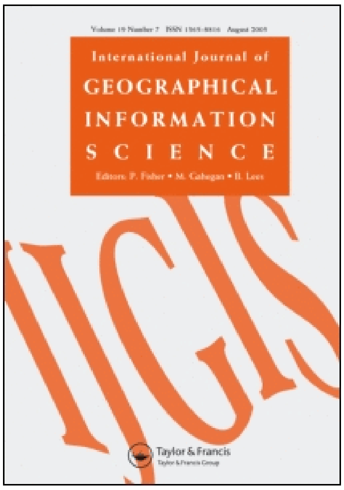

International Journal of Geographical Information Science
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713599799

Assessing housing growth when census boundaries change

Alexandra D. Syphard [a]; Susan I. Stewart [b]; Jason Mckeefry [a]; Roger B. Hammer [c]; Jeremy S. Fried [d]; Sherry Holcomb [a]; Volker C. Radeloff [a]

[a] Department of Forest Ecology and Management, University of Wisconsin, Madison, WI, USA [b] USDA Forest Service, Northern Research Station, Evanston, IL, USA [c] Department of Sociology, Oregon State University, Corvallis, OR, USA [d] USDA Forest Service, Pacific Northwest Research Station, Portland, OR, USA

First Published on: 03 November 2008

To cite this Article Syphard, Alexandra D., Stewart, Susan I., Mckeefry, Jason, Hammer, Roger B., Fried, Jeremy S., Holcomb, Sherry and Radeloff, Volker C.(2008)'Assessing housing growth when census boundaries change',International Journal of Geographical Information Science,99999:1,

To link to this Article: DOI: 10.1080/13658810802359877

URL: http://dx.doi.org/10.1080/13658810802359877

International Journal of Geographical Information Science
2008, iFirst Article, 1–18

Research Article

Assessing housing growth when census boundaries change

ALEXANDRA D. SYPHARD*†, SUSAN I. STEWART‡,
JASON MCKEEFRY†, ROGER B. HAMMER§, JEREMY S. FRIED¶,
SHERRY HOLCOMB† and VOLKER C. RADELOFF†

†Department of Forest Ecology and Management, University of Wisconsin, Madison, WI, 53706, USA
‡USDA Forest Service, Northern Research Station, Evanston, IL, 60201, USA
§Department of Sociology, Oregon State University, Corvallis, OR, 97331, USA
¶USDA Forest Service, Pacific Northwest Research Station, Portland, OR, 97205, USA

(*Received 15 August 2007; in final form 09 July 2008*)

The US Census provides the primary source of spatially explicit social data, but changing block boundaries complicate analyses of housing growth over time. We compared procedures for reconciling housing density data between 1990 and 2000 census block boundaries in order to assess the sensitivity of analytical methods to estimates of housing growth in Oregon. Estimates of housing growth varied substantially and were sensitive to the method of interpolation. With no processing and areal-weighted interpolation, more than 35% of the landscape changed; 75–80% of this change was due to decline in housing density. This decline was implausible, however, because housing structures generally persist over time. Based on aggregated boundaries, 11% of the landscape changed, but only 4% experienced a decline in housing density. Nevertheless, the housing density change map was almost twice as coarse spatially as the 2000 housing density data. We also applied a dasymetric approach to redistribute 1990 housing data into 2000 census boundaries under the assumption that the distribution of housing in 2000 reflected the same distribution as in 1990. The dasymetric approach resulted in conservative change estimates at a fine resolution. All methods involved some type of trade-off (e.g. analytical difficulty, data resolution, magnitude or bias in direction of change). However, our dasymetric procedure is a novel approach for assessing housing growth over changing census boundaries that may be particularly useful because it accounts for the uniquely persistent nature of housing over time.

Keywords: US Census; Housing density; Interpolation; MAUP; Census boundary; Land use/land cover change

1. Introduction

Housing development has been occurring at unprecedented rates, both globally and in the United States. In fact, due to the declining average household size, housing development is growing substantially faster than the population (Liu *et al.* 2003). From 1945 to 2002, the population in the United States doubled, but urban area quadrupled (Lubowski *et al.* 2006). The fastest-growing type of development in the

*Corresponding author. Email: asyphard@yahoo.com

International Journal of Geographical Information Science
ISSN 1365-8816 print/ISSN 1362-3087 online
http://www.tandf.co.uk/journals
DOI: 10.1080/13658810802359877

United States, low-density residential, also consumes the most land, occupying approximately five times more area than all other urban and suburban developments combined (Crump 2003, Theobald 2004).

The pace and extent of housing growth is linked to serious ecological and social issues, so understanding and documenting this change is becoming an urgent research priority in a number of disciplines (Radeloff *et al.* 2001, Brown *et al.* 2005, Foley *et al.* 2005). Housing development is a critical concern for conservation biologists because it is a primary cause of biodiversity loss (Sala *et al.* 2000, Theobald *et al.* 2000, Shochat *et al.* 2006). Both plant and animal species are directly impacted by the habitat loss and fragmentation caused by housing development (Liu *et al.* 2003), but housing development also indirectly threatens ecological integrity and native species diversity by contributing to factors such as exotic species invasions, increased resource consumption, and recreational impacts (McKinney 2002). Furthermore, ecologists are concerned about the locations of habitat conversion because a substantial proportion of new residential development is now occurring at the boundaries of public lands and sensitive conservation areas (Miller and Hobbs 2002).

Many of the ecological issues related to housing development have social implications. For example, dispersed housing development located within or near wildland vegetation increases the wildland urban interface, which in turn increases human vulnerability to wildfire (Radeloff *et al.* 2005). Other social issues related to urban sprawl include traffic congestion, public health concerns and fiscal disparities (Downs 1999). In fact, urban sprawl is now one of the top concerns of US citizens (Pew Center for Civic Journalism 2000), who are increasingly voicing demands for more open space (Kline 2006).

Considering these wide-reaching impacts, there is growing demand for fine-resolution, spatially explicit social data that can be used to document how the pattern and extent of housing density changes over time (Brown *et al.* 2005). The US Census Bureau provides the primary source of social and housing data used in the health, social, ecological and geographic sciences in the United States (Krieger 2006). Through the topologically integrated geographic encoding and referencing (TIGER) system, data for various political and statistical geographies are provided nation-wide every decade at the resolution of census blocks, which are the smallest geographic areas for which the Bureau of the Census collects and tabulates data. The boundaries of census blocks occur along roads, railroads, streams or other bodies of water, other visible and cultural features, and legal or administrative boundaries (US Census Bureau 1994). While these fine-resolution census block data provide a wealth of information, the geography of block boundaries changes over time. This is a major drawback; in many states, more than 50% of the census blocks changed boundaries from 1990 to 2000 (figure 1).

The need to analyze spatially explicit social data over time is internationally pervasive and not limited to the United States. Many other countries also collect social, economic and demographic data aggregated within administrative zones. These data may be in the form of census maps (e.g. United States, Japan, United Kingdom), or they may be available as national longitudinal geographic information systems (GIS) (e.g. China, Belgium, Ireland and Russia) (Sadahiro 2000, Martin 2003, Knowles 2005, Gregory and Ell 2006). Some of these census products and historical datasets provide the only long-term, spatially explicit social data available for some countries, while for other countries, these data are available

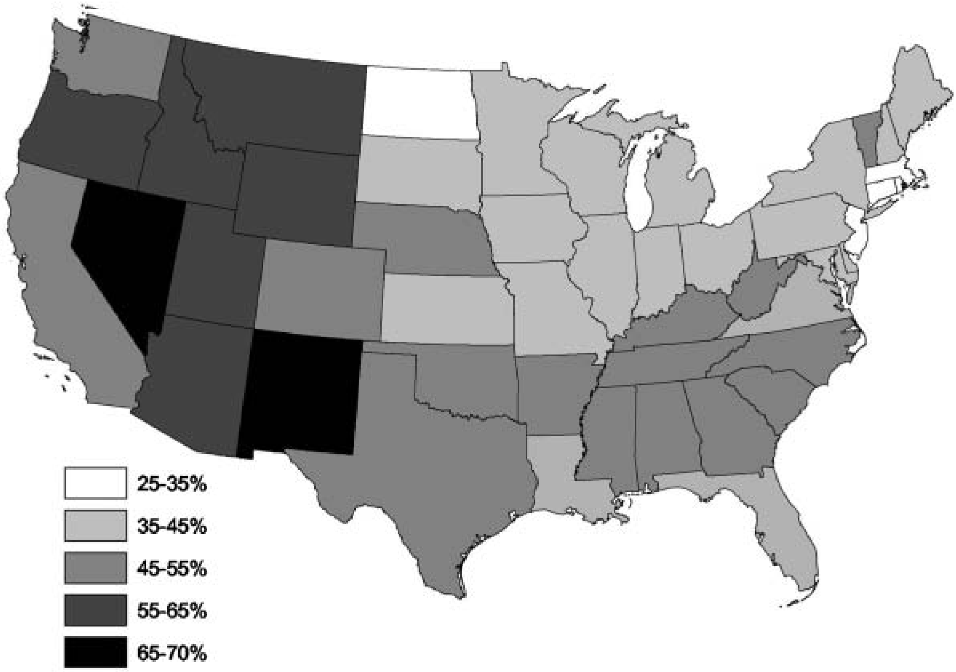

Figure 1. The percentage of 2000 US Census blocks that had different boundaries in the 1990 US Census.

at a finer resolution than other available data. All countries have a common challenge: to examine temporal characteristics of housing density while administrative boundaries change over time.

Changes in census boundaries (or other administrative zones) are problematic because they result in incompatible areas, or spatial bases (Goodchild *et al.* 1993), for analyzing changes between census enumerations. Spatial basis incompatibility problems occur any time data are available for one spatial basis (e.g. census geography), but analysis is required for data in a different spatial basis (e.g. ecoregions). Problems also occur in longitudinal analysis, when geography changes over time. The types of analytical issues related to the need for interpolation between incompatible spatial bases (or 'modifiable' areal units) were first described in the statistical and geographic literature and are collectively referred to as the modifiable areal unit problem (MAUP) (Openshaw 1984). The MAUP describes how the results of an analysis may vary if areal data are combined into sets of increasingly larger or smaller areas (the 'scale' problem); or, if boundaries are defined differently within areas of the same geographic extent (the 'zoning' problem) (Jelinski and Wu 1996). The zoning problem only applies to spatially extensive variables. Both scale and zoning problems occur when census boundaries change over time.

The problems inherent in changing boundaries between census dates seriously limit any benefits of having finer-resolution data in more recent years. To date, many of the methods that have been developed to overcome this problem have involved some level of sacrifice, or trade-off, in the analysis. For example, some approaches simply avoid direct geographic overlay in the analysis between dates. One of these approaches, referred to as 'contemporaneous census tracts', assesses change by using the boundaries from one census date to analyze data from that year and then uses boundaries from other census dates to analyze data from those years (e.g. Yang and Jargowsky 2006). While this approach offers a unique way to quantify change without incurring errors due to changing census boundaries, it also prohibits spatial analyses that are only possible through direct overlay.

When direct geographic overlays are performed, data within census block boundaries are sometimes merged or aggregated for one date based on the boundaries of another date without considering the MAUP and potential for misinterpretation of the analysis (e.g. Heidkamp and Lucas 2006). Other approaches apply analyses at coarser scales than census blocks to assure consistent zonal boundaries over time (e.g. Hammel and Wyly 1996), although even census tract boundaries may shift over time. Although analyses using these coarser scales are appropriate for some broad-scale applications (e.g. national land use change, Brown *et al.* 2005), other applications may require data to be analyzed at finer resolutions (e.g. economic change in housing markets, Steif 2004).

One method that allows direct overlay at a finer resolution is areal interpolation, a common spatial interpolation technique used to transfer data from one zonal system (the source zone) into another zonal system (the target zone). The simplest form of this technique is areal-weighting, which assumes that the data to be interpolated are distributed homogenously within each source and target zone (Goodchild and Lam 1980, Flowerdew and Green 1992). Because social characteristics are often distributed heterogeneously instead of homogenously, errors inherent in areal-weighting have been widely documented (Sadahiro 2000, Gregory 2002, Gregory and Ell 2006). To overcome these problems, some analysts have developed dasymetric mapping techniques, which are a form of areal interpolation. In dasymetric mapping, the analyst uses ancillary data to develop a weighting scheme that determines how the source zone data should be distributed in the target zone boundaries. Depending on the application and type of data, dasymetric mapping may or may not improve upon areal-weighted interpolation (Chen *et al.* 2004, Hay *et al.* 2005).

Most of census interpolation focuses on population data as the primary variable of interest. In some cases, population data are used as ancillary information to inform the interpolation of other demographic variables (e.g. Gregory 2002). Most applications, however, use other ancillary data to improve the interpolation of population (e.g. Mennis 2003, Reibel and Bufalino 2005, Langford 2006, GeoLytics (http://geolytics.com/Default.asp)). Manipulating raster-based ancillary data (such as remotely sensed land cover) so as to distribute population data into polygon-based target zones is challenging; so the use of dasymetric mapping has been limited (Chen *et al.* 2004). Using vector-based data, such as streets, has improved the results of areal interpolation while easing the difficulties involved with raster-based interpolation (e.g. Reibel and Bufalino 2005). However, this approach assumes that local roads are reasonable indicators of where population is distributed. Another approach used to estimate the spatial distribution of population density is geospatial Kernel Density Estimation. Kernel Density Estimation techniques typically use point-based data to derive a continuous density surface in which population densities across grid cells are weighted according to their distance from points with known population values (Goodchild *et al.* 1993). Kernel Density Estimation methods depend on the availability and resolution of population data, so they work best in urban areas where points of known population are located close together (Martin *et al.* 2000).

While these different approaches have been used to analyze change in population density, there is a strong and increasing need to map housing density patterns over time due to the major ecological and social implications of such profound landscape change. Considering that interpretations of population change have been sensitive to the method of interpolation, our objective was to develop and compare alternate methods of interpolating housing density to determine whether estimates of housing growth share this same sensitivity. Housing data are different from population data because housing structures are stationary and persistent over time, whereas people tend to migrate. Therefore, in addition to areal weighting and interpolation over aggregated boundaries, we developed a dasymetric technique that exploited this unique aspect of housing data. For this approach, we redistributed 1990 housing data into 2000 census boundaries based on the assumption that 1990 housing was distributed proportionately to 2000 housing. We compared estimates of housing growth using all three of these techniques to estimates of housing growth based on no interpolation in order to assess the differences between the results using these different methods.

2. Methods

2.1 *Data*

We developed and tested our methods using the TIGER census data for Oregon, a state where more than 60 percent of block boundaries changed between 1990 and 2000, making it a particularly complex and challenging case. The majority of boundary changes that occurred in Oregon, as well as the other states in the nation, occurred because the 2000 blocks tended to be finer in resolution, although in a few places the 1990 blocks had finer resolution than the 2000 blocks (figure 2). The most widespread consequence of these boundary shifts was that, without reconciliation, large areas appeared to change from low housing density in 1990 to no housing in 2000 (figure 3A).

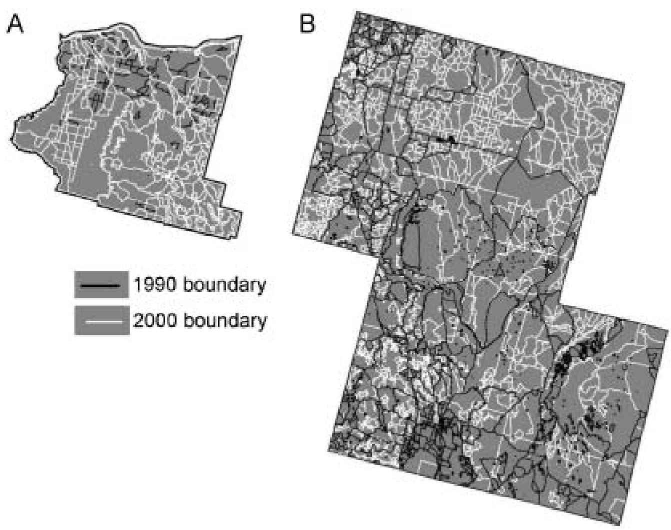

Figure 2. Maps of Sherman, Gilliam (A), and Lake counties (B) in Oregon illustrating the mismatch in census boundaries between 1990 and 2000.

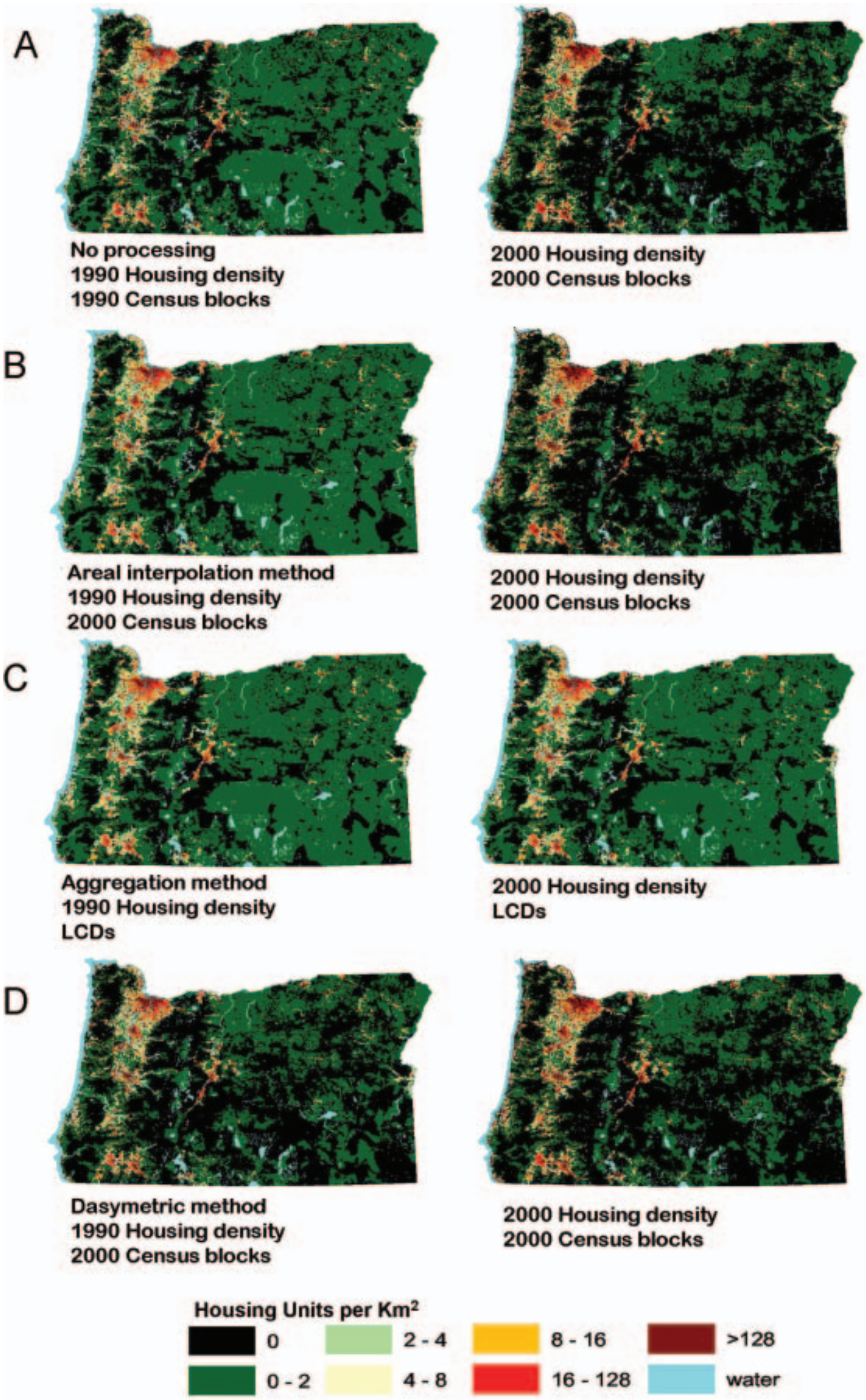

Figure 3. Maps of Oregon showing the spatial differences in housing density between 1990 and 2000. (A) shows maps of housing density using 1990 and 2000 Census blocks (with no reconciliation between block boundaries). Three methods of reconciliation between Census block boundaries are shown with (B) areal weighting, (C) aggregation, and (D) dasymetric procedure.

2.2 *Interpolation methods*

In order to evaluate the sensitivity of housing density estimates to different interpolation methods, we compared an overlay of 1990 and 2000 census boundaries with no processing to three alternative interpolation approaches. We compared approaches by visualizing the differences in the spatial distribution of housing density classes in the 1990 and 2000 maps and by calculating percent change in housing density from 1990 to 2000. Most of the area in the state had low housing density (less than 8 housing units per km^2). Therefore, for the analysis, we binned housing density into classes with a finer resolution at lower densities so we could better distinguish change among the classes in this range.

2.2.1 Areal weighting. Our areal weighing approach involved redistributing the housing data in the 1990 census blocks (the source zones) into the 2000 census blocks (the target zones) (as shown in the overlay in figure 4A). The procedure depended upon two assumptions: first, the analysis had to occur across the same geographic extent for 1990 and 2000 (the extent in figure 4B); and second, the 1990 data had to be normalized into densities. To start the areal weighting, we first performed a union with the two geographies. Then, using the 2000 geography, we computed the area-weighted average for all of the 1990 component densities that occurred within each 2000 block's boundaries (figure 4B). In other words, we multiplied the number of houses per unit area in the 1990 census blocks by the area of the 2000 census blocks to arrive at estimates of 1990 density in the 2000 census blocks.

2.2.2 Aggregation. Due to the multiple ways that boundaries can shift between census enumerations, we defined four possible relationships that can exist between datasets, assuming that the extents of the two incompatible zonal systems are

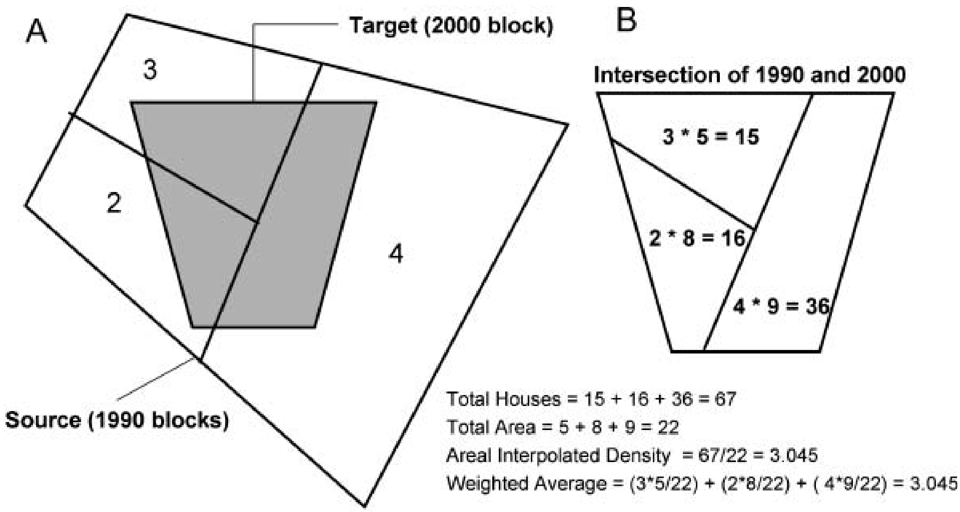

Figure 4. Conceptual diagram of the areal weighting method to redistribute 1990 census data on housing into 2000 census block boundaries. (A) shows an overlay of the 1990 and 2000 census boundaries and the distribution of the number of houses in 1990. (B) shows how the houses in 1990 are area-weighted within the intersection of 1990 and 2000 to determine the areal interpolated density of the 2000 target zone.

congruent (such as the extent in figure 4B common to both the 1990 and the 2000 blocks). Zone A represents the 1990 blocks, and zone B the 2000 census blocks.

(1) If there is an exact correspondence between a zone in A and B, then the relationship of A to B is one to one (1–1).
(2) If several zones in B can be aggregated to exactly match a zone in A, the relationship of A to B is one to many (1–M). For example, a 1990 block has been broken into several pieces in the 2000 geography.
(3) If several zones in A can be aggregated to exactly match a zone in B, the relationship of A to B is many to one (M–1). For example, a group of 1990 blocks has been combined into one 2000 block (figure 4B).
(4) If none of the previous apply, then the relationship of A to B is many to many (M–M) such that >1 zone in both A and B must be aggregated to create congruent shapes (figure 5).

In all of these cases, the shapes created by these aggregations are referred to as 'least common denominator' polygons (LCD polygons), in other words, the smallest possible combination of blocks required to match the 1990 boundaries to the 2000 boundaries. In all except the 1–1 case, LCD polygons are larger (lower resolution) than the zones of either A and/or B (illustrated by the LCD polygon in figure 5).

In the areal weighting method (figure 4B), only the M–1 case was performed at the level of the LCD polygons. Otherwise, averaging occurred at the resolution of the 2000 census blocks. For the aggregation method, we explicitly identified LCD polygons, then summed the number of houses over the aggregated component zones to arrive at values for the LCD polygons (figure 5). Thus, for aggregation, we summed extensive variables (i.e. counts) instead of relying on intensive values (i.e. proportions, percentages, or rates). To obtain extensive variables for the LCD polygons, we multiplied the intensive variables by the areas of the component zones and summed them for the LCD polygon.

2.2.3 Dasymetric method. We designed the dasymetric mapping method to take advantage of the unique characteristic of housing data. Because housing units are rarely lost or moved (particularly on a time scale of a decade), the distribution of

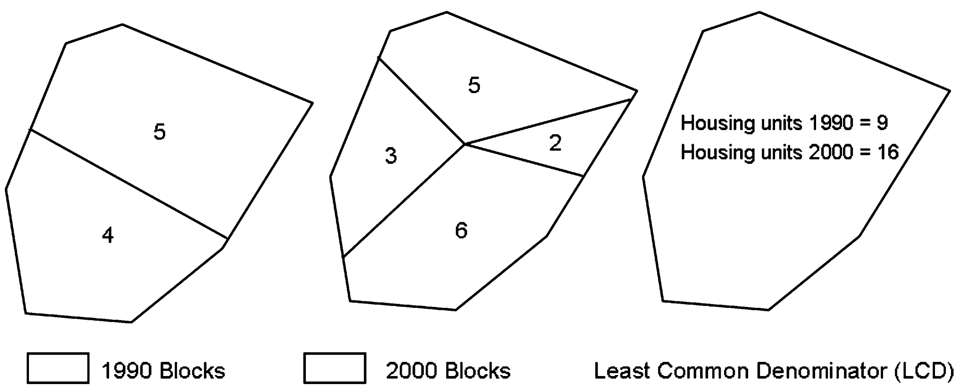

Figure 5. Conceptual diagram of the aggregation method to create least common denominators (LCDs), which are the smallest combined areas between 1990 and 2000 census block boundaries. The numbers of houses were summed for 1990 and 2000 to determine the total number within the common boundary.

houses at the beginning of a time period will be largely unchanged at the end of the time period. Thus, distribution of 2000 housing depends on that of 1990 housing. Our dasymetric approach used a bootstrapping technique in which the 2000 housing density distribution served as ancillary data to determine how to distribute the 1990 housing density data. In other words, we allocated the 1990 housing units into the 2000 census blocks according to the distribution of the 2000 housing units, assuming that the 1990 housing was distributed proportionately to the 2000 housing.

Like the aggregation approach, our first step in the dasymetric procedure was to identify LCD polygons. Like the areal weighting method, we used the 2000 TIGER as our output geography instead of the LCD polygons. If the LCD polygons represented a 1–1 relationship, no processing was necessary because the houses at both census dates were distributed within the same geography.

If the relationship was 1–M (the most frequent scenario), we used the housing distribution across the component zones in 2000 to redistribute 1990 houses across the same zones. The first step in this process was to sum the extensive variable (i.e. count of houses) over the *n* 2000 component zones (*b*) within each LCD polygon to get *T*, where:

$$T_b = \sum_{i=1}^{n} b_i \tag{1}$$

Next, we let *a* represent the extensive variable (i.e. count of houses) in the LCD polygon for 1990, and p_i be the proportion b_i / T_b. Then, for each of the component zones in the LCD polygon, the apportioned value of housing count for 1990 (again, maintaining the 2000 geography) is:

$$c_i = ap_i \tag{2}$$

For the M–1 relationship, we simply summed the extensive variable (i.e. count of houses) in 1990 over the LCD polygon to maintain compatible geography with 2000.

In the M–M case (figure 6), $\geqslant$2 zones in both 1990 and 2000 had to be aggregated; however, our dasymetric method was similar to that for the 1–M relationship except that we first summed the extensive variable for the 1990 components in the LCD polygon. Therefore, we let *a* represent the extensive variable referenced to the *n* zones in A composing an LCD polygon. We let *b* represent the extensive variable referenced to the *m* zones in B composing the same LCD polygon. The total expected value of *a* for the LCD polygon is:

$$T_a = \sum_{i=1}^{n} a_i \tag{3}$$

and for *b*

$$T_b = \sum_{i=1}^{m} b_i \tag{4}$$

Now, we let the apportioned value $c_i = T_a p_i$.

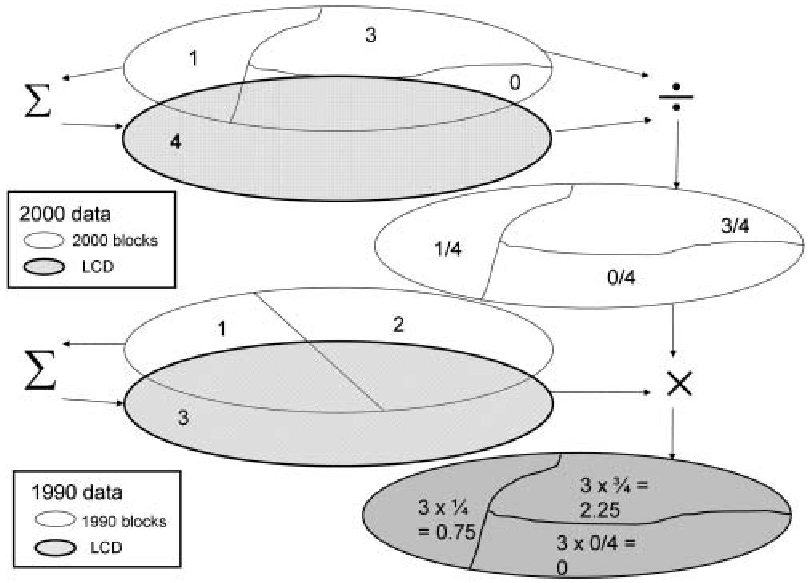

Figure 6. Conceptual diagram of the dasymetric method in which 1990 housing density is allocated into 2000 census block boundaries according to the distribution of 2000 housing density. The houses in each unit of the finer-resolution 2000 boundaries contained within the LCD (least common denominator, or smallest combined area between 1990 and 2000 census block boundaries) were divided by the sum of houses in the 2000 LCD to determine their proportional distribution. The total number of houses in the 1990 LCD was then multiplied by these proportions to distribute the 1990 housing density into the 2000 boundaries.

In some cases p_i may be undefined because $T_b=0$. This happens when the extensive variable in 2000 totals 0 in a given LCD polygon. In such a case, one could set the apportioned value $c_i=0$ or $c_i=T_a$. For density calculations, an LCD polygon's area never includes water block areas. In our work, when the value referenced to a zone in 2000 is 0 in an LCD polygon that is not water, $c_i=T_a$. If the zone is water, then $c_i=0$.

3. Results

3.1 *No processing and areal weighting*

The distribution of census block sizes was strongly skewed, and the vast majority of census blocks in both 1990 and 2000 were smaller than 250 m^2 (0.025 km^2). However, the resolution of the 2000 blocks was much finer than that of the 1990 blocks, with more than 58,300 blocks smaller than 250 m^2 in 2000 compared to approximately 46,500 such blocks in 1990 (figure 7). Although the data resolution for the 1990 housing density map produced through areal weighting was finer than the original 1990 census boundary map, the maps look very similar (figure 3A and B).

With no processing and with areal weighting, a large area went from >0–2 housing units/km^2 in 1990 to 0 housing units/km^2 in 2000. With no processing, 36% of the land in the state showed an apparent housing density change during the decade, of which 75% was due to a housing density decline. With areal weighting, 38% of the landscape changed with 80% due to a housing density decline. Much of

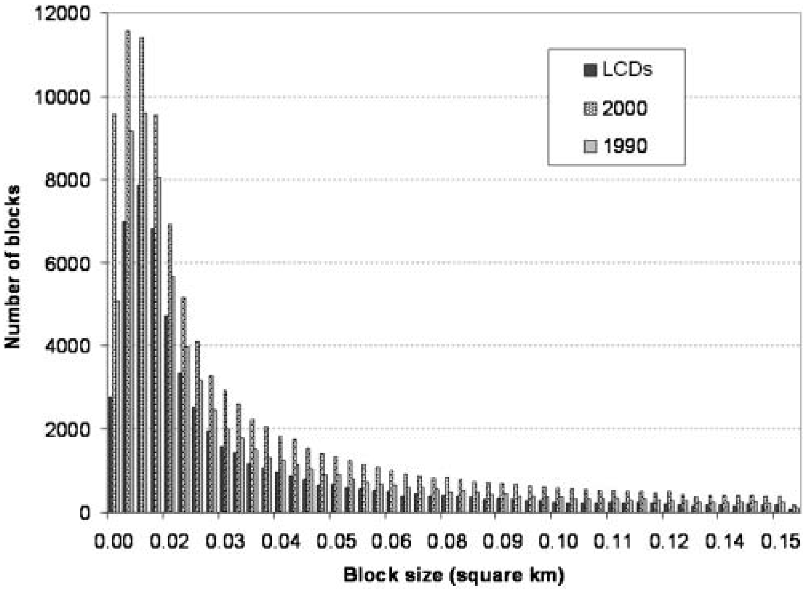

Figure 7. Histogram showing the number of census blocks within different size ranges based on three levels of resolution: (A) least common denominator (LCD) polygons, (B) 2000 census boundaries, and (C) 1990 census boundaries. The histogram is truncated to show only the small block sizes, which constitute the majority of blocks for all datasets.

this apparent decline occurred in the western part of the state in areas interspersed with areas of growth (figure 8). For both methods, there was minimal change in either direction among higher housing density classes (figure 9).

3.2 *Aggregation*

Although the 1990 census block boundaries were coarser than the 2000 boundaries, the boundaries used for the 1990 housing density data for the aggregation method (the LCD polygons) were coarser than the 1990 data. For the LCD polygons, only approximately 35,100 blocks were smaller than 250 m^2, compared to 46,500 blocks in 1990 and 58,300 blocks in 2000 (figure 7). Like the areal weighting method, the 1990 map showing results of the aggregation procedure looks very similar to the original 1990 census block boundaries map, despite the differences in resolution (figure 3C).

Because the 2000 map resulting from the aggregation method was also at the resolution of the LCD polygons, there was much less indication of change between 1990 and 2000 than there was with no processing or areal weighting. Apparent housing density decline was substantially minimized with aggregation, particularly in the eastern part of the state. However, there was also a strong expansion in the area where housing density increased (figure 8). The aggregation method indicated a housing density change in 11% of the land in the state from 1990 to 2000. Of this change, only 4% was decline from >0–2 to 0 housing units/km^2. Approximately 60% of the land in the state stayed in the >0–2 housing units/km^2 class from 1990 to 2000; a relatively small area changed from one density class to another using this method (figure 9).

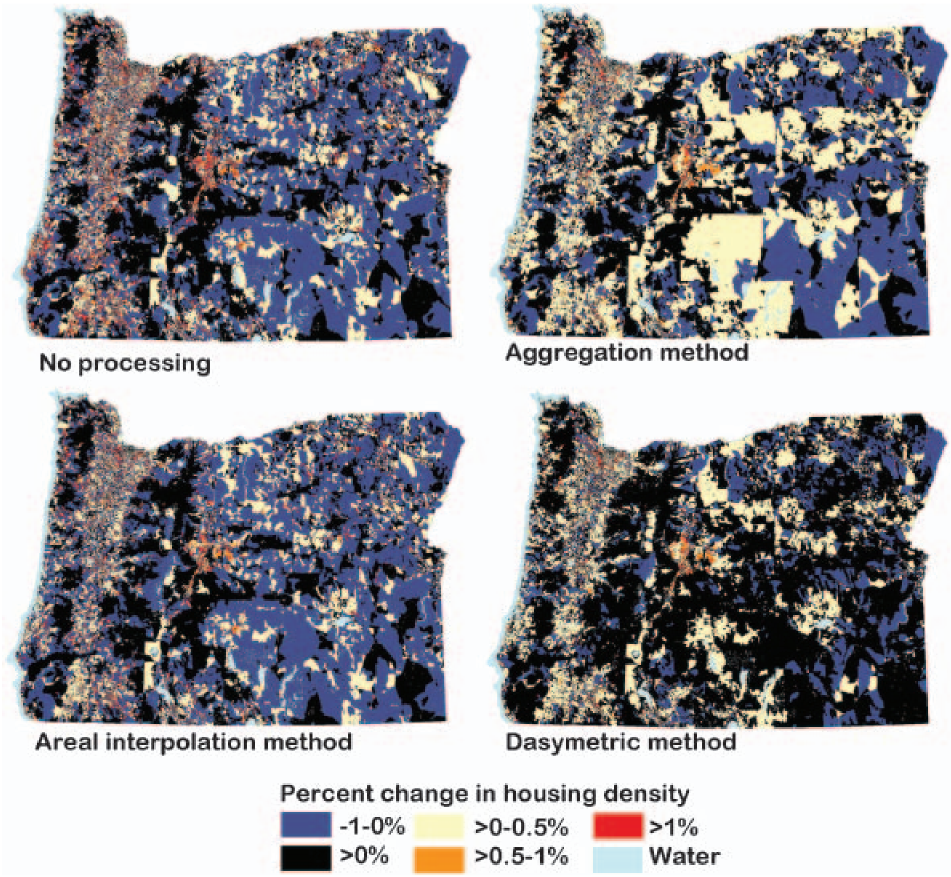

Figure 8. Percentage change in housing density in Oregon between 1990 and 2000 using four methods: (A) no processing, (B) areal weighting, (C) aggregation, and (D) dasymetric. Change in the 1–0% class indicates that housing density declined from 1990 to 2000; the 0% class indicates that housing density remained the same over time; and the remaining classes indicate different percentage increases in housing density from 1990 to 2000.

3.3 *Dasymetric*

Like the areal weighting method, the dasymetric procedure redistributed the 1990 housing units to the 2000 census blocks so there were approximately 58,300 blocks smaller than 250 m^2, compared to 46,500 blocks in 1990 and 35,100 blocks using the LCDs as boundaries (figure 7). The 1990 housing density map produced through dasymetric estimation looked different from the other 1990 maps, with substantially more area in the 0 housing units/km^2 class (figure 3D). This 1990 map looked the most similar to the 2000 census block map.

Like the aggregation method, the dasymetric procedure indicated minimal change between 1990 and 2000. In fact, the dasymetric method was more conservative than aggregation; less area increased in housing density (figure 8). Only 9% of the land in the state experienced either an increase or a decrease in housing density from 1990 to 2000 using the dasymetric procedure, and of that change, 4% was due to housing density decline. Whereas approximately 60% of the land stayed in the >0–2 housing units/km^2 class from 1990 to 2000 using aggregation, approximately 30% of the land

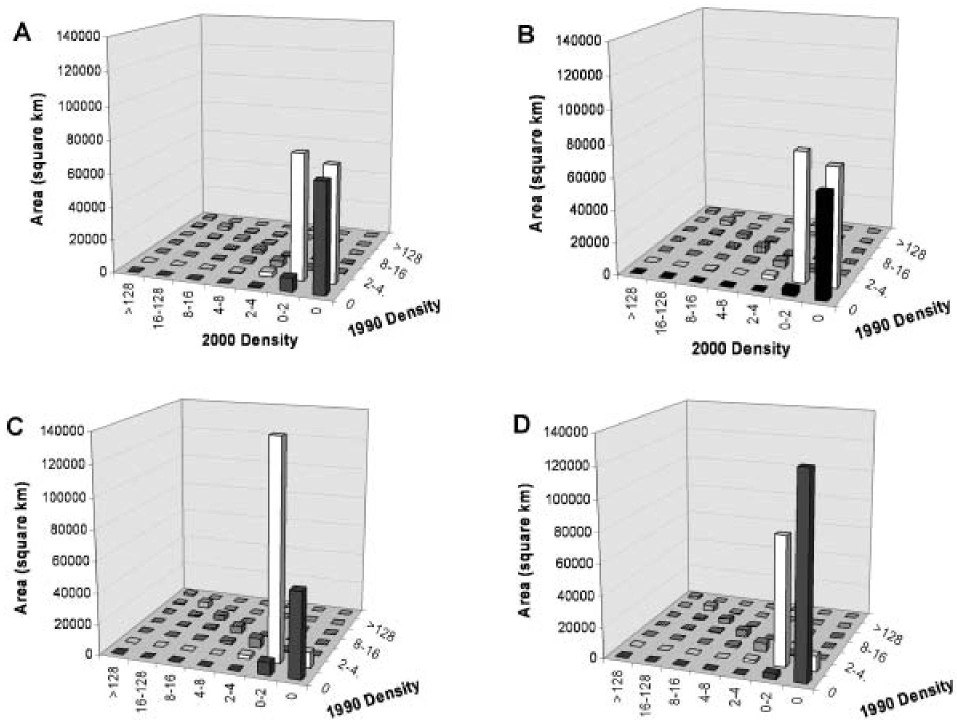

Figure 9. Transition matrices representing how much area changed from 1990 housing density classes to 2000 housing density classes using four methods: (A) no processing, (B) areal weighting, (C) aggregation, and (D) dasymetric procedure.

stayed in that class using the dasymetric method, while more than 50% of the land stayed in the 0 housing units/km^2 class (figure 9).

4. Discussion

Estimates of change in housing density varied in the state of Oregon, showing substantial sensitivity to the method of data interpolation used to reconcile 1990 and 2000 census block boundaries. All methods involved some level of trade-off, including analytical difficulty, data resolution, magnitude or bias in direction of change. It is important for analysts to recognize that housing density change estimates are sensitive to the method of interpolation and to choose between the trade-offs in accordance with their specific objectives and research questions.

The resolution, completeness, and accuracy of US Census data is consistently improving; one of the advantages of the 2000 census data over the 1990 data is the finer resolution of block boundaries (Krieger 2006, Hammer *et al.* 2007). Therefore, from the perspective of data resolution, the areal weighting method provided advantages over the no processing alternative because the 1990 data were redistributed into the finer-resolution boundaries of the 2000 data. Areal weighting was also relatively easy to calculate. With an area as large as the state of Oregon, a simple method greatly enhances efficiency. Nevertheless, errors in housing density change estimates can become substantial across such a large area.

Despite the improvement in resolution stemming from the 2000 census block improvements, the widespread housing density decline observed using areal weighting was likely an artifact of the processing method rather than actual loss in housing. Widespread housing density decline is unrealistic, except for cases of, for example, natural disasters, the effects of which tend to be limited in space and/or time (e.g. wildfires and hurricanes are almost always followed by rebuilding). Recent analyses have shown that the only places where sustained, substantial loss of housing in the US occurs are in areas of widespread property abandonment at the center of large urban areas that are decentralizing (Lee and Leigh 2005); these areas affect a very small proportion of the landscape compared to areas where houses remain or are becoming more numerous. Urban core property abandonment is even more unlikely in Oregon because the state encourages new development within city centers (Wassmer and Baass 2006).

The dramatic housing density decline indicated by areal weighting is likely due to the increase in spatial resolution of the 2000 blocks combined with the assumption that housing data were homogenously distributed within the source and target zones. For example, some of the 2000 block boundaries may have been updated to delineate small neighborhoods located within large, sparsely populated areas. Thus, when a large polygon with low housing density in 1990 was split into two polygons in 2000, one of the 2000 polygons would be comparatively densely populated with houses, whether or not there was any change in houses on the ground, while the other would have no houses in the area that was formerly represented as populated with low density housing. An overlay of the 1990 and 2000 boundaries in the second polygon would therefore show a false decline in housing density (when in reality, it was probably undeveloped in both 1990 and 2000).

The primary advantage of using the aggregation method was that housing density change estimates were much more conservative than those observed in areal weighting. Minimal housing density decline occurred because aggregating the boundaries allowed us to directly analyze the data through a 1–1 relationship, and data did not have to be interpolated from one resolution into another. Aggregation was also relatively straightforward in terms of analytical complexity.

Although aggregation avoided the drawbacks of interpolating over census blocks of different sizes, the weakness of the approach was its coarse resolution. The substantial expansion of the 0–50% growth category may be an artifact of summarizing and comparing housing data over areas that exceeded the size of the census boundaries. In other words, increases in housing density that really only occurred in smaller areas were attributed to the entire extent of the LCDs. Therefore, while aggregation did produce a much more conservative estimate of change than areal weighting, the distortions introduced by the coarse resolution may affect the results of the change analysis and limit its utility where grain size is important.

Unlike the other two methods of interpolation, the dasymetric approach allowed us to calculate conservative estimates of housing change within the fine-resolution boundaries of the 2000 census blocks. This approach actually capitalized on the uniquely persistent nature of housing development and enabled us to assume that the 1990 housing density was distributed proportionally to the 2000 housing density. When population data are compared over time, the distributions of the initial and final populations across a newly divided block (1–M) are often independent because people are dynamic and migrate. Houses, however, remain in place over time; their final distribution depends on their initial distribution. Therefore, although processing housing data for the dasymetric approach was slightly more complex than it was for the other reconciliation methods, the bootstrapping approach of using data internal to the process of investigation kept the methods much simpler than if we were working with population or its characteristics.

One special feature of census data that facilitated both the aggregation and the dasymetric interpolation methods was the limit on the extent of the 1990–2000 LCD polygons. All census data below the county level sums to the county level, so for any census reconciliation at a multi-county scale, the county is the largest possible LCD polygon (assuming counties do not change, which is almost always true). If there were no shared, common boundaries limiting the possible extent of aggregation necessary to find the LCD polygons, our methods for interpolating between census boundaries would be more error-prone, and even impractical. However, the accuracy of areal interpolation methods is generally better with smaller source and target zones that are similar in shape (Sadahiro 2001). Therefore, interpolation across finer-scale census boundaries should be expected to produce more reliable change analyses.

In most states, census boundaries will probably increase in resolution, reflecting continuing improvement in the accuracy and precision of housing counts over time (Krieger 2006), as they did in most of the state of Oregon between 1990 and 2000. Therefore, the dasymetric method of redistributing census data from an earlier year based on the housing distribution of more current years should generally be an appropriate method for census boundary reconciliation across the country. However, users should be careful if more relationships in their areas of analysis are M–1 (i.e. the resolution becomes coarser over time). Another consideration is that, because census boundaries are usually altered to account for new housing growth, it is inevitable that the areas that change the most are often also those most prone to error due to interpolation across incompatible zones. This artifact of data collection compounds the problem of trying to accurately estimate and evaluate change over time. However, boundary changes in the US Census from 1990 to 2000 may be an exception to this rule because many changes in census boundaries occurred in areas with very little housing growth, as was the case in central and eastern Oregon. In other words, there was widely dispersed growth across the vast majority of the state, but in most cases, the magnitude of change was not as substantial as the amount of area that changed. One reason for these widespread changes may be that the 1990 Census was the first to be released in GIS format, and many improvements were made to increase the resolution for the 2000 Census data; so the 2000 Census represented a major improvement. This may also explain why the dasymetric approach was particularly useful for these data because it minimized potential errors in calculation across such a large geographic extent.

Housing development trends are expected to continue, and the amount of developed land in the United States is expected to increase by 79% in the next 25 years (Alig *et al.* 2004). Therefore, it will continue to be important for scientists to understand where and how housing growth is occurring so that land-use planners and conservation biologists can anticipate how to best develop comprehensive regional plans (Lenth *et al.* 2006). All of our methods involved some level of trade-off, and housing change estimates were sensitive to the method of interpolation. Although boundary resolution and ease of processing are important factors to consider when choosing a method of interpolation, the dasymetric bootstrapping approach may be particularly useful for assessing housing growth across changing census boundaries because it provides the most conservative estimate of change and minimizes issues related to apparent housing density decline when census units become finer over time.

Acknowledgments

We gratefully acknowledge support by the Northern Research Station of the US Forest Service and the Forest Inventory and Analysis program of the Pacific Northwest Research Station of the US Forest Service for this research. Two anonymous reviewers and associate editor Dr M. Yuan provided very thorough and helpful comments that greatly improved our manuscript.

References

Alig, R.J., Kline, J.D. and Lichtenstein, M., 2004, Urbanization on the US landscape: looking ahead in the 21st century. *Landscape and Urban Planning*, **69**, pp. 219–234.

Brown, D.G., Johnson, K.M., Loveland, T.R. and Theobald, D.M., 2005, Rural land-use trends in the conterminous United States, 1950–2000. *Ecological Applications*, **15**, pp. 1851–1863.

Chen, K., McAneney, J., Blong, R., Leigh, R., Hunter, L. and Magill, C., 2004, Defining area at risk and its effect in catastrophe loss estimation: a dasymetric mapping approach. *Applied Geography*, **24**, pp. 97–117.

Crump, J.R., 2003, Finding a place in the country. *Environment and Behavior*, **35**, pp. 187–202.

Downs, A., 1999, Some realities about sprawl and urban decline. *Housing Policy and Debate*, **10**, pp. 955–974.

Flowerdew, R. and Green, M., 1992, Developments in areal interpolation methods and GIS. *Annals of Regional Science*, **26**, pp. 67–78.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N. and Snyder, P.K., 2005, Global consequences of land use. *Science*, **309**, pp. 570–574.

Goodchild, M.F. and Lam, N.S.-N., 1980, Areal interpolation: a variant of the traditional spatial problem. *Geo-Processing*, **1**, pp. 297–312.

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

Gregory, I.N., 2002, The accuracy of areal interpolation techniques: standardising 19th and 20th century census data to allow long-term comparisons. *Computers, Environment and Urban Systems*, **26**, pp. 293–314.

Gregory, I.N. and Ell, P.S., 2006, Error-sensitive historical GIS: identifying areal interpolation errors in time-series data. *International Journal of Geographical Information Science*, **20**, pp. 135–152.

Hammel, D. and Wyly, E.K., 1996, A model for identifying gentrified areas with census data. *Urban Geography*, **17**, pp. 248–268.

Hammer, R.B., Radeloff, V.C., Fried, J.S. and Stewart, S.I., 2007, Wildland-Urban Interface growth during the 1990s in California, Oregon and Washington. *International Journal of Wildland Fire*, **16**, pp. 255–265.

Hay, S., Noor, A., Nelson, A. and Tatem, A., 2005, The accuracy of human population maps for public health application. *Tropical Medicine and International Health*, **10**, pp. 1073–1086.

Heidkamp, C.P. and Lucas, S., 2006, Finding the gentrification frontier using census data: the case of Portland, Maine. *Urban Geography*, **27**, pp. 101–125.

Jelinski, D.E. and Wu, J., 1996, The modifiable areal unit problem and implications for landscape ecology. *Landscape ecology*, **11**, pp. 129–140.

KLINE, J.D., 2006, Public demand for preserving open space. *Society and Natural Resources*, **19**, pp. 645–659.

KNOWLES, A.K., 2005, Emerging trends in historical GIS. *Historical Geography*, **33**, pp. 134–158.

KRIEGER, N., 2006, A century of census tracts: health & the body politic (1906–2006). *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, **83**, pp. 355–361.

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.

LEE, S. and LEIGH, N.G., 2005, The role of inner ring suburbs in Metropolitan smart growth strategies. *Journal of Planning Literature*, **19**, pp. 330–346.

LENTH, B.A., KNIGHT, R.L. and GILBERT, W.C., 2006, Conservation value of clustered housing developments. *Conservation Biology*, **20**, pp. 1445–1456.

LIU, J., DAILY, G.C., EHRLICH, P.R. and LUCK, G.W., 2003, Effects of household dynamics on resource consumption and biodiversity. *Nature*, **421**, pp. 530–533.

LUBOWSKI, R.N., VESTERBY, M., BUCHOLTZ, S., BAEZ, A. and ROBERTS, M.J., 2006, Major land uses in the United States, 2002. United States Department of Agriculture Economic Research Service, Economic Information Bulletin Number 14, 54 pp., Electronic Report. Available online at: http://www.ers.usda.gov/publications/EIB14/eib14.pdf.

MARTIN, D., 2003, Extending the automated zoning procedure to reconcile incompatible zoning systems. *International Journal of Geographical Information Science*, **17**, pp. 181–196.

MARTIN, D., TATE, N.J. and LANGFORD, M., 2000, Refining population surface models: experiments with Northern Ireland census data. *Transactions in GIS*, **4**, pp. 343–360.

MCKINNEY, M.L., 2002, Urbanization, biodiversity, and conservation. *BioScience*, **52**, pp. 883–890.

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

MILLER, J. and HOBBS, R.J., 2002, Conservation where people live and work. *Conservation Biology*, **16**, pp. 330–337.

OPENSHAW, S., 1984, The modifiable areal unit problem. CATMOG 38. GeoBooks, Norwich, England.

PEW CENTER FOR CIVIC JOURNALISM, 2000, Straight talk from Americans 2000: a national survey for the Pew Center for Civic Journalism conducted by Princeton Survey Research Associates. College Park, MD: Pew Center for Civic Journalism.

RADELOFF, V.C., HAMMER, R.B., VOSS, P.R., HAGEN, A.E., FIELD, D.R. and MLADENOFF, D.J., 2001, Human demographic trends and landscape level forest management in the northwest Wisconsin Pine Barrens. *Forest Science*, **47**, pp. 229–241.

RADELOFF, V.C., HAMMER, R.B., STEWART, S.I., FRIED, J.S., HOLCOMB, S.S. and MCKEEFREY, J.F., 2005, The wildland-urban interface in the United States. *Ecological Applications*, **15**, pp. 799–805.

REIBEL, M. and BUFALINO, M., 2005, Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environment and Planning A*, **37**, pp. 127–139.

SADAHIRO, Y., 2000, Accuracy of count data transferred through the areal weighting interpolation method. *International Journal of Geographical Information Science*, **14**, pp. 25–50.

SALA, O.E., CHAPIN, III, F.S., ARMESTO, J.J., BERLOW, R., BLOOMFIELD, J., DIRZO, R., HUBER-SANWALD, E., HUENNEKE, L.F., JACKSON, R.B., KINZIG, A., LEEMANS, R., LODGE, D., MOONEY, H.A., OESTERHELD, M., POFF, N.L., SYKES, M.T., WALKER, B.H., WALKER, M. and WALL, D.H., 2000, Global biodiversity scenarios for the year 2100. *Science*, **287**, pp. 1770–1774.

SHOCHAT, E., WARREN, P.S., FAETH, S.H., MCINTYRE, N.E. and HOPE, D., 2006, From patterns to emerging processes in mechanistic urban ecology. *Trends in Ecology and Evolution*, **21**, pp. 186–191.

STEIF, K., 2004, Creating a model for geodemographic representations of housing market activity: a research note with possible public policy implications. *Middle States Geographer*, **37**, pp. 116–121.

THEOBALD, D.M., 2001, Land-use dynamics beyond the urban fringe. *Geographical Review*, **91**, pp. 544–564.

THEOBALD, D.M., 2004, Placing exurban land-use change in a human modification framework. *Frontiers in Ecology and the Environment*, **2**, pp. 139–144.

THEOBALD, D.M., HOBBS, N.T., BEARLY, T., ZACK, J.A., SHENK, T. and RIEBSAME, W.E., 2000, Incorporating biological information in local land-use decision making: designing a system for conservation planning. *Landscape Ecology*, **15**, pp. 35–45.

WASSMER, R.W. and BAASS, M.C., 2006, Does a more centralized urban form raise housing prices? *Journal of Policy Analysis and Management*, **25**, pp. 439–462.

YANG, R. and JARGOWSKY, P.A., 2006, Suburban development and economic segregation in the 1990s. *Journal of Urban Affairs*, **28**, pp. 253–273.