International Journal of Remote Sensing
Vol. 31, No. 21, 10 November 2010, 5571–5586

Taylor & Francis
Taylor & Francis Group

Evaluation of small-area population estimation using LiDAR, Landsat TM and parcel data

PINLIANG DONG*, SATHYA RAMESH and ANJEEV NEPALI
Department of Geography, University of North Texas / 1155 Union Circle #305279, Denton, TX 76203, USA

This paper presents methods and results of small-area population estimation using a combined Light Detection And Ranging (LiDAR), Landsat Thematic Mapper (TM) and parcel dataset for a study area in Denton, Texas, USA. A normalized digital surface model (nDSM) was created from a digital surface model (DSM) and a digital elevation model (DEM) built from LiDAR point data. Residential and commercial parcels were selected from parcel data and used as a mask to remove non-residential and non-commercial pixels from the nDSM. Classification results of residential areas from Landsat TM images acquired on two dates were used to further refine the nDSM. Using continuous and random census blocks as samples, building count, building area and building volume were calculated from the nDSM through mathematical morphological operations, zonal statistics, data conversion and spatial joining in a geographic information system (GIS). Combined with census 2000 data, a total of 10 ordinary least squares (OLS) regression models and geographically weighted regression (GWR) models were built and applied to the census blocks in the study area. Finally, accuracy assessments were carried out. The results show that the sign and magnitude of the relative estimation errors at the census-block level lead to underestimation of the total population in the study area. Possible reasons for the relatively low accuracies and problems for further investigation are also discussed.

1. Introduction

To support land-use planning and better understanding the complex interactions between human and the environment, it is essential to estimate population distribution in a timely and efficient manner. Traditionally, population information is mainly obtained through a census, which is time consuming, labour intensive and expensive. Many national census agencies and international organizations use four methods to update census data and estimate population size (Smith and Mandell 1984): (1) component II (COMP-II) – using vital statistics such as birth and death data to measure the natural increase from the last census, (2) ratio-correlation (R-CORR) – using regression methods to relate changes in population to changes in indicators of population change, such as school enrolment, the number of voters, the number of passenger car registrations and the number of occupied housing units, (3) administrative record (AD-REC) – using births, deaths, school enrolment, social insurance, building permits, driver licences, voter registration and tax returns to estimate population size and (4) cohort component method – tracing people born in a given year through their lives.

*Corresponding author. Email: pdong@unt.edu

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online
http://www.tandf.co.uk/journals
DOI: 10.1080/01431161.2010.496804

From the above methods, it seems that ratio-correlation is the only method in which remote sensing can play a role for population estimation using image spectral and spatial information. Since the 1970s, remotely sensed data from various platforms have been used for population estimation, including low-spatial-resolution image data (Welch and Zupko 1980, Sutton *et al*. 1997, 2001), medium-spatial-resolution Landsat Thematic Mapper (TM) images (Harvey 2002a,b, 2003, Lo 2003, Qiu *et al*. 2003, Wu and Murray 2003, Wu 2004, Li and Weng 2005, Lu *et al*. 2006) and high-spatial-resolution aerial photographs (Lo and Welch 1977, Lo 1986a,b). In Sutton *et al*. (1997), nighttime satellite images were used for estimating population density for the continental US. The application of nighttime images was extended to the estimation of global human population (Sutton *et al*. 2001). Qiu *et al*. (2003) used Landsat TM images and Topologically Integrated Geographic Encoding and Referencing (TIGER) system road data for modelling urban population growth. Lu *et al*. (2006) derived impervious surface information from Landsat TM images for population estimation. In terms of algorithms, while ordinary least squares (OLS) regression is a popular method, some other methods such as kriging (Wu and Murray 2003) have also been used for population estimation. To deal with the spatial non-stationarity problem in population estimation, Lo (2008) investigated the application of geographically weighted regression (GWR) models for population estimation in Atlanta and produced more accurate results than using OLS models. In recent years, population estimation in emergencies has become a new application area of remote sensing and geographic information systems (GISs) (Henderson 2006, Stone 2008). Based on the reviews by Lo (1986b) and Wu *et al*. (2005), existing population methods using remote sensing and GISs can be put into two general categories: areal interpolation and statistical modelling (figure 1).

From existing literature, it seems that Landsat TM images are a popular choice for population estimation, mainly because of their relatively rich spectral information for mapping land-cover types and moderate spatial resolution to cover large study areas. For population estimation at census-block group or census-block level, Landsat TM has two major limitations: (1) the spatial information is not high enough and (2) the

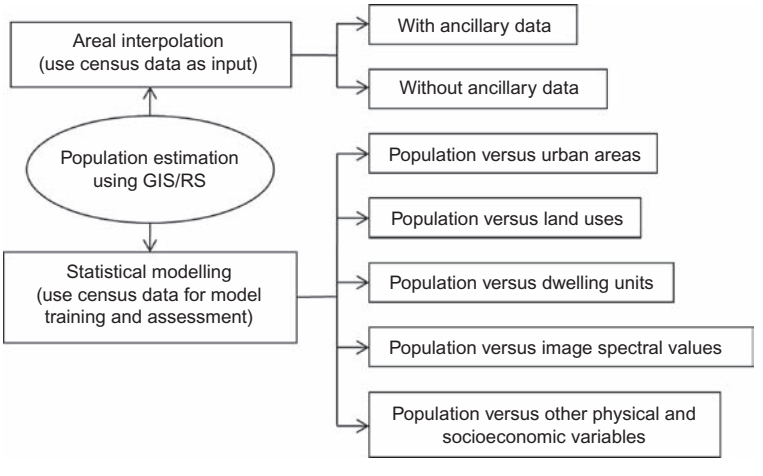


Figure 1. Methods of population estimation using a geographic information system (GIS) and remote sensing (RS).

lack of elevation information to separate spectrally similar objects. For example, gas-well pads in Texas may not be separated from buildings on a Landsat TM image. With the ability to provide elevation information, Light Detection And Ranging (LiDAR) data have been widely used in topographic mapping and forest-structure modelling, among other applications. The potential of building extraction from LiDAR data to support population estimation has been mentioned in several papers (e.g. Wu *et al.* 2005, Lu *et al.* 2006). In recent years, many papers have been published on high-resolution satellite image and LiDAR data analysis (Zhou *et al.* 2004, Bork and Su 2007, Brandtberg 2007, Koetz *et al.* 2007, Sohn and Dowman 2007, Antonarakis *et al.* 2008, Xie *et al.* 2008, Aubrecht *et al.* 2009, Chen *et al.* 2009). While most of the papers address the issue of information extraction using spatial information or elevation information, further investigations of the extracted information for real-world applications remain limited. Evaluation of LiDAR data for population estimation would be an important addition to traditional remote-sensing methods in this application field.

This paper presents an evaluation of LiDAR, Landsat TM and parcel data for small-area population estimation. A small area can be defined as a subcounty area such as census tracts, block groups and blocks, and the areas that can be aggregated from them. Numerous studies have been carried out for small-area population estimation using traditional demographic characteristics (Verma *et al.* 1984, Platek *et al.* 1987, Bracken 1991, Wolter and Causey 1991, Verma 1992, Cai 2007, Jarosz 2008). According to figure 1, this study belongs to the statistical modelling category with a focus on population versus dwelling units derived from LiDAR, Landsat TM and parcel data. The major objective is to test if a combination of LiDAR, Landsat TM and parcel data provide accurate population estimations for small areas. The basic ideas is to build OLS and GWR models based on sampling census blocks using (1) population versus building count, (2) population versus building area and (3) population versus building volume, and evaluate the models for population estimation with support from census 2000 data.

2. Study area and data

The study area is located in the eastern part of the City of Denton, Texas. The area has 8 census tracts, 30 census-block groups and 764 census blocks (figure 2).

Data for this study include (1) Landsat TM images acquired on 4 March 2000 and 10 July 2000, (2) LiDAR data acquired during leaf-on season (4 September 2001), (3) parcel data of Denton County and (4) census 2000 data. The purpose of using Landsat TM images of two dates in spring and summer is to obtain more accurate land-cover classification to support population estimation. LiDAR data were post-processed through clustering and filtering and output as ASCII text files for ground and off-ground points with a point spacing of 3–5 m (unfortunately, the original LiDAR point clouds were not available). Parcel data are maintained by the Central Appraisal District of Denton County and pre-2000 (including 2000) parcels are selected from the whole dataset. Finally, census 2000 data were used to build regressions models based on sampling census blocks and to evaluate the accuracy of the estimated population. Figure 3 shows the above datasets. Figure 4 shows continuous and random census blocks selected for developing regression models.

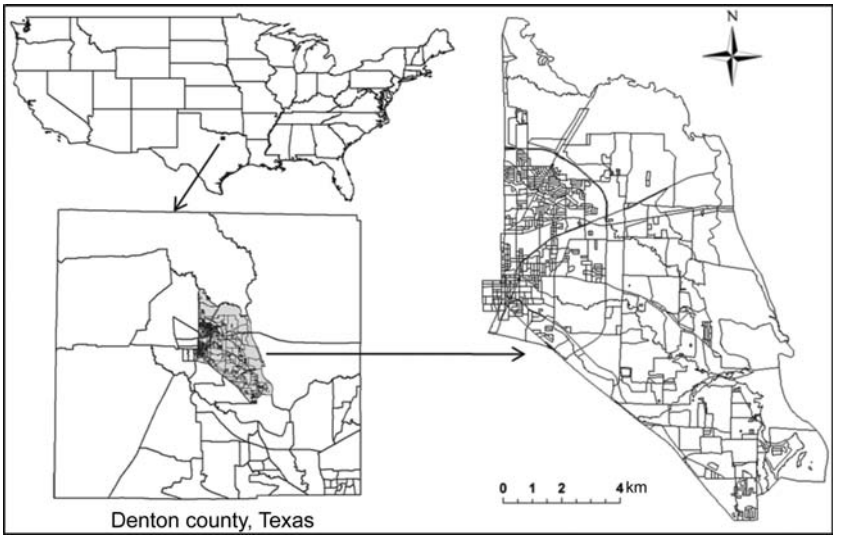


Figure 2. Location of the study area.

3. Methodology

The methodology flowchart is shown in figure 5, and some major steps are briefly described below.

(1) Landsat TM image classification. Landsat TM images of 12 bands (two dates, 30 m × 30 m resolution) were used as input for land-use/land-cover classification. Maximum likelihood classification was conducted using training samples for residential areas, commercial / industrial / transportation (CIT) areas, soils, vegetation and water. 256 random sample points were selected for classification-accuracy assessment. An overall accuracy of 80.2% was obtained.

(2) LiDAR data processing. Twenty-nine tiles of LiDAR text files for digital elevation models (DEMs) and digital surface models (DSMs) were imported into ArcGIS to create point-shape files and converted to rasters with 1 m × 1 m cell size by inverse distance weight (IDW) interpolation. The LiDAR dataset with 1 m × 1 m cell size were created for a separate study using LiDAR and 1 m resolution IKONOS images. Although the 1 m × 1 m cell size is finer than the LiDAR point spacing of 3–5 m in this study and does not contain more information than a 3 m × 3 m cell (MacEachren and Davidson 1987, Chow and Hodgson 2009), no efforts were made to resample the LiDAR dataset to a coarser cell size. A DEM mosaic and a DSM mosaic were created from the rasters. A normalized digital surface model (nDSM) (Aubrecht *et al.* 2009, Hill and Broughton 2009) was obtained by subtracting the DEM from the DSM. Theoretically, nDSM values are from zero to the maximum height of the surface features. Because the LiDAR points were post-processed to create a point spacing of 3–5 m, noise pixels were obvious on the nDSM. These noise pixels were removed by grey-scale closing operations in mathematical morphology, using structural elements of 3 × 3 pixels for dilation operation followed by erosion operation (Dong 1997). Based on the Manufactured Home Construction and Safety Standards published by the

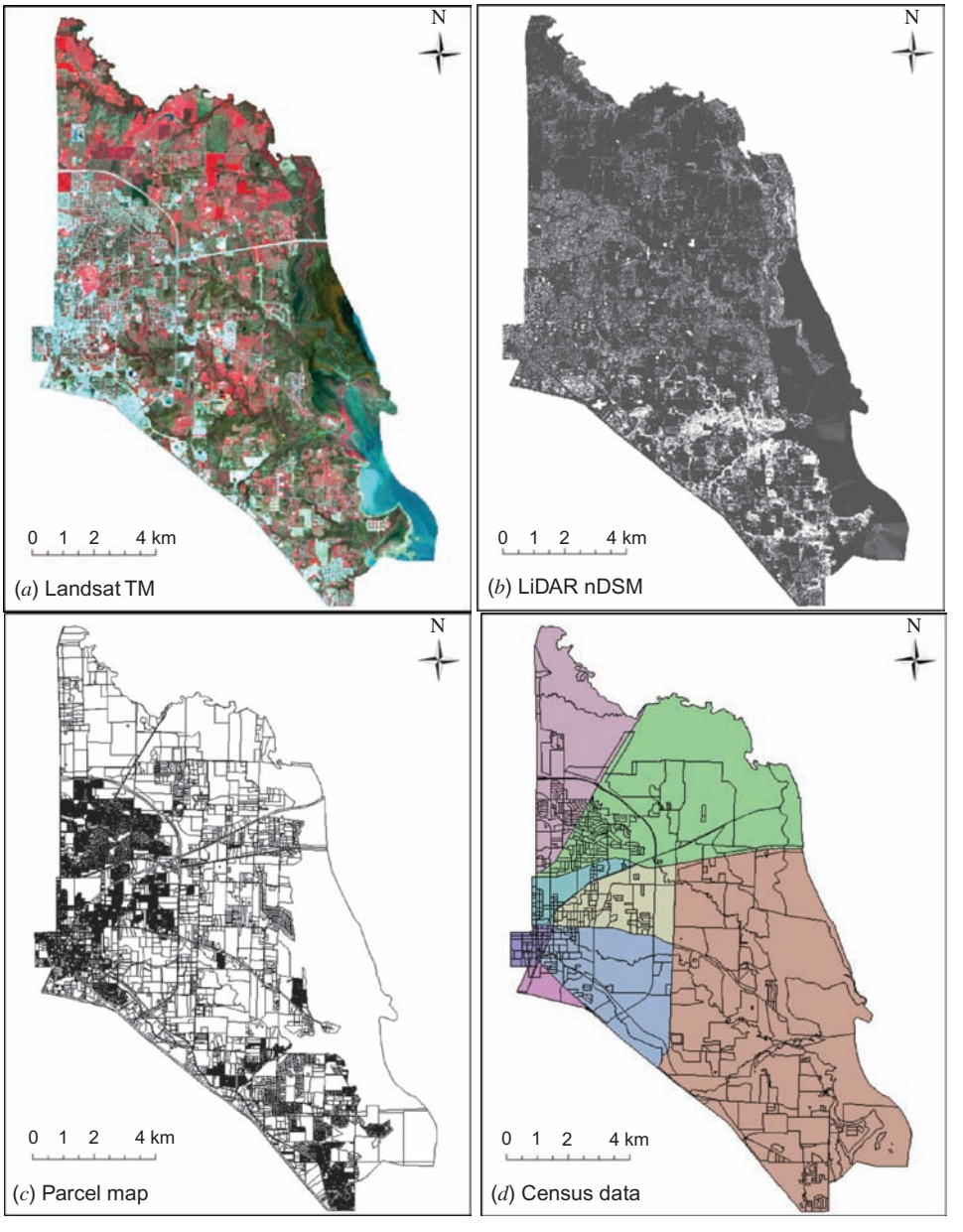


Figure 3. Datasets used in the study: (*a*) Landsat TM, (*b*) normalized digital surface model (nDSM) derived from LiDAR data, (*c*) parcel map and (*d*) census tracts are shown in different colours.

Texas Department of Housing & Community Affairs (TDHCA 2009), a threshold of 2.2 m was used to separate buildings and trees from other ground objects. Pixels lower than 2.2 m on the nDSM were set to zero.

(3) Creation of the final nDSM map. Using pre-2000 (including 2000) residential and commercial parcels as the first mask (parcel mask) and residential pixels

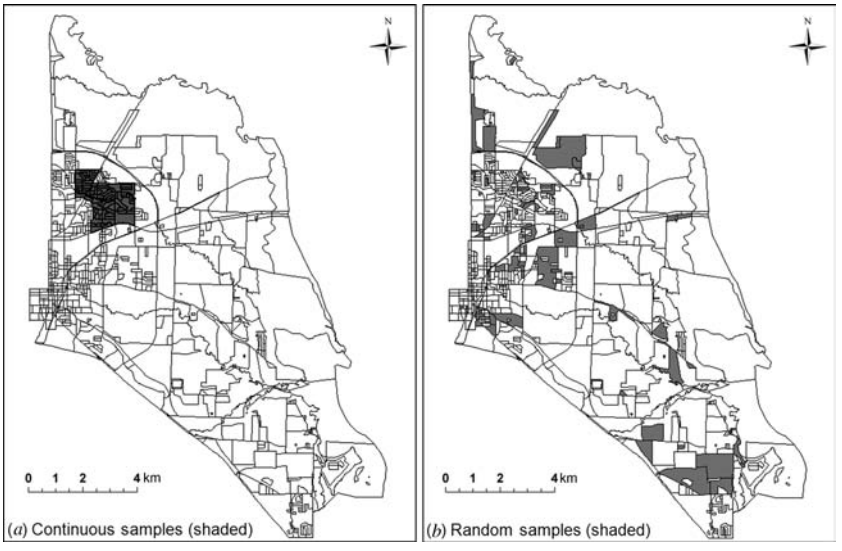


Figure 4. Sampling census blocks for regression analysis: (*a*) 91 continuous samples and (*b*) 91 random samples.

from Landsat TM image classification as the second mask (TM mask), the final nDSM was created by removing cells outside the masks. Since the LiDAR data were collected on 4 September 2001 (17 months after the census 2000), the use of the parcel mask and the TM mask can also help eliminate buildings that were built after 2000 when creating the final nDSM.

(4) Building area, volume and count derived from nDSM. Based on the final nDSM, a binary raster can be created in which non-zero values in the input nDSM will be output as 1, while other values in the input nDSM will be output as 0. Using census blocks as zones and the binary raster as input raster, a zonal sum can be calculated for each census block using zonal statistics. This zonal sum is the building area of that census block because the cell size is 1 m $\times$ 1 m. Similarly, using census blocks as zones and the final nDSM as the input raster, building volume can be obtained by calculating zonal sum in each census block. Building count is obtained by converting the buildings in the nDSM to polygons (polygons with an area less than 10 m^2 were removed) and joining building centroids to census-block shapefile through spatial joining. After that, the number of buildings in each block is readily available.

(5) Development of regression models. A total of 10 regression models were created based on census 2000 population data and independent variables for building count, building area and building volume: (i) three OLS regression models derived from building count, building area and building volume of 91 continuous census blocks (see figure 4(*a*)), (ii) three OLS regression models derived from building count, building area and building volume of 91 random census blocks (see figure 4(*b*)), (iii) one multiple linear-regression model developed using building count and building area as independent variables from the 91 random census blocks (it was found that building area is highly correlated with building volume, therefore building volume was not used as an independent

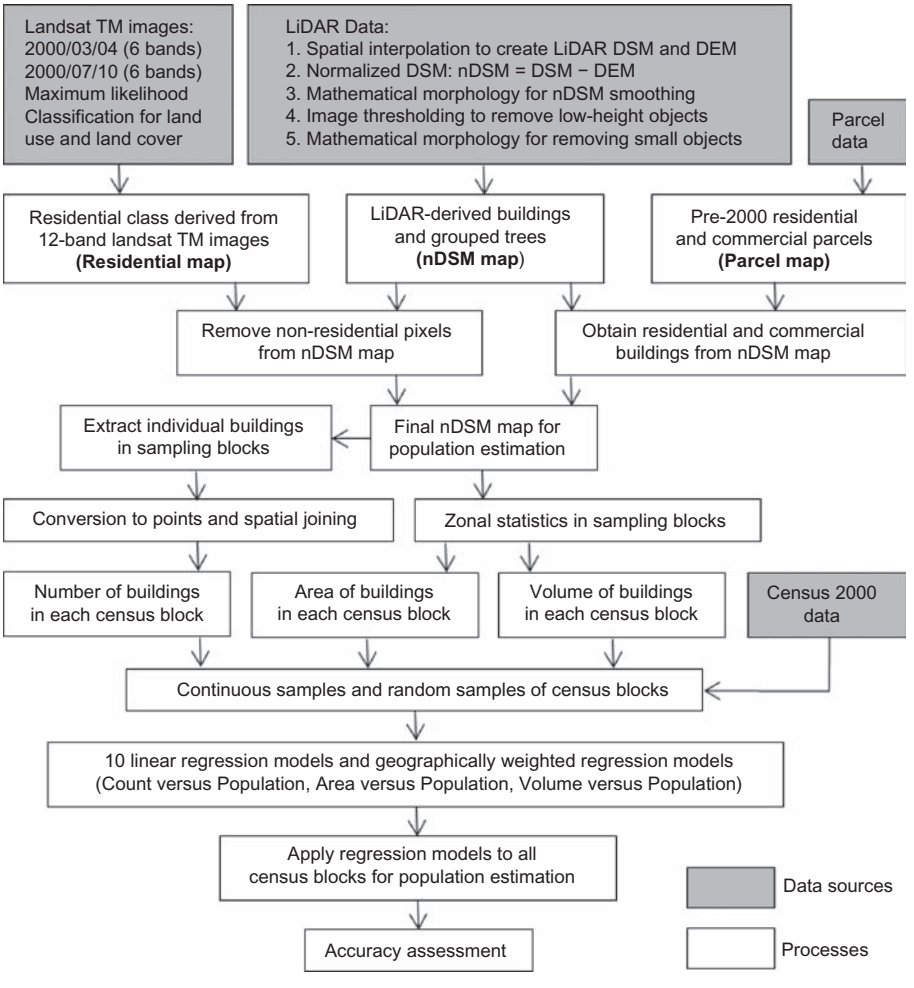


Figure 5. Flowchart of population estimation using GIS and remotely sensed data.

variable for multiple linear regression) and (iv) three GWR models derived from building count, building area and building volume of the 91 random census blocks.

(6) The basic idea of GWR for modelling heterogeneous processes is that parameters may be estimated anywhere in the study area given a dependent variable (population) and a set of one or more independent variables (building count, building area and building volume) measured at known locations. A GWR model can be expressed as (Lo 2008):

$$Y_i = a_{i_0} + \sum_{k=1}^{n} a_{ik} x_{ik} + e_i \quad (i = 1, 2, \ldots, n), \tag{1}$$

where Y_i and x_{ik} are the dependent and independent variables at i, $k = 1, 2, \ldots, n$, e_i are normally distributed error terms (with zero mean and constant variance at point i) and a_{ik} are the values of the kth parameters at locations i. In this

study, adaptive kernels are used and the bandwidth is determined using cross validation (CV). More details of GWR models can be found in Fotheringham *et al.* (2002), Lo (2008) and ESRI (2009).

(7) Accuracy assessment. Accuracy assessment is an important step in population estimation using remote sensing and GISs. Similar to Lu *et al.* (2006), here we use three measures for accuracy assessment:

$$\text{relative error: RE} = (P_e - P_g)/P_g \times 100, \tag{2}$$

where P_e and P_g are the estimated and reference population in each census block, respectively. Values of RE can be stored in a field in the census-block attribute table to support convenient mapping using the GIS. Additionally, mean relative error (MRE) can be used to quantify model performance:

$$\text{mean relative error: MRE} = \frac{\sum_{k=1}^{n} |\text{RE}_k|}{n}. \tag{3}$$

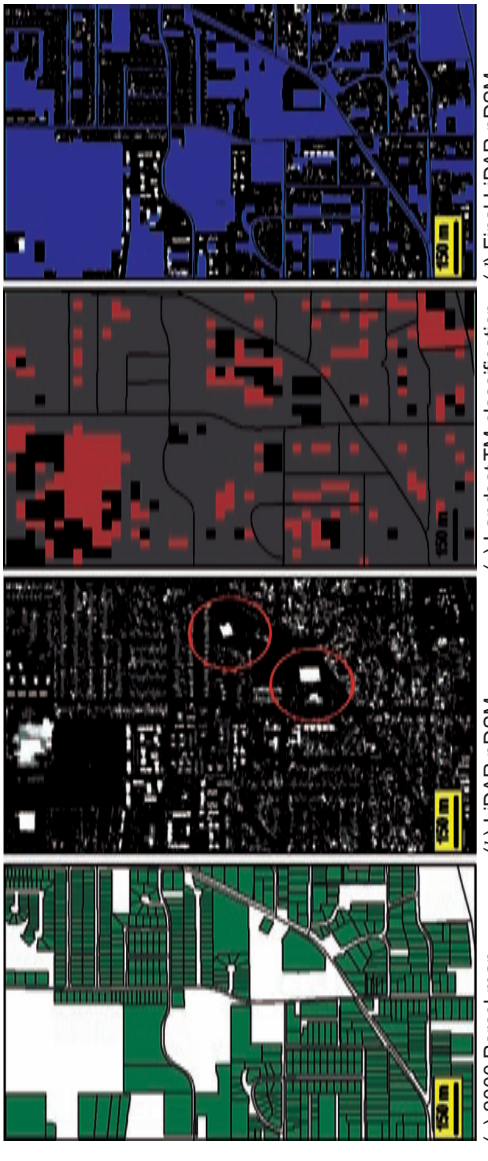
To reduce the influence of extreme values, a median absolute relative error (MARE) is also used.

As an example of data pre-processing, figure 6 shows the input layers and the final nDSM. From the parcel records, pre-2000 (including 2000) residential and commercial parcels are selected to create a parcel map (figure 6(*a*)). Commercial parcels are selected at this stage because apartment complexes are listed as commercial parcels. This parcel map is then used as a mask to remove non-residential and non-commercial areas from the LiDAR nDSM map. As a result, some commercial buildings may exist in the LiDAR nDSM map (see red circles in figure 6(*b*)). Due to the relatively large sizes and spectral properties of the roofs and neighbouring car parks, these buildings can be classified as CIT on Landsat TM images. Figure 6(*c*) shows a classified Landsat TM image in which grey pixels are residential areas, red pixels are CIT, yellow pixels are soils and green pixels are vegetation. It can be seen that the two isolated buildings in figure 6(*b*) are classified as CIT. Even though the Landsat TM images have a much lower resolution compared with LiDAR nDSM, the spectral information provided by Landsat TM images can help remove non-residential areas from the nDSM. Figure 6(*d*) is the final nDSM after combining the parcel map, the original nDSM and the Landsat TM image classification results. This final nDSM is the base map for population estimation using regression models for building count, building area and building volume.

4. Results and discussions

Figure 7 shows the regressions models derived from (1) 91 continuous sampling census blocks (see shaded blocks in figure 4(*a*)) and (2) 91 random sampling blocks (see shaded blocks in figure 4(*b*)). A summary of model results are listed in tables 1, 2, 3 and 4.

To better understand the error distribution of the regression models, scatter diagrams were generated to show the relationship between relative estimation error and population density at the census-block level (figure 8). In figure 8, the relative errors were obtained from OLS regression models built from random samples using (a) building count as the independent variable, (b) building area as independent variable



(*a*) 2000 Parcel map (*b*) LiDAR nDSM (*c*) Landsat TM classification (*d*) Final LiDAR nDSM

Figure 6. Samples of input data layers and final LiDAR nDSM. Several large buildings in the red circles in (*b*) are classified as CIT in (*c*), where grey is residential buildings, red is CIT, yellow is soil, and green is vegetation, and removed from the final LiDAR nDSM in (*d*), where blue areas represent a non-residential mask.

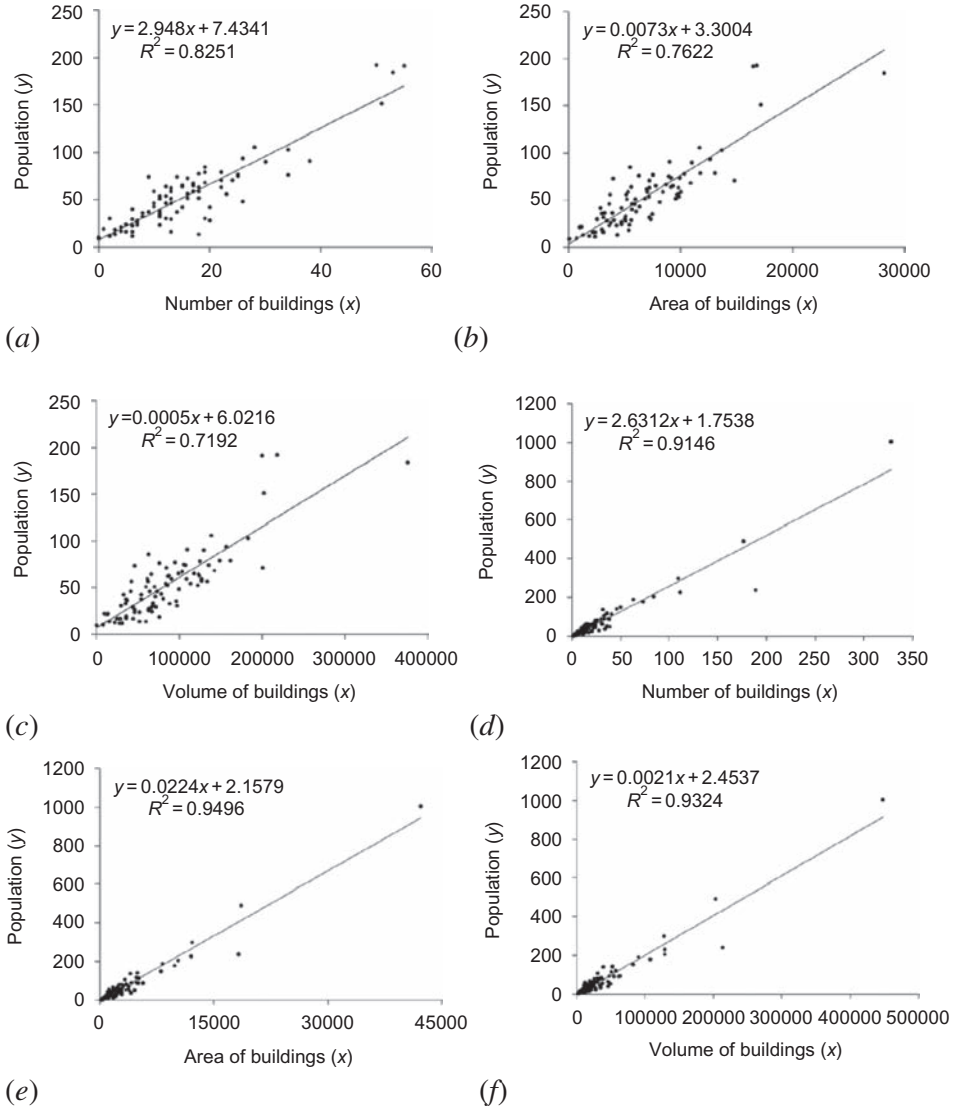


Figure 7. Regression models derived from sampling blocks: (*a*), (*b*) and (*c*) are from continuous sampling blocks shown in figure 4(*a*), while (*d*), (*e*) and (*f*) are from random sampling blocks shown in figure 4(*b*).

Table 1. Summary of continuous sampling model results.

Independent variable	Regression model	R^2	MRE (%)	MARE (%)	TE (%)
nDSM building count	$y = 2.948x + 7.4341$	0.8251	53.46	29.43	–25.42
nDSM building area	$y = 0.0073x + 3.3004$	0.7622	52.84	41.75	–35.29
nDSM building volume	$y = 0.0005x + 6.0216$	0.7192	55.77	47.12	–38.18

MRE: mean relative error for the overall dataset; MARE: median absolute relative error for the overall dataset; TE: total population estimation error (%) based on the overall dataset in the study area; R^2: coefficient of determination.

Table 2. Summary of random-sampling model results.

Independent variable	Regression model	R^2	MRE (%)	MARE (%)	TE (%)
nDSM building count	$y = 2.6312x - 1.7538$	0.9146	55.80	26.14	–23.03
nDSM building area	$y = 0.0224x - 2.1579$	0.9496	38.87	24.20	–26.57
nDSM building volume	$y = 0.0021x - 2.4537$	0.9324	46.41	23.92	–24.06

Table 3. Summary of multiple-regression model based on random sampling.

Independent variables	Regression model	R^2	MRE (%)	MARE (%)	TE (%)
nDSM building area (x_1) and building count (x_2)	$y = 0.0299x_1 - 10.9131x_2 - 1.5613$	0.9520	36.12	23.66	–27.48

Table 4. Summary of geographically weighted regression (GWR) results.

Independent variable	Regression model	R^2	MRE (%)	MARE (%)	TE (%)
nDSM building count	GWR	0.9701	44.55	21.19	–24.93
nDSM building area	GWR	0.9789	38.09	15.96	–27.16
nDSM building volume	GWR	0.9741	42.30	16.62	–25.36

and (c) building volume as the independent variable. It should be noted that scatter diagrams created from the multiple regression model and GWR models also show error distribution patterns very similar to figure 8, although these diagrams are not included in the paper.

A few observations of the above results are discussed below.

(1) The R^2 values for OLS models and GWR models suggest that population count is strongly correlated with residential building count, area and volume derived from the final nDSM. Although other studies show that the traditional housing-unit method (HUM) offers a number of advantages over other population-estimation methods (Smith and Lewis 1980, 1983, Smith and Cody 1994), there is no obvious pattern to show that building count outperformed building area and volume in this study. High-resolution image data and LiDAR data are need for more detailed evaluation of the three independent variables for small-area population estimation.

(2) Figure 8 indicates that population count is often overestimated when population density at the census-block level is less than approximately 300 persons km^{-2}, whereas population count is always underestimated when population density is greater than approximately 3500 persons km^{-2}, regardless of the independent variable. Similar observations have been reported by Lu *et al.* (2006) for population-density estimation using Landsat TM data. It seems that the total estimation error is controlled by both the sign and magnitude of the relative estimation errors. In this study, underestimation has a larger magnitude because

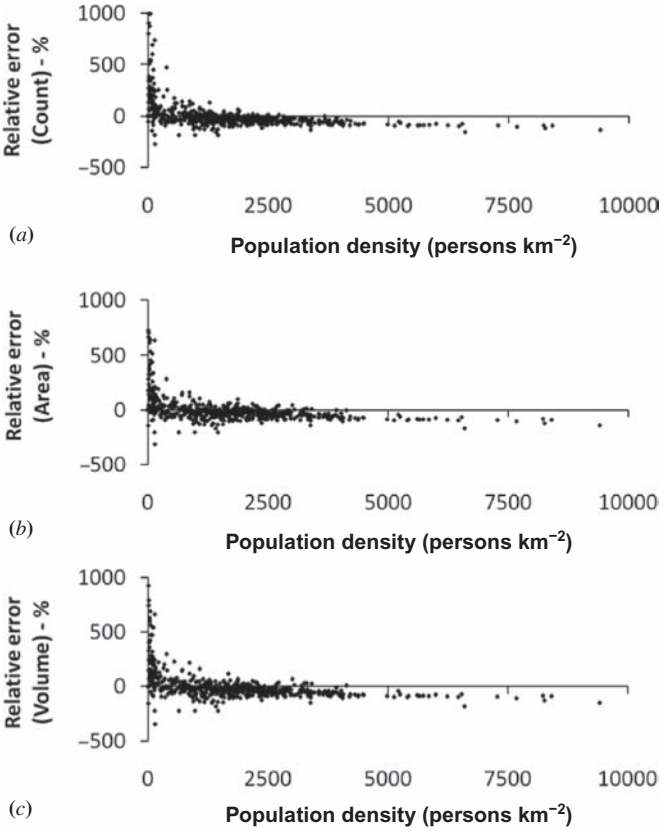


Figure 8. Scatter diagrams of relative population-estimation error versus population density at the census-block level. The relative errors were obtained from OLS models built from random samples using (*a*) building count as the independent variable, (*b*) building area as the independent variable and (*c*) building volume as the independent variable.

it mainly happens in census blocks with a high population density. As a result, total population is underestimated, as shown in tables 1 to 4.

(3) Models derived from random samples seem to generate more accurate results compared with those derived from continuous samples because random samples can better represent data distribution in the study area than clustered samples. Values of the median absolute relative error (MARE) indicate that: (a) the OLS models and GWR models built from randomly selected samples provide more accurate estimates than the OLS model built from continuous samples at the census-block level and (b) the GWR models provide more accurate estimates than the OLS models because spatial heterogeneity can be better modelled in GWR models than in OLS models. The better performance of GWR models than OLS models is also reported by Lo (2008) for population estimation in Atlanta.

(4) For census blocks with a low population density (for example, less than 100 persons km^{-2}), relative error of population estimation in percentage can be misleading. For example, for a census block with actual population of 2 and

estimated population of 6, the relative error is 200%, but the actual error of 4 persons may be insignificant compared with the total population of the census tract. The small-number problem needs be taken into account for population estimation at census-block level.

(5) Contrary to (4), an extremely high population density in a small census block (such as high-rise apartment complexes) would also affect the total estimation error. Separation of high-rise apartment buildings from commercial/industrial ones using remote sensing remains a challenge. Further study is required to test the use of physical properties of LiDAR-derived buildings and other models of population estimation such as kriging (Wu and Murray 2005).

(6) The LiDAR data in this study were collected during leaf-on season (4 September 2001), which makes it difficult to separate residential buildings from tree canopies. This may not be a serious problem for relatively new communities where most trees are small, but can be a factor affecting population estimation in older communities with many mature trees. It is anticipated that high-resolution LiDAR data can help separate tress from buildings based on measures of slope, aspect and three-dimensional (3D) shape. A recent study suggests that 3D shape signatures derived from LiDAR data can be used to discriminate different tree crowns (Dong 2009). It is reasonable to expect that 3D shape signatures of buildings and trees can be different, which may help separate trees from buildings.

(7) As shown in figure 6, integration of spatial information provided by LiDAR and spectral information provided by Landsat TM images may be very useful for information extraction in some areas. However, because of the big difference in spatial resolution of Landsat TM and LiDAR data, using Landsat TM image classification as a mask may also cause information loss in the final nDSM data. The integration of spatial information from LiDAR and size, shape and texture information from high-resolution images (such as IKONOS) could help separate trees from buildings and improve the accuracy of population estimation.

(8) The LiDAR data used in this study were resampled to a point spacing of 3–5 m, which affects accurate representation of buildings. It is possible that the relatively low accuracies for small-area population estimation were partly caused by the relative low spatial resolution of the LiDAR data. It would be interesting to compare area-based, volume-based and count-based OLS models and GWR models when high-resolution LiDAR data are available.

5. Conclusions

Based on LiDAR, Landsat TM and parcel data, a refined nDSM was created and used for small-area population estimation. Using census 2000 data and census-block samples, 10 OLS models and GWR models were built from independent variables of building count, building area and building volume obtained from the nDSM. The regression models were then applied to the census blocks in the study area, and accuracy assessments were carried out. Models derived from random samples seem to generate more accurate results compared with those derived from continuous samples because random samples can better represent data distribution in the study area than clustered samples. Median absolute relative errors suggest that the GWR models outperformed the OLS regression models because spatial heterogeneity in

population density is better handled in GWR models. The results show that the total accuracy of population estimation in the study area is controlled by the sign and magnitude of relative errors at the census-block level. Since underestimation usually happens in census blocks with high population density, the total population count in the study area is underestimated, with a minimum estimation error of about –23%. Such an accuracy is not high enough for small-area population estimation. Possible reasons behind the relatively low accuracies are (1) lack of high-resolution LiDAR data and image data and (2) difficulty in separating trees and buildings using Landsat TM images and LiDAR data with 3–5 m point spacing. It would be interesting to compare the results with those derived from high-resolution LiDAR data and image data (such as IKONOS), which would help evaluate the application of LiDAR and remotely sensed image data for small-area population estimation. Alternative models such as kriging should be evaluated to address issues caused by small numbers and variations in population density. Problems related to tree and building separation, shape indexes and signatures of trees and buildings and spatial/spectral integration of high-resolution satellite images and LiDAR data need to be further investigated.

Acknowledgements

The authors would like to thank the North Central Texas Council of Governments (NCTCOG) for providing LiDAR data and the US Geological Survey (USGS) for providing Landsat TM images. Thanks are also due to two anonymous reviewers for their comments and suggestions.

References

Antonarakis, A.S., Richards, K.S. and Brasington, J., 2008, Object-based land cover classification using airborne LiDAR. *Remote Sensing of Environment*, **112**, pp. 2988–2998.

Aubrecht, C., Steinnocher, K., Hollaus, M. and Wagner, W., 2009, Integrating earth observation and GIScience for high resolution spatial and functional modeling of urban land use. *Computers, Environment and Urban Systems*, **33**, pp. 15–25.

Bracken, I. 1991, A surface model approach to small area population estimation. *Town Planning Review*, **62**, pp. 225–237.

Bork, E.W. and Su, J.G., 2007, Integrating LiDAR data and multi-spectral imagery for enhanced classification of rangeland vegetation: a meta analysis. *Remote Sensing of Environment*, **111**, pp. 11–24.

Brandtberg, T., 2007, Classifying individual tree species under leaf-off and leaf-on conditions using airborne LiDAR. *ISPRS Journal of Photogrammetry and Remote Sensing*, **61**, pp. 325–340.

Cai, Q., 2007, New techniques in small area population estimates by demographic characteristics. *Population Research and Policy Review*, **26**, pp. 203–218.

Chen, Y., Su, W., Li, J. and Sun, Z., 2009, Hierarchical object oriented classification using very high resolution imagery and LiDAR data over urban areas. *Advances in Space Research* **43**, pp. 1101–1110.

Chow, T.E. and Hodgson, M.E., 2009, Effects of LiDAR post-spacing and DEM resolution to mean slope estimation. *International Journal of Geographic Information Science* **23**, pp. 1277–1295.

Dong, P., 1997, Implementation of mathematical morphological operations for spatial data processing. *Computers and Geosciences*, **23**, pp. 103–107.

Dong, P., 2009, Characterization of individual tree crowns using three-dimensional shape signatures derived from LiDAR data. *International Journal of Remote Sensing* **30**, pp. 6621–6628.

Environmental Systems Research Institute (ESRI), 2009, ArcGIS 9.3 Desktop Help, software.

FOTHERINGHAM, A.S., BRUNSDON, C. and CHARLTON, M., 2002, *Geographically Weighted Regression: the Analysis of Spatially Varying Relationships* (Chichester, UK: Wiley).

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, UK: Taylor & Francis).

HENDERSON, A., 2006, Estimating population size in emergencies. *Bulletin of the World Health Organization*, **83**, pp. 161–240.

HILL, R.A. and BROUGHTON, R.K., 2009, Mapping the understorey of deciduous woodland from leaf-on and leaf-off airborne LiDAR data: a case study in lowland Britain. *ISPRS Journal of Photogrammetry and Remote Sensing*, **64**, pp. 223–233.

JAROSZ, B., 2008, Using assessor parcel data to maintain housing unit counts for small area population estimates. In *Applied Demography in the 21st Century*, S. Murdock and D. Swanson (Eds), pp. 89–102 (Dordrecht, The Netherlands: Springer).

KOETZ, B., SUN, G., MORSDORF, F., RANSON, K.J., KNEUBÜHLER, M., ITTEN K., and ALLGÖWER, B., 2007., Fusion of imaging spectrometer and LiDAR data over combined radiative transfer models for forest canopy characterization. *Remote Sensing of Environment*, **106**, pp. 449–459.

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, NY: Longman).

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, UK: Taylor & Francis).

LO, C.P., 2008, Population estimation using geographically weighted regression. *Journal GIScience and Remote Sensing*, **45**, pp. 131–148.

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

LU, D., WENG, Q. and LI, G., 2006, Residential population estimation using a remote sensing derived impervious surface approach. *International Journal of Remote Sensing*, **27**, pp. 3553–3570.

MACEACHREN, A.M. and DAVIDSON, J.V., 1987, Sampling and isometric mapping of continuous geographic surfaces. *The American Cartographer*, **14**, pp. 299–320.

PLATEK, R., RAO, J., SARNDAL, C. and SINGH, M., 1987, *Small Area Statistics: An International Symposium* (New York City, NY: John Wiley and Sons).

QIU, F., WOLLER, K.L. and BRIGGS, R., 2003, Modeling urban population growth from remotely sensed imagery and TIGER GIS Road Data. *Photogrammetric Engineering and Remote Sensing*, **69**, pp. 1031–1042.

SMITH, S.K. and CODY, S., 1994, Evaluating the housing unit method: a case study of 1990 population estimates in Florida. *Journal of the American Planning Association*, **60**, pp. 209–221.

SMITH, S.K. and LEWIS, B., 1983, Some new techniques for applying the housing unit method of local population estimation: further evidence. *Demography*, **20**, pp. 407–413.

SMITH, S.K. and LEWIS, B., 1980, Some new techniques for applying the housing unit method of local population estimation. *Demography*, **17**, pp. 323–339.

Smith, S.K. and Mandell, M. 1984, A comparison of population estimation methods: housing unit versus component II, ratio correlation and administrative records. *Journal of American Statistical Association*, **79**, pp. 282–289.

Sohn, G. and Dowman, I., 2007, Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, **62**, pp. 43–63.

Stone, G., 2008, Methods for measuring the population after a disaster: household population surveys in post-Katrina New Orleans, October 2005–October 2006. In *Applied Demography in the 21st Century*, S. Murdock and D. Swanson (Eds), pp. 113–123 (Dordrecht, The Netherlands: Springer).

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.

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.

Texas Department of Housing & Community Affairs (TDHCA), 2009, Manufactured home construction and safety standards. Available online at: http://www.tdhca.state.tx.us/mh/ (accessed 14 September 2009).

Verma, R., 1992, Regression-based estimates of total population for local areas: then and now. In *Readings in Population Research: Policy*, *Methods and Perspectives*, P. Krishnan, C. Tuan and K. Mahadevan (Eds), pp. 167–178 (Delhi, India: B.R. Publishing).

Verma, R., Basavarajappa, K. and Bender, R., 1984, Estimation of local area population: an international comparison. In *1984 Proceedings of the Social Statistics Section*, pp. 324–329 (Alexandria, VA: American Statistical Association).

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

Wolter, K. and Causey, B., 1991, Evaluation of procedures for improving population estimates for small areas. *Journal of the American Statistical Association*, **86**, pp. 278–284.

Wu, C. and Murray, A.T., 2005, A cokriging method for estimating population density in urban areas. *Computers, Environment and Urban Systems*, **29**, pp. 558–579.

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.

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

Xie, Z., Roberts, C. and Johnson, B., 2008, Object-based target search using remotely sensed data: a case study in detecting invasive exotic Australian Pine in south Florida. *ISPRS Journal of Photogrammetry and Remote Sensing*, **63**, pp. 647–660.

Zhou, G., Song, C., Simmers, J. and Cheng, P., 2004, Urban 3D GIS from LiDAR and digital aerial images. *Computers and Geosciences*, **30**, pp. 345–353.