An Evaluation of Land Parcel-Weighted Areal Interpolation in Small Area Estimates

Kyle Reese-Cassal
Washington State Office of Financial Management
Forecasting Division

Manuscript prepared for presentation at the
2007 Annual Meeting of the Population Association of America
March 29 – 31, New York, NY

Kyle Reese-Cassal, Geo-Demographer, Office of Financial Management - Forecasting Division. GA Building, Suite 318 G, 210 11th Ave. SW, Olympia, WA 98504-3113. (360) 902-9815 – Phone, (360) 725-5174 – Fax. Kyle.Reese-Cassal@ofm.wa.gov. www.ofm.wa.gov/Forecasting.
This work was partially supported by the Washington State Office of Financial Management. Thanks to Theresa Lowe, Mike Mohrman, and Tom Kimpel for their contributions.

Abstract

Small area population estimates for user-defined boundaries often require that administrative units be split and then re-aggregated using areal interpolation techniques. The spatial allocation process can benefit from a detailed understanding of the distribution of attributes at a scale greater than that of the finest areal unit. Rather than assuming that attributes are distributed uniformly across areal units, estimate accuracy can be improved by creating a population surface based on additional sources of disaggregate data that more closely resemble the true population distribution. Many complex strategies have been developed for estimating population based on land use and land cover classification systems. A more simplistic strategy is available to applied demographers. Rich sets of administrative data, such as land parcels, can be utilized to geospatially reference actual housing units which can then be used as an areal weight. This paper describes the process of creating areal weights from land parcel information and transforming the weights into a population surface. Estimates derived from four different interpolation algorithms—area-weighted areal interpolation (AAI), parcel point-weighted areal interpolation (PPAI), residential parcel point-weighted areal interpolation (RPPAI), and residential parcel point surface interpolation (RPPSI)—are tested for accuracy against a set of known benchmark values. All three parcel-weighted allocation methods produce less than half the error produced by area-weighted areal interpolation. When freed from the restraints of census geography, RPPSI is the most flexible interpolation technique and produces the least error in small area estimates.

Key words: Areal interpolation, dasymetric mapping, small area estimates.

The population unit within Washington State's Office of Financial Management (OFM) routinely produces population estimates for small areas. Accuracy is of paramount importance since the small area estimates are used in revenue allocation and policy decisions by elected officials and local planners.

OFM develops small area population and housing estimates biannually at the census block level based on building permit data, assessor records, postal statistics, and information from the decennial census. Block level estimates are then joined with census TIGER/Line blocks to form a base geography. Blocks within a user-defined boundary are aggregated to produce small area estimates. When the user-defined boundary is not conterminous with census block boundaries, small area estimates are forced to split and re-aggregate census blocks. This boundary discrepancy necessitates the use of areal interpolation techniques—the transfer of information from one set of areal units to another.

The most common method of areal interpolation is area-weighted interpolation—where population is allocated based on the proportion of the area of the source unit within the estimate boundary. This method is based on the erroneous assumption of a uniform distribution of population across the areal unit. Allocating population based on the size of the geographic area is a viable solution only when no other information is available about the actual distribution of population within the areal unit.

Dasymetric mapping strategies attempt to overcome the erroneous assumption of evenly distributed attributes by using supplementary data that are more indicative of the actual population distribution. This paper explores the use of land parcels as a source of ancillary data for use in dasymetric mapping. Geospatially referenced land parcel information can yield important spatial information that is indicative of the true distribution of population and housing across the landscape. This paper describes the process of creating areal weights from land parcel information and transforming the weights into a population surface. Estimates derived from four different interpolation algorithms—area-weighted areal interpolation (AAI), parcel point-weighted areal interpolation (PPAI), residential parcel point-weighted areal interpolation (RPPAI), and residential parcel point surface interpolation (RPPSI)—are tested for accuracy against a set of known benchmark values.

Areal Units

Applied demographers are confronted with two main problems when dealing with socioeconomic phenomena tabulated by administrative areal units. The first relates to the size of the areal unit. As size increases, local specificity and precision decreases. The US Census Bureau's smallest level of tabulation is the census block. Census blocks are bounded on all sides by visible features such as roads and rivers and by invisible boundaries, such as city boundaries and property lines[1]. In urban settings a census block corresponds with an actual city block. In rural areas a census block can be quite large, sometimes encompassing several square miles. Although the actual distribution of attributes within a block varies, analysts often incorrectly assume a uniform distribution of attributes within areal units. This assumption can introduce error, especially in larger blocks where the spatial distribution of attributes is less likely to be evenly distributed.

The second issue associated with data attached to areal units is one of delineation. Enumeration districts are delineated to match community patterns and to correspond roughly with population thresholds. The way that these arbitrary boundaries are drawn directly affects tabulation outcomes. Because socioeconomic phenomena are not constrained by administrative

[1] While this is a valid strategy, the location of city boundaries and roads can change over time. Furthermore, when comparing census block boundaries to aerial imagery or locally-derived GIS parcel boundaries, there are often large discrepancies due to spatial inaccuracies.

boundaries (i.e., attributes do not start and stop at zonal borders), they are perhaps better conceptualized as surfaces that are draped over the geographical landscape.

Statistical inference is dependent upon the partitioning and the size of areal units (Gehlke and Biehl 1934). Openshaw (1983) described the variability of statistical outcomes of data grouped into areal units as the Modifiable Area Unit Problem (MAUP) where the results are largely a function of size (the scale effect) and shape (the zoning effect). It is often difficult to discern whether the outcomes from aggregated areal units are reflective of local actuality or the result of the size and/or shape of the areal unit. Analysts have employed a host of techniques to lessen the effect of the MAUP and improve on the local specificity of areal units. Knowledge about the actual patterns of distribution at a finer level of resolution than the smallest areal unit is vital to the process of accurately allocating population and housing in small areas. For example, a block may be entirely composed of housing units and vacant land. If the block is split by a user-defined boundary, taking into account the location of the housing units would greatly increase the accuracy of the estimate.

Areal Interpolation

The most accessible and simplistic method of transferring data between two incompatible zones is area-weighted areal interpolation or cross-area estimation (Goodchild and Lam 1980), where data are transferred from a source zone to a target zone (Flowerdew and Green 1989). In the areal weighting method, intersecting zones are allocated data based on the geometric area of the target zone within the source zone (Fisher and Langford 1996):

$$\mathrm{P}_t = \sum_{r=1}^{q} \frac{\mathrm{A}_{ts\,r}\mathrm{P}_{s\,r}}{\mathrm{A}_{s\,r}} \tag{1}$$

where P_t is the estimated target zone population total, q is the number of source zones that intersect the target zone, P_s is the population of the rth intersecting source zone, A_s is the area of the rth source zone, and A_{ts} is the area of intersection between the rth source zone and the target zone (see Map 1). This method meets the minimum criterion of the pycnophylactic property (Tobler 1979), where records from the source zone are neither created nor destroyed by weighing and allocating parts of records to target zones, i.e., new parts sum to previous totals. Area-weighted areal interpolation assumes even population distribution across areal units and therefore can produce erroneous results when communities are clustered or land features prohibit inhabitation. The objectivity of this technique can be justified, but it should only be used when there is no other information available about the distribution of a population.

Dasymetric Mapping

Supplemental information about the distribution of population can be added to choropleth maps to lessen the burden of the MAUP and to better represent the local specificity of socioeconomic data. This technique, known as dasymetric mapping (Wright 1936), changes the assumption that data are uniformly distributed over space by adding local value to arbitrary administrative areal units. The widespread adoption of geographic information systems (GIS) has greatly increased the accessibility and prevalence of dasymetric mapping (Eicher and Brewer 2001).

Finding ancillary data that corresponds to population distribution can be difficult and costly. A fair amount of research has been undertaken using the properties of the areal unit to distribute population. Researchers have assumed that population is clustered around the centroid of an areal unit or enumeration district (Martin 1989; Martin and Bracken 1989) and thus decreases in a uniform fashion as the distance from the centroid increases. This method ignores geographic features that may disburse population away from the center of an enumeration district such as the presence of a wetland near the center or a desirable waterfront setting near the edge of the district.

Other analysts have explored dasymetric mapping with vector based ancillary data. Reibel and Bufalino (2005) found street-weighted interpolation to produce significantly lower and more consistent errors than simple areal interpolation. Street-weighted interpolation produced larger errors in areas of rapid development where streets and population expansion were not in synch. Street-weighted interpolation also results in biased estimates in areas with a high concentration of industrial activity where the quantity of roads is not correlated with number of residential structures.

A fairly common strategy is the use of raster data in the form of land use or land cover (LU/LC) information produced from satellite imagery. In short, LU/LC data are used to classify areas based on inhabitability. The simplest form of classification is binary—land is either inhabitable or non-inhabitable. Population is then distributed to inhabitable areas based on areal allocation. Langford (2006) found this method to be 33 percent more accurate than simple areal allocation. Langford also found inconclusive results when extending the binary model to a 3-class dasymetric model taking into account urban residential density.

Many complex models have been developed that incorporate aerial photography and satellite imagery (Langford and Unwin 1994; Mennis 2003; Reibel and Agrawal 2006; Langford et al. 1991; Langford 2006; Weichselbaum et al. 2005). Satellite remote sensing can not directly identify population density, but can describe urbanized land-cover patterns. Pixels are classified as either residential or nonresidential in a binary model, or pixels are classified according to land use types that are assigned varying population densities. Classification categories are subjective and classification error is highly correlated with image quality and resolution (Foody 2002). Since there is no standard convention for classifying land types, adjustments to any model must be made on a case-by-case basis.

All of the above methods represent an improvement over area-weighted areal interpolation yet they each contain their own shortcomings. Although imagery can be an excellent proxy determinant of population distribution, knowing the specific location of housing units would likely yield more accurate population estimates. While understanding that a slope is too steep to be built upon is helpful, the ability to precisely locate housing units avoids issues surrounding image resolution and the subjective classification of land types. Surprisingly little dasymetric mapping has been undertaken using information about the location of actual housing units to approximate the actual population distribution.

Moon and Farmer (2001) allocated census block counts based on residential structure data from the Arkansas State Highway Transportation Department. They found that when freed from the census-derived boundaries, the data produced more accurate estimates for user-defined boundaries. In addition, San Diego Association of Governments (SANDAG) is using tax parcel records to georeference housing units for use in their small area estimate system. In order to build a parcel-level housing unit inventory, SANDAG had to expend considerable effort to reconcile any discrepancies between census records and local assessor records using aerial imagery. Although the process of reconciliation took nearly a year, SANDAG believes that this process has greatly increased the flexibility and utility of their small area estimate program (Jarosz 2007).

Land Parcels

Land parcel information provides a realistic means for identifying the distribution of housing activity across geographic space. GIS-based tax parcel information is readily available for most urban areas and in some cases on a statewide basis[2]. Land parcels are typically maintained at the county-level by the assessor, planning, or GIS department. The most common parcel attributes consist of parcel identification number, lot and building size, site address, ownership information, land use, and property valuation metrics. Parcels are used for various administrative functions including taxation, tracking building permits, and land use planning. Applied demographers and GIS analysts can usually obtain parcel coverages for free or for a nominal fee. For these reasons, parcels are uniquely suited for dasymetric mapping.

Parcel-weighted areal interpolation can solve many of the problems attributed to the various dasymetric mapping methods mentioned previously. Where other weighting efforts struggled to adequately classify industrial locations, land parcels are uniquely suited to properly weight these areas. Land parcels are a good indicator of population density since parcel size is often inversely related to population density. Also, non-residential (e.g., agricultural and industrial parcels) are typically larger than residential parcels and would therefore receive less weight.

Adding to their utility, land parcels often contain attribute data that can add to the accuracy of population distribution. For instance, most parcel data can be easily classified as residential or nonresidential based on a land use classification variable[3]. This simple binary distinction proves to be very helpful when allocating population across areal units. Parcel data are also an excellent proxy determinant of population and housing distribution in areas of rapid development. When areas are being developed, plat lines are drawn before housing units and roads are built and parcels layers are updated frequently.

Methods and Data

This study tests the accuracy of using land parcel information as ancillary data in dasymetric mapping of population and housing. Estimates derived from four different interpolation algorithms—area-weighted areal interpolation (AAI), parcel point-weighted areal interpolation (PPAI), residential parcel point-weighted areal interpolation (RPPAI), and residential parcel point surface interpolation (RPPSI)—are tested for accuracy against a set of known benchmark values. Tests of allocation accuracy are performed by interpolating population counts from 1990 census tract geography to 2000 census tract geography. Errors are estimated using the 1990 census block counts aggregated to census 2000 tracts as a benchmark. This method, borrowed from Reibel and Bufalino (2005), allows allocation errors to be analyzed by subtracting the 1990 benchmark counts from the 1990 estimates for each target zone. A detailed description of the process of aggregating 1990 census blocks to 2000 census tracts can be found in Appendix A. The pattern and magnitude of estimate errors for the different allocation techniques is then compared. Four types of allocations are tested: area-weighted, parcel point-weighted, residential parcel point-weighted and parcel surface.

The study area consists of Clark County, Washington. Clark County is part of the Portland, Oregon metropolitan region. The City of Vancouver is highly urban while northeastern Clark County is mostly rural. Clark County was the fastest growing county in Washington between 1990 and 2000—growing by nearly 45% (Lowe 2000). This growth resulted in the addition of 1,537 new census blocks, 29 new census tracts, and 36 realigned census tracts in 2000. Clark County is an excellent test case as it offers a range of disparate geographical features including urban, rural, industrial, and rapidly developing areas.

All spatial processes were performed with ArcInfo 9.1 (ESRI 2006) and ET GeoWizards 9.3.1 (Tchoukanski 2005). Land parcels from April 2001 were obtained from Clark County GIS

[2] Oregon, Massachusetts, and California are all in the process of developing statewide parcel layers.
[3] The primary exception being parcels classified as mixed-use. In some cases a jurisdiction will assign a non-residential land use to what is essentially a parcel that contains at least some residential units.

department (Clark County 2001). Parcels from 2001 where chosen because they most closely correspond to the housing distribution in 2000.

Parcel Point-Weighted Areal Estimation

Parcel Point-weighted Areal Interpolation (PPAI) is accomplished by converting the feature geometry of land parcels from polygons to points. Parcel points are then spatially joined with the source layer (1990 tracts) so that each source tract has an aggregated total count of parcel points that fall within its boundary. An overlay is then performed with the target layer (2000 tracts). Source zones are split and a spatial join is performed with the parcel points to calculate a new aggregated total count for the target zones. The parcel point count in the source zone divided by the parcel point count in the target zone creates a ratio that acts as a weight for distributing attributes to the new target zones (see Map 2). PPAI is defined in Equation 2:

$$\mathrm{P}_t = \sum_{r=1}^{q} \frac{D_{ts\,r}\mathrm{P}_{s\,r}}{D_{s\,r}} \qquad (2)$$

where P_t is the estimated target zone population total, q is the number of source zones that intersect the target zone, P_s is the population of the rth intersecting source zone, D_s is the count of all parcel points in the area of the rth source zone, and D_{ts} is the count of all parcels in the area of intersection between the rth source zone and the target zone.

Residential Parcel Point-Weighted Areal Estimation

Land parcels commonly contain attributes distinguishing parcels as either residential or non-residential. This distinction is important for taxing and zoning purposes. Being able to omit non-residential parcels helps avoid many of the faults of the previously mentioned dasymetric classification schemes. Residential Parcel Point-weighted Areal Interpolation (RPPAI) can be calculated by using Equation 2 and omitting non-residential parcel points. For RPPAI, D_s is the count of all *residential* parcels in the area of the rth source zone, and D_{ts} is the count of all *residential* parcels in the area of intersection between the rth source zone and the target zone (see Map 3).

Residential Parcel Point Surface Estimation

Residential Parcel Point Surface Interpolation (RPPSI) is the final method of interpolation tested. RPPSI aims to lessen the MAUP by creating a surface of data irrespective of census geography. Population and housing distributions are best envisioned as a surface that does not stop and start at zonal boundaries. In the study, a surface takes the form of a vector grid where each cell has a share of the source zone's (1990 tract) population and housing values. A surface can be created using a level of precision that is generally much finer than the smallest areal unit provided by the census bureau, thereby potentially increasing the accuracy of the small area estimates.

Grid cell size must be selected prior to the creation of the surface. In this analysis a cell size of 500 x 500 feet, or 250,000 sq. feet is selected. This scale is roughly the size of a baseball field or large factory. Each grid cell is assigned a unique identifier to be used later in the process for attribute aggregation.

The surface is created in much the same way as the RPPAI method, but now the grid functions as an intermediate target zone. Residential parcel points are spatially joined with the source layer (1990 tracts) so that each source tract has an aggregated total count of the parcel points that fall within its boundary. An overlay is then performed with the target layer (grid).

Source zones are split and a spatial join is again performed with the parcel points to calculate a new aggregated total count for the target zones (grid cells). The parcel point count in the source zone divided by the parcel point count in the target zone creates a ratio that acts as a weight for distributing attributes to the new target zones (grid cells).

Grid cells are often divided between tracts. In these instances the individual grid cell is divided into parts. Each part of the grid cell has a separate source zone and target zone count of residential parcel points. These counts are used to calculate a ratio which is used as a weight to allocate population and housing to each part of the grid cell. Once the grid cell parts have estimates of population and housing, then the parts are dissolved into the original unbroken grid cell by a unique identifier and estimates for the parts are aggregated. This process produces a continuous surface of grid cells with one record for population and one record for housing.

Grid cells are then converted from polygons to points. This point surface is spatially joined with the actual target zone (2000 tracts). Attributes from the point surface are aggregated if the points are completely within the target zone to produce population and housing estimates. RPPS is defined in Equation 3:

$$\mathrm{P}_t = \sum_{r=1}^{a} \left(\sum_{r=1}^{g} \left(\sum_{r=1}^{q} \frac{D_{ts\,r}\mathrm{P}_{s\,r}}{D_{s\,r}} \right) \right) \tag{3}$$

where P_t is the estimated target zone population total, a is the number of point within the source zone, g is the unique identifier of the grid cell, q is the number of source zones that intersect the target zone, P_s is the population of the rth intersecting source zone, D_s is the count of all residential parcels points in the area of the rth source zone, and D_{ts} is the count of all parcels in the area of intersection between the rth source zone and the target zone (see Map 4).

Estimate Error

Benchmark population and housing estimates are computed by allocating Clark County 1990 tract-level population and housing counts to 2000 tract boundaries, these estimates are assumed to be accurate for purposes of estimation comparison. Estimates for the four types of interpolation algorithms—area-weighted areal interpolation (AAI), parcel point-weighted areal interpolation (PPAI), residential parcel point-weighted areal interpolation (RPPAI), and residential parcel point surface interpolation (RPPSI)—are tested for accuracy against the benchmark estimates. Table 1 summarizes the error measures and error distributions in the population estimates for the four interpolation methods. Table 2 summarizes the error measures and error distributions in the housing unit estimates.

Reports of estimate error are given as the root mean squared error (RMSE) as defined in Equation 4:

$$RMSE = \sqrt{\frac{\sum_{r=1}^{q}\left(P_{s\,r} - P_{t\,r}\right)^2}{q}} \tag{4}$$

where P_{tr} is the estimated population of the rth target zone, P_{sr} is the estimated population of the rth source zone, and q is the number of target zones in the analysis. RMSE is a measure of total estimate error and is highly sensitive to large errors/outliers making it an ideal measure for small area estimates which attempt to avoid large errors relative to population size.

Mean Algebraic Percent Error (MALPE) and Mean Absolute Percent Error (MAPE) are presented as additional measures of estimate error. Both measures account for population size by using percents rather than numeric error. MALPE is often used as a measure of bias since positive and negative errors offset one another. A Negative MALPE shows that estimates are too low while a positive MALPE shows that estimates are too high. MALPE is defined in Equation 5:

$$MALPE = \frac{\sum PE_t}{n} \tag{5}$$

where PE is the percent error at time t and n is the number of cases.

Positive and negative errors measured by the MAPE do not offset one another and therefore the MAPE is a good measure of total error regardless of the direction of the error. MAPE is defined in Equation 6:

$$MAPE = \frac{\sum |PE_t|}{n} \tag{6}$$

where |PE| is the absolute value of the percent error at time t and n is the number of cases.

Examining the error distributions and the RMSE, it is apparent that there are larger errors in the allocation of population than housing units in all four allocation algorithms. Aside from the magnitude of the errors, error patterns are strikingly similar when comparing population estimate error and housing unit estimate error. Therefore, patterns of error can be discussed in general without having to refer specifically to Tables 1 and 2.

Significant gains in allocation accuracy are attained when using any of the parcel point weighing algorithms over area-weighted areal interpolation. As expected, areal allocation based on parcel information is superior to areal allocation based on geographic area. A fifty percent RMSE error reduction is gained when using PPAI rather than AAI in population estimates. The RMSE shows that RPPAI and RPPSI are more accurate yet, but produce only marginal gains in error reduction. Looking at table 1, RMSE error reduction provided by PPAI over AAI is 51%. When using only residential parcels (RPPAI) we see an additional 1% improvement over PPAI. Converting RPPAI into a surface (RPPSI) provides an additional .7% RMSE error reduction over RPPAI. As the parcel-weighting algorithms become more complex, we do not see a corresponding improvement in RMSE error reduction.

The MALPE shows that PPAI, RPPAI, and RPPSI outperform AAI, but RPPSI produces slightly higher error than RPPAI. The MALPE shows that all interpolation schemes produce modest overestimates but again we see an almost 50% error reduction in error over AAI.

The MAPE shows that RPPSI produces the lowest overall error when positive and negative errors offset one another. This trend becomes more apparent when looking at the breakdown of percentile error. Although RPPSI produces errors throughout the range of percentiles, these errors have the smallest standard deviation of any interpolation method, i.e., RPPSI produces more errors, but the errors are smaller in magnitude than the other methods. These qualities make RPPSI an excellent choice as an interpolation algorithm for small area population and housing estimates.

Discussion

From the previous analysis of error distribution it is apparent that using land parcels in an areal weighting algorithm produces estimates with considerably less error than area-weighted areal interpolation. When omitting non-residential land parcels, estimate error is reduced but not as much as expected. This finding suggests that similar weighting ratios were calculated when using both all parcels and when using only residential parcels. Therefore many tracts must have relatively equal distributions of residential and non-residential land parcels. Distinguishing a land parcel as residential may in itself introduce error into the analysis when the actual status of the land parcel is unknown.

Because RPPAI does not provide much error reduction over PPAI, it follows that a surface build from residential parcel points (RPPSI) would also fail to produce considerable gains in RMSE error reduction. When using residential land parcels to create a surface estimate, error is decreased in all measures except the MALPE. Because some grid cells represent the sum of the attributes of multiple tracts, it is likely that estimates would be positively skewed when grid cells with little population pick up the attributes of neighboring grid cells with higher population. Adjusting grid cell size may have considerable effects on RPPSI allocation accuracy. A slightly higher MAPLE is of little concern as the RPPSI is truly the most accurate interpolation method tested. Not only is RPPSI an improvement over the other interpolation techniques that are tested, but it has additional intrinsic value in that it produces population and housing estimates for grid cells irrespective of census geography. Having estimates at this resolution yields flexibility in small area estimates that the other models can not approach.

Parcel Limitations

Using land parcels in a weighting algorithm provides more accurate estimates than area-weighed areal allocation when areal units are split and attributes are reallocated. Although this method is superior and fairly simplistic, it does have its limitations.

There is often a considerable time lag between the permitting and completion of a residential unit and the time the unit is represented in the parcel layer. In Washington, this lag can be anywhere from one week to over one year. When using land parcel attributes, analysts must be especially cautious. There are no standard conventions for tracking land parcel attributes; therefore jurisdictions will have different methods and programs to track land parcels. If a jurisdiction only differentiates between residential and non-residential land parcels, analysts will not have access to information about unit type. When unit type is not considered, as in this analysis, a single family residence receives the same weight as a multi-family structure. Knowing what type of housing unit sits on the land parcel is also important when dealing with group quarters such as prisons and nursing homes. The misallocation of group quarters is often the source of egregious errors in small area estimates (Mohrman 2007).

Analysts must also be aware of the occupancy status of housing units. In areas of seasonal housing, housing units may be vacant for the majority of the year. Allocating resident population to these areas is likely to decrease estimate accuracy. Special housing units such as RV parks and marinas also pose a unique problem. Although these special units may be occupied for the majority of the year, they are unlikely to be tracked as taxable residential land parcels. Other areas that may not be tracked in a parcel database are sovereign jurisdictions such as: military bases, Native American reservations, state or federal owned land, and other non-taxable lands (Jarosz 2007).

Conclusion

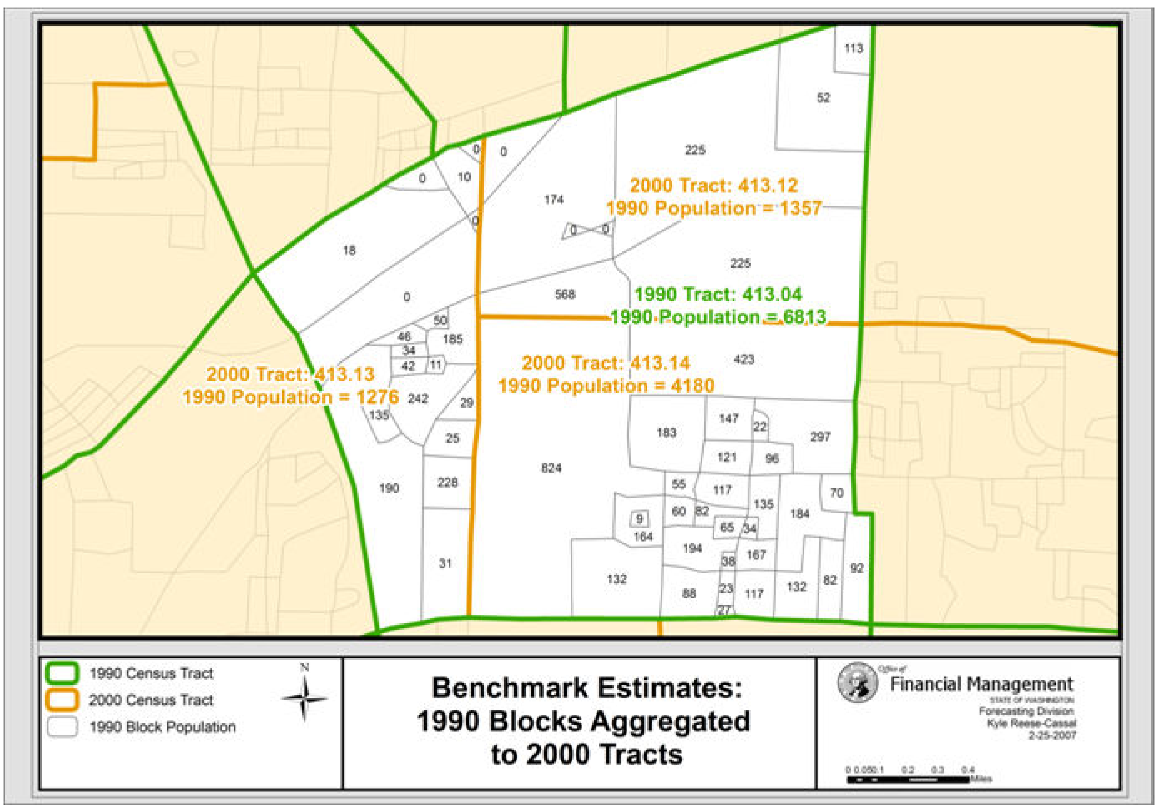
Analysts performing small area population estimates will continue to struggle with socioeconomic data tabulated to areal units. In order to lessen the MAUP analysts will continue to explore data than can provide a detailed understanding of the distribution of attributes at a scale greater than that of the finest areal unit. Many complex strategies have been developed to estimate population and housing distributions. Although some of these strategies have been fruitful, many are not practical for the applied demographer who is tasked with producing small area estimates. For these applied demographers it is important not to overlook the utility of land parcels as ancillary data for use in dasymetric mapping of population and housing.

Appendix A

Benchmark counts for census 2000 tracts with 1990 block population

This analysis uses a method of measuring estimate error based on benchmark counts of 1990 block population aggregated to 2000 tracts. This approach borrows heavily from Reibel and Bufalino's (2005) analysis of street-weighted interpolation techniques. As population grows, census tracts are divided to maintain tabular homogeneity in relation to population thresholds. Tracts that are split represent dense, rapidly growing areas. New divisions are non-arbitrary and in most cases follow 1990 block boundaries. Therefore, when no significant changes are warranted, 1990 block boundaries are the building blocks of 2000 tracts.

When overlaying 1990 blocks onto 2000 tracts many small gaps or slivers are produced due to the increased accuracy of the 2000 boundaries. Issues concerning sliver polygons are avoided with point-in-polygon aggregation. 1990 census blocks are aggregated to 2000 tracts if their label point is within the tract boundaries, attributes are then aggregated at the 2000 tract level. These counts of population and housing are assumed to be accurate for purposes of estimation comparison.



Citations

Bracken, Ian. and David Martin. 1989. "The Generation of Spatial Population Distributions from Census Centroid Data." *Environment and Planning A* 22:1079-1089.

Bracken, Ian. 1993. "An Extensive Surface Model Database for Population-related Information: Concept and Application." Environment and Planning B: Planning and Design 20(1):13-27.

Clark County GIS. 2001. Tax Parcel Layer. Clark County Geographic Information System (CCGIS) Database, Retrieved April 25, 2001 (http://gis.clark.wa.gov/applications/gishome/index.cfm).

Eicher, Cory L. and Cynthia A. Brewer. 2001. "Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation." *Cartography and Geographic Information Science* 28(2):125-138.

ESRI. Environmental Systems Research Institute. ArcGIS: Release 9.2. Redlands, California: Environmental Systems Research Institute, 1999-2007.

Fisher, Peter F. and Mitchel Langford. 1996. "Modeling Sensitivity to Accuracy in Classified Imagery: A Study of Areal Interpolation by Dasymetric Mapping." *Professional Geographer* 48(3):299-309.

Flowerdew, Robin and Michael Green. 1989. "Statistical Methods for Inference Between Incompatible Zonal Systems." Pp. 239-47 in *The Accuracy of Spatial Databases,* edited by M. Goodchild and S. Gopal, London: Taylor and Francis.

------. 1994. "Areal Interpolation and Types of Data." Pp. 121-145 in *Spatial Analysis and GIS,* edited by S. Fotheringham and P. Rogerson, London: Taylor and Francis.

Foody, Giles M. 2002. "Status of Land Cover Classification Accuracy Assessment." *Remote Sensing of Environment* 80:185-201.

Goodchild, Michael F. and Nina S. Lam. 1980. "Areal Interpolation: A Variant of the Traditional Spatial Problem." *Geoprocessing* 1:297-331.

Gehlke, C.E. and Katherine Biehl. 1934. "Certain Effects of Grouping Upon the Size of the Correlation Coefficient in Census Tract Material." *Journal of the American Statistical Association* 29(185):169-170.

Langford, Mitchel., D. Maguire and D.J. Unwin. 1991. "The Areal Interpolation Problem: Estimating Population Using Remote Sensing in a GIS Framework." Pp. 55-77 in *Handling Geographical Information: Methodology and Potential Applications,* edited by I. Masser and M. Blakemmore, New York: Longman.

Langford, Mitchel. and D.J. Unwin. 1994. "Generating and Mapping Population Density Surfaces Within a Geographical Information System." *The Cartographic Journal* 31:21-26.

Langford, Mitchel. 2006. "Obtaining Population Estimates in Non-Census Reporting Zones: An Evaluation of the 3-Class Dasymetric Method." *Computers, Environment and Urban Systems* 30(2):161-180.

Lowe, Theresa J. 2000. "State, County, City Populations." Pp. 9 in *Population Trends*. Office of Financial Management, State of Washington Press.

Jarosz, Beth. 2007. "Using Assessor Parcel Data to Maintain Housing Unit Counts for Small Area Population Estimates." Paper presented at the 2007 Applied Demography Conference in San Antonio, TX.

Martin, David. 1989. "Mapping Population Data from Zone Centroid Locations." *Transactions of the Institute of British Geographers* 14(1):90-97.

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

Mohrman, Mike. 2007. "Tracking Group Quarter Facility Data: The Washington State Experience." Paper presented at the 2007 Applied Demography Conference in San Antonio, TX.

Moon, Zola K. and Frank L Farmer. 2001. "Population Density Surface: A New Approach to an Old Problem." *Society and Natural Resources* 14:39-49.

Openshaw, Stan. 1983. *The Modifiable Areal Unit Problem.* Norwich, UK: Geobooks.

Reibel, Michael and Aditya Agrawal. 2006. "Land Use Weighted Areal Interpolation." Presented at the 2006 Population Association of America annual meeting, Los Angeles, CA.

Reibel, Michael and Michael E. Bufalino. 2005. "Street-Weighted Interpolation Techniques for Demographic Count Estimation in Incompatible Zone Systems." *Environment and Planning* 37(1):127-139.

Tchoukanski, Ianko. 2007. ET Spatial Techniques. ET GeoWizards: Release 9.6. Pretoria, South Africa: ET Spatial Techniques, 2003-2007.

Tobler, Waldo. 1979. "Smooth Pycnophylactic Interpolation for Geographical regions." *Journal of the Americal Statistical Association* 74:519-530.

Wright, John K. 1936. "A Method of Mapping Densities of Population with Cape Cod as an Example." *Geographical Review* 26:103-110.

Weichselbaum, J., F. Petrini-Monteferri, M. Papathoma, W. Wagner, and N. Hackner. 2005. "Sharpening Census Information in GIS to Meet Real-World Conditions - The Case for Earth Observation." Pp.143-153 in *Sustainable Development and Planning II Vol 1.*, edited by C.A. Brebbia, and A. Kungolos. UK: WIT Press.

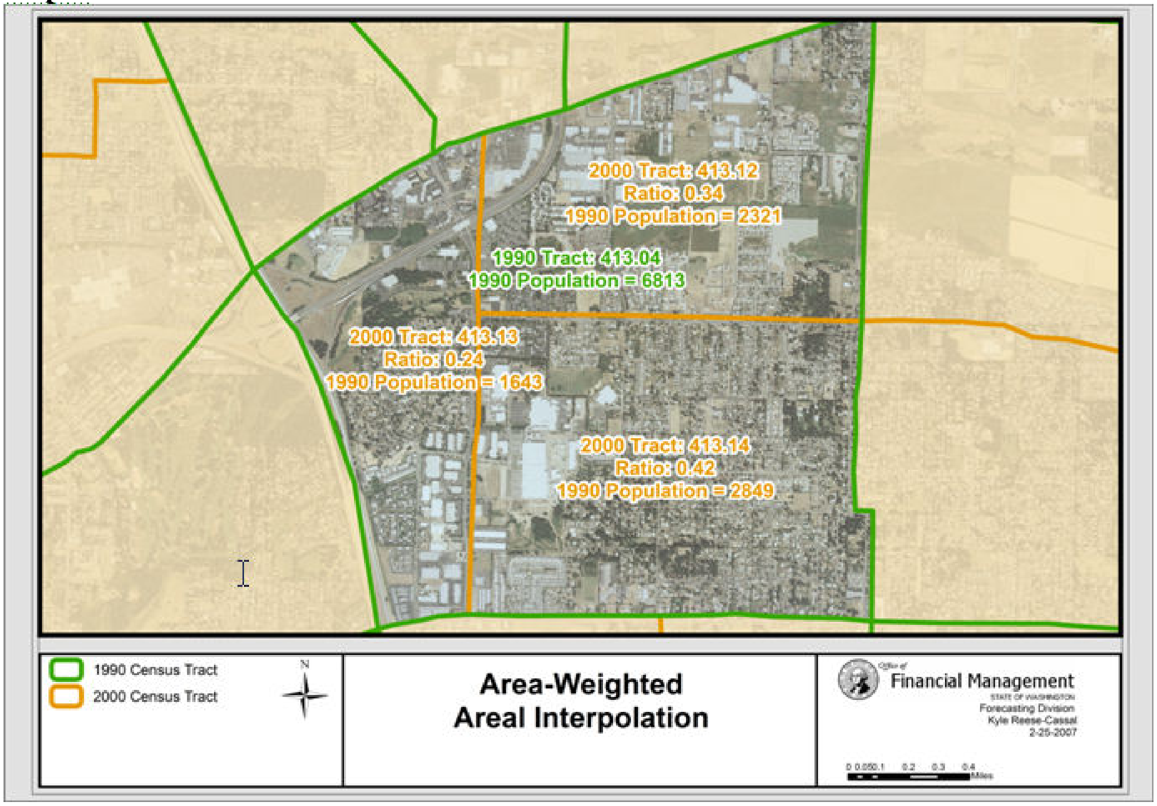
Table 1. Population Estimate Error

	Areal-Weighted Areal Interpolation	Parcel Point-Weighted Areal Interpolation	Residential Parcel Point-Weighted Areal Interpolation	Residential Parcel Point Surface Interpolation
	AAI	**PPAI**	**RPPAI**	**RPPSI**
N=65				
RMSE	1,147.19	565.75	553.49	545.86
RMSE Error Reduction		50.68%	51.75%	52.42%
MAPLE	10.16	5.48	5.23	5.41
MAPE	36.52	20.96	20.37	5.41
Standard Deviation	1,156	570	558	550
Percentiles				
1	-2,742	-1,234	-1,261	-1,344
10	-1,371	-701	-730	-711
20	-756	-446	-385	-383
30	-398	-253	-230	-225
40	-45	-60	-30	-63
50	1	0	0	19
60	25	31	30	41
70	343	257	144	108
80	695	385	385	401
90	1,371	658	576	642
99	2,833	2,050	2,037	1,926

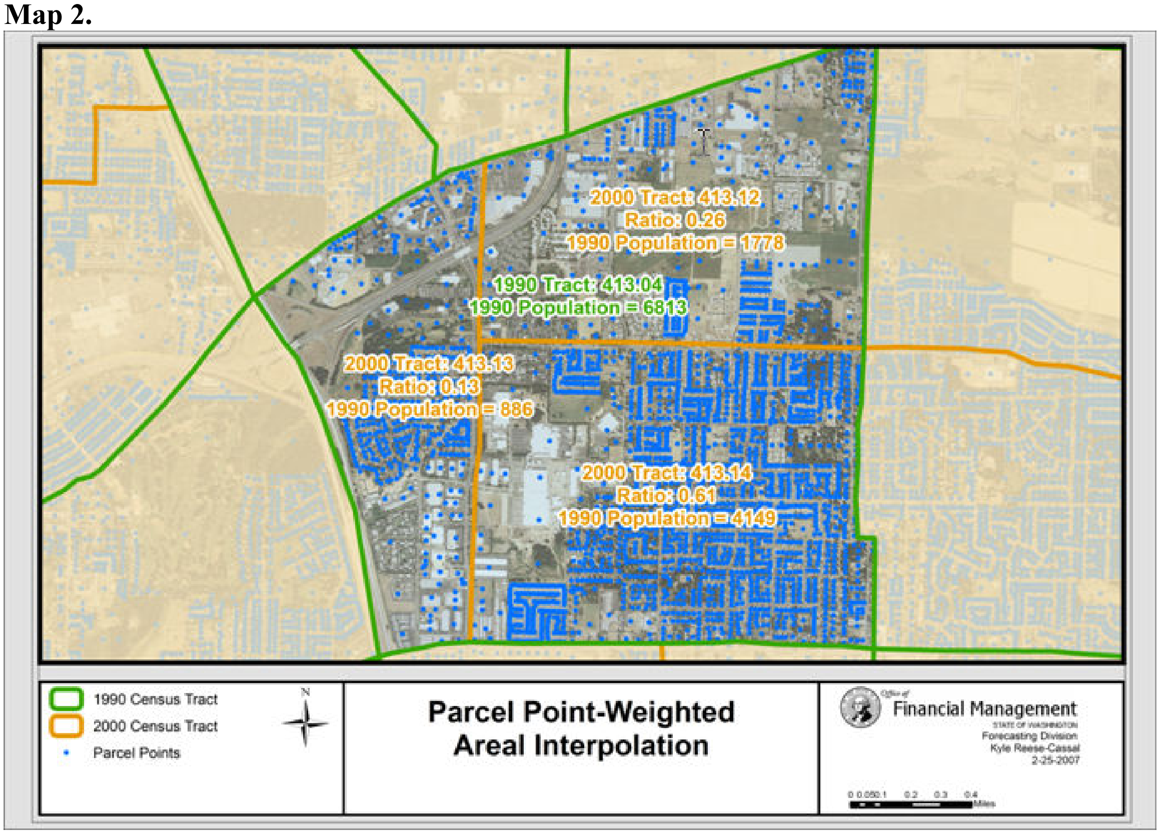
Table 2. Housing Unit Estimate Error

	Areal-Weighted Areal Interpolation	Parcel Point-Weighted Areal Interpolation	Residential Parcel Point-Weighted Areal Interpolation	Residential Parcel Point Surface Interpolation
	AAI	**PPAI**	**RPPAI**	**RPPSI**
N=65				
RMSE	394.74	198.34	198.02	197.33
RMSE Error Reduction		49.75%	49.83%	50.01%
MAPLE	11.02	6.21	6.07	6.30
MAPE	37.96	21.33	20.99	6.30
Standard Deviation	398	200	200	198
Percentiles				
1	-1,005	-448	-458	-489
10	-452	-261	-276	-287
20	-252	-169	-131	-130
30	-173	-57	-47	-67
40	-31	-12	-4	-13
50	1	0	0	7
60	13	8	10	17
70	128	69	42	37
80	230	124	102	99
90	452	207	253	248
99	1,005	705	701	662

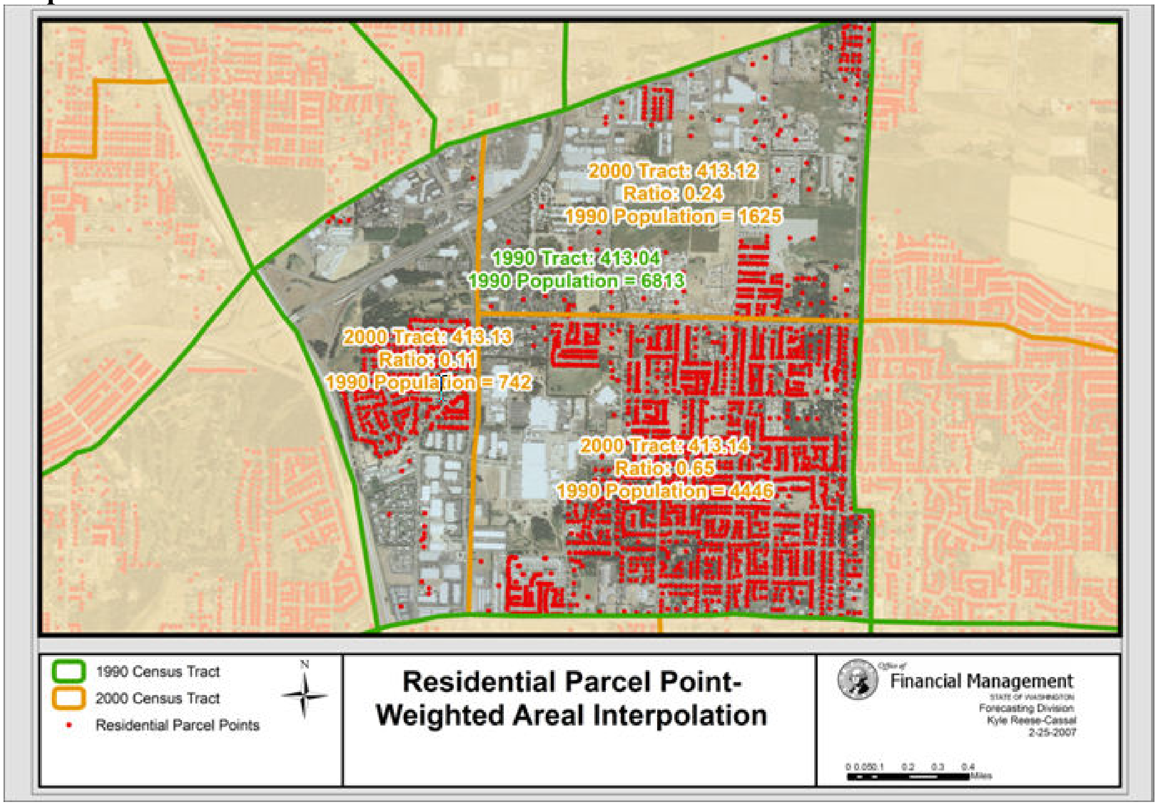
Map 1.


2000 Tract: 413.12
Ratio: 0.34
1990 Population = 2321
1990 Tract: 413.04
1990 Population = 6813
2000 Tract: 413.13
Ratio: 0.24
1990 Population = 1643
2000 Tract: 413.14
Ratio: 0.42
1990 Population = 2849
1990 Census Tract
2000 Census Tract
N
Area-Weighted
Areal Interpolation
Financial Management
Forecasting Division
Kyle Reese-Cassal
2-25-2007
Miles

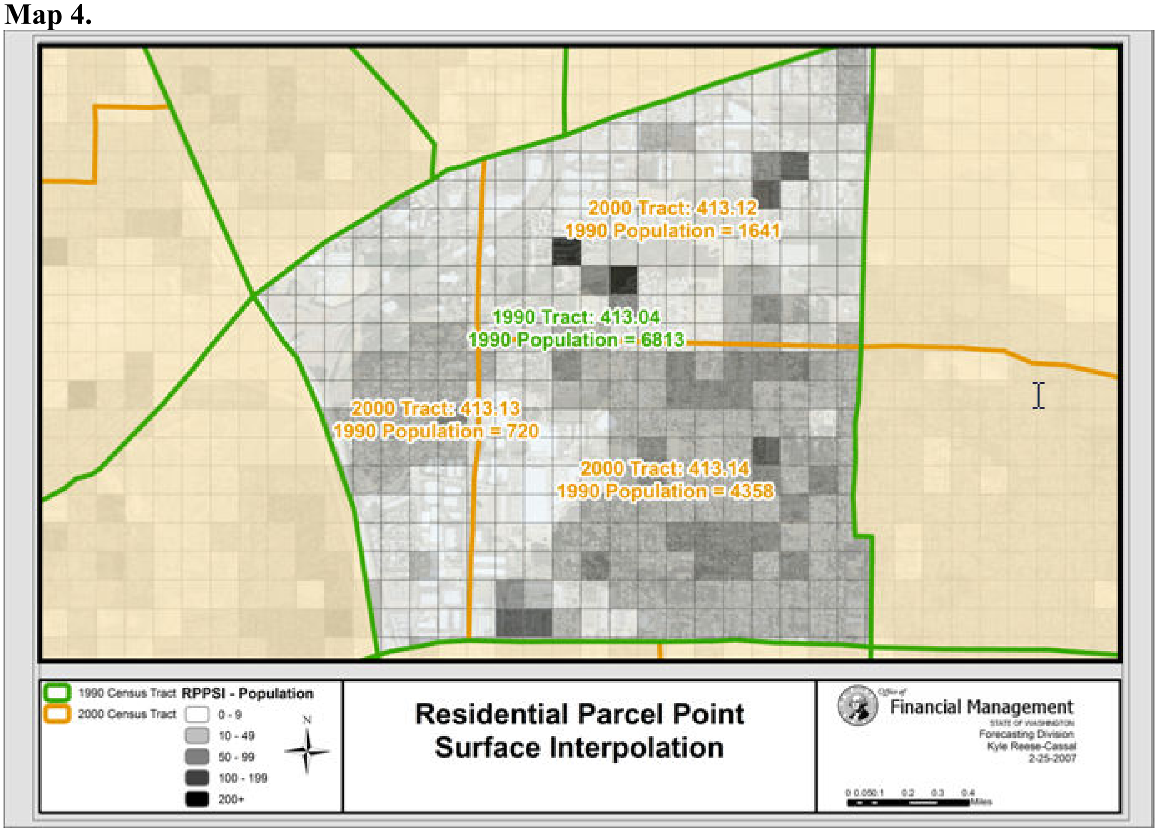
Map 2.


2000 Tract: 413.12
Ratio: 0.26
1990 Population = 1778
1990 Tract: 413.04
1990 Population = 6813
2000 Tract: 413.13
Ratio: 0.13
1990 Population = 886
2000 Tract: 413.14
Ratio: 0.61
1990 Population = 4149
1990 Census Tract
2000 Census Tract
Parcel Points
N
Parcel Point-Weighted
Areal Interpolation
Financial Management
Forecasting Division
Kyle Reese-Cassal
2-25-2007
Miles

Map 3.


2000 Tract: 413.12
Ratio: 0.24
1990 Population = 1625
1990 Tract: 413.04
1990 Population = 6813
2000 Tract: 413.13
Ratio: 0.11
1990 Population = 742
2000 Tract: 413.14
Ratio: 0.65
1990 Population = 4446
1990 Census Tract
2000 Census Tract
Residential Parcel Points
N
Residential Parcel Point-
Weighted Areal Interpolation
Office of
Financial Management
STATE OF WASHINGTON
Forecasting Division
Kyle Reese-Cassal
2-25-2007
Miles

Map 4.


2000 Tract: 413.12
1990 Population = 1641
1990 Tract: 413.04
1990 Population = 6813
2000 Tract: 413.13
1990 Population = 720
2000 Tract: 413.14
1990 Population = 4358
1990 Census Tract
2000 Census Tract
RPPSI - Population
0 - 9
10 - 49
50 - 99
100 - 199
200+
N
Residential Parcel Point
Surface Interpolation
Office of
Financial Management
STATE OF WASHINGTON
Forecasting Division
Kyle Reese-Cassal
2-25-2007
Miles