Review of Economics & Finance
Submitted on 18/04/2018
Article ID: 1923-7529-2018-04-25-14

Use of Kernel Density and Raster Manipulation in GIS to Predict Population in New Mexico Census Tracts

Dr. ***Srini Vasan*** (Correspondence author)
Department of Mathematics/Statistics, University of New Mexico,
1700 Lomas Boulevard NE, Albuquerque, NM, U.S.A.
E-mail: svasan01@unm.edu

Dr. ***Jack Baker***
Health Fitness Corporation, Albuquerque, NM, U.S.A.

Dr. ***Adélamar Alcántara***
Institute for Geospatial and Population Studies, University of New Mexico,
1700 Lomas Boulevard NE, Albuquerque, NM, U.S.A.

***A*bstract:** Shifts in small-area boundaries (such as blocks and tracts) between censuses create a significant challenge for applied demographers wishing to make small-area population projections. To date, most methods for reformulating data across incongruent boundaries have revolved around various methods of areal interpolation and recent research has suggested that this may result in substantial error and bias in small-area population projections. This paper attempts to improve the accuracy with which such transformations are made by introducing the application of kernel density functions based upon raster data within a GIS-based interface. Population density for 1990 and 2000 census are allocated to the level of pixels, which allows nearly infinitesimally small summations to be made that more accurately transfer data across incongruent geographies. Small-area population densities are thereby reconstructed for 1990 and 2000 Censuses then projected forward to 2010. The resulting census tract-level projections are then compared to the results of the 2010 Census in an ex-post-facto evaluation of error and bias. It has been possible to predict census tract population with an overall Mean Absolute Percent Error (MAPE) of 29.51% and a Mean Algebraic Percent Error (MALPE) of 6.92%. Predictions for the 2020 census tract population are then made.

Keywords: Census tract; Kernel density; Population; Prediction; Projection; Small area

JEL Classification: C02, C10, C52, C88

1. Introduction

Modeling longitudinal population change through demographic projection faces a fundamental challenge in dealing with inter-censal boundary changes (Voss *et al*., 1999; Baker *et al*., 2014a). Since historical datasets provide the basis for most demographic projection methodologies (Smith *et al*., 2001; George, 2004), a lack of concordance between historical and target geographies can introduce important biases into small-scale population projections (Zandbergen and Ignizio, 2010; Tobler, 1979; Voss *et al*., 1999; Fisher and Langford, 1995; Sadahiro, 2000; Swanson and Tayman, 2012; Baker *et al*, 2014). This bias may be substantial as high as 35% (Zandbergen and Ignizio 2010).

Demographers and population geographers have typically dealt with this issue through the process of geographic normalization (Tobler, 1979; Voss *et al*., 1999) — the most common variant of which involves *areal interpolation* (Zandbergen and Ignizio, 2010; Tobler, 1979; Fisher and Langford, 1995). The substantial errors associated with this method are directly attributable to the imprecision associated with the assumption of *a homogeneous distribution of population density within the intersection of donor and recipient geographies* (Zandbergen and Ignizio, 2010; Tobler, 1979). Housing units have been shown to improve accuracy in studies in both population geography and demography (Zandbergen and Ignizio, 2010; Baker *et al*., 2014a). The Census Bureau, who produces official normalization factors based on the census microdata itself in *Tract Relationship Files* (https://www.census.gov/geo/maps-data/data/relationship.html), however, this microdata is unavailable for non-census geographies, placing important limitations (Swanson and Tayman, 2012; Smith *et al*., 2001; Baker *et al*., 2013).

In this article, we address the fundamental problem of geographic normalization. We develop a Kernel-density function approach to create raster-based data from census block counts reported in the 1990, 2000, and 2010 US censuses. This approach permits the development of "surfaces" of population density within census tracts and allows the normalization of 1990 and 2000 census data into 2010 census geographies without reliance upon geometric intersection between the three sets of geographies. By developing a raster-based density using the Kernel function approach, we propose to relate directly census 1990 and 2000 data to 2010 census geographies using elementary concepts from integral calculus (Riemann sums — see Keisler, 2013 for a review). Sums of small-scale densities within 2010 census tracts during 1990 and 2000 are used to estimate 1990 and 2000 census populations within the 2010 geographies and rates of change are used to project these populations to 2010. The precision of this method is expected to reduce both error and bias in 10-year horizon demographic projections.

The resulting vintage 2010 projections are evaluated against the census tract populations enumerated in the 2010 Census. Deviations from Census 2010 counts are conceptualized as measurement error (Hogan, 2003; Hogan and Mulry, 2014) and are reported in standard statistics of error and bias commonly employed in applied demography including: mean absolute percentage error (MAPE) and mean algebraic percentage error (MALPE) (Tayman, 1999; Tayman *et al*., 1998; Swanson and Tayman, 2012). These measures provide information on error (MAPE) and bias (MALPE) (Smith, 1987; Smith and Sincich, 1990; Smith *et al*., 2001). The implications of these findings are reviewed in light of their potential impact within the related fields of applied demography and population geography.

2. Materials and Methods

2.1 Study area & dataset

There is a direct relationship between the size of geographic unit and the accuracy with which it can be represented with an areal measurement. It follows that the population counts within smaller units, such as census blocks, will be better represented by a centroid as opposed to a polygon. Once the centroid of a census block represents its decennial year census population as an attribute, then it is possible to calculate an areal population density based on the geographical locations of various centroids with their associated population as the weighting factor. This enables conversion of population into a population density raster. Year 2010 census tracts in New Mexico (n=499) comprise the study area for this research. Figure 1 displays this geography for Bernalillo County, the most populous in New Mexico. Datasets were converted from block-level census counts (1990 and 2000) to population density rasters using the Kernel Density/Raster Manipulation approach described below. A cell (pixel) size of 150 meters and a search radius of 1,500 meters were used for the kernel density computations. These parameters were kept fixed for all other density calculations.

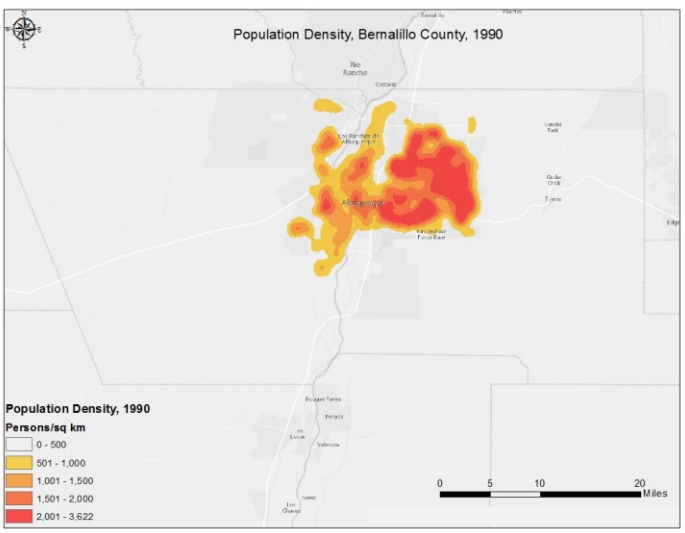

Figure 1. Raster-based kernel density surface, Bernalillo County, 1990 (Source: by endeavors of authors)

Densities for Bernalillo County were observed to change between 1990 (Fig. 1) and 2000 (Fig. 2 on page 28), at the pixel level — and changes in these densities were extrapolated forward to 2010 to construct a historical series of raster-based densities for census blocks from 1990 to 2000. The difference raster is shown in Figure 3 on page 28, which depicts the change in population density between 1990 and 2000 at each pixel level. These densities are standardized to persons per square km to facilitate area-based modeling, but form a semi-continuous surface in space that can be subdivided at a small enough level of resolution to permit re-aggregation to census tracts. These data result in a density function that is area-based and changes between 1990 and 2010. These densities form the basis of the raster-based projection of 2010 populations, based on shifts in density per pixel between 1990 and 2010. Once the method is validated, then the 2000 and 2010 decennial population density raster information is used to extrapolate forward to enable 2020 census tract population predictions.

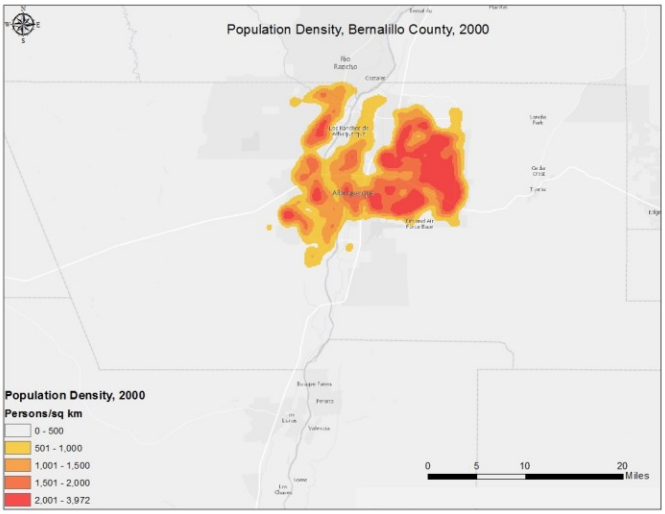

Figure 2. Raster-based kernel density surface, Bernalillo County, 2000 (Source: by endeavors of authors)

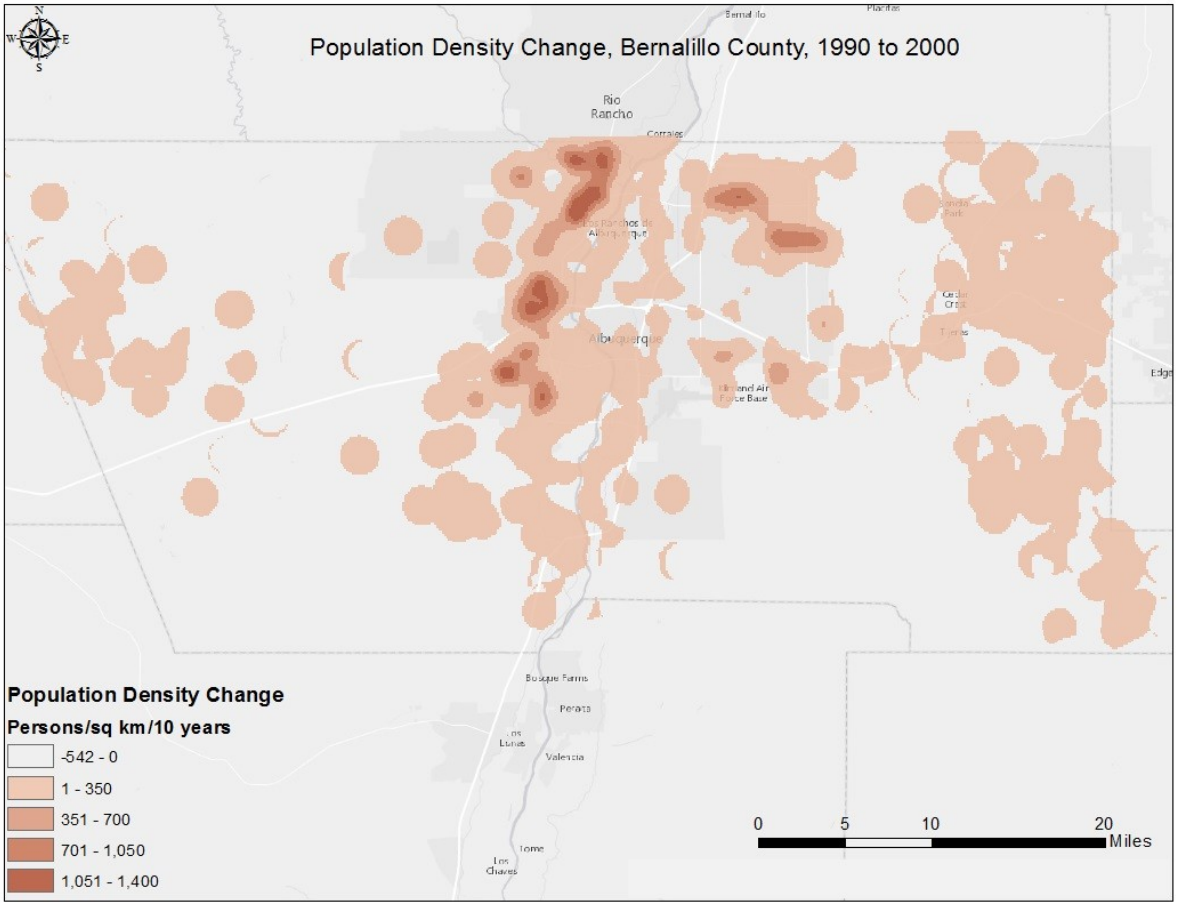

Figure 3. A difference raster computed by subtracting the 1990 population density by pixel from the 2000 value. This difference is used to project the density of 2010, and population projections at the census block/tract level are arrived at using Riemann sums. (Source: by endeavors of authors)

2.2 Kernel density model and raster manipulation in GIS

In order to project population, we need to understand local changes of population with respect to both space and time. Using fundamental concepts of calculus, the local change can be written as the partial derivative

$$\text{Local population change, } \Delta = \frac{\partial P(t,A)}{\partial t \partial A} \tag{1}$$

where $P(t, A)$ represents the population at a location at time t and A is the area over which the population lives. While data from decennial population census are available at various geographies including counties, census tracts, census block groups and census blocks, it is the smaller areas such as blocks and tracts that are of focus in this report. We first look at Bernalillo County, the most populated in New Mexico. Per the 2010 census, the county had 12,946 census blocks (average area 0.23 km^2) in 153 census tracts with an average of 85 census blocks per census tract. In order to evaluate the derivative of equation (1), the first step is to calculate the centroid of each block, assigning the value of the block's population as an attribute. Once this is accomplished, the partial derivative with respect to area A, namely $\frac{\partial P(t,A)}{\partial A}$ can be evaluated using kernel density. Thus while census geographies such as block boundaries change constantly from one decennial census to the next, it is possible to use the population (kernel) density and study its change with time at any location. Let us say that we have the exact block populations for two decennial time periods t_1 and t_2 available with $t_2 > t_1$. Equation (1) can be approximated using the finite difference, as

$$\Delta = \frac{\partial P(t,A)}{\partial t \partial A} \backsimeq \frac{1}{(t_2 - t_1)} \left[\frac{\partial P(t_2,A)}{\partial A} - \frac{\partial P(t_1,A)}{\partial A}\right] \tag{2}$$

Both the population densities on the right hand side (RHS) of equation (2) can be calculated at the two census times. Since the kernel density is a raster, it is easy to calculate the difference of two rasters. Now we make the assumption that the quantity $\Delta = \frac{\partial P(t,A)}{\partial t \partial A}$ remains constant from t_1 to t_2, and from t_2 to t_3 the next time period. Implicit in this assumption is that the change in birth, death and net migration of people has remained the same from t_1 to t_2 to t_3. This is strictly a mathematical extrapolation model to simplify matters and the justification for this assumption is *a posteriori* based on data validation discussed later. Thus,

$$\frac{\partial P(t_3,A)}{\partial A} = \Delta(t_3 - t_2) + \frac{\partial P(t_2,A)}{\partial A} \tag{3}$$

Again, since both the quantities on the RHS of equation (3) are rasters, they can be easily added in Arc GIS resulting in a predicted kernel density raster for the future time period t_3 making sure that any negative population densities are reset to zero..

Now that we have the population density predicted at time t_3, we obtain the population by small area (census blocks) for t_3 as follows. The population density at time t_3 is converted from raster to vector and then intersected with the census blocks geography defined for time period t_2. It is assumed that the census geography does not change from t_2 to t_3 since it is impossible to predict a future census blocks geography. Upon intersection a multitude of very small parcels much smaller than the size of an average block are created. These parcels are typically of a raster pixel size, which in this case was chosen as 150 meters square and about ten times smaller than an average block area. Let A_i represent this small parcel area. Then the population at time period t_3 is obtained by integration of equation (3), as

$$P(t_3, A) = \int \frac{\partial P(t_3,A)}{\partial A} \, \mathrm{d}A \tag{4}$$

For each census block *j*, the integral on the RHS is approximated by the Riemann sum over all the small parcels *i* in a census block *j*, as given by

$$P(t_3, A_j) \simeq \sum_{i=1}^{n} \frac{\partial P(t_3, A_j)}{\partial A} \Delta A_i \qquad (5)$$

where the summation occurs over all the parcels *n* inside the census block *j*, and repeating the evaluation of equation (5) for all the *j* census blocks. The partial derivative inside the summation in equation (5) is derived from equation (3) and the LHS of equation (3) is evaluated from adding the difference raster (as in Figure 3) to the kernel density raster of the previous period. The value of *n* in equation (5) is different for each block since each block is of a different size. The validity of equation (5) for predicting the population of the next decennial census for blocks using the past two census data was verified by using the 1990 and 2000 census block population for Bernalillo County to predict the 2010 block population which was then compared with the published 2010 census block populations. The Mean Absolute Percent Error was computed for each block. In addition, the census blocks were dissolved by census tract to compare predicted census tract population to published values and MAPE evaluated.

2.3 Software

ArcGIS software is used for mapping and computations. The authors are grateful for the reference maps supplied with the site software license with redistribution rights granted by the vendors that include ESRI, HERE, DeLorme, OpenStreetMap contributors, MapmyIndia and the GIS user community.

2.4 Modeling error and bias in demographic projections

Error and bias are measured in applied demography in an *ex-post-facto* fashion — by measuring various discrepancy statistics between a set of population estimates or projections and a reference standard, usually the decennial census counts (Hogan, 2003; Hogan and Mulry, 2014; Tayman, 1999; Tayman *et al*., 1998; Swanson and Tayman, 2012; Smith *et al*., 2001). Error is typically quantified as the mean absolute error, in numeric or percentage terms (MAPE), while bias — a tendency of error to be directional in nature — is typically measured using signed or "algebraic" errors (Cavanaugh, 1981; Smith, 1987; Smith and Sincich, 1990). This strategy is also employed in the current study, but due to the large differences in population sizes across tracts (which can tend to make errors difficult to compare across tracts), we employed mean absolute (MAPE) and algebraic *percentage* errors (MALPE) (Tayman *et al*., 1998; Tayman, 1999; Swanson and Tayman, 2012).

These results are reported by county, as well for groupings of counties by types including: Metropolitan, Micropolitan, and Rural. Metropolitan counties include Bernalillo, Dona Ana, San Juan, Sandoval, Santa Fe, Torrance and Valencia — each of which contains a larger urban center. Micropolitan counties are those with a smaller town centers such as Chaves County (city of Roswell) and Otero County (Alamogordo). Metropolitan and Micropolitan tracts have typically been growing between 1990 and 2010. Rural counties included sparsely populated areas such as Catron, De Baca, and Union counties — all of which have in recent times experienced population loss.

3. Results

Table 1. Error and Bias, 2010 Normalized Census Tract Population Projections

Error Type / Area	Mean Absolute Percentage Error (MAPE), %	Mean Algebraic Percentage Error (MALPE), %
Metropolitan	**31.75**	**9.52**
Bernalillo	22.71	0.34
Dona Ana	33.05	-3.18
San Juan	30.2	8
Sandoval	41.54	-5.26
Santa Fe	30.9	17.79
Torrance	27.99	24.07
Valencia	35.87	24.91
Micropolitan	**29.24**	**7.06**
Chaves	38.91	1.26
Cibola	16.09	-13.46
Curry	38.6	14.1
Eddy	31.03	1.22
Grant	29.15	14.17
Lea	27.19	-14.03
Lincoln	19.12	13.32
Los Alamos	14.29	5.07
Luna	31.5	17.04
McKinley	45.64	28.46
Otero	55.03	17.23
Rio Arriba	28.26	8.19
Roosevelt	26.34	-4.31
San Miguel	19.82	8.51
Taos	17.56	9.06
Rural	**28.47**	**5.09**
Catron	9.85	-9.85
Colfax	13.12	7.45
De Baca	1.58	-1.58
Guadalupe	1.66	1.66
Harding	43.6	-43.6
Hidalgo	59.44	45.66
Mora	21.82	21.82
Quay	11.52	-5.34
Sierra	80.91	48.13
Socorro	42.55	18.76
Union	27.1	-27.1
ALL	**29.51**	**6.92**

When calculations for MAPE were carried out at the census block level, the MAPE was high, around 142%. It is a well-known fact in demography that larger enumeration areas have lower prediction errors. Therefore, the projected population data by census block was aggregated to census tracts using a dissolve operation in Arc GIS. The block populations were summed for each tract and based on 153 census tracts, the MAPE for the 2010 census tract was calculated to be 22.71% for Bernalillo County, a large improvement in the projection accuracy. The process was repeated for the other 32 counties in the New Mexico.

Table 1 above reports error measures in an ex-post-facto evaluation of Vintage 2010 projections produced through construction of a raster-based kernel density function based on changes in density per pixel from 1990 to 2000. For the state as whole, mean absolute percentage errors at the census tract level were 29.51 percent and mean algebraic percentage errors suggested a small positive bias overall (6.92%). When broken out by the Metropolitan, Micropolitan, and Rural distinctions (see Materials and Methods), the errors vary from highest in Metropolitan tracts to lowest in rural tracts although the differences are small. Mean absolute percentage errors for Metropolitan census tracts ranged between a low of 22.71% (Bernalillo county) to a high of 41.54% (Sandoval County), centering upon 31.75% across all tracts. Mean algebraic percentage errors indicated small amounts of bias in Bernalillo County (0.34%) to large positive biases in Torrance and Valencia counties (24.07% and 24.91% respectively). Overall, bias was small, but positive across all Metropolitan census tracts (9.52%). Within micropolitan census tracts, both error and bias measures were somewhat similar (MAPE = 29.24 and MALPE = 7.06). Mean absolute percentage errors ranged between a low of 14.29% in Los Alamos County to a high of 55.03% in Otero County. Mean algebraic percentage errors were both positive and negative. Bias ranged from a high in the positive direction of 28.46% (McKinley County) to a high in the negative direction of -14.03% (Lea County). Rural tracts while smaller on the average in terms of bias (MALPE = 5.09%), however had a huge range from a high positive bias of 45.66% (Hidalgo county) to a high negative bias of -43.60% (Harding county).

Figure 4 shows a scatter plot of predicted versus actual census tract population for Bernalillo County for 2010. Figure 5 on the next page shows a map of 2010 tract population as predicted by our methodology for Bernalillo County.

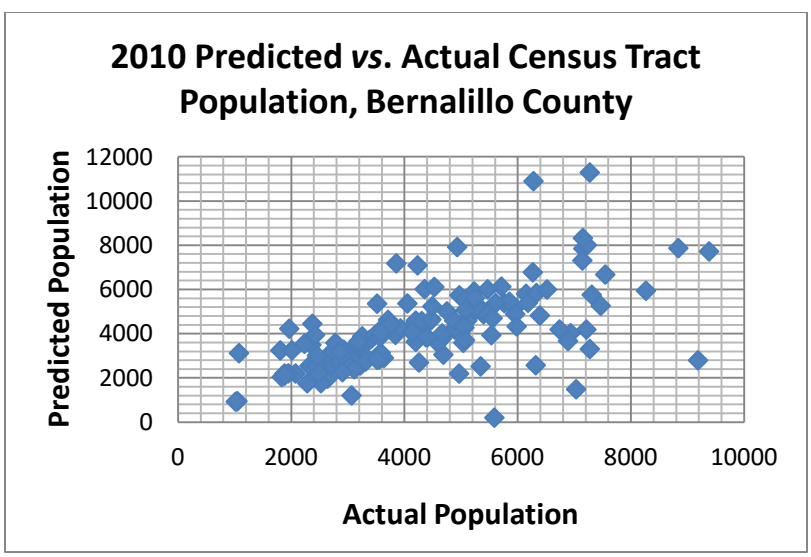

Figure 4. Predicted (vertical axis) versus actual (horizontal axis) census tract population, 2010

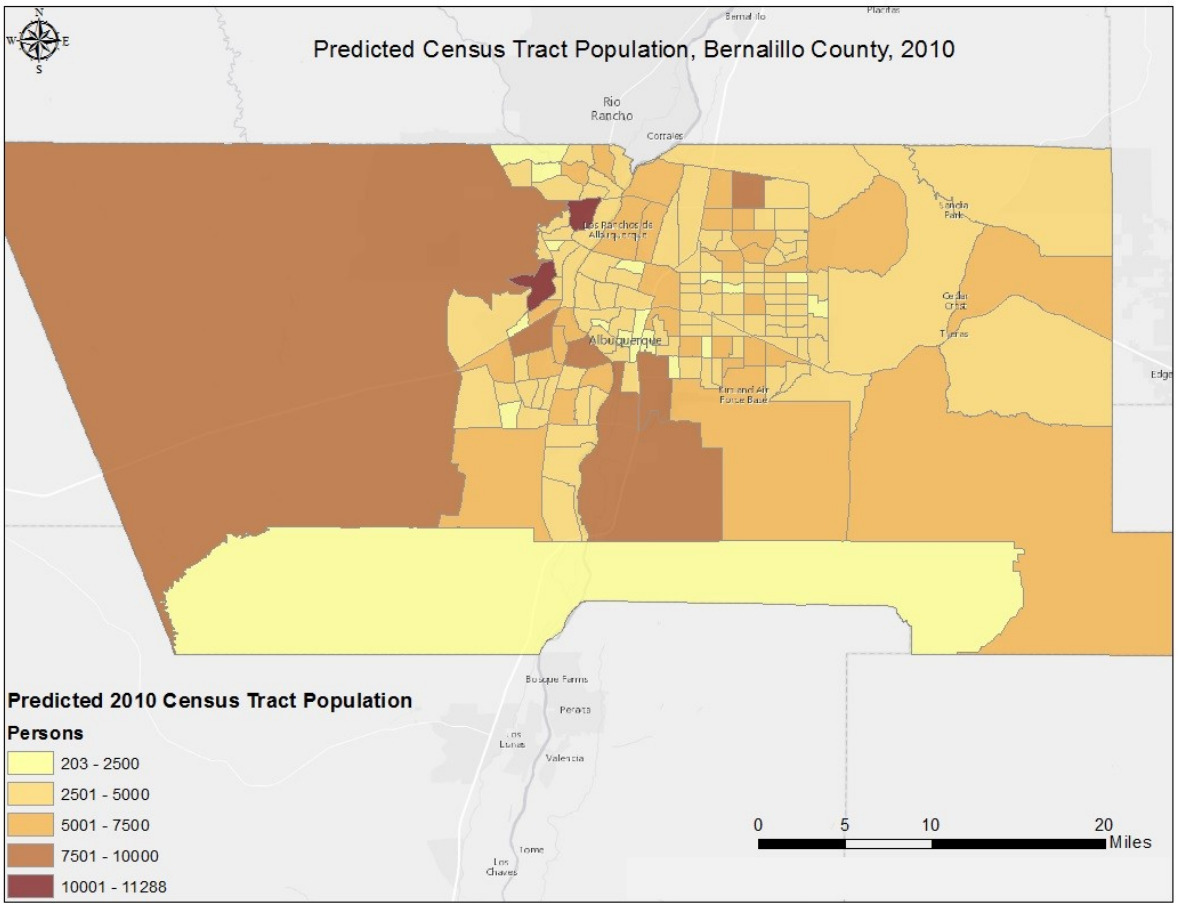

Figure 5. Predicted 2010 census tract population, Bernalillo County

When the predicted census tract populations were aggregated over all the tracts, the resulting predicted 2010 county population for Bernalillo was calculated to be 629,045. The actual 2010 county population was 662,564, a difference of 5.06%. This is consistent with our early observation that as the size of the geographical unit is increased, the prediction error decreases (for example, going from block level to tract level). Summing the predicted 2010 census tract populations over each county, the 2010 predicted county population was obtained and compared with the actual 2010 decennial census population data as shown in Table 2 on the next page. The MAPE is seen to be 13.64% and the MALPE 4.95%, a small positive bias. These numbers are well within the values reported by Hoque (2010) in his study of Texas counties.

Having established that the procedure as outlined above gives a reasonable prediction of populations at the census tract level, we use the same procedure to predict the 2020 census tract population for Bernalillo county using the actual 2000 and 2010 census block population data. The result of the prediction is shown in Figure 6 on page 35. The procedure was then repeated for other counties. Since no comprehensive population projection numbers at the census tract level are available for the year 2020, the tract population predicted for 2020 using our method was aggregated to the county level and compared with published projections (GPS, 2015) that use traditional population balance component methods. The Mean Absolute Percent difference between our values and the traditional method is 11.80% across counties.

Table 2. Comparison of prediction errors for various counties, 2010

County	2010 Number of Census Tracts	2010 Predicted County Population	2010 Actual County Population	% Error, County Population 2010
Bernalillo	153	629,045	662,564	-5.06
Catron	1	3,358	3,725	-9.85
Chaves	16	63,974	65,645	-2.55
Cibola	7	24,436	27,213	-10.20
Colfax	3	15,004	13,750	9.12
Curry	12	48,726	48,376	-0.21
De Baca	1	1,990	2,022	-1.58
Dona Ana	41	202,552	209,233	-3.19
Eddy	12	53,862	53,829	0.06
Grant	8	32,925	29,514	11.55
Guadalupe	1	4,765	4,687	1.67
Harding	1	392	695	-43.60
Hidalgo	2	6,829	4,894	39.54
Lea	18	54,732	64,727	-15.44
Lincoln	5	24,416	20,497	19.12
Los Alamos	4	18,427	17,490	2.66
Luna	6	31,243	25,095	24.50
McKinley	17	85,898	71,492	20.15
Mora	1	5,946	4,881	21.82
Otero	16	72,950	63,737	14.45
Quay	3	8,575	9,041	-5.15
Rio Arriba	9	44,655	40,246	10.96
Roosevelt	5	18,471	19,846	-6.93
San Juan	33	133,477	130,044	2.64
San Miguel	7	33,163	29,393	12.83
Sandoval	28	115,377	131,561	-12.30
Santa Fe	50	160,725	144,170	11.48
Sierra	4	16,172	11,988	34.90
Socorro	6	20,094	17,866	12.47
Taos	6	36,220	32,937	9.97
Torrance	4	21,820	16,383	33.19
Union	1	3,316	4,549	-27.10
Valencia	18	87,087	76,569	13.74
MAPE %				13.64
MALPE %				4.96

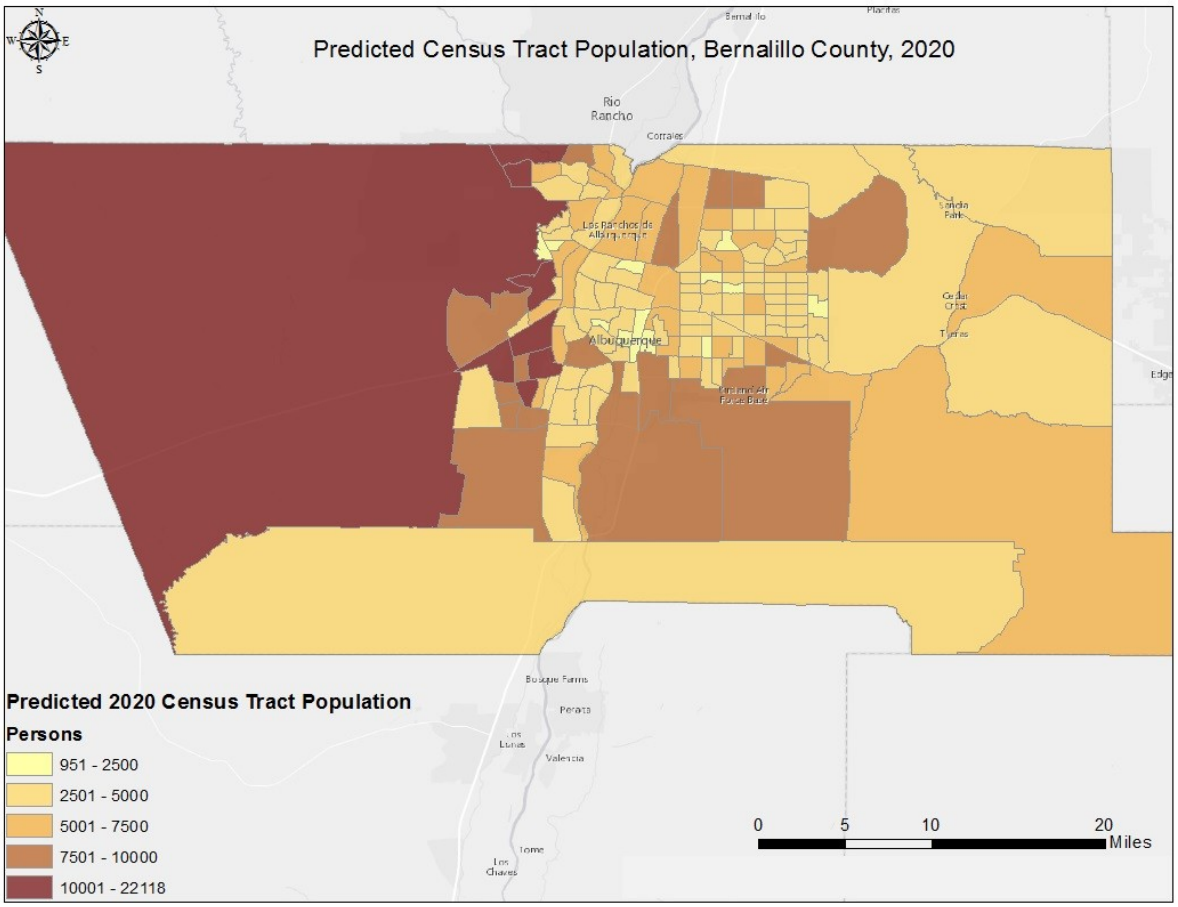

Figure 6. Predicted 2020 census tract population, Bernalillo County

4. Discussion

The overall picture of error and bias in the raster-based kernel density function-based projections reported here is one of surprisingly large errors or in other terms, substantial inaccuracy and bias. This result has been observed in every study to date of projection errors for US census tracts or similarly sized geographies (Baker *et al*., 2012, 2013, 2014a, 2014b; Hoque, 2010; Harper *et al*., 2003; Voss and Kale, 1985; Smith and Shahidullah, 1995 — but see a notable English exception in Lunn *et al*., 1998). In general, errors across such studies have ranged widely, from as little as 10.00% (Lunn *et al*., 1998) to as large as 100.00 percent or more (Baker *et al*., 2014b; Hoque, 2010). In general, errors ranging between 18.00 percent and 25.00 percent are to be expected for census tract total population projections over a ten-year projection horizon (Smith *et al*., 2001).

Hoque (2010) reports an analysis of population estimates (produced using the same methods one might use for projection population and therefore applicable to understanding the results presented here — see Baker *et al*., 2013 for a justification of this) that suggests for Texas counties that are sized similarly to census tracts (0 to about 5,000 persons) errors in total population projections may range anywhere between 4.70 percent at a low end and 19.85 percent at a high. Harper *et al*. (2003) report errors ranging between 7.50 percent and 35.10% for municipalities of a similar size (which might tend to be more accurate due to the use of administrative data on housing

— see Baker *et al*., 2012). Smith (1987) reports errors from 18.00 percent to 25.00 percent at the census tract level in Florida and Voss and Kale (1985) find similar error ranges (15.00 percent to 20.00 percent) in their study of Wisconsin census tracts. Smith and Shahidullah (1995) report 17.30% to 27.61% error, with a mean absolute percentage error centered upon 24.59% across all Florida census tracts included in their analysis. In general, population projections across all of these studies are less accurate and more biased as population size decreases. All of these findings are similar to those presented here, where errors are large across census tracts in general, but tend to be more accurate in more urbanized areas (Review Table 1).

Several studies previously conducted in New Mexico census tracts and block groups are directly relevant to the findings of this study. Baker *et al*., (2012) conducted a study of demographic estimates (equivalent to projections made using a housing unit method that updates a 2000 base forward to 2010) at the block group level within the City of Albuquerque that suggested much smaller errors than observed among Bernalillo County census tracts in this study (MAPE = 12.94%). This study involved geographic normalization of Census 2000 data into 2010 boundaries as well, but for smaller geographies than utilized in this study and with substantially more accurate results (nearly 10 percentage points). Baker *et al*., (2013) further present results of the equivalent of ten year projections made using two variants of the component method (Swanson and Tayman, 2012; Shyrock and Siegel, 1980; Hoque, 2010) and found errors ranging between 56.74% and 58.35% in a representative sample of New Mexico census tracts (n=182) that is broadly comparable to the findings presented here. Baker and coworkers note the presence of a significant number of outliers in their analysis (see also Hoque, 2010) and report much smaller *median* errors therefore — between 26.06 and 28.35 percent. In a follow-up study of census tract level errors in a 10 year population forecast that include spatially weighting but were excluded to urbanized areas (n=221), Baker *et al*. (2014b) report much smaller errors: from 5.19 percent to a high of 10.98 percent depending upon the method of weighting utilized.

5. Conclusions

The errors reported in this study are generally similar to the previous studies of at similar geographic levels. The errors reported in similar studies are smallest in more population (Metropolitan) census tracts and become larger as the size of the population to be projected decreases (Smith, 1987; Smith and Sincich, 1990; Smith *et al*., 2001). While the dependence of prediction accuracy on the population size of the county was not very significant in our studies, the overall magnitude of error and bias observed in this study falls within the range of those of previous studies. Alternatively, it does not suggest dramatic differences in error or bias are observed using this procedure and the facility with which data may be recast in alternative geographies is a significant advantage over polygon-based models that *all such studies reviewed here depend upon inherently*. Zandbergen and Ignizio (2010) have suggested that normalization bias can account for as much as 40.00 percent or greater errors in polygon-based models (see also Tobler, 1979; Fisher and Langford, 1995; Sadahiro, 2000; Leyk *et al*., 2013). Although the current study is incapable of isolating normalization error specifically, the results presented here suggest that the ease of normalization associated with the raster-based model does not negatively affect accuracy or bias and, therefore, may be of substantial future importance to applied demographers and population geographers interested in small-area population modeling and dynamics. The results of the current study are subject to the same limitations as any trend-based extrapolation procedure utilized to project future populations based on historical patterns. Under such conditions, rapidly changing

areas will tend to be estimated incorrectly: areas of unprecedented growth will tend to be under-projected while those characterized by reduced growth from the preceding inter-censal period will tend to be over-estimated (Hoque, 2010; Baker *et al*., 2013; Shyrock and Siegel, 1980; Keyfitz, 1981). Addressing geographic normalization bias has no impact on this fundamental limitation of trend-extrapolation. Progress in this areas will be made through exploration of how to use ancillary data to improve projections by incorporating current data (Zandbergen and Ignizio, 2010; Swanson and Tayman, 2012; Baker *et al*., 2014b).

The results of this study suggest that further exploration of similar raster-based methods for addressing geographic normalization issues will be worthwhile. In the context of population projections for small areas such as census tracts, further studies should attempt to isolate the error associated with geographic normalization rather than bundling this form of error remediation into the overall error measured in the population projection. This could be accomplished by using the same approach in the context of converting population data across alternative geographies that are incongruent, but for which census microdata are equivalently-reported (Zandbergen and Ignizio, 2013; Voss *et al*., 1999). For example, one might normalize census tract data to zip codes, and then compare the resulting estimates to the published zip-code level tabulations based on microdata from the decennial census. With a better grasp of the contributing components of error in population projections (Hogan and Mulry, 2014; Baker *et al*., 2013), improved use of GIS-based methods such as the one outlined here may lead to long-awaited improvements in demographic projection accuracy. The high degree of migration among certain populations such as natives as well as from across the international border from Mexico makes the task more difficult.

References

[1] Baker J., Alcántara A., Ruan X.M., Watkins K. (2012). "The Impact of Incomplete Geocoding on Small-Area Population Estimates", *Journal of Population Research*, 29(1): 91-112.

[2] Baker J., Alcántara A., Ruan X.M., Watkins K, Vasan S. (2013). "A Comparative Evaluation of Error and Bias in Census Tract-level Age/Sex-Specific Population Estimates: Component I (Net-migration) vs. Component III (Hamilton-Perry) ", *Population Research and Policy Review*, 32(6): 919-942.

[3] Baker J., Alcántara A., Ruan X.M., Ruiz D., Crouse N. (2014a). "Sub-County Component Estimates Using Administrative Records: A Case-study in New Mexico", In: *Emerging Techniques in Applied Demography*. Nazrul Hoque and Lloyd Potter (Eds.), New York: Springer. Chapter 6. pp. 63-80.

[4] Baker J., Alcántara A., Ruan X.M., Watkins K., Vasan S. (2014b). "Spatial Weighting Improves Accuracy and Reduces Bias in Small-Area Demographic Forecasts of Urban Populations", *Journal of Population Research*, 31(4): 345-359.

[5] Cavanaugh F. (1981). "The Census Bureau's 1980 Census Test of Population Estimates", In: *Small Area Population Estimates — Methods and Their Accuracy and New Metropolitan Definitions and Their Impact on the Private and Public Sector, Series GE-41, No. 7.* Washington DC: Government Planning Office.

[6] Fisher P.F., and Langford M. (1995). "Modeling the Errors in Area Interpolation Between Zonal Systems and GIS", *The Annals of Regional Science*, 27(2): 67-78.

[7] George M.V. (2004). "Population Projections", In: J. Siegel and D. Swanson (Eds.), *The Methods and Materials of Demography*. New York: Springer.

[8] GPS (2015). "Population Projections", Bureau of Business and Economic Research, University of New Mexico Data Bank. [Online]Available at: https://gps.unm.edu/pru/

[9] Harper G., Coleman C., and Devine J. (2003). "Evaluation of 2000 Subcounty Population Estimates", Population Estimates Branch, Population Division. US Census Bureau.

[10] Hogan H. (2003). "Accuracy and Coverage Evaluation: Theory and Design", *Survey Methodology*, 29(2): 129-138.

[11] Hogan H., and Mulry M. (2014). "Assessing Accuracy in Postcensal Estimates: Statistical Properties of Different Measures", In: Nazrul Hoque and Lloyd Potter (Eds.), *Emerging Techniques in Applied Demography*, New York: Springer. Chapter 9. Pp. 119-136.

[12] Hoque N. (2010). "An Evaluation of Small-Area Population Estimates Produced by Component Method II, Ratio Correlation, and Housing Unit Methods for 1990", *The Open Demography Journal*, 6(3): 18-30.

[13] Keisler H.J. (2013). *Elementary Calculus: An Infinitesimal Approach* (Third edition). Mineola: Dover.

[14] Keyfitz N. (1981). "The Limits of Population Forecasting", *Population and Development Review*, 7(4): 579-593.

[15] Lunn D., Simpson S., Diamond I., Middleton L. (1998). "The Accuracy of Age-Specific Population Estimates for Small Areas in Brittain", *Population Studies*, 52(3): 327-344.

[16] Sadahiro Y. (2000). "Accuracy of Count Data Transferred Through the Areal Weighting Interpolation Method", *International Journal of Geographical Information Science*, 14(1): 25-50.

[17] Shyrock H. and Siegel J. (1980). *The Methods and Materials of Demography* (Vol. II), Washington DC: US Bureau of the Census.

[18] Smith S. (1987). "Tests of Forecast Accuracy and Bias for County Population Projections", *Journal of the American Statistical Association*, 82(400): 991-1003.

[19] Smith S. and Shahidullah M. (1995). "An Evaluation of Projection Errors for Census Tracts", *Journal of the American Statistical Association*, 90(429): 64-71.

[20] Smith S. and Sincich T. (1990). "The Relationship Between Length of the Base Period and Population Forecast Errors", *Journal of the American Statistical Association*, 85(410): 367-375.

[21] Smith S., Tayman J, Swanson D. (2001). *State and Local Population Projections: Methodology and Analysis*, New York: Plenum.

[22] Swanson D. and Tayman J. (2012). *Subnational Population Estimates*, New York: Springer.

[23] Tayman J., Schafer E., Carter L. (1998). "The Role of Population Size in the Determination of Prediction of Population Forecast Errors: An Evaluation Using Confidence Intervals for Subcounty Areas", *Population Research and Policy Review*, 17(1): 1-20.

[24] Tayman J. (1999). "On the Validity of MAPE as a Measure of Forecast Accuracy", *Population Research and Policy Review*, 18(4): 299-322.

[25] Tobler W.R. (1979). "Smooth pycnophylactic interpolation for Geographical Regions", *Journal of the American Statistical Association*, 74(367): 519-530.

[26] Voss P. and Kale B. (1985). "Refinements to Small-Area Population Projections Models: Results of a test based on 128 Wisconsin communities", Presented at the Annual Meeting of the Population Association of America, 28-30 March, 1985.

[27] Voss P., Long D., Hammer R. (1999). "When Census Geography Doesn't Work: Using Ancillary Information to Improve the Spatial Interpolation of Spatial Demographic Data", Center for Demography and Ecology, University of Wisconsin, Working Paper No. 99-26.

[28] Zandbergen P. and Ignizio D. (2010). "Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates", *Cartography and Geographic Information Science*, 37(3): 199-214.