DASYMETRIC MAPPING TECHNIQUES FOR THE SAN FRANCISCO BAY REGION, CALIFORNIA

Rachel Sleeter
U.S. Geological Survey
Western Geographic Science Center
345 Middlefield Rd.
Menlo Park, CA. 94025
rsleeter@usgs.gov

Abstract: Demographic data are commonly represented by using a choropleth map, which aggregates the data to arbitrary areal units, causing inaccuracies associated with spatial analysis and distribution. In contrast, dasymetric mapping takes quantitative areal data and attempts to show the underlying statistical surface by breaking up the areal units into zones of relative homogeneity. This thesis applies the dasymetric mapping method to the 1990 U.S. Census block-group populations of Alameda County, California using the U.S. Geological Survey's 1992 National Land Cover Data Set and other ancillary land-cover sources to redistribute the block-group populations into a 30-m grid based on categorical zones relative to population distribution. To test the accuracy of the dasymetric approach, census block populations were compared with the dasymetric mapping distributions; the results yield high correlation coefficients (between 0.80-0.88), indicating that the dasymetric mapping method produced more accurate population distributions than the choropleth method relative to the census block.

INTRODUCTION

Demographic data and socioeconomic information are commonly displayed cartographically using choropleth mapping techniques. For example, the choropleth map is used to display U.S. Census data, a geographic standard for demographics, and is used as a medium by virtually all geographers and many nongeographers (Slocum and Egbert 1993). The choropleth map spatially aggregates data into geographic areas or areal units (e.g., county, census tract, block group, etc.). If the spatial units are too large, the data's spatial variation tends to be reduced or averaged out. Because the value in the enumeration unit is spread uniformly throughout the areal unit, continuous geographic phenomena cannot be displayed. Dorling (1993) noted that choropleth maps of population by administrative areal unit give the notion that population is distributed homogeneously throughout each areal unit, even when proportions of the region are, in reality, uninhabited. This discrepancy is greatest in areas with mixed urban, undeveloped, and agricultural land uses.

The U.S. Census Bureau (2001) collects demographic data at the block geographic level. Census blocks are areas bounded on all sides by visible features, such as streets, roads, streams, and railroad tracks, and by invisible boundaries, such as city, and county limits. Generally, census blocks are small in area; for example, a block bounded by city streets. However, census blocks in sparsely populated areas may be large and contain many square miles. The block group is the next

geographic level of U.S. Census delineations, consisting of a cluster of census blocks generally containing from 600 to 3,000 persons (U.S. Census Bureau 2001). The population of each block group is an aggregate of the cluster of blocks. The boundaries of the census delineations are chosen on the basis of linear features and administrative boundaries, causing discrepancies between the enumeration units and the relevancy of population distributions.

Dasymetric mapping is a potential solution for the dilemma of portraying population data that have been aggregated to areal units. Eicher and Brewer (2001:125) stated that "Dasymetric mapping depicts quantitative areal data using boundaries that divide the mapped area into zones of relative homogeneity with the purpose of best portraying the underlying statistical surface." This type of mapping has been described as an intelligent approach to choropleth mapping in an attempt to improve area homogeneity. Thus, new zones are created that directly relate to the function of the map, which is to show spatial variations in population density. Land-cover data can indicate residential areas for the delineation of new homogeneous zones. The census block-group populations can be redistributed to the new zones, resulting in a more accurate portrayal of where people live within an administrative boundary.

This study explores a surface-based representation of population, using a dasymetric mapping technique that incorporates land-cover classifications as a means to redistribute the original census block-group population value into a surface grid based on levels of urbanization and undeveloped land. Through areal interpolation, the distribution is depicted semicontinuously, where multiple datasets redefine the populated surface. We hypothesize that this method will provide a more accurate representation of where people live in Alameda County, Calif., within a given block group than would choropleth maps of the same area. The greatest improvements in the accuracy of population-density values should be seen in block groups with various land-cover types and significant amounts of undisturbed land. In block groups that are heavily urbanized, the dasymetric mapping method may not show much difference from the choropleth method due to the smaller size of block groups and the better correlation of the block-group boundaries in these areas to the actual population distribution.

To determine the accuracy of the dasymetric mapping technique, we analyzed 1990 U.S. Census blocks to see how closely the population density of the urbanization zones by block group match the populations of the census block. We hypothesize that the block groups on the dasymetric map will show a statistically superior match to the census-block populations over those on the choropleth map.

Dasymetric-Mapping Approaches

An essential step in dasymetric mapping is the creation of zones within the areal unit that correspond to the variable being mapped. To create intraunit zones of relative homogeneity among population, ancillary data must be used to interpret relative levels of habitation. Past approaches have focused on using ownership records, topography, and land-cover classifications to identify and mask uninhabited areas. Holloway et al. (1997) used multiple datasets to detect and remove uninhabited lands from the area of analysis. Four types of area were ruled out, including census blocks with zero population, all lands owned by local, State, or Federal government, all corporate timberlands, and all water or wetlands, as well as all open and wooded

areas with elevation data that have a slope of at most 15% (Holloway et al. 1997). To redistribute the census population to the ancillary feature classes, a predetermined percentage was assigned to each class. The subjectivity and accuracy of this percentage assignment (e.g., 80% of the population to urban polygons, 10% to open polygons, and 5% to agriculture and wooded polygons) can be argued because of the absence of empirical evidence.

In contrast, Mennis (2003) used a three-tier raster classification of urban land cover derived from the Landsat Thematic Mapper as ancillary data. Within the remotely sensed land-cover data, urban features were put into three classes of high density, low density, and nonurban, with no distinction of wooded areas, agriculture, or slope. Initially, all census data were converted into a 100-m raster surface that was used for areal interpolation. The population statistics derived from the 1990 Census block-group data were distributed via areal interpolation to each 100-m grid cell on the basis of two factors: “the relative difference in population densities among the three urbanization classes…and the percentage of total area of each block group occupied by each of the three urbanization classes. (Mennis 2003:36)” An empirical sampling of population density between urbanization classes helps determine what percentage of the census block-group population should be assigned to each urbanization class. Also, an area-based weighting addresses the relative differences between each urbanization class within the census block group.

Data Models and Representation

The relative merit of object versus field models for quantitative representation is a subject of ongoing debate in the fields of cartography and geography. Michael Goodchild (1992) has written extensively about object versus field models in a geographic information system (GIS). In the object model, features are generally represented as points, lines, or polygons, and so this mode is known as the vector data model. The field model, which typically represents square features as a set of uniform-sized cells, is known as the raster data model. The advantages and disadvantages for visualization and quantitative representation in both models have become evident and depend on the scale and quality of the data. Mennis (2003) determined that a field representation of population data, where the data are modeled onto a continuous surface, works well with the transformation of population data from census block groups.

In our study, we used a combination of the object and field models. If the object represents the same area as the field, then the two models can be used interchangeably. For example, in a vector representation of points where each point represents 30m, the points can be converted to 30m pixels without a loss of data. This conversion meets the ideal of the pycnophylactic property (Tobler 1979), where “the summation of population data to the original set of areal units is preserved in the transformation to a new set of areal units. (Mennis 2003:32)” Therefore, the Modifiable Areal Unit Problem is avoided during the areal interpolation.

METHODS

Study Area

The Alameda County, Calif., study area, which is part of the greater San Francisco Bay region was chosen because we sought an area with which we were already familiar to enhance the depth of understanding that can be brought to the spatial relations being analyzed (Figure 1). As Eicher and Brewer (2001:125) pointed out: “The cartographer generates dasymetric zones by using ancillary information. This information can be both objective and subjective, depending on other available data and the cartographer’s knowledge of the area.” Furthermore, Alameda County has a widely varying topography and a mix of land-cover types from undeveloped and agricultural to heavily urbanized. Alameda County is home to the city of Oakland, the eighth-largest city in California, with a 2000 U.S. Census population of 399,484, and the Port of Oakland, one of the major container ports on the west coast of the United States. Many city centers within Alameda County contribute to the greater San Francisco Bay region, such as Livermore, Pleasanton, Berkeley, Alameda, Hayward, and Fremont. Conversely, the network of East Bay Regional Parks snakes through Alameda and neighboring Contra Costa counties, preserving 94,500 acres of open space. Although the west side of Alameda County is heavily populated and urbanized, the land cover changes drastically toward the southeast into rugged hills and toward the east into the agricultural landscape of the California Central Valley. Alameda County, which typifies the urban-rural fringe of the San Francisco Bay region, is an important but complex area to understand and portray demographically.

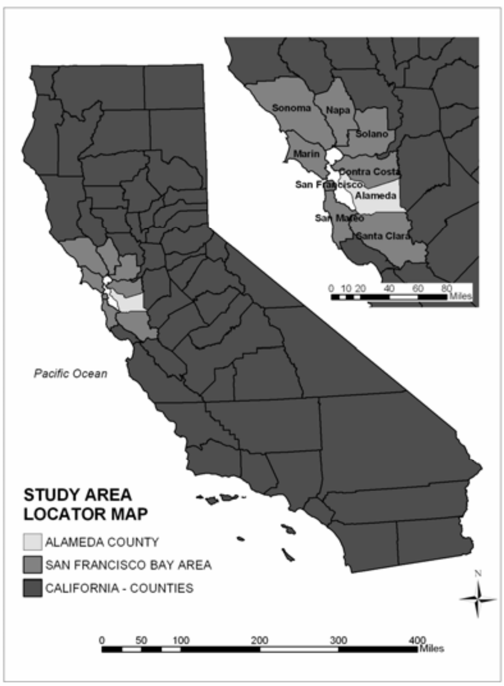

Figure 1.California, showing location of Alameda County study area.

Data Preparation

Our approach combined the methodologies of Mennis (2003) and Holloway et al. (1997) by choosing four land-cover classes, using a three-tier urbanization classification and adding an excluded class representing zero population. The 21 class National Land Cover Data Set (NLCD) (Volgelmann et al. 2000) were recoded into four classes; high-intensity residential, low-intensity residential, nonurban, and water: the nonurban class consists of all remaining 18 classes, representing all lands that are not residential but may be sparsely populated. The undeveloped layer, contributed by GreenInfo Network (2003), incorporates lands that have some recreational, open-space, habitat-protection, or agricultural-protection value in the San Francisco Bay region. These lands either are owned by a public agency or a nongovernmental organization (NGO), or have an easement on them held by a public agency or an NGO. The uninhabited layer was merged with the recoded NLCD layer to produce a comprehensive land-cover layer. From this

dataset, the classes were merged and reconfigured into classes of high-intensity residential, low-intensity residential, nonurban and exclusion; the exclusion class is a combination of all water and undeveloped areas (Table 1). The advantage of incorporating an exclusion class is to more accurately display population density by weeding out large areas of the areal interpolation, allowing the visual depiction of population to be strictly within those areas that are actually populated.

Class Code	Class Definition for Recoding	
0	No data	
1	High intensity	
2	Low intensity	
3	Nonurban	
4	Water	
10	Undeveloped	
11	Undeveloped + high intensity	
12	Undeveloped + low intensity	← Classes 10-14 recoded to class code 4 after summation
13	Undeveloped + nonurban	
14	Undeveloped + water	

Table 1. Combining ancillary land-cover layers for final re-coding

After the recoding process, the new zones of relative homogeneity were in a raster format. Before the areal interpolation, the raster land-cover data were converted to points for two reasons: (1) to provide an easy way to spatially join the land-cover dataset to the census block data and (2) to efficiently create and calculate new fields in a tabular structure. As mentioned above, this method allows us to use both raster and vector data interchangeably.

At this stage, each point representing a 30-m pixel has an associated land-cover code but no census block-group information. To attach the census data, each point also needs an associated block-group identifier. This step is required for our dasymetric mapping approach because each calculation is performed on a block-group-by-block-group basis. The census polygon data are joined spatially to the land-cover points, and the data are now ready for areal interpolation.

Areal Interpolation

Areal interpolation, which is the process by which data from one set of source polygons are redistributed to another set of overlapping target polygons, is used primarily when a project contains data from various sources covering the same area but with differing internal boundaries. We adapted the four equations from Mennis (2003) (see Appendix I) to address our addition of a zero-population zone or exclusion class. The removal of the spatial area of the exclusion class from the total possible area of population distribution should contribute more accurate results overall relative to Mennis' approach because of the addition of a fourth class covering all areas of zero population, such as water.

To quantify the urban land-cover variable within each subcounty subdivision, a sampling method was used to calculate the relative difference in population density among urbanization classes within each unit (Mennis 2003). Three representative block groups were selected for each urbanization class (block groups that clearly had mostly high-intensity residential, low-intensity residential, or nonurban points within them). The total population and area were calculated for each urbanization-class sample, resulting in an aggregated population density (Table 2).

Urban-Class	*Subcounty*	*Area(mi²)*	*Populated Area (mi²)*	*Population Density (persons/mi²)*	*Sum of Density*	*Population Density Fraction*
High	Oakland	0.0516	1,863	36,104.651	57,150.053	63.18
Low	Oakland	0.080782	1,180	14,607.214	57,150.053	25.56
NonUrban	Oakland	0.0817	526	6,438.188	57,150.053	11.26
High	Berkeley	0.108026	2,389	22,115.046	38,097.222	58.05
Low	Berkeley	0.070769	944	13,339.174	38,097.222	35.01
NonUrban	Berkeley	0.18199	481	2,643.002	38,097.222	6.94
High	Alameda	0.27888	2,715	9,735.37	25,626.627	38
Low	Alameda	0.068526	842	12,287.307	25,626.627	48
NonUrban	Alameda	1.09158	3,934	3,603.95	25,626.627	14
High	Hayward	0.106925	1,127	10,540.0981	19,209.8786	54.87
Low	Hayward	0.129875	1,116	8,592.8777	19,209.8786	44.73
NonUrban	Hayward	5.82553	448	76.9028	19,209.8786	0.4
High	Fremont	0.126942	1,070	8,431.6459	25,341.69808	33.27
Low	Fremont	0.08667	1,436	16,572.40106	25,341.69808	65.4
NonUrban	Fremont	6.35665	2,146	337.65112	25,341.69808	1.33
High	Livermore-Pleasanton	0.179661	1,029	5,727.453371	13,911.78012	41.17
Low	Livermore-Pleasanton	0.105797	865	8,176.035237	13,911.78012	58.77
NonUrban	Livermore-Pleasanton	12.42234933	103	8.291507285	13,911.78012	0.06

Table 2. Sampled population-density fractions for the subcounties

The population density fraction was then calculated for each urbanization class within each subcounty. This fraction indicates the proportion of the block-group population that should be assigned to each urbanization class within a given block group (see Appendix I for equations). To alter the population-density fractions, an area ratio was calculated for each block group, showing the proportion of area that each of the three urbanization classes occupies within a given block group. This calculation was performed on every urbanization class within every block group. At this point, the exclusion class was inserted back into the analysis as the entire areas of each urbanization class within the block-group total were summed into “area ratios” including the exclusion-class points.

Next, the population-density fraction was multiplied by the area ratio to give the fraction of the original block-group population that was distributed to each urbanization class within each block

group; then that result was divided by the sum of the same expression of all three urbanization classes. The total fraction is the underlying solution to the interpolation. Once the total fraction was calculated, part of the original block-group population could be assigned to each point within the block group according to its urbanization class. This calculation was the final step in the areal interpolation, resulting in a surface-based representation of population density (Figure 2). The final distribution was completed by multiplying the total fraction of an urbanization class by the total block-group population and then dividing the result by the number of points within the urbanization class (see Appendix I). The utilization of a point-feature class in a GIS made the areal interpolation efficient because each new field could be created and calculated in a semiautomated mode, using the GIS field calculator.

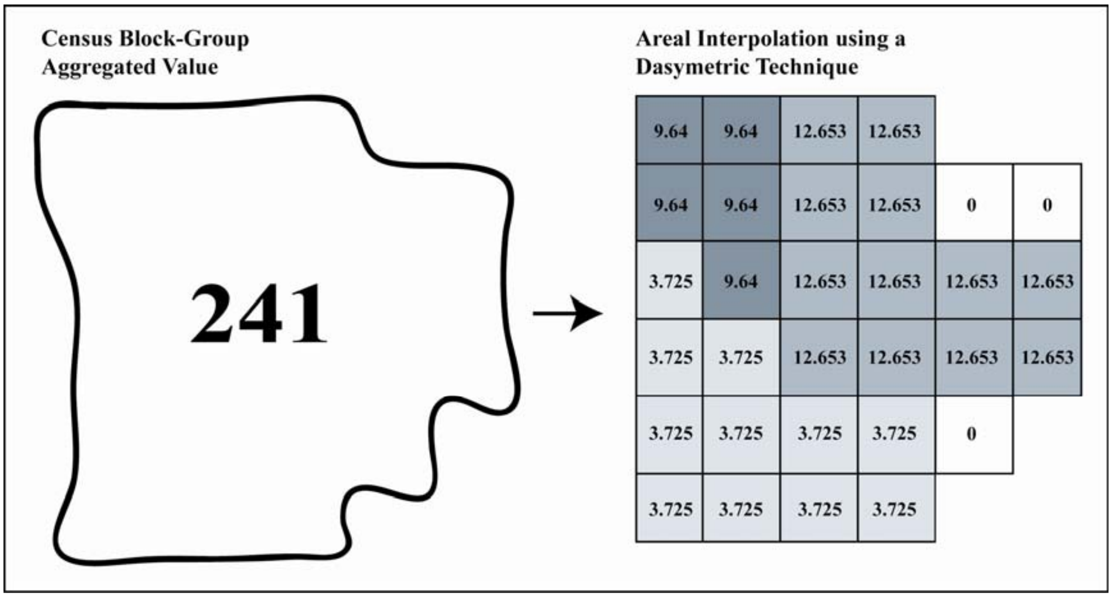

Figure 2. Example of input and output values of areal interpolation. Dasymetric technique uses land-cover urbanization classes (dark gray = high-intensity urban, gray = low-intensity urban, light gray = nonurban, and white = exclusion or zero population), relative area and density of classes within block groups, which are used to redistribute population values

ANALYSIS

Comparing Census Block Populations to Dasymetric Mapping Results

Our hypothesis is that the dasymetric mapping method will more accurately represent where people live within a given block group. To test the precision of the dasymetric map distribution, the census block totals were evaluated as an indicator of how well the population was distributed within the block groups. The dasymetric map consists of 30-m points, each representing a population total for that area. To obtain a value comparable to the block populations, a sum of the dasymetric points that are located within a given block was generated.

Figure 3 is a histogram of the absolute differences between the two variables, calculated by subtracting the block populations from the summation of all the dasymetric points for each block. If the difference between the block populations and the dasymetric comparative values equals zero, the results indicate that the dasymetric map distributions equals the census-block population. The histogram shows an approximately normal distribution, with most of the difference values near zero, affirming our claim that the dasymetric map preserves an accurate block level summation of census population data.

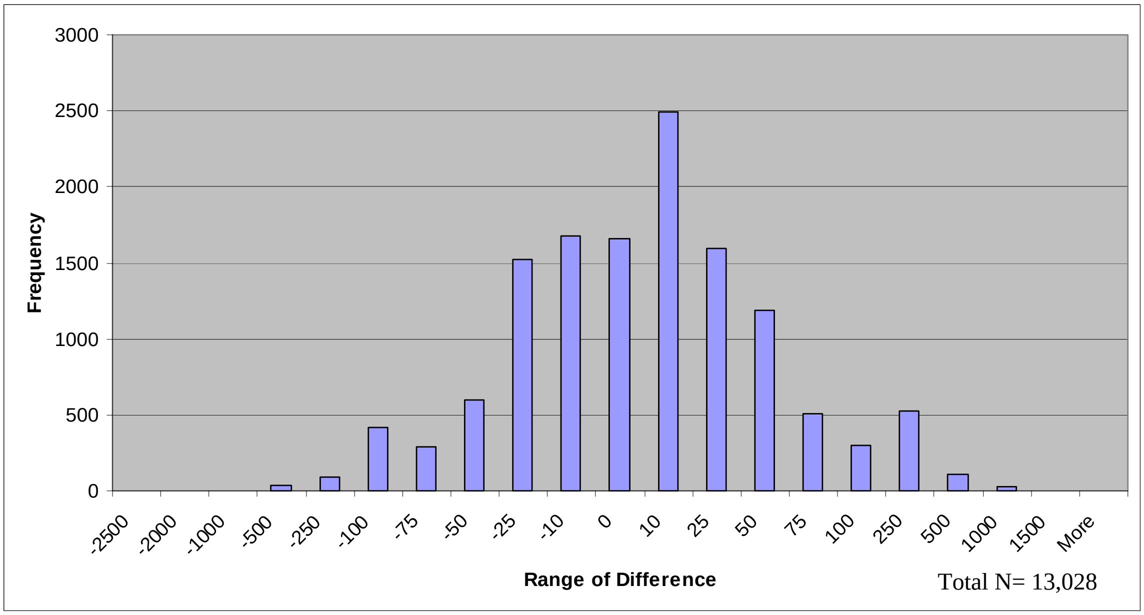

Figure 3. Histogram showing difference between block populations and dasymetric map values for Alameda County, an approximately normal distribution, with most of the difference near zero

Correlation Coefficients

To further test a positive association between the block-population totals and the dasymetric mapping distributions, we conducted a correlation analysis. The correlation coefficient, denoted by *r* of the pairs (*x* , *y*), is calculated as

$$r = \frac{\sum d_x d_y}{\sqrt{\left(\sum d_x^{2} \sum d_y^{2}\right)}}$$

Here the strength of the relation between the estimated dasymetric population per block and the observed block population is tested by using a bivariate or simple correlation analysis (Burt and Barber 1996). Our hypothesis requires a positive correlation between the two arrays, which would indicate that “the relationship between *x* and *y* is such that small values of *y* tend to go with small values of *x* and large values of *y* tend to go with large values of *x*” (Freund and Simon 1997:528). The correlation coefficients for each subcounty are high, ranging from 0.80 to 0.88 (Table 3). This statistic can be interpreted as a standardized measure of areal association and the degree of similarity of the two maps in the individual statistics (Burt and Barber 1996).

Subcounty	
Berkeley	0.84
Oakland	0.82
Alameda	0.88
Hayward	0.87
Fremont	0.87
Livermore_Pleasanton	0.80
Alameda County	0.85

Table 3. Correlation coefficient between block populations and dasymetric populations for each subcounty

We also computed a correlation coefficient to compare the choropleth map of block-level summations derived from block-group population densities to the actual block population densities. The results support our initial hypothesis. The subcounties that scored the lowest in the choropleth-to-block comparison were Alameda, Fremont, and Livermore-Pleasanton (Table 4). These three subcounties also have a lower ratio of urbanized to nonurban and undeveloped land cover (Table 5). Our hypothesis is that the dasymetric mapping method would be more effective in areas with more land-cover variation and less concentrated urbanization as was true for all but one subcounty (Hayward), which is highly urbanized but has some large undeveloped areas. The large undeveloped area in Hayward may have contributed to the lower percentage of residential land cover.

Subcounty	Dasymetric : Block	Choropleth : Block
Berkeley	0.84	0.83
Alameda	0.88	0.67
Oakland	0.81	0.79
Hayward	0.87	0.79
Fremont	0.87	0.56
Livermore-Pleasanton	0.8	0.57
ALAMEDA COUNTY	0.85	0.7

Table 4. Correlation coefficients comparing the dasymetric method to the choropleth method for each subcounty

Subcounties	Residential	Total	RATIO	%Residential
Berkeley	27482	33691	0.815707	81.57
Alameda	14575	36407	0.400335	40.03
Oakland	94215	174206	0.540825	54.08
Hayward	103602	380763	0.272091	27.21
Fremont	94850	338532	0.28018	28.02
Livermore-Pleasanton	62553	1206282	0.051856	5.19
ALAMEDA COUNTY	397277	2169881	0.183087	18.31

Table 5. Area ratio between residential and nonurban/undeveloped by subcounty (in pixels)

Discussion

Figure 3 shows that the deviations between block and block-group totals aggregated by dasymetric mapping techniques were approximately normally distributed, indicating minimal difference between the two datasets. Also, all of the correlation coefficients exceeded 0.80. The correlation coefficients between the choropleth map and the block populations ranged from 0.56

to 0.83, unambiguously lower than for the dasymetric map, confirming our hypothesis that the dasymetric mapping method of representing block-group population density is more accurate than the choropleth mapping method.

The dasymetric map also produces a superior visual enhancement of the data, a fact most evident when focusing on water features (Figure 4). For example, the city of Alameda is an island in the northwestern part of the county that is adjacent to San

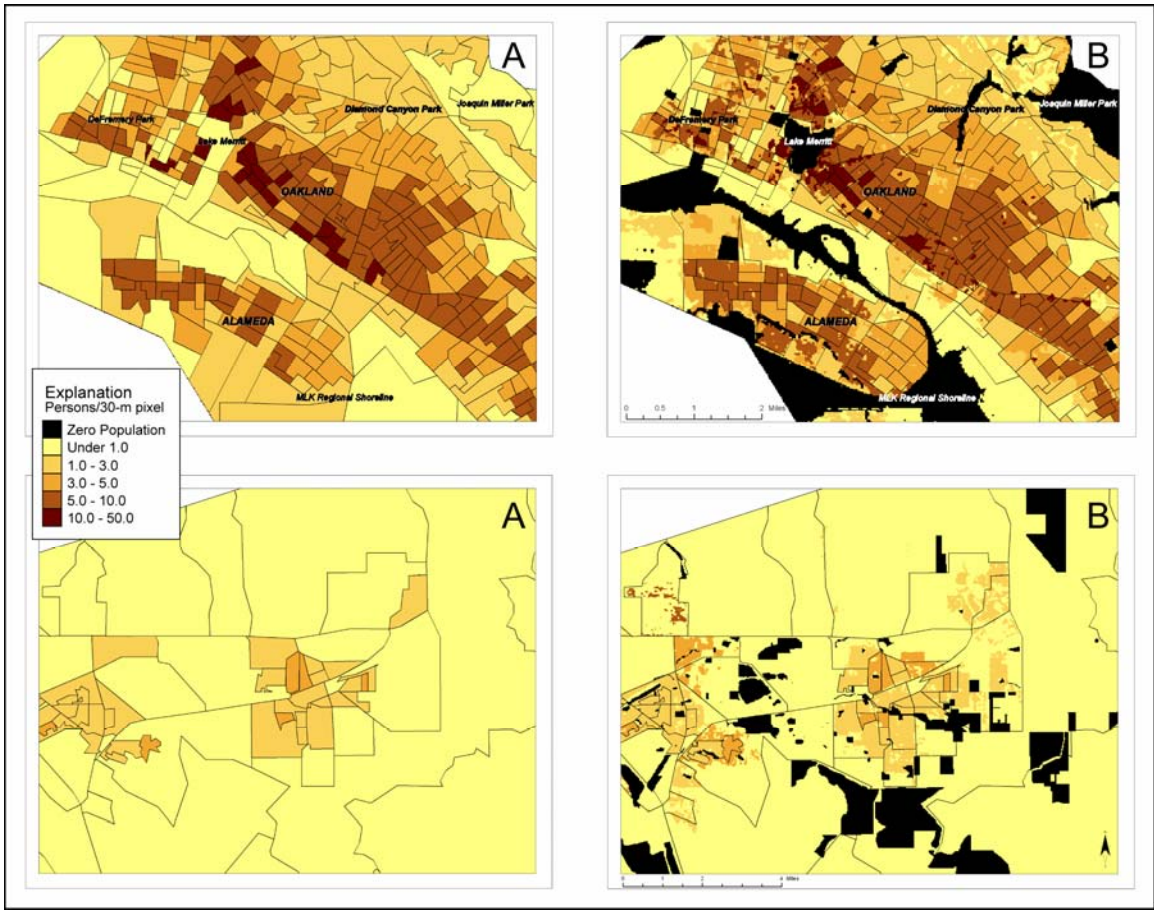

Figure 4. Choropleth (A) and dasymetric (B) maps of 1990 population density in Alameda County, Calif.

Francisco Bay, which the choropleth map conceals entirely and also shows nonzero population levels where there is water. Also, Lake Merritt, the largest urban water feature in Oakland, appears to be populated on the choropleth map, whereas the dasymetric map correctly identifies this area as uninhabited. Other features, such as parks, have been designated as uninhabited areas as well, adding to the overall visual realism of the dasymetric map.

CONCLUSION

Our study has shown that the dasymetric mapping technique is a viable approach for defining the underlying statistical surface from spatial data that are aggregated and attributed to large areal units. The process may seem laborious to some geographers for mapping population density because urban core areas typically show the same distributions. However, in large block groups

with sparse population (Figure 4) the dasymetric map demonstrates an intuitive and more informative distribution. Although processing time is minimized with the GIS field calculator, the overall task of areal interpolation is time consuming, although it could be automated into a programming interface if desired. The inclusion of enhanced ancillary data could improve accuracy within all land-cover types, owing to the identification and elimination of all areas with zero population. The decision to use a dasymetric mapping technique should be made on the basis of the purpose and the intended audience of the map.

About the Author: Rachel Sleeter recently finished her M.A. in Geography from San Jose State University, California. She has worked for the U.S. Geological Survey for 5 years and currently plays a roll in the Research and Technology Team of the Western Geographic Science Center. Her interests include urban morphology relative to land-use change; regional impacts of urbanization; and GIS applications for visualization.

Acknowledgements: Dr. Richard Taketa, San Jose State University/USGS and Dr. Leonard Gaydos, USGS.

REFERENCES

Burt, J. E. and G. M. Barber, 1996, *Elementary Statistics for Geographers* (New York: Guilford Press).

Dorling, D., 1993, Map Design for Census Mapping. *Cartographic Journal*, 30, 167-83.

Eicher, C. L., and C. A. Brewer, 2001, Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. *Cartography and Geographic Information Science*, 28(2), 125-38.

Freund, J. and G. Simon, 1997, *Modern Elementary Statistics* (New Jersey: Prentice Hall, Inc.).

Goodchild, M. 1992, Geographical Data Modeling. *Computers and Geosciences*, 18(4), 401-8.

GreenInfo Network, 2003, Bay Area Public and Privately Protected Lands, Shapefile.

Holloway, S., J. Schumacher, and R. Redmond, 1997, People & Place: Dasymetric Mapping Using Arc/Info. *Cartographic Design Using ArcView and Arc/Info*, Missoula: University of Montana, Wildlife Spatial Analysis Lab.

Mennis, J., 2003, Generating Surface Models of Population Using Dasymetric Mapping. *The Professional Geographer*, 55(1), 31-42.

Slocum, T. A., and S.L. Egbert, 1993, Knowledge acquisition from choropleth maps. *Cartography and Geographic Information Systems*, 20(2), 83-95.

Tobler, W., 1979, Smooth Pyncnophylactic Interpolation for Geographic Regions. *Journal of the American Statistical Association*, 74, 519-30.

Unwin, D.J., 1996, GIS, Spatial Analysis and Spatial Statistics. *Progress in Human Geography*, 20(4), 540-441.

U.S. Census Bureau, 2001, Census 2000 Geographic Definitions, August 2004 (http://www.census.gov/geo/www/geo_defn.html)

Vogelmann, J.E., S.M. Howard, L. Yang, C.R. Larson, B.K. Wylie, and N. Van Driel, 2000, Completion of the 1990s National Land Cover Dataset for the Conterminous United States from Landsat Thematic Mapper Data and Ancillary Data Sources. *Photogrammetric Engineering and Remote Sensing*, 67(6), 650-661.

APPENDIX I – Mennis (2003)

a) $d_{uc} = P_{uc}/(P_{bc} + P_{lc} + P_{nc})$ (1)
where d_{uc} = population-density fraction of urbanization class *u* in county *c*, P_{uc} = population density (persons/900 m^2) of urbanization class *u* in county *c*,
P_{bc} = population density (persons/900 m^2) of urbanization class *h* (high) in county *c*,
P_{lc} = population density (persons/900 m^2) of urbanization class *l* (low) in county *c*, and
P_{nc} = population density (persons/900 m^2) of urbanization class *n* (nonurban) in county *c*.

b) $ba_{ub} = (n_{ub}/n_b)/0.33$ (2)
where a_{ub} = area ratio of urbanization class *u* in block-group *b*,
n_{ub} = number of grid cells of urbanization class *u* in block-group *b*, and
n_b = number of grid cells in block-group *b*.

c) $f_{ubc} = (d_{uc}*a_{ub})/$
$[(d_{hc}*a_{hb}) + (d_{lc}*a_{lb}) + (d_{nc}*a_{nb})]$ (3)
where f_{ubc} = total fraction of urbanization class *u* in block-group *b* and in county *c*,
d_{uc} = population-density fraction of urbanization class *u* in county *c*, a_{ub} = area ratio of urbanization class *u* in block-group *b*,
d_{hc} = population-density fraction of urbanization class *h* (high) in county *c*, d_{lc} = population-density fraction of urbanization class *l* (low) in county *c*, d_{nc} = population-density fraction of urbanization class *n* (nonurban) in county *c*,
a_{hb} = area ratio of urbanization class *h* (high) in block-group *b*,
a_{lb} = area ratio of urbanization class *l* (low) in block-group *b*, and
a_{nb} = area ratio of urbanization class *n* (nonurban)in block-group *b*.

d) $pop_{ubc} = (f_{ubc}*pop_b)/n_{ub}$ (4)
where pop_{ubc} = population assigned to one grid cell of urbanization class *u* in block-group *b* and in county *c*,
f_{ubc} = total fraction for urbanization class *u* in block-group *b* and in county *c*, pop_b = population of block-group *b*, and
n_{ub} = number of grid cells of urbanization class *u* in block-group *b*.