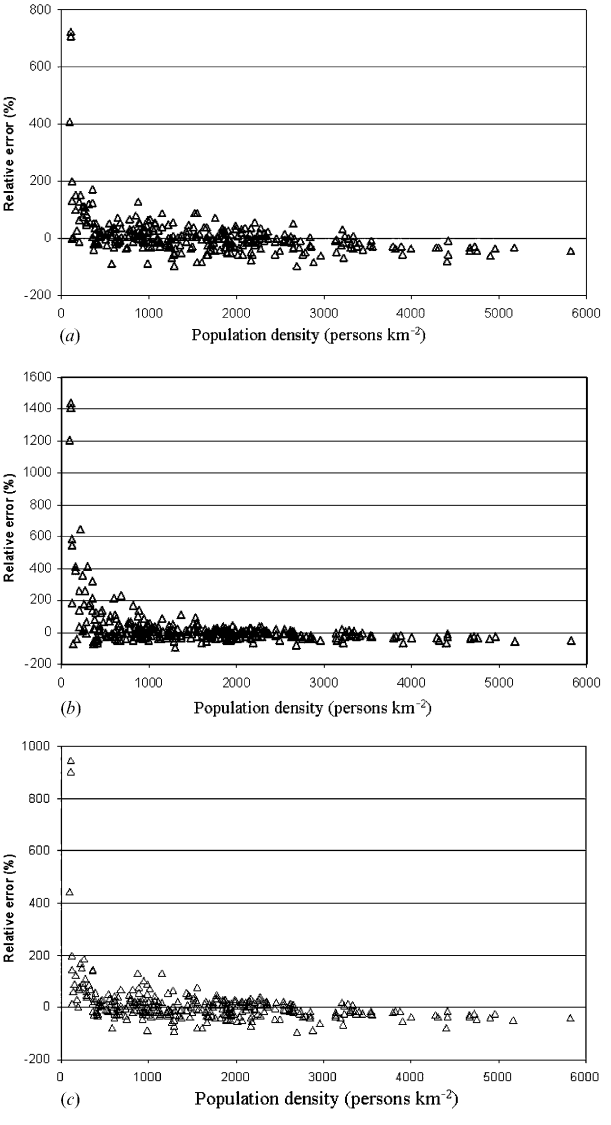

Figure 8. Comparison of population estimation errors based on different remote-sensing derived variables: (*a*) percentage of residential area; (*b*) adjusted impervious surface; (*c*) residential impervious surface.

be overestimated. Conversely, higher population density areas tend to be underestimated. Figure 8 indicates that when population density is less than approximately 300 (persons per km^2), population densities are often overestimated, no matter which model is assessed, whereas when population density is greater than approximately 3500 (persons per km^2), population densities are often underestimated. A comparison of figures 8(*a*), 8(*b*) and 8(*c*) indicates that the use of percentage of residential area in the population estimation can improve estimation results in low population density areas.

Figure 9 plots reference against estimated population density data. It indicates that use of percentage of residential area and residential impervious surface

improves population estimation results, especially for low-density areas. However, all three models have a common problem, in that the population density is underestimated in high-density areas.

6. Discussion and conclusions

Remote sensing-based population estimation is still a challenging task, especially for residential regions with extremely high or low population densities. The residential impervious surface provided the best population density estimation performance with a median relative error of 23% and a total population error of −0.97%. A major advantage of this approach is that the impervious surface is stable, and almost independent of seasonal change and atmospheric conditions. Another advantage for using the residential impervious surface is that the models developed may be transferable to other study areas.

The impervious surface is a complex feature that can be made up of different kinds of materials and can be of different types such as buildings, roads, parking lots and sidewalks. The impervious surface is spectrally easily confused with bare soils. Many factors may affect the success of extraction of impervious surfaces from satellite imagery. For example, the complexity and mixture of different kinds of impervious surfaces result in difficulty in identifying suitable endmembers. The limited spectral channels and the high correlations among certain channels limit the number of endmembers that can be used in the SMA. Moreover, the image-based endmember selection method cannot ensure that sufficient numbers of typical endmembers are obtained. This method assumes that typical endmembers are always contained in the dataset used, but this assumption is not always true, especially with complex urban environments. Hence, it is necessary to associate image endmembers with reference endmembers of actual target materials (Adams *et al.* 1995, Roberts *et al.* 1998*a*). Selection of reference endmembers from spectral libraries or from field measurements is flexible, but it is difficult to account for all possible features and processes due to the many factors influencing the data spectra. For example, image reflectance spectra may not be accurately correspondent with library reflectance spectra due to the effects of different atmospheric correction methods. Conversion of spectral library data to the units of the image (e.g. reflectance, radiance, or DN) or conversion of the image data to the units of spectral library data may also be necessary. It is important to make sure that a spectral library data approximates the image spectra in terms of wavelengths, bandwidths and band shapes (Roberts *et al.* 1998*a*). To solve the dilemma between limited image channels and the large number of endmembers required, a possible method is to use multiple endmember models (Roberts *et al.* 1998*b*). The multiple-endmember SMA approach permits a large number of endmembers to be modelled across a scene, and has shown better performance than the standard SMA approach (Painter *et al.* 1998, Roberts *et al.* 1998*b*, Okin *et al.* 2001, Dennison and Roberts 2003). This approach starts with a series of candidate two-endmember models and then evaluates each model based on fraction values, RMSE and residual threshold, and finally produces fraction images with the lowest error (Roberts *et al.* 1998*b*). The multiple-endmember SMA approach may be more suitable for use in urban environments.

The adjusted impervious surface has shown an improvement for population estimation over the original impervious surface. It is important to select suitable thresholds to remove those pixels with non-residential areas. The thresholds of 25%

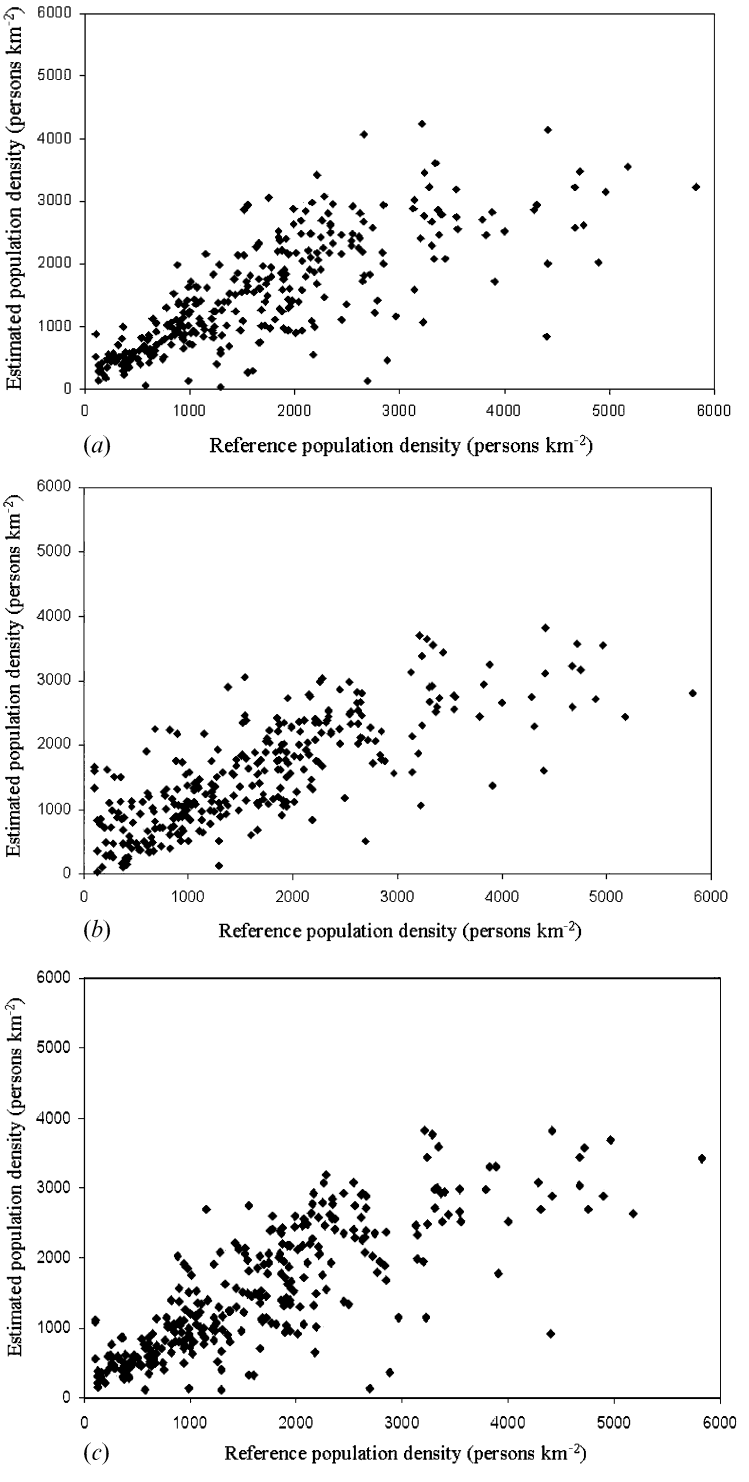

Figure 9. Comparison of population reference data and estimated population data based on different remote-sensing derived variables: (*a*) percentage of residential area; (*b*) adjusted impervious surface; (*c*) residential impervious surface.

and 75% used in this study are based on the limited training samples of the CIT areas and the agricultural areas. Some high-rise apartment buildings with very high population densities may be masked out, but some CIT areas with less than 75% of impervious surface values may remain in the impervious surface image. A similar problem occurred in the areas with less than 25% of impervious surface. This is an important factor influencing the population estimation performance. Integration of the impervious surface image and the residential class is recommended to generate a good residential impervious surface image, if an accurate classification map is available.

This research suggests that with remote sensing-based methods, population is often overestimated in residential areas with low population density, but underestimated in areas with high population density. Stratification of population density has been demonstrated to be effective in improving estimation results (Lo 2003, Li and Weng 2005). The pixel-based population estimation approach may provide a better result than those based on census measurements (Harvey 2003). Based on the findings of this study, we consider that a stratification of population densities, combined with the pixel-based estimation models using the residential impervious surface, may further improve population estimation performance. One important source of errors for population estimation is from high-rise apartment buildings. The optical sensor data such as TM/ETM+ can only provide land surface information and cannot provide height and intra-building information. The incorporation of building height information in population estimation models may improve model performance. Light Detection And Ranging (LiDAR) data, for example, have been shown to be capable of extracting building height information (Maas and Vosselman 1999, Barnsley *et al.* 2003), and may provide new insight for population estimation through integration of LiDAR-derived information and ETM+-derived variables. Some possible solutions have been proposed for improving population estimation accuracy in low- or high-population density areas (Harvey 2003, Lo 2003).

Acknowledgements

We acknowledge the financial support of the Indiana Space Grant Consortium to Q.W., and NASA funds through Purdue University (Grant no. NGTS-40114-4) for a project entitled 'Indiana Impervious Surface Mapping Initiative'. D.L. acknowledges support of the Center for the Study of Institutions, Population, and Environmental Change, Indiana University, through the National Science Foundation (Grant 99-06826). We also thank the two anonymous reviewers who provide many constructive suggestions in revising this paper.

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*, **52**, pp. 137–154.

BARNSLEY, M.J., STEEL, A.M. and BARR, S.L., 2003, Determining urban land use through an analysis of the spatial composition of buildings identified in LIDAR and multispectral image data. In *Remotely Sensed Cities*, V. Mesev (Ed.), pp. 83–108 (London: Taylor & Francis).

BATESON, A. and CURTISS, B., 1996, A method for manual endmember selection and spectral unmixing. *Remote Sensing of Environment*, **55**, pp. 229–243.

BAUER, M.E., HEINERT, N.J., DOYLE, J.K. and YUAN, F., 2004, Impervious surface mapping and change monitoring using Landsat remote sensing. In *ASPRS Annual Conference Proceedings*, 23–28 May 2004, Denver, CO (American Society for Photogrammetry and Remote Sensing: Bethesda, MD).

DENNISON, P.E. and ROBERTS, D.A., 2003, Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE. *Remote Sensing of Environment*, **87**, pp. 123–135.

GARCIA-HARO, F.J., GILABERT, M.A. and MELIA, J., 1996, Linear spectral mixture modeling to estimate vegetation amount from optical spectral data. *International Journal of Remote Sensing*, **17**, pp. 3373–3400.

GARCIA-HARO, F.J., GILABERT, M.A. and MELIA, J., 1999, Extraction of endmembers from spectral mixtures. *Remote Sensing of Environment*, **68**, pp. 237–253.

HARVEY, J.T., 2002a, Estimating census district populations from satellite imagery: some approaches and limitations. *International Journal of Remote Sensing*, **23**, pp. 2071–2095.

HARVEY, J.T., 2002b, Population estimation models based on individual TM pixels. *Photogrammetric Engineering and Remote Sensing*, **68**, pp. 1181–1192.

HARVEY, J.T., 2003, Population estimation at the pixel level: developing the expectation maximization technique. In *Remotely Sensed Cities*, V. Mesev (Ed.), pp. 181–205 (London: Taylor & Francis).

LANGFORD, M., MAGUIRE, D.J. and UNWIN, D.J., 1991, The areal interpolation problem: estimating population using remote sensing in a GIS framework. In *Handing Geographical Information: Methodology and Potential Applications*, L. Masser and M. Blakemore (Eds), pp. 55–77 (New York: Longman Scientific & Technical/John Wiley & Sons, Inc.).

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

LO, C.P., 1986a, Accuracy of population estimation from medium-scale aerial photography. *Photogrammetric Engineering and Remote Sensing*, **52**, pp. 1859–1869.

LO, C.P., 1986b, *Applied Remote Sensing* (New York: Longman).

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

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

LO, C.P., 2003, Zone-based estimation of population and housing units from satellite-generated land use/land cover maps. In *Remotely Sensed Cities*, V. Mesev (Ed.), pp. 157–180 (London: Taylor & Francis).

LO, C.P. and WELCH, R., 1977, Chinese urban population estimation. *Annals of the Association of American Geographers*, **67**, pp. 246–253.

LU, D., MAUSEL, P., BRONDÍZIO, E. and MORAN, E., 2002, Assessment of atmospheric correction methods for Landsat TM data applicable to Amazon basin LBA research. *International Journal of Remote Sensing*, **23**, pp. 2651–2671.

LU, D. and WENG, Q., 2004, Spectral mixture analysis of the urban landscapes in Indianapolis with Landsat ETM+ imagery. *Photogrammetric Engineering and Remote Sensing*, **70**, pp. 1053–1062.

MASELLI, F., 2001, Definition of spatially variable spectral endmembers by locally calibrated multivariate regression analysis. *Remote Sensing of Environment*, **75**, pp. 29–38.

MAAS, H.-G. and VOSSELMAN, G., 1999, Two algorithms for extracting building models from raw laser altimetry data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **54**, pp. 153–163.

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

OKIN, G.S., ROBERTS, D.A., MURRAY, B. and OKIN, W.J., 2001, Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments. *Remote Sensing of Environment*, **77**, pp. 212–225.

PAINTER, T.H., ROBERTS, D.A., GREEN, R.O. and DOZIER, J., 1998, The effects of grain size on spectral mixture analysis of snow-covered area from AVIRIS data. *Remote Sensing of Environment*, **65**, pp. 320–332.

QUARMBY, N.A., TOWNSHEND, J.R.G., SETTLE, J.J. and WHITE, K.H., 1992, Linear mixture modeling applied to AVHRR data for crop area estimation. *International Journal of Remote Sensing*, **13**, pp. 415–425.

RASHED, T., WEEKS, J.R., GADALLA, M.S. and HILL, A.G., 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**, pp. 5–15.

RIDD, M.K., 1995, Exploring a V–I–S (Vegetation–Impervious Surface–Soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities. *International Journal of Remote Sensing*, **16**, pp. 2165–2185.

ROBERTS, D.A., BATISTA, G.T., PEREIRA, J.L.G., WALLER, E.K. and NELSON, B.W., 1998a, Change identification using multitemporal spectral mixture analysis: applications in eastern Amazonia. In *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*, R.S. Lunetta and C.D. Elvidge (Eds), pp. 137–161 (Ann Arbor, MI: Ann Arbor Press).

ROBERTS, D.A., GARDNER, M., CHURCH, R., USTIN, S., SCHEER, G. and GREEN, R.O., 1998b, Mapping chaparral in the Santa Monica mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment*, **65**, pp. 267–279.

SETTLE, J.J. and DRAKE, N.A., 1993, Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing*, **14**, pp. 1159–1177.

SHIMABUKURO, Y.E. and SMITH, J.A., 1991, The least-squares mixing models to generate fraction images derived from remote sensing multispectral data. *IEEE Transactions on Geoscience and Remote Sensing*, **29**, pp. 16–20.

SMITH, M.O., USTIN, S.L., ADAMS, J.B. and GILLESPIE, A.R., 1990, Vegetation in Deserts: I. A regional measure of abundance from multispectral images. *Remote Sensing of Environment*, **31**, pp. 1–26.

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*, **22**, pp. 3061–3076.

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 and Remote Sensing*, **63**, pp. 1303–1313.

THESEIRA, M.A., THOMAS, G., TAYLOR, J.C., GEMMELL, F. and VARJO, J., 2003, Sensitivity of mixture modeling to endmember selection. *International Journal of Remote Sensing*, **24**, pp. 1559–1575.

TOMPKINS, S., MUSTARD, J.F., PIETERS, C.M. and FORSYTH, D.W., 1997, Optimization of endmembers for spectral mixture analysis. *Remote Sensing of Environment*, **59**, pp. 472–489.

VAN DER MEER, F., 1999, Iterative spectral unmixing (ISU). *International Journal of Remote Sensing*, **20**, pp. 3431–3436.

WELCH, R. and ZUPKO, S., 1980, Urbanized area energy utilization patterns from DMSP data. *Photogrammetric Engineering and Remote Sensing*, **46**, pp. 1107–1121.

WU, C., 2004, Normalized spectral mixture analysis for monitoring urban composition using ETM+ imagery. *Remote Sensing of Environment*, **93**, pp. 480–492.

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