TOWARDS A MORE CONTEXTUALLY-SENSITIVE UNDERSTANDING OF

COMMUNITY COLLECTIVE EFFICACY:

AN APPLICATION OF "INTELLIGENT DASYMETRIC MAPPING"

by

TAYLOR CHRISTIAN JOHNSON

(Under the Direction of Steven R. Holloway)

ABSTRACT

Previous contextual research relies heavily on data collected by the U.S. Census Bureau due to its widespread availability, but this incorrectly assumes that populations are equally distributed within the arbitrarily-drawn Census boundaries. Areal interpolation research suggests that methods employing ancillary data assist in correcting the problem. This thesis posits that the augmentation of census variables through "Intelligent Dasymetric Mapping," and the redefinition of neighborhoods as circular or road network-based buffers around interview locations, provides more accurate contextual information for modeling neighborhood effects. This hypothesis is tested through the multivariate regression analysis used to explain individual perceptions of community collective efficacy which is theorized to mediate specific of criminal behavior. Effects of neighborhood size (i.e. buffer distance) are also explored. Results indicate improved model explanatory power compared to that of models using traditional measures of context while significance of model parameters estimates varies by scale.

INDEX WORDS: Intelligent Dasymetric Mapping, Community collective efficacy, Areal interpolation, Population, Neighborhood

TOWARDS A MORE CONTEXTUALLY-SENSITIVE UNDERSTANDING OF COMMUNITY COLLECTIVE EFFICACY: AN APPLICATION OF "INTELLIGENT DASYMETRIC MAPPING"

by

TAYLOR CHRISTIAN JOHNSON

BS, University of Georgia, 2007

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2011

TOWARDS A MORE CONTEXTUALLY-SENSITIVE UNDERSTANDING OF

COMMUNITY COLLECTIVE EFFICACY:

AN APPLICATION OF “INTELLIGENT DASYMETRIC MAPPING”

by

TAYLOR CHRISTIAN JOHNSON

Major Professor: Steven R. Holloway

Committee: Thomas R. Jordan
Lan Mu

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2011

ACKNOWLEDGEMENTS

I would like to briefly acknowledge the individuals who helped make the completion of this thesis possible. First, I must thank Dr. Steven Holloway for his guidance, patience, and helpful insights as my major professor. Next, I would like to acknowledge and thank my committee members, Dr. Thomas Jordan and Dr. Lan Mu. I am very appreciative of my committee's help and constructive criticism throughout the thesis preparation process. I would also like to thank Dr. Ron Simmons for allowing me to work with and contribute to the Family and Community Health Study project which was funded in part by the Center for Disease Control and Prevention. Finally, I am indebted to my family and friends whose constant encouragement and advice kept me motivated and focused. Specifically, I would like to acknowledge the following people for their impact during my time as a master's student: Katherine Schweitzer, Carey Burda, Seth Gustafson, Genevieve Holdridge, Peter Hossler, Amber Ignatius, Woo Jang, Ellen Kohl, Nick Kruskamp, Ryan Lash, Matthew Miller, Jake McDonald, Dustin Menhart, Matthew Mitchelson, Priscilla McCutcheon, Richard Milligan, Michelle Palma, Andrew Parker, Chris Ploetz, Shadrock Roberts, Tom Vanderhorst, Mark Weaver, Shanqi Zhang and many others.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

New Techniques for Examining the Urban "Mosaic"

> "The processes of segregation establish moral distances which make the city a mosaic of little worlds which touch but do not interpenetrate."
>
> - Robert E. Park (1915), p. 608

Almost a century has passed since Robert Park outlined a novel methodology for studying the relatively new urban phenomenon. Park, a guiding figure of the Chicago School of Sociology, left a legacy of empirically-based research in the urban context referred to as "human ecology." Much of that research relied on the perspective that the city was divided into "natural areas" which could be easily distinguished and should be measured to gain insight into social processes such as crime and vice. The notion of segregated pieces of glass in a larger urban "mosaic" is a useful analogy for describing these areas. Indeed, a single street can affect how residents perceive the urban environment just as a freeway can forever bifurcate a once unified neighborhood. Individuals from the "bad" side of town or the "wrong" side of the [railroad] tracks can be viewed with suspicion and mistrust. However, the lead separating those glass tiles is far from impenetrable. Like pollution blown downwind from the coal stack, social processes do not necessarily stop in the face of such barricades or arbitrarily-drawn administrative demarcations. This type of social segregation identified by Park is useful but often fails to address how power – and those who wield it – shapes the processes driving urban segregation, thus yielding a mosaic that is far from "natural." Social science research today still relies on information collected in such mosaic pieces (i.e. data within census units) while failing to note the effects of processes acting just beyond the subjective partitions dividing the urban milieu. Therefore, augmenting our conception of geographical, or ecological, context may provide a better way of understanding the ever-penetrating processes at work in the city.

Traditional neighborhood effects research typically relies on census data to predict individual outcomes. This type of analysis often fails to address the theoretical concerns of the modifiable areal unit problem (MAUP) when using aggregated individual-level data. A potentially more important fallacy in the utilization of census data is the assumption that the population, and its social and economic characteristics, is evenly distributed throughout the zones in which the data are reported. As urban social and demographic trends remain dynamic and urban policy implementation remains limited by insufficient information, social scientists and activists will require methods for understanding and alleviating urban inequalities. Specifically, sociologists are interested in the mechanisms and structures which affect levels of various crimes. Robert J. Sampson and fellow researchers (Sampson, Raudenbush and Earls 1997) explain the theory of community collective efficacy, defined as "social cohesion among neighbors combined with their willingness to intervene on behalf of the common good," as a mechanism that potentially mediates certain crime within specific neighborhoods. This thesis is interested in exploring the role that geographic context plays in influencing collective efficacy. Dasymetric modeling, a spatial data interpolation technique, was initially utilized for mapping population densities (Wright 1936), but has recently been used in the creation of more spatially-aware variables for geographical analysis. This thesis furthers that trend by using "Intelligent Dasymetric Mapping" (Mennis and Hultgren 2006) to generate more contextual-sensitive community-level variables from census data in order to better model collective efficacy. This research illustrates the importance of developing more meaningful contextual characteristics and better estimating their effects within a geographic framework. In particular, this thesis aims to increase the understanding of the role that geography plays in shaping collective efficacy.

Research Purpose

Drawing from population disaggregation literatures, this thesis examines the efficacy of "Intelligent Dasymetric Mapping" in generating more contextually-sensitive variables as inputs for neighborhood effects modeling. To test the hypothesis of areal interpolation offering researchers the ability to augment regression variables based on the neighborhood's geographic context, this thesis specifies the notion of community collective efficacy as an object of study worth pursuing separately from its usual criminology perspective. While the implications of this theory (as pertaining to criminogenesis) are hardly trivial, the geographic and social construction of collective efficacy deserves more direct scrutiny and will serve as an application of "Intelligent Dasymetric Mapping" for the improvement of neighborhood effects analysis. Therefore, the thesis addresses three separate, but intertwined, research questions:

1.) Can "Intelligent Dasymetric Mapping" offer an improved method (compared to more traditional measures of geographic context) of augmenting social and economic variables typically used in neighborhood effects analysis?

2.) Can the differential operationalization of neighborhoods through size and shape (e.g., Census block groups versus circular buffers versus road network buffers of varying radii) have an effect in that type of analysis? In order words, to what degree is the MAUP present in understanding collective efficacy through neighborhood effect analysis?

3.) Finally, based on those augmented Census variables, in what ways do social and economic characteristics of neighborhoods positively or negatively impact the formation of collective efficacy (i.e. community cohesion and control) after the unit of analysis is redefined around the individuals interviewed regarding their own neighborhood?

This thesis proposes that, indeed, geography matters in shaping community collective efficacy and more contextually-sensitive measures, or measures that more accurately reflect the neighboring variations due to geography, of socio-economic characteristics will improve its understanding and the understanding of other social processes acting in the local environment to influence individual-level responses.

Expected Significance

This thesis not only addresses a new methodology application for dasymetric modeling but also aids in the theoretical understanding of factors that influence community collective efficacy for African Americans residing in the United States. Furthermore, this thesis illustrates the importance of developing improved contextually-sensitive measures within a more geographical-explicit framework (i.e., context based on proximity rather than arbitrarily-defined administrative unit) by successfully presenting marked improvements to regression results. Finally, this thesis expands our understanding of the role that geography plays in shaping community collective efficacy, thereby allowing policy makers to focus their efforts in areas nearly able to facilitate their own informal controls without the need for hyper-policing tactics targeting the urban poor.

Thesis Design

In the following chapters, I will present a review of the pertinent literature in the areas of neighborhood effects, community context, collective efficacy, and areal interpolation, a detailed methodology regarding the application of "Intelligent Dasymetric Mapping" to the creation of more contextually-sensitive regression variables, a summary of the analysis and results from the proposed methodology, and general conclusions. Chapter 2 examines the history and application of areal interpolation in relation to the modeling of social processes. This literature review bridges the methodological innovation of dasymetric modeling with the theoretical considerations in understanding the role location plays in the development of community collective efficacy through neighborhood effects analysis. Chapter 3 outlines the methods and data used in this thesis. The necessary sources of data area detailed in addition to the methods behind redefining “neighborhoods” in the context of the Family and Community Health Study. Next, the creation, and evaluation, of new, more contextually-sensitive socio-economic variables through dasymetric modeling is described. Finally, that chapter concludes with the methodology for incorporating those augmented variables into multivariate regression models. Chapter 4 presents and summarizes the thesis results including mapped variables, regression model parameter estimates, interpolation comparisons, and the final specified models for explaining community collective efficacy. Finally, Chapter 5 summarizes the thesis, relating the results presented in Chapter 4 to the research questions, acknowledges methodological limitations, and suggests directions for future study beyond the scope of this thesis.

CHAPTER 2

LITERATURE REVIEW

Evaluating Neighborhood Effects and Community Context

Blalock (1984) describes contextual effects as "the allowance for macro processes that are presumed to have an impact on the individual actor over and above the effects of any individual-level variables that may be operating." Contextual effects does not suggest that individual-level variables are ineffectual, but rather, that processes beyond the scope of the individual can have a real and measurable effect upon that person. There are numerous examples in the recent literature of contextual effects analysis in many disciplines, including public health and epidemiology (Chaix et al. 2009, Cummins 2007, Flowerdew, Manley and Sabel 2008, Lebel, Pampalon and Villeneuve 2007, Weiss et al. 2007), residential choice analysis (Guo and Bhat 2007), and criminology (Duncan et al. 2003, Mayer and Jencks 1989, Sampson, Morenoff and Gannon-Rowley 2002). When the place in which a person resides provides the context neighborhood effects analysis attempts to model that relationship. Neighborhood effects is the framework in which this thesis will explore the relationship between modified measures of community context and the outcome, or response, of community collective efficacy. One of the main methodological issues within neighborhood effects research has been the operationalization of "neighborhood" (Chaix et al. 2009, Flowerdew et al. 2008, Guo and Bhat 2007, Lebel et al. 2007). Frequently, census enumeration units are used as surrogates for "neighborhood" although their nearly-arbitrary delineation has hardly any theoretical appeal. This arbitrary quality can lead to issues of the modifiable areal unit problem where redrawing the boundaries of enumeration units or aggregating units differently could yield significantly different results (Openshaw 1983). Using census boundaries to represent neighborhood is referred to as having "fixed boundaries" or being a non-overlapping "territorial neighborhood" (Chaix et al. 2009). Chaix et al. contrast that with the notion of "ego-centric neighborhoods" that have "sliding boundaries." In this instance, neighborhood boundaries are allowed to adjust in relation to an individual's location and context. Sliding boundaries can be created using circular buffers or network bands (on a street network) radiating from some origin such as a residence (Guo and Bhat 2007). While conceptually and computationally simple, the circular buffer approach is criticized as "naïve" since it assumes that context in all direction has an equal chance of affecting the individual outcome. Network bands are preferred since they conform to the transportation grid, restricting context along known paths. This is logical in urban contexts but may be problematic in rural settings. Once it is decided to utilize an "ego-centric neighborhood," the buffer or network band width must be decided upon. Spielman and Yoo (2009, p. 1104) criticize performing the analysis at various scales and then selecting the size with the best model fit value. Based on their simulations, regression results become inaccurate "if the spatial dimension of the independent variable(s) does not match the areal extent of the environmental influences on the outcome regression." Chaix et. al. (2009, p. 1309) disagree in part and suggest that after preliminary theoretical considerations, modification of buffers based on empirical evidence "is part of a more global process aimed at reducing multiple sources of bias and measurement error in eco-epidemiology." Spielman and Yoo (2009) also warn separately that by not taking into account individual observation variability, static buffers underestimate the relationship between context and outcome and increase the standard error of that estimated parameter. Therefore, neighborhoods should be allowed to vary in size and shape from individual to individual.

Towards an Understanding of Community Collective Efficacy

The theory of community collective efficacy was developed as part of the framework of social disorganization theory that dominated criminological research in the later 20th and early 21st centuries. Social disorganization theory itself grew from the quantitative work of the Chicago School whose influence was rooted in the early European urban social theory of Ferdinand Tönnies, Emile Durkheim, and Georg Simmel. One of criminology's seminal quantitative works based in social disorganization theory is <u>Juvenile Delinquency and Urban Areas</u> (Shaw and McKay 1969). Similar to Simmel's concept of "dynamic density" (Simmel 1903/1971), Shaw and McKay, using Burgess's (1925) concentric zone model of urban spatial structure and socio-economic data, found in Chicago that rates of crime increased towards the central business district as social control waned. Also, they found that locations with high crime rates were typified by physical deterioration, decreasing population, and mixed land use. Important to this thesis (and neighborhood effects research up to present), they additionally found that areas characterized by high levels of crime maintained those levels through time even when the population changed (in terms in ethnic immigrants and nation of origin). This evidence supports the notion that certain types of individuals are not inherently criminal, but that *place* matters and certain places act to reduce "social control." Those places are therefore criminogenic themselves – or more likely to produce criminal behavior amongst whomever resides there.

Focusing on physical deterioration and disorder, "broken windows" theory (Wilson and Kelling 1982) gained significant interest in both criminology and urban policy beginning during the 1980s. Wilson and Kelling use a shattered window as an analogy for initial signals of social disorder that can accumulate over time. Too many of these visual cues (e.g. excessive littering, loitering, panhandling, prostitution, and property damage) and the perception of disorder will increase. More windows are broken (presumably by restless white teens), which only reinforces that idea that no one, or no authority, successfully "controls" this location. These visual clues continue to increase as local residents begin to avoid this location, thus effectively reducing the actual social control in the area. Without local residents preventing these small, but numerous, criminal signals actual crime is given the opportunity to flourish (Kelling and Coles 1996). Wilson and Kelling describe how communities can deteriorate quickly if this process is not kept in check. They suggest that over the decades, the role of the police in the city has shifted towards protecting individual rights rather than protecting the community as a whole. In response, Kelling has worked with cities across the United States from New York to Seattle to implement zero-tolerance measures in an effort to deter drug use, homelessness, and criminal activity with mixed results and even criticism when faced with questions regarding the subjective nature of observing "disorder." What happens when policy makers or police officers have their own social construction of what defines "disorder" and which type of people are "disorderly"?

Due to the racialized concentration of poverty in U.S. cities (Massey 1990, Wilson 1987), many of these "broken windows"-based policies were targeted (indirectly or directly) towards black urbanites. Sampson and Raudenbush (2004) argue this is partially due to the perception of negative black ghetto stereotypes as disorderly within social psychology. When examining individual perceptions of disorder, they found that their own observations of disorder did predict the level of disorder perceived by residents, "but racial and economic context matter more." So, where physical disorder does somewhat shape how residents perceive their community, their observations were already "imbued with social meanings" indicating that the local characteristics of surroundings can play a more important role in predicting levels of crime.

Often identified as an opponent to broken windows theory (Sampson and Raudenbush 2004, St. Jean 2007), "collective efficacy" theory, also based in social disorganization theory, has grown in recognition within the past fifteen years as an alternative explanation of criminogensis. Like broken windows, collective efficacy theory relies on an indirect link for the development of criminal activity. Whereas broken windows theory suggests that physical disorder leads to actual disorder through a reduction in social control, collective efficacy relies on increased social cohesion and mutual trust amongst neighbors in an effort to enforce social control (an example of informal regulation), thereby limiting certain types of criminal behavior within a community (Sampson et al. 1997). While perhaps invoking a more uplifting characterization of social disorganization theory, the two components of cohesion and control must be in place in order for collective efficacy to facilitate change within a community. The relationship between concentrated urban poverty and violence is mediated by those two social processes of cohesion and control. Sampson, Raudenbush et al. (1997) explicitly identify the following three neighborhood characteristics as influencing collective efficacy: *concentrated disadvantage, immigrant concentration,* and *residential stability*. Supporting their hypothesis, concentrated disadvantage and immigrant concentration were found to significantly lower levels of collective efficacy within neighborhoods while residential stability significantly increased those levels, net of personal-level control variables. The social implications of these diverging theoretical frameworks are equally different. Rather than harsher punishments for criminal activities such as aggressive panhandling, loitering, and public drunkenness – while they may be disruptive, such acts hardly warrant arrest – collective efficacy suggests that efforts should be focused in building relationships within existing communities. Increasing these ties provide residents the ability to facilitate their own solutions from within the local community.

Small-Area Population Estimation through Disaggregation Techniques

The modeling of social processes necessitates contextual data of higher spatial resolution to produce more reliable results (Sampson, Morenoff and Earls 1999). Social scientists are often limited, however, to data that has already been produced. Regardless of the limiting factor, financial or practical, census data has traditionally been used to produce contextual measures of neighborhoods. Just because census data exist and are relatively accessible does not make them (and their spatial extent) the most appropriate for the task at hand. Generating contextual data of higher resolution requires a set of techniques that allow for the disaggregation of census information into more meaningful spatial units. Spatial interpolation, or the act of inferring information from known data into the geographical boundaries of another dataset, is one such technique, and there are a variety of methods in which it is performed.

Wu et al. (2005) and Zandbergen and Ignizio (2010) outline the specific methodologies and their various implementations of areal interpolation and statistical modeling in the context of population estimation. Areal weighting is cited as "the most basic form of areal interpolation" (Zandbergen and Ignizio 2010) and has been used frequently in the past (Flowerdew and Green 1991, Kim and Yao 2010). The data (i.e. population) is proportionally allocated to the target zones from the source zones according to the areal overlap of those zones. This method, like choropleth mapping, assumes that the population is equally distributed throughout the target zone which is rarely true in practice, but is relatively easy to implement nonetheless. Tobler (1979) improved upon the areal weighting method by introducing a "mass-preserving" function, or pycnophylactic approach. This method allows for "smooth" interpolation within *and* across areal units, and more importantly, that smooth interpolation can be correctly re-aggregated to the original source units – one would be given the exact values provided by the original dataset. The pycnophylactic method, therefore, is useful in transferring data from one set of boundaries to a smooth surface. This approach uses context from surrounding areal units, but other modeling techniques, such as dasymetric modeling, introduce ancillary datasets to more accurately predict the underlying population density surface.

Dasymetric modeling was first introduced in the United States by John K. Wright (1936) as a method for producing more accurate population density maps of Cape Cod, Massachusetts. Wright understood the local geography and realized that population distribution maps failed to acknowledge the vast uninhabited regions. This fallacy leads map readers to assume a lower overall population density for the populated regions. By excluding uninhabited regions of the cape and limiting population densities to expected values, Wright produced estimates of population density that more accurate reflected those in reality. Wright's approach is referred to as the limiting variable method. Eicher and Brewer (2001) compared this method with two others: the binary and three-class methods. The binary method is computationally simple, but only excludes uninhabited regions. The three-class method assigns population to three separate classes of density, but the density values are arbitrarily set (unless known *a priori*). This method can yield unrealistic high values if one class contains the majority of the population (e.g. urban), but is relatively small. Eicher and Brewer found the limiting variable method to be the most accurate statistically when compared against known population density values. Various datasets are suitable as the ancillary layer in population dasymetric modeling, including, but not limited to: land use / land cover, zoning boundaries, road network densities, and night-time lights (Zandbergen and Ignizio 2010). Land cover datasets such as the National Land Cover Dataset 2001 have been used successfully in population modeling and are well-suited for nation-wide projects given their spatial resolution and availability.

Recently, more “intelligent” forms of interpolation have improved our ability to model population density (Kim and Yao 2010, Mennis and Hultgren 2006). Mennis and Hultgren (2006) developed the “Intelligent Dasymetric Mapping” to address the flaws of the three-class and limiting variable dasymetric modeling techniques (i.e. the reliance upon *a priori* population density estimates). The method automatically searches the population layer and ancillary layer for source zones to sample population density values. For example, if the population layer is data from the U.S. Census Bureau in the form of census tracts and the ancillary layer is from the National Land Cover Dataset, the “Intelligent Dasymetric Mapping” method searches for census tracts that contain at least a certain proportion (set by the researcher) of a specific land cover. Many similar tracts are sampled, providing a population density value for that particular land cover class. This process is repeated for each class in the ancillary layer. Research can use this approach to effectively model population distributions without knowing *a priori* the actually population density of land cover classes. The method also takes measures to ensure pycnophylactic quality of disaggregation allowing target zones to be re-aggregated in any configuration. Therefore, this approach allows for the creation of new “neighborhoods.”

Kim and Yao (2010) present a hybridized approach combining the binary dasymetric and pycnophylactic models. The final product is well suited for cartographic purposes with smoothed zonal boundaries and maintains the volume of the population within each source zone, though smoothed boundaries are not required for raster-based spatial analysis. They found that this hybrid approach outperformed both areal weighting interpolation and the binary dasymetric approach. However, the “Intelligent Dasymetric Mapping” method combines automatic land cover sampling, the pycnophylactic property, and multiple classes of land cover, which in theory should provide a more realistic distribution of population than the binary approach.

Summary

In conclusion, “Intelligent Dasymetric Mapping” will provide a method for augmenting Census variable values in an effort to better model community collective efficacy through neighborhood effects analysis. “Intelligent Dasymetric Mapping” combines the necessary detail of the 3 (or more) -class method for modeling population in a varied urban context with the pycnophylactic property to preserve original count values while removing the subjectivity of *a priori* population density values for each ancillary class. Areal weighting interpolation will be employed as well for comparison. This thesis will conform to the ego-centric definition of neighborhood. Sliding boundaries will be generated around the residence of interview respondents reporting on perceptions of community collective efficacy. Both circular buffers and road-based network bands will be generated around each location based on the area of the block group in which the interview respondent lives. Each location, or observation, will be allowed to vary in size following Spielman and Yoo (2009). However, since perceptions of community collective efficacy are clearly tied to the definition of “neighborhood” itself, whereas other health outcomes in epidemiology might not be as explicit, I hypothesis that the specified “neighborhood” size is likely to approximate individual perceptions of his or her neighborhood. However, two additional buffer sizes will be created to examine the influence of scale upon the contextual analysis. Buffers will be created at 75% and 150% of the size of the host block group. The goal of adding additional buffers sizes is to simply observe how the relationship between community context and collective efficacy changes if social science research theorize collective efficacy as operating at scale greater than or less than that of the original host block group. Chapter 3 will explain in detail the methodology behind performing the areal interpolation of census variable values and their inclusion in neighborhood effects regression analysis.

CHAPTER 3

STUDY AREA AND METHODOLOGY

Family and Community Health Study

The Family and Community Health Study (FACHS) began in 1995 as a longitudinal study to examine family, community, and genetic influences on general well-being, as well as depressive and antisocial behavior, in children as they transitioned into adulthood. FACHS includes responses from African-American youth and their corresponding family and friends living initially in Iowa and Georgia. Since wave one in 1997, some individuals have moved throughout the country so that by wave four in 2006 the FACHS respondents resided in twenty-six states across the United States, though most individuals were still clustered around the original sampling locations in North Georgia (near Atlanta and Athens) and the Iowa cities of Des Moines, Waterloo, and Cedar Rapids (see Figure 3.1). The original families recruited for the FACHS project were selected from 1990 Census block groups with greater than ten percent of the families identifying as African-American and with poverty levels (as defined by the Census Bureau) between 10% and 100%. Collected in 1997, wave one included 867 children (around the ages of 10 to 12), from both Iowa and Georgia. Wave four was collected between 2006 and 2007 and contained 714 of the original individuals interviewed with a retention rate of approximately 82%. At the time of wave four data collection, the individuals, now young adults, ranged in age from 16 to 21, and lived in a variety of community types with various levels of racial composition and economic affluence. New to wave four, interviewers carried hand-held Global Positioning System (GPS) devices to record the latitude and longitude of the primary residence, the best friend's residence, and the favorite hangout location of the interview respondent. This study will focus only on the primary residence of each individual and not the context of hangouts or friends locations, though they should be investigated further.

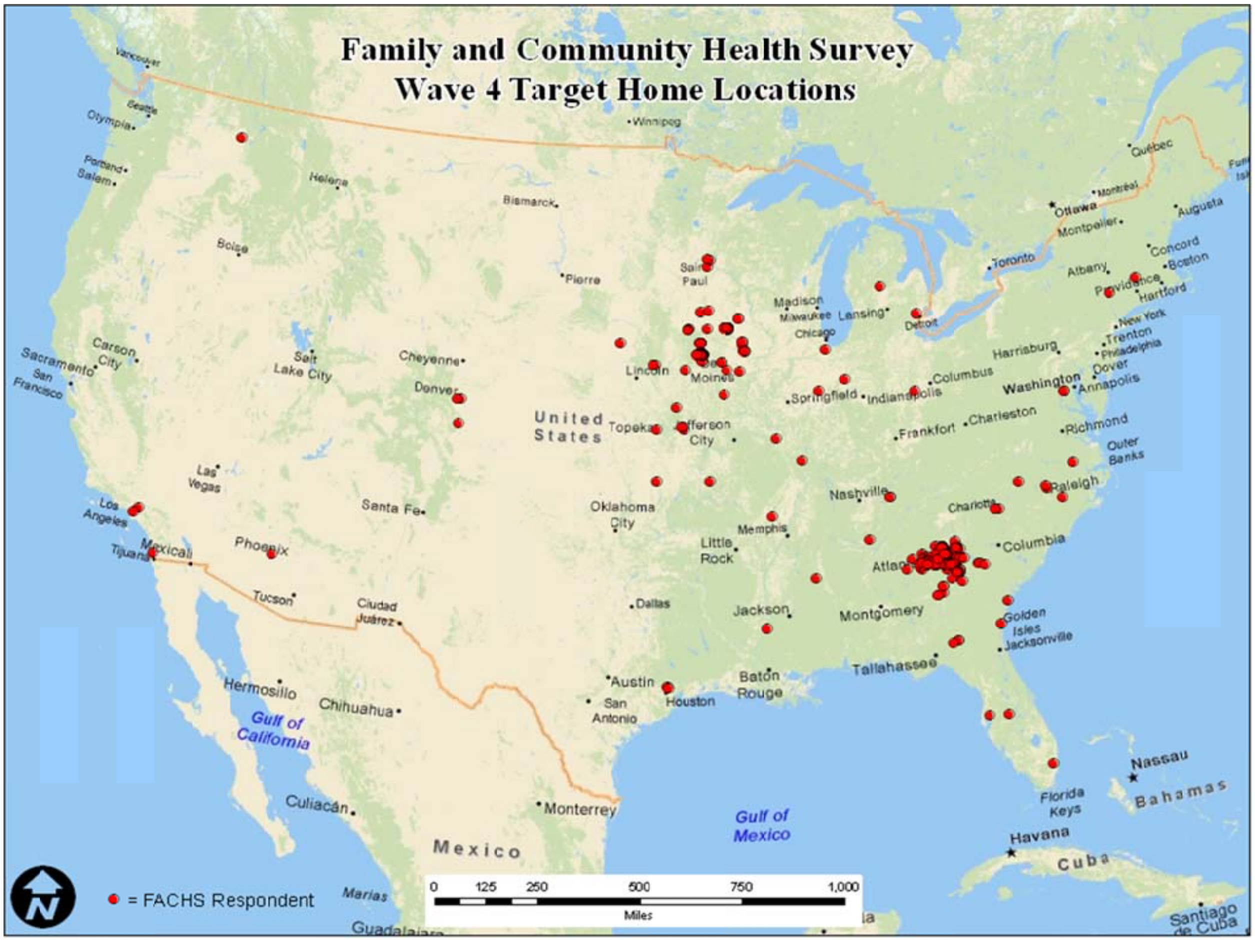

Figure 3.1: Spatial distribution of FACHS wave 4 respondent residences. The majority of participants are still clustered around Iowa and Georgia.

Data

This thesis utilizes individual-level, community-level, and nation-wide data derived from three main sources including the Family and Community Health Study, the United States Census Bureau, and the United States Multi-Resolution Land Characteristics Consortium (MRLC). Specifically, responses from the fourth wave of FACHS interviews will provide the measure of community collective efficacy. Additional data collected from those individuals interviewed included: age, gender, family income, and educational attainment. Community context measures will be derived from the 2000 Census long-form sampled data aggregated to block group level geography. The MRLC, through nine federal agencies, including the U.S. Geologic Survey, the Environmental Protection Agency, the U.S. Forest Service, the National Oceanic and Atmospheric Administration, the National Aeronautics and Space Administration, the Bureau of Land Management, the Natural Resources Conservation Service, and the U.S. Fish and Wildlife Service, generated the 2001 National Land Cover Dataset from Landsat satellite images which will be used as the ancillary layer for dasymetric modeling.

Dependent Variable: Perceived Collective Efficacy

Multiple regression analysis relies on many independent, or X, variables to explain or predict one dependent, or Y, variable. This thesis explains individual-levels of perceived community collective efficacy in terms of other individual, familial, and community contextual variables. Following (Sampson et al. 1997), two dimensions of community collective efficacy (i.e. social cohesion and willingness to intervene) were measured by FACHS using a 14-item additive scale. Respondents were asked a series of questions regarding the behavior of other individuals living within their neighborhood of primary residence over the previous twelve months. Specifically, respondents were asked how likely an adult was to intervene if: (i) they saw someone breaking the law, (ii) a teenage who showed disrespect to an adult, (iii) teenagers got loud or disorderly, (iv) a group of teens were fighting with each other; (v) they saw public intoxication or drug use. These five questions measured the willingness to intervene dimension of community collective efficacy. Additionally, to assess the social cohesion, or social ties, dimension, respondents assessed the likelihood of the following scenarios within their neighborhood: (i) when there was a problem, the people in the area got together and dealt with it;

(ii) the people in the area were a fairly close-knit group; (iii) no one in the area really cared much about what happened to anyone else; (iv) there were adults in the area that teens looked up to; (v) People were willing to help each other out; (vi) many of the adults didn't get along with each other; (vii) people in the area shared the same values; (viii) people trusted each other; (ix) people in the area mostly went their own way. Responses for each dimension were coded so that higher scores corresponded to higher levels of social cohesion and a greater willingness to intervene (responses to questions (iii) and (vi) were reversely coded).

Independent Variables: Individual and Family-Level Controls

In order to take into account individual and family-level variation among FACHS respondents, FACHS included control variables of participant age, gender, family income, and educational attainment. Other previous research employed additional individual controls including parenting type and both violent and depressed behavior (Stewart, Simons and Conger 2002, Simons et al. 2005, Brody et al. 2001, Natsuaki et al. 2007). This thesis will not include such variables from the FACHS project, not as an omission, but rather to simplify analysis and emphasize the effects of geographic context in understanding community collective efficacy. Additionally, the control variables were not found to significantly impact collective efficacy in preliminary analysis. Their exclusion also preserved a greater number of degrees of freedom with which to test the resulting regression models. Table 3.1 below lists the independent variables and their expected contribution to community levels of collective efficacy.

Independent Variables: Community Context

Collective efficacy theory suggests that certain constructs within a geographic location act to mediate specific types of criminal behavior (Sampson et al. 1997). Additionally, the same researchers posit that structures within communities act to either promote or discourage the formation of collective efficacy between residents. This thesis employs five community-level variables to measure geographic contextual information: concentrated economic disadvantage, percent of the population identifying as African-American, percent of the population born outside of the United States, percent of population maintaining their residence over the past five years, and settlement type. These five variables were derived from the United States 2000 Census using data aggregated at the block group level.

Concentrated disadvantage was constructed as an index calculated from six Census 2000 variables including: per capita income, percentage of households under the poverty level, percentage of population without a high school degree, percentage of households receiving public assistance, percentage of single motherhood, and percentage of males not in the labor force. For each block group (and the subsequently recreated interpolated neighborhoods), the six variables were standardized and then summed (per capita income was reversely coded) to determine the concentrated disadvantage for each area. Factor analysis from preliminary research shows that this index has a reliability of 0.89.

Previous studies included the percentage of African-American population as a component for concentrated disadvantage (Morenoff, Sampson and Raudenbush 2001, Sampson et al. 1997). However, this thesis excludes this variable from the disadvantage index. While evidence exists that clearly demonstrates the structured inequality between predominately white and predominately black neighborhoods (Massey 1990, Wilson 1987), not all African-Americans reside in economically disadvantaged communities and there exists numerous examples of middle-class black communities in cities across the United States. Factor analysis from this dataset also supports this logic since the percentage of African-American population does not load well with the other components of concentrated disadvantage.

Additionally, since this thesis employs a sample of African-American adolescents, it is expected, contrary to previous studies of community collective efficacy, that a higher proportion of African-Americans in the respondent's neighborhood might improve their perception of community collective efficacy by fostering population homogeneity. Conversely, many social theorists, including Sampson et al. (1997), suggest that the introduction of immigrants, or foreign-born populations, yields an increase in population heterogeneity that then acts to decreased community collective efficacy. Therefore, the percentage of foreign-born populations will be included in this analysis as an additional partial indicator of community context. However, this category of individuals might be underrepresented within the sample of block groups if FACHS respondents reside in highly racially segregated areas.

As a proxy for residential stability, the percentage of individuals residing in the same location for the previous five years will be included in this project. It is predicted that a higher rate of residential turnover (i.e. a relatively large rate of in-migration and out-migration) should provide neighborhood residents fewer opportunities to create lasting relationships built on mutual trust. As the percentage of the population who maintained their residence over the past five years increases in a neighborhood, the residential stability will be higher which should in turn yield higher levels of community collective efficacy perceptions among individuals.

Finally, community, or settlement type, will also be introduced to partially explain the variations in community collective efficacy. Social disorganization theory suggests that different spatial contexts have varying impacts on social ties (Shaw and McKay 1969). For example, as population density increases, individuals are less likely to form long-lasting bonds since economic specialization limits interpersonal transaction times, and increases the quantity, but not necessarily quality, of personal interactions (Simmel 1903/1971). This study employs a system of settlement classification based on two location type characteristics defined by the U.S. Census Bureau: population size and urban place designations. Settlement type is treated as a continuous variable measured from "rural" coded as 0 up to "large, incorporated urbanized area" coded as 8. Following social disorganization theories, it is expected that an increase in settlement type size will act to decrease community collective efficacy. Settlement type was determined for each interview location and was not augmented during the dasymetric process. This variable can therefore be considered a control for each observation taking into account variations in local population size and urban incorporation.

Table 3.1: Multiple regression variables and their relationship to collective efficacy (CE).

Independent Variable	**Expected Relationship**
Settlement Type (Coded Continuously)	Strong; more "urban" regions will produce lower CE.
Concentrated Disadvantage (Index)	Strong; higher economic disadvantage will yield lower CE.
African American Population (%)	Moderate; higher %-age of black residents yields higher CE.
Residentially "Stable" Population (%)	Strong; a more "stable" neighborhood produces higher CE.
Foreign-born Population (%)	Moderate; higher %-age of foreign-born yields lower CE.

Dasymetric Modeling Ancillary Layer

Lastly, this thesis incorporates the second version of the 2001 National Land Cover Dataset generated by the MLCR. The ancillary layer for type of dasymetric modeling attempts to correct the assumption of areal data that population (or any other units of study) is evenly distributed throughout the boundaries of the areal units (i.e., population is clustered through census block group regions). The ancillary layer provides an estimate of where the population is actually distributed since one would expect to find greater human population densities in an urban area of high-rise apartments compared to rural farmland. The NLCD was initially classified into 24 land cover classes ranging from "Highly urbanized" to "Forested" to "Ice Covered." Table 3.2 contains the original NLCD classes and the resulting combination of recoded values to be used in this study. Numerous regions, including "Water" and any

Table 3.2: NLCD to thesis project land cover classification cross-walk table.

National Land Cover Dataset 2001*	Project Classification
11. Open Water	1. No Population
12. Perennial Ice/Snow	1. No Population
21. Developed, Open Space	2. Urban, class 1
22. Developed, Low Intensity	3. Urban, class 2
23. Developed, Medium Intensity	3. Urban, class 2
24. Developed, High Intensity	4. Urban, class 3
31. Barren Land	1. No Population
41. Deciduous Forest	5. Forested
42. Evergreen Forest	5. Forested
43. Mixed Forest	5. Forested
52. Shrub/Scrub	6. Agriculture/Grassland
71. Grassland/Herbaceous	6. Agriculture/Grassland
72. Sedge/Herbaceous	6. Agriculture/Grassland
81. Pasture Hay	6. Agriculture/Grassland
82. Cultivated Crops	6. Agriculture/Grassland
90. Woody Wetlands	1. No Population
95. Emergent Herbaceous Wetlands	1. No Population

* The NLCD classification is modified from the Anderson Level I classification (Anderson et al. 1976).

"Wetland" category, were recoded to "no population" since it is very rare one would find a FACHS respondent residing in those land covers. The NLCD is a raster image (30 meter pixel resolution) derived from composite low-cloud cover Landsat satellite images over the course of three seasons around 2001. The MLCR used the true-color (Red, Green, and Blue) and the near-infrared image bands collected by the satellite sensor to classify pixels according to a modified version of the Anderson Level I classification system used in vegetation mapping.

Methodology

The methodology for this thesis consists of five separate stages. The first stage redefined the unit of study in relation to neighborhood effects. For this work a "neighborhood" was operationalized as either a circular buffer or a road network-based buffer around the FACHS respondent's home. To examine the impact of the scale of that "neighborhood," buffers of three sizes were generated at 75, 100, and 150% of the host block group's radius. The second stage disaggregates Census data either through areal weighting (AW) interpolation or "Intelligent Dasymetric Mapping" (IDM). The disaggregation process occurs separately for each numerator and denominator of associated Census ratio values. The interpolated data will then be re-aggregated to the newly defined "neighborhoods." Next, stage three evaluates the performance

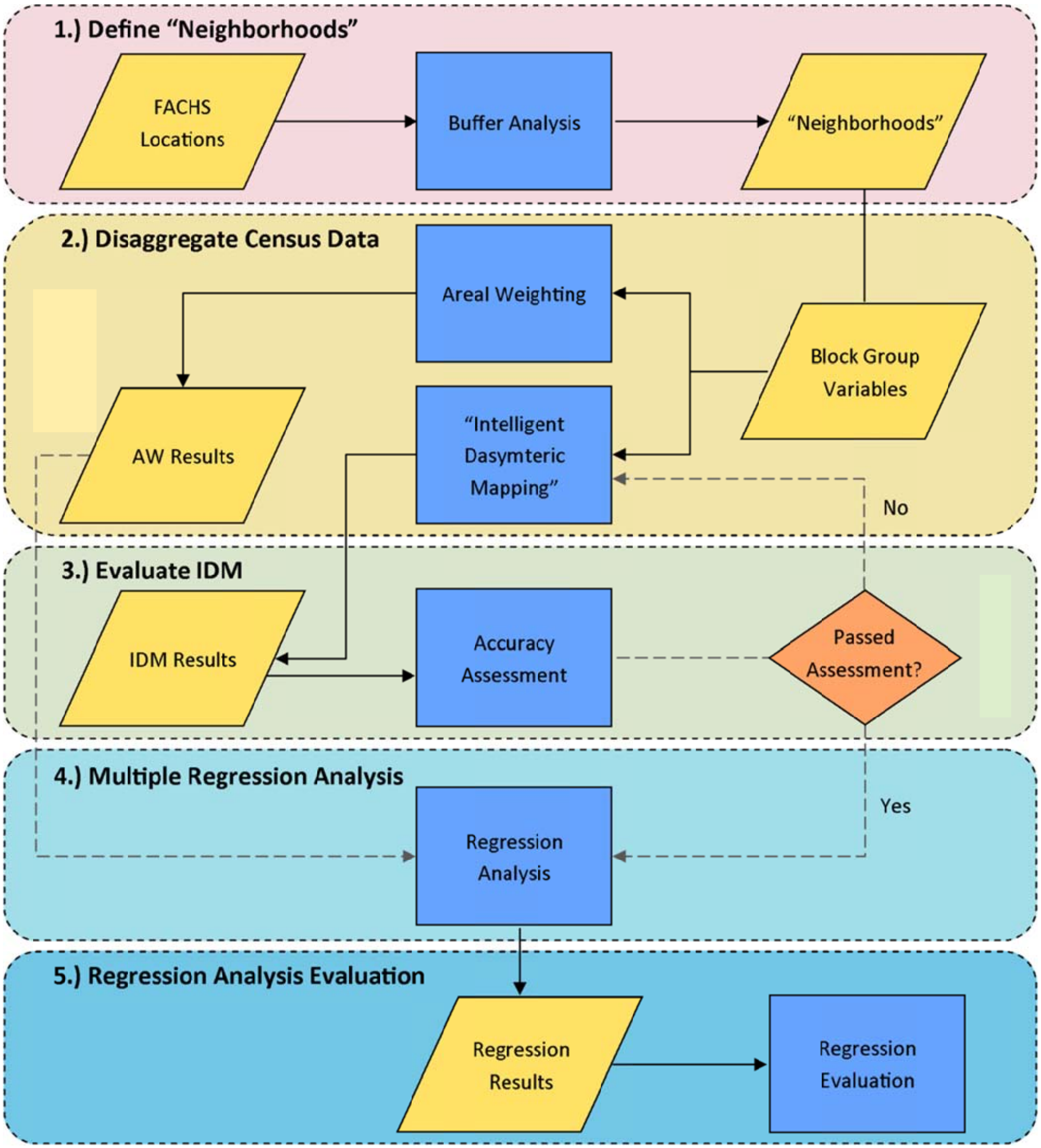

Figure 3.2: Flow diagram illustrating the analysis and evaluation of thesis methodology.

of IDM through comparison with Census block level data. Visual inspection and quantitative error analysis will provide metrics for comparison with previous studies employing IDM. The fourth stage performs linear regression analysis in an attempt to model community collective efficacy. A total of thirteen simple models were specified to examine the changes in model parameter estimation and overall model fit when varying contextual technique (i.e., traditional block group measures, AW or IDM interpolation), buffer type (i.e., circular or network), and scale (i.e., buffers at 75%, 100%, or 150%). Finally, the last stage evaluates the resulting regression models to determine if any of the theoretical assumptions of linear regression were violated. Figure 3.2 above outlines the datasets, processes, and decisions involved in the thesis.

"Neighborhood" Definition

Previous neighborhood effects studies relying on census-based data have been typically limited to the geography in which the data was collected (Comstock et al. 2010, Duncan et al. 2003, Holloway et al. 1998, McNulty and Holloway 2000, Mennis 2006, Morenoff et al. 2001, Sampson et al. 2002, Sampson and Raudenbush 2004, Sampson et al. 1997, Simons et al. 2002). In the case of the United States, that means utilizing block groups or tracts to represent a "neighborhood." Block-level data is the smallest unit in which the Census Bureau releases population count data, but in the interest of maintaining individual privacy certain household economic statistics are not reported. Block groups, which are aggregations of census blocks, are the smallest units of analysis which release all social and economic variables to the general public and encompass 1,500 people, optimally, whereas census tracts are aggregated from block groups and are even larger in size, containing roughly 4,000 people per tract (U.S. Census Bureau 2000).

Since census boundaries are drawn with the goal of maintaining similar population sizes between administrative units, block group sizes typically decrease in area as one travels towards more densely populated regions (i.e. urban and suburban neighborhoods). Block group-level data will be employed for this study in an effort to obtain more accurate population characteristics across small areas. This type of contextual analysis has traditionally assumed that the populations residing within each census boundary are homogeneous which is not necessarily true. For example, a location near the edge of hypothetical block group A may be influenced by the population characteristics of adjacent block group B. This concept, along with the issue of arbitrarily defined administrative units (i.e. the scale effect and the zone effect), is commonly referred to as the modifiable areal unit problem and contextual analysis research should attempt to incorporate a solution to this complex problem (Green and Flowerdew 1996, Openshaw 1983, Wong 2009, Wong and Lee 2005).

This research departs from what has been considered traditional community contextual analysis by redefining the unit of study. Previous studies in areas such as environment justice have already shown the benefit of redefining administrative boundaries to create more contextually-sensitive variables (Maantay 2002, Maantay 2007, Maantay and Maroko 2009, Maantay, Maroko and Herrmann 2007, Mennis 2002, Boone 2008). One of the most common, and most easily computed, methods for transferring data from administrative boundaries, or source zones, to newly created "neighborhoods," or target zones, is called areal weighting interpolation and will be used in this study (Kim and Yao 2010). A more complex method of interpolation, IDM (Mennis and Hultgren 2006), which incorporates ancillary data to model the actually population distribution, will also be used and compared against the more traditional measures of geographic context (e.g. derived from census tracts or block groups).

The dependent variable of this study, perceived community collective efficacy, was reported by FACHS respondents. The latitude and longitude of each interview location with FACHS participants was collected by field personnel using a hand-held GPS device and then incorporated into a geographic information system (GIS) where a circular buffer was generated around each interview location. The buffer radius was calculated so that the resulting buffer area will equal the area of the original census block group in which the FACHS respondent resided. Each circular buffer is generated to represent a new, more contextually-sensitive (relative to an arbitrarily defined Census block group) neighborhood for each individual following Waldo Tobler's "first law of geography" that nearer locations have a greater effect than locations at a greater distance (Tobler 1970). Next, that same buffer radius was used to create what are called "service areas" in network analysis. A service area is the area around a

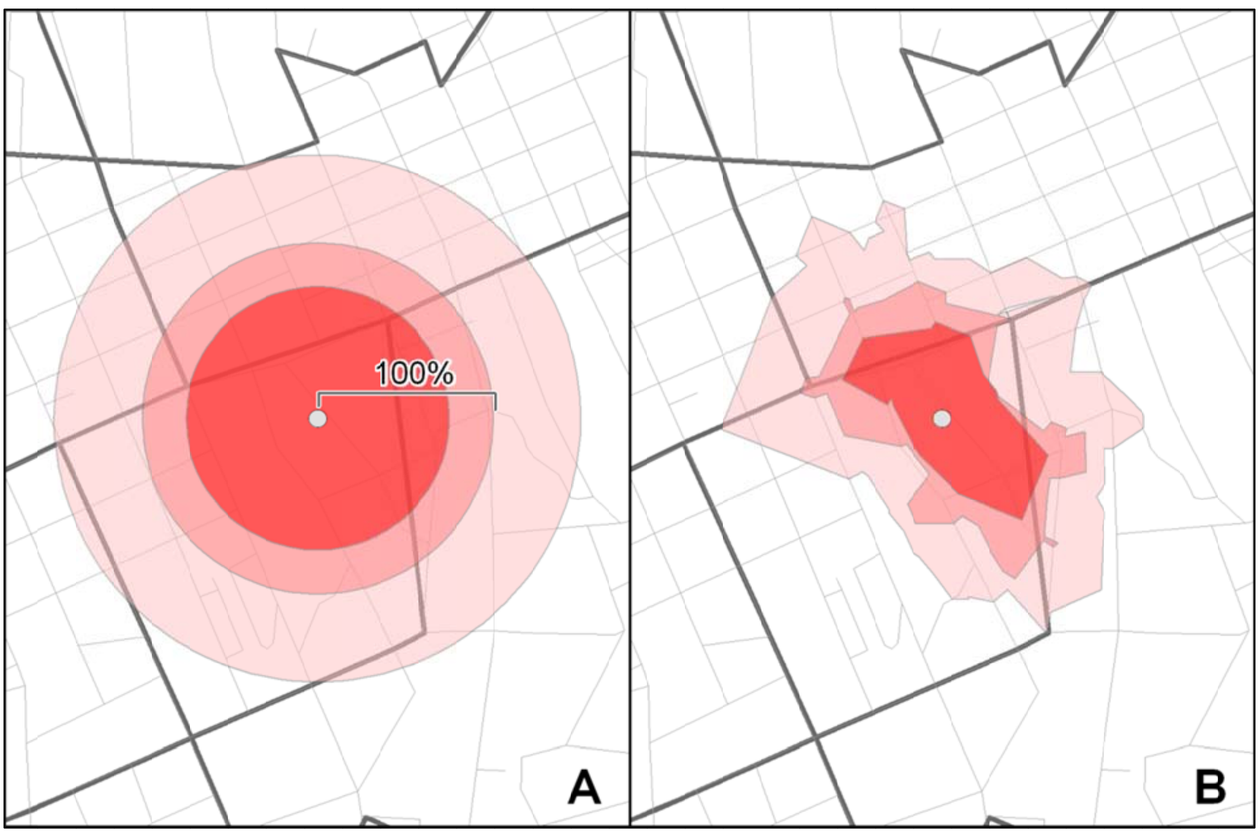

Figure 3.3: Comparison of A) circular and B) network-based "neighborhoods."

location that is accessible within a set time or distance. From the interview location, the road network-based buffer radiates outward following the road until the buffer radius is reached. At that point, the ends of the roads are connected to create a polygon. For this analysis, that polygon was trimmed to only include areas within 100 meters of the road itself, assuming most persons will not live too far beyond that distance. In an effort to understand the effects of scale on the notion of community collective efficacy, eight additional concentric "neighborhoods," or relative focal areas (rFA) were generated by adjusting the buffer area to 75% and 150% of the original buffer size. Like the changing the aperture on a camera lens, adjustment of the focal areas (i.e., geographic areas of interest) changes how the interpolation techniques utilize contextual information. The focal areas are constructed relative to each FACHS location. The three rFA are depicted in Figure 3.3 (Figure 3.3A represents circular buffer neighborhoods while Figure 3.3B represents network-based buffer neighborhoods). Instead of radiating equally from the FACHS respondent's residence, the network-based buffers expand outward only along nearby streets. This research assumes that the generated circular and road network-based buffers effectively approximate "actual" community boundaries on the ground and that the responses from FACHS members apply within those same boundaries.

Small-Area Population Estimation

Estimation Based on Areal Weighting Interpolation

Once the new "neighborhoods" were generated through buffer analysis, they were be overlaid in ArcGIS 10 with the original Census block group boundaries containing the necessary social and economic variables. The intersection then contained values of one or more Census block groups. If the buffer was completely within one block group, the variable values remained the same. However, if more than one block group intersected the circular or network-based buffer, the values of those separate block groups were weighted (proportionally to the area of the individual block group within the buffer) and summed, providing a new areal weighted census variable value. This process was performed for each the 16 unique numerators and denominators. Those values were then divided appropriately to generate the nine ratio values.

Estimation Based on "Intelligent Dasymetric Mapping"

The process for generating the second set of interpolated census variables is similar to that of areal interpolation, however, rather than combining the buffer neighborhoods with the original Census boundaries, they were overlain against disaggregated areas produced from IDM. The boundaries of those new areas are associated with the selected ancillary layer and are therefore 30 meters by 30 meters raster grid cells (the same as the NLCD 2001). The IDM process follows the methodology first written and scripted by Mennis and Hultgren (2006) in Python geoprocessing programming language. Rather than relying on *a priori* population density estimations (except for zones of population exclusion where it is assumed population density is zero), the script selected Census block groups associated with the remaining land cover classes. The population density samples were used to disaggregate population counts from block groups to pixels based on the land cover. This process was repeated for every census variable. For example, to produce a value for the proportion of foreign-born residents within a block group (a ratio value), one must divide the total number of foreign-born residents by the total number of residents in that block group (both integer values). IDM interpolation was first performed separately on both the total number of foreign-born residents (the ratio numerator) and the total number of residents (the ratio denominator). Then, the IDM pixels were re-aggregated into the newly created buffer neighborhoods using the "zonal attributes" tool in ArcGIS 10. Finally, new ratio estimates were created by dividing the new numerator by the new denominator. This process was repeated for each of the community-level census variables and their individual components. Figure 3.4 delineates the exact regions in the United States where IDM was performed. Block groups were grouped together to ensure a large enough region was available for the IDM process to sample population densities. The NCLD raster image was clipped to the 45 block group clusters (BGC) and the IDM process was performed in each BGC iteratively through Python geoprocessing.

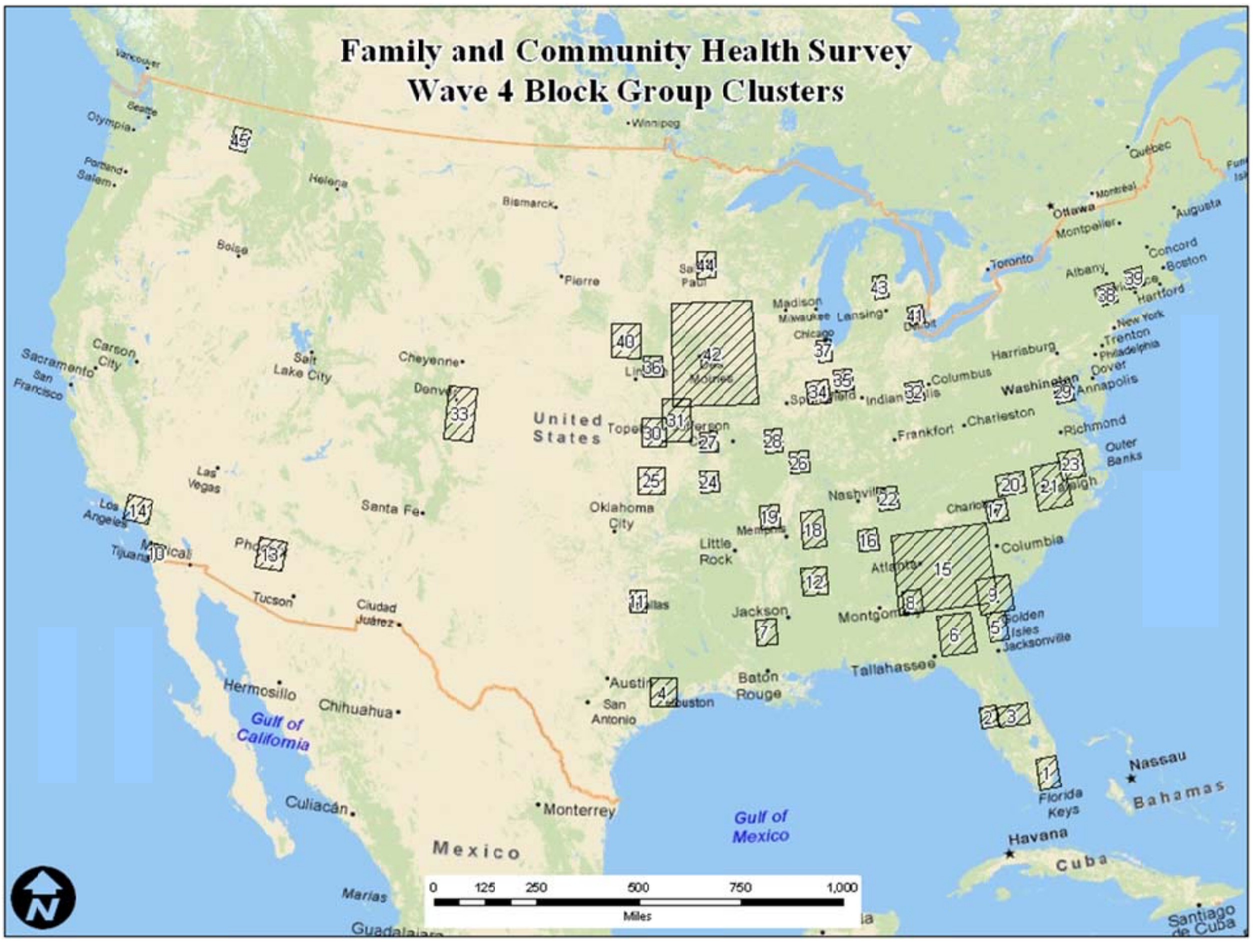

Figure 3.4: Location of 45 Block Group Clusters. Block group clusters were generated by selecting the Census block groups within 20 miles of the FACHS participant providing more than enough geographic contextual information for analysis within the new buffer neighborhoods. Overlapping regions were aggregated together (example BGCs 15 and 42).

Accuracy Assessment of "Intelligent Dasymetric Mapping"

Following Eicher and Brewer (2001) and Mennis and Hultgren (2006), this thesis evaluates the IDM performance by examining the coefficient of variation of the root mean square error, *CV*(RMSE), against Census data of a smaller geographic level with known values. Where previous studies have disaggregated Census tracts and compared the dasymetric results to block groups, this thesis disaggregated Census block groups and compared IDM results to block level data. For privacy reasons, block level sample data is only available for population counts by race and a limited number of other variables, but not for most economic variables utilized in this thesis. However, an evaluation of the total population does provide an indication of the overall accuracy of the IDM method. Rather than re-aggregating to the circular or network-based "neighborhoods," the accuracy assessment involved re-aggregating pixel-sized IDM results to Census blocks (for which the population counts are known). The RMSE and *CV*(RMSE) were calculated for each block group. Appendix A contains the descriptive statistics for the resulting ratio variables generated by both interpolation techniques.

Multiple Regression Analysis

Multiple Ordinary Least Squares (OLS) regression analysis provides estimates for the individual effects of the independent variables (i.e. the control and contextual variables) upon the dependent variable of community collective efficacy. After generating two new sets of community contextual variables (in addition to the original Census variables), regression analysis was performed with each set for each buffer size. The specified regression model is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$$

where: β_0 = y-intercept, X_1 = settlement type, X_2 = concentrated economic disadvantage, X_3 = residential stability, X_4 = % African American population, X_5 = % foreign-born population, and ε = error term. This model was specified exactly the same after each set of variables was generated for each buffer size and then model fit parameters were compared, including: coefficient of determination, RMSE, global *F*-test, and the Akaike information criterion (AIC). Individual coefficients were assessed as well using the Student's *t*-test and reporting the corresponding *p*-value. A total of 13 regression models were assessed including an initial model using traditional context measures, six models with AW generated variables, and six models with IDM generated variables (including in those six are two sets of three models using 75%, 100%, and 150% circular and network based rFA). Model performance was evaluated empirically and theoretically. The final models were compared using global fit criteria and parameter estimate significance. Statistically insignificant variables were not removed from the models to ensure that differences in these metrics were caused by changes in the interpolation technique, buffer type, or scale and not due to model specification.

Initial Regression Diagnostics

The final portion of this methodology focuses on assessing the theoretical assumptions associated with linear regression. OLS regression relies on a set of assumptions to ensure that the model results from a **b**est (i.e. most efficient), **l**inear (form of relationship), and **u**nbiased **e**stimator (BLUE) (Hamilton 1992). The following five assumptions that must be met for OLS to be theoretically BLUE: "1.) fixed X [values], 2.) errors have zero mean, 3.) errors have constant variance (homoskedasticity), 4.) errors are uncorrelated with each other (no autocorrelation), and 5.) errors are normally distributed." Or simply stated: "assume the linear model is correct, with normal, independent, and identically distributed (normal i.i.d.) errors" (Hamilton 1992). These theoretical assumptions are rarely completely fulfilled in practice, however, so the resulting models were examined for the following conditions: having a linear functional form, influential observations, heteroskedasticity, and multicollinearity. Any violations of these assumptions must be handled accordingly.

Appendix B contains pair-wise bivariate correlations between variables generated by each technique. These values, along with variance inflation factors (VIFs) below 1.7 suggest that multicollinearity was not an issue with these models. Scatterplot diagrams between each independent variable and the dependent variable, along with partial regression plots (added variables plots in STATA 10), suggest that the models have a linear form but that influential observations and heteroskedasticity could be problematic for the settlement type variable and possibly the percent black population variable. White's general test for heteroskedasticity (as calculated in STATA) confirmed that suspicion. It is believed that the heteroskedasticity for the percent black population was caused by influential observations in the form of dependent variable outliers. Since there was no evidence of measurement error, those observations were left in the model. Transforming the dependent variable (raised by power of two) pulled in those outliers and removed significant heteroskedasticity, but made the regression results difficult to interpret. Instead, robust standard errors ("vce(hc3)" in STATA) were used to better ensure success when performing inferential hypothesis testing of parameter estimates. STATA 10 documentation suggest following Davidson and MacKinnon's (1993) recommendation of using HC3 when heteroskedasticity is present in the regression model.

CHAPTER 4

DESCRIPTIVES AND RESULTS

Areal Interpolation Results

"Intelligent Dasymetric Mapping" was performed across all forty-five block group clusters in the United States. Block group cluster 15 was selected for quantitative error assessment since it contained the largest share of FACHS interview points and the second largest land area. Examples of the dasymetric inputs (i.e., Census values and the NLCD, respectively) and total population output are illustrated in Figures 4.1-4.3. Atlanta is labeled for reference.

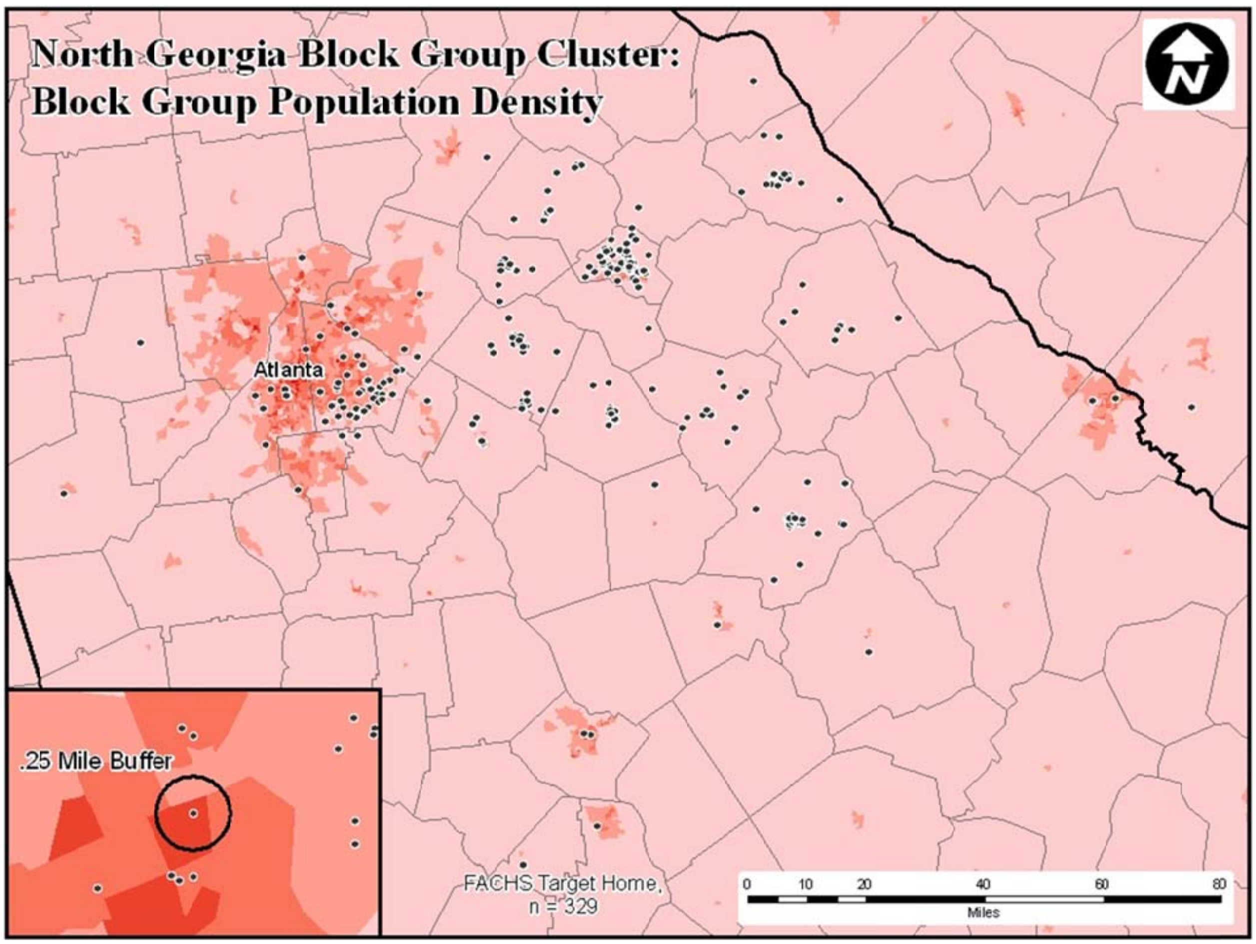

Figure 4.1: Block Group Cluster 15 Population Density by Census Block Group.

Darker shades of red indicate a higher population density within each block group. Urban areas including Atlanta, Macon, and Augusta, Georgia are clearly visible on the map. Unfortunately, this choropleth map assumes that population density is the same within each block group, which is not the case. The inset map in Figure 4.1 illustrates the relative size of a 0.25 mile radius circular buffer surrounding a FACHS location in relation to urban block groups. In this instance, the new "neighborhood" includes parts of at least four block groups and social analysis should therefore include their attributes. Figure 4.2 illustrates the variability of land cover compared to FACHS locations and the same circular buffer. Within the example buffer are at least three different land cover classes each with their own associated population density.

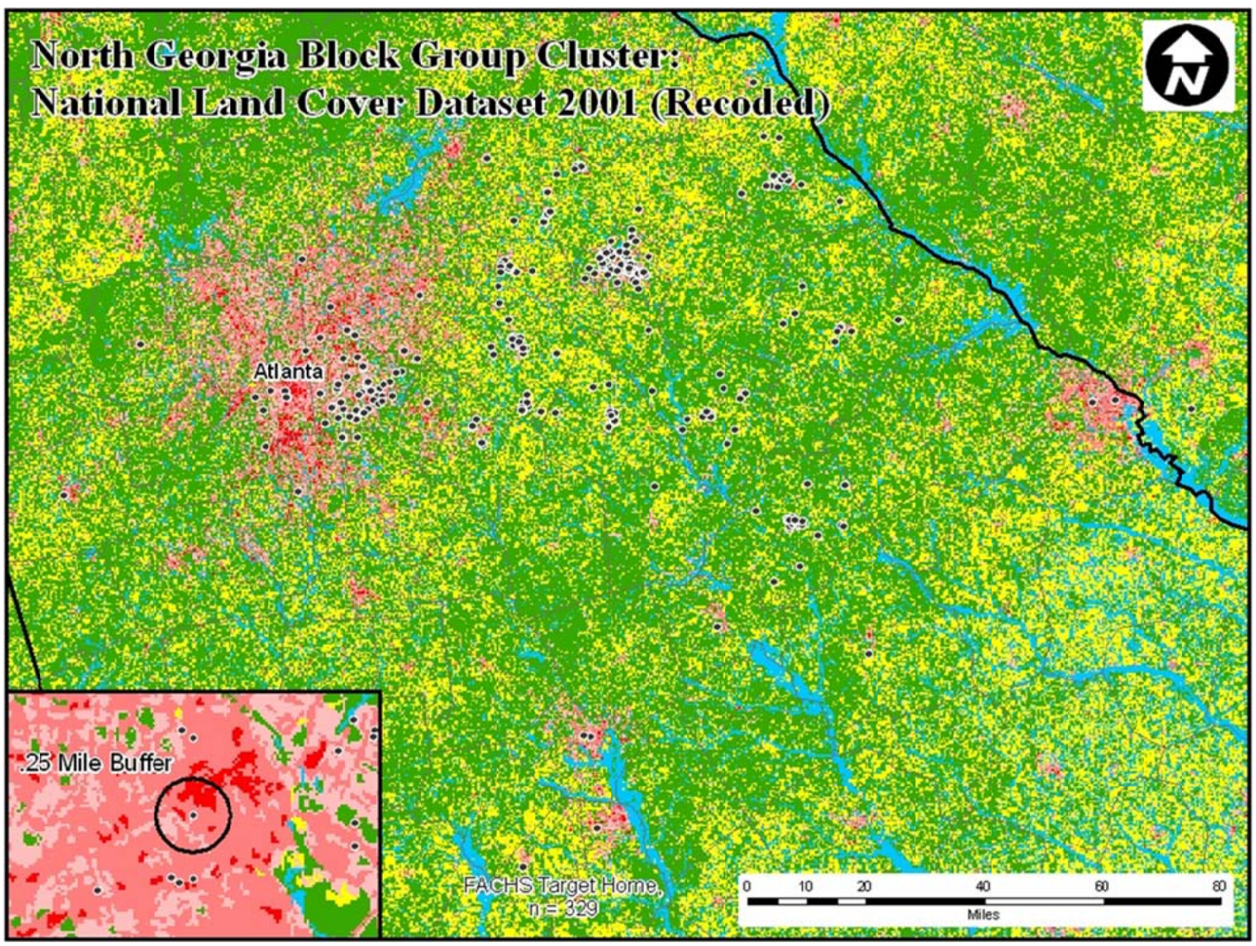

Figure 4.2: Block Group Cluster 15 Land Cover.

Finally, when merged together though the dasymetric process, census and land cover data produced the population density map in Figure 4.3. Again, darker shades of red correspond to areas of greater population density. The difference in this map is that each raster pixel, measured 30 meters by 30 meters, illustrates density at a much finer spatial scale than previously allowed. Those raster image pixels act as the building blocks of the newly created area buffers (i.e., "neighborhoods") when aggregated together. The summation of those pixel values provides an estimate for population or household count values in the new target zones.

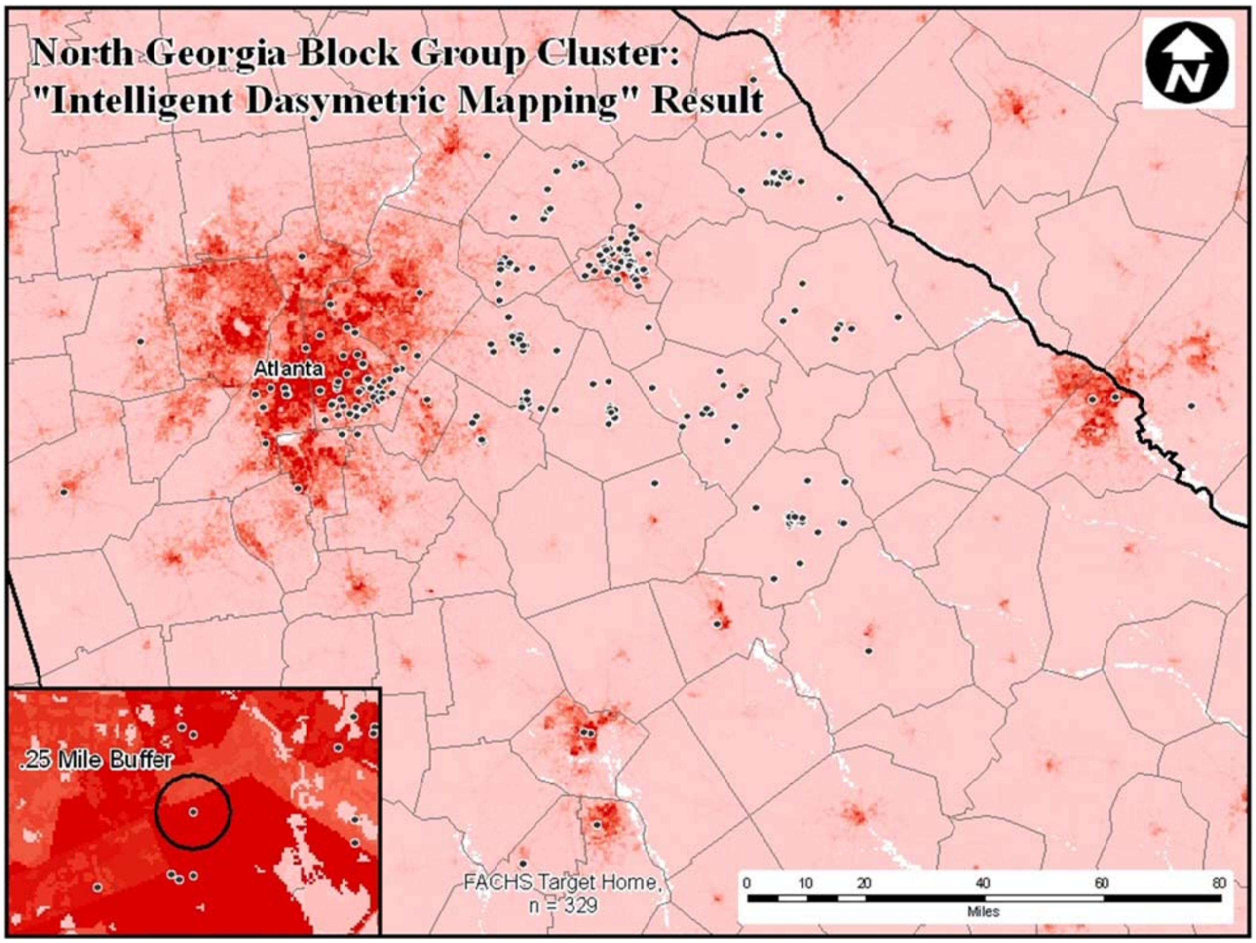

Figure 4.3: Block Group Cluster 15 "Intelligent Dasymetric Mapping" of Total Population.

The accuracy of the IDM process was measured with the coefficient of variation of the root mean square error between estimated and known Census block values averaged across Census block groups. The average *CV*(RMSE) (approximately 0.05) was similar in magnitude to that reported in previous studies, though it did not perform as well as IDM interpolation results presented by Mennis and Hultgren (2006). It did perform better than road network-based interpolation performed by Reibel and Bufalino (2005). Table 4.1 compares the average *CV*(RMSE) of the results of this thesis and two similar studies. There are likely two reasons for the slightly weaker results in this analysis. Firstly, the NLCD, while useful in its consistency

Table 4.1: Comparison of mean *CV*(RMSE) of various interpolation techniques.

Technique	***CV*(RMSE)**	**Source**
IDM – 50.1%	0.05	Present Study
IDM – 70%, 80%, 90%	0.019, 0.0215, 0.0185	Mennis and Hultgren (2006)
Street-weighted Interpolation	0.207	Reibel and Bufalino (2005)

and availability across the entire United States, cannot compare to the higher classification accuracy of manually-interpreted land-use/land-cover datasets derived from USGS digital orthophotograph quadrangles with a raster cell size of 1 meter square. Secondly, the sampling parameters were relaxed compared to those of Mennis and Hultgren (2006) in order to ensure an approximately statistically significant sample of at least thirty block groups per land cover class. The IDM process selected all of the block groups in the study area that have a user-specified threshold of a particular ancillary layer class. For this thesis, that threshold value was relaxed to a majority, or 50.1% cover, meaning that for block groups to be used in the population density sample of ancillary class 1, they must contain at least 50.1% ground cover of that class. Mennis and Hultgren (2006) tested 19 separate areal interpolation techniques, three of which used the same selection method, but the tested threshold values of 70%, 80%, and 90% cover. In theory, raising the threshold value restricts potential population units (e.g., Census block groups or tracts) to more realistic values. Higher threshold values yield more restrictive selections and thus fewer samples are used in calculating the relative population density of each ancillary class.

Another issue resulting from the IDM process was the creation of the new ratio values. The Census variables all potentially range from 0% to 100% (except for per captia income). For both AW and IDM, the numerator and denominator of those ratio values were processes separately and then division produced the final value. In rare instances (less than 1% of observations) the value for percent black population exceeded 100% (typically between 101% and 109%). Generally this occurred when the proportion of black residents in the host block group was very high. Since these values (i.e., the numerator and denominator) are processed separately, it is possible that during the block group selection for ancillary class relative densities, the density generated for one variable in one particular class could be different than the density generated for another variable in another class. This error could be explained if African-Americans are living in greater concentration compared to the overall population in spatially-restricted, but still densely-populated, land covers (such as "Urban, class 3" in this study). While no ratio values were below 0%, those values larger than 100% were truncated at the logical limit.

In all studies comparing dasymetric modeling (with any type of ancillary layer) to traditional areal weighting interpolation, dasymetric modeling outperforms areal weighting in terms of total population counts. The interpolation methodology in this thesis (IDM) produced similar results which serve to validate dasymetric modeling as an appropriate technique for estimating small-area population values. As this discussion suggests, there are numerous strategies for improving this methodology, however. Accurate datasets with high spatial resolution yield superior results in terms of properly allocating population across the landscape. Many recent studies compare the validity of different ancillary layers from land parcels (Maantay et al. 2007) to imperviousness (Wu et al. 2005), but studies of this spatial extent still require consistently produced datasets at the national scale.

General Descriptive Patterns and Bivariate Analysis

Before analyzing multivariate regression results, it was important to understand the underlying univariate distribution and bivariate pair-wise correlations between variables and how that changed after augmenting those variables through various interpolation techniques, buffer types, and scales. Appendices A and B below include summaries for descriptive statistics and bivariate correlations, respectively, for the variables used in this analysis. It is important here to note the strength and direction of the predictor variables in relation to collective efficacy in order to check whether they correspond to theoretical expectations acknowledged above. Additionally, pair-wise correlations between independent variables should be investigated as high correlations could yield problematic levels of multicollinearity within the specified regression model.

Variations in Descriptive Statistics

Tables A.1 and A.2 provide the descriptive statistics for the augmented census variable values. Overall, the new values correlate highly with the original census values (e.g., percent black population ranged from $r = .9348$ to $r = .9660$) which suggests that the interpolation techniques did in fact produce meaningful ratio values. Additionally, the values become more distinct (i.e., the correlation coefficient decreases) as rFA increases from 75% to 150% of the host block group area. Two main trends emerge from the twelve new data sets. First, across all variables (except for concentrated disadvantage), interpolation techniques, and buffer types, as rFA increases, the standard deviation (and the normalized standard deviation, or coefficient of variation) decreases in value. The larger rFA size encapsulates a greater geographic area and thus aggregates more pixel values together. This acts to reduce the effect of extreme values within the buffers. Concentrated disadvantage exhibits the opposite pattern. As rFA increases, so does the standard deviation. A possible explanation is offered in the next sections. Second, for percent black population and percent foreign-born population, as rFA increases in size, the mean consistently decreases in value. This can be explained by each variable's highly skewed distribution. As rFA increases, the mean shifts in the opposite direction of skew.

Variations in Bivariate Correlations

Analyzing bivariate correlations (Pearson's product-moment correlation coefficient, r) offer an initial step to understanding each variable's relationship to the dependent variable, community collective efficacy. Appendix B lists the 13 pair-wise correlation tables with significance levels. Table B.1 lists the correlations for the original block group values while B.2 through B.13 list coefficients for each of the interpolation techniques. Checking column one in each table helps check the expected direction and strength of the relationship between collective efficacy and the independent variables. The other columns and rows can help check for possible multicollinearity in the form of high correlation coefficients. Multicollinearity should not be an issue between variables since the highest coefficient is $r = 0.53$ between concentrated disadvantage and percent black population. This moderate correlation was expected, however, since previous studies included this race category as a component of the disadvantage index. Variance inflation factors do not suggest high multicollinearity, validating the exclusion of percent black population from the index.

All correlation coefficients have expected sign directions though multiple patterns emerge from the tables. The next highest correlation is a positive relationship between percent foreign-born population and settlement type indicating that immigrant populations would tend to reside in more urbanized areas. That relationship is probably more nuanced since this sample of modified Census values spans the United States. Foreign-born populations in certain parts of the country could represent various social and economic classes, from students to day laborers.

Other mild negative correlations include: foreign-born population and residential stability, residential stability and settlement type, and foreign-born and percent black populations. These relationships agree with the findings of social disorganization theorists Shaw and McKay (1969), who suggest that population groups (immigrants especially) are more transient in urban areas and that African-Americans are subject to residential segregation. However, these relationships, while statistically significant, hardly explain all conditions across the entire country with correlation coefficients ranging in absolute value of 0.19 to 0.53.

Lastly, special attention should be paid to the correlation coefficients between collective efficacy and the explanatory variables and their statistical significance across scale. Settlement type consistently exhibited a mild negative correlation with collective efficacy, as expected (r = -0.20). Again, settlement type is a control variable for each location and was not modified by either interpolation process. Percent black population maintained a weak negative correlation with collective efficacy. The strength of that relationship weakened as the rFA increased from 75% to 150% of host block group area. Additionally, the significance of the correlation decreased with rFA as well. Percent foreign-born population maintained a correlation coefficient of around -0.12 through all interpolation techniques, though the significance increased slightly with rFA. Residential stability maintained a correlation coefficient between 0.12 and 0.15 though its strength and significance decreased with rFA. Finally, concentrated disadvantage exhibited a weak negative relationship with collective efficacy, though both its strength and significance actually increased with rFA. This difference between concentrated disadvantage and the other explanatory variables, and increasing spread with rFA indicates that this variable might operate at a different geographic scale than the others. Whereas percent black population might have a greater influence upon perceived collective efficacy at closer distances (i.e., one's immediate neighbors), perhaps concentrated disadvantage acts at the neighborhood level or slightly larger. These relationships are parsed out further in the next section through multiple regression analysis.

Multiple Regression Analysis and Comparison of Interpolation Techniques

The next three sections present the results of the multiple regression analysis attempting to explain community collective efficacy through five measures of community context. Contextual measures were derived through three methods: 1) traditional Census block group values (BG), 2) interpolation through AW and 3) interpolation through IDM. Furthermore, the two interpolation techniques operationalized "neighborhood" two different ways: 1) circular Euclidean buffers and 2) road network-based service areas. Finally, additional "neighborhoods" were generated by scaling the buffer radius to 75% and 150% of the original host block group area creating three rFA for each location. A total of 13 different multiple regression models were generated (each with the same variable specification) and the results of those models are summarized in Table 4.2 and Appendix C and presented in the following sections.

Overall Model Performance

Figure 4.4 below summarizes the overall mode performance in terms of the coefficient of determination (R^2), the root mean square error, the global *F*-statistic, and the Akaike Information Criterion. The R^2 value is the ratio of the variance of community collective efficacy explained by the model compared to the total variance and ranges from 0% to 100%. The RMSE is the square root of the average squared model residuals and is measured in units of the dependent variable. Smaller values correspond to less model error and a better overall fit. The *F*-statistic is an indicator of whether or not any or all of the model parameters have any significant impact on the dependent variable. Lastly, the AIC is a measure of relative model fit which allows one to compare two similarly constructed regression models while controlling for complexity. Lower AIC values indicate a better performing regression model.

Examination of the four summary graphs in Figure 4.4 yields substantial findings and each measure agrees overall. First, and most obvious, all interpolation techniques yield better fitting regression models compared to BG values alone. IDM with network buffers at 100% rFA yields a 10% improvement over the BG method. All interpolation techniques results in lower RMSE and AIC values. *F*-statistics for all methods are statistically significant with all but one model more significant than the BG method. When compared across rFA, interpolation techniques generally performed better at 75% or 100% of the host block group area. IDM with network buffers consistently outperformed the other interpolation methods. However, the next best performing method depending upon the scale at which the analysis was performed. At the 75% rFA, IDM with network buffers was followed by AW with network buffers. Next was IDM with circular buffers and finally AW with circular buffers. These results continued into the 100% rFA. However, at the 150% rFA, IDM with network buffers was followed by IDM with circular buffers, then AW with circular buffers and lastly AW with network buffers. The relative performance was the same when measured by the coefficient of determination, the RMSE, and the AIC. When performing contextual analysis at or less than 100% of the initial areal units, network buffers outperform circular buffers. However, when performing contextual analysis at scales greater than 100%, it appears that IDM outperforms AW. Again, IDM with network buffers results in a better performing model regardless of the scale at which the social process is assumed to operate. Finally, from the graphs, it looks as though the curves for buffer type parallel one another. This suggests that at these three scales of analysis, buffer type is the main determining factor of model performance and interpolation technique plays a secondary role.

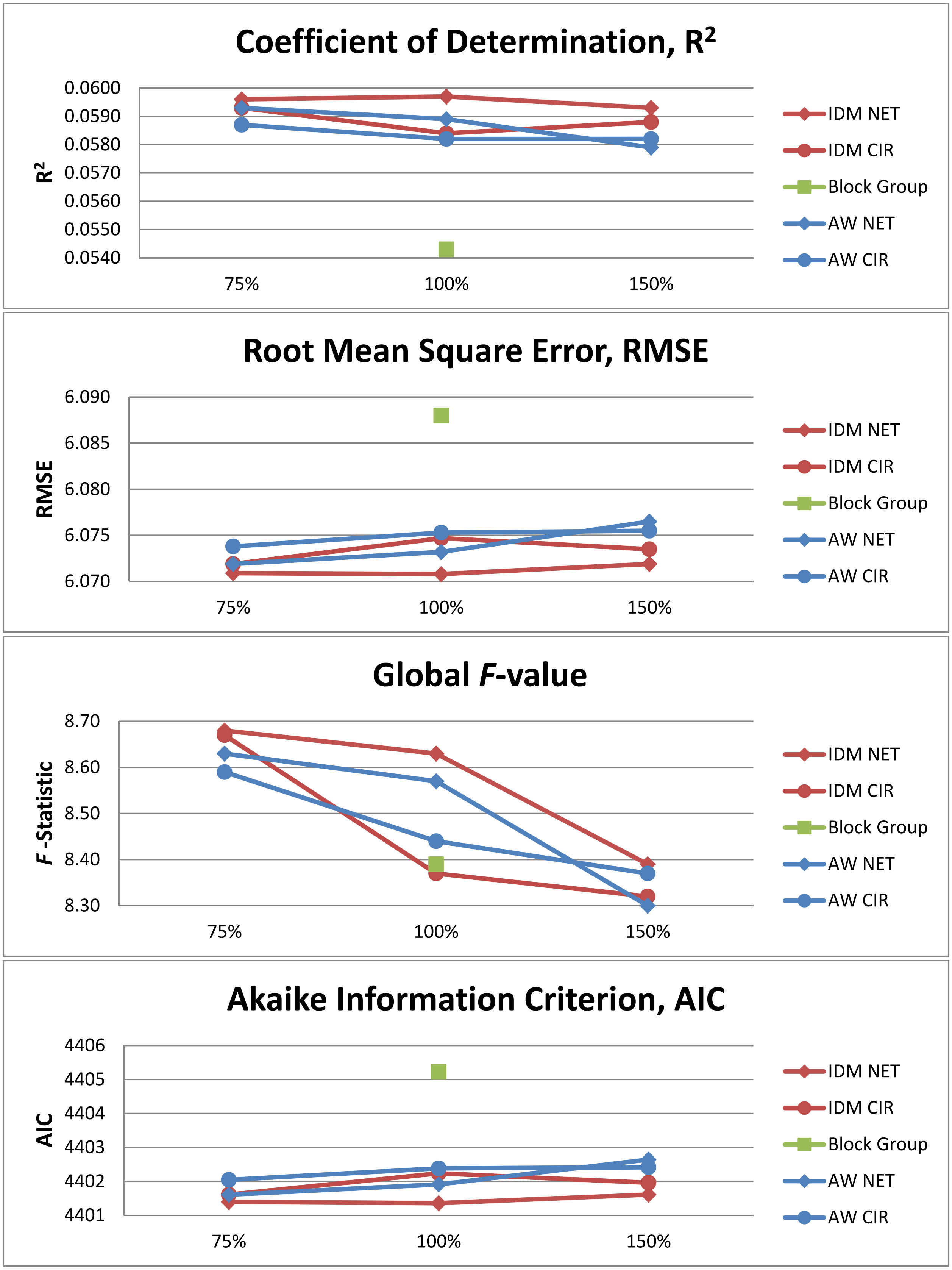

Figure 4.4: Model Fit Across Interpolation Technique, Buffer Type, and rFA

Variations in Partial Slope Coefficients

This section addresses changes in the partial slope coefficient for each explanatory variable seen in Figure 4.5. Combinations of interpolation methodology, buffer type, and rFA yields parameter estimates that impact the level of collective efficacy to a different extent. Four of the explanatory variables were statistically significant in some or all of the models and will be discussed in the next two sections. Percent foreign-born population did not significantly impact collective efficacy with this model specification or with these variable interpolation techniques. Settlement type was consistently the most influential variable across the models (i.e., it had the largest magnitude standardized coefficient). It should be noted that settlement type was a control variable and not subject to the value augmentation through AW or dasymetric interpolation. Compared to the BG method, interpolated methods yield larger slope coefficients (in absolute value). Also, as rFA increases, the effect of settlement type on collective efficacy increases, net of the other variables. Since settlement type was not augmented, its change in effect is attributed to changes in the other variables, which alter the multivariate covariance patterns between the independent variables. Concentrated disadvantage and percent black population exhibit the same pattern. Both have a greater partial effect on collective efficacy after interpolation and that effect increases with rFA (though the effect of percent black population seems to peak at the 100% rFA). However, the opposite is true for residential stability. Not only does its partial effect decrease with rFA, but at 100%, it has approximately the same effect as the BG method alone. It should also be noted that, again, the graphed curves of the different methods match for buffer types and not interpolation techniques. For example, the percent black population slope coefficient changes little between 75% and 100% but changes greatly between 100% and 150% for circular buffers. However, the values peak at 100% for network buffers.

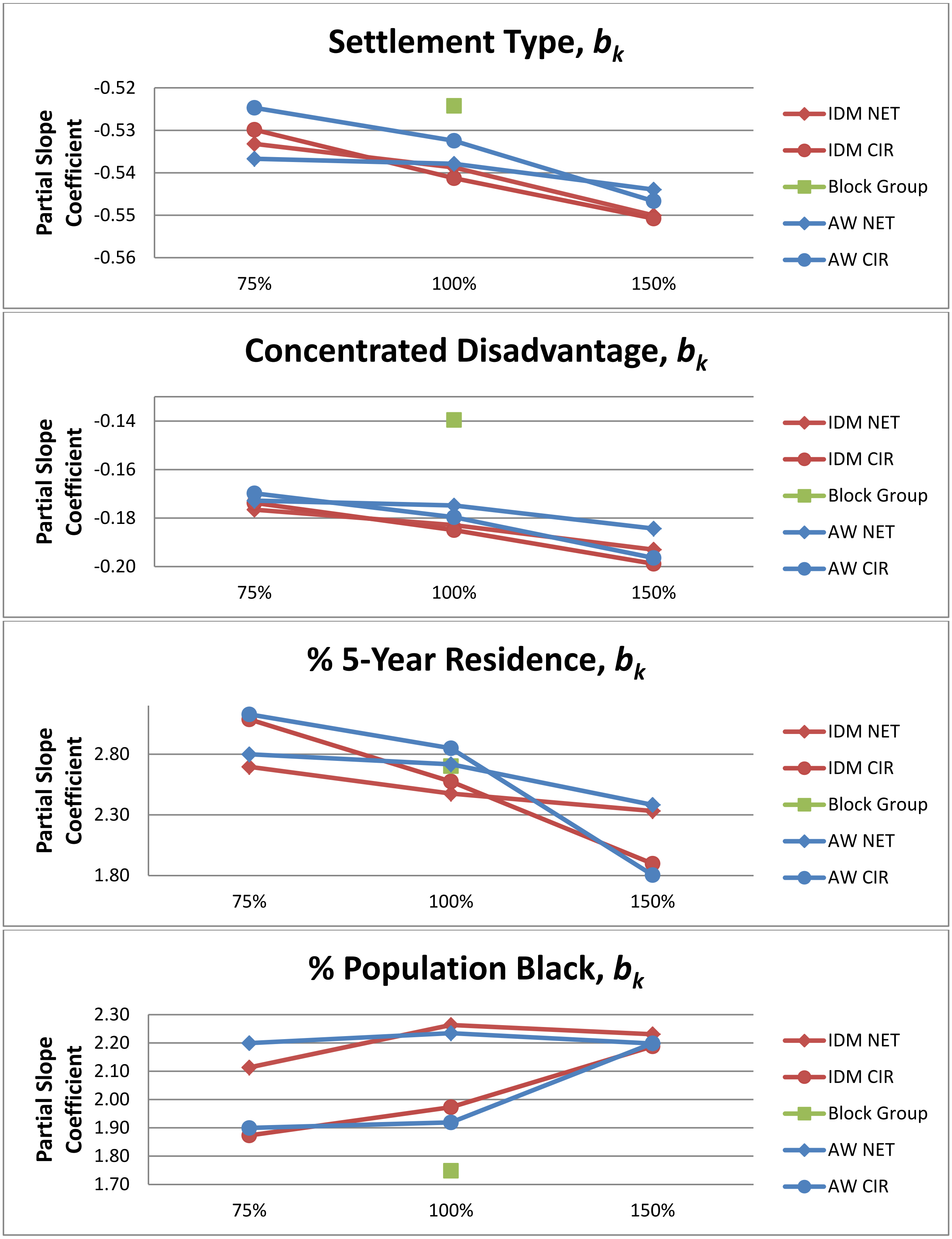

Figure 4.5: Model Coefficients Across Interpolation Technique, Buffer Type, and rFA

Variations in Coefficient Significance-Level

The last section compares the Student's *t*-value for each parameter estimate across interpolation technique, buffer type, and rFA. In inferential hypothesis testing, sample sizes fewer than 30 typically follow a *t*-distribution. That distribution approaches the normal, or *z*-distribution, as the sample size increases. *T*-values are used to test the null hypothesis that parameter estimates are statistically no different than zero, or that the explanatory variable has a null effect on the dependent variable. A *t*-value of 1.96 indicates that the parameter estimate is correct 95 times out of 100 when a random sample is taken. Figure 4.6 displays the *t*-value for each parameter estimate. Values greater than the absolute value of 1.96 indicate statistical significance at the 95% confidence level. Settlement type, again the control variable, is always significant regardless of the method employed. However, the interpolated models yield a parameter estimate with a lower level of significance. Again, since settlement type was not augmented its change must be attributed to changes in the other explanatory variables. The IDM with network buffers produces the highest level of significance for settlement type. Concentrated disadvantage is also statistically significant regardless of approach used. Unlike settlement type, disadvantage is more significant when interpolation methods are employed rather than the BG method. Also, each of the interpolation techniques increases statistical significance of this variable as rFA increases. Therefore, at a larger rFA, one can be more confident in the partial slope coefficient estimate, or more confident in the effect concentrated disadvantage has on collective efficacy. Circular buffers produce more confident estimates compared to network buffers for this variable at these rFA. Residential stability does not reach statistical significance at the 95% confidence level with any technique, but it does approach the 90% confidence level with smaller rFA. However, there is little change from the BG method.

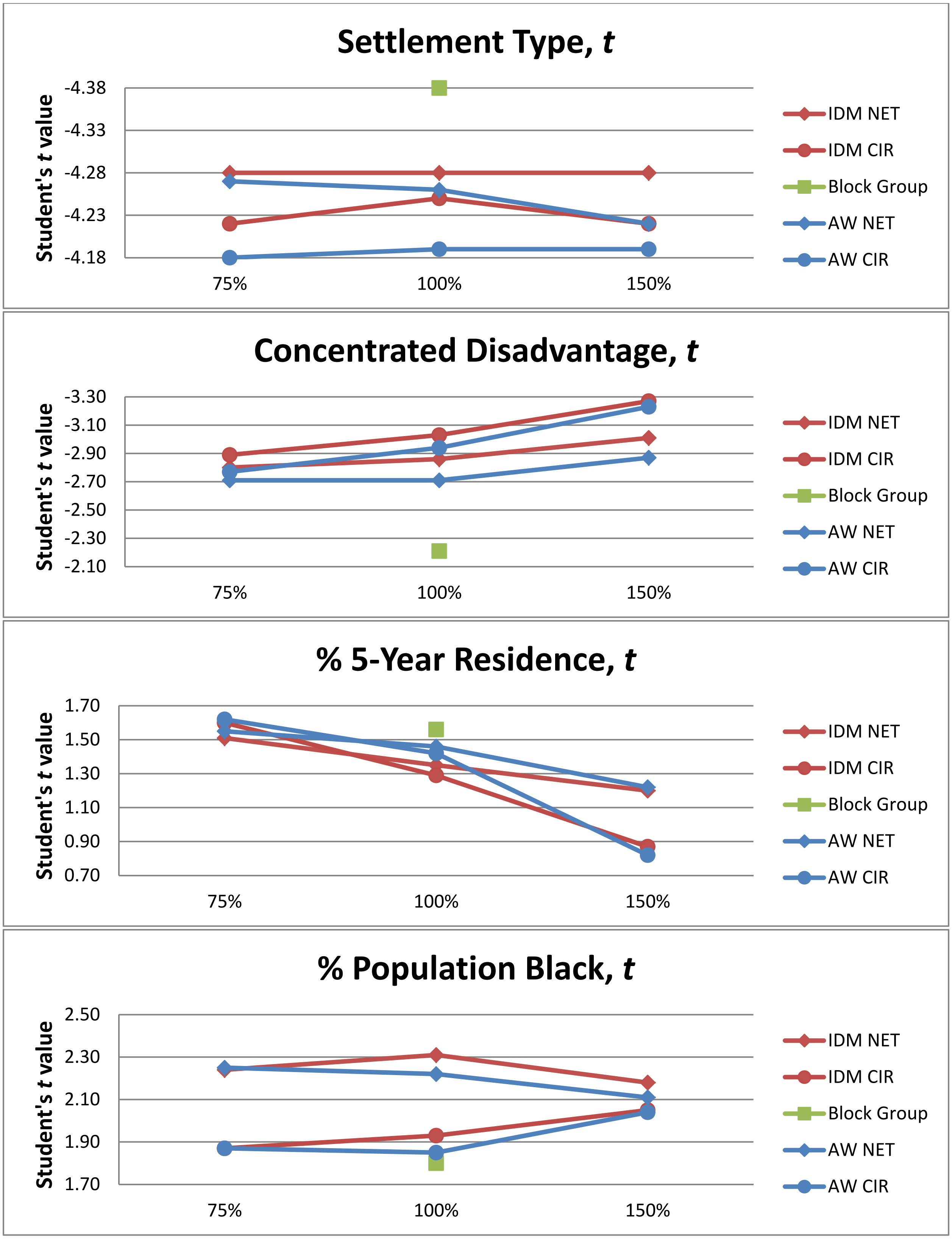

Figure 4.6: Parameter Significance Across Interpolation Technique, Buffer Type, and rFA

Finally, the percent black population variable exhibits greater estimate significance for the interpolation techniques compared to the BG method. The circular buffer methods are just below the 95% confidence level until the 150% rFA while the network buffer methods are more significant than the circular buffer methods. Also, the network buffer methods are more significant at the 75% and 100% rFA and approach the significance level of the circular buffer methods at the 150% rFA. It should be noted that the shape of the significance curves match very well those of the partial slope coefficient. Contrary to previous studies regarding community collective efficacy, this research found that percent black population had a significant positive impact on collective efficacy. This result is not surprising, however, since the perspective of the FACHS data is that of African-American youth. Even still, this suggests that a further exploration on the effect of race, racial identity, and racial segregation is needed in the context of community collective efficacy. It should be noted that the previous graphs were scaled to highlight the relative differences in methods, rather than to misled readers.

Table 4.2: OLS Regression Results of 13 Methods of Context

Method	Settlement Type	Con. Disad.	% 5-Year Residence	% Pop Black	%Foreign -born	Constant	R^2
Block Group	-0.5242*	-0.1395*	2.7024	1.7484	0.1571	30.7267*	0.0543*
AW CIR 75	-0.5247*	-0.1698*	3.1292	1.8995	0.0200	30.5010*	0.0587*
AW CIR 100	-0.5324*	-0.1796*	2.8497	1.9191	0.1028	30.6713*	0.0582*
AW CIR 150	-0.5467*	-0.1964*	1.8039	2.1988*	-0.5109	31.2129*	0.0582*
AW NET 75	-0.5367*	-0.1728*	2.8001	2.1994*	0.1495	30.5869*	0.0593*
AW NET 100	-0.5379*	-0.1748*	2.7173	2.2347*	0.1876	30.6286*	0.0589*
AW NET 150	-0.5439*	-0.1843*	2.3826	2.1980*	0.0566	30.8600*	0.0579*
IDM CIR 75	-0.5298*	-0.1737*	3.0877	1.8732	-0.0051	30.5532*	0.0593*
IDM CIR 100	-0.5413*	-0.1849*	2.5754	1.9730	0.0542	30.8261*	0.0584*
IDM CIR 150	-0.5508*	-0.1988*	1.8977	2.1880*	-0.3334	31.1759*	0.0588*
IDM NET 75	-0.5332*	-0.1765*	2.6966	2.1134*	0.1794	30.6269*	0.0596*
IDM NET 100	-0.5387*	-0.1829*	2.4759	2.2636*	0.1509	30.7176*	0.0597*
IDM NET 150	-0.5500*	-0.1930*	2.3323	2.2308*	0.3425	30.8620*	0.0593*

Note: n=682; * $p < 0.05$

Summary of Results

This chapter covered a range of topics summarizing an effort to contribute methodologically to neighborhood effect research by first redefining the unit in which context is derived and then by interpolating the social and economic information collected by the U.S. Census Bureau. In the framework of community collective efficacy, buffered "neighborhoods" and dasymetric modeling offer an improved approach to conceptualizing community context. Dasymetric interpolation through the IDM approach and the NLCD as ancillary data provide a way to augment Census variables across the United States at the sub-block group scale. This combination illustrates both a micro (i.e., block group) and macro (i.e., nationwide) approach to understanding community collective efficacy. Dasymetric modeling produced valid results that conformed to standards set by previous studies given the challenge of working at both the micro and macro-scale. Suggestions for improving this approach in relation to deriving contextual measures will be addressed in the Conclusions chapter.

Statistical analysis of the resulting modified Census variables validated the dasymetric approach. Multiple regression analysis allowed for comparison of 13 methods for generating community context. IDM interpolation with network buffers improved overall model fit compared to traditional Census unit context and areal weighting interpolation and "neighborhoods" as circular buffers. Depending on the scale of analysis, either the network-based buffer method, or the IDM method yielded the highest improvement. Relative focal areas of 75% to 100% of the host block group area support IDM as the dominating factor for improved model fit. On the other hand, at rFA greater than 100% of the host block group, network-based buffers were the dominating factor for improved model fit. The selection and combination of these factors must be taken into consideration when performing regression analysis with community context.

Comparing parameter estimates and their corresponding statistical significance yielded varying results. Both parameter estimates and significance levels varied across interpolation technique, buffer type, and rFA, but not all variables changed concurrently. When greater geographic context was taken into consideration, settlement type (the control variable) became less influential and less significant to the model overall. The decrease in parameter effect was at the increase of influence from other variables. Both concentrated disadvantage and percent black population increased in parameter *and* statistical significance, but in different ways. Disadvantage appears to have a greater, and more significant, effect at rFA larger than 100% while percent black population seems best suited at the same scale as the host block groups while residential stability is more significant and more influential at smaller rFA. These results indicate that the processes measured by these variables operate socially at various spatial scales. Residential stability influences perceptions of community collective efficacy at a more restrictive spatial scale than racial dominance. Segregated black populations influence perceptions of community collective efficacy at a more restrictive spatial scale than concentrated disadvantage. One can infer that if you do not recognize or associate with your immediate neighbors then your perception of your "community" would be held in lower regards. At a larger rFA, perhaps, the scale at which you perceive to be your "neighborhood," if you tend to see fellow residents with whom you identify more closely racially then your perception of "community" might improve. Finally, if you are limited to an area of town (beyond the boundaries of your "neighborhood") in which economic disadvantage is concentrated then you might perceive lower collective efficacy.

These results could be viewed as providing partial empirical evidence for Suttles' (1972) scales of neighborhood. He describes the four scales of neighborhood at which households interact socially, including: "block face," "community of limited liability," "expanded community of limited liability," and "sector of a city." Residential stability, in terms of the FACHS dataset, could affect collective efficacy at the block face scale – the scale at which you are most likely to interact with your immediate neighbors. A rapid turnover of neighbors could prevent the formation of social bonds necessary to feel included in one's own neighborhood. Similarly, residential segregation could affect collective efficacy at the scale of community of limited liability or beyond while economic concentrated disadvantage could affect collective efficacy at the scale of a unique sector of a city.

CHAPTER 5

CONCLUSIONS

Dasymetric Modeling and Community Context

Regardless of the subject, contextual analysis often requires information from zones in which it was not collected. Used initially as a cartographic technique for improving the display of population densities, dasymetric modeling has the ability to interpolate a variety of count values. The methodological contribution of dasymetric modeling to contextual analysis provides social scientists an opportunity to better estimate social processes thought to operate in various spatial extents. Furthermore, dasymetric modeling allows for the infinite aggregation and zonation of small, homogenous regions of data into meaningful, theoretically supported spatial units in an attempt to counteract arbitrarily drawn administrative boundaries and the modifiable areal unit problem. This thesis compared "Intelligent Dasymetric Mapping" and other methods of measuring community context in order improve how community collective efficacy is modeled. Specifically, I used data from FACHS to address issues regarding the performance of IDM in creating improved contextual measures, how the operationalization of "neighborhoods" effecting this type of analysis (i.e., the extent to which the MAUP is present in this dataset), and how certain neighborhood contexts positively or negatively impacted the formation of community collective efficacy. My research found that, indeed, IDM does generate variable values better suited for modeling community collective efficacy. Multiple regression analysis suggests that overall, creation of contextually-sensitive variables via IDM improves model fit over other methods of community context including traditional block group measures and areal weighted interpolation. There were differences in model performance based on the scale at which the model was analyzed. IDM appears better suited for modeling at geographic scales below the initial population unit (e.g., Census block groups) whereas network-based buffers seem better suited for scales at or above the initial population unit – regardless of the interpolation technique. However, IDM almost always performed better than AW.

The use of more contextually-aware values in regression analysis shifted how influentially and significantly explanatory variables impacted community collective efficacy. Concentrated disadvantage and percent black population increased in relative important over the traditional block group (i.e., contextually-unaware) method. However, these changes depended upon the scale of the analysis. This led to the conclusion that these variables measure social processes that operate at a variety of spatial extents. This finding, while not surprising, and supported by the literature in fact, was unexpected from such simplistic regression models. IDM can therefore aid in the creation of more meaningful "sliding neighborhoods" that may adjust according to the social process that is under investigation.

Methodological and Theoretical Limitations

Addressed in this thesis are a number of methodological and theoretical limitations that must be taken into consideration. From a methodological standpoint, this project methodology (i.e., IDM) was reliant upon a computationally-intense process for completing the dasymetric maps and an ancillary layer that needs to be as accurate as possible. Since the FACHS dataset has a spatial extent as large as the continental United States, an appropriate ancillary layer was required. Additionally, the study area needed to be subdivided into 45 block group clusters. The whole methodology was scripted to run on separate computers in a single lab. If, however, the study area were focused on one metropolitan area, and the research specified the exact interpolation technique, buffer type, and rFA, then the process would take considerably less time.

As previously noted, the selection of appropriate ancillary data is an arduous task requiring consistently across the study area, accuracy in the assume population distribution, and the appropriate scale for measure the target zones. The NLCD, with 30 meter pixel resolution, did create target zones much smaller than typical block groups. However, this dataset only contained information on the land *cover*, not what the land was being used for. Problems arise in classification when heavily urbanized areas are assumed to be densely populated even though that particular region is used for commercial or industrial purposes. Finally, to aid in IDM accuracy, certain block groups should be restricted from the sampling procedure within IDM or risk biases population density measures with extreme values. Such extreme values in terms of population would include block groups that contain institutionalized persons, dormitories, or military barracks. These geographically restricted individuals are clustered in large numbers and are not representative of populations in typical block groups, but still need to be accounted for.

From a theoretical standpoint, this thesis relies on a number of assumptions that may or may not affect the results. First, violating the theoretical assumptions regarding Ordinary Least Square regression could alter not only the parameter estimates, but also the confidence on which those estimates are evaluated. For example, the models specified in this thesis were relatively simple in order to test the effects of interpolation techniques, buffer types, and scale. While variable selection was guided by previous studies in collective efficacy, variable misspecification can lead to biased parameter estimates and unreliable significance values. Additionally, if the IDM process produces a pattern of population count errors that is reflected in the residual of the model, heteroskedasticity could also render significance values unreliable (Hamilton 1992).

Another theoretical assumption in this analysis is that of “neighborhood” creation. Galster (2008) recently suggested that determining the appropriate scale at which to define “neighborhood” was the first of six major challenges facing neighborhood effects analysis. This research allowed a neighborhood to be operationalized as either a circular buffer or network service area. The buffer width or network band size was determined by the block group in which the FACHS respondent lived. Larger block groups produced larger buffer sizes. This allowed the neighborhoods to vary in area across each observation. It was assumed that these buffers reflect how the FACHS respondent conceptualized his or her own “neighborhood” and how that conceptualization frames his or her responses to the FACHS questionnaire. Perhaps a better measure of “neighborhood” through propinquity would be the amount of time spent commuting to work or school each day, or whether or not the individual relied on a personal automobile or public transit as the dominant mode of transportation. Still, the use of road-network buffer-based “neighborhoods” represents a theoretical step in the right direction away from the static definition of community imposed by the Census Bureau since individuals typically must conform to the transportation network while either commuting or visiting friends. Further theoretical and empirical support comes from Grannis (1998) and Guo and Bhat (2007).

Recommendations for Future Research and Final Thoughts

Understanding how location affects individuals is the central talk of neighborhood effects research. Quantitative analysis of these effects relies on measurement of geographic or community context. Areal interpolation, specifically IDM provides social scientists the opportunity to redraw the boundaries in which readily available data already exists to conform to a variety of studies. In terms of community collective efficacy, “neighborhoods” based on proximity to road networks offer a suitable surrogate for the unknown boundaries conforming to individual perceptions of community. Due to this ambiguity, the notion of scale as defined by the relative focal area requires further exploration. Setting the scale of analysis to anything other than 100% should be based on theoretical expectations of how social processes operate across space. However, the findings of this thesis are supported by previous studies (Buck 2001, Bolster et al. 2007) which also found that regression models predicting the relationship between the neighborhood effects and various outcomes was sometimes better or the coefficients estimated a larger effect (i.e., regression models had a larger coefficient of variation or parameter coefficients were larger) when using smaller geographic areas to operationalize the context space. Galster (2008), drawing from previous research, agrees that “neighborhood” might require definition at multiple scales simultaneously, depending on the processes at work. This and the findings of this thesis suggest that once a theoretical or empirical-based scale of operation is set, a relative focal area could be defined separately for each variable of the neighborhood effects analysis. This multi-scalar neighborhood effects scenario should provide a more theoretically (and empirically) meaningful operationalization of “neighborhood” and a more appropriate definition of how various social processes (e.g., segregation, concentrated disadvantage, etc.) interact with such an area. Areal interpolation could then provide the contextual values at whatever rFA specified. Obviously, how one defines and operationalizes “neighborhoods” has an impact on neighborhood effects analysis. The use of pre-defined Census brought uniformity to previous studies in the field. However, that uniformity came as the expense of theoretical appropriateness. Other than utilizing the approach outlined here (i.e., “ego-centric neighborhoods”), future analysis could examine other uniform, or “territorial” conceptualizations of neighborhood. For instance, planning departments often subdivide cities into neighborhoods associated with some historic significance. These politically established boundaries might provide a more useful operationalization of neighborhood than census boundaries since a more local state entity, supposedly with more intimate knowledge of the community, delineated them. Though this method is still subject to the question of whether individuals who reside in those pre-defined neighborhoods actually associate and interact socially with those same boundaries. Finally, the IDM methodology should be tested in other arenas of research. Public health or environmental injustice are other fields in which individual outcomes can be a direct results of environmental (i.e., contextual) factors and in which this methodology could provide substantial measurement improvements.

In terms of community collective efficacy, this thesis illustrated the importance of utilizing increasingly contextually-sensitive variables that can improve the overall explanatory power of existing models and provide a more precise measure of the direct net effect of each contextual variable. The next step requires careful consideration of how each of those explanatory variables, and others not addressed in this thesis operate across space. Specifying the correct model not only requires the appropriate variable specification, but also the appropriate scale at which each variable must be understood. Once determined, these improved models should help social science researchers assist policy-makers in identifying areas once considered "broken" and stimulate positive change through community building rather expanding zero-tolerance policy.

REFERENCES

Anderson, J. R., E. E. Hardy, J. T. Roach & R. E. Witmer. 1976. A Land Use and Land Cover Classification System For Use with Remote Sensor Data. In *USGS Professional Paper 964. A revision of the land use classification system as presented in the USGS Circular 671*. Washington: United States Geologic Survey.

Blalock, H. M. (1984) Contextual-Effects Models: Theoretical and Methodological Issues. *Annual Review of Sociology,* 10**,** 353-372.

Bolster, A., S. Burgess, R. Johnston, K. Jones, C. Propper & R. Sarker (2007) Neighbourhoods, households and income dynamics: a semi-parametric investigation of neighbourhood effects. *Journal of Economic Geography,* 7**,** 1-38.

Boone, C. G. (2008) Improving resolution of census data in metropolitan areas using a dasymetric approach: applications for the Baltimore Ecosystem Study. *Cities and the Environment,* 1**,** 25.

Brody, G. H., X. Ge, R. Conger, F. X. Gibbons, V. M. Murry, M. Gerrard & R. L. Simons (2001) The Influence of Neighborhood Disadvantage, Collective Socialization, and Parenting on African American Children's Affiliation with Deviant Peers. *Child Development,* 72**,** 1231-1246.

Buck, N. (2001) Identifying neighbourhood effects on social exclusion. *Urban Studies,* 38**,** 2251-2275.

Chaix, B., J. Merlo, D. Evans, C. Leal & S. Havard (2009) Neighbourhoods in eco-epidemiologic research: Delimiting personal exposure areas. A response to Riva, Gauvin, Apparicio and Brodeur. *Social Science & Medicine,* 69**,** 1306-1310.

Comstock, N., L. M. Dickinson, J. A. Marshall, M. J. Soobader, M. S. Turbin, M. Buchenau & J. S. Litt (2010) Neighborhood attachment and its correlates Exploring neighborhood conditions, collective efficacy, and gardening. *Journal of Environmental Psychology,* 30**,** 435-442.

Cummins, S. (2007) Commentary: Investigating neighbourhood effects on health - avoiding the 'Local Trap'. *International Journal of Epidemiology,* 36**,** 355-357.

Davidson, R. & J. G. MacKinnon. 1993. *Estimation and inference in econometrics*. Oxford University Press.

Duncan, T. E., S. C. Duncan, H. Okut, L. A. Strycker & H. Hix-Small (2003) A multilevel contextual model of neighborhood collective efficacy. *American Journal of Community Psychology,* 32**,** 245-252.

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

Flowerdew, R. & M. Green. 1991. Data Integration: Statistical Methods for Transferring Data Between Zonal Systems. In *Handling Geographical Information,* eds. I. Masser & M. Blakemore. Essex: Longman Scientific & Technical.

Flowerdew, R., D. J. Manley & C. E. Sabel (2008) Neighbourhood effects on health: Does it matter where you draw the boundaries? *Social Science & Medicine,* 66**,** 1241-1255.

Galster, G. (2008) Quantifying the Effect of Neighborhood on Individuals: Challenges, Alternative Approaches, and Promising Directions. *Journal of Applied Social Science Studies,* 128.

Grannis, R. (1998) The importance of trivial streets: Residential streets and residential segregation. *American Journal of Sociology,* 103**,** 1530-1564.

Green, M. & R. Flowerdew. 1996. New Evidence on the Modifiable Areal Unit Problem. In *Spatial Analysis: Modelling in a GIS Environment,* eds. P. Longley & M. Batty. Cambridge: GeoInformation International.

Guo, J. Y. & C. R. Bhat (2007) Operationalizing the concept of neighborhood: Application to residential location choice analysis. *Journal of Transport Geography,* 15**,** 31-45.

Hamilton, L. C. 1992. *Regression with Graphics: A Second Course in Applied Statistics*. Belmont: Duxbury Press.

Holloway, S. R., D. Bryan, R. Chabot, D. M. Rogers & J. Rulli (1998) Exploring the effect of public housing on the concentration of poverty in Columbus, Ohio. *Urban Affairs Review,* 33**,** 767-789.

Kelling, G. L. & C. M. Coles. 1996. *Fixing broken windows: restoring order and reducing crime in our communities*. New York: Martin Kessler Books.

Kim, H. & X. B. Yao (2010) Pycnophylactic interpolation revisited: integration with the dasymetric-mapping method. *International Journal of Remote Sensing,* 31**,** 5657-5671.

Lebel, A., R. Pampalon & P. Y. Villeneuve (2007) A multi-perspective approach for defining neighbourhood units in the context of a study on health inequalities in the Quebec City region. *International Journal of Health Geographics,* 6**,** 15.

Maantay, J. (2002) Mapping Environmental Injustices: Pitfalls and Potential of Geographic Information Systems in Assessing Environmental Health and Equity. *Environmental Health Perspectives,* 110**,** 161-171.

Maantay, J. (2007) Asthma and air pollution in the Bronx: Methodological and data considerations in using GIS for environmental justice and health research. *Health & Place,* 13**,** 32-56.

Maantay, J. & A. Maroko (2009) Mapping urban risk: Flood hazards, race, & environmental justice in New York. *Applied Geography,* 29**,** 111-124.

Maantay, J. A., A. R. Maroko & C. Herrmann (2007) Mapping Population Distribution in the Urban Environment: The Cadastral-based Expert Dasymetric System (CEDS). *Cartography and Geographic Information Science,* 34**,** 77-102.

Massey, D. S. (1990) AMERICAN APARTHEID - SEGREGATION AND THE MAKING OF THE UNDERCLASS. *American Journal of Sociology,* 96**,** 329-357.

Mayer, S. E. & C. Jencks (1989) GROWING UP IN POOR NEIGHBORHOODS - HOW MUCH DOES IT MATTER. *Science,* 243**,** 1441-1445.

McNulty, T. L. & S. R. Holloway (2000) Race, crime, and public housing in Atlanta: Testing a conditional effect hypothesis. *Social Forces,* 79**,** 707-729.

Mennis, J. (2002) Using geographic information systems to create and analyze statistical surfaces of population and risk for environmental justice analysis. *Social Science Quarterly,* 83**,** 281-297.

--- (2006) Mapping the results of geographically weighted regression. *Cartographic Journal,* 43**,** 171-179.

Mennis, J. & T. Hultgren (2006) Intelligent Dasymetric Mapping and Its Application to Areal Interpolation. *Cartography and Geographic Information Science,* 33**,** 179-194.

Morenoff, J. D., R. J. Sampson & S. W. Raudenbush (2001) Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology,* 39**,** 517-559.

Natsuaki, M. N., X. Ge, G. H. Brody, R. L. Simons, F. X. Gibbons & C. E. Cutrona (2007) African American Children's Depresive Symptoms: The Prospective Effects of Neighborhood Disorder, Stressful Life Events, and Parenting. *American Journal of Psychology,* 39**,** 163-176.

Openshaw, S. 1983. *The Modifiable Areal Unit Problem*. Norwich: Geo Books.

Park, R. E. (1915) The City: Suggestions for the Investigation of Human Behavior in the City Environment. *The American Journal of Sociology,* 20**,** 577-612.

Park, R. E., E. W. Burgess & R. D. McKenzie. 1925. *The City*. University of Chicago Press.

Reibel, M. & M. E. Bufalino (2005) Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environment and Planning A,* 37**,** 127-139.

Sampson, R. J., J. D. Morenoff & F. Earls (1999) Beyond social capital: Spatial dynamics of collective efficacy for children. *American Sociological Review,* 64**,** 633-660.

Sampson, R. J., J. D. Morenoff & T. Gannon-Rowley (2002) Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology,* 28**,** 443-478.

Sampson, R. J. & S. W. Raudenbush (2004) Seeing disorder: Neighborhood stigma and the social construction of "Broken windows". *Social Psychology Quarterly,* 67**,** 319-342.

Sampson, R. J., S. W. Raudenbush & F. Earls (1997) Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science,* 277**,** 918-924.

Shaw, C. R. & H. D. McKay. 1969. *Juvenile delinquency and urban areas; a study of rates of delinquency in relation to differential characteristics of local communities in American cities*. Chicago: University of Chicago Press.

Simmel, G. 1903/1971. The Metropolis and Mental Life. In *On Individuality and Social Forms,* ed. D. N. Levine. Chicago & London: University of Chicago Press.

Simons, R. L., K.-H. Lin, L. C. Gordon, G. H. Brody, V. Murry & R. D. Conger (2002) Community Differences in the Associations Between Parenting Practices and Child Conduct Problems. *Journal of Marriage and Family,* 64**,** 331-345.

Simons, R. L., L. G. Simons, C. H. Burt, G. H. Brody & C. Cutrona (2005) Collective efficacy, authoritative parenting and delinquency: a longitudinal test of a model integrating community- and family-level processes. *Criminology,* 43**,** 989-1029.

Spielman, S. E. & E. H. Yoo (2009) The spatial dimensions of neighborhood effects. *Social Science & Medicine,* 68**,** 1098-1105.

St. Jean, P. K. B. 2007. *Pockets of crime: broken windows, collective efficacy, and the criminal point of view*. Chicago: University of Chicago Press.

Stewart, E. A., R. L. Simons & R. D. Conger (2002) Assessing Neighborhood And Social Psychological Influences on Childhood Violence in an African-American Sample. *Criminology,* 40**,** 801-829.

Suttles, G. D. 1972. *Social Construction of Communities*. Chicago: University of Chicago Press.

Tobler, W. R. (1970) COMPUTER MOVIE SIMULATING URBAN GROWTH IN DETROIT REGION. *Economic Geography,* 46**,** 234-240.

--- (1979) SMOOTH PYCNOPHYLACTIC INTERPOLATION FOR GEOGRAPHICAL REGIONS. *Journal of the American Statistical Association,* 74**,** 519-530.

U.S. Census Bureau. 2000. 2000 Census of Population and Housing: Summary File 3. Washington: The Bureau.

Weiss, L., D. Ompad, S. Galea & D. Vlahov (2007) Defining neighborhood boundaries for urban health research. *American Journal of Preventive Medicine,* 32**,** S154-S159.

Wilson, J. Q. & G. L. Kelling. 1982. The Police and Neighborhood Safety: Broken Windows. In *The Atlantic Monthly*, 29-38.

Wilson, W. J. 1987. *The truly disadvantaged: the inner city, the underclass, and public policy*. Chicago: University of Chicago Press.

Wong, D. 2009. The Modifiable Areal Unit Problem (MAUP). In *The SAGE Handbook of Spatial Analysis,* eds. A. S. Fotheringham & P. A. Rogerson. London: SAGE Publications Ltd.

Wong, D. W. S. & J. Lee. 2005. *Statistical Analysis of Geographic Information with ArcView GIS and ArcGIS*. Hoboken: John Wiley & Sons, Inc.

Wright, J. K. (1936) A Method of Mapping Densities of Population: With Cape Code as an Example. *Geographical Review,* 26**,** 103-110.

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

Zandbergen, P. A. & D. A. Ignizio (2010) Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates. *Cartography and Geographic Information Science,* 37**,** 199-214.

APPENDIX A: DESCRIPTIVE STATISTICS

Table A.1: Descriptive Statistics for Individual, Block Group, and IDM Variables.

			Variable	Mean	Std. Dev.	C.V.	Min	Max
Individual-Level Responses			Collective Efficacy	30.3031	6.2439		2	42
			Gender (Dummy)	0.5622	0.4965		0	1
			Age (Years)	18.8185	0.9057		16	21
			Income	$38,270.51	$28,212.98		$2,689.54	$225,000
			Educational Level	12.1362	2.0436		1	16
			Settlement Type	4.6413	2.2325		0	8
Block Group			% Black Pop	0.3812	0.3151	0.8267	0.0000	1.0000
			% Foreign-born Pop	0.0773	0.0941	1.2184	0.0000	0.6274
			% 5 Year Residence	0.4878	0.1593	0.3265	0.0000	0.8514
			Disadvantage	0.0124	4.5719		-9.9581	11.8431
Intelligent Dasymetric Mapping	Circular Buffers	75%	% Black Pop	0.3723	0.2974	0.7988	0.0000	1.0000
			% Foreign-born Pop	0.0766	0.0833	1.0869	0.0000	0.5971
			% 5 Year Residence	0.4857	0.1396	0.2874	0.0518	0.8250
			Disadvantage	4.98E-09	4.7812		-10.4431	13.0720
		100%	% Black Pop	0.3646	0.2891	0.7929	0.0000	1.0000
			% Foreign-born Pop	0.0768	0.0811	1.0558	0.0000	0.5713
			% 5 Year Residence	0.4866	0.1332	0.2736	0.0605	0.7644
			Disadvantage	-5.29E-09	4.8180		-10.7166	15.2113
		150%	% Black Pop	0.3515	0.2765	0.7866	0.0000	1.0000
			% Foreign-born Pop	0.0756	0.0757	1.0021	0.0001	0.6024
			% 5 Year Residence	0.4884	0.1228	0.2514	0.0679	0.7497
			Disadvantage	-5.09E-09	4.9101		-10.8701	16.9524
	Network Buffers	75%	% Black Pop	0.3861	0.3151	0.8161	0.0000	1.0000
			% Foreign-born Pop	0.0761	0.0848	1.1138	0.0000	0.5706
			% 5 Year Residence	0.4854	0.1454	0.2996	0.0416	0.8835
			Disadvantage	4.46E-09	4.7435		-10.3569	16.0634
		100%	% Black Pop	0.3801	0.3057	0.8043	0.0000	1.0000
			% Foreign-born Pop	0.0761	0.0824	1.0831	0.0000	0.5494
			% 5 Year Residence	0.4853	0.1408	0.2901	0.0554	0.8588
			Disadvantage	-9.71E-10	4.7860		-10.7945	18.1682
		150%	% Black Pop	0.3678	0.2906	0.7900	0.0000	1.0000
			% Foreign-born Pop	0.0760	0.0791	1.0402	0.0000	0.5656
			% 5 Year Residence	0.4865	0.1317	0.2707	0.0648	0.7542
			Disadvantage	4.03E-09	4.8399		-11.5382	20.4775

Table A.2: Descriptive Statistics for Individual, Block Group, and AW Variables.

			Variable	Mean	Std. Dev.	C.V.	Min	Max
Individual-Level Responses			Collective Efficacy	30.3031	6.2439		2	42
			Gender (Dummy)	0.5622	0.4965		0	1
			Age (Years)	18.8185	0.9057		16	21
			Income	$38,270.51	$28,212.98		$2,689.54	$225,000
			Educational Level	12.1362	2.0436		1	16
			Settlement Type	4.6413	2.2325		0	8
Block Group			% Black Pop	0.3812	0.3151	0.8267	0.0000	1.0000
			% Foreign-born Pop	0.0773	0.0941	1.2184	0.0000	0.6274
			% 5 Year Residence	0.4878	0.1593	0.3265	0.0000	0.8514
			Disadvantage	0.0124	4.5719		-9.9581	11.8431
Areal Weighting Interpolation	Circular Buffers	75%	% Black Pop	0.3674	0.2919	0.7946	0.0000	0.9948
			% Foreign-born Pop	0.0768	0.0854	1.1119	0.0000	0.6970
			% 5 Year Residence	0.4875	0.1391	0.2853	0.0498	0.7988
			Disadvantage	5.09E-09	4.7638		-10.7708	13.4126
		100%	% Black Pop	0.3612	0.2854	0.7902	0.0000	0.9917
			% Foreign-born Pop	0.0767	0.0828	1.0804	0.0000	0.6954
			% 5 Year Residence	0.4875	0.1328	0.2724	0.0584	0.7580
			Disadvantage	-1.43E-09	4.8112		-10.8816	14.8163
		150%	% Black Pop	0.3505	0.2753	0.7856	0.0000	0.9853
			% Foreign-born Pop	0.0754	0.0773	1.0246	0.0001	0.6896
			% 5 Year Residence	0.4891	0.1226	0.2506	0.0697	0.7588
			Disadvantage	-5.60E-09	4.9142		-10.8645	16.4677
	Network Buffers	75%	% Black Pop	0.3721	0.3006	0.8079	0.0000	0.9968
			% Foreign-born Pop	0.0765	0.0871	1.1395	0.0000	0.6653
			% 5 Year Residence	0.4874	0.1443	0.2961	0.0414	0.8514
			Disadvantage	-1.30E-09	4.7344		-10.7521	16.6903
		100%	% Black Pop	0.3670	0.2937	0.8001	0.0000	0.9962
			% Foreign-born Pop	0.0763	0.0851	1.1156	0.0000	0.6633
			% 5 Year Residence	0.4872	0.1399	0.2871	0.0534	0.8333
			Disadvantage	5.45E-12	4.7621		-10.9651	19.0937
		150%	% Black Pop	0.3586	0.2841	0.7922	0.0000	0.9921
			% Foreign-born Pop	0.0758	0.0810	1.0687	0.0000	0.6494
			% 5 Year Residence	0.4879	0.1314	0.2694	0.0671	0.7591
			Disadvantage	-8.50E-09	4.8342		-11.5103	20.6906

APPENDIX B: BIVARIATE CORRELATIONS

Table B.1: Correlations of Original Block Group Values

Block Group	1.	2.	3.	4.	5.	6.
1. Collective Efficacy	1.00					
2. Settlement Type	-0.19***	1.00				
3. % Black Pop	0.05$^{\times}$	0.00	1.00			
4. % Foreign-born Pop	-0.11**	0.38***	-0.21***	1.00		
5. % 5-Year Residence	0.14***	-0.26***	0.28***	-0.35***	1.00	
6. Disadvantage	-0.04	-0.11**	0.52***	0.01	0.01	1.00

Note: n=683; *** p<0.001; ** p<0.01; * p<0.05; $^{+}$ p<0.10; $^{\times}$ p<0.20

Table B.2: Correlations of Dasymetric Interpolation with Circular Buffers at 75%

IDM CIR 75	1.	2.	3.	4.	5.	6.
1. Collective Efficacy	1.00					
2. Settlement Type	-0.20***	1.00				
3. % Black Pop	0.04	0.01	1.00			
4. % Foreign-born Pop	-0.12**	0.42***	-0.22***	1.00		
5. % 5-Year Residence	0.15***	-0.28***	0.30***	-0.35***	1.00	
6. Disadvantage	-0.06^{+}	-0.12**	0.52***	0.01	0.01	1.00

Note: n=683; *** p<0.001; ** p<0.01; * p<0.05; $^{+}$ p<0.10; $^{\times}$ p<0.20

Table B.3: Correlations of Dasymetric Interpolation with Circular Buffers at 100%

IDM CIR 100	1.	2.	3.	4.	5.	6.
1. Collective Efficacy	1.00					
2. Settlement Type	-0.20***	1.00				
3. % Black Pop	0.03	0.00	1.00			
4. % Foreign-born Pop	-0.12**	0.43***	-0.21***	1.00		
5. % 5-Year Residence	0.13***	-0.29***	0.29***	-0.37***	1.00	
6. Disadvantage	-0.07^{+}	-0.11**	0.52***	0.01	0.02	1.00

Note: n=683; *** p<0.001; ** p<0.01; * p<0.05; $^{+}$ p<0.10; $^{\times}$ p<0.20

Table B.4: Correlations of Dasymetric Interpolation with Circular Buffers at 150%

IDM CIR 150	1.	2.	3.	4.	5.	6.
1. Collective Efficacy	1.00					
2. Settlement Type	-0.20***	1.00				
3. % Black Pop	0.02	0.01	1.00			
4. % Foreign-born Pop	-0.13***	0.45***	-0.20***	1.00		
5. % 5-Year Residence	0.12**	-0.31***	0.27***	-0.39***	1.00	
6. Disadvantage	-0.08*	-0.10**	0.53***	0.01	0.02	1.00

Note: n=683; *** p<0.001; ** p<0.01; * p<0.05; $^{+}$ p<0.10; $^{\times}$ p<0.20

Table B.5: Correlations of Dasymetric Interpolation with Network Buffers at 75%

IDM NET 75	1.	2.	3.	4.	5.	6.
1. Collective Efficacy	1.00					
2. Settlement Type	-0.20***	1.00				
3. % Black Pop	0.06$^{\times}$	-0.03	1.00			
4. % Foreign-born Pop	-0.12**	0.40***	-0.23***	1.00		
5. % 5-Year Residence	0.14***	-0.26***	0.30***	-0.33***	1.00	
6. Disadvantage	-0.06$^{\times}$	-0.11**	0.53***	0.00	0.00	1.00

Note: n=683; *** p<0.001; ** p<0.01; * p<0.05; $^{+}$ p<0.10; $^{\times}$ p<0.20

Table B.6: Correlations of Dasymetric Interpolation with Network Buffers at 100%

IDM NET 100	1.	2.	3.	4.	5.	6.
1. Collective Efficacy	1.00					
2. Settlement Type	-0.20***	1.00				
3. % Black Pop	0.06$^{\times}$	-0.02	-1.00***			
4. % Foreign-born Pop	-0.12**	0.41***	-0.22***	1.00		
5. % 5-Year Residence	0.14***	-0.27***	0.30***	-0.35***	1.00	
6. Disadvantage	-0.06$^{\times}$	-0.11**	0.53***	0.01	0.00	1.00

Note: n=683; *** p<0.001; ** p<0.01; * p<0.05; $^{+}$ p<0.10; $^{\times}$ p<0.20

Table B.7: Correlations of Dasymetric Interpolation with Network Buffers at 150%

IDM NET 150	1.	2.	3.	4.	5.	6.
1. Collective Efficacy	1.00***					
2. Settlement Type	-0.20***	1.00				
3. % Black Pop	0.04	-0.01	1.00			
4. % Foreign-born Pop	-0.12**	0.42***	-0.21***	1.00		
5. % 5-Year Residence	0.13***	-0.29***	0.29***	-0.37***	1.00	
6. Disadvantage	-0.07^{+}	-0.11**	0.53***	0.01	0.01	1.00

Note: n=683; *** p<0.001; ** p<0.01; * p<0.05; $^{+}$ p<0.10; $^{\times}$ p<0.20

Table B.8: Correlations of Areal Weighting with Circular Buffers at 75%

AW CIR 75	1.	2.	3.	4.	5.	6.
1. Collective Efficacy	1.00					
2. Settlement Type	-0.20***	1.00				
3. % Black Pop	0.04	0.01	1.00			
4. % Foreign-born Pop	-0.12**	0.42***	-0.20***	1.00		
5. % 5-Year Residence	0.15***	-0.29***	0.29***	-0.37***	1.00	
6. Disadvantage	-0.06^{+}	-0.10**	0.51***	0.02	0.01	1.00

Note: n=683; *** p<0.001; ** p<0.01; * p<0.05; $^{+}$ p<0.10; $^{\times}$ p<0.20

Table B.9: Correlations of Areal Weighting with Circular Buffers at 100%

AW CIR 100	1.	2.	3.	4.	5.	6.
1. Collective Efficacy	1.00					
2. Settlement Type	-0.20***	1.00				
3. % Black Pop	0.03	0.01	1.00			
4. % Foreign-born Pop	-0.12**	0.42***	-0.20***	1.00		
5. % 5-Year Residence	0.14***	-0.30***	0.29***	-0.38***	1.00	
6. Disadvantage	-0.07^{+}	-0.09*	0.52***	0.02	0.02	1.00

Note: n=683; *** p<0.001; ** p<0.01; * p<0.05; $^{+}$ p<0.10; $^{\times}$ p<0.20

Table B.10: Correlations of Areal Weighting with Circular Buffers at 150%

AW CIR 150	1.	2.	3.	4.	5.	6.
1. Collective Efficacy	1.00					
2. Settlement Type	-0.20***	1.00				
3. % Black Pop	0.02	0.02	1.00			
4. % Foreign-born Pop	-0.13***	0.45***	-0.19***	1.00		
5. % 5-Year Residence	0.12**	-0.32***	0.27***	-0.39***	1.00	
6. Disadvantage	-0.08*	-0.09^{+}	0.53***	0.01	0.03	1.00

Note: n=683; *** p<0.001; ** p<0.01; * p<0.05; $^{+}$ p<0.10; $^{\times}$ p<0.20

Table B.11: Correlations of Areal Weighting with Network Buffers at 75%

AW NET 75	1.	2.	3.	4.	5.	6.
1. Collective Efficacy	1.00					
2. Settlement Type	-0.20***	1.00				
3. % Black Pop	0.05$^{\times}$	0.01	1.00			
4. % Foreign-born Pop	-0.12**	0.41***	-0.20***	1.00		
5. % 5-Year Residence	0.14***	-0.27***	0.28***	-0.36***	1.00	
6. Disadvantage	-0.05$^{\times}$	-0.10**	0.52***	0.01	0.03	1.00

Note: n=683; *** p<0.001; ** p<0.01; * p<0.05; $^{+}$ p<0.10; $^{\times}$ p<0.20

Table B.12: Correlations of Areal Weighting with Network Buffers at 100%

AW NET 100	1.	2.	3.	4.	5.	6.
1. Collective Efficacy	1.00					
2. Settlement Type	-0.20***	1.00				
3. % Black Pop	0.05	0.02	1.00			
4. % Foreign-born Pop	-0.12**	0.41***	-0.19***	1.00		
5. % 5-Year Residence	0.14***	-0.28***	0.28***	-0.37***	1.00	
6. Disadvantage	-0.06^{+}	-0.10*	0.52***	0.02	0.01	1.00

Note: n=683; *** p<0.001; ** p<0.01; * p<0.05; $^{+}$ p<0.10; $^{\times}$ p<0.20

Table B.13: Correlations of Areal Weighting with Network Buffers at 150%

AW NET 150	1.	2.	3.	4.	5.	6.
1. Collective Efficacy	1.00					
2. Settlement Type	-0.20***	1.00				
3. % Black Pop	0.03	0.03	1.00			
4. % Foreign-born Pop	-0.12**	0.42***	-0.19***	1.00		
5. % 5-Year Residence	0.13***	-0.30***	0.27***	-0.37***	1.00	
6. Disadvantage	-0.07*	-0.9*	0.53***	0.01	0.02	1.00

Note: n=683; *** p<0.001; ** p<0.01; * p<0.05; $^{+}$ p<0.10; $^{\times}$ p<0.20

APPENDIX C: MULTIPLE REGRESSION RESULTS

Table C.1: Comparison of regression parameter estimates and model fit between block groups and areal weighting interpolation.

		Block Group	Areal Weighting Interpolation					
			Circular Buffers			Network Buffers		
			75%	100%	150%	75%	100%	150%
Settlement Type	*b*	-0.5242	-0.5247	-0.5324	-0.5467	-0.5367	-0.5379	-0.5439
	t	-4.380	-4.180	-4.190	-4.190	-4.270	-4.260	-4.220
	*b**	-0.1874	-0.1876	-0.1903	-0.1954	-0.1919	-0.1923	-0.1945
Concentrated Disadvantage	*b*	-0.1395	-0.1698	-0.1796	-0.1964	-0.1728	-0.1748	-0.1843
	t	-2.210	-2.770	-2.940	-3.230	-2.710	-2.710	-2.870
	*b**	-0.1023	-0.1298	-0.1386	-0.1548	-0.1313	-0.1336	-0.1429
% 5-Year Residence	*b*	2.7024	3.1292	2.8497	1.8039	2.8001	2.7173	2.3826
	t	1.560	1.620	1.420	0.820	1.550	1.460	1.220
	*b**	0.0686	0.0698	0.0607	0.0355	0.0648	0.0610	0.0502
% Pop Black	*b*	1.7484	1.8995	1.9191	2.1988	2.1994	2.2347	2.1980
	t	1.800	1.870	1.850	2.040	2.250	2.220	2.110
	*b**	0.0883	0.0889	0.0878	0.0971	0.1060	0.1052	0.1001
% Pop Foreign-born	*b*	0.1571	0.0200	0.1028	-0.5109	0.1495	0.1876	0.0566
	t	0.050	0.010	0.030	-0.130	0.040	0.050	0.010
	*b**	0.0024	0.0003	0.0014	-0.0063	0.0021	0.0026	0.0007
Constant	*b*	30.7267	30.5010	30.6713	31.2129	30.5869	30.6286	30.8600
R2		0.0543	0.0587	0.0582	0.0582	0.0593	0.0589	0.0579
F-value		8.39	8.59	8.44	8.37	8.63	8.57	8.3
RMSE		6.088	6.0738	6.0753	6.0755	6.0719	6.0732	6.0765
AIC		4405.23	4402.05	4402.39	4402.42	4401.61	4401.91	4402.64

Table C.2: Comparison of regression parameter estimates and model fit between block groups and "Intelligent Dasymetric Mapping" interpolation.

		Block Group	"Intelligent Dasymetric Mapping"					
			Circular Buffers			Network Buffers		
			75%	100%	150%	75%	100%	150%
Settlement Type	*b*	-0.5242	-0.5298	-0.5413	-0.5508	-0.5332	-0.5387	-0.5500
	t	-4.380	-4.220	-4.250	-4.220	-4.280	-4.280	-4.280
	*b**	-0.1874	-0.1894	-0.1935	-0.1969	-0.1906	-0.1926	-0.1966
Concentrated Disadvantage	*b*	-0.1395	-0.1737	-0.1849	-0.1988	-0.1765	-0.1829	-0.1930
	t	-2.210	-2.890	-3.030	-3.270	-2.800	-2.860	-3.010
	*b**	-0.1023	-0.1333	-0.1430	-0.1566	-0.1343	-0.1404	-0.1499
% 5-Year Residence	*b*	2.7024	3.0877	2.5754	1.8977	2.6966	2.4759	2.3323
	t	1.560	1.600	1.290	0.870	1.510	1.350	1.200
	*b**	0.0686	0.0692	0.0550	0.0374	0.0629	0.0559	0.0493
% Pop Black	*b*	1.7484	1.8732	1.9730	2.1880	2.1134	2.2636	2.2308
	t	1.800	1.870	1.930	2.050	2.240	2.310	2.180
	*b**	0.0883	0.0893	0.0915	0.0970	0.1068	0.1110	0.1040
% Pop Foreign-born	*b*	0.1571	-0.0051	0.0542	-0.3334	0.1794	0.1509	0.3425
	t	0.050	0.000	0.010	-0.080	0.050	0.040	0.080
	*b**	0.0024	-0.0001	0.0007	-0.0043	0.0024	0.0020	0.0043
Constant	*b*	30.7267	30.5532	30.8261	31.17592	30.6269	30.7176	30.8620
R2		0.0543	0.0593	0.0593	0.0588	0.0596	0.0597	0.0593
F-value		8.39	8.67	8.67	8.32	8.68	8.63	8.39
RMSE		6.088	6.0719	6.0719	6.0735	6.0709	6.0708	6.0719
AIC		4405.23	4401.612	4401.61	4401.96	4401.40	4401.36	4401.61