

8-10-2013

Evaluating The Performance Of A Filtered Area Weighting Method In Population Estimation For Public Health Studies

Andrew Chiang
Graduate Student

Follow this and additional works at: http://scholarworks.gsu.edu/geosciences_theses

Recommended Citation

Chiang, Andrew, "Evaluating The Performance Of A Filtered Area Weighting Method In Population Estimation For Public Health Studies." Thesis, Georgia State University, 2013.
http://scholarworks.gsu.edu/geosciences_theses/62

This Thesis is brought to you for free and open access by the Department of Geosciences at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Geosciences Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

EVALUATING THE PERFORMANCE OF A FILTERED AREA WEIGHTING METHOD IN POPULATION
ESTIMATION FOR PUBLIC HEALTH STUDIES

by

ANDREW CHIANG

Under the Direction of Dr. Dajun Dai

ABSTRACT

Areal interpolation is a geospatial analysis method that allows researchers to estimate the incidence of a phenomenon in one set of areal units given data based on different areal units. One practice implemented in conjunction with areal interpolation is known as filtered area weighting, in which ancillary data is introduced to exclude specific areas from the areal units based on certain criteria, thus providing a more accurate representation of population distribution. This thesis examines the benefits that filtered area weighting can provide to population estimation using a hospital accessibility case study. The study shows that filtered area weighting does not always improve population estimation as expected, and suggests that the ancillary data and the criteria employed to exclude areas from analysis needs particular attention in future research when the filtered area weighting method is used.

INDEX WORDS: Geospatial, Public health, Areal interpolation, GIS, Filtered area weighting, Estimation

EVALUATING THE PERFORMANCE OF A FILTERED AREA WEIGHTING METHOD IN POPULATION
ESTIMATION FOR PUBLIC HEALTH STUDIES

by

ANDREW CHIANG

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2013

Copyright by
Andrew Franklin Chiang
2013

EVALUATING THE PERFORMANCE OF A FILTERED AREA WEIGHTING METHOD IN POPULATION
ESTIMATION FOR PUBLIC HEALTH STUDIES

by

ANDREW CHIANG

Committee Chair: Dajun Dai

Committee: Jeremy Diem

Timothy Hawthorne

Electronic Version Approved

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2013

DEDICATION

This thesis, which represents a culmination of knowledge in my academic and professional career to date, is dedicated to my parents, who have constantly pushed me to learn new things and look forward to the next step of my life, and to my loving wife, Catherine, who has been there for me through the entire process of balancing my academic, professional, and personal life.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Dajun Dai for providing guidance and assistance throughout the process of me developing, writing, and proofreading this thesis. In addition, I would like to acknowledge Dr. Jeremy Diem and Dr. Tim Hawthorne for serving on my thesis committee and providing additional guidance and support of my work. I would also like to acknowledge the Geospatial Research, Analysis, and Services Program (GRASP) at the Centers for Disease Control and Prevention in Atlanta, GA, who provided key guidance and technological resources to help me complete this work. In particular, I wish to acknowledge Andy Dent for his guidance regarding geospatial analysis techniques and for providing me fun and engaging work through which I could develop the ideas for this thesis, Elaine Hallisey and Efomo Woghiren for assistance testing the tools that were created for this thesis and for providing insight into numerous geospatial analysis and statistical concepts, and Jeff Henry for providing me a framework from which I could develop his prior work on area proportion and areal interpolation analysis into this thesis concept, and a powerful and useful tool for the good of people everywhere.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	v
LIST OF TABLES	vii
LIST OF FIGURES.....	viii
1. INTRODUCTION	1
2. LITERATURE REVIEW.....	2
3. METHODOLOGY.....	6
4. DATA AND SOFTWARE OVERVIEW	16
5. RESULTS	19
6. DISCUSSION.....	24
7. CONCLUSION.....	28
REFERENCES	30

LIST OF TABLES

Table 1: Improvement towards LandScan Benchmark provided by Filtered Area Weighting	20
Table 2: Mean benchmarks and estimates calculated in Non-Stratified Random Simulation	20
Table 3: Mean benchmarks and estimates calculated in Stratified Random Simulation	23

LIST OF FIGURES

Figure 1: Area Proportion Methodology Flowchart 8

Figure 2: User Interface of the Created Automated Area Proportion Tool..... 8

Figure 3: Visualization of the Catchment Area Created by Buffering Features of Interest 10

Figure 4: DeKalb County blocks and hospitals/medical centers with Land Use and Water Areas 11

Figure 5: DeKalb County blocks with Excluded Areas Erased 12

Figure 6: Catchment Area with Excluded Areas Erased..... 13

Figure 7: Population Density of DeKalb County Census blocks 15

Figure 8: LandScan USA dataset for DeKalb County, GA 19

Figure 9: Improvement towards LandScan in for Half-Mile Catchment Area 21

Figure 10: Improvement towards LandScan in for One-Mile Catchment Area 22

Figure 11: Improvement towards LandScan in for Two-Mile Catchment Area 22

1. INTRODUCTION

In public health studies, geographic information systems (GISs) are often utilized to perform quantitative analysis and calculations quickly. One method that has seen increasing use in public health studies is known as *areal interpolation*. Areal interpolation is a form of spatial interpolation used by analysts in order to estimate the amount of a phenomenon (usually population) in a particular catchment area given the distribution of that same phenomenon using different spatial boundaries or areal units (Cromley, McLafferty 2012). The basic principle of estimation through areal interpolation is based on calculating the percentage of an areal unit that resides within a catchment area in a study area. It is then assumed that a proportionate amount of the phenomena associated with that areal unit also resides within the catchment area (Goodchild, Lam 1980). The following proportional equality is applied to all areal units in the study area...

$$\frac{\text{Area of Areal Unit within Catchment Area}}{\text{Total Area of Areal Unit}} = \frac{\text{Phenomena of Areal Unit within Catchment Area}}{\text{Total Phenomena in Areal Unit}}$$

It therefore follows that the total number of phenomena in the catchment area is the sum of all phenomena in each areal unit that lies at least partially in the catchment area. Population units that are wholly encompassed in the catchment area are assumed to have 100% of their associated population included in the catchment area. Similarly, population units that are entirely outside the catchment area are assumed to have 0% of their associated population included in the catchment area.

However, areal interpolation makes one key assumption about the areal unit features that are being used: that the phenomena within that areal unit are equally distributed across its entire area (Goodchild, Lam 1980). When observing the reality of population distributions in most areas, this is not a correct assumption to make. Some areal units in urban areas may be more densely populated, whereas the distribution in other areas can be more distorted due to natural geographic features or land use

constraints (Fisher, Langford 1996). Historically, one method that geographers and analysts have used to compensate for the error that can occur with this assumption of homogeneity is known as *filtered area weighting* (Flowerdew, Green 1989; Maantay, et al 2007). Filtered area weighting involves the use of an ancillary dataset to filter out areas in which the phenomenon is not expected to occur. This is often used to exclude areas that are considered uninhabited or uninhabitable when estimating population (Reibel, Agrawal 2007). These areas can include areas such as parks, industrial areas, water bodies, and other similar features (Eicher, Brewer 2001). By using filtered area weighting, these uninhabited areas are erased from the areal units, which exclude them from area calculations in areal interpolation, and thus hopefully provide a more accurate estimation (Maantay, et al 2007).

In this thesis, I will evaluate the performance of the filtered area weighting technique in population estimation in a public health context. Public health studies have used GIS and geospatial analysis to conduct population estimation. Practices such as disease mapping (Cromley, McLafferty 2012), pollutant dispersion forecasting (Dent 2000), and mapping socioeconomic data (Goodchild, Anselin, Deichmann 1992) rely on geospatial data and analysis practices to produce accurate predictions and estimations on future disease or public health activity. However, detailed exploration of areal interpolation and filtered area weighting in particular is limited. I seek to evaluate the accuracy of the filtered area weighting method over regular areal interpolation without filtered area weighting to determine if its use in public health population estimation can yield improved results.

2. LITERATURE REVIEW

Much of the discussion on spatial interpolation and population estimation is grounded in the practice of dasymetric mapping. Dasymetric mapping is the practice of mapping certain discrete or continuous phenomena together with areal units that may not coincide, and has its origins in population surface and density estimation (Wright 1936). Research on applying dasymetric mapping to population

estimation continues even in modern research. A common application of population estimation using dasymetric mapping techniques is modeling populations using Census data (Holt, Lo, Hodler 2004). Holt, et al, utilize these techniques on vector polygon data by overlaying Census blocks from two different years and using the overlapping areas and populations to calculate approximate populations of particular areas in between Census years. The same dasymetric practices have also been used to generate continuous surface population data as well through the use of remote sensing techniques (Wu, et al 2005). Remote sensing classification is one method that is used to identify uninhabitable areas for the purposes of filtered area weighting, and when used in conjunction with ancillary data, this can provide much more refined population distribution estimates (Mennis 2008). The basic practice of dasymetric mapping has also been expanded to include consideration for the magnitude at which different data or variables can affect the distribution of population or phenomena, as in so-called “intelligent” dasymetric methods (Mennis, Hultgren 2006). This discussion of the ancillary data that is incorporated into dasymetric mapping applications is of particular relevance to this thesis because I intend to show a relationship between the ancillary data itself and the improvement realized through filtered area weighting.

Research on dasymetric mapping has also yielded much discussion on different methods of spatial interpolation. Multiple methods of spatial interpolation have been discussed, including areal interpolation based on Wright’s dasymetric mapping, Tobler’s pycnophylactic method, and Xie’s geometric network-based hierarchical interpolation method (Hawley, Moellering 2005). Areal interpolation utilizes dasymetric mapping techniques to estimate populations based on the intersections or coincidence of vector data that uses different areal units (Goodchild, Lam 1980), and is the basis for this thesis. Compared to the other methods of interpolation, areal interpolation is usually considered the simplest method because it does not rely on complex mathematical and statistical distributions like pycnophylactic interpolation does (Tobler 1979), and it can easily be done binomially as opposed to

relying on a hierarchy of weighting classifications like road network interpolation does (Xie 1995). However, despite their differences, these methods can still be examined and considered simultaneously. For example, a population surface generated through pycnophylactic interpolation can be refined using dasymetric practices, such as filtering uninhabitable areas, and this improved population estimation surface can be considered more desirable to analysts for its inclusion of ancillary data (Kim 2010). In another study, Comber, et al, use a combined dasymetric and pycnophylactic approach to map agricultural land use from a combination of datasets (Comber, et al 2008). The idea of combining multiple interpolation approaches has proven to be valuable in the development of more intelligent modeling and data derivation algorithms. This is demonstrated in various research generating population models using various data such as land cover data and remote sensing techniques (Briggs 2007), population classification and distribution data (Langford et al 2008), and specific cadastral data that delineates areas of interest or exclusion (Maantay 2007). The application of multiple interpolation techniques is also illustrated in the LandScan USA dataset, which combines several techniques to generate a continuous population surface estimation (Bhaduri, et al 2007).

Across all of these applications of spatial interpolation, ancillary data that is introduced to the analysis remains an important factor in the overall quality of the population estimation itself. As with any population model that relies on an assumption about the population's distribution, error in the population estimation is inevitable. Previous studies have sought to quantify this error using simulations in which an experiment with an infinite number of possible parameters is repeated multiple times, and in each of the iterations, certain parameters are randomly generated (Fisher, Langford 1995). In Fisher and Langford's work on Monte Carlo simulations and areal interpolation, catchment areas were randomly generated using the same parameters for each of the iterations, and a population estimate for the catchment area was generated. The population estimates across all trials were then analyzed to calculate the overall error. Multiple methods of interpolation were used to generate the population

estimates, and areal interpolation based on dasymetric mapping yielded the lowest error (Fisher, Langford 1995). When using “intelligent” interpolation methods, these errors can potentially be reduced further based on the accuracy and relevance of the ancillary data being used. The closer that the ancillary data fits the realistic population distribution, the lower the error will be (Flowerdew, Green 1989; Sadahiro 1999). The importance of the relevance of ancillary data will be demonstrated in this thesis.

In addition to understanding areal interpolation as a practice, we must also consider the significance of areal interpolation to the realm of public health research. In recent years, GIS has played an increasing role in public health research (Baum 2003). Analysis methods such as disease mapping, geographic correlation studies, and cluster detection analysis (Aylin, et al 1999) have allowed health researchers to identify patterns in disease and phenomena incidence, which has led to many advances in our knowledge of how to handle disease outbreaks and endemics (Wakefield 1999). The spatial interpolation techniques described earlier are often used to estimate at-risk populations in specially defined catchment areas. The definitions of these catchment areas vary greatly depending on the indicators being studied in the research. One important indicator that is often used as a measure of public health in a community is geographic proximity to certain features of interest (Besleme 2007). These features of interest may be sources of air pollution (Dent, et al 2000; Chakraborty, et al 2011; Maroko 2009) and sites that store or use toxic and hazardous materials (Lu, et al 2000), including agricultural pesticides (Reynolds, et al 2004). Studies have identified positive correlations between poor health and proximity to perceived environmental risk areas, which make examining the proximity of populations to these high-risk areas particularly relevant to public health and safety.

In a similar vein to estimating at-risk populations using proximity analysis, GIS has also become a key tool in defining the notion of “accessibility” in public health. Proximity to features associated with

positive community health, such as parks (Bedimo-Rung et al 2005), hospitals and medical providers (Rosero-Bixby 2004, Hendryx et al 2002), and community centers and recreation centers (Norman et al 2006), is a relevant indicator that is focused on in studies regarding social justice and community health. Proximity to features associated with negative community health is also used as a key indicator of assessing overall health levels. These features are often not only used as indicators of health, but also as indicators of community affluence and social status as well (Chakraborty, et al 2011). This leads many to also analyze the effects of proximity and access to certain features and services on community social capital (Mittelmark 2001) as well as community health. These influences of population estimates on social capital and policy decisions illustrate the importance of accurate calculations when performing proximity analysis. The accuracy of the population model generated through spatial interpolation has a direct correlation to the accuracy of the estimate of an at-risk population or an accessible population (Hay, et al 2005).

Although areal interpolation has been used many times in prior public health studies for population estimation, the use of filtered area weighting has not been examined in as much detail. This thesis will seek to evaluate the performance of the filtered area weighting approach for population estimation in public health through empirical data analysis. Findings from this research will provide information for future use of this method regarding its improvement on population estimation.

3. METHODOLOGY

To evaluate the performance of the filtered area weighting method in a relevant public health application, a hypothetical case study was devised, which sought to determine the total number of people within one (1) mile, one half (0.5) mile, and two (2) miles of a hospital or a general medical services provider in DeKalb County, Georgia. Hospitals and medical centers were selected as the features of interest because of their relevance to healthy community design studies and their positive

correlation to overall community health (McGlynn, et al 2003). DeKalb County was chosen for this study partially because of geographic familiarity, but also because DeKalb County encompasses a wide variety of areas of development and population density. It contains some densely populated urban areas in the metropolitan Atlanta area (including the Decatur neighborhood and areas west), suburban areas such as Chamblee, Dunwoody, Tucker, and Stone Mountain, and more sparsely populated rural areas, such as Lithonia and points east. DeKalb County also contains a high number of land use areas that can be included in an exclusion set, including golf clubs, country clubs, and major park and water area features, which tend to take up a large unpopulated area. As for the catchment area buffer distances, although the buffer distances that define a catchment area are often determined ad hoc depending on a study, I chose to use a half mile, one mile, and two mile distances because they are typical standard distances for catchment areas used in many proximity studies at the Centers for Disease Control (CDC). In addition, two miles was the largest catchment area definition used because catchment areas much larger than two miles from a hospital or medical center would have encompassed most, if not all of the area of DeKalb County. At that point, calculating the catchment areas that include most of DeKalb County would not provide us any meaningful data. This study case will use 2010 United States Census blocks as the areal unit, as Census blocks are the most detailed areal unit readily available.

In order to expedite the execution of areal interpolation in the various cases that will be used in this thesis, an automated tool for performing areal interpolation was created. To do this, a geoprocessing algorithm was created to define not only how the calculations are performed, but also how the tool will function. The algorithm that defines this tool's function to automate the areal interpolation analysis follows below. **Figure 1** shows a flow chart of the algorithm used for this tool.

1. The user inputs the input parameters of the tool. **Figure 2** shows the user interface that was created to define the inputs for the tool. These inputs include...

a. The feature class to be analyzed, hereafter referred to as the “features of interest”.

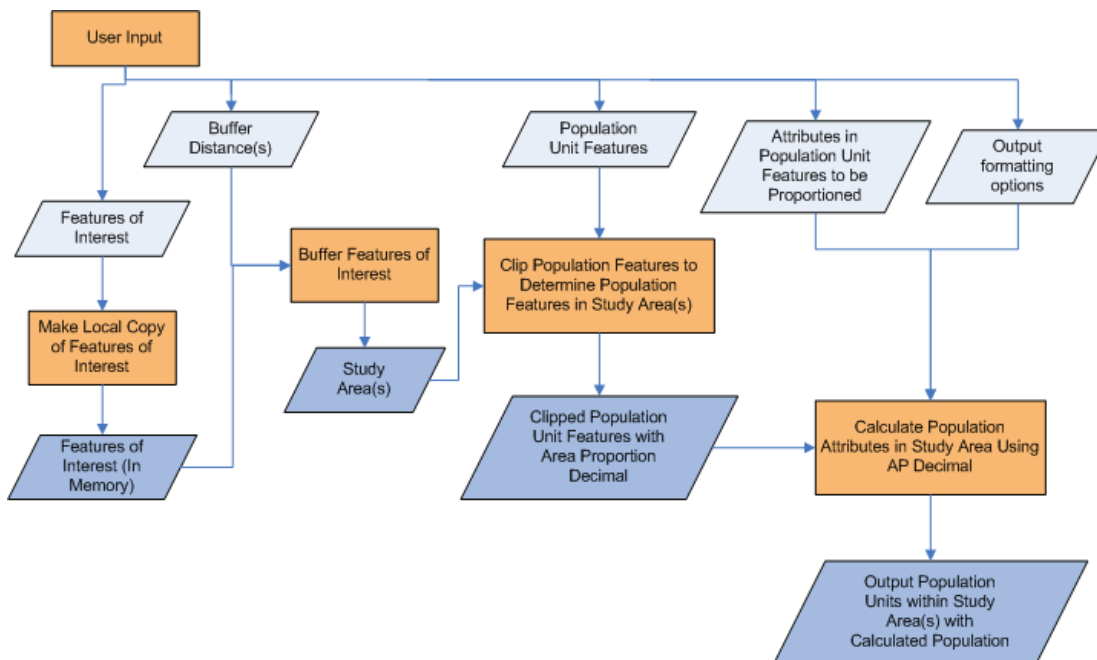


Figure 1: Areal Interpolation Methodology Flowchart

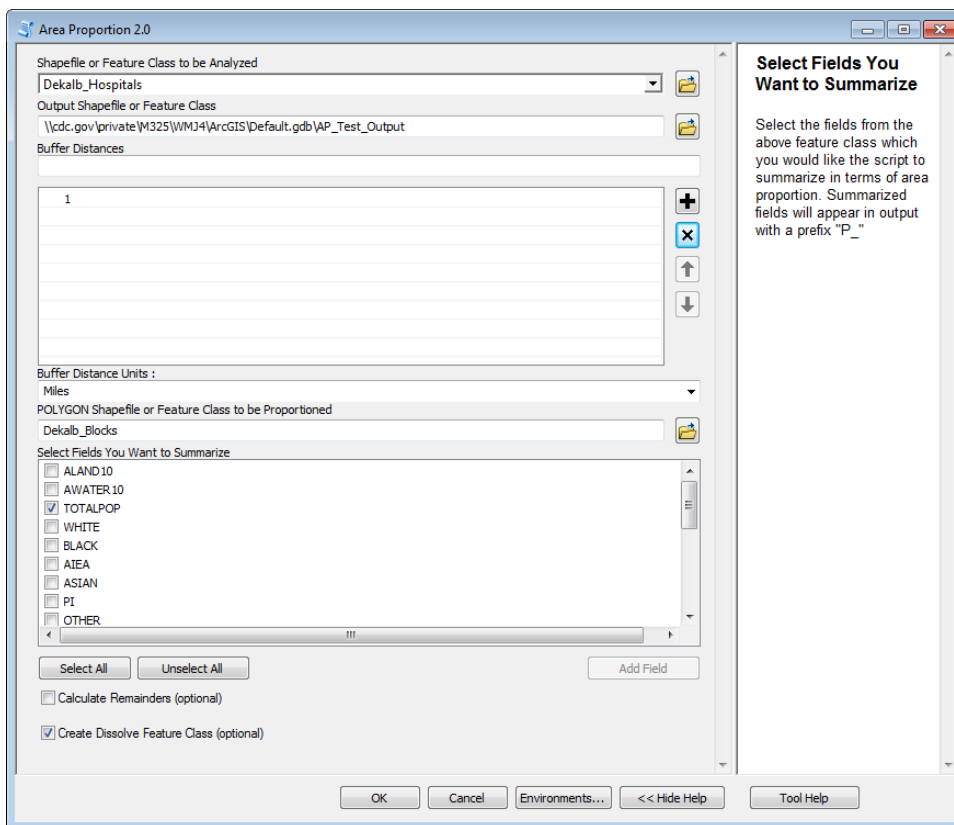


Figure 2. User Interface of the Created Automated Area Proportion Tool

This feature class contains the features that are of interest to the particular user, and to which proximity of the population to will be analyzed.

- b. One or more distances from the features of interest, which is used to create buffer areas around the features of interest, therefore creating the catchment area(s)
 - c. The feature class to be proportioned using the catchment area parameters, or the areal units. These will usually be defined as Census blocks, Census tracts, counties, or any other areal unit with a population attribute.
 - d. A list of attributes in the areal units feature class that will be proportioned based on the area proportion of the population unit that is in the catchment area. These represent the phenomenon/phenomena being studied.
 - e. Output formatting options, including polygon dissolving options
2. The user confirms the input parameters, and starts the tool's operation
 3. An in-memory copy of the features of interest feature class is created. The features of interest are buffered using the Buffer function based on the distances specified by input 1.b. above.
 4. For each buffer distance specified, a separate buffer feature is created. The buffers created by buffering each separate feature of interest are then dissolved together based on the buffer distance, thereby creating one (potentially multi-part) polygon feature that represents all areas within the study area that are within the catchment area. For the purposes of this study, areas in DeKalb County that are within the defined distance of a hospital or medical center that is outside of DeKalb County will still be considered as within the catchment area (see **Data and Software Overview** section). **Figure 3** shows the resulting catchment area created with this process.

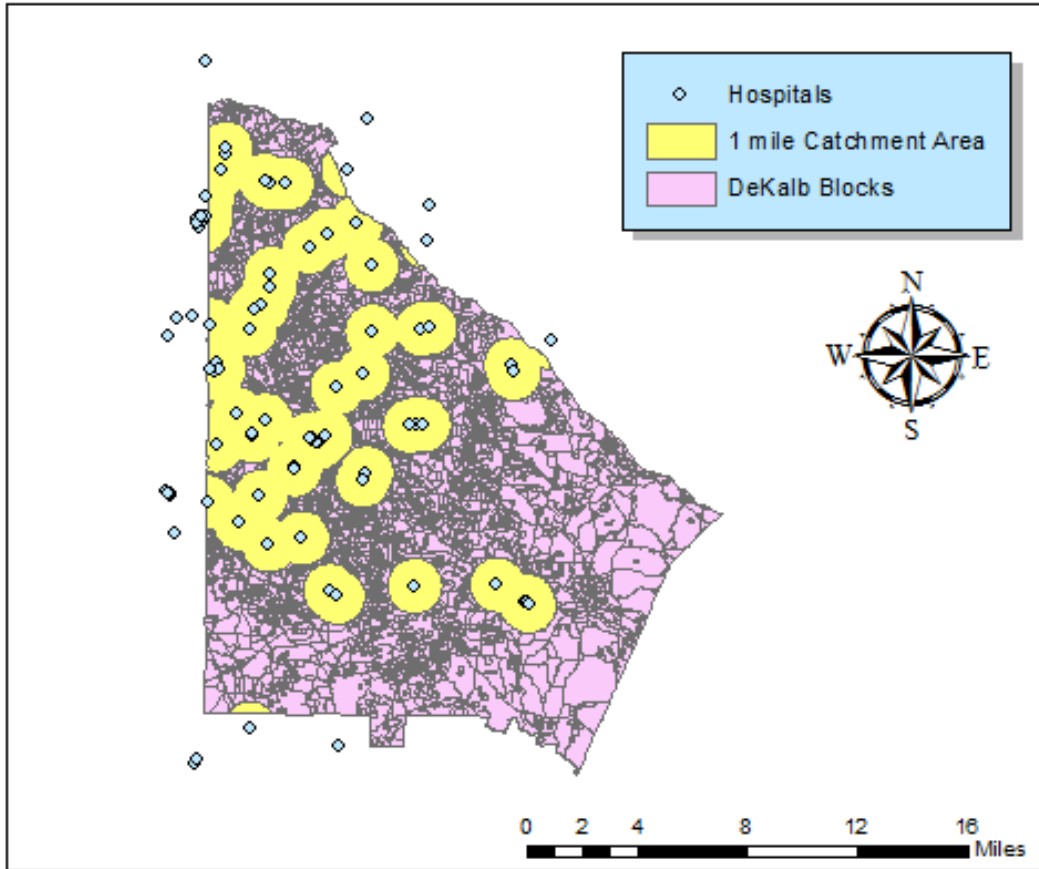


Figure 3. A Visualization of the Catchment Area Created by Buffering the Features of Interest

5. The areal unit features are then clipped using the Clip function by the buffer polygon features created in Step 4. The output of this step will be a set of areal unit polygon features that represent the areas of the areal units that fall within the catchment area.
6. The percentage of each areal unit that falls within the catchment area is then calculated. This is done by taking the area of each areal unit that resides within the catchment area, and dividing it by the original area of the areal unit. This will yield a decimal number between 0 and 1, inclusive, which we will refer to as the “area proportion decimal”
7. For each areal unit, take the population associated with the areal unit, and multiply it by the area proportion decimal. This will yield the population of that unit that resides within the catchment area. This number is rounded to the nearest whole person.

8. Finally, calculate the sum of the population from all areal units that reside within the catchment area. This will yield an estimate of the total population in the entire study area that resides within the defined proximity of the target features.

The end result of this development was a fully functional Python-based script for ArcGIS that will perform a full areal interpolation analysis and create feature classes that represent the catchment area cartographically. Using this automated areal interpolation tool, a population estimate was calculated using the Census blocks as they are, with no areas removed. Then, to implement filtered area weighting, the ancillary data was introduced to the analysis, and areas defined as being uninhabitable or having zero population were erased from the Census blocks. These areal features included water area and land use features from the TomTom Multinet and Points of Interest datasets (see **Data and Software Overview** section). **Figure 4** shows a map of the original DeKalb County Census blocks with the ancillary

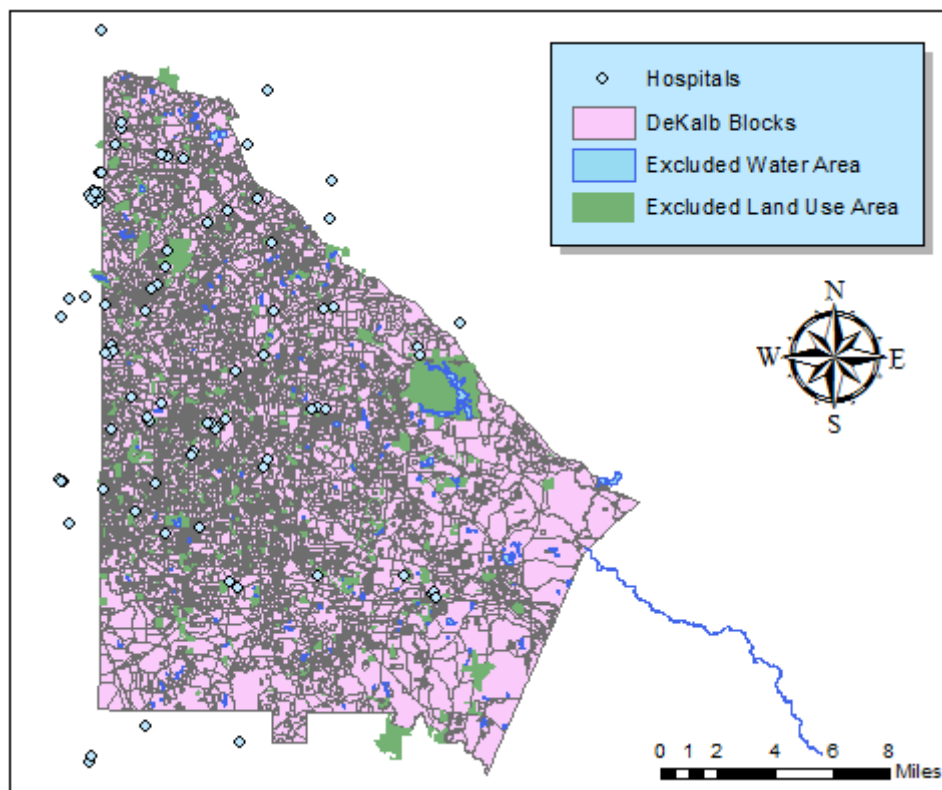


Figure 4. DeKalb County blocks and hospital/medical centers with exclusion area features overlaid on top

data mapped over it. These zero population area features will be known as the “exclusion set”. This exclusion set was then filtered from the original DeKalb County blocks using the Erase task in ArcGIS.

Figure 5 shows the resulting areal units. Using the new areal units with the defined exclusion set erased, the areal interpolation tool was run again using the new block geometries. The resulting calculations for the total number of people that reside within our specified catchment areas now reflect the removal of the exclusion set. **Figure 6** shows the new area that represents where people are expected to reside within the defined study area.

In order to determine the accuracy of the population estimates obtained from the analysis using regular areal interpolation and filtered area weighting, a benchmark population estimate was also calculated. This was done using areal interpolation and the LandScan USA 2011 dataset, and using the

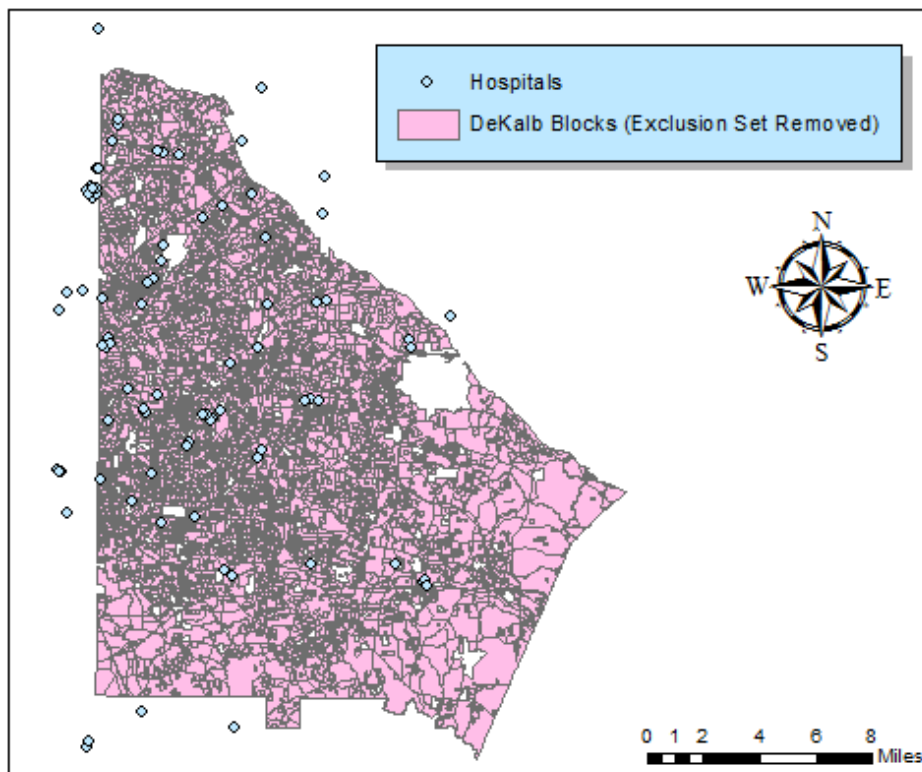


Figure 5. DeKalb County blocks with defined exclusion areas erased. The resulting block features represent only the areas in which people are assumed to reside

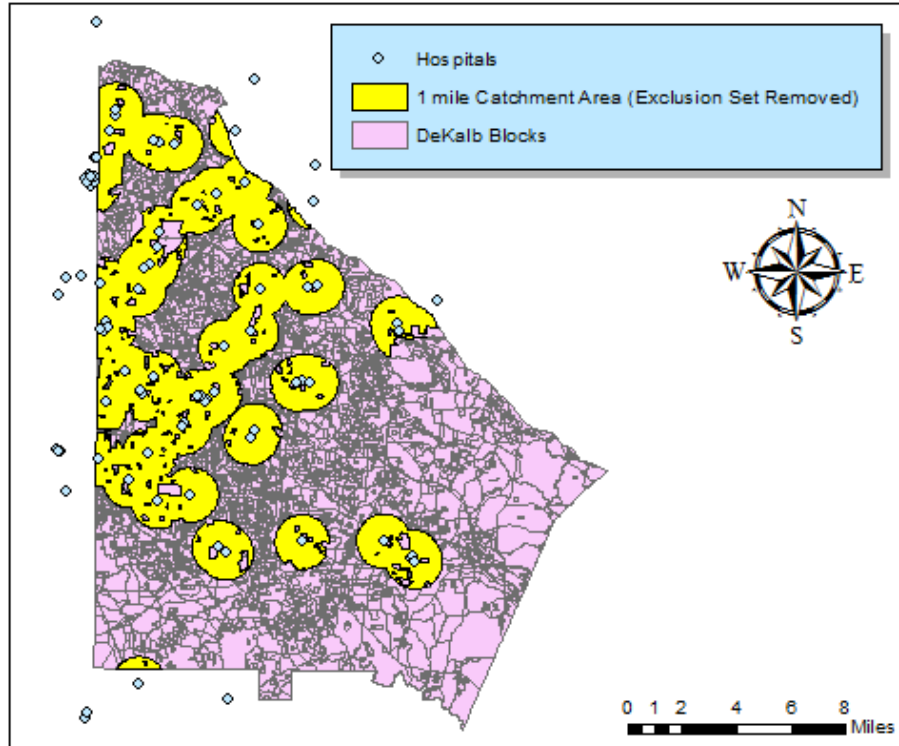


Figure 6. The catchment area defined as one mile from a hospital or medical center, with the exclusion areas removed

Zonal Statistics tool present in ArcGIS Spatial Analyst toolbox. LandScan USA was used as the benchmark for population estimation because it implements multiple methods of spatial interpolation to calculate population estimates, and also adequately takes into account many ancillary datasets to filter uninhabitable areas. Because of the extensive algorithm used to generate the population surface, this study will assume that LandScan USA is the best estimate of population. A detailed explanation of how LandScan USA calculates its population estimates follows in the **Data and Software Overview** section of this thesis. The population estimates obtained from regular areal interpolation and filtered area weighting were compared to the benchmark estimates calculated using the LandScan USA data to determine if improvement or regression from the benchmark occurred. Using these differences, an “improvement” metric was calculated to determine if filtered area weighting (denoted FAW) yielded an estimate closer to the benchmark than areal interpolation (denoted AI). The improvement number was calculated using the following formula...

$$\text{Improvement} = (\text{Estimate}_{AI} - \text{Estimate}_{LandScan}) - (\text{Estimate}_{FAW} - \text{Estimate}_{LandScan})$$

Using this improvement formula, a positive improvement value indicates that filtered area weighting yielded an estimate closer to the LandScan benchmark, and a negative improvement value indicates that filtered area weighting yielded an estimate farther away from the LandScan benchmark. An improvement value of zero indicates that areal interpolation and filtered area weighting returned the same population estimate.

The methodology up to this point focused specifically on the hypothetical case study of population estimation near hospitals and medical centers. In order to determine if filtered area weighting provides overall improvement over regular areal interpolation, two simulations in which population estimates were generated for randomly generated catchment areas were also performed. For the first simulation, in each iteration of the simulation, a random sampling of points in DeKalb County, GA equal to the number of hospitals and medical centers defined in the original case study ($n = 97$) was generated. The same catchment area definitions were used (0.5, 1, and 2 miles), and a population estimate for each catchment area was calculated using both areal interpolation and areal interpolation with filtered area weighting. In addition, a different benchmark population estimate was calculated using the randomly points and the LandScan USA dataset, and the difference between the LandScan benchmark and the calculated values using areal interpolation and filtered area weighting were determined. The total simulation consisted of 1,000 iterations, and was conducted using Python automation in ArcGIS. At the end of the simulation, the mean population estimates using regular areal interpolation and filtered area weighting were calculated. In addition, the mean benchmark using the LandScan data across all 1,000 iterations was calculated. For each of these mean population estimates (regular areal interpolation, filtered area weighting), the standard deviation of the difference from the benchmark using each method was also calculated. The results of this simulation indicate whether or

not across a large number of iterations filtered area weighting provides an improvement on average over regular areal interpolation.

After performing the first simulation, a second simulation was performed using the same parameters as the first with the exception of the generation of the random sample of points of interest. Unlike in the first simulation, the second simulation utilized a stratified random sampling technique, in which the points of interest were placed only in blocks that were in the top 25% of blocks in terms of population density. This was done to observe if filtered area weighting would yield more consistent improvement patterns based on population density. This also provided us with a more true-to-life sampling method, as using pure random generation of points of interest could yield a disproportionate amount of points in areas with lower population. As we observe in **Figure 7**, hospitals and medical centers are, in reality, clustered more heavily around more densely populated areas.

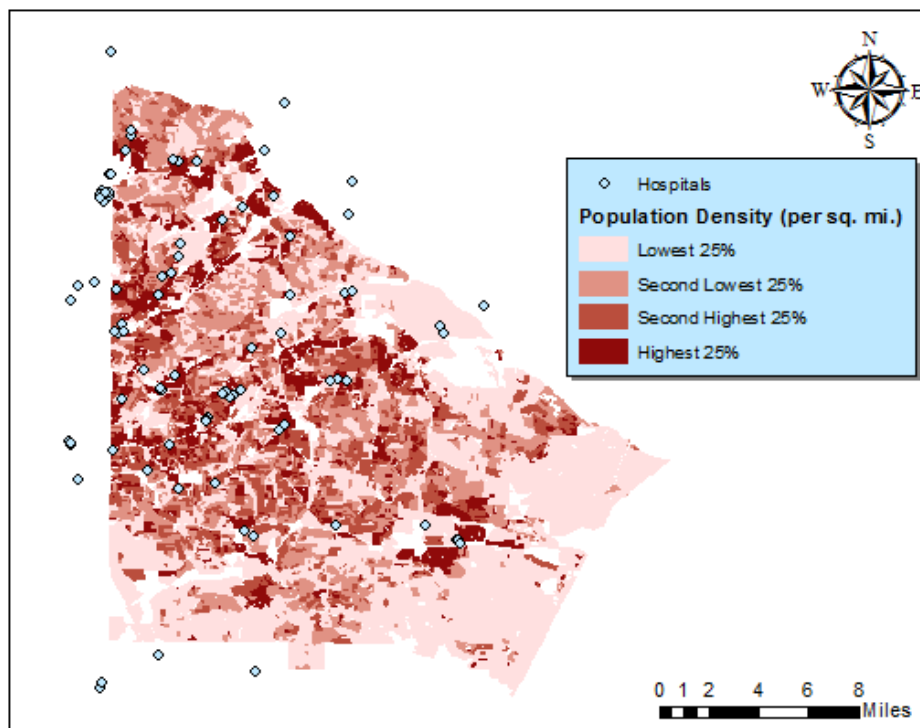


Figure 7: Population Density of DeKalb County Census blocks, classified by quartiles

4. DATA AND SOFTWARE OVERVIEW

For this experiment, the ArcGIS suite of GIS products, created by the Environmental Science Research Institute (ESRI), was used to create the automated tool that performs the areal interpolation analysis given specific input data, and to perform all geoprocessing operations on GIS data. ArcGIS Version 10.0 with Service Pack 4 was used to create the automated areal interpolation tool used in this analysis. The automated script to run the areal interpolation analysis and the simulations described in this thesis was created using Python 2.6 and ArcGIS's arcpy geoprocessing library. This geoprocessing library provides functionality to automate large-scale geoprocessing tasks, allowing users to execute long scripts and data processing with minimal user input and effort.

All of the data that was used for this analysis was provided through and used with permission from the Geospatial Research, Analysis, and Services Program (GRASP) at the CDC. The 2010 Census block geometries and attributes were provided through the United States Census Bureau, and use the block geometries and population values from the 2010 Census. For this research, the attributes for the Census blocks were joined with the geometries based on a field defined by the Census Bureau called "GEOID", which is an aggregation of a state FIPS code, a county FIPS code, a Census tract ID, and a Census block ID. The resulting Census block feature class contains an attribute called TOTALPOP, which represents the total people that reside in that Census block. This TOTALPOP field was calculated by adding together the total population field (P0010001) from each Census block in DeKalb County. There are a total of **7,591** Census blocks in DeKalb County, with a total population of 691,893.

The ancillary data that was introduced to implement filtered area weighting includes certain land use and water area polygons which we assume have zero population or are uninhabitable. The land use polygons and points of interest datasets were provided by TomTom, an Amsterdam-based creator and supplier of map data, traffic analysis program, and leading supplier of navigation and location

devices for automobiles in over forty countries, including the United States. The land use area and water area feature classes were obtained from TomTom's 2012 Multinet dataset, and the points of interest feature class used to get the hospital and medical center data for this research was obtained from TomTom's 2012 Local Points of Interest (LPOI) dataset. The Multinet and LPOI datasets are professional datasets provided by TomTom designed for navigation and feature locating in GIS applications. The land use features used in this study were obtained from the MN_LU feature class in the Multinet dataset, and include polygons that represent schools, major shopping centers, cemeteries, research institutions, airports and airport runways, stadiums and event venues, hospitals, medical centers, golf courses and country clubs, amusement parks, industrial and company property, and parks. For this experiment, all of the land use features provided in this feature class was used except colleges, universities, and islands. These features were not used in the research because there are areas that fall within the boundaries of colleges, universities, and islands where people still reside on a permanent basis. In addition, smaller commercial areas were not included in this study because there was not a reliable collection of commercial retail areas that included every available commercial area accessible. This limitation may affect the accuracy of our results, but the inclusion of major public and commercial land use areas should still provide noticeable and meaningful results. By using definition queries, these three feature types were able to be excluded from the land use feature class for the purposes of this research. After excluding these three feature types, there were a total of **339** land use area features in DeKalb County. The water area features were obtained from the MN_WA feature class in the same Multinet dataset, and a total of **141** water area features are present in DeKalb County.

The hospitals and general medical facility points of interest are obtained from the LPOI_PI feature class in the LPOI dataset, and are defined as those with a subcategory code (SUBCAT) of 7321002. According to TomTom, this subcategory includes major hospitals, trauma centers, urgent care clinics, and family care centers. In this research, I decided to include hospitals and medical center

features that are not only in DeKalb County, but also within two miles of DeKalb County. This was done because consideration should still be given to hospitals and medical centers that are outside of DeKalb County, but to which a resident in DeKalb County could still be within close proximity of. This way, DeKalb County residents that are still within range of a hospital or medical center, no matter what county it is in, are still considered in range. However, it is important to note that including features of interest outside of the study area is not always possible. Due to differing state health care laws, a study area that is along the state border may not include hospitals outside the state border in the analysis. However, for this particular study, discretion was used in including hospitals and medical centers just outside of DeKalb County. Including these features that are in or within two miles of DeKalb County, there are **97** total hospitals and medical center points in this study.

As a benchmark to judge the accuracy of the calculations of the areal interpolation and filtered area weighting methods against, the 2011 LandScan USA population density and distribution raster dataset was used as a benchmark population estimate. The 2011 LandScan raster dataset is created and maintained by the Oak Ridge National Laboratory. The LandScan dataset is at approximately a 1 km resolution, and represents ambient population in an area, averaged over a 24-hour period. **Figure 8** shows the LandScan USA dataset visualization for DeKalb County, which is the subject of the case study for this thesis. The population values for each 1 km² area is an aggregation of various data, including population density, population distribution, satellite imagery classification, and land use data, compiled using a combination of data and imagery analysis techniques and multivariate dasymetric modeling. The data is provided in the WGS 1984 geographic coordinate system, which is what the analysis was performed in. This dataset was used as a benchmark for accuracy of our methods because of its comprehensive population calculation algorithm that includes multiple methods of spatial interpolation and multiple ancillary datasets and area classification methods are used to generate this population

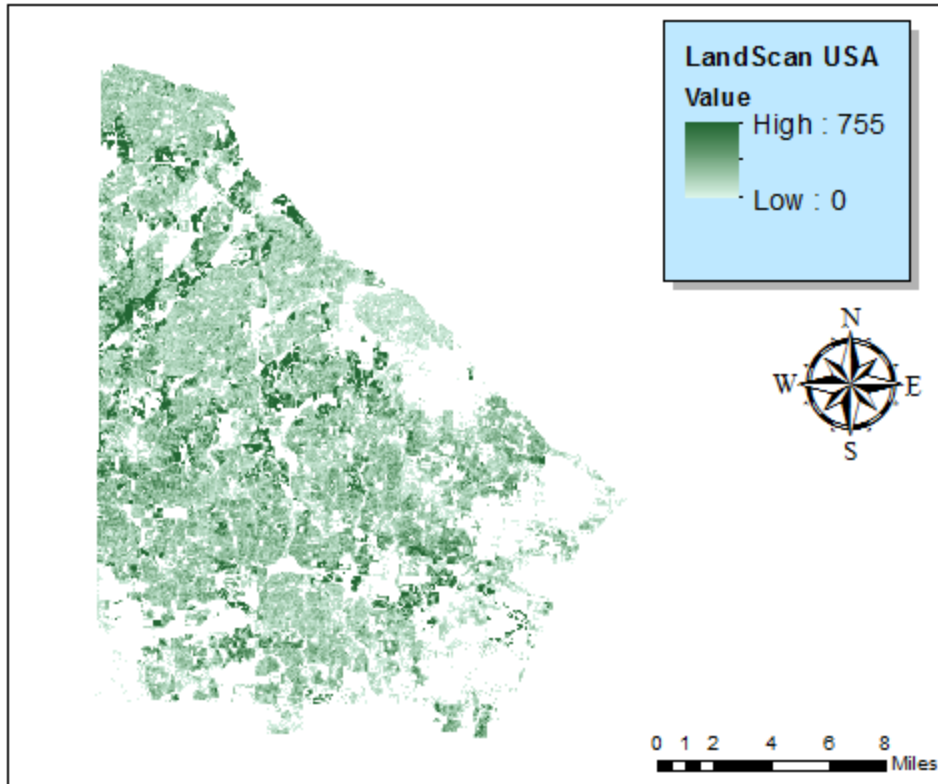


Figure 8: LandScan USA dataset for DeKalb County, GA

estimate, which accounts for all of the major population distribution factors that we are interested in (Bhaduri, et al 2007). This dataset represents the most precise population estimate dataset that is available for this analysis. However, despite the complexity of this dataset, certain degrees of error may still exist in the results obtained using this dataset due to the nature of LandScan USA still being simply an estimate of population distribution.

5. RESULTS

The results from the areal interpolation and filtered area weight analysis yielded mixed results in terms of improvement towards the defined benchmark of estimation, but the results of the simulations confirm that filtered area weighting does on average yield improved accuracy over regular areal interpolation. **Table 1** shows the differences between the areal interpolation estimate, the filtered

Table 1: Improvement towards LandScan Benchmark provided by Filtered Area Weighting

Distance	Method	Population Estimate Calculated	Population Estimated using LandScan USA	Difference between Interpolation and LandScan	Improvement Towards LandScan
0.5 miles	AI	122,752	122,873	-121	47
	FAW	122,799		-74	
1 mile	AI	314,111	314,351	-240	-326
	FAW	313,785		-566	
2 miles	AI	545,127	544,991	136	-93
	FAW	545,220		229	

area weighting estimate, and the LandScan benchmark population estimate number for each catchment area distance. For each catchment area distance, the “improvement” number is also indicated. The results show that in the cases of the 0.5 mile catchment area, filtered area weighting yielded a population estimate closer to the LandScan benchmark estimate. In the cases of the 1 and 2 mile catchment areas, filtered area weighting yielded a population estimate further away from the LandScan benchmark estimate.

Table 2 shows the mean calculated population estimate values and the mean benchmark estimate values for the first simulation in which a non-stratified random sampling of hospital and medical center points was generated. The results in the table show the mean population estimates calculated across all 1,000 iterations of the simulation using areal interpolation, filtered area

Table 2: Mean benchmarks and estimates calculated in Non-Stratified Random Simulation

Distance	Method	Mean Estimate Calculated Using Interpolation	Mean Value Calculated Using LandScan	Mean Diff. btn Interpolation and LandScan	Standard Deviation of Difference	Improvement
0.5 miles	AI	115,014	114,900	114	1496	32
	FAW	114,982		82	1500	
1 miles	AI	358,000	357,766	234	1608	124
	FAW	357,876		110	1607	
2 miles	AI	655,136	654,603	533	731	187
	FAW	654,949		346	733	

weighting, and the LandScan benchmark. The results show that when using filtered area weighting over regular areal interpolation, the population estimates on average are closer to the LandScan benchmark estimates. This is the case for all three of the catchment areas defined. When looking at the standard deviation of the differences between the LandScan benchmark and the estimate generated through interpolation across all of the iterations of the simulation, we see that although we do get improvement towards the benchmark on average using filtered area weighting, we are still likely to see mixed results from trial to trial. This is reinforced by **Figure 9**, which graphs the improvement realized with filtered area weighting over regular areal interpolation specifically for the half-mile catchment area. **Figure 9** shows that among individual iterations of the simulation, both improvement and regression from the LandScan benchmark were observed roughly the same amount of times. However, looking at the standard deviation of the differences, the value decreases greatly when looking at the two-mile catchment area. This seems to indicate that with larger catchment areas, the potential variance of the

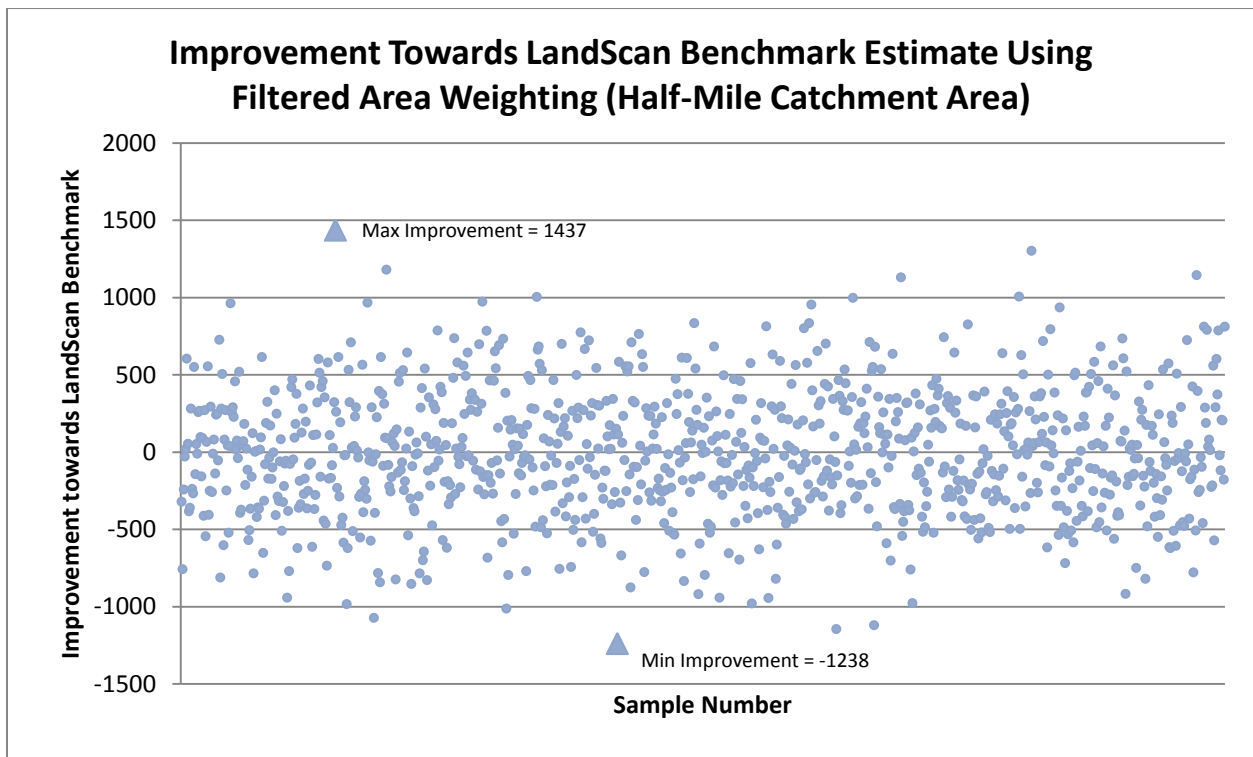


Figure 9: Improvement towards LandScan for Half-Mile Catchment Area

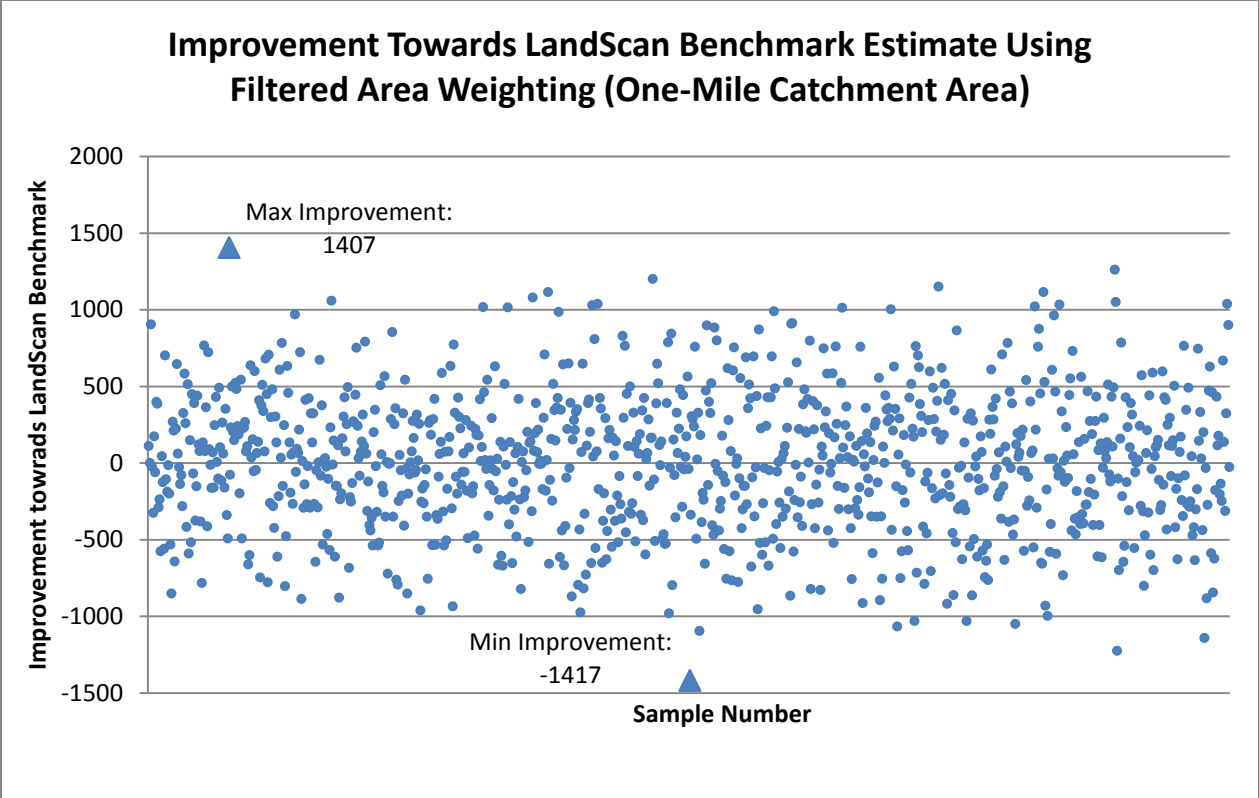


Figure 10: Improvement towards LandScan for One-Mile Catchment Area

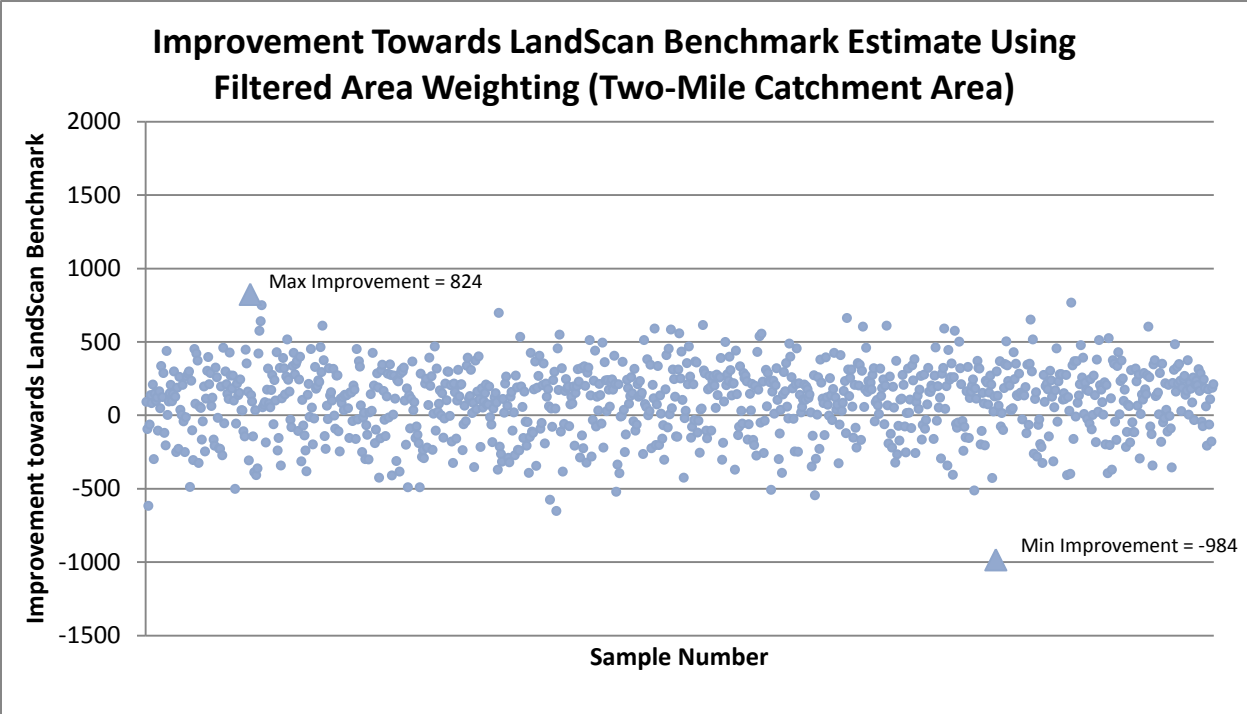


Figure 11: Improvement towards LandScan in for Two-Mile Catchment Area

differences between the interpolated estimation values and the LandScan values are smaller. **Figure 10** shows the improvement metrics for the same simulation for the one-mile catchment area, and **Figure 11** shows the improvement metrics for the two-mile catchment area. Compared to the plot for the half-mile catchment area and the one-mile catchment area, the plot for the two-mile catchment area is more clustered, and the minimum and maximum observed improvement values are smaller in magnitude. In addition, there appear to be fewer trials that yield negative improvement with the two-mile catchment area. This shows us that with higher catchment areas, we can expect to see improvement with filtered area weighting more consistently than with smaller catchment areas.

The second, stratified simulation showed very similar patterns to the first non-stratified simulation. The estimates using filtered area weighting showed improvement towards the LandScan benchmark estimates for all three catchment areas, and as the catchment area increases in size, the standard deviation of the difference between the interpolated estimates and the LandScan estimates decreased. This indicates that the locations of the features of interest did not make a difference in the pattern of overall improvement using filtered area weighting. In addition, the results still support the observation from the first simulation that with larger catchment areas, this improvement is more evident and more consistent. However, more catchment area sizes are warranted in order to evaluate the improvement.

Table 3: Mean benchmarks and estimates calculated in Stratified Random Simulation

Distance	Method	Mean Estimate Calculated Using Interpolation	Mean Value Calculated Using LandScan	Mean Diff. btn Interpolation and LandScan	Standard Deviation of Difference	Improvement
0.5 miles	AI	226,759	227,247	488	1420	43
	FAW	226,802		445	1435	
1 miles	AI	454,076	453,552	524	1087	12
	FAW	454,064		512	1095	
2 miles	AI	638,683	637,981	702	848	52
	FAW	638,631		650	818	

6. DISCUSSION

The main question that emerges from these results is why a higher degree of accuracy is observed with filtered area weighting in larger catchment areas versus smaller catchment areas. When considering prior literature on areal interpolation and filtered area weighting, the most likely explanation for this correlation is that larger catchment areas incorporate more ancillary data that is introduced through filtered area weighting. This is seen in work involving population estimation using dasymetric methods such as areal interpolation (Briggs 2007) and combined dasymetric/pycnophylactic approaches to interpolation (Comber 2008; Kim 2010). For example, Briggs, et al, use dasymetric mapping methods to create a small-area population distribution model using land use data similar to what is used in this case study. However, the model was shown to be more accurate when additional ancillary data in the form of nighttime light emissions data was introduced to the catchment area of the study (Briggs 2007). For this particular study, larger catchment areas incorporate more of the ancillary data used to define uninhabitable areas. Across all 1,000 iterations of the simulations performed in this study, the pattern of overall improvement was still observed. This reinforces the notion that more ancillary data is more likely to provide improved accuracy because the population distribution is better defined in a catchment area that includes more of that ancillary data.

The notion of Bayesian statistical theory can also be used to support this notion that more ancillary data makes filtered area weighting more effective (Farrell 1997). Parallels between spatial interpolation and multivariate and covariate statistical models have been made in past research, and although they are different in the sense that spatial population modeling incorporates location and proximity into calculating estimates and correlation, they both still rely on existing data and known values to make estimates of data given unknown parameters (Farrell 1997; Wakefield 1999). To illustrate this connection, one can visualize that the probability of correctly locating a person anywhere in Earth within a few feet is lower when only high-level information about their location is known.

Similar to the laws of conditional probability, when more information on that person's possible location is known, such as state, county, or city, the probability of correctly locating that person increases (Shafer 1985). Areal units that are larger in size, such as counties, states, or countries, can locate certain people as being within those larger areal units. However, the precision and accuracy with which one can locate the point where an individual person actually lives using those more coarse areal units is lower than if one used more detailed areal units, such as Census tracts or Census blocks. The process of introducing ancillary data to coarse areal units and estimating the population of a more finely defined area is similar to improving the probability of correctly locating a person geographically by limiting the potential areas in which they can be located. When this process is repeated multiple times using various data, a greater level of detail in the population estimation model can be achieved (Mugglin, et al 1999). However, when that ancillary dataset does not introduce a greater level of detail regarding the population distribution, the likelihood of a population model locating a particular person does not increase. Farrell and Mugglin's previous work supports this notion, because there can be no refinement of coarsely defined areal units without geographically relevant ancillary data. When examined this way, the effective use of filtered area weighting becomes contingent on the ancillary datasets in comparison to the original dataset.

Despite the use of the ancillary land use data in this experiment, the results of this study shows that filtered area weighting can still yield population estimates further away from the benchmark estimate than regular areal interpolation. This mixture of improvement towards and regression from the LandScan benchmark estimate seems to conflict with the assertion that ancillary data provides more accurate population distribution data. This inconsistency in the improvement metrics can be tied to the quality of the ancillary data itself, not just its inclusion in the interpolation process. For instance, the introduction of certain land use features to filter certain Census block areas as uninhabited may not yield any difference in the population estimate because those filtered Census blocks may already be defined as having zero population. In this case, the inclusion of zero population Census blocks in areal

interpolation is inconsequential to the estimate produced. However, significant differences in the population estimates calculated can be seen in cases where areas that are uninhabited are not filtered by the ancillary data, or areas that are inhabited by people are filtered incorrectly. While it is impossible to catch every instance of error that may occur when performing filtered area weighting, the amount of error present can still be mitigated through the use of multiple ancillary datasets in conjunction with one another (Larsen 2003). In this particular study, only one dataset was used: the TomTom Multinet land use and water area polygons. The introduction of additional datasets, such as additional land use and water area polygons from other data providers or a continuous land cover raster dataset, may have provided more consistent improvement towards the LandScan benchmark estimate because more uninhabited or uninhabitable areas could have been filtered from the original areal units.

Considering this fact, the use of Census blocks as the basic areal unit for this study can also be brought into question. If a coarser areal unit, such as Census tracts, was used in the study as opposed to Census blocks, the population estimates yielded could represent a greater improvement towards a benchmark estimate if an accurate ancillary dataset is used to mask out the unpopulated areas. However, the quality of the ancillary data used in filtered area weighting is also relevant. Although the choice of areal population unit for a study is certainly not trivial when considering the computational resources required to analyze a large number of areal units and the desired scope of the study (Boscoe 2003), the actual observed improvement with filtered area weighting would appear to be more directly linked to the quality and accuracy of the ancillary data. This is likely because no matter the areal unit used, the same amount of overall area is being used to model the population distribution of the same amount of people. In addition, accessibility to population data at the desired areal units can inhibit the ability to obtain an optimally accurate population estimate. The accessibility of such detailed datasets such as LandScan USA may be restricted based on the resources available to the group or individual performing the study. Although LandScan USA could have easily been used as the primary population

estimate, not all groups performing areal interpolation may have access to such a dataset, and thus must derive population models through other datasets (Bhaduri, et al 2007). Evaluating the degree of estimate accuracy improvement when different levels of areal units are used and datasets from different vendors are used would be a logical next step for research regarding spatial interpolation.

When all of these observations from the results are taken together, the instances in which the use of filtered area weighting can be used should be put under scrutiny. Through repetition and analysis, this study reinforces the assertion that the introduction of ancillary data through filtered area weighting can yield an improved population estimate, but does not always do so depending on the catchment areas and features of interest. Despite this observation and the relatively sparse amount of ancillary data used in this experiment, an overall improvement towards our accuracy benchmark was still observed. Considering these seemingly conflicting observations, further analysis of filtered area weighting in a public health context is warranted.

These points of discussion highlight some key limitations with the case study used for this thesis. The use of only three catchment area definitions (one mile, half mile, and two mile) provides only a limited scope of the trends in population estimation improvement. A future case study can examine other catchment area definitions, such as one-and-a-half (1.5) miles. Other catchment area definitions other than pre-determined distances can also be used for future research. The catchment area may be determined through a dispersion model or a geometric network analysis. In either case, a better population estimation method can translate into a more accurate depiction of health trends, especially in an age where studying a health indicator's effect includes knowing and understanding its effect on populations at large (Baum 2003). Another limitation of this thesis was the use of only one ancillary dataset (in this case, the TomTom Multinet dataset). Prior research highlights the ability to obtain improved population estimates through the use of multiple ancillary datasets as opposed to just one

(Comber, et al 2008; Briggs 2007). Implementing additional datasets with filtered area weighting should yield improved population estimates as well. Additional datasets themselves should also be scrutinized for their quality. If the data is not sufficiently detailed, the improvement in a population estimate becomes limited. The use of the LandScan USA dataset as the benchmark estimate for the study was also a key limitation of this study. Despite the intensive algorithm and comprehensive data sources used to derive the LandScan USA dataset (Bhaduri, et al 2007), this dataset remains a population estimate, and therefore should also be subject to scrutiny. The analysis will also need to be tested in other regions to examine whether the findings hold true or not.

7. CONCLUSION

This thesis examined the performance of filtered area weighting and areal interpolation for use in public health population estimation. The results suggested that on average, the use of filtered area weighting is able to improve population estimation in various catchment areas than areal interpolation alone. However, the improvement from using filtered area weighting is dependent on the ancillary datasets used. Depending on the geography and features of interest that are used in a study, the performance of filtered area weighting depends on whether or not a significant amount of unpopulated areas can be masked off using ancillary data.

Although the case study in this paper was limited to DeKalb County, Georgia, a metropolitan Atlanta county with a relatively urban and suburban population, this analysis of filtered area weighting and spatial interpolation in general needs to be expanded to include more areas of varying population densities and land use presence to test its performance. By examining more areas with varying characteristics, patterns of performance of filtered area weighting can be identified. There are a myriad of different situations and studies that filtered area weighting can be examined with, but with all of them, we must always consider the data that is available to us, and how it can be used most effectively.

Although we had the LandScan 2011 dataset available to us, this data will not always be available for all public health studies. More analysis needs to be done using other supplementary data sets to test whether filtered area weighting can provide more accurate estimates.

Public health analysis is becoming a more important part of the development of modern society. The findings from public health studies on particular diseases or phenomena have helped shape social policy and direct medical research. Public health studies of various types use proximity analysis to determine at-risk populations and accessible populations (Bedimo-Rung, et al 2005; Rosero-Bixby 2004; Hendryx, et al 2002; Norman, et al 2006). As GIS and spatial analysis become a more integral part of public health studies, the importance of knowing and understanding the effects, uses, and limitations of spatial interpolation as they apply to estimating populations becomes critical to identifying patterns in public health. This thesis evaluates the performance of filtered area weighting for use in population estimates for public health. Further exploration of filtered area weighting and spatial interpolation can continue to establish the degree to which it can further refine population estimation as a practice in general.

REFERENCES

- Aylin, P., Maheswaran, R., Wakefield, J., Cockings, S., Jarup, L., Arnold, R., Wheeler, G., & Elliott P. (1999). A national facility for small area disease mapping and rapid initial assessment of apparent disease cluster around a point source: the UK Small Area Health Statistics Unit. *Journal of Public Health, 21*(3), 289-298.
- Baum, F. (2003). *The New Public Health*. Oxford University Press.
- Bedimo-Rung, A., Mowen, A., & Cohen, D. (2005). The significance of parks to physical activity and public health: A conceptual model. *American Journal of Preventive Medicine, 28*(2), 159-168.
- Besleme, K., & Mullin, M. (2007). Community Indicators and Healthy Communities. *National Civic Review, 86*(1), 43-52.
- Bhaduri, B., Bright, E., Coleman, P., & Urban, M. (2007). LandScan USA: a high-resolution geospatial and temporal modeling approach for population distribution and dynamics. *GeoJournal, 69*(1-2), 103-117.
- Boscoe, F., & Pickle, L. (2003). Choosing Geographic Units for Choropleth Rate Maps, with an Emphasis on Public Health Applications. *Cartography and Geographic Information Science, 30*(3), 237-248.
- Briggs, D., Gulliver, J., Fecht, D., & Vienneau, D. (2007). Dasymetric modeling of small-area population distribution using land cover and light emissions data. *Remote Sensing of Environment, 108*(4), 451-466.
- Chakraborty, J., Maantay, J., & Brender, J. (2011). Disproportionate Proximity to Environmental Health Hazards: Methods, Models, and Measurement. *American Journal of Public Health, 101*(S1), S27-S36.

- Chakraborty, J., & Maantay, J. (2011). Proximity Analysis for Exposure Assessment in Environmental Health Justice Research. *Geospatial Analysis of Environmental Health*, 4(1), 111-138.
- Comber, A., Proctor, C., & Anthony, S. (2008). The Creation of a National Agricultural Land Use Dataset: Combining Pycnophylactic Interpolation with Dasymetric Mapping Techniques. *Transactions in GIS*, 12(6), 775-791.
- Cromley, E., & McLafferty, S. (2012). *GIS and Public Health: 2nd edition*. New York, NY: The Guilford Press.
- Dent, A., Fowler, D., Kaplan, B., Zarus, G., & Henriques, W. (2000). Using GIS to Study the Health Impacts of Air Emissions. *Drug and Chemical Toxicology*, 23(1), 161-178.
- Eicher, C., & Brewer, C. (2001). Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. *Cartography and Geographic Information Science*, 28(2), 125-138.
- Farrell, P., MacGibbon, B., & Tomberlin, T. (1997). Empirical Bayes Small-Area Estimation Using Logistic Regression Models and Summary Statistics. *Journal of Business & Economic Statistics*. 15(1), 101-108.
- Fisher, P., & Langford, M. (1995). Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning*. 27(2), 211-224.
- Fisher, P., & Langford, M. (1996). Modelling Sensitivity to Accuracy in Classified Imagery: A Study of Areal Interpolation by Dasymetric Mapping. *The Professional Geographer*. 48(3), 299-309.
- Goodchild, M., & Lam, N. (1980). Areal Interpolation: A Variant of the Traditional Spatial Problem. *Geo-Processing*. 1(1980), 297-312.
- Goodchild, M., Anselin, L., & Deichmann, U. (1992). A framework for the areal interpolation of socioeconomic data. *Environment and Planning*. 25, 383-397.

- Hawley, K., & Moellering, K. (2005). A Comparative Analysis of Areal Interpolation Methods. *Cartography and Geographic Information Science*, 32(4), 411-423.
- Hay, S., Noor, A., Nelson, A., & Tatem, J. (2005). The accuracy of human population maps for public health application. *Tropical Medicine & International Health*, 10(10), 1073-1086.
- Hendryx, M., Ahern, M., Lovrich, N., & McCurdy, A. (2002). Access to Health Care and Community Social Capital. *Health Services Research*, 37(1), 85-101.
- Holt, J., Lo, C., & Hodler, T. (2004). Dasymetric Estimation of Population Density and Areal Interpolation of Census Data. *Cartography and Geographic Information Science*, 31(2), 103-121.
- Kim, H., & Yao, X. (2010). Pycnophylactic interpolation revisited: integration with the dasymetric-mapping method. *International Journal of Remote Sensing*, 31(21), 5657-5671.
- LandScan Home. Retrieved 21 February, 2013, from <http://www.ornl.gov/sci/landscan/>.
- Langford, M., Maguire, D. J., & Unwin, D. J. (1991). The areal interpolation problem: estimating population using remote sensing in a GIS framework. *Handling geographical information: Methodology and potential applications*, 55-77.
- Langford, M., Higgs, G., Radcliffe, J., & White, S. (2008). Urban population distribution models and service accessibility estimation. *Computers, Environment and Urban Systems*, 32(1), 66-80.
- Larsen, M. (2003). Estimation of small-area proportions using covariates and survey data. *Journal of Statistical Planning and Inference*, 112(1-2), 89-98.
- Lu, C., Fenske, R., Simcox, N., & Kalman, D. (2000). Pesticide Exposure of Children in an Agricultural Community: Evidence of Household Proximity to Farmland and Take Home Exposure Pathways. *Environmental Research*, 84(3), 290-302.

- Maantay, J., Maroko, A., & Herrmann, C. (2007). Mapping Population Distribution in the Urban Environment: The Cadastral-based Expert Dasymetric System (CEDS). *Cartography and Geographic Information Science*, 34(2), 77-102.
- Maroko, A. (2009). Using air dispersion modeling and proximity analysis to assess chronic exposure to fine particulate matter and environmental justice in New York City. *Applied Geography*, 34(1), 533-547.
- McGlynn, E., Asch, S., Adams, J., Kessey, J., Hicks, J., DeCristofaro, A., & Kerr, E. (2003). The Quality of Health Care Delivered to Adults in the United States. *The New England Journal of Medicine*, 348(1), 2635-2645.
- Mennis, J. (2003). Using Geographic Information Systems to Create and Analyze Statistical Surfaces of Population and Risk for Environmental Justice Analysis. *Social Science Quarterly*, 83(1), 281-297.
- Mennis, J. (2008). Generating Surface Models of Population Using Dasymetric Mapping. *The Professional Geographer*, 55(1), 31-42.
- Mennis, J. (2009). Dasymetric Mapping for Estimating Population in Small Areas. *Geography Compass*, 3(2), 727-745.
- Mennis, J., & Hultgren, T. (2006). Intelligent Dasymetric Mapping and Its Application to Areal Interpolation. *Cartography and Geographic Information Science*. 33(3), 179-194.
- Mittelmark, M. (2001). Promoting Social Responsibility for Health: Health impact assessment and healthy public policy at the community level. *Health Promotion International*, 16(3), 269-274.

- Mugglin, A., Carlin, B., Zhu, L., & Conlo, E. (1999). Bayesian areal interpolation, estimation, and smoothing: an inferential approach for geographic information systems. *Environment and Planning*. 31(8), 1337-1352.
- Norman, G., Nutter, S., Ryan, S., Sallis, J., Calfas, K., & Patrick, K. (2006). Community Design and Access to Recreational Facilities as Correlates of Adolescent Physical Activity and Body-Mass Index. *Journal of Physical Activity and Health*, 3(1), S118-128.
- Reibel, M., & Agrawal, A. (2007). Areal Interpolation of Population Counts Using Pre-Classified Land Cover Data. *Population Research and Policy Review*. 26(5-6), 619-633.
- Reynolds, P., et al (2004). Residential proximity to agricultural pesticide use and incidence of breast cancer in the California Teachers Study cohort. *Environmental Research*. 96(2), 206-218.
- Rosero-Bixby, L. (2004). Spatial access to health care in Costa Rica and its equity: a GIS-based study. *Social Science & Medicine*, 58(1), 1271-1284.
- Sadahiro, Y. (1999). Accuracy of areal interpolation: A comparison of alternative methods. *Journal of Geographic Systems*. 1(4), 323-346.
- Shafer, G. (1985). Conditional Probability. *International Statistical Review*. 53(3), 261-275.
- Tobler, W. (1979). Smooth Pycnophylactic Interpolation for Geographical Regions. *Journal of the American Statistical Association*. 74(367), 519-530.
- Wakefield, J., & Elliott, P. (1999). Issues in the Statistical Analysis of Small Area Health Data. *Statistics in Medicine*. 18(17-18), 15-30.
- Wright, J. (1936). A Method of Mapping Densities of Population with Cape Cod as an Example. *Geographical Review*. 26(1), 103-110.

Wu, S., Qiu, X., & Wang, L. (2005). Population Estimation Methods in GIS and Remote Sensing: A Review. *GIScience & Remote Sensing*. 42(1), 80-96.

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