International Journal of Remote Sensing
Vol. 31, No. 21, 10 November 2010, 5605–5634

Assessing fine-spatial-resolution remote sensing for small-area population estimation

JOSE L. SILVÁN-CÁRDENAS†, LE WANG*†, PETER ROGERSON†, CHANGSHAN WU‡, TIANTIAN FENG† and BENJAMIN D. KAMPHAUS§

†Department of Geography, State University of New York, 105 Wilkeson Quad, Buffalo, NY 14261, USA

‡Department of Geography, University of Wisconsin-Milwaukee, Bolton Hall 482, P.O. Box 413, Milwaukee, WI 53201-0413, USA

§Department of Anthropology, State University of New York, 380 MFAC-Ellicott Complex Buffalo, NY 14261, USA

Small-area population estimation is an important task that has received considerable attention from the remote-sensing community in the past four decades. The wealth of related studies reveals that the notion of *living space* had been considered a key linkage between population and remote-sensing measurements. Unfortunately, a formal definition for this important variable has proved difficult, due, in part, to the relatively coarse spatial resolution of the remote-sensing data used for population estimation. The advent of airborne Light Detection And Ranging (LiDAR) sensors for measuring elevation at fine spatial resolutions has provided new opportunities for considering the three-dimensional nature of living space in urban environments and for improving small-area population estimations. In this study, we assess the potential of fine-spatial-resolution LiDAR measurements (1 m) coupled with automated techniques for building extraction and land-use classification. The study seeks to provide an answer to the question: what level of information extracted from fine-spatial-resolution LiDAR and aerial photographs can be realistically translated into improved small-area population estimation? This question is addressed through a comparative study of up to seven linear models with building count, building area and/or building volume as explanatory variables at one of two land-use levels: single-family dwelling, multi-family dwelling and other types, versus residential and other types. Results show that, while building volume fits more naturally the population figures, it also represents the most challenging variable to measure by automated means. Because of this, a simple model expressed in terms of residential-building counts results in more reliable population estimates.

1. Introduction

Small-area population estimates are essential for understanding and responding to many social, political, economic and environmental problems (Liu 2003), such as resource allocation (Smith *et al.* 2002), public-transit route design (Benn 1995), customer-design analysis, market-area delineation and site-location identification (Martin and Williams 1992, Plane and Rogerson 1994). Billions of dollars from public funds are allocated every year based on diagnostic indicators such as unemployment rates, mortality and morbidity rates, etc., which have population estimates as a common denominator. Despite its great significance to many applications, detailed

*Corresponding author. Email: lewang@buffalo.edu

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2010 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/01431161.2010.496800

and accurate population information is only available on one date per decade in most countries, a sampling rate that is well below the changing rates of fast-growing cities. Therefore, the development of new techniques for generating accurate and timely population estimates is of great importance (Smith *et al.* 2002).

Population estimation by means of remote sensing has a long tradition that can be traced back to the 1950s (Porter 1956, Hsu 1971, Kraus *et al.* 1974, Lo and Welch 1977), when aerial photographs were used to count dwelling units. Since then, the field has evolved both in sophistication of sensors used for data acquisition, as well as in methods and models involved in estimation. Lo (2006) identified four main approaches for population estimation using remote sensing. These approaches are: (1) counting dwelling units, (2) measuring urban built-up areas, (3) measuring areas of different land use/land cover (LULC) and (4) automated digital analysis. A similar scheme was also presented by Wu *et al.* (2005), who also discussed various interpolation approaches pertinent to population estimation. Arguably, Lo's approaches (2) and (3) are essentially the same, with the major difference being the scale. Nonetheless, the LULC provides thematic information apart from area, which is generally used to mask out non-residential land uses, or for model stratification of various kinds (e.g. Watkins and Morrow 1985, Lo 1995, 2003, Langford 2006).

Traditionally, population estimates are based on dwelling-unit counts from aerial photography (Hsu 1971, Lo and Chan 1980, Watkins and Morrow 1985, Lo 1986, Taragi *et al.* 1994). This approach, which largely relies on visual identification and enumeration of dwelling units, is a relic of the housing-unit (HU) method of population geography that is used routinely by the US Census Bureau for estimating sub-county (census-tract and census-block) population in non-census years (Smith and Cody 2004). The HU method assumes that almost every person (except the homeless) lives in a housing structure, and it is therefore feasible to generate population estimates for a small area by counting the number of houses within that area (Smith and Cody 2004). From a remote-sensing perspective, small-area population estimates (P) are calculated by multiplying the number of housing units (U) determined from the aerial photography with the average occupancy rate (r) and the average household size (h), i.e. $P = rhU$. The approach can be extended to consider various housing types with different household sizes and vacancy levels. For instance, Watkins and Morrow (1985) considered the number of single-family (U_1) and multi-family (U_2) units and expressed the total population in the form $P = r_1h_1U_1 + r_2h_2U_2$, where the model parameters (r and h) are determined from field surveys or from a prior census, or even through a regression approach.

A second group of methods relates the planimetric area of living spaces to population. Examples of this group include those methods that consider the areas of cities (Lo and Welch 1977, Lo 1986, 2001, Sutton *et al.* 2001), areas of land uses (either residential or non-residential, e.g. Kraus *et al.* 1974, Lo 2003, 2008, Langford 2006) and, more recently, areas of impervious surface (Lu *et al.* 2006a, Wu and Murray 2007, Morton and Yuan 2009). Most of these studies relied on Landsat Thematic Mapper (TM) or Landsat Enhanced Thematic Mapper (ETM+) coupled with some digital image-analysis techniques for LULC or impervious-surface extraction, with the exceptions of Lo (2001) and Sutton *et al.* (2001), who applied thresholds to nighttime light from the Defense Meteorological Satellite Programme Operational LineScan System (DMSP OLS) to determine urban built-up areas. The initial motivation for relating area to population was the biological law of allometric growth introduced by Huxley (1932) and examined by Nordbeck (1965) in a geographical context. In a sense, the urban allometric growth law loosely states that the relative population growth is proportional to the relative growth of the *living space*. More precisely, $P = aA^b$, where A is the inhabited area, i.e. the urban area/built-up area at one scale, or the residential land-use area at another scale, and a and b are the model parameters. Because population is essentially a three-dimensional variable (i.e. a measure of the volume occupied by human bodies) and the inhabited area is a two-dimensional variable, Nordbeck (1971) conjectured that the exponent of the relation must always be $b = 3/2$, which appeared to be the case at city level, but is not quite generalizable for other various urbanization levels or finer spatial units (e.g. Lo 1986, Sutton *et al.* 2001, Morton and Yuan 2009). Since a general theory leading to the allometric growth equation is lacking, and because the implications from the modifiable areal unit problem (MAUP; Openshaw and Taylor 1981) question its generality, others forms of the nonlinear, area–population relationship can be used. For instance, instead of using logarithms, Lu *et al.* (2006a) linearized the relationship by taking the square root of population and found a high correlation coefficient ($R^2 \approx 0.8$) with various impervious surfaces. To this end, the complexity of the area–population relationship is recognized to be strongly scale dependent. At the fine levels, the issues of spatial autocorrelation and non-stationarity come to play a significant role for small-area population estimation (Langford 2006, Hardin 2008, Lo 2008).

A third group of methods seeks to directly relate physical measurements made by the sensors (or their transformed values) to the population. The approach was first proposed by Iisaka and Hegedus (1982) and was extensively used by Lo in his studies of Chinese cities (Lo 1989, 1995). The rationale for this approach is that some spectral bands are more sensitive to urban reflectance than others. For instance, Lo (1995) observed strong negative correlations of population density at city level with radiances of Satellite Pour l'Observation de la Terre (SPOT) High Resolution Visible (HRV) band three (infrared) and band one (green) and a positive correlation with band two (red). Because bands one and three are associated with vegetation density, they serve to indicate absence of or low-density population, whereas band two is associated with the high albedo of urban built-up areas, resulting in higher correlation with population density. Naturally, researchers have also successfully related nighttime light energy from the DMSP OLS to population, but at a much coarser spatial resolution (Sutton *et al.* 1997, Dobson *et al.* 2000, Lo 2001). While relating urban radiometric measurement to population seems to work for large areas, it has been noted that, for small areas, this approach tends to produce poor results (Wu *et al.* 2005). In light of this, many researchers have also tested transformed values such as principal components and band ratios (Harvey 2002a, b, Li and Weng 2005), texture descriptors from semi-variances and the grey level co-occurrence matrix (Wu *et al.* 2005, Liu *et al.* 2006), as well as higher order spatial metrics derived from thematic classifications (Liu *et al.* 2006, Hardin 2008). All these studies have shown promising results for small-area population estimations (down to aggregation levels of block group and block, in US census units), but are still limited by the poor spatial resolution of remote-sensing data used.

Although very-fine-spatial-resolution satellites (IKONOS with 1 m and QuickBird with 0.65 m in their respective panchromatic bands) have been available for around a decade now (IKONOS launched in 1999 and QuickBird launched in 2001), studies taking advantage of this fine spatial resolution for small-area population estimations are still scarce and generally limited to the application of traditional visual interpretation for HU count (e.g. Yagoub 2006). Likewise, airborne Light Detection And Ranging (LiDAR) devices have allowed rapid access to vertical information of urban structures, but the integration of this new level of information for population estimation has not been fully investigated. Recently, it was suggested that the volume information provided by LiDAR could serve to best improve small-area population estimations (Wu *et al.* 2008). Whether or not LiDAR measurements coupled with automated techniques for building extraction and land-use classification can lead to improved small-area population estimation is still an unresolved matter.

This study sought to provide an answer to the question: what level of information extracted from LiDAR measurements and fine-spatial-resolution imagery can be realistically translated into improved small-area population estimation? This question was addressed through a comparative study of up to seven linear models expressed in terms of building count, building area and/or building volume, at one of two land-use levels. At the most detailed land-use level, extracted buildings were classified as single-family dwelling, multi-family dwelling or other types, whereas at the least detailed level, they were classified as either residential or other types. The models based on building count and building area were meant to represent two of the most prominent categories of population-estimation models discussed above, whereas models that used building volume served as a means for testing the hypothesis that residential-building volume can more naturally fit population figures, as it provides a better representation of the three-dimensional nature of living space. In order to provide and answer the question posed above, a number of automated building extraction and land-use classification methods were developed, and the impact of the inaccuracies from remote-sensing methods in the final population estimation was quantified.

The remainder of the manuscript is organized as follows. Section 2 introduces the study area and data used in this study, highlighting sources and purpose of data. Section 3 summarizes the methods used for data preparation, building detection, land-use classification, population estimation and accuracy assessment. Section 4 presents the main results by providing an answer to the specific question posed in this study and §5 presents a discussion and conclusions.

2. Study area and data used

The study area is located in the City of Austin, capital of Texas (figure 1). Austin is currently the third fastest growing large city in the US with a population of 750 000. According to the US censuses of years 1990 and 2000, the city grew an impressive 41%, from 465 622 in 1990 to 656 562 in 2000, with an average annual growth rate of 3.5%. The city's population has been projected to top 800 000 by 2010 (City of Austin 2009a). The selected area for this study covers approximately 4.8 km x 6.4 km, representing 4.5% of the entire city area (figure 1). A major interstate highway, IH-35, runs S–N and splits this area in two sides. The west side is dominated by civic, commercial, as well as some residential land uses located in the northern and southern ends (figure 2). Civic land use is defined as the land covered by semi-institutional housing, hospital, government services, educational meeting and assembly, and cemetery (City of Austin 2009b). In the study area, the civic land-use areas correspond to the University of Texas (UT) campus, whereas the commercial land, located south of the UT campus, corresponds to the fringes of the city's central business district, which is located at the far left of the study boundary. The east side of the study area is dominated by residential land uses, with some instances of industrial land use located along major roads. A total of 1153 census blocks, with nearly 20 000 buildings, cover this area. Most buildings are residential dwellings (85%), of which 94% are

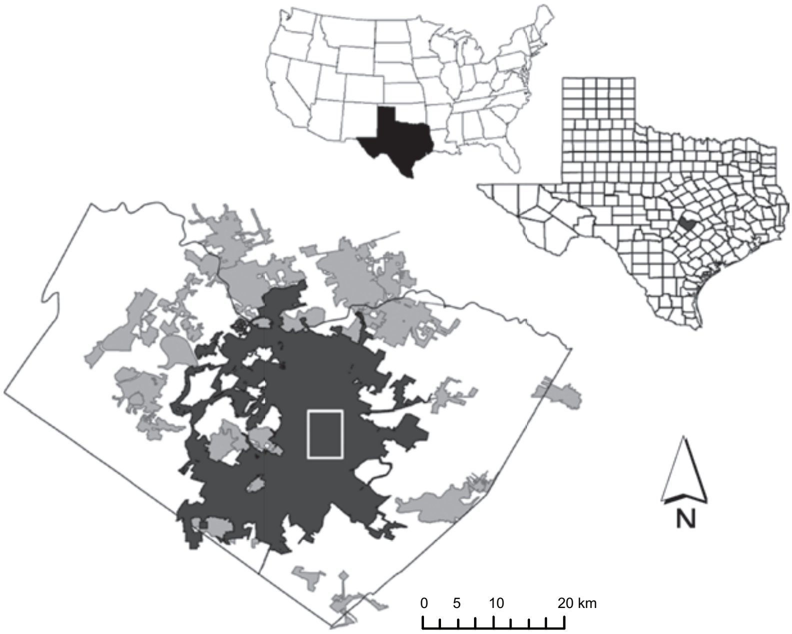

Figure 1. Geographic location of the study area. The key maps show the location of Texas State in the conterminous US (top), the location of Travis County in Texas State (right), and the location of Austin city in Travis County (centre).

single-family detached and two-family attached, and only 6% are multi-family three-/four-plex (dwelling of three or four units) and apartment/condo (individually owned units in a single building with common facilities).

In order to perform detailed analyses of building-detection and land-use classification methods, we selected four small study areas within the larger study area. The inset boundaries are shown in figure 2. These sites were carefully selected to represent the spectrum of living environments found in the study area:

- inset 1 contains multi-family dwelling units located in a sparsely vegetated area,
- inset 2 contains single-family dwelling units located in a densely vegetated area,
- inset 3 contains single-family dwelling units located in sparsely vegetated area and
- inset 4 contains both single-family and multi-family dwelling units located in a sparsely vegetated area.

Datasets acquired for the study area include LiDAR altimetry measurements, demographic and geographic census data, building footprints and land-use layers, aerial photography and a Landsat TM image. All datasets were contemporarily acquired around year 2000. The LiDAR data was provided by the Bureau of Economic Geology of the University of Texas at Austin, and was acquired in 2000 using an Optech, Inc. Airborne Laser Terrain Mapper (ALTM; Optech, Inc., Vaughan, Ontario, Canada) 1225 instrument mounted on a single engine craft. The ALTM instrument delivers a cloud of three-dimensional points for the first and last return of a laser pulse. For each return, the backscatter intensity was also recorded. The average point density was around 3 points m^{-2}, and there were around 40 million points in the entire study

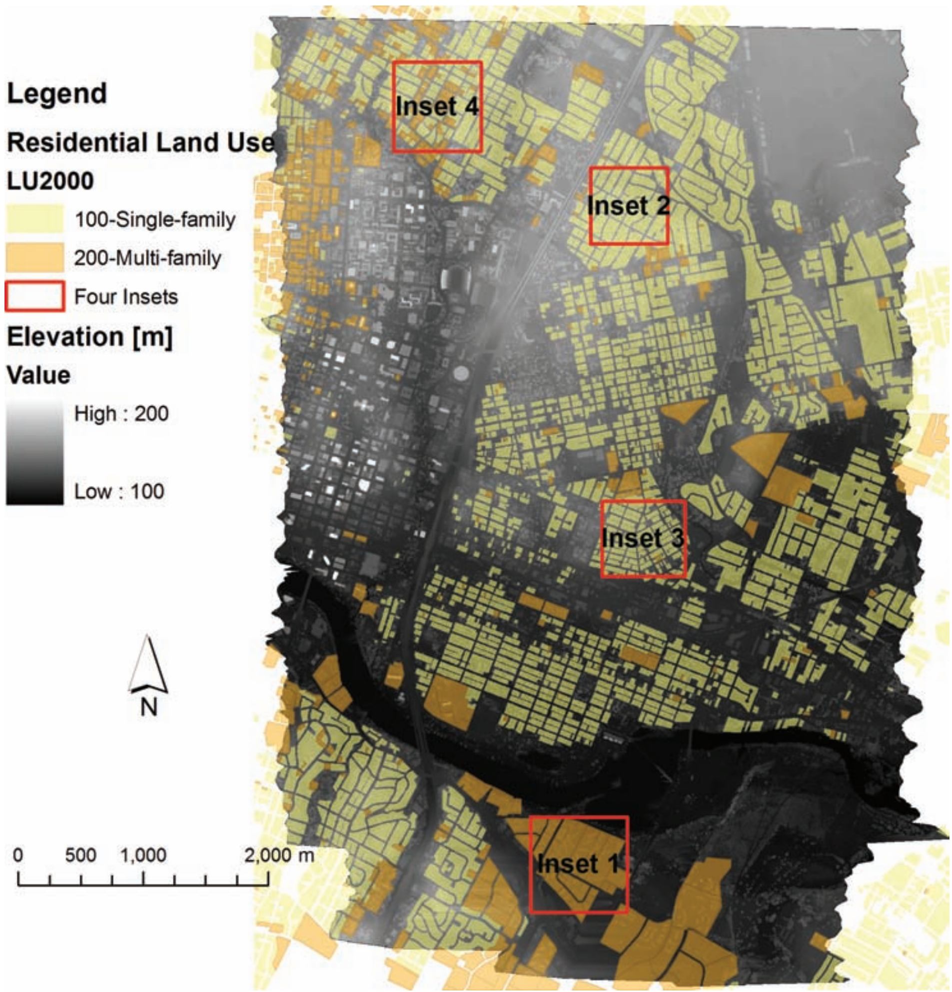

Figure 2. This map shows the location of four insets within the study area, the residential land-use areas and the LiDAR elevation measurements as a background image.

area. Demographic and geographic data were acquired at census-block level through the US Census Bureau's American FactFinder (US Census Bureau 2009a) and TIGER/Line shapefiles (US Census Bureau 2009b) web sites, respectively. Building footprint and land-use layers, together with a fine-spatial-resolution (0.6 m) colour–infrared (CIR) aerial photography, were acquired through the City of Austin Neighborhood Planning and Zoning Department (NPZD; City of Austin 2009b). A Landsat TM image acquired in the spring of 2000 was downloaded from the US Geological Survey's GloVis data distribution portal (US Geological Survey 2009).

3. Methods

Figure 3 shows the workflow of the study. This workflow involved methods for: (1) data pre-processing, (2) building extraction, (3) land-use classification, (4)

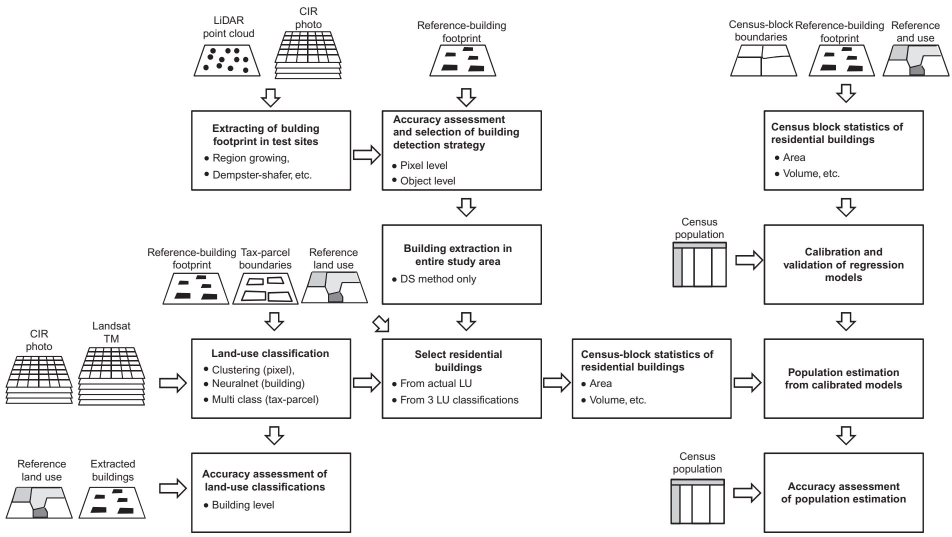

Figure 3. This chart shows the workflow of the study, indicating major inputs and processing steps.

population estimation and (5) accuracy assessment. These methods are briefly described below, with much technical details obviated due to constraints in the length of the manuscript. All implementation were based on ESRI's ArcGIS suite (ESRI, Redlands, CA, USA), ENVI v4.5 (ITT Visual Information Solutions, Boulder, CO, USA, 2008), eCognition (Definiens, Inc., München, Germany), the open-source data mining software WEKA (Witten and Frank 2005) and custom developments in MATLAB v7.3 programming language (The MathWorks, Inc., Natick, MA, USA, 2006).

3.1 *Data pre-processing*

3.1.1 Derived raster layers. Several raster layers were derived from the LiDAR point cloud and the CIR photograph (see table 1), which served as inputs to various automatic building detection and land-use classification methods. The layers derived from the CIR photograph included the following masks: vegetation, bare ground, impervious surface and pervious, non-bare ground. The vegetation mask was defined by applying a threshold to the normalized difference vegetation index (NDVI). Specifically, a pixel was considered vegetation if NDVI > 0.5, or non-vegetation otherwise. The other masks were produced through the maximum likelihood classification (MLC) of the CIR photograph. Training samples of up to five land-cover classes (bare ground, building roof, road/parking lot, tree and grass) were manually delineated in ENVI software and passed to the MLC routine to produce a classification image. Then, the bare-ground mask was extracted from the classification result, the impervious-surface mask was produced by clumping the classes paved roads, parking lots and building roofs, whereas the pervious, non-bare-ground mask was produced by clumping the classes tree and grass. The later can be considered as an alternate vegetation mask.

Table 1. Input–output relationship among various raster layers used in the study. All output layers had a cell size of 1 m, with the exceptions of the Landsat image, with a cell size of 30 m, and the building density, with a cell size of 500 m.

Input layer(s)	Output layer(s)
LiDAR last return (elevation value)	Digital surface model (DSM)
DSM	Ground mask
	Angular second moment
DSM & ground mask	Digital terrain model (DTM)
DSM & DTM	Feature height
LiDAR two returns (intensity value)	Intensity difference
CIR photo	Normalized difference vegetation index (NDVI)
	Impervious surface mask
	Bare-ground mask
	Pervious, non-bare-ground mask
NDVI	Vegetation mask
Landsat	Digital numbers (six bands)
Building footprint	Building count grid[a]

[a]Number of footprint centroids per cell of 500 m x 500 m

The datasets derived from LiDAR consisted of the feature height, ground mask, intensity difference and the grey level co-occurrence matrix (GLCM) angular second moment of a digital surface model (DSM). The DSM was produced at a spatial resolution of 1 m by applying a point to raster conversion tool to the point-cloud elevation values in ArcGIS. Elevation values from the last return of the laser pulse were selected due to its likelihood of penetrating through tree canopy. Because the ground-sampling distance of the point data was not uniform, the number of points within each 1 m cell varied from none to several points. The raster conversion assigned the minimum elevation from the points within each cell, or an interpolated value from adjacent neighbour cells in the case that no point occurred inside the cell. Once in raster format, a ground mask was produced using the multi-resolution ground-filtering approach described by Silván-Cárdenas and Wang (2006). Following such a method, the DSM was first decomposed into coefficients of a multi-resolution transform, termed the multi-scale discrete Hermite transform (Silván-Cárdenas and Escalante-Ramírez 2006). Then, the transform coefficients were eroded, and the inverse transformation was applied to recover a bare-earth digital terrain model (DTM). The recovered DTM was then subtracted from the DSM to produce a feature-height layer, which was in turn used for producing a ground mask. Cells in the ground mask were set to one or zero, depending on whether the height was less than 1 m. The GLCM angular second moment texture layer was also produced from the DSM with a moving window of 3 x 3 cells and a shift parameter of one horizontal increment per interval. The angular second moment calculation is a strong measure of homogeneity versus dissimilarity, and, when applied to a DSM, is extremely sensitive to changes in elevation, making it an ideal layer for facilitating building-surface segmentation.

The process used to produce the DSM was also followed for the production of two intensity layers, one using the first return and one the last, in which the value assigned to the resulting raster was that of the intensity field rather than the elevation field. After these two intensity raster layers were computed, a simple intensity-difference layer was produced using map algebra in ArcGIS, in which the value of the last-pulse intensity layer was subtracted from the value of the first-pulse intensity layer.

3.1.2 Reference-building footprint. The reference-building-footprint layer was an edited version of the NPZD's building-footprint layer. Since the latter was produced by manual digitization using aerial photos and LiDAR datasets collected in 2003 (City of Austin 2009b), we edited this layer to match the aerial photography and the LiDAR data acquired in 2000. Specifically, we calculated the building height from the LiDAR data and deleted all buildings with an average height below 2.5 m. Although the city regulation permits a minimum first-floor height of 4.5 m, the much lower threshold was used due to possible discrepancy in building-height definition. In this study, we calculated the building height as the average LiDAR elevation within the building footprint minus the minimum elevation within a 2 m buffer of the building footprint. Close inspection of deleted building footprints with the help of the CIR photography confirmed that these buildings were built after the census year. A total of 18 453 buildings remained in this layer.

Besides building height (H), we calculated, for each building, its footprint area (A), footprint perimeter (P), volume (V) and shape (S). The building volume was calculated as the product of footprint area and its height, i.e. $V = AH$. The building shape was defined through the two-dimensional compactness index:

$$S = 4\pi \frac{A}{P^2}. \tag{1}$$

Other building attributes, including neighbourhood and zonal statistics, were also attached to the reference layer. The complete set of building attributes calculated is provided in table 2. Attributes related to build-up density were based on Thiessen polygons around the buildings. Thiessen polygons were first constructed around all building vertices and then, each of those belonging to a single building, were merged into one polygon. Merged Thiessen polygons were meant to provide a stand in for parcels as a source of information about density of buildings. Attributes related to building neighbourhood correspond to spatial statistics on a neighbourhood defined through a Delaunay triangulation of building-footprint centroids.

The reference-building-footprint layer was also used to produce a raster layer of building count. The cell value in this raster layer was the number of building centroids within each cell of 500 m x 500 m.

Table 2. Attributes calculated for the building layer.

Category	Attributes
Building geometry	1. Building footprint area 2. Building footprint perimeter 3. Building footprint shape 4. Building height[a] 5. Building volume
Built-up density	6. Area of Thiessen polygon[b] 7. Footprint area to Thiessen polygon area ratio 8. Impervious surface percentage to Thiessen polygon area
Building neighbourhood	9–24. Average and standard deviation for attributes above within neighbourhood[c]

[a]Average elevation in footprint less minimum elevation in a 2 m buffer
[b]Based on footprint vertices and merged for each footprint
[c]Based on a Delaunay triangulation of building centroids

3.1.3 Reference land-use layer. The spatial unit of the NPZD's 2000 land-use layer is the tax parcel. Its classification scheme consists of 15 land-use classes (City of Austin2009b). The classes (with codes in parentheses) are: large-lot single-family (50), single-family (100), mobile homes (113), multi-family (200), commercial (300), office (400), industrial (500), mining (560), civic (600), open space (700), transportation (800), utilities (870), undeveloped/rural (900) and water (940). This classification scheme was reduced to nine classes by considering only non-empty classes at the 100 level of the City of Austin's land-use code. For instance, class mobile homes (113) was reclassified as single-family (100), whereas class large-lot single-family was deleted, as it was empty. The reclassification led to two residential classes (single-family and multi-family) and up to seven non-residential classes (commercial, office, industrial, civic, open space, transportation and undeveloped/rural). Note that quarter population living in civic land use, such as university accommodation, was not estimated in this study.

The land-use layer was also edited to minimize discrepancies with other data sources. Inspection of consistency between land use and population revealed that some census blocks that corresponded to non-residential (nor civic) land uses had non-zero population. This inconsistency could have been due to a temporal mismatch between acquisition dates of census and the land-use data. Hence, the land-use parcels within such census blocks were reclassified as either residential or other types based on a visual inspection of the CIR photography and with the help of Google Earth v4.3 (Google, Inc., Mountain View, CA, USA, 2008).

A number of attributes were also calculated for each parcel in the edited land-use layer. A complete list of attributes is provided in table 3. The various percent covers

Table 3. Attributes calculated for the tax-parcel layer.

Category	Attributes
Parcel geometry	1. Parcel area
	2. Parcel perimeter
	3. Parcel shape
Neighbour parcels	4. Distance to nearest neighbour
	5. Similarity of parcel area to nearest neighbour's[a]
	6. Similarity of parcel perimeter to nearest neighbour's[a]
	7. Similarity of parcel shape to nearest neighbour's[a]
Vegetation	8. Average NDVI
	9. Standard deviation of NDVI
	10. Percent cover of vegetation
Impervious surface	11. Percent cover of impervious surfaces
	12. Percent cover of bare ground
	13. Percent cover of pervious, non-bare ground
Building	14. Number of buildings in parcel
	15. Building area to parcel area ratio
	16–30. Average, minimum and maximum of building area, perimeter, shape, height and volume in parcel
Neighbour buildings	31–33. Average, minimum and maximum of distance between buildings in parcel and their nearest neighbour[b]
	34. Building density within a cell of 500 m x 500 m[c]

[a]Absolute difference normalized by parcel value
[b]Distances based on footprint centroid
[c]Average taken when multiple cells overlapped with the parcel

were computed using ArcGIS zonal statistic operations on the corresponding masks. Distances among parcels and among buildings were defined from polygon centroids, and nearest neighbours were found using a Delaunay triangulation on centroids. The building density for each parcel was calculated as a zonal average of the building-count grid divided by the area of the cell.

3.2 *Building-detection methods*

In order to select an appropriate building-detection method for population estimation, four building-extraction methods were tested on each of the insets shown in figure 2. These methods were based on region-growing segmentation (Zhang *et al.* 2006), the Hermite transform (Silván-Cárdenas and Escalante-Ramírez 2006), Dempster–Shafer theory of evidence (Shafer 1976, Lu *et al.* 2006b) and Definiens' eCognition segmentation method.

3.2.1 Region-growing segmentation. The first method consisted of a segmentation of the height layer by means of the region-growing algorithm as proposed by Zhang *et al.* (2006). The region-growing segmentation (RGS) is an iterative method that applies a plane-fitting technique to grow regions from seed pixels.

For each non-ground measurement area, inside and boundary points are identified. If at least one of the eight neighbours of a point is a ground measurement, the point is defined as a boundary point. Otherwise, the point is an inside point. The following residual is calculated for each inside point $p_0(x_0, y_0, z_0)$ and its eight neighbours:

$$R = \sum_{k \in M} [a(x_k - x_0) + b(y_k - y_0) + c - z_k]^2, \qquad (2)$$

where M is a set for the inside point and its neighbours, and a,b and c are plane parameters estimated through least-squares analysis. The point with the minimum residual R is labelled and selected as the first seed point for region growing. All neighbours of a seed point are labelled as belonging to the same segment if the deviation between its height and the plane height is under a threshold. A threshold of 0.1 m was used in our implementation. The plane parameters are then updated including the new labelled points. The neighbours of the grown area are examined further, and the process is continued until no additional points can be added into the segment. Then, the unlabelled point with the minimum R is selected as the next seed point. The process is repeated until all non-ground points are labelled.

After the RGS algorithm was run (and following Zhang *et al.* (2006)), small segments (with less than five pixels) were removed, holes were filled and contiguous segments were merged to form building footprints.

3.2.2 Hermite transform. The second method tested was based on the computation of a rotated discrete Hermite transform (DHT) from the height layer. A detailed description of this method will be provided in another article and, therefore, only a summary is provided here.

The rotated DHT coefficients, denoted as $z_{n,m}$, for $n, m = 1, \ldots, N$, correspond to scale–space derivatives of order n with respect to a variable varying along the local surface gradient, and of order m with respect to a variable varying in the direction orthogonal to the local gradient. The parameter N (which defines the length of the discrete approximation of Gaussian derivative filters; Silván-Cárdenas and

Escalante-Ramírez 2006) was set to $N = 4$. At each pixel location of the height layer, a residual energy was defined from a subset of DHT coefficients as:

$$E^2 = \sum_{i=2}^{N} \sum_{j=1}^{i} z_{i-j,j}^2. \tag{3}$$

This residual energy measures the degree to which the local surface (within a neighbourhood of 5 x 5 cells) does not conform to a one-dimensional signal embedded in three-dimensional space. More specifically, since the scale–space derivatives are sensitive to polynomial variations of the order of the derivative or superior, then the energy of the coefficients $z_{i-j,j}$, with $i \geq 2$, provides a measure of deviation from the planar surface, such as roofs with planar surfaces. On the other hand, the use of coefficients of order $j \geq 1$ account for deviation from one-dimensional patterns, such as linear edges formed by building walls. Thus, E is typically greater for trees than for buildings because canopy heights are neither uniformly distributed along one preferred direction nor planar. Therefore, a building mask was computed in which building pixels were assumed wherever a small E was found (i.e. $E < 0.1$).

3.2.3 Dempster–Shafer method. The third method was the Dempster–Shafer (DS) data-fusion technique introduced by Lu *et al.* (2006b) for building detection. This method is based on the Dempster–Shafer theory (DST) of evidence (Shafer 1976). The DST uses a belief (or mass) function and plausible reasoning to combine separate pieces of information (evidence) to calculate the probability of a proposition or event (such as 'the patch is building' or 'the patch is non-building').

Let S denote a set of basic propositions and 2^S its power set, including the empty set Φ. The theory of evidence assigns a belief mass to each element of the power set. Formally, a function $m : 2^X \mapsto [0, 1]$ is a belief or mass function if it satisfied two conditions:

$$\text{(i)} \qquad m(\Phi) = 0, \tag{4}$$

$$\text{(ii)} \sum\nolimits_{A \in 2^S} m(A) = 1. \tag{5}$$

For each member of the power set, the DST provides representation of both imprecision and uncertainty through the definition of two measures called support (Sup) and plausibility (Pls). These measures are defined as follows:

$$\text{Sup}(A) = \sum\nolimits_{B \in A} m(B), \tag{6}$$

$$\text{Pls}(A) = \sum\nolimits_{B \cap A \neq \Phi} m(B). \tag{7}$$

For two independent mass assignments m_1 and m_2, a joint mass assignment m is defined using the Dempster's rule of combination:

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - \sum_{B \cap C \neq \Phi} m_1(B) m_2(C)}. \tag{8}$$

This rule, and its generalization to multiple mass assignments (see Lu *et al.* (2006b), for a general formula), is the key for combining multiple sources of evidence.

The DS method, as applied in this study, combined three sources of evidence and, hence, used three mass assignments: (1) from a mask of preliminary buildings detected

Table 4. Calculation of probability masses, support and plausibility for Dempster–Shafer data fusion for three data sources. The proposition A take values in the power set of basic propositions or classes, whereas the classes are: building (C_1), vegetation (C_2) and ground (C_3). Mass assignments from each source are denoted by m_1 for initial building mask, m_2 for vegetation mask and m_3 for the ground mask. The normalization factor $k = 4(q_1q_2 + q_1q_3 + q_2q_3 - 9q_1q_2q_3)$, where $q_i = 1 - b_i$ and b_i are belief of patch being of class C_i. See the text for details on calculation of belief parameters.

A	m_1	m_2	m_3	m	Sup(A)	Sup($\bar{A}$)
C_1	b_1	$\frac{1}{3}q_2$	$\frac{1}{3}q_2$	$\frac{1}{k}b_1q_2q_3$	$\frac{1}{k}b_1q_2q_3$	$\frac{4}{k}q_1(q_2 + q_3 - 2q_2q_3)$
C_2	$\frac{1}{3}q_1$	b_2	$\frac{1}{3}q_2$	$\frac{1}{k}q_1b_2q_3$	$\frac{1}{k}q_1b_2q_3$	$\frac{4}{k}q_2(q_1 + q_3 - 2q_1q_3)$
C_3	$\frac{1}{3}q_1$	$\frac{1}{3}q_2$	b_3	$\frac{1}{k}q_1q_2b_3$	$\frac{1}{k}q_1q_2b_3$	$\frac{4}{k}q_3(q_1 + q_2 - 2q_1q_2)$
$C_1 \cup C_2$	0	0	$\frac{1}{3}q_2$	0	$\frac{4}{k}q_3(q_1 + q_2 - 2q_1q_2)$	$\frac{1}{k}q_1q_2b_3$
$C_1 \cup C_3$	0	$\frac{1}{3}q_2$	0	0	$\frac{4}{k}q_2(q_1 + q_3 - 2q_1q_3)$	$\frac{1}{k}q_1b_2q_3$
$C_2 \cup C_3$	$\frac{1}{3}q_1$	0	0	0	$\frac{4}{k}q_1(q_2 + q_3 - 2q_2q_3)$	$\frac{1}{k}b_1q_2q_3$
$C_1 \cup C_2 \cup C_3$	0	0	0	0	1	0

though the DHT method above, (2) from a mask of pixels detected as vegetation with the help of the NDVI from the CIR photograph and (3) from a mask of pixels detected as ground during the filtering stage of the LiDAR data. Three basic propositions or classes were considered: building (C_1), vegetation (C_2) and ground (C_3). Hence, the power set included up to seven non-null propositions: C_1, C_2, C_3, $C_1 \cup C_2$, $C_1 \cup C_3$, $C_2 \cup C_3$ and $C_1 \cup C_2 \cup C_3$. Then, for each building in the initial building layer, the support of each proposition was calculated following Lu *et al.* (2006b). Table 4 summarizes the calculation of probability masses, support and plausibility for each proposition. These calculations are based on the belief parameters b_1, b_2 and b_3. The belief that a patch in the initial building layer is indeed a building patch was set to $b_1 = 0.33$. This relatively low value was chosen to allow detection only if there was a strong support from other data sources. The belief that a patch in the initial building layer is a vegetation patch (b_2) or a ground patch (b_3), were set to the proportion of vegetation and ground pixels in the bounding box of the building patch. These proportions were determined from the vegetation and ground masks, respectively. Finally, a building patch in the initial layer was retained if and only if the support for the proposition C_1 (building) was the largest and greater than the support to the proposition $\overline{C_1}$ (not building).

3.2.4 eCognition. The fourth building-detection method tested was based on Definiens' eCognition segmentation of LiDAR derived layers followed by a supervized object-level classification in MATLAB.

The segmentation was based on a stacked image containing the ground mask, the feature height and the angular second moment layers. The segmentation algorithm was set up with a scale parameter of 4, and with a weight of 0.8 on layer values and 0.2 on segment shape, with a 0.6 bias towards segment smoothness over 0.4 towards segment compactness. The layer weights assigned for the segmentation process were 5 for ground mask, 4 for angular second-moment and 2 for feature height. The segmented image was exported from Definiens software, along with the mean and standard deviation for each of the layers used for segmentation plus the intensity-difference layer.

The segments were further classified into building and non-building using a back-propagation neural network (BPNN) with two hidden layers (of 25 and 15 neurons) from the MATLAB Neural Network Toolbox (v5.0.1). The network inputs consisted of per-segment mean and standard deviation of feature height, angular second moment, intensity difference and ground mask, whereas the output neurons consisted of two activation values in the interval 0–1, one for building and one for non-building. As such, all neurons had a logarithmic sigmoid function (logsig). The network was trained using the resilient back-propagation algorithm available in the toolbox. This method required a training sample, which consisted of pairs of segment properties and segment class (building, non-building). Ten percent of the total segments were randomly selected for this purpose. The reference building/non-building labels used for training were assigned to each segment based on a spatial query of whether or not the segment's centroid occurred within the boundaries of a building polygon (this was found to provide the best fit to building shape and avoid edge-match errors incurred when intersection or partial-overlap methods were used).

3.3 *Land-use classification*

The ability for automatic discrimination between residential and non-residential areas, and among the various dwelling types, within the urban environment is what makes remote sensing appealing for population estimation. Such a discrimination task is known as land-use classification, and is considered a critical component for accurate small-area population estimation. In order to test the impact that land-use accuracy had on the final population estimation, we selected three classification methods that represented three accuracy levels: low accuracy (overall accuracy less than 70%), moderate accuracy (overall accuracy between 70% and 80%) and high accuracy (overall accuracy higher than 80%).

3.3.1 Clustering-TM pixel. The low classification accuracy was yielded by a per-pixel supervized clustering approach (Silván-Cárdenas 2003). This method uses ISODATA to estimate centroids from a training sample. Then, it calculates a Delaunay tessellation from the centroids. The tessellation is used to build a neural network that performs a nearest-neighbour classification in which each class is defined through a number of centroids. This method was applied to six Landsat TM bands to classify pixels into single-family residential (SF), multi-family residential (MF) and other land-use types (OTs). This method is hereafter referred to as the clustering-TM pixel.

3.3.2 Neuralnet-building. The moderate classification accuracy was yielded by a BPNN trained using 24 building attributes (table 2). These attributes consisted of building geometry (area, perimeter, height, shape and volume) and built-up density (area of Thiessen polygon, footprint area to Thiessen polygon area ratio and impervious-surface percentage to Thiessen polygon area), which were computed on a per-building and on a per-neighbourhood basis (average and standard-deviation statistics). The 24 attributes assigned to each building footprint were imported into MATLAB, and a random sample was drawn to train a BPNN with 24 inputs, two hidden layers with 35 and 25 neurons, and nine output neurons. As such, the output vectors for each class were assigned values (1,0,0,0,0,0,0,0,0), (0,1,0,0,0,0,0,0,0), etc. Thus, in a classification mode, the class was determined according to the highest activation of the output neurons. The BPNN was trained with the resilient back-propagation algorithm, which ran for 500 iterations.

In this case, the classification was performed at the 100 level of the City of Austin's land-use code, which resulted in nine separate classes. However, the non-residential land-use classes were clumped into one class in the final classification result. This classification method is hereafter referred to as the neuralnet-building method.

3.3.3 Multi-class-tax parcel. The high classification accuracy was the hardest to attain. After an extensive test of classification methods implemented in WEKA software, the multi-class method in the classifier catalogue called Meta was selected. This Meta classifier transforms a multi-class problem into several binary problems (Ichino 1979).

The multi-class method was applied to classify tax parcels in nine classes, which were subsequently combined to form three classes, as with the previous method. The use of tax parcel as the classification unit is justified because it is the legal land-use unit, its limits seldom change and a geographic information system (GIS) layer is readily available. The attributes used in this case (see table 3) were derived from parcel boundaries (area, perimeter and shape), neighbour parcels (distance to nearest parcel, similarity of area between parcel and the nearest parcel, similarity of perimeter between parcel and the nearest parcel, and similarity of shape between parcel and the nearest parcel, where similarity was defined as a normalized absolute difference), NDVI zonal statistics (average and standard deviation), vegetation mask (percent of vegetation cover in parcel), land-cover proportions (percent of impervious surface, percent of bare ground, percent of other land-cover type), building-footprint layer (number of buildings in parcel, fraction of building cover in parcel, average, minimum and maximum statistics of area, perimeter, shape, volume and height of buildings within the parcel), neighbour buildings (average, minimum and maximum of distance to nearest neighbours) and building density. This method is hereafter referred to as the multi-class-tax-parcel method.

3.3.4 Training samples. Table 5 summarizes the training sample used for each land-use-classification method. For the clustering-TM pixel method, the training sample consisted of 950 randomly selected pixels, which yielded 50 centroids (16 for SF, 16 for MF and 18 for OTs). Likewise, the neuralnet-building method was trained using a random sample of 100 buildings per class, 900 in total. For the multi-class-tax-parcel method, the training sample consisted of around 40% of parcels from each land-use class, with the exception of the single-family class, for which only 20% were selected due to its larger proportion in the study area (see table 5 for sample sizes).

3.4 *Population-estimation models*

Seven linear models of population estimations were tested in this study, with the explanatory variables used being the fundamental difference among them. It is through the explanatory variables that the models incorporated information about building structure and land-use type.

3.4.1 Model characteristics. Table 6 summarizes the building statistics and land-use information required by each model. The first six models are generated by combining three building statistics at block level (building count, footprint area and total volume) with two levels of land-use information (residential versus SF and MF). Thus, model 1 uses the per-block counts of residential buildings (N) regardless of

Table 5. Characteristics of training sample for each land-use-classification method tested. All sample sets were randomly selected

Method	Sample attributes[a]	Sample size
Clustering-TM pixel	Landsat pixels (6) Land-use class (3)	950 pixels: 300 single-family, 300 multi-family and 350 other
Neuralnet-building	Building geometry (5) Built-up density (3) Neighbourhood (16) Land-use class (9)	900 buildings: 100 buildings per class
Multi-class-tax parcel	Parcel geometry (3) Neighbour parcels (4) Vegetation (3) Impervious surface (3) Building (17) Neighbour building (4) Land-use class (9)	3451 parcels: 1600 single-family[b], 240 multi-family, 470 commercial, 260 office, 212 industrial, 209 civic, 58 open space, 133 transportation and 269 undeveloped

[a]Number of attributes for each category is in parenthesis
[b]Proportion of sample was 20% for single-family, and 40% for other classes

Table 6. Population-estimation models tested. Per-block building statistics and land-use information required by each model are also provided

Name	Equation[a]	Building statistics	Land use
Model 1	$P_1 = \alpha_1 N + \varepsilon$	Building count	Residential
Model 2	$P_2 = \alpha_1 N_1 + \alpha_2 N_2 + \varepsilon$	Building count	SF and MF
Model 3	$P_3 = \alpha_1 A + \varepsilon$	Footprint area	Residential
Model 4	$P_4 = \alpha_1 A_1 + \alpha_2 A_2 + \varepsilon$	Footprint area	SF and MF
Model 5	$P_5 = \alpha_1 V + \varepsilon$	Total volume	Residential
Model 6	$P_6 = \alpha_1 V_1 + \alpha_2 V_2 + \varepsilon$	Total volume	SF and MF
Model 7[b]	$P_7 = \alpha_1 A_1 + \alpha_2 N_2 + \alpha_3 V_2 + \varepsilon$	Building count, footprint area and total volume	SF and MF

[a]α parameters are to be estimated by minimizing the squared residuals, ε^2
[b]A variable selection strategy was applied for this model (see the text)

whether it is SF or MF, whereas model 2 uses the split of the count of residential building into the count of single-family buildings (N_1) and the count of multi-family buildings (N_2). These two models are inspired in the HU method (Watkins and Morrow 1985, Smith and Cody 2004), where the number of buildings replaces the number of housing units. Likewise, models 3 and 4 are inspired in the broadly used area-based methods, but with a linear form. We used the linear form of the area-population relationship because a preliminary exploratory analysis confirmed that the linear form fitted the data better than the allometric form. Models 4 and 5 were proposed under the premise that building volume can better describe the living space, and thus may allow for more accurate population estimates.

In addition to the six models above, an optimal linear model (model 7) was constructed by selecting a few explanatory variables out of 16 variables. The set of initial variables included building count, area, volume, perimeter, shape and height, from both single-family and multi-family buildings. The variable selection strategy sought to include the lowest number of explanatory variables while maintaining a high correlation with the dependent variable.

3.4.2 Construction of model 7. The procedure for building model 7 was based in the part-correlation statistics (Selvin 1995). The squared part correlation $C^2_{(i)}$ is equal to the increase in the squared multiple-correlation coefficient achieved by adding a specific variable (say x_i) to the regression equation, i.e. $C^2_{(i)} = R^2_{x_i-\text{included}} - R^2_{x_i-\text{excluded}}$. The procedure for the generation of model 7 can be summarized as follows:

(i) build the regression line considering all 16 explanatory variables,

(ii) compute the part correlation for each variable through:

$$C^2_{(i)} = \frac{t_i(1 - R^2)}{n - (k + 1)}, \tag{9}$$

where t_i is the t-statistics for the ith variable, R^2 is the coefficient of determination from the regression, n is the number of samples used in the regression and k is the number of explanatory variables or degrees of freedom and

(iii) make the variable with the least part correlation unavailable for further analysis and repeat from step 1 until the lowest part correlation is greater than a preset threshold.

3.4.3 Parameter estimation. The coefficients for each of the seven models were estimated through the ordinary least-squares procedure (Selvin 1995). We also calculated normalized coefficients or path coefficients, which correspond to the regression coefficients multiplied by the standard deviation of the explanatory variable divided by the standard deviation of the dependent variable. The normalized forms are useful for inter-comparisons as they represent the sensitivity of the dependent variable to the variation of the independent variable (Selvin 1995).

3.5 *Methods of accuracy assessment*

Methods of accuracy assessment include the accuracy assessment of building detection, land-use classification and population estimations.

3.5.1 Statistics. Building-detection accuracy assessment was performed at both pixel and object level. At the pixel level, we derived the overall accuracy (percent of correctly classified pixels) and the kappa statistics, both from the standardized confusion matrix (Congalton and Green 1998). At the object level, we calculated the detection rate, i.e. the percentage of correctly detected buildings to total number of reference buildings, and the commission error, i.e. percentage of false detections to total number of detected buildings.

For the accuracy assessment of land-use classifications, both the reference and the extracted land-use labels were first attached to each building footprint. Then, a confusion matrix was built by comparing the reference and extracted land-use labels.

Finally, we calculated the overall and per-class accuracy ratios and the kappa coefficient of agreement from the confusion matrix.

The goodness-of-fit of population-estimation models was assessed through the coefficient of determination (R^2), whereas the model validation was based on statistics derived from the absolute difference between census and estimated populations, i.e. absolute error. The statistics included the mean, standard deviation, median (50 percentile), maximum (100 percentile), lower quartile (25 percentile), upper quartile (75 percentile) and interquartile range (the difference between upper and lower quartiles) of the absolute error. Among these error measures, the median absolute error (MAE) and the interquartile range (IQR) were more extensively considered as they are most common in population estimation studies. Relative errors were not considered due to sensitivity issues in areas of low population density, or even indetermination in non-populated areas.

3.5.2 Sampling. The accuracy assessment of building-detection methods was based on insets 1 through to 4. The sample sizes in term of pixels were 498 525, 374 519, 398 463 and 500 500 pixels for insets 1 through to 4, whereas in terms of objects, the sample sizes were 145, 476, 427 and 490 buildings for insets 1 through to 4, respectively. No random sample was drawn for accuracy assessment of classification methods. Instead, the assessment was based on the entire study area.

Random samples were also drawn from the total number of blocks in the study site (1153 blocks) for calibration (90 blocks) and validation (85 blocks) purposes. Table 7 summarizes the land-use characteristics and number of blocks used for calibration, validation and estimation. Note that while the estimation was based on the entire dataset, the sample sizes for validation and calibration were sufficiently large to be statistically significant, but not so large that it would not have been realistically collectible, i.e. in case census data were not available.

In addition to the above, an independent sample was selected for construction of model 7. This sample consisted of 150 randomly selected blocks, drawn from a total of 775 blocks containing at least one residential building. The relatively large sample size was necessary given the number of coefficients (16) to be estimated during the construction of the model. In order to ensure that the sample captures the variability in composition of dwelling units within the blocks, 60 blocks were selected from pure SF land use, 30 from pure MF land use and 60 from mixed SF/MF land use.

4. Results

4.1 *Building-detection results*

The results from each building-detection method are illustrated in figure 4. These error maps were built through comparing the detection mask from each method with the reference-building-footprint layer in raster format. Errors of omission and commission are coloured in blue and red for easy identification. In addition, the overall per-pixel accuracy, the kappa statistics, the detection rate and the commission error were calculated for each method and inset. Average statistics over all insets are provided in table 8. It was observed that although most methods had acceptable performance, with average overall accuracy ranging from 87 to 92% and detection rate ranging from 84 and 90%, there was considerable variability across insets. Specifically, inset 1 was the most accurately classified into building and non-building

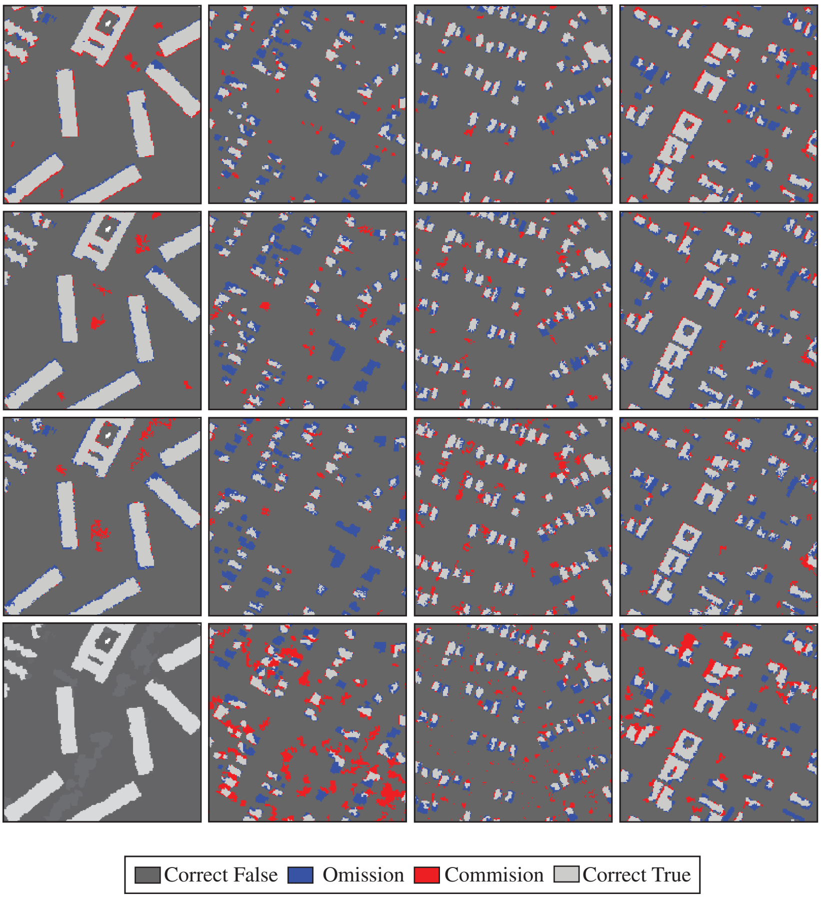

Figure 4. Error maps from each building-detection method and inset. Rows from top to bottom correspond to Dempster–Shafer, region growing, Hermite transform and eCognition, whereas columns from left to right correspond to subsets from inset 1 through to inset 4, respectively.

by any method, with a per-pixel accuracy ranging from 92 to 95%. This was due to the relatively high and large structure of multi-family buildings. On the other hand, inset 2 represented the most challenging area due to the relatively small size of single-family buildings and the high chance of occlusions by trees. In this case, per-pixel accuracies ranged from 83 to 89%. In between these two extreme were insets 3 and 4.

Based on the accuracy levels and its consistency over the various insets, the DS method was selected for detecting buildings over the entire study area. The accuracy of the detected buildings for the entire study area was not assessed, but it is reasonable to assume a detection rate of around 90%, as this was the average performance from selected insets (table 8). The building mask for the entire study area was imported into ArcGIS and converted to vector format (yielding 15211 building polygons) for further analysis.

Table 7. Number and land-use characteristics of census blocks selected for model calibration and validation, and for population estimation.

Land-use mixture	Calibration	Validation	Estimation	Percent
Pure single-family (SF)	30	30	594	52%
Pure multi-family (MF)	30	25	55	5%
Mixed SF/MF	30	30	126	11%
Non-residential	0	0	378	33%
Total	90	85	1153	100%

4.2 *Land-use classification results*

The three selected land-use classification methods (multi-class-tax parcel, neuralnet-building, clustering-TM pixel) were used to produce three land-use layers of different accuracy levels. For building-driven land-cover classification methods (multi-class-tax parcel and neuralnet-building), training samples were based on the reference-building-footprint layer. Figure 5 shows classification results from each classification method and from reference land-use layer (top row) for portions of insets 2–4. Note that these maps are in their native classification unit.

The land-use information, from both the reference and classifications, was attached to the extracted buildings (overlaid on figure 5) for population estimation and accuracy assessment purpose. In this step, the class with the largest overlap was assigned to the building. Although each land use classification result was produced in different spatial units, namely tax-parcel, building footprint and Landsat pixel, the accuracy assessment was performed on a per building basis. For this matter, detected building footprints were considered in order to account for detection errors. Table 9 summarizes various accuracy measures for each classification method.

The land-use layers extracted from each method, as well as the reference land-use layer, were used to mask out non-residential buildings from the detected building layer. This process made available up to four layers of residential buildings, which were used with each population-estimation model.

4.3 *Population-estimation results*

4.3.1 Construction of model 7. The construction of model 7 followed the procedure outlined in the previous section using a threshold of 0.1 for the part correlation. The initial set of explanatory variables consisted of the following block-level statistics: building count, total building-footprint area, total building volume, maximum building volume, average building height, maximum building height, average building-footprint perimeter and average building-footprint shape, for both single-family and multi-family dwelling types. The variable selection procedure yielded a model of the form:

$$P_7 = \alpha_1 A_1 + \alpha_2 N_2 + \alpha_3 V_2 + \varepsilon, \tag{10}$$

where A_1 is the area of single-family dwellings, N_2 and V_2 are count and volume of multi-family dwellings, ε is the residual not explained by the linear model, and α values are model parameters.

Interestingly, the building perimeter, building shape and building height did not provide further explanation to the variability of block population. Equation (10) also

Figure 5. Land-use classification result. Rows from top to bottom correspond to reference land use, multi-class-tax parcel, neuralnet-building and clustering-TM pixel, whereas columns from left to right correspond to subsets from insets 2 through to 4, respectively.

Table 8. Two-level accuracy assessment of building detection. Values denote average statistics over four insets with 498 525, 374 519, 398 463 and 500 500 pixels, and 145, 476, 427 and 490 buildings, respectively. Values were rounded to the nearest integer.

	% pixels		% objects	
Method	Overall accuracy	Kappa	Detection rate	Commission
Dempster–Shafer	92	66	90	30
Region growing	91	64	83	19
Hermite transform	91	62	85	21
eCognition	87	55	84	43

Table 9. Building-level accuracy assessment of land-use classification methods. Statistics were based on a total of 15 211 extracted buildings in the study site. Entries under columns SF, MF and res correspond to per-class producer accuracies of classes single-family, multi-family and residential, respectively. Columns labelled as OA and kappa correspond to the overall accuracy and the kappa coefficient of agreement. Values were rounded to the nearest integer.

	% Correct				
Method	SF	MF	Res	OA	Kappa
Multi-class-tax parcel	91	55	85	83	71
Neuralnet-building	93	17	80	73	48
Clustering-TM pixel	75	67	73	66	47

tells us that the volume information from single-family buildings is not as important as its area and suggests that the contribution from multi-family units depends both on the number and the volume.

4.3.2 Calibration of models. The estimated coefficients for all the models are provided in table 10, together with the normalized form ($\tilde{\alpha}$). The values for the normalized coefficients suggest that population estimates are more sensitive to multi-family than to single-family building characteristics.

The bar chart of figure 6 compares the goodness-of-fit in terms of R^2 values obtained from each model. As it turns out, building-volume information, at both land-use levels, fits better than building area and building counts. It is also noticeable

Table 10. Estimated coefficients for each model defined in table 6. The path coefficients are the regression coefficients multiplied by the standard deviation of the explanatory variable and divided by the standard deviation of the dependent variable.

	Regression coefficients			Path coefficients		
Model	α_1	α_2	α_3	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	$\tilde{\alpha}_3$
1	3.4335	–	–	0.65	–	–
2	1.8129	11.8573	–	0.35	0.75	–
3	0.0303	–	–	0.76	–	–
4	0.0153	0.0416	–	0.32	0.87	–
5	0.0036	–	–	0.81	–	–
6	0.0025	0.0039	–	0.04	0.89	–
7	0.0153	4.7673	0.0031	0.32	0.30	0.72

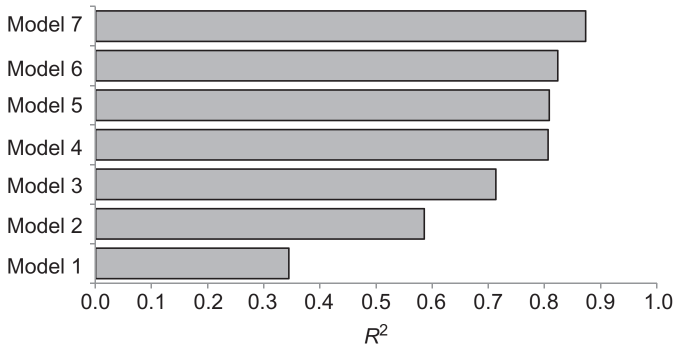

Figure 6. Model-calibration results.

that all models that distinguish between SF and MF dwelling types yielded superior goodness-of-fit with respect to their counterpart that did not made this distinction. This is largely due to the fact that such models have greater degrees of freedom and thus are more flexible. Interestingly, the improvement achieved when making the distinction of dwelling types was most notable for building counts (from $R^2 = 0.34$ to $R^2 = 0.59$), and building area (from $R^2 = 0.71$ to $R^2 = 0.81$) than for building volume (from $R^2 = 0.81$ to $R^2 = 0.82$). This result suggests that building volume already accounts for the dwelling-type characteristic. The observation is important because, unlike area-based or count-based models, a volume-based model would not rely on detailed land-use information (at least not to the level of distinguishing between dwelling types), which is generally difficult to extract by automated means.

4.3.3 Validation of models. In order to confirm the fitting trends observed during the calibration stage, a validation sample (table 7) was used to test each model. The absolute-error statistics are summarized in table 11. The various statistics reveal the form of the error distribution in terms of central values (mean and median), dispersion (standard deviation, maximum and interquartile range (IQR)) and non-central values (25 and 75 percentile). The R^2 and MAE with interquartile ranges are plotted in figure 7 for a visual comparison. Contrary to expectation, models 2 and 4 performed comparably to, or even better than, models that incorporate volume information (models 5–7). This may be due to the fact that trees hanging above the buildings, especially for single-family residential buildings, cast significant errors on the estimation of building height and volume. The inaccurate height information negatively impacted the population estimations from

Table 11. Model validation based on statistics from absolute error. Values were rounded to the nearest integer. The minimum value for each column is indicated in boldface.

Model	Mean	Standard deviation	Median	Maximum	25 percentile	75 percentile	IQR
1	26	76	26	421	10	73	62
2	17	56	17	272	8	41	33
3	30	46	30	**224**	14	59	45
4	15	**43**	15	239	6	**33**	**27**
5	19	58	19	354	8	46	38
6	**14**	60	**14**	388	**6**	36	31
7	16	51	16	320	8	45	38

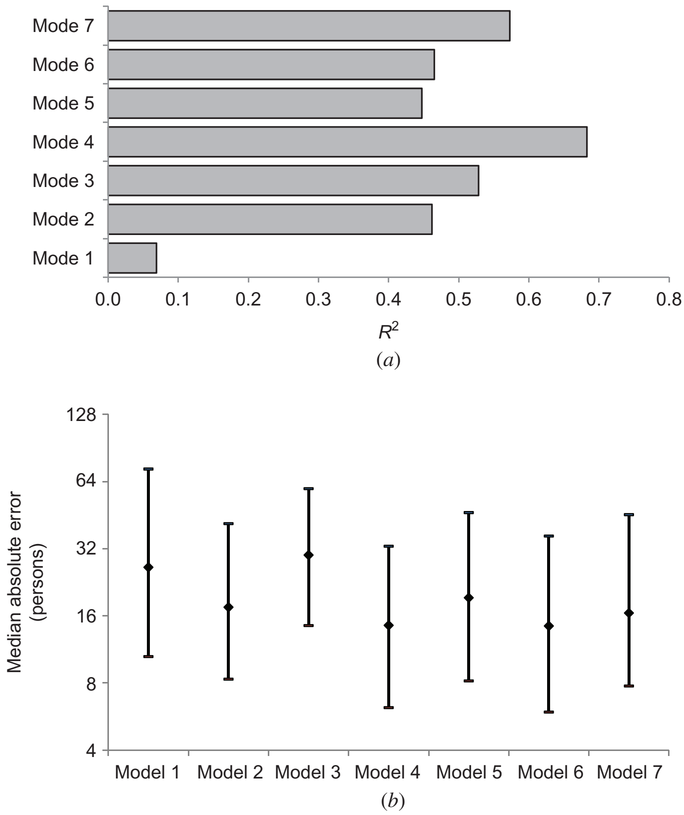

Figure 7. Model-validation results. (*a*) R^2 statistics and (*b*) median absolute error, with interquartile range (vertical bars) based on a validation sample. Note the vertical axis in (*b*) is on a logarithmic scale.

building volumes. Presumably, this also explains why model 7 was no longer the best-performing model. Nevertheless, the differential improvement by the incorporation of the finest land-use information seemed to hold in the validation sample (compare model 1 with model 2, model 3 with model 4 and model 5 with model 6).

4.3.4 Population estimation and accuracy assessment. The calibrated models were applied to the entire study area using the extracted buildings combined with the various land-use layers of varying accuracy. The goal here was to determine the effect of inaccuracies from building detection and land-use classification.

In a first test, the detected buildings layer replaced the reference-building-footprint layer, while the reference land-use layer was still used for assigning the type of dwelling. Then, geometric attributes were calculated from the extracted buildings and aggregated at census-block level. Figure 8 shows the scatter plots of estimated versus measured building attributes at census-block level. These scatter plots evidence that detected buildings generally yielded lower values than those derived from the reference-building-footprint layer. This trend was the strongest for blocks dominated by single-family dwelling units. Because of this, population counts were also underestimated from all the models. The sensitivity of each model to the underestimation

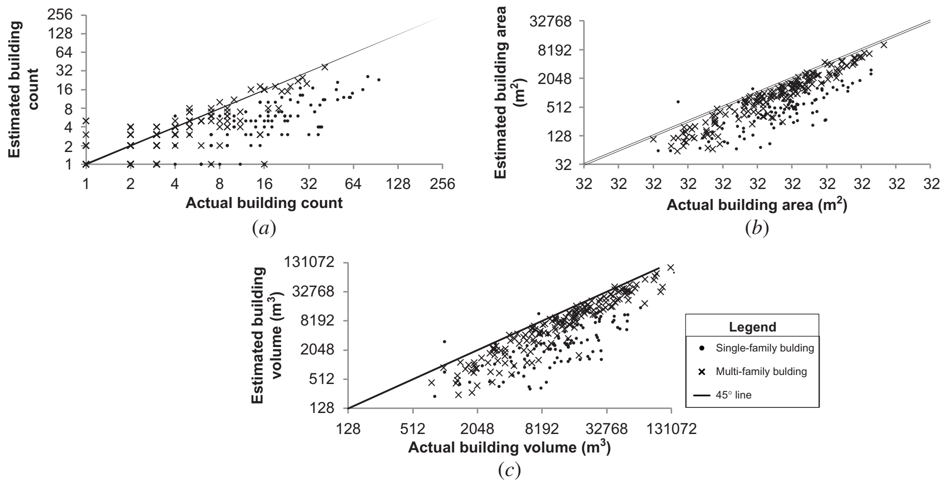

Figure 8. Scatter plots of estimated versus reference data at the aggregation level of census block: (*a*) building count, (*b*) building area and (*c*) building volume. Axes are on a logarithmic scale.

errors of explanatory variables was then assessed through the MAE of population estimates.

Figure 9 compares the MAE yielded by each model when using the reference buildings and when using the extracted buildings. The two series are negatively correlated ($R^2 = 0.86$, $n = 7$, $p < 0.05$), showing that the trend observed in the model fitting stage (figure 6) is inverted due to building-detection errors alone. In other words, although building volume fits block population better than building area, and building area fits better than building count, it appears that building count is less sensitive to building-detection errors than building area, and building area is less sensitive than building volume. Since both building area and building-volume measurements heavily depend on LiDAR measurements, they are more severely affected by detection and measurement errors. This is the case especially for single-family residential buildings located in densely vegetated areas. Moreover, the models that distinguish between dwelling types (models 2, 4 and 6) tend to have larger errors than their counterparts that consider residential buildings altogether (models 1, 3 and 5). This result suggests that the best strategy is to use the minimal amount of information: residential-building counts.

In a second test, the extracted land-use information was incorporated in addition to the extracted buildings. The increments of MAE due to land-use inaccuracies, for each land-use classification method, are plotted in figure 10. As expected, the accuracy of population estimation can be seriously affected by the accuracy of land use, especially when using very sensitive explanatory variables such as area and volume. Interestingly, model 1 showed the least sensitivity to land-use inaccuracies. Even when the least accurate (overall accuracy = 66%, $\kappa = 0.47$) land-use data was used, the increase of the MAE was less than two people per block for model 1. This value was much lower than that obtained for the volume-based models, even when the most accurate (overall accuracy = 83%, $\kappa = 0.71$) land-use data was used. The same trend was observed for the mean and maximum absolute error (data not shown).

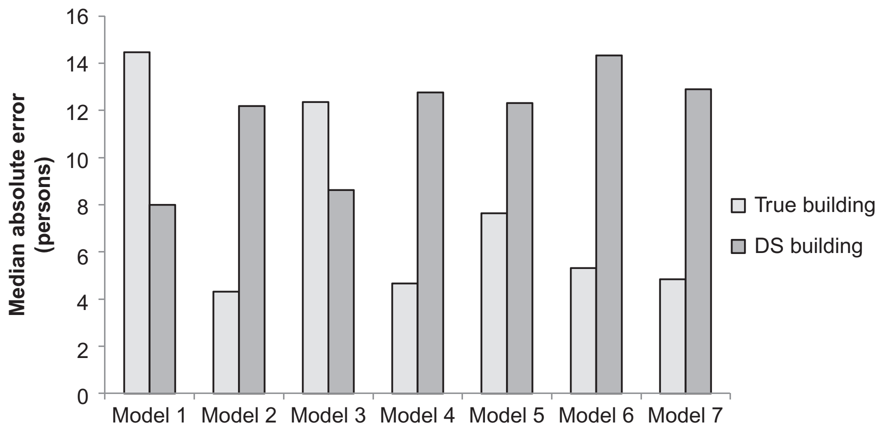

Figure 9. Effect of building-detection accuracy on population estimation. This chart shows the median absolute error from the reference buildings (True Bldg) and from the extracted buildings (DS Bldg). In the two cases, the reference land-use layer was used to discriminate residential from non-residential buildings. The extracted buildings were produced using the Dempster–Shafer method, with an accuracy of around 90%.

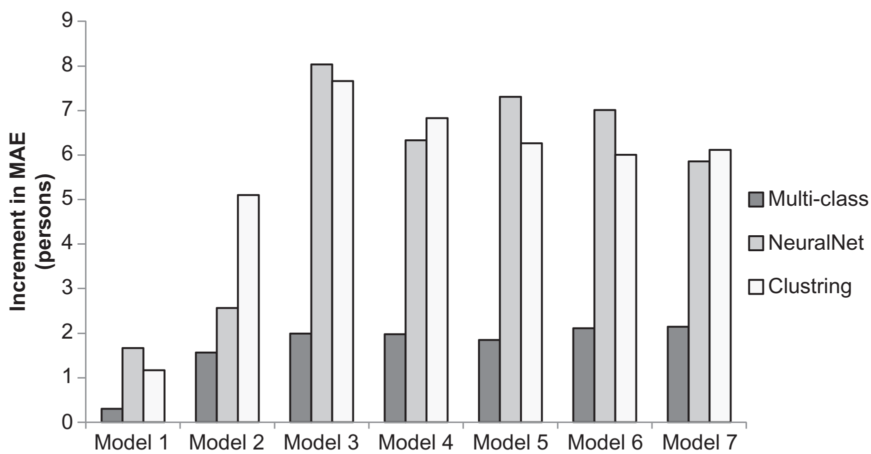

Figure 10. Effect of land-use accuracy on population estimation. This chart shows increments in the median absolute errors (MAEs) yielded by each model when using land-use information extracted from three classification methods of varying accuracy. Refer to table 9 for land-use classification accuracy.

5. Discussion and conclusions

Small-area population estimation is an important task that has received considerable attention by the remote-sensing community in the past four decades (Hsu 1971, Kraus *et al.* 1974, Lo and Welch 1977, Lo and Chan 1980, Watkins and Morrow 1985, Lo 1986, Harvey 2002b, Li and Weng 2005, Wu *et al.* 2005, Liu *et al.* 2006, Lu *et al.* 2006a, Wu and Murray 2007, Hardin 2008, Morton and Yuan 2009). The wealth of related studies reveals that the notion of living space had been considered a key linkage between population and remote-sensing measurements. Unfortunately, a formal definition for this important variable has proved difficult, due, in part, to the relatively coarse spatial resolution of the remote-sensing data used for population estimation. The advent of fine-spatial-resolution satellite images (1–5 m) coupled with LiDAR measurements opened new opportunities for considering the three-dimensional nature of living space in urban environments and for improving small-area population estimations.

In the study reported here, we tested the potential of fine-spatial-resolution LiDAR measurements coupled with automated and semi-automated techniques for building extraction and land-use classification. We compared seven linear models for small-area population estimations, each of which is expressed in terms of one, two or three explanatory variables representing building statistics on a per-block basis (count, area and volume) at one of two land-use classification levels (residential/non-residential versus single-family residential/multi-family residential/non-residential). These explanatory variables were meant to more closely represent the living space because the great majority of population lives inside buildings. Interestingly, when considering other geometric characteristics of buildings, such as perimeter, shape and height, their contribution to the regression was negligible (model 7).

At the model-fitting stage, it was observed that the incorporation of either the detailed land-use information or the volume information led to higher correlation coefficients. At a validation stage, however, the differential improvement by building volume appeared not as important as that of using detailed land-use information. Presumably, this was due to errors introduced by the method used for calculating the height information from LiDAR data. At the estimation stage, we first replaced the reference buildings by the detected buildings and tested the effect on the population estimation errors. The original trend observed during the fitting state was totally inverted. The sensitivity of each model to the errors in the land-use information further favoured the simplest model based on counts of residential buildings. While the reason for this inversion appeared to be the violation of the fundamental assumption that the high-quality calibration sample was representative of the remote-sensing-derived buildings, it was also apparent that the most important building characteristics for population estimation, namely dwelling type and volume, were also the most difficult to accurately extract from automated means.

Overall, the results suggested that the incorporation of detailed land use or building volume (or even building area) does not necessarily imply an improvement over the traditional approach that relies on unit counts. Hence, for fine-spatial-resolution remote sensing to play a significant role in improving small-area population estimation, building extraction and classification methods must be improved. Improvements should not be focused on increasing the detection accuracy alone, but also on reducing the bias of estimated building attributes. Unbiased estimations of building attributes may improve population estimates, as errors tend to cancel out during the aggregation at the spatial unit of the census.

Acknowledgements

This study was supported by grants to Le Wang from the National Science Foundation (BCS-0822489, DEB-0810933), and from National Key Basic Research and Development Program, China (2006CB701304). The final form of the manuscript was achieved thanks to comments from two anonymous reviewers and the Editor.

References

Benn, H., 1995, TCRP Synthesis of transit practice 10: bus route evaluation standards. Technical report, Transportation Research Board, National Research Council, Washington, DC.

City of Austin, 2009a, City of Austin demographics. Available online at: http://www.ci.austin.tx.us/demographics/ (accessed 12 May 2009).

City of Austin, 2009b, City of Austin GIS data sets. Available online at: ftp://coageoid01.ci.austin.tx.us/GIS-Data/Regional/coa_gis.html (accessed 08 May 2009).

Congalton, R. and Green, K., 1998, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices* (Boca Raton, FL: Lewis).

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

Hardin, P., 2008, Modelling housing unit density from land cover metrics: a midwestern US example. *Geocarto International*, **23**, pp. 393–411.

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

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

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

Huxley, J., 1932, *Problems of Relative Growth* (London, UK: Metheuen).

Ichino, M., 1979, A nonparametric multiclass pattern classifier. *IEEE Transactions on Systems, Man and Cybernetics*, **9**, pp. 345–352.

Iisaka, J. and Hegedus, E., 1982, Population estimation from Landsat imagery. *Remote Sensing of Environment*, **12**, pp. 259–72.

Kraus, S., Senger, L. and Ryerson, J., 1974, Estimating population from photographically determined residential land use types. *Remote Sensing of Environment*, **3**, pp. 35–42.

Langford, M., 2006, Obtaining population estimates in non-census reporting zones: an evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems*, **30**, pp. 161–180.

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.

Liu, X., Clarke, K. and Herold, M., 2006, Population density and image texture: a comparison study. *Photogrammetric Engineering and Remote Sensing*, **72**, pp. 187.

Liu, X., 2003, Estimation of the spatial distribution of urban population using high spatial resolution satellite imagery. PhD thesis, University of California, Santa Barbara, CA.

Lo, C., 1986, Accuracy of population estimation from medium-scale aerial photography. In *Proceedings of the American Congress on Surveying and Mapping and American Society for Photogrammetry and Remote Sensing, Annual Convention,* Washington, DC, pp. 1–10.

Lo, C., 1989, A raster approach to population estimation using high-altitude aerial and space photographs. *Remote Sensing of Environment*, **27**, pp. 59–71.

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

Lo, C., 2003, Zone-based estimation of population and housing units from satellite-generated land use/land cover maps. In *Remotely Sensed Cities*, p. 157 (London, UK: Taylor and Francis).

Lo, C., 2006, Estimating population and census data. In *Remote Sensing of Human Settlements, Manual of Remote Sensing* (3rd edn), Vol. 5, pp. 337–377 (Bethesda, MD: American Society for Photogrammetry and Remote Sensing).

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

Lo, C. and Chan, H., 1980, Rural population estimation from aerial photographs Hong Kong. *Photogrammetric Engineering and Remote Sensing*, **46**, pp. 337–345.

Lo, C. and Welch, R., 1977, Chinese urban population estimates. *Annals of the Association of American Geographers*, **67**, pp. 246–253.

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

Lu, Y., Trinder, J. and Kubik, K., 2006b, Automatic building detection using the Dempster–Shafer algorithm. *Photogrammetric Engineering and Remote Sensing*, **72**, pp. 395–403.

Martin, D. and Williams, H., 1992, Market-area analysis and accessibility to primary health-care centres. *Environment and Planning A*, **24**, pp. 1009–1019.

Morton, T. and Yuan, F., 2009, Analysis of population dynamics using satellite remote sensing and US census data. *Geocarto International*, **24**, pp. 143–163.

Nordbeck, S., 1965, The law of allometric growth. *Michigan Inter-University Community of Mathematical Geographers*, Discussion Paper 7.

Nordbeck, S., 1971, Urban allometric growth. *Geografiska Annaler, Series B, Human Geography*, **53**, pp. 54–67.

Openshaw, S. and Taylor, P., 1981, The modifiable areal unit problem. In *Quantitative Geography: a British View*, N. Wrigley and R. Bennett (eds.), pp. 60–69 (London, UK: Routledge and Kegan Paul).

Plane, D. and Rogerson, P., 1994, *The Geographical Analysis of Population: With Applications to Planning and Business* (New York, NY: Wiley & Sons).

Porter, P., 1956, Population distribution and land use in Liberia. PhD thesis, London School of Economics and Political Science, University of London, UK.

Selvin, S., 1995, *Practical Biostatistical Methods* (Belmont, CA: Duxbury Press Belmont).

Shafer, G., 1976, *A Mathematical Theory of Evidence* (Princeton, NJ: Princeton University Press).

Silván-Cárdenas, J. and Escalante-Ramírez, B., 2006, The multiscale Hermite transform for local orientation analysis. *IEEE Transactions on Image Processing*, **15**, pp. 1236–1253.

Silván-Cárdenas, J. and Wang, L., 2006, A multi-resolution approach for filtering LiDAR altimetry data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **61**, pp. 11–22.

Silván-Cárdenas, J.L., 2003, Optimal design of neural networks for land-cover classification from multispectral imagery. In *Proceedings of the SPIE, Conference on Image and Signal Processing for Remote Sensing IX*, L. Bruzzone (Ed.), February, pp. 420–431 (Barcelona, Spain: International Society for Optical Engineering).

Smith, S. and Cody, S., 2004, An evaluation of population estimates in Florida: April 1, 2000. *Population Research and Policy Review*, **23**, pp. 1–24.

Smith, S., Nogle, J. and Cody, S., 2002, A regression approach to estimating the average number of persons per household. *Demography*, **39**, pp. 697–712.

Sutton, P., Roberts, D., Elvidge, C. 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. 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.

Taragi, R., Bisht, K. and Sokhi, B., 1994, Generating population census data through aerial remote sensing. *Journal of the Indian Society of Remote Sensing*, **22**, pp. 131–138.

US Census Bureau, 2009a, American FactFinder. Available online at: http://factfinder.census.gov (accessed 08 May 2009).

US Census Bureau, 2009b, TIGER, TIGER/Line and TIGER-related products. Available online at: http://www.census.gov/geo/www/tiger (accessed 08 May 2009).

US Geological Survey, 2009, Global visualization viewer. Available online at: http://glovis.usgs.gov/ (accessed 30 March 2009).

Watkins, J. and Morrow, H., 1985, Small area population estimates using aerial photography. *Photogrammetric Engineering and Remote Sensing*, **51**, pp. 1933–1935.

Witten, I.H. and Frank, E., 2005, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd edn (San Francisco, SF: Morgan Kaufmann).

Wu, C. and Murray, A., 2007, Population estimation using Landsat Enhanced Thematic Mapper imagery. *Geographical Analysis*, **39**, pp. 26–43.

Wu, S., Wang, L. and Qiu, X., 2008, Incorporating GIS building data and census housing statistics for sub-block-level population estimation. *The Professional Geographer*, **60**, pp. 121–135.

Wu, S., Qiu, X. and Wang, L., 2005, Population estimation methods in GIS and remote sensing: a review. *GIScience and Remote Sensing*, **42**, pp. 80–96.

Yagoub, M., 2006, Application of remote sensing and geographic information systems (GIS) to population studies in the Gulf: a case of Al Ain City (UAE). *Journal of the Indian Society of Remote Sensing*, **34**, pp. 7–21.

Zhang, K., Yan, J. and Chen, S., 2006, Automatic construction of building footprints from airborne LIDAR data. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, pp. 2523–2533.