Incorporating GIS building data and census housing statistics for sub-block population estimation

Shuo-sheng “Derek” Wu

Department of Geography, Texas State University-San Marcos

601 University Drive

San Marcos, Texas 78666-4616

Phone: 512-586-2115.

Fax: 512-245-8353

E-mail: derekwu@txstate.edu

Advisor: Le Wang

Phone: 512-245-1333.

Fax: 512-245-8353

E-mail: lw15@txstate.edu

Incorporating GIS building data and census housing statistics for sub-block population estimation

Abstract: Past population estimation methods are usually designed for a scale larger than the census block because of the limitation of data resolution and/or the errors associated with population estimation for small areas. In addition, there is no efficient way to assess sub-block estimates without sub-block population data. This article presents a deterministic model for population estimation at the sub-block level using GIS building data and census housing statistics. Furthermore, a simulation approach is proposed to assess how well the model estimates sub-block population by aggregating every two to twenty blocks as the base units of modeling. In this way, the actual sub-unit populations, i.e. populations of individual census blocks or block aggregations from census data, can be compared with the estimates correspondingly. The results reveal that the deterministic model can estimate block-level population with mean absolute relative errors (percentages) ranging from 0.1 to 0.15. The assessment of sub-block estimates using a simulation approach indicates that the smaller the sub-unit areas, the higher the estimation errors, estimation errors for the same percent of different base unit areas do not relate to the base unit sizes, and the rescaling for volume-preservation improves all scales of sub-unit estimates but not in a systematic way. The assessment results also show that estimation errors are higher for single family land use areas than those for multi-family land use areas, and are higher for mixed-land-use areas than those for residential land use areas. The results of this study are important for those who need to estimate populations for small areas with arbitrary boundaries and/or need to assess estimation accuracies.

Key Words: dasymetric mapping, population estimation, population interpolation

Introduction

In the U.S., the finest census population data available to the public is at the block level. However, in practical applications, people may need to estimate population counts for areas not coinciding with boundaries of census blocks. For example, a subdivision for a proposed redevelopment project does not always have the same boundaries as those drawn for census blocks, but for cost and benefit analyses, developers may need to estimate how many people live in the subdivision. Watersheds, as well, usually do not share the same boundaries as those of census blocks, yet environmental scientists may need to estimate watershed populations for environmental impact assessment. Another example of is that city planners may need to estimate populations within a half-mile buffer of a proposed railroad route or landfill site to assess how many local residents would be affected by the projects. All these examples illustrate the demand for fine-scale population estimation methods.

This article presents a deterministic model for sub-block population estimation, based on building volumes from GIS data and housing statistics from Census 2000 block-level data. Furthermore, a base-unit simulation approach was used to assess how well the deterministic model computes sub-unit populations, how base unit sizes affect sub-unit estimates, how the rescaling for volume preservation affects sub-unit estimates, and how estimation errors vary with the land use of the estimation areas.

The results of this study are important for those who need to estimate populations for small areas with arbitrary boundaries and/or need to assess estimation accuracies. For example, city planners need to estimate how many residents live within floodplains for evacuation planning purposes. Researchers need sub-block population data to study urban sprawl in suburban areas where census blocks are usually large. Businessmen, too, need to model the distribution of potential customers at a fine spatial scale for store location analysis. And knowing the associated errors for the estimated populations is also critical for them for accuracy assessment.

Population Estimation in GIScience

According to Wu, Qiu, and Wang (2005), past studies relevant to population estimation in GIScience literatures can be grouped into two categories -- areal interpolation and statistical modeling -- depending on the intended goal and the required information. Areal interpolation studies use census population data as the input and apply certain interpolation techniques to obtain refined population data, usually for the purpose of transforming population data from one set of spatial unit to the other. In contrast, statistical modeling studies are interested in inferring a statistical relationship between population and other physical and socio-economic variables, usually for the purpose of estimating intercensal populations for urban areas or populations of areas difficult to conduct the census.

Areal interpolation of population may be further separated into two categories depending on whether the interpolation is based on mathematical functions or ancillary information (Wu, Qiu, and Wang 2005). For areal interpolation based on mathematical functions, there are point-based methods and areal-based methods, depending on whether the population data for computational input is in the digital form of points or areas. Examples of point-based interpolation include Martin (1989), Bracken (1991), and Martin (1996). Examples of area-based interpolation include Tobler (1979) and Rase (2001). Areal-based methods usually have the volume-preserving property -- the aggregated populations to census units from the interpolated populations are not changed. Volume preserving is an important requirement for accurate interpolation because the total population counts for census units should be maintained after the interpolation.

Population is related to other information such as land use, transportation, and topography. Ancillary information relevant to population distribution, therefore, can be used to assist areal interpolation of population. Areal interpolation using ancillary information can be referred to as the dasymetric mapping method, and this approach is generally volume-preserving (Wu, Qiu, and Wang 2005). The most commonly-used ancillary information for population interpolation is land use and/or land cover data (e.g., Yuan, Smith, and Limp 1997; Mennis 2003; Holt, Lo, and Hodler 2004). Other ancillary information includes topographic data (Wright 1936), election districts demographic data (e.g., Flowerdew and Green 1989; Flowerdew and Green 1991), road network data (e.g., Xie 1995; Reibel and Bufalino 2005; Hawley and Moellering 2005), and remote sensing image spectral and textural statistics (Harvey 2002b; Wu, Qiu, and Wang 2006).

Another category of literature relevant to population estimation is statistical modeling of population. Depending on the scales of analysis, researchers have used five types of predictor variables to infer populations (Wu, Qiu, and Wang 2005), including urban areas (e.g., Tobler 1969; Lo and Welch 1977; Prosperie and Eyton 2000), land use areas (e.g., Kraus, Senger, and Ryerson 1974; Weber 1994; Lo 2003), dwelling units count (e.g., Hsu 1971; Lo and Chan 1980; Lo 1989), image pixel statistics (e.g., Webster 1996; Harvey 2002a; Liu, Clarke, and Herold 2006), and other physical or socioeconomic characteristics (e.g., Green and Monier 1959; Dobson et al. 2000; Liu and Clarke 2002).

Past population estimation methods, whether they are based on areal interpolation or statistical modeling approaches, usually utilized coarse scale GIS and remote sensing data for analysis. For example, land use data used for population estimation may be classified from Landsat images with a spatial resolution of 30 meters. As a result, these methods are only suitable for population estimation at a coarse scale or for large areas. Even some studies utilized high-resolution images and derived high correlation coefficients between population density and image statistics, the population modeling and accuracy assessment were built upon aggregated census data, and the accuracy of using the derived models for sub-block population estimation is not certain. Therefore, the applicability of the models for fine-scale, small-area population estimation is questionable. For example, Wu, Qiu, and Wang (2006) inferred population densities from semivariance image texture statistics using high-resolution aerial photographs and census block population data. Although they obtained high R^2 between block population densities and image statistics, the robustness and applicability of the models to sub-block population estimates were not quantitatively verified.

Methodology

This article makes contributions in sub-block population estimation and accuracy assessment by presenting a deterministic model for population estimation and a simulation approach to assess the model for sub-block population estimation. Specifically, the deterministic model infers populations from building volumes and three block-level housing statistics, the former being derived from GIS data of building footprints, and the latter being derived from the U.S. Census 2000 data:

$$\text{Pop} = \text{BdV} / \text{HuSpace} * \text{OccRate} * \text{HdSize} \tag{1}$$

where Pop = populations, BdV = building volumes, HuSpace = average space per housing unit, OccRate = housing unit occupancy rate, and HdSize = average household size.

When total building volumes are divided by the average space per housing unit within a census block, the derived figure should be close to the total number of housing units within the block. Then, when the total number of housing units is multiplied by the occupancy rate of the block, the derived figure should be close to the total number of households within the block. Further, when the total number of households is multiplied by the average household size of the block, the derived figure should be close to the total populations within the block. Therefore, equation 1 provides a close estimate of the actual block population, and an approximate estimate of sub-block population depending on the uniformity of housing statistics within the block.

To assess sub-block population estimation accuracies, we simulated the base unit by aggregating blocks and then estimated sub-unit populations from the deterministic model. In this way, we were able to compare the sub-unit estimates with the actual sub-unit populations calculated from aggregating census block populations. Specifically, we simulated a base unit of twenty-block-aggregations and estimated populations for sub-units of one- to nineteen-block-aggregations. Then we compared the model estimates with the ground truth populations from the census. Using this simulation approach, we also assessed how different scales of base unit aggregations affect their sub-unit estimates. Specifically, we simulated the base units from two- to twenty-block-aggregations, and then compared the population estimation errors for their 50 percent base-unit areas. We also used the simulation approach to assess how the rescaling for volume preservation affected different scales of sub-unit estimates, and how land use information could improve the accuracy of population estimation.

Compared to related population estimation studies, our approach is unique in three aspects. Firstly, past studies generally relied on ancillary GIS and remote sensing data for population estimation. In contrast, we incorporated census housing statistics for population estimation. By incorporating block-level housing statistics, sub-block population estimation would be relatively reliable and accurate.

Secondly, most studies estimated populations by land use areas or image pixel characteristics. In contrast, this study estimated populations by building volumes, because people usually live and work in buildings. Further, the census aims to survey the amount of people living in residential buildings. Population estimation by buildings is consistent with census data that are used for population modeling. Buildings are also generally contained within man-made or natural zones of which populations need to be estimated. As a result, population modeling by building volumes would be more reliable, and sub-block population estimation by building volumes would be more accurate.

Lastly, past studies generally built population models from high-level census population data and then verified the models from low-level census data. For example, population models can be built from census tract or block group data, and census block data was used to verify the models. In this study, for the purpose of estimating sub-block populations, the deterministic model is based upon block-level population data, because block-level statistics are more similar to sub-block statistics than the statistics from block groups or census tracts. We then used the simulation approach to assess the model for sub-block population estimation.

Data Source

We used census geographic and demographic data, building data, elevation data, and land use data of the city of Austin to test the deterministic model for sub-block population estimation and accuracy assessment. The data are all for the year 2000. The census data was obtained from the U.S. Census Bureau’s American FactFinder website (U.S. Census Bureau 2006). The other three datasets were obtained from the City of Austin Neighborhood Planning and Zoning Department (NPZD), either directly downloaded from their FTP server (City of Austin 2005a) or acquired through personal contact.

In addition to census block geographies and populations, three block-level housing statistics were directly downloaded or computed from relevant census statistics and GIS data. The three block-level housing statistics include housing units occupancy rate, average household size, and average space per housing unit.

The building data is building footprints in the vector polygon format. The department only has building datasets for the years 1997 and 2003, but by referencing with high-spatial-resolution aerial photographs taken during the year 2000 we combined them into a single dataset for the year 2000. The building data contains information of the average altitude for individual building roofs. The building data and the elevation data are both generated by the Analytical Surveys Incorporation (ASI) contracted with the city. ASI first manually digitized building footprints from aerial photographs. Then, by referencing with remote sensing Light Detection and Ranging (LIDAR) data (2-feet spatial resolution), ASI estimated the altitude for individual building roofs. From the LIDAR data, ASI also generated the ground surface elevation data. The elevation data measures the elevation for ground surface in 0.61-meter (2-feet) contour lines. With the building roof altitude and the ground surface elevation collected, we were then able to estimate the average height for individual buildings. Building volumes were further derived by multiplying building footprint areas with a building’s height.

The land use data is in the vector polygon format. It is generated and updated by the NPZD based on a variety of sources, including historical land use data, Texas Central Appraisal

District (TCAD) tax parcel data, the city parcels database, natural preserves GIS data, aerial photographs, building footprint data, and field check information (City of Austin 2005b).

Assessing Block-Level Population Estimation

We first assessed how the deterministic model estimates populations for residential blocks. By referencing the land use data, 120 multi-family blocks and 600 single-family blocks were selected. The sample blocks were selected subjectively from the entire Austin area, with the goal that they were representative of a variety of housing patterns and population densities. Aerial photographs, building footprint data, land use data, and block-level population density choropleth maps were utilized to assist the selection process.

After sample blocks were selected, we calculated the total building volumes for individual blocks by overlaying with the building footprints data. We then applied the deterministic model (equation 1) to estimate populations for the sample blocks. The mean absolute relative error was calculated to compare model-estimated block populations with original block populations. The mean absolute relative error (MARE) was calculated as:

$$MARE = \frac{1}{m}\sum_{i=1}^{m}\frac{|P_i - Y_i|}{Y_i} \times 100\% \qquad (2)$$

where P_i is the model-estimated population for the *i*th census block, Y_i is the census reported population for the *i*th census block, and m is the number of census blocks under investigation. The MARE gives us an overall estimate of how many percent of original census block populations were underestimated or overestimated. This measure was adopted because it is easy to interpret. The MARE was calculated, respectively, for the 600 single-family blocks, the 120 multi-family blocks, and the combined 720 residential blocks (Figure 1). The results show that multi-family blocks have higher MARE than single-family blocks, and the combined residential blocks are in between. We can explain that single-family land use is more homogeneous than multi-family land use, and, therefore, the relationships between populations and the building volume variable are more uniform. In addition, by examining the standard deviation of population counts for the multi-family blocks and the single-family blocks, we can observe that multi-family blocks have more varied population distribution with a standard deviation of 573 persons compared to that of single-family blocks of thirty four persons, which also explains the higher estimation errors.

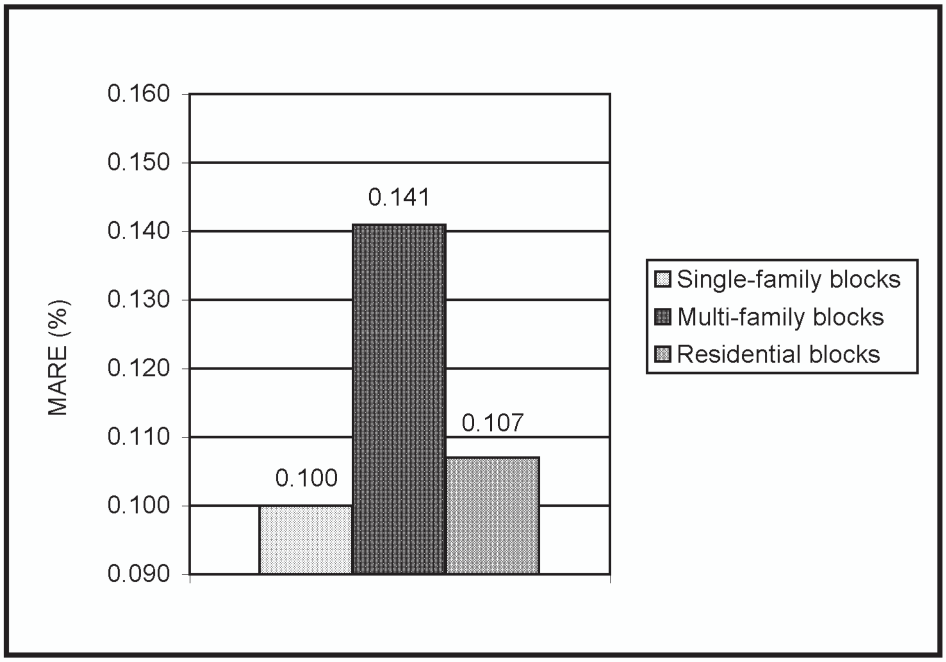

Figure 1. Population Estimation Errors for Single-Family Blocks, Multi-Family Blocks, and Combined Residential Blocks

Assessing Sub-block Population Estimation

The deterministic model ideally estimates areal populations based on the total building volumes and three block-level housing statistics. When applying the model for sub-block population estimation, the block-level housing statistics may not be applied to the entire block areas, and the sub-block estimates are likely to have higher errors. Since sub-block populations were unavailable from the census for estimation assessment, we assessed how well the model would estimate sub-block populations using a simulation approach that simulates the base unit for modeling by aggregating blocks. Firstly, we aggregated every twenty neighboring blocks as the base unit. The housing statistics of HuSpace, OccRate, and HdSize for the base units were calculated from their original block-level housing statistics. We then estimated populations for sub-units of one- to nineteen-block-aggregations based on the total building volumes within the sub-units and associated base-unit-level housing statistics. The actual populations for the sub-units were calculated by adding census block populations accordingly. Lastly, we calculated the MARE by comparing estimated sub-unit populations with their actual census populations. The processes of aggregating every twenty blocks as the base units, estimating populations for sub-units of one- to nineteen-block-aggregations, and comparing estimated sub-unit populations with actual sub-unit populations from the census were performed, respectively, for the 600 single-family blocks, the 120 multi-family blocks, and the 720 residential blocks (Figure 2 and Figure 3). In Figure 2 and Figure 3, the X-axis is the sub-unit area represented as the percentage of the base-unit area. For example, when the base unit is by twenty-block-aggregations, a sub unit of ten-block-aggregations would be at the 50 percent level. From Figure 2 and Figure 3, it is observed that generally the smaller the sub-unit areas, the higher the estimation errors. Estimates for multi-family areas have higher errors and uncertainties than for single-family areas due to their more heterogeneous nature, as previously discussed. The error graphs in Figure 2 and Figure 3 provide us with estimates of what the errors would be if the block-level deterministic model were used to estimate sub-block populations for residential land use areas.

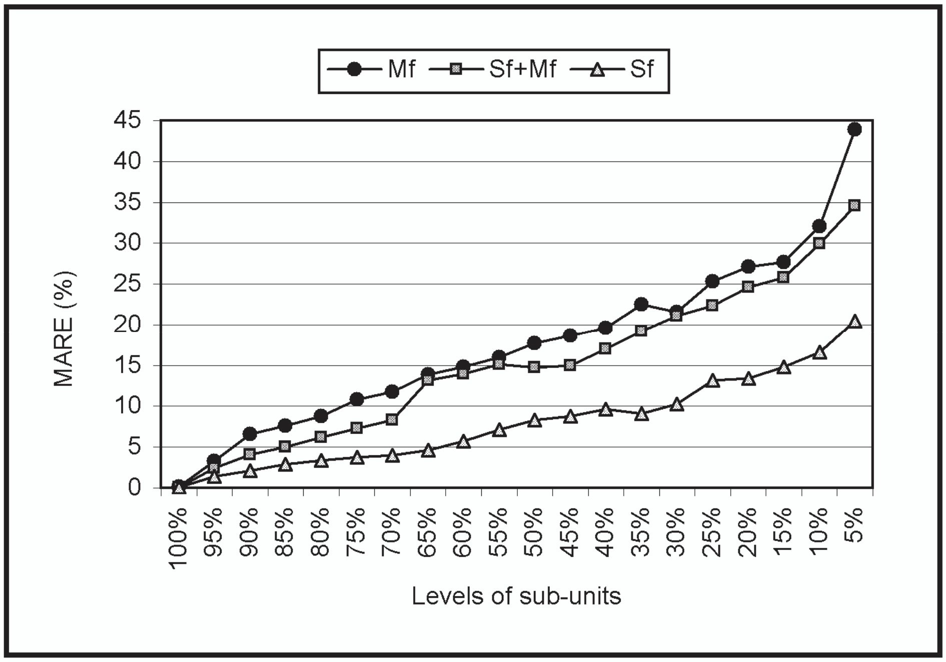

Figure 2. Sub-Unit Population Estimation Errors for Single-Family (SF), Multi-Family (MF), and Residential (Sf+Mf) Areas

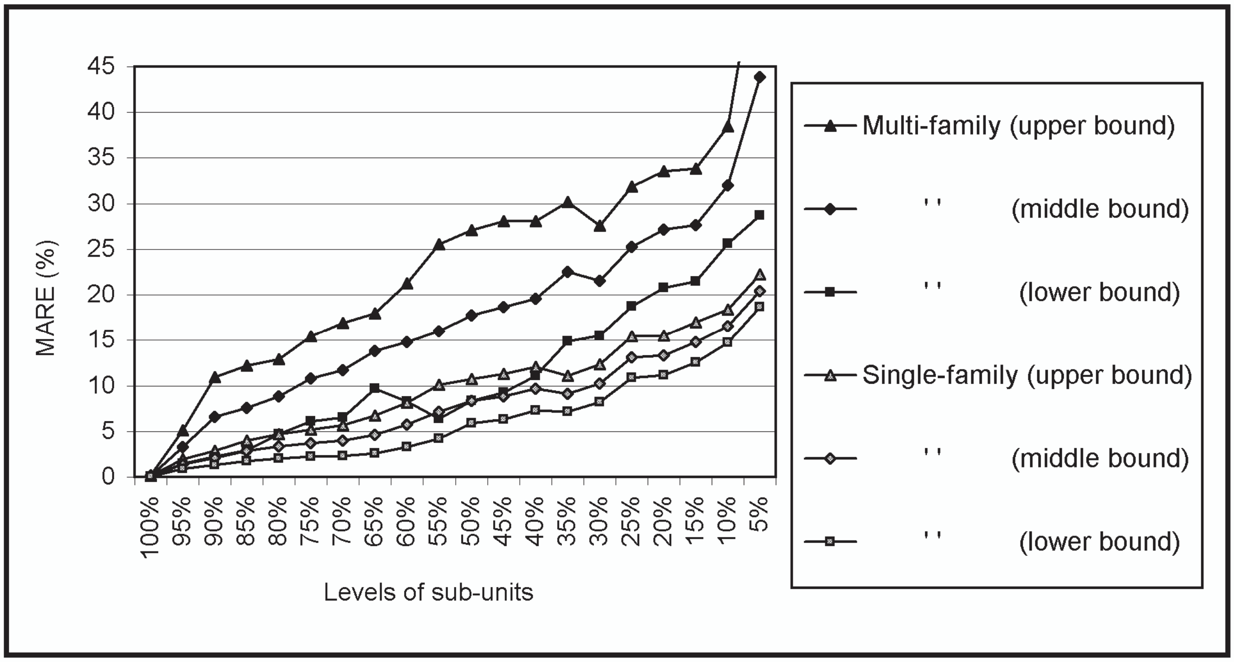

Figure 3. Sub-Unit Population Estimation Errors and Error Bounds for Single-Family and Multi-Family Areas

Assessing Aggregation Effects on Sub-Unit Estimates

To investigate whether the error graphs in Figure 2 and Figure 3, which are based on a base-unit of 20 blocks, can be applied in a context of a single block as the base unit, we tested to see if the estimation errors for the same level of sub-unit estimates vary with the base units used. Specifically, we simulated the base units from two- to twenty-block-aggregations, using the 720 residential blocks. Population estimates for the 50 percent level of sub-units were then calculated from the deterministic model as previously performed. The estimation errors for the 50 percent level of sub-units at different aggregations of base units (Figure 4) show that the errors do not relate to the size of the base units. The greatest difference between the highest MARE and the lowest MARE was approximately 3.3(percent). Therefore, we concluded that the error graphs at a base unit of 20-block-aggregations (Figure 2 and Figure 3) can be approximately applied to the standard interpolation when the base unit is a single block.

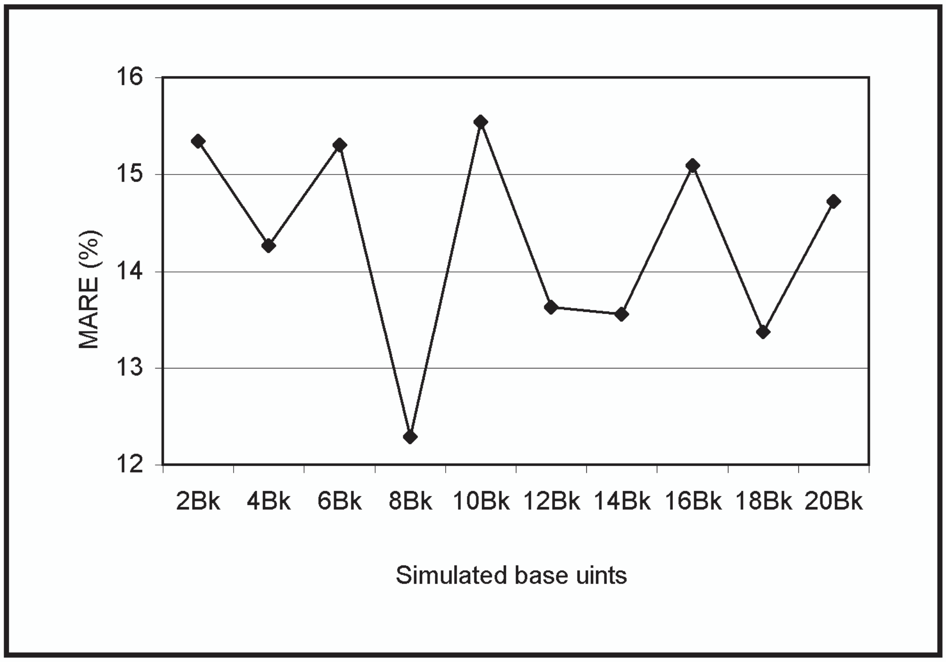

Figure 4. Population Estimation Errors for 50 Percent Base-Unit Areas with Different Unit Sizes

(Bk = Blocks)

Assessing Rescaling Effects on Sub-Unit Estimates

An important characteristic of dasymetric interpolation is volume preservation, which means that the sum of sub-unit population estimates is equal to the original base-unit population after interpolation. Volume preservation can be achieved by applying a single scaling factor to all sub-unit estimates. After the rescaling, the MARE for base-unit estimates is equal to zero, and sub-unit estimates may still have errors because sub-units are likely to have different population and housing characteristics, as well as different scaling factors from those of the base-unit. To investigate whether the rescaling improves all levels of sub-unit estimates equally, we calculated the difference of MARE between original sub-unit estimates and rescaled sub-unit estimates using the 720 residential blocks with twenty-block-aggregations as the base units and one- to nineteen-block-aggregations as the sub-units. The results in Figure 5 show that the rescaling improved all levels of sub-unit estimates but not in a systematic way. The rescaling greatly improved the base-unit estimates because the base-unit populations were maintained after the rescaling. We concluded that the rescaling effect on the improvement of sub-unit estimates does not relate to the levels of sub-units.

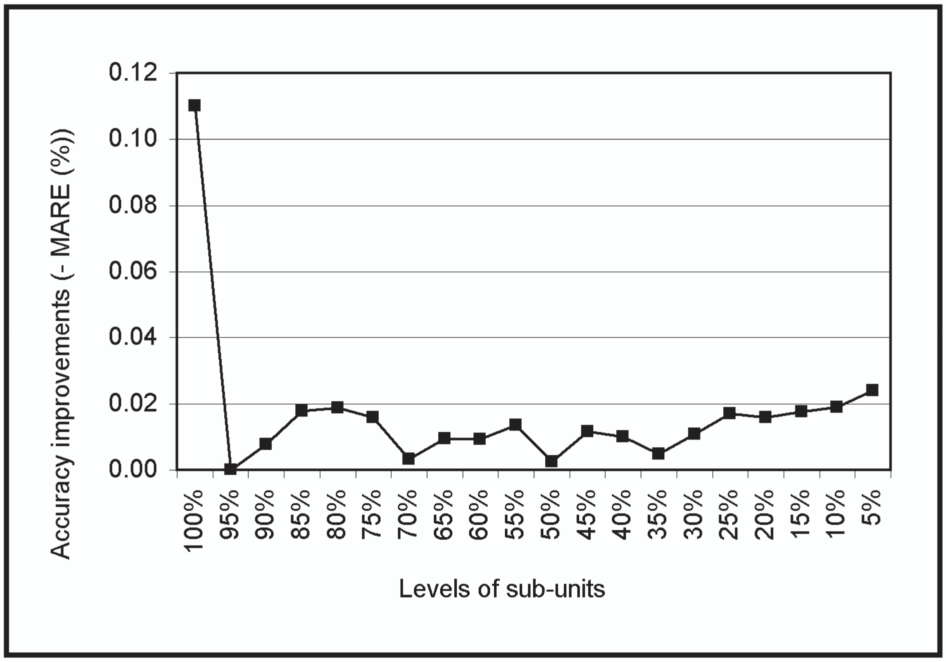

Figure 5. The Improvements of Sub-Unit Population Estimation after the Rescaling for Base-Unit Population Preservation

Assessing Estimation Errors for Mixed-Land-Use Areas

Residential land use information is not always available for estimating sub-block populations. Therefore, we tested sub-unit population estimation for mixed land use areas. A total of 1,320 blocks that contain mixed residential and non-residential land use were selected. We then aggregated every 20 neighboring blocks as the base unit and calculated the average housing statistics for each individual base unit. Populations of the sub-units of one- to nineteen-block-aggregations were estimated from the deterministic model and further rescaled to maintain base-unit populations. The MARE for each of the rescaled sub-unit estimates was calculated and compared with those calculated from residential blocks (Figure 6). The results show that sub-unit estimates for mixed-land-use areas have higher errors than for residential areas, particularly at lower levels of sub-units.

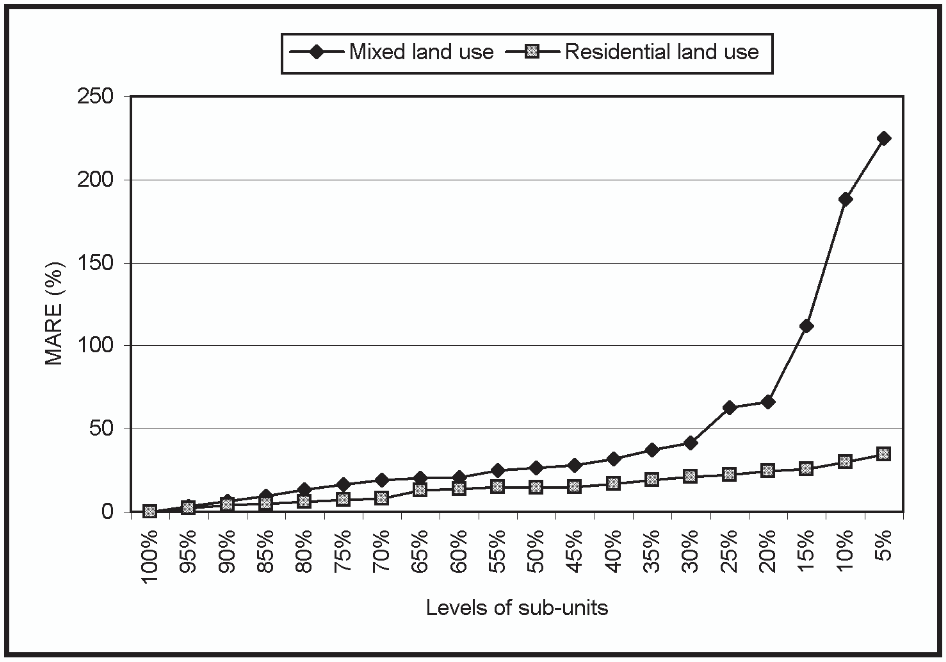

Figure 6. Sub-Unit Population Estimation Errors for Mixed Land Use Areas and for Residential Areas

Discussions

This deterministic model is a robust model that can be applied to other U.S. cities because it makes use of local block-level census statistics. On the other hand, the error graphs of sub-unit population estimates for single-family, multi-family, residential, and mixed land use areas (Figure 2, Figure 3, and Figure 6) may vary between cities, depending on the homogeneity of local land use, housing, and population distribution patterns. Therefore, to predict sub-unit population estimation errors, researchers need to re-calculate the error graphs using the proposed base-unit simulation approach.

The proposed deterministic model can estimate block-level population with a high degree of accuracy but not much for sub-block estimates. For population estimation for 50% block areas, the estimated MARE is 17.7% for multi-family area, 14.7% for residential area, and 8.3%

for single-family area (figure 2). For population estimation for 5% block areas, the estimated MARE is 43.9%, 34.6%, and 20.5% respectively. Generally a building is smaller than 5% of a block. So there would be a high degree of errors to estimate populations by individual buildings. As a result, if accurate population counts are needed for small, sub-block areas, deliberate field surveys of sub-block housing statistics would be necessary. For example, if we want to estimate populations for a sub-block area consisting of twenty houses, we may survey the housing statistics for every four houses, and then apply the model to estimate populations for these areas. The underlying assumption is that neighboring houses have similar housing statistics.

Another limitation of the presented fine-scale population estimation model is that it relies on building volumes data for model input. In the past, no effective way of automatically extracting building areas and heights existed. Researchers relied on manually identifying and counting dwelling units from high-spatial-resolution aerial photographs, even though visual interpretation is laborious and time consuming. With the advance of very high spatial resolution satellite images, such as IKONOS and QuickBird, and the improvement of feature extraction techniques (Haverkamp 2004), automatic extraction of dwelling units from satellite images has become possible. Another prospect for automatic building extraction has come with the advancement of 3D object extraction techniques from LIDAR data (Rottensteiner 2003; Rottensteiner et al. 2004). With these new remote sensing data and building extraction techniques, population estimation by building volumes would become a feasible approach.

Another limitation regarding data source is that the model relies on existing census housing statistics. In other words, if the model is to be applied for non-census years, an assumption regarding the timeliness of input data must be made and/or additional up-to-date data must be acquired. For example, if we want to infer fine-scale populations for the year of 2005 using the deterministic model, we may either use housing statistics from the Census 2000 data or collect up-to-date housing statistics for model input.

There may be several sources of errors in this study due to data quality/accuracy issues, such as spatial misalignment between census block geographies and building footprints data, census data miscounting, and houses unoccupied or otherwise under construction. Nevertheless, when the sub-block estimates are constrained by census block population totals, all errors are also constrained within block-level estimates. Therefore, errors due to data quality/accuracy issues would not have significant impact on population mapping and estimation for large areas.

Conclusions

This article presents a deterministic model to estimate populations for small areas based on building volume statistics and three block-level housing statistics, including the average space per housing unit, the housing unit occupancy rate, and the average household size. Model assessment for block-level estimates shows that the deterministic model can estimate block-level populations within 0.15 percent accuracy. Population estimates for multi-family areas have higher errors than that for single-family areas because of their more heterogeneous housing characteristics and population distributions. Model assessment for sub-block estimates, using a base-unit simulation approach, indicates that the smaller the sub-unit areas, the higher the estimation errors. The assessment also reveals that estimation errors do not relate to the sizes of base units. Therefore, it is reasonable to infer errors from a base unit of 20 blocks to a base unit of one block. Assessment of the rescaling effect on sub-unit population estimation indicates that the rescaling improves all levels of sub-unit estimates but not in a systematic way. Further, the assessment results show that population estimation for mixed-land-use areas have higher errors than that for residential-land-use areas, due to their more heterogeneous nature. We concluded that land use information allows us to obtain more reliable estimates of the potential errors associated with population estimation, particularly for sub-block estimates.

Future research is needed to study the estimation of sub-block populations for the current year based on the deterministic model. Here are some questions to ask. How have the local housing statistics of the average space per housing unit, housing unit occupancy rate, and average household size changed since the year 2000? Can the changes be estimated/projected? Is it feasible to conduct field survey to obtain up-to-date housing statistics? How do we design the sampling scheme to estimate populations for the areas under investigation?

Future research can use the base-unit simulation approach to investigate the modifiable area unit problem in analyses of socio-economic variables. Here are some questions to ask. How do R^2 and coefficients from standard regression or geographically weighted regression analysis of socio-economic variables change with the scale and the shape of the units used? By using the base-unit simulation approach, how does aggregation by location proximity versus by attribute similarity affect R^2 and coefficients? How does spatial autocorrelation play a role here?

References

Bracken, I. 1991. A surface model approach to small area population estimation. *Town Planning Review* 62 (2): 225–237.

City of Austin. 2005a. City of Austin GIS Data Sets. ftp://coageoid01.ci.austin.tx.us/GIS-Data/Regional/coa_gis.html (last accessed 6 June 2006).

______. 2005b. Land Use Survey Methodology. http://www.ci.austin.tx.us/landuse/survey.htm (last accessed 6 June 2006).

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

Flowerdew, R., and M. Green. 1989. Statistical methods for inference between incompatible zonal systems. In *Accuracy of Spatial Databases*, eds. M. Goodchild, and S. Gopal, 239-247. London: Taylor and Francis.

______. 1991. Data integration: Statistical methods for transferring data between zonal systems. In *Handling Geographical Information: Methodology and Potential Applications*, eds. I. Masser, and M. Blakemore, 38–54. New York: Wiley.

Green, N. E., and R. B. Monier. 1959. Aerial photographic interpretation of the human ecology of the city. *Photogrammetric Engineering* 25: 770-773.

Harvey, J. T. 2002a. Estimating census district populations from satellite imagery: Some approaches and limitations. *International Journal of Remote sensing* 23 (10): 2071–2095.

______. 2002b. Population estimation models based on individual TM pixels. *Photogrammetric Engineering and Remote Sensing* 68 (11): 1181-1192.

Haverkamp, D. 2004. Automatic building extraction from IKONOS imagery. In *Proceedings of ASPRS 2004 Conference, Denver, Colorado, May 23-28*. Available at http://www.spaceimaging.com/techpapers/default.htm (last accessed 6 June 2006).

Hawley, K., and H. Moellering. 2005. A comparative analysis of areal interpolation methods. *Cartography and Geographic Information Science* 32 (4): 411-423.

Holt, J. B., C. P. Lo, and T. W. Hodler. 2004. Dasymetric estimation of population density and areal interpolation of census data. *Cartography and Geographic Information Science* 31 (2): 103-121.

Hsu, S. Y. 1971. Population estimation. *Photogrammetric Engineering* 37: 449-454.

Kraus, S. P., L. W. Senger, and J. M. Ryerson. 1974. Estimating population from photographically determined residential land use types. *Remote Sensing of Environment* 3 (1): 35-42.

Liu, X., and K. C. Clarke. 2002. Estimation of residential population using high resolution satellite imagery. In *Proceedings of the 3rd Symposium in Remote Sensing of Urban Areas, June 11-13, 2002*, eds. D. Maktav, C. Juergens, and F. Sunar-Erbek, 153-160. Istanbul, Turkey: Istanbul Technical University Press.

Liu, X., K.C. Clarke, and M. Herold. 2006. *Population density and image texture: a comparison study*. *Photogrammetric Engineering & Remote Sensing* 72(2): 187-196.

Lo, C. P. 1989. A raster approach to population estimation using high-altitude aerial and space photographs. *Remote Sensing of Environment* 27 (1): 59-71.

______. 2003. Zone-based estimation of population and housing units from satellite-generated land use/land cover maps. In *Remotely sensed cities*, ed. V. Mesev, 157-180. New York: Taylor & Francis.

Lo, C. P., and H. F. Chan. 1980. Rural population estimation from aerial photographs. *Photogrammetric Engineering and Remote Sensing* 46 (3): 337-345.

Lo, C. P., and R. Welch. 1977. Chinese urban population estimates. *Annals of the Association of American Geographers* 67 (2): 246-253.

Martin, D. 1989. Mapping population data from zone centroid locations. *Transactions of the Institute of British Geographers* 14 (1): 90-97.

______. 1996. An assessment of surface and zonal models of population. *International Journal of Geographical Information Systems* 10 (8): 973-989.

Mennis, J. 2003. Generating surface models of population using dasymetric mapping. *The Professional Geographer* 55 (1): 31-42.

Prosperie, L., and R. Eyton. 2000. The relationship between brightness values from a nighttime satellite image and Texas county population. *The Southwestern Geographer* 4: 16-29.

Rase, W. 2001. Volume-preserving interpolation of a smooth surface from polygon-related data. *Journal of Geographical Systems* 3 (2): 199-213.

Reibel, M., and M. E. Bufalino. 2005. Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environment and Planning A* 37 (1): 127-139.

Rottensteiner, F. 2003. Automatic generation of high-quality building models from Lidar data. *IEEE Computer Graphics and Applications* 23 (6): 42-50.

Rottensteiner, F., J. Trinder, S. Clode, K. Kubik, and B. Lovell. 2004. Building detection by Dempster-Shafer fusion of lidar data and multispectral aerial imagery. *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)* 2: 339-342.

Tobler, W. R. 1969. Satellite confirmation of settlement size coefficients. *Area* 1 (3): 30-34.

______. 1979. Smooth pyncophylactic interpolation for geographical regions. *Journal of the American Statistical Association* 74 (367): 519-30.

U.S. Census Bureau. 2006. The U.S. Census American FactFinder. http://factfinder.census.gov/home/saff/main.html?_lang=en (last accessed 6 June 2006).

Weber, C. 1994. Per-zone classification of urban land use cover for urban population estimation. In *Environmental remote sensing from regional to global scales*, eds. G. M. Foody, and P. J. Curran, 142-148. New York: Wiley.

Webster, C. J. 1996. Population and dwelling unit estimation from space. *Third World Planning Review* 18 (2): 155-176.

Wright, J. K. 1936. A method of mapping densities of population. *The Geographical Review* 26 (1): 103-110.

Wu, S., X. Qiu, and L. Wang. 2005. Population estimation methods in GIS and remote sensing: a review. *GIScience & Remote Sensing* 42 (1): 58-74.

______. 2006. Using semi-variance image texture statistics model population densities. *Cartography and Geographic Information Science* 33 (2): 127-140.

Xie, Y. 1995. The overlaid network algorithms for areal interpolation problem. *Computers Environment and Urban Systems* 19 (4): 287-306.

Yuan, Y., R. M. Smith, and W. F. Limp. 1997. Remodeling census population with spatial information from LandSat TM imagery. *Computers, Environment and Urban Systems* 21 (3-4): 245-258.