AN ALTERNATIVE APPROACH TO ANCILLARY DATA IN DASYMETRIC MAPPING TECHNIQUES

By

Almaz Sium Naizghi

A THESIS

Submitted to
Michigan State University
In partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Geography

2010

ABSTRACT

AN ALTERNATIVE APPROACH TO ANCILLARY DATA IN DASYMETRIC MAPPING TECHNIQUES

By

Almaz Sium Naizghi

The United States Census Bureau conducts a count and representative sample of the American population at the beginning of each decade. Census data are important for congressional redistricting, distribution of governmental funds and planning purposes. Unfortunately, these extremely robust decadal surveys are still rife with limitations. For example, aggregated census population data have several analytical and cartographic problems because the partitioning of areal units is not always based on natural geographical features. Conventional dasymetric mapping techniques attempt to mitigate this effect by employing an areal interpolation technique that disaggregates spatial data using ancillary information (i.e., land use or land cover data).

Previous research has demonstrated the utility of dasymetric techniques; however, many improvements remain possible. This research identified the major errors involved in conventional dasymetric mapping and examined the methods to extend conventional dasymetric mapping techniques by using alternative sources of ancillary data, including municipal zoning data and building-level population data. The alternative ancillary data helped to refine population estimates. Each technique was validated using the pycnophylactic property and it was found that the building level dasymetric method based on census data had preserved the population most accurately and with RMSE of zero. In addition, this research discussed about future development of dasymetric mapping.

Dedicated to my younger brother:
Engineer Mussie S. Naizghi
This would not have been possible
without your support.

ACKNOWLEDEGMENTS

I wish to express my deep appreciation to my advisor Dr. Kirk Goldsberry for your encouragement and support from the first day I came to MSU and your efforts to finish my thesis. I would like to thank my committee members Dr. Ashton Shortridge and Dr. Sue Grady for all their advice and support. I also owe a special thanks to my family members: my mother Alemash Tewelde who spent all her time and energy for me, my husband Weldekidan Hailu who was always besides me, my daughter Ruta Weldekidan for her patience to do my study, my brother Berhane Seyoum (M.D.) from Wayne State University who helped me morally and financially. I appreciate for all the technical support I have received from my former professor Johnathan Bascom from Calvin College and my friends Linda, Tanita, Sarah, Mahalakshmi, Kristie, Lindsay, and Carolyn.

TABLE OF CONTENTS

LIST OF TABLES vii

LIST OF FIGURES viii

Chapter One – INTRODUCTION 1
- 1.1 Introduction 1
- 1.2 Research goals 4
- 1.3 Research questions. 5
- 1.4 Relevance of research 5

Chapter Two – LITERATURE REVIEW 7
- 2.1 Introduction and background 7
- 2.2 Census background 8
  - 2.2.1 Significance of U.S. Census 10
- 2.3 Choropleth mapping 11
  - 2.3.1 Limitations of choropleth mapping 12
- 2.4 Significance of dasymetric mapping 13
- 2.5 Methods used in dasymetric mapping 13
  - 2.5.1 Areal interpolation method 14
  - 2.5.2 Binary method 17
  - 2.5.3 Three class method 19
  - 2.5.4 The limiting variable method 20
- 2.6 Land use/land cover classification 22
- 2.7 Limitations of dasymetric mapping 23

Chapter Three – METHODS 25
- 3.1 Study area 25
- 3.2 Data 26
  - 3.2.1 U.S. Census data 26
  - 3.2.2 Land cover data 27
  - 3.2.3 Municipal zoning data 29
  - 3.2.4 MSU building footprints 30
  - 3.2.5 Supplementary campus population data 30
  - 3.2.6 Data limitations 33
- 3.3 Overview of methods 33
  - 3.3.1 Dasymetric map using land use and land cover as ancillary information 34
  - 3.3.2 Case study 1: Dasymetric map using zoning data as ancillary information 39

3.3.3 Case study 2: Dasymetric map using Michigan State University building footprint data as ancillary information 43
3.3.4 Dasymetric map using supplementary campus population data 45
3.4 Accuracy assessment 46

Chapter Four – RESULTS 48
4.1 Overview 48
4.2 Conventional dasymetric mapping 48
4.2.1 Separating human population from water bodies 53
4.2.2 Separating human population from forest 56
4.3 Case study 1: Municipal zoning data as ancillary information 59
4.3.1 Overview of zoning dasymetric map (ZDM) 60
4.4 Case study 2: Building level dasymetric mapping (BLDM) 65
4.4.1 Building level population of census blocks 65
4.4.2 Building level population of supplementary campus data 70
4.5 Statistical and visual analysis of errors in dasymetric maps 72
4.5.1 Validation of dasymetric maps 72
4.5.2 Visual analysis of spatial errors 74

Chapter Five – DISCUSSION 82
5.1 Mapping human population 82
5.2 Uncertainties and errors in dasymetric mapping 83
5.3 Limitations of conventional dasymetric mapping 84
5.3.1 U.S. Census aggregation problem 84
5.3.2 Uncertainties in classification of remotely sensed imagery 84
5.3.3 Mixed pixel problem 85
5.3.4 Conversion from vector to raster 86
5.3.5 Subjective population weighting 87
5.4 Advantage of zoning data as ancillary information 89
5.4.1 Limitations of ZDM 93
5.5 Advantage of building data as ancillary information 94
5.5.1 Limitations of building level population 97
5.6 Daytime vs. nighttime population distribution 98
5.7 The future of dasymetric mapping 100
5.7.1 LiDAR building extraction 100
5.7.2 Digital Elevation Model (DEM) as ancillary information 102
5.7.3 Population weighting based on zoning data 102
5.7.4 Real-time population 103
5.8 Summary of limitations 104
5.9 Conclusions 105

REFERENCES 109

LIST OF TABLES

Table 3.1 Land use and land cover classification values according to NOAA C-CAP. 29

Table 3.2 Number of people living in MSU residence halls and apartments in 2010. 32

Table 3.3 Municipal zoning data categories assigned by the city of East Lansing. The zoning data help to differentiate the residential areas from non-residential areas which, the satellite images were not able to differentiate. 40

Table 4.1 Summary result of pycnophylactic property..74

Table 5.1 Summary of errors and uncertainties in dasymetric mapping techniques. 98

LIST OF FIGURES

Figure 2.1 Example of ancillary data (land cover) with different classes such as urban, agriculture, and forest. The spatial resolution of the data is 30 meters by 30 meters. 16

Figure 2.2 Illustration of binary method showing inhabitable and uninhabitable land cover in a grid format. 18

Figure 3.1 Illustration of how the land cover data and census data were integrated in GIS to produce a dasymetric map. 34

Figure 3.2 Flow diagram of the proposed dasymetric mapping process. 38

Figure 3.3 Illustration of how the census, land cover, and zoning data are integrated in GIS to produce a ZDM. 39

Figure 3.4 Illustration integrating land cover and zoning data and the resulting land cover refined residential areas only by removing pixels that are non–residential areas based on the municipal zoning data. 42

Figure 4.1 Urban land cover of Ingham County where 70%, 20%, and 10% of the census block population is weighted to high, medium, and low-intensity developed and other land cover class are assigned zero population density. 50

Figure 4.2 Choropleth map of 2000 U.S. Census block population of Ingham County, showing a homogenous distribution of population within each block. However, in reality the land cover and population is heterogeneous. 51

Figure 4.3 Conventional dasymetric mapping disaggregated the census block population to only urban land cover classes. Zero population is assigned to other land cover such as water, agriculture, and forest. Box A and B indicate the area of detail shown in Figure 4.4 and Figure 4.7 respectively. 52

Figure 4.4 Detail of census blocks specified in Figure 4.2, as box A, showing the difference between a disaggregated residential population in a dasymetric map (a) compared to a uniformly distributed population in a choropleth map (b) in which the central enumeration unit includes a small lake. 54

Figure 4.5 Small lake by Lake View apartments, Haslett, Michigan. The lake and the residential building are clearly distinguished, as represented on Figure 4.4a. April 2010, photo by author. 55

Figure 4.6 This aerial image identifies the lae and buildings clearly. The geographic distribution of population implied by the image is best represented in Figure 4.4a. The Google Earth [*last accessed 04-20-2010*] 55

Figure 4.7 Detail of census blocks specified in Figure 4.2 as box B, showing the difference between dasymetric map (a) and choropleth map (b) representation of population density. 57

Figure 4.8 The choropleth map underestimates the population density in the dormitories of MSU. April 2010, photo by author.. 58

Figure 4.9 The choropleth map overestimates the population density in forest areas of MSU. April 2010, photo by author.. 58

Figure 4.10 Dormitories and forest aerial view at MSU. The choropleth map is misleading as large area of the census block is covered by forest. Google Earth [*last accessed 04-20-2010*]... 59

Figure 4.11 Choropleth map of East Lansing showing population density homogenously distributed within census blocks. In reality, we could not get such uniformity, as some places are non-residential, such as commercial and business districts. 61

Figure 4.12 Land cover of East Lansing first integrated with zoning data where 70%, 20% and 10% of the census block population is assigned to high, medium and low-intensity developed respectively. Other types of land cover class would be assigned a population of zero. 62

Figure 4.13 Residential and non-residential zones designated by the city of East Lansing. Residential areas have “R” in the zoning codes, and include housing types such as low density single-family and medium density multiple-family. Non-residential areas are all places without “R” of the zoning codes, such as commercial and university areas... 63

Figure 4.14 ZDM of East Lansing showing human population density more disaggregated to residential areas while all non-residential areas such as commercial or business zones and roads have been assigned zero population density. 64

Figure 4.15 Choropleth map of MSU shows homogenous population density within census blocks. In reality, the population is only concentrated in the residential buildings of the university halls and apartments. 67

Figure 4.16 Building footprints of MSU that were used as ancillary information in the process of dasymetric mapping. Residential buildings include all university residence halls and apartments while non-residential buildings are all offices, classrooms and parking structures. 68

Figure 4.17 The building dasymetric map shows more refined population density, since it controls the total number of people from the census block to only residential areas of MSU. 69

Figure 4.18 Building dasymetric map based on 2009 supplementary population data. It was able to show the population density of each individual building 71

Figure 4.19 Dasymetric map (a) using only land cover as ancillary information in Ingham County showing a percent error map (b), and count error map (c). Percent error is error relative to the total population of a polygon in the census blocks while count error is a value calculated using the actual population of the census block subtracted from the estimated population data. 78

Figure 4.20 ZDM (a), percent error map (b), and count error map (c) in East Lansing. Percent error is error relative to the total population of a polygon in the census blocks while count error is a value calculated using the actual population of the census block subtracted from the estimated population data. 79

Figure 4.21 Illustration of problems related to BLDM based on the supplementary population data. The building shapefile overlaid on the BLDM shows how the shape and size of the buildings are distorted and how this method was not able to preserve the population using the assessment of the pycnophylactic property. 81

Figure 5.1 The ZDM was prepared based on population weighting of 70%, 20%, and 10% to high, medium, and low urban land cover respectively (a). The ZDM was modified to 80%, 15%, and 5% of the population assigned to high, medium, and low urban land cover respectively (b). The 70-20-10 population weighting was subtracted from 80-15-5 population weighting to show how the two dasymetric maps are different (c). 88

Figure 5.2 ZDM of East Lansing improved human population density compared to the conventional dasymetric map. Box A in indicates the area of detail shown in Figure 5.3. 91

Figure 5.3 Detail of census blocks specified in Figure 5.2 as box A, showing the difference between ZDM (a) where the Seven Eleven area was designated a zero population, and conventional dasymetric map (b) in which, commercial areas and roads were assigned higher population inaccurately. 92

Figure 5.4 Commercial areas (Seven Eleven) and roads in East Lansing. This area was classified as high-intensity developed in land cover and the conventional dasymetric map allocated population inaccurately. April 2010, photo by author. 93

Figure 5.5 Building dasymetric map produced by distributing population to only residential building of MSU. It improved human population density compared to the zoning and conventional dasymetric maps. Box A indicates the area of detail shown in Figure 5.6. 95

Figure 5.6 Detail of census blocks specified in Figure 5.2 as box A, showing the difference among a, conventional dasymetric map b, ZDM and c, BLDM. Both land cover and zoning ancillary information alone were not able to display the population distribution accurately. The BLDM was able to allocate the population to only residential building as it is found on the ground. 96

Chapter One – INTRODUCTION

1.1 Introduction

United States Census data are used mainly to determine the number of seats each state is allotted in the House of Representatives. In addition to this it has other applications such as planning purposes, distributing governmental funds, and identifying trends over time (U.S. Census Bureau). However, geographic analyses of the data are limited for many reasons. The U.S Census Bureau's counts and samples data at the individual level, but the data are aggregated to enumeration units prior to being published. The boundaries of census enumeration units represent areas within which to monitor population change, both in number and composition. These boundaries are constructed through hierarchical aggregations from blocks to block groups to census tracts to counties and to states. However, the boundaries are primarily designed to ease the aggregation process rather than to represent the most appropriate geographical distribution of the population or any socio-economic variables, and they do not necessarily correspond to official geographical boundaries (Wu & Murray, 2005; & Liu, 2003). As a consequence, aggregation tends to smooth local spatial variability and depicts population as distributed uniformly across enumeration units (Moon & Farmer, 2001). There are important reasons for reporting census information in aggregated ways, such as privacy, but it is also done for convenience, to minimize data volumes (Yaun *et al*., 1997; Sadahiro, 1999; Langford, 2007).

Besides the assumption that the population is distributed uniformly, census-derived boundaries may or may not correspond with the actual settlement patterns or communities (Goodchild *et al.,* 1993). Therefore, zone-based population data are related to incompatible spatial information layers since different departments and agencies (e.g. school districts, transportation analysis zones, and metropolitan statistical areas) collect and distribute data in varying zonal arrangements (Moon & Farmer, 2001). As a consequence, a significant problem arises in regional analysis and modeling of a population, in which multiple data sources must be integrated before analysis can be implemented (Goodchild *et al.*, 1993). Since the boundaries of areal census units are based on enumeration and reporting rather than derived data, census data are also vulnerable to the effect of modifiable areal unit problem (MAUP), where changing the areal units or scales of the data can significantly affect subsequent analyses (Openshaw, 1984).

Conventionally, cartographers have used choropleth mapping to represent census - derived population counts. A choropleth map is a thematic map in which geographic areas, usually political units, are shaded (or patterned) in accordance with the measurement of some geographical variable (Holt *et al*., 2004). As a result, within a choropleth map, each enumeration unit behaves as a map symbol and the depiction includes variables uniformly distributed within the enumeration units comprising the map (Slocum *et al.*, 2009).

The distribution of human population is commonly mapped using choropleth techniques (Suchan *et al.*, 2007). The U.S. Census regularly produces choropleth maps of population density, which are easily derived by dividing a given unit's total population by its area. Unfortunately, choropleth approaches to mapping human population have several inherent problems (Holt *et al*., 2004). First, due to the MAUP, changing the areal extent of a given enumeration area will change the appearance of population density depending on how the boundary of the enumeration area is delineated (Openshaw, 1984). Second, choropleth maps give the impression of abrupt changes based on boundaries of the administrative areas such as census blocks, census block groups or census tracts (Eicher & Brewer, 2001). According to the principles of symbolization (Slocum *et al*., 2009), human population is not a continuous spatial phenomenon; however, population density, which is the number of people per unit area, is a continuous variable because estimated values can be defined everywhere. Lastly, choropleth approaches graphically imply homogenous distribution of population although in reality the distributions are quite heterogeneous (Holt *et al.*, 2004; Liu *et al.,* 2008; Maantay *et al.*, 2007). Many census zones are likely to have residential as well as non-residential lands such as woodlands, open water bodies, parks, industrial premises, or commercial districts (Mennis, 2003; Langford, 2003).

One approach for dealing with the limitations of choropleth maps is to transform aggregated census data into grid-based population estimates using areal interpolation, which is designed to transform data from source zones to target zones. In the context of population distribution, census units, such as census blocks or census block groups usually serve as the source zones; the target zones are typically grid cells or land use zones (Liu *et al.,* 2008; Mennis, 2003; Wu & Murray, 2005).

Dasymetric mapping is an areal interpolation technique that disaggregates spatial data to a finer unit of analysis using additional (or ancillary) data to help refine locations of population or other phenomena (e.g., Eicher & Brewer, 2001; Langford & Unwin, 1994; Maantay *et al.*, 2007; Mennis, 2003; Wu & Murray, 2005). The ancillary data often consist of land use and land cover information, which indicate where human beings are most or least likely to inhabit. For example, since human beings do not inhabit lakes, a dasymetric approach excludes population counts from pixels classified as water in land cover imagery. Previous research has demonstrated the utility of dasymetric techniques; however, many improvements remain possible. This thesis identifies the major errors involved with dasymetric mapping and refines population mapping techniques using various forms of ancillary data.

1.2 Research goals

The goal of this research is to extend conventional dasymetric techniques by demonstrating the abilities of various ancillary data sources to produce more refined population maps. Although the conventional way to produce dasymetric maps involves land use and land cover as ancillary information, researchers have found it is vulnerable to errors and uncertainty (Mennis, 2009; Yuan *et al.,* 1997). These shortcomings result from different types of sources. First, even if the smallest U.S. Census enumeration unit, the block is used the data are aggregated, which introduces errors. Second, there is no ideal classification of land cover. Although it uses an objective numerical approach classification, the process itself tends to be subjective due to different perspectives (Anderson *et al.,* 1976). Third, so far dasymetric mapping has not yet developed standardized techniques that are accessible to all (Mennis, 2009) and the weights used to distribute the population are subjective. Using a case study, this research will examine dasymetric techniques by incorporating land cover with additional information such as municipal zoning data and residential building footprints. The study has the following goals:

1. To determine the taxonomy of uncertainty and errors in dasymetric mapping.
2. To create a dasymetric map using municipal zoning data as ancillary information and compare it with the conventional dasymetric map.
3. To create a dasymetric map using building footprint data as ancillary information and compare it with the conventional and zoning dasymetric map (ZDM).
4. To validate and verify the output of dasymetric mapping techniques using the pycnophylactic property and determine which ancillary data preserve the population more accurately.

1.3 Research questions.

To achieve the overall goal of extending dasymetric mapping techniques, the following research questions were investigated:

1. Do other ancillary data such as municipal zoning data or building footprint data help improve the accuracy of dasymetric mapping?
2. Are there additional types of errors and uncertainties associated with these new kinds of ancillary data?

1.4 Relevance of research

Many methods have been practiced in GIS and remote sensing fields to estimate population. By combining vector population data with ancillary information, dasymetric mapping has the potential to depict human population distribution with increased accuracy. However, this potential remains unrealized. Hence, this research extends the conventional dasymetric approaches by identifying the errors and uncertainties involved in that and assess the use of other ancillary data. Further, the study provides a practical illustration by examining a case study and demonstrating a new, more accurate, method of mapping. The broader impacts of this research will contribute to the development of new dasymetric mapping techniques. Accurate population estimation has many useful applications to fields such as public health, crime mapping, and risk assessment. For example, emergency management operations and implementation can be improved by providing more precise information for the actual positions of susceptible populations. In addition, accurate population distribution can also be valuable for urban planning by identifying the characteristics of target populations. These characteristics can be used to arrive at more equitable resource allocation (Maantay *et al.,* 2007).

This thesis is organized into five chapters. Chapter one introduces the research goals and research questions of the study. Chapter two reviews the background of dasymetric mapping and methods applied in the literature. The overall structure of the methods is outlined in chapter three. Chapter four discusses the results of each dasymetric mapping technique. Finally, chapter five presents the discussion, conclusion, and recommendations for future research.

Chapter Two – LITERATURE REVIEW

2.1 Introduction and background

This chapter discusses the historical background of dasymetric mapping and explains more about significance of the U.S. Census data and its origin. Additionally, it reviews choropleth mapping limitations and how dasymetric mapping improves those shortcomings. The final section discusses the areal interpolation techniques and outlines the methods used in dasymetric mapping by explaining the ancillary information of land use and land cover data and states the limitation of dasymetric mapping.

Historically, dasymetric maps are associated with mapping human populations. Early dasymetric maps demonstrate the motives behind the technique (MacEachren, 1979). The first dasymetric map, created in 1833 by George Poulett Scrope depicted world population density (Maantay *et al*., 2007). The second one, produced in 1837 by Henry Drury Harness, displayed the population density of Ireland for the Second Report of the Railway Commissioners. Although by that time the term itself had not yet been invented and the authors did not claim to be the originators of the dasymetric technique, both maps applied basic dasymetric methods by shading the magnitude of the population density that did not correspond consistently with the administrative boundaries (Mennis, 2009).

In recent literature, there is a great deal of misunderstanding on the origin of dasymetric mapping. Mennis (2009) suggested that one reason for this misunderstanding may be due to the variety of academic backgrounds for those practicing dasymetric mapping. Second, there were long periods of inactivity from the time of its invention until the development of technologies to assist its application. In addition, some early examples and the use of the term "dasymetric mapping" were published in Russian geography journals and reports which made it difficult for English speaking practitioners to access the information (Mennis, 2009). Many researchers of dasymetric mapping refer to the 1936 Geographical Review article written by John K. Wright and Wright has occasionally been credited as the person who invented dasymetric mapping, although Wright mentioned its Russian origin and identified the term dasymetric, which means "density measuring" (Langford, 2003; Mennis, 2009). Following Wright's (1936) publication, cartographers tried to develop dasymetric methods; however, it progressed significantly more slowly than choropleth mapping.

Recent developments in GIS and remote sensing have transformed dasymetric mapping from an obscure cartographic technique to a much more popular investigation topic (Langford & Unwin, 1994; Mennis, 2009; Poulsen & Kennedy 2004; Yuan, 1997). Contemporary GIS analysis makes it possible to integrate a variety of spatial information to create a new, relatively homogenous area of target units from the original census units (Robinson *et al*., 1995). For these reasons, dasymetric mapping will continue to be an important research topic within geographic information science.

2.2 Census background

Since ancient times, humans have used censuses to help understand human population. Historically, emperors and kings used them to evaluate the strength of their kingdom. Early censuses were conducted sporadically and their main purpose was to measure the tax or military capacity of a particular area. Unlike the modern census, which counts every individual, early censuses tended to count only adult men who were responsible for military services or liable for taxes. The modern census approach started during the seventeenth and eighteenth centuries with the colonial powers in Western Europe, and was used to determine the success of their colonies overseas (Anderson, 2000).

By definition, a census is a count or enumeration of everybody or everything in a country as of a fixed date (MacDonald & Peters, 2004). National governments conduct censuses to determine how many people live throughout the country in order to assess whether the population is growing, stable, or declining as a whole, or in particular geographical area within the country. In addition, a census typically provides demographic characteristics about a given population such as age, sex, ethnic background, and marital status as well as numerous indicators of socio-economic information such as wealth and health (Langford, 2003). Governments collect the information either by sending a questionnaire in the mail to every residential addresses or interviewing every household (Anderson, 2000).

The U.S. Census has been conducted decennially since 1790, in years ending with zero (Lavin, 1996). In 1790, congress passed a law which directed how and by whom the census shall be conducted. This was the first census law, the ACT of March 1, 1790 (1 Stat. 101) and established a residence rule "a person is to be counted where he or she usually resides" (Anderson, 2000, p. 56).

The 2000 U.S. Census collected data using two types of questionnaires: the short form and the long form. The short form (100-percent characteristics) had a limited number of questions and is asked to every individual and housing unit. Information, such as age, sex and race could be derived. The long forms (sample characteristics) had additional questions and were asked of a sample of persons and housing units (basically one in six households). The data gathered through long form questionnaires were more detailed including socio-economic and migration information. The results from the 2000 U.S. Census are available on the following format (U.S. Census Bureau).

1. Summary File 1 (SF 1) contains basic information on the U.S. population and the file's products are published with the census's smallest geographical area, which is the census block level.
2. Summary File 2 (SF 2) contains basic information on the U.S. population and the file's products are focused on examining the results by race and ethnicity. The smallest geographical area reported is the census tract. Both SF 1 and SF 2 are gathered from the short form questionnaire.
3. Summary File 3 (SF 3) contains the richest and most complete statistical data available on U.S. residents, and the smallest geographical level available is the census block group.
4. Summary File 4 (SF 4) contains data compiled from a sample of approximately 19 million housing units and is focused primarily on examining the results by race and ethnicity, and the smallest geographical area available is the census tract.

2.2.1 Significance of U.S. Census

The decennial census is the foundation of the U.S. democratic system of government. Historically, political power allocation among the states was a sensitive issue as smaller states had been concerned by the larger and faster growing states. Taking into consideration the differences among the state's population size, land mass, and natural and economic resources was a way to help to determine each state's financial contribution to the federal government (Anderson, 2000). Population size was used as a proxy of wealth and a basis for the state's political representation and financial support of the federal government. Therefore, the original purpose of the U.S. Census was to determine how many representatives each state would be allotted in the House of Representatives. However, census data have become very useful to many users outside of the federal government such as planners, human resource managers, lawyers, and academic researchers (Lavin, 1996).

2.3 Choropleth mapping

From their early development, both choropleth and dasymetric mapping were mainly designed to map population. These two mapping methods became more clearly differentiated in the 1900s. Choropleth mapping became far more popular, both in modern cartography and for general use outside the discipline (Eicher & Brewer, 2001). Choropleth mapping is most appropriate for a phenomenon that is uniformly distributed within enumeration unit boundaries. In addition, the enumeration units are preferably without significant variation in size and shape (Slocum *et al*., 2009) in order to prevent the effect of modifiable areal unit problem (MAUP) and ecological fallacy.

Openshaw (1984) discussed MAUP and ecological fallacy in his review. MAUP has two distinct, but closely related problems. First, a scale problem occurs from variation in the output that has been obtained when data for one set of areal units are progressively aggregated into smaller or larger units for the purpose of analysis. For example, when census enumeration units are aggregated from census block, into census block groups, census tracts, counties and states, the results change completely with an increasing scale. Second, the aggregation problem occurs from any variation in results due to the use of alternative units of analysis. MAUP is also closely related to the ecological fallacy problem. “An ecological fallacy occurs when it is inferred that results based on aggregate zonal (or grouped) data can be applied to the individuals who form the zones or groups being studied (Openshaw, 1984, p. 8)." However, census units are largely heterogeneous and so are particularly vulnerable to the effect of MAUP and ecological fallacy.

Langford (2003) discussed the two distinct purposes of mapping population: first, to create a cartographic expression by representing the spatial distribution and pattern of population across the geographical area; second, to extract quantitative estimates of population density for use in subsequent spatial analytical modeling. Population density is usually computed by dividing the total population for a given enumeration area by its total land area. This is easily done in GIS and the result will be displayed in the form of choropleth maps. Most people are familiar with these maps and they can easily understand and interpret them. In addition, it is easy to compare population densities in different areas (Holt *et al*., 2004; Langford, 2003; Langford & Unwin, 1994; Sutton *et al*., 2003).

2.3.1 Limitations of choropleth mapping

Population density is typically represented by choropleth maps in which the densities are calculated from a series of polygonal areas (normally administrative or census enumeration zones) and displayed using various shading schemes (Langford & Unwin, 1994). The spatially discontinuous or abruptly changing choropleth map is not a good representation of the underlying continuous or smoothly changing distribution of population density (Holt *et al*., 2004; Langford & Unwin, 1994; Mennis, 2003). Choropleth maps of population density derived from aggregated census data have several problems. First, modifying the areal units or scale of the data can significantly affect the result; this is referred to as the MAUP (Openshaw, 1984). Second, choropleth maps give the impression that population is distributed homogeneously within census zones, even though distributions are usually heterogeneous (Holt *et al*, 2004; Langford, 2003; Liu, 2003; Maantay *et al.*, 2007). Third, they frequently give the false impression of abrupt spatial changes depending on the arbitrarily created boundaries of census enumerations units (Eicher & Brewer, 2001).

2.4 Significance of dasymetric mapping

Reliable information on population distribution is significant for providing services and assessing risks (Maantay *et al.*, 2007; Briggs *et al.*, 2007). Dasymetric mapping is one way to represent the underlying population distribution more accurately, by reflecting the population with greater precision at finer spatial scale than choropleth mapping (Eicher & Brewer, 2001; Holloway *et al*., 1996; Holt *et al.*, 2004; McCleary, 1969; Mennis, 2003; Sleeter, 2004). Choropleth maps are developed based on existing administrative boundaries that are usually independent of the phenomena to be mapped. However, in dasymetric mapping, ancillary data are used to divide the original administrative units into smaller spatial units that are more homogenous (Eicher & Brewer, 2001; Holt *et al*., 2004; Maantay *et al*., 2008; Sleeter & Wood, 2006).

2.5 Methods used in dasymetric mapping

As a means to reduce some limitations of choropleth techniques, previous research has employed ancillary information to assist human population mapping. Ancillary data are additional information, commonly land use/land cover data, which further refines the distribution of population. John K. Wright (1936) demonstrated dasymetric mapping in Cape Cod, Massachusetts by first redistributing population from different sets of areal units into inhabited and uninhabited regions as indicated on the United States Geological Survey (USGS) topographic maps. He further subdivided the inhabited areas into smaller portions using settlement pattern data. On certain parts population density values were assigned derived from "controlled guesswork" which was more realistic and reveals the reality on ground. Wright's ideas are still being used, except today the ancillary information is most commonly derived from remotely sensed imagery (Liu *et al*., 2008). Various types of satellite imagery have been used to examine population distribution, including Thematic Mapper (TM), (Langford & Unwin, 1994; Mennis, 2003), and ETM+ (Enhanced Thematic Mapped Plus), (Wu & Murray, 2005). Although TM or ETM+ images provide valuable information for estimating population, their 30-meter spatial resolution limits their application, especially in large-scale urban analyses. Therefore, higher resolution imagery (0.5-5meter) such as IKONOS has been recommended to acquire detailed population density estimation (Jensen & Cowen, 1999). Further Eicher and Brewer (2001) stated that even if detail ancillary data are used for preparing the dasymetric zones, the cartographer's knowledge of the area is still very important.

2.5.1 Areal interpolation method

A common method to calculate disaggregated population values is areal interpolation. Mennis (2003) defined areal interpolation as the transformation of geographical data from one set (source units) to another set (target units). In the context of population distribution, source zones are typically census units such as census block, census block groups or census tracts while target zones are usually grid cells or land use/land cover zones (Liu *et al.*, 2008). Many areal interpolation methods have been developed and generally grouped into simple or intelligent depending on whether supplementary (ancillary) data have been applied (Okabe & Sadahiro, 1997). Simple interpolation methods do not use any ancillary data, other than the source-zone population. Areal weighting is an example of a simple interpolation which allocates population according to the amount of area proportional in the source zones verses the target zone (Langford, 2003; Liu *et al.*, 2008, Maantay, 2007). Goodchild and Lam (1980) discussed that the simple method of areal interpolation is to weight the variable's values by a ratio derived from the relative areal measurements of the two types of zones (source and target). Areal weighting is derived on the assumption that population is distributed uniformly. However, this method is susceptible to the major sources of error as on the ground there are zones that are uninhabitable such as water bodies, wetlands, and forests (Figure 2.1).

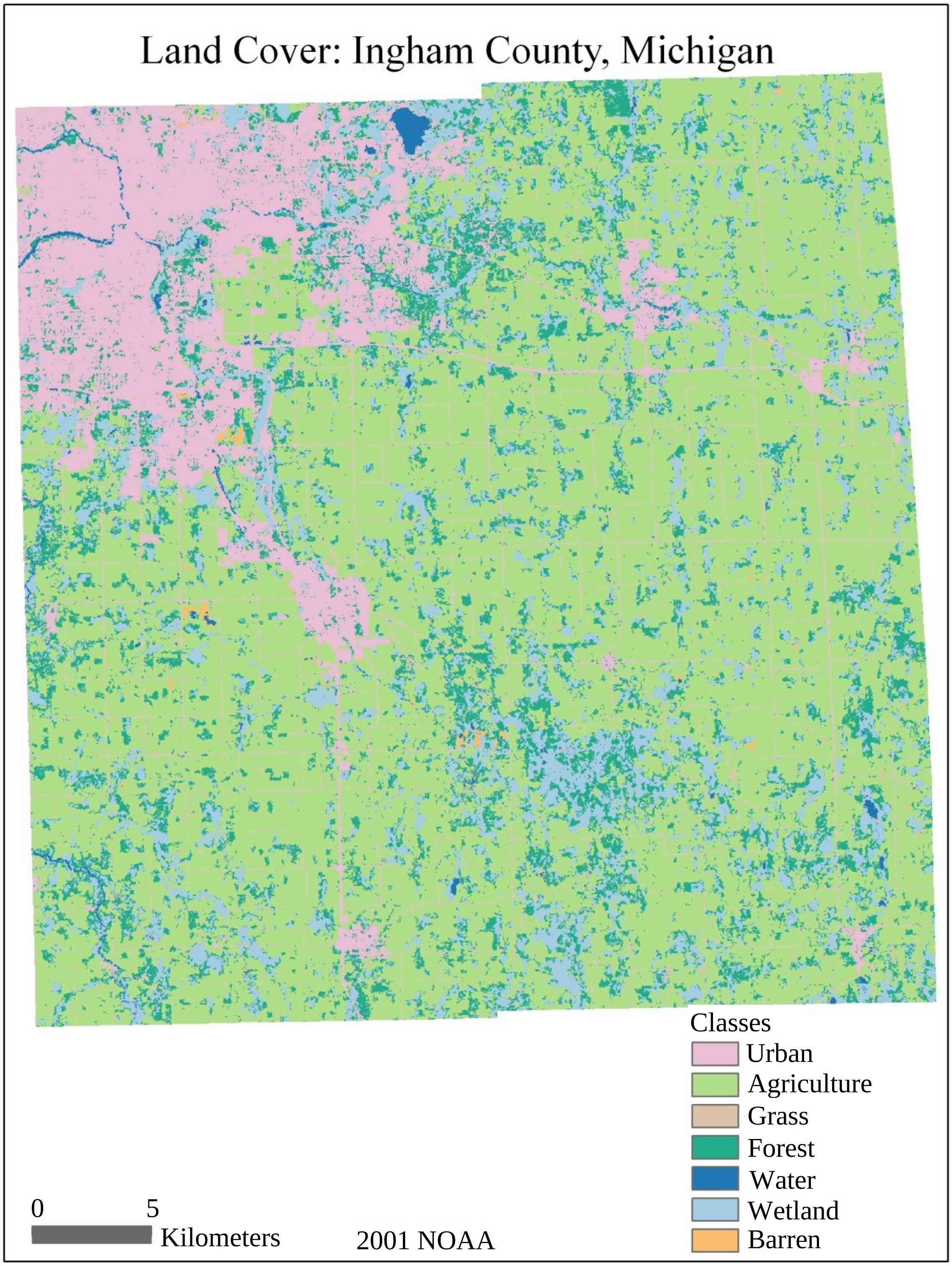

Figure 2.1 Example of ancillary data (land cover) with different classes such as urban, agriculture, and forest. The spatial resolution of the data is 30 meters by 30 meters.

** "For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this thesis."

Goodchild *et al.* (1993) implemented areal weighing for various socio-economic variables such as employment, income and population of the 58 counties of California as source zones and the state's 12 major hydrological basins as the target zones in order to conduct economic impact study for water usage and policy-making. Even though, the boundaries of the two sets of spatial units were non-coincident, they assumed that densities in the source zone were uniform. Further, they compared the result of the areal weighing method with other statistical approaches and found that the result of areal weighting had a higher mean percentage error.

Areal interpolation is closely related to dasymetric mapping of population densities (Holt *et al.*, 2004). According to Eicher and Brewer (2001), the main difference between areal interpolation and dasymetric mapping was that in the dasymetric approach, the data were not re-aggregated into desired enumeration units as they were in the areal interpolation. Various areal interpolation methods can be incorporated into dasymetric mapping in order to overcome the assumption that people are evenly distributed in the areal weighted method (Fisher & Langford, 1996), since disaggregating the population would give a better depiction of the geographical reality rather than assuming homogeneity.

2.5.2 Binary method

In previous dasymetric mapping studies, areal weighting was used as a starting point and then a filter was applied to the data using ancillary information. The ancillary data often consist of land use and land cover data which indicates where the uninhabitable areas are, then excludes these areas and redistributes the population in the remaining areas (Maantay, 2007). This method, referred to as the "binary" method (Eicher & Brewer, 2001), commonly uses remotely sensed data or land use/land cover polygon data as a filter or mask to identify the location of “occupied” and “unoccupied” areas. It is considered binary because the surface land is identified as either inhabited or uninhabited. Uninhabited land includes parks, cemeteries, water bodies, and so forth. Figure 2.2 gives a better illustration how the land use and land cover data were used as a filter or mask to eliminate the areas considered to be uninhabitable or non-residential areas.

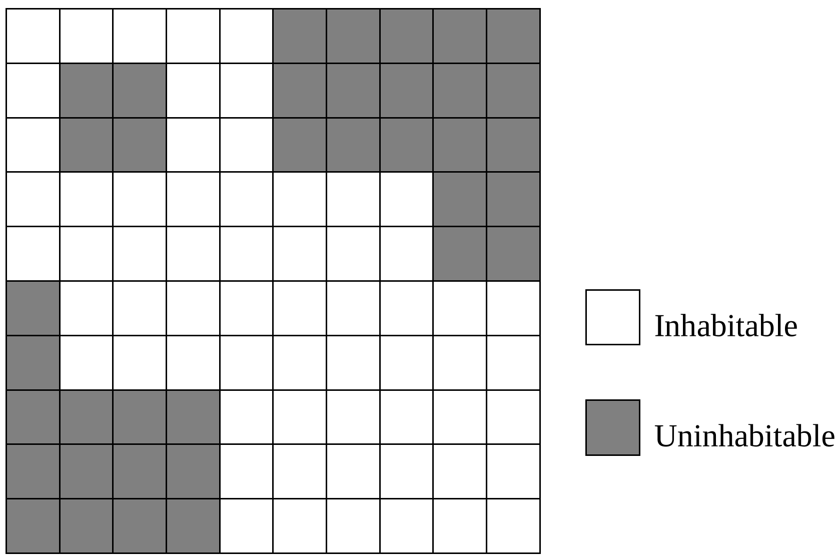

Figure 2.2 Illustration of binary method showing inhabitable and uninhabitable land cover in a grid format.

Langford and Unwin (1994) used TM images to classify the populated areas of residential housing as “occupied” while all other land uses are recognized as “unoccupied.” The dasymetric map they produced overcame the difficulties related with choropleth map representation such as generalization, arbitrary zonation and the resulting false spatial continuities/discontinuities. Their simple binary division was readily integrated into a GIS based on the raster data structure.

The results of filtered areal weighting generally are an improvement compared to simple areal weighting; however, there are still considerable deficiencies in this method. For example, the filtered method assumes that all residential areas have the same density of housing units or population even though; it is unlikely to find all residential areas to be of homogenous size and density. In reality, there are high-rise, low-rise buildings, and suburban housing developments, yielding a heterogeneous population density. Even those zones described as non-residential might often have population, but the binary method completely eliminates them (Maantay *et al*., 2007).

2.5.3 Three class method

Refinements of the binary method, including a three-class (class percent) method, were implemented by Holloway *et al.* (1996), who mapped Missoula County, Montana. They assigned 80% of each tract population to urban polygons, 10% to open polygons, and 5% each to agricultural and forested areas. Based on Holloway *et al.* (1996), Eicher and Brewer (2001) applied the three class method weighting scheme to assign population or housing data to three land-use classes within each county. In their study, they assigned a predetermined percentage of a county’s population for a given land-use area in that county. For instance, if the three major land use categories are said to be urban, agricultural/woodland, and forest, these classes were assigned 70%, 20% and 10% respectively. However, the assigning of the percentages is fairly subjective and usually not based on any demographic evidence (Maantay, 2007).

Eicher and Brewer (2001) acknowledge that the major weakness of the three class method was that it did not account for the area of each particular land use in each county. For instance, if a county had only one or two small urban polygons, 70% of the population would still be distributed in those areas. This could be misleading since it results in the urban areas being given higher densities and the other land use areas lower densities. The problem with this approach is that, although it accounts for the difference between land-use classes, it did not recognize the differences within a land-use class. Not all residential areas have identical population densities (Liu *et al*., 2006).

Mennis (2003) used remote sensing TM and derived ancillary data to redistribute population. But instead of binary assignment of population based on residential/non-residential pixel classification, he used a three-tier classification of urban land-cover as ancillary data. This is similar to what Eicher and Brewer (2001) described as the three-class method. Wu and Murray (2005) also used ETM+ images for estimating urban population density, but the information used was the residential impervious surface fraction as an effective replacement for land use/land cover data. Impervious surface fraction, calculated as the proportion of impervious surface over a small area has been found to provide more information about built-up areas compared to land use and land cover classification (Ji & Jensen, 1999). Especially, for population estimation in residential areas, the impervious surface generally corresponds to housing, which serves as indicator of people (Wu & Murray, 2005). However, satellite imagery also has limitations because it cannot differentiate between residential buildings, industrial buildings, and commercial buildings. Additional information, such as municipal zoning, is necessary to specify residential areas from non-residential.

2.5.4 The limiting variable method

The limiting variable method is based on the three-class method, but it differs by setting a threshold density that limits the population assigned to each category of land use polygons (Maantay *et al*., 2007). Eicher and Brewer (2001) used the approach of the limiting variable method according to the work of Wright (1936). The first step they stated on this method is to distribute the data by simple areal weighing to all inhabitable land-use types within a county (urban, agriculture/woodland, and forested polygons), and subsequently setting thresholds of maximum density in each habitable category of a county. If any of the land use polygons exceeded the predefined threshold for its class of land use, the remaining data would be removed and redistributed to the other land use polygons in the study area (Eicher & Brewer, 2001).

For example, if an enumeration unit has 5,000 people with an area of 50 square kilometers and if this area has the same land cover, then to determine population density of an enumeration unit the following equation would be used (5,000/50 or 100 people per square kilometer). However, if the land cover is different, the threshold value for the maximum density allowed in each habitable category should be specified. Eicher and Brewer (2001) limited the values to 15 and 50 people per square kilometer for forested and agricultural/woodland categories, respectively. If the value of population density exceeded the maximum threshold to those categories, then the remaining population were removed from the area and assigned to other categories. In this example, because the value of 100 exceeds both thresholds, the remaining values have to be distributed to the urban category uniformly (Slocum *et al*., 2009).

Eicher and Brewer (2001) evaluated the accuracy of the three approaches (binary, three-class, and limiting variable method) for assigning population in the dasymetric zones using both statistical analyses and visual presentations of error. They found that the limiting variable method produced significantly lower error when distributing the population than maps made using other methods.

2.6 Land use/land cover classification

Land use and land cover are often used as interchangeable terms, but they have unique definitions. Land use refers to the use of land by humans, while land cover refers to the natural or man-made surface of the earth. Land use/land cover information is a significant element for forming policies and decision making regarding economic, demographic, and environmental issues (Jensen & Cowen, 1999; Compbell, 2002).

Estimating population using remote sensing data started in the mid-1950s. The initial motivation was to address the shortcoming of the decennial census such as high cost, low frequency and, intense labor (Liu, 2003). Land use/land cover data collected by Landsat and SPOT sensors were used as ancillary information to study population distribution. Remote sensing data cannot indicate population density directly, but it can be used to describe the morphology of developed and non-developed areas (Liu *et al.,* 2006; Yuan *et al*., 1997). Generally, there is a positive relationship between population density and the degree of urban development that has been acquired by satellite imagery. Nevertheless, there are several issues associated with the use of remote sensing data which can result in inaccurate mapping of populations (Mennis, 2003; Liu *et al.,* 2006).

Compbell (2002) and Forster (1985) explained that errors are presented in any land use/land cover classification. These can be caused by misidentification of parcels, excessive generalization, registration errors, attenuation of electromagnetic signal due to the earth's atmosphere, and variations in sensor calibration and platform-target geometry over time. In addition, the "mixed pixel" is another problem that occurs when the resolution of an element fall in the boundaries between two land cover and registers a digital value unlike either of the two categories represented. This will result in misclassified pixels and produce errors even in the most robust and accurate classification procedure (Compbell, 2002).

The remote sensing literature explains that accuracy measures can be derived from an error matrix or a confusion matrix (Compbell, 2002). The confusion matrix is a table that compares the classified cover type on the map and actually present on the ground or the "reference" by doing cross tabulation. This table identifies not only the overall errors for each category but also the misclassifications by category. Overall accuracy in a classified scene, expressed in percentage, is the common measure of accuracy since it represents the proportion of sites where the classified and actual land cover data are coincident (Compbell, 2002). According to Fisher and Langford (1996), the overall accuracy rarely exceeds 90%, however it is also generally acceptable if it is greater than 80%. One of the standard references (Anderson *et al.*, 1976) suggests that 85% accuracy is acceptable in mapping land cover from remote sensing.

2.7 Limitations of dasymetric mapping

Dasymetric mapping provides a refined estimation of population distribution in the real world but, "it is not a totally adequate solution" (Langford & Unwin, 1994, p.22). Additional errors are introduced in each step of the process, which produce uncertainties in mapping population (Yuan *et al.*, 1997). These errors could be explained in three ways. First, Fisher and Langford (1996) discussed inaccurate land cover information could greatly affect the accuracy of the dasymetric mapping. For example, determining the pixels of the residential housing may be confusing sometimes as houses could have the size of the pixel and are mixed with other land cover classification. Second, there is no specific rule for its implementation. Any available sources of spatial information that can provide internal homogeneity in the source zones and appropriate methodology can be used (Langford, 2003). Moreover, the weighting for population distribution is subjective. Third, uncertainty present in all geospatial information as it is an abstraction of reality that makes complex geospatial information usable and understandable (MacEachern, 1992; Roth, 2009).

This chapter discussed the background of dasymetric mapping and how it improves the limitations of choropleth maps. However, dasymetric mapping also involves its own errors and uncertainties. This research implemented different dasymetric mapping models by employing new types of ancillary data to accurately represent population distribution, and additional errors and uncertainties were evaluated for those models.

Chapter Three – METHODS

3.1 Study area

This chapter explains each data source and discusses the steps for the implementation of the different dasymetric mapping techniques by employing zoning and building footprints as ancillary information. Using case studies at a fine spatial resolution will highlight the limitations of only applying land cover data as ancillary information for estimating population distributions, and these errors will be closely investigated. Further an accuracy assessment of each dasymetric mapping would be conducted to validate using pycnophylactic property which would help to summarize the errors and uncertainties in dasymetric techniques

The study area, Ingham County, Michigan, includes a large portion of city of East Lansing, and Michigan State University (MSU). Ingham County has a total area of 561 mi^2; of which 559 mi^2 are land and 2 mi^2 are water. Lansing, Michigan's state capital is located in Ingham County and Clinton County and has a total population of 279,260 according to the 2000 U.S. Census records (U.S. Census Bureau).A large part of city of East Lansing is located in Ingham County and the rest expands to Clinton County. The case study of East Lansing used for this research only includes areas located in Ingham County. Furthermore, MSU which is located in the city of East Lansing was used as the large-scale case study area. All study areas overlap one another, which makes it easy to compare the results. My knowledge and familiarity of Ingham County was helpful for verifying study results.

3.2 Data

The data sets used for this research are from different sources, and they are meant for different purposes. As it will be explained in detail in the following sections, each data set contributes to the development of accurate population estimation. The datasets which will be discussed include: U.S. Census data, land cover data, municipal zoning data, MSU building footprints and supplementary campus population data.

3.2.1 U.S. Census data

Population data were obtained from the 2000 United States Census Bureau aggregated at different levels such as block, block group, and census tract. The data are available free of charge via the World Wide Web from the United States Census Bureau website (http://www.census .gov).

The goal of the research is to map human population more accurately by overcoming basic limitations involved in conventional population mapping. Using only census data, conventional approaches would produce choropleth or population symbol maps of population; however, as mentioned previously these maps have significant limitations. This study used the finest resolution of population data available, which is aggregated at the block level and was collected from the Census 2000 Summary File 1 (SF 1) 100-percent data. According to the U.S. Census Bureau, a block is defined as "a subdivision of a census tract (or, prior to 2000, a block numbering area); a block is the smallest geographic unit for which the Census Bureau tabulates 100-percent data (U.S. Census Bureau)."

Several blocks make up block groups, which in turn make up census tracts. Blocks typically have a four-digit number where the first number indicates to which block group the block belongs. Many blocks correspond to individual city blocks bounded by streets, but blocks in rural areas may include many square miles and may have some boundaries that are not streets (Anderson, 2000).

As the U.S. Census Bureau indicated there are errors involved in a large-scale survey of the 2000 U.S. Census that occurred through human and computer-related errors. The errors arose from different sources such as non-recording of a household or an individual person in the population, not getting appropriate information from respondents, or obtaining non reliable information. Additionally there are errors that occur during field review of enumerators' work, by misunderstanding of census questionnaires by respondents, and during electronic processing of questionnaires. Also, the Census Bureau has purposely modified some data for confidentiality reasons. According the United Stated Code the Census Bureau is prohibited from publishing results in which individuals can be easily identified

3.2.2 Land cover data

The land cover data were downloaded from the National Oceanic and Atmospheric Administration (NOAA) Coastal Services Center (CSC). The data were collected in 2001 and are available free of charge via the World Wide Web from the NOAA Coastal Service Center website (http://csc.noaa.gov). The land cover classification was derived from Landsat TM scenes with a spatial resolution of 30 meters by 30 meters. The overall accuracy estimate of the land cover data was 87.7%, which is acceptable from a remote sensing perspective (Anderson *et al.*, 1976). The land cover data for this study were produced from a preexisting classification system: the Integrated Forest Monitoring, Assessment, and Prescription (IFMAP), created under contract with the Michigan Department of Natural Resources, and recoded to the specifications of the

Costal Change Analysis Program (C-CAP). Anderson *et al*. (1976) discussed how each land use/land cover classification is made for a specific purpose in order to suit the needs of the user. As a result, few users are satisfied with an inventory. In order to satisfy the requirement of the majority of the users a certain set of criteria for evaluation should first be established and a good classification system using remote sensing techniques is also required. The intended purpose for the 2001 land cover C-CAP was to improve the understanding of coastal uplands and wetlands, and their linkages with the distribution, abundance, and health of living marine resources. The different classes of C-CAP land cover are shown in Table 3.1.

Values	Class name	Modified Anderson Reclassification Level 1	Description
0	Background	No Data	No Data
1	Unclassified (Cloud, Shadow, etc)	No Data	No Data
2	High Intensity Developed	1	Urban
3	Medium Intensity Developed	1	
4	Low Intensity Developed	1	
5	Open Spaces Developed	1	
6	Cultivated Land	2	Agriculture
7	Pasture/Hay	2	
8	Grassland	3	Grass\Shrubs
9	Deciduous Forest	4	Forest
10	Evergreen Forest	4	
11	Mixed Forest	4	
12	Scrub/Shrub	3	Grass\Shrubs
13	Palustrine Forested Wetland	6	Wetlands
14	Palustrine Scrub/Shrub Wetland	6	
15	Palustrine Emergent Wetland	6	
16	Estuarine Forested Wetland	6	
17	Estuarine Scrub/Shrub Wetland	6	
18	Estuarine Emergent Wetland	6	
19	Unconsolidated Shore	7	Barren
20	Bare Land	7	
21	Open Water	5	Water
22	Palustrine Aquatic Bed	6	Wetland

Table 3.1 Land use and land cover classification values according to NOAA C-CAP.

3.2.3 Municipal zoning data

Zoning refers to the division of land by government (usually by a municipality) for a particular purpose such as residential, commercial, agricultural, and so forth. The zoning data have been derived from the city of East Lansing's Zoning Office for the year 2009. The main purpose of zoning for the city of East Lansing is "to promote the public health, safety, convenience, economic and general welfare of the residents of the city by regulating the development of land and establishing districts in which uses of land and structures are regulated (Municode Library, 2009)."

The zoning data are ancillary information to land cover data. Since remotely sensed data could not differentiate non-residential places such as commercial and industrial areas from residential ones, the zoning information improves the land cover data by disaggregating to only residential areas. As a consequence, land cover data are more refined to only residential areas and may minimize the kind of errors generated in conventional dasymetric mapping.

3.2.4 MSU building footprints

MSU has GIS data of all building footprints on campus in shapefile format. From the 2009 data requested, usage of all buildings, including university residence halls and apartments, were distinguished. The buildings were then categorized as either residential or non-residential areas.

Building footprints provide the exact dimensions of all buildings and, therefore, were used as a substitute for land use/land cover data in the dasymetric mapping method. This is because errors and uncertainties caused by applying land cover and zoning data can be minimized. In addition, assigning the population only for residential buildings provide a better depiction of population distribution.

3.2.5 Supplementary campus population data

The number of people living in each residential building of MSU was collected to compare with the dasymetric map produced from 2000 U.S. Census data. The data were gathered in February, 2010 by contacting the managers of the university residence halls and apartments. Unlike the census, these data had not been collected per a specific date, and most of the managers stated that it was difficult to know the exact number of residents at the time of a request, as tenants are always moving in and out. Based on the responses of the managers, the data were chosen either the maximum capacity of each residential building or the number of tenants residing at the time of request. Although Spartan Village, Cherry Lane and Faculty Bricks accommodate families with children, only the number of bedrooms available in each area was reported due to confidentiality reasons. Table 3.2 shows number of people living in each MSU residence hall or apartment.

Name	2010 Residents
Case Hall	890
Wonders Hall	970
Wilson Hall	984
Holden Hall	1069
Owen Hall	800
Cherry Lane	496
Faculty Bricks	244
Spartan Village	1102
Williams Hall	186
Gilchrist Hall and Yakeley Hall	437
Landon Hall	270
Campbell Hall	240
Mayo Hall	213
Mason Hall	300
Abbot Hall	300
Phillips Hall	300
Snyder Hall	300
Shaw Hall	825
Van Hoosen Hall	200
McDonel Hall	858
Holmes Hall	1164
Akers Hall	1019
Hubbard Hall	933
Butterfield Hall	400
Rather Hall	400
Bryan Hall	400
Armstrong Hall	400
Bailey Hall	400
Emmons Hall	400
University Village	304
Total	**16,804**

Table 3.2 Number of people living in MSU residence halls and apartments in 2010.

All dormitories and university apartments had population reports from their census block of the 2000 U.S. Census data, except for Shaw, Mason, and Abbot Halls. It was discovered that Shaw Hall was being remodeled from 1999-2000 and students were not living there at the time of census data collection, so it was not included in the census data. On the other hand, the Mason and Abbot Halls were serving as residential areas for approximately 640 students in 2000. As a result, it could be stated that there might be errors in the census count.

3.2.6 Data limitations

As discussed in the previous sections, the datasets used in this research were from different years. U.S. Census data were collected in 2000 and land cover data were collected in 2001. Although these two datasets were collected only one year apart, it would have been better for the data to have been collected at the same time. In reality, there are constant changes that affect the distribution of population such as moving of people within the city and the establishment of new housing developments. Both the municipal zoning and MSU building footprint data were from 2009, and the supplementary data were from 2010. Therefore, integrating this information with the older census and land cover data affected the results obtained from the dasymetric output. The temporal difference of all the datasets was a major limitation of this research. Had the project been conducted in 2011, using the newly-collected 2010 U.S. Census data and recent satellite images, the result would have depicted a more accurate population distribution. With regards to spatial resolution, the dasymetric mapping process requires all the vector data to be converted to raster data with the same resolution of its ancillary data.

3.3 Overview of methods

The goal of this study is to extend conventional dasymetric mapping techniques by using various kinds of ancillary information such as zoning data and building data as well as land use and land cover. Many methods have been reported in the literature depending on the approach and ancillary information required. Liu (2003) grouped these methods into two categories: areal interpolation and statistical modeling. This study focused mainly on areal interpolation, which transforms data from one set of spatial units to another.

3.3.1 Dasymetric map using land use and land cover as ancillary information

Remotely sensed imagery, used as ancillary information, was integrated with the census data to apportion the population into finer grid cells. The basic description of the dasymetric technique is illustrated by the addition operation and its output map (Figure 3.1).

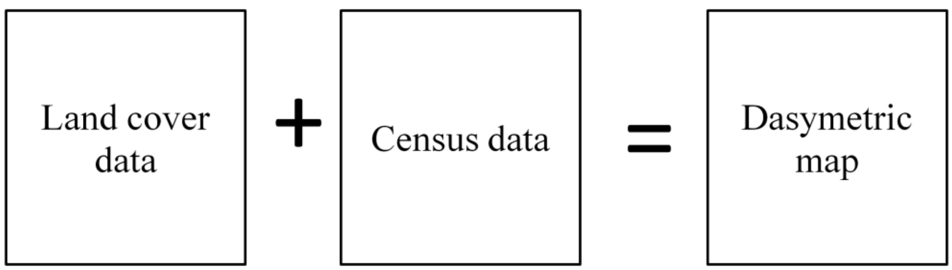

Figure 3.1 Illustration of how the land cover data and census data were integrated in GIS to produce a dasymetric map.

This research combined the methodologies of Holloway *et al*. (1996), Eicher and Brewer (2001) and Mennis (2003) by using the three land-cover classes. The C-CAP has 22 classes, but only three urban classes were used: high-intensity developed, medium-intensity developed, and low-intensity developed. This research assumed that all other categories such as forest, agriculture, and water bodies are uninhabitable by human populations, which are rarely expected to have a residential population. The main reason for this assumption was because the 30 meter by 30 meter resolution of the Landsat TM images could be fine enough to detect all residential buildings bigger than 900 square meters. Smaller buildings, however, cannot be accounted for and result in a “mixed pixel” problem, which causes error in the output dasymetric map. Eicher and Brewer (2001) reviewed different dasymetric mapping techniques, including the use of the raster-based approaches to areal interpolation, which they called the “grid three-class” method. In their study, they assigned a subjective and predetermined percentage of a county’s population to a given land use and land cover: 70% of the population to urban, 20% to agriculture/woodland, and 10% to forested area. Mennis (2003) proposed to use urban land-cover data with high, low and nonurban classes by addressing the weakness of Eicher and Brewer (2001). This research is considering the same kind of population weighting as Eicher and Brewer (2001) and Mennis (2003), but instead of assigning population to urban, agriculture and forest areas, it assigns 70% of the population to the land cover of high-intensity developed, 20% to medium-intensity developed and 10% to low-intensity developed.

The population of a cell can be estimated by the following equation, modified from (Holloway, 1996)

$$\mathbf{P_C = (R_A * N * P_A) / (E * A_T)} \tag{1}$$

Where, P_C is the population of a cell, R_A is the relative density of a cell with land cover-type A, N is the actual population of an enumeration unit (i.e., census block), P_A is the proportion of cells or cell size of land-cover type A in the enumeration unit, and E is the expected population of the enumeration unit calculated using the relative densities. E equals the sum of the products of relative density and the proportion of each land-cover type in each enumeration unit. A_T is the total square meters of all cells classified as inhabitable areas in the enumeration unit.

The relative difference in population densities among the three urban land cover classifications is one of the factors that control each census block population distribution to each grid cell contained in those census units. The land cover data of high, medium, and low-intensity developed was reclassified to 70%, 20%, and 10% respectively. A grid cell in the high-intensity developed class has a higher population density than a grid cell in the low-intensity developed class. This means any grid cell that is higher intensity developed should receive a greater share of total population assigned to a block than a grid cell with medium or low-intensity developed in the same block. For example, if a block in the county had a population of 100 people, 70 (70% x 100 people) of those people would be assigned to the portion of that block that was classified as a high-intensity developed area, followed by 20 (20% x 100) and 10 (10% x 100) in the medium and low-intensity developed areas, respectively.

To facilitate the dasymetric process, vector-based census block data must be converted to a grid cell size with the same resolution as the raster-based land cover data. The resolution of the land cover grid cell (30 meters by 30 meters) serves as the resolution for the eventual dasymetric map. The resolution of the grid cell size should be fine enough to capture the desired spatial variation of population within the area of interest. If the size of the grid cell is greater than any small area of the population data (census block) then, the data would be lost in the process of vector-to-raster transformation.

The expected population of a census block was calculated using cross-tabulated areas between two datasets and by summing the products of the relative densities of high, medium and low-intensity developed and the proportion of each urban land cover type within each census block boundary. The output was in a table format, which requires joining it with the census block vector data in order to acquire a geographical location. The total number of cells in the inhabitable area of the enumeration unit is also calculated using cross-tabulates areas. The output table reports the total counts of all those pixels classified as inhabitable areas of the land cover by removing other classes such as water, and forests. Furthermore, the expected population and the total number of cells in every enumeration unit were transformed into raster data. The overall structure of the methods was summarized in the following diagram (Figure 3.2).

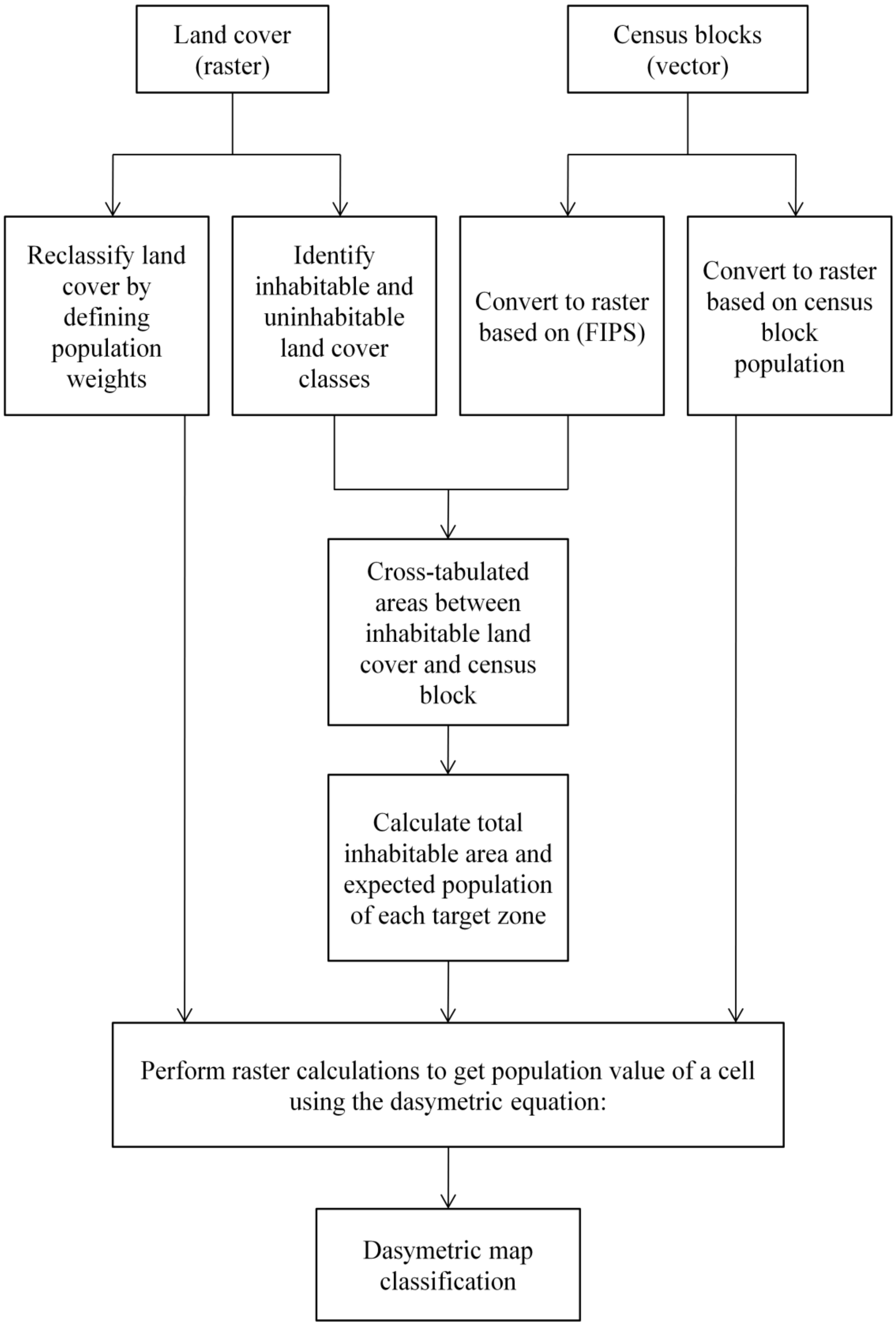

Figure 3.2 Flow diagram of the proposed dasymetric mapping process.

3.3.2 Case study 1: Dasymetric map using zoning data as ancillary information

The technique of integrating zoning data with the land cover data was applied to a case study in the city of East Lansing in Ingham County. A dasymetric map using zoning data incorporates census block, land cover and zoning data. The basic description of the zoning dasymetric technique is illustrated by the addition operation and its output map (Figure 3.3).

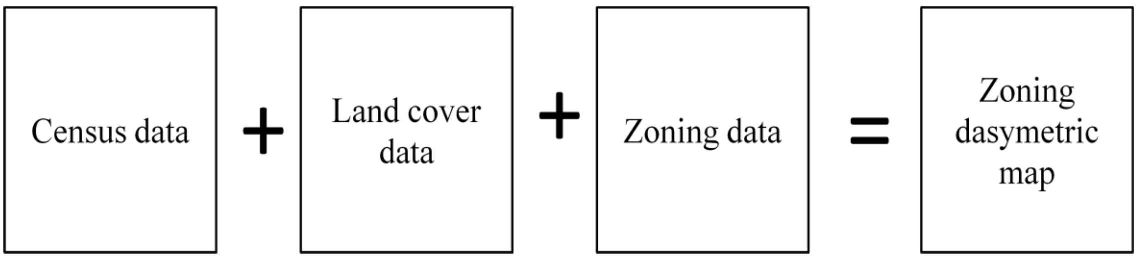

Figure 3.3 Illustration of how the census, land cover, and zoning data are integrated in GIS to produce a ZDM.

The zoning classification was assigned by the city of East Lansing for the different uses of the land. Table 3.3 shows the different categories of zoning, accompanied by a description of each code. All codes with a character of “R” were aggregated into residential activities depending on the number and density of residential units. Additionally, all other codes were aggregated for non-residential areas, including business district, commercial district and so on.

Codes	Categories of zoning
B100	B-1 General Office Business District
B200	B-2 Retail Sales Business District
B300	B-3 City Center Commercial District
B400	B-4 Restricted Office Business District
B5	B-5 Community Retail Sales Business District
BC00	Business, Community - Dewitt Township
C000	C - Community Facilities District
C-1	Commercial District - Meridian Township
D	Development District - Bath Township
M400	Multiple Family - Dewitt Township
OIP0	OIP - Office Industrial Park
P000	P – Parking
R100	R-1 Low Density Single-Family Residential
R1RO1	R-1 Low Density Single-Family with Rental Overlay
R1RO2	R-1 Low Density Single-Family with Rental Overlay
R1RO3	R-1 Low Density Single-Family with Rental Overlay
R200	R-2 Medium Density Single-Family Residential
R2RO1	R-2 Medium Density Single-Family Residential with Rental Overlay
R2RO3	R-2 Medium Density Single-Family Residential with Rental Overlay
R300	R-3 Single Family and Two Family Residential
RA00	RA -Residential Agricultural
RDD	Multiple Family - Low Density - Meridian Township
RM08	RM-8 Planned Unit Development District
RM14	RM-14 Low Density Multiple Family Residential
RM22	RM-22 Medium Density Multiple Family Residential
RM32	RM-32 City Center Multiple Family Residential
RM54	RM-54 University Oriented Multiple Family Residential
U000	U – University

Table 3.3 Municipal zoning data categories assigned by the city of East Lansing. The zoning data help to differentiate the residential areas from non-residential areas which, the satellite images were not able to differentiate.

The zoning data were classified into residential and non-residential, converted to raster data using a binary classification, and then assigned a value of 1 and 0 for residential and non-residential, respectively. Areas that were classified as high, medium, and low-intensity developed by the land cover data had to be disaggregated further using the zoning data. Incorporating these two datasets together created three urban classes in order to weight the total population of the census block. The following equation was applied to obtain areas zoned as residential as well as the three urban classes from the land cover data. In GIS, this function multiplies the two raster's values on a cell-by-cell basis and provides an output of data that met the criteria from both data sources.

$$\mathbf{R=Z*L} \tag{2}$$

Where, R equals the output land cover data that meets the criteria of residential as well as high, medium, and low-intensity developed urban land cover classes. Z is the zoning data classified as binary information of residential and non-residential areas. L is the three urban land cover classes and others such as forest and agriculture. Figure 3.4 illustrates how the equation performs to disaggregate only the residential areas of the urban land cover. For example, even if there is a large area classified as high-intensity developed in the land cover data. The zoning data were categorized as non-residential which could be a commercial district, then, in the output land cover data this area was no longer considered as urban and did not share population.

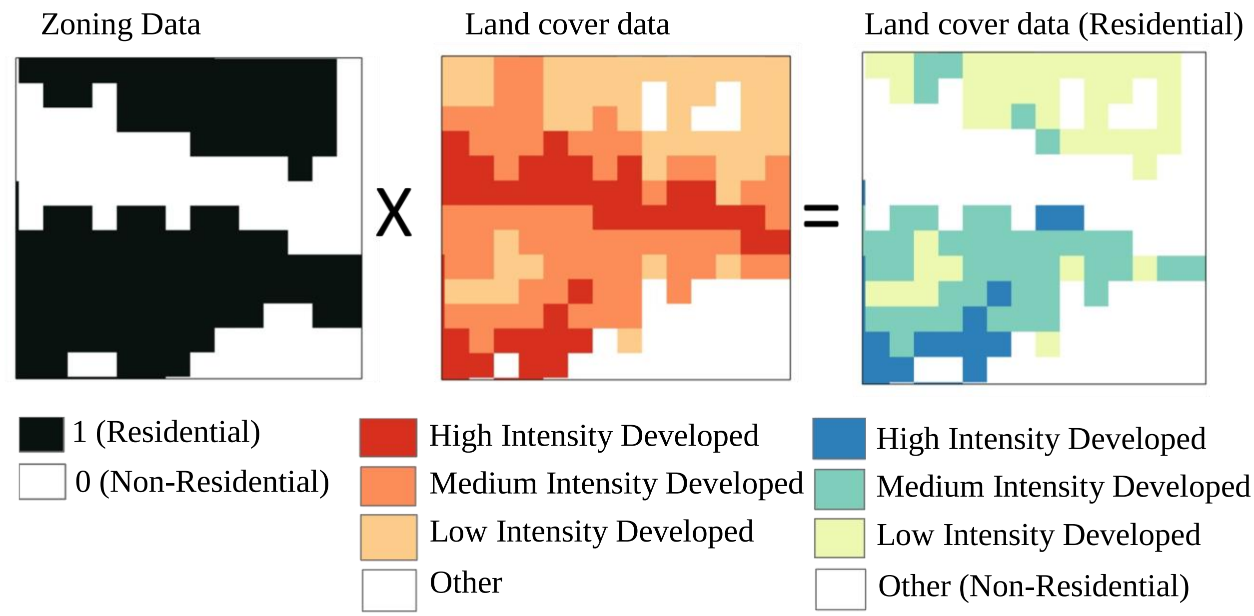

Figure 3.4 Illustration integrating land cover and zoning data and the resulting land cover refined residential areas only by removing pixels that are non–residential areas based on the municipal zoning data.

All non-residential areas such as roads, parks, business and commercial areas were removed by integrating zoning data with land cover. As a consequence, the output land cover data had improved greatly by identifying only residential areas, which in turn, provided more accurate ancillary information. In highly urbanized areas, land cover data derived from satellites may not provide a true picture of the population density, due to limitation in available pixel resolution and intra-pixel heterogeneity of urban areas (Forster, 1985). A refined land cover data using zoning information would serve an input layer to proceed to the ZDM. Finally, the equation and population weighting scheme used for ZDM is the same as the conventional dasymetric mapping though the urban classes in this technique are more refined to residential areas only.

$$\mathbf{P_C = (R_A * N * P_A) / (E * A_T)} \tag{3}$$

3.3.3 Case study 2: Dasymetric map using Michigan State University building footprint data as ancillary information

The building footprints of MSU were stored in a vector format. These data were categorized into residential and non-residential buildings and converted to raster data to allocate the population at each grid cell of all residential areas. Some buildings are small in size and have an irregular shape and can easily be lost due to their shape and size in the process of conversion from vector to raster format. The solution for the conversion problem can be reached by applying the resolution of the cell to a smaller value, which helps to maintain the information. A 10 meter by 10 meter resolution was able to maintain the shape and size of some buildings; all other input data must be with the same resolution to provide the same pixel resolution. Since the building level data contains detailed information, at this stage there is no need to use land cover data as ancillary information. If building footprint data are available for any area, it can be used as a more accurate substitute for land use/land cover data.

The population of a cell for the building level of the census block can be estimated by the following equation.

$$\mathbf{P_C = (B^* \; N^* \; P) \, / \, (E^* A_T)} \tag{4}$$

Where, P_C is the population of a cell, B is all residential buildings in the enumeration unit (census block), N is the actual population of enumeration unit, P is the proportion of cells or cell size of the residential buildings in the enumeration unit, and E is the expected population of the enumeration unit calculated using the residential buildings. A_T is the total area in square meter of all cells in the enumeration unit.

Unlike the 70%, 20% and 10% weighting of population when using land cover and zoning as ancillary data, the MSU residential buildings allocate 100% of the reported census block population to only those residential buildings in the enumeration unit. Because the data were collected from the residential buildings, themselves, it should be distributed equally. However, the land cover data had classified all urban areas either as high, medium or low-intensity developed, which would make it difficult to distribute the total population of census block to one type of class. As a result, there is no need to weight the population based on the classes of urban land cover. Nevertheless, distributing the total population to all the building equally has its own shortcomings as some buildings are high-rise while others are low-rise buildings. For future studies, Light Detection and Ranging (LiDAR) data could be integrated with zoning data and applied as ancillary data to the process of dasymetric mapping. LiDAR data have the advantage of providing the exact location and height of all buildings, thereby making land cover data unnecessary in dasymetric mapping.

The actual population of the census block should be converted to the grid cell size with the same resolution as the MSU residential buildings. The grid cell for the building, 10 meter by 10 meter resolution, serves as the resolution for the final dasymetric map. The expected population and the total area of the enumeration units for the MSU building data were calculated using the cross-tabulated areas similar to land cover and zoning data. However, the expected population of enumeration units for the residential buildings in MSU would be 100% or a value of 1. In addition, the total area was just the total number of pixels in each square meter found in that enumeration unit. Both of those values are in a table format and should be joined by using the common Federal Information Processing System (FIPS) ID into the census block of MSU and converted to raster data at the same resolution as the MSU residential buildings.

3.3.4 Dasymetric map using supplementary campus population data

The population data collected for each residential building in MSU was assigned to its specific building by adding a new column. Building population totals are the population to be modeled at every building level instead of assuming the population is homogeneously distributed in all the residential buildings of each enumeration unit.

All residential buildings were allocated a value of 1 to differentiate from non-residential buildings with a value of 0. Like the approach of dasymetric mapping using the MSU buildings as ancillary data for the census block population, the building level population was converted to raster at a resolution of 10 meters by 10 meters. The individual residential building footprint would serve as the enumeration unit to precede the dasymetric mapping.

The population of a cell controlling at building level can be estimated by the following equation.

$$\mathbf{P_C = (R * N * P) / (E * A_T)} \tag{5}$$

Where, P_C is the population of a cell, R is individual residential buildings, N is the actual population of each building unit, P is the proportion of cells or cell size of the residential buildings, and E is the expected population at each residential building. A_T is the total area, in square meters, of all the cells in each building unit.

The population data collected in February 2010 for each MSU residential building would be assigned to each pixel of that unit. The heights of the buildings were not taken into consideration, as the building footprints do not include volume. All buildings would have an internally homogenous population distribution as it was allocated 100% of the population. The expected population also uses the same weighting. The expected population and the total area were acquired by creating a cross-tabulated area, similar to the previously discussed procedures. The main difference with this approach, however, was that the areal unit was converted to raster data based on the individual building ID to define the boundary.

3.4 Accuracy assessment

Since dasymetric mapping is an areal interpolation technique, it is necessary to evaluate whether the original census data are preserved in the process of redistribution to targeted zones. One mechanism to closely investigate the errors and uncertainties related to each dasymetric technique applied on this research is to use a pycnophylactic property (Tobler, 1979). The pycnophylactic property can be defined as follows, "when the summation of population data to the original set of areal units is preserved in the transformation to a new set of areal units" (Mennis, 2003 p. 32). In their discussion of the pycnophylactic property, Langford and Unwin (1994) stressed that in the process of population re-distribution the total number of people should be preserved. They emphasized this point by stating that people are not destroyed or manufactured.

The total population of the original areal units in the census block should be preserved after the areal transformation into dasymetric output. However, if the population is not preserved, it can be stated as error. In GIS, it is easy to calculate the statistics and summarize the values of the dasymetric raster data within the boundary of the census block and report them in a table. Using the FIPS ID, the census block and the table reported from the Spatial Analyst Tool can be joined together to compare the total population of the 2000 U.S. Census block with the statistics of the summation values from the table. If the difference of the total population from the original data (census block) and the sum from the table is zero, then that means the number of people were preserved, suggesting the result is accurate. Any negative or positive values would mean that a number of people have been destroyed or manufactured in the analysis, which involves errors and uncertainties. A detailed analysis of the errors associated in this process with the percent and count error will be discussed in chapter four.

To further differentiate the performance of the different dasymetric techniques, the root mean square error (RMSE), which was applied by Eicher and Brewer (2001), Fisher and Langford (1996), Liu *et al.* (2008), and Wu and Murray (2005) was calculated. RMSE provides a summary of the difference between the “actual” (measured) and estimated population values. A large RMSE value means the errors are widely spread and small value of RMSE means the errors are packed tightly around the mean value (Bolstad, 2005). RMSE of the population estimates is defined by the following equation.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(\widehat{P}_i - P_i\right)^2}{N}} \tag{6}$$

Where $\widehat{P}_i$ is the estimated population of the census block,

P_i is the actual, census-reported population of a block,

N is the number of census blocks

Chapter Four – RESULTS

4.1 Overview

This chapter will present the results of conventional dasymetric mapping using only land cover as ancillary information. Furthermore, by applying alternative sources of ancillary data of municipal zoning and building footprints, dasymetric maps using zoning and building-level population were produced. Statistical and visual analyses of errors were implemented using the phycnophylactic property and the RMSE for each dasymetric technique was calculated.

Areal interpolation techniques were applied for population distribution in the dasymetric mapping approach. Areal interpolation methods usually use census enumeration units as a source zones and apply ancillary information for disaggregation or interpolation methods to acquire finer-scale population estimation. One benefit of areal interpolation was volume preservation, in which the total population of each source zone or census unit was preserved (Liu, 2003; Mennis, 2003; Mennis, 2009).

4.2 Conventional dasymetric mapping

Different land use/land cover types have different residential population densities. For instance, multiple-unit residential areas have high population density compared to forest and wetland. Therefore, a correlation exists between population density and land use/land cover and it is one of the bases of all methods that use remotely sensed imagery to improve the estimation of population distribution (Liu, 2003).

Figure 4.1 illustrates land cover, where high, medium and low-intensity developed urban classes and the rest of the categories were aggregated into one class for which the population was assigned a value of zero. As discussed in previous chapters, 70%, 20% and 10% of the census population had been weighted to high, medium and low-intensity developed urban classes respectively. Figure 4.2 shows a choropleth map depicting the population density of Ingham County; this map has significant limitations. The intensity of color for each census unit category was determined by population density (i.e., the darker the color, the higher the population density). As a result the distribution of population within each category is homogenous; however, in reality the land cover has a heterogeneous pattern. In addition, the MAUP described by Openshaw (1984) and the abrupt change discussed by Eicher and Brewer (2001) were manifested on this map.

A dasymetric map was produced by disaggregating the 2000 U.S. Census block data using ancillary information from the 2001 land cover data of NOAA (Figure 4.3). However, instead of assigning the population using a binary method based on residential/non-residential areas like Langford and Unwin (1994), this research used the three classes of urban land cover (Eicher & Brewer, 2001; Mennis, 2003). The dasymetric map of Ingham County shows that the urban core areas of Lansing and East Lansing do not appear to differ significantly from the vector census block map (Figure 4.2). However, in urban areas where there are parks or water bodies the raster grid population data disaggregate significantly. Sleeter (2004 p. 9) discussed "the dasymetric mapping method would be more effective in areas with more land-cover variation and less concentrated urbanization." If the land cover within a census block has the same urbanization class, the dasymetric map will have a homogenous population density in a raster grid format of that enumeration unit. The same problem of uniform population distribution as a choropleth map could exist in dasymetric maps as well.

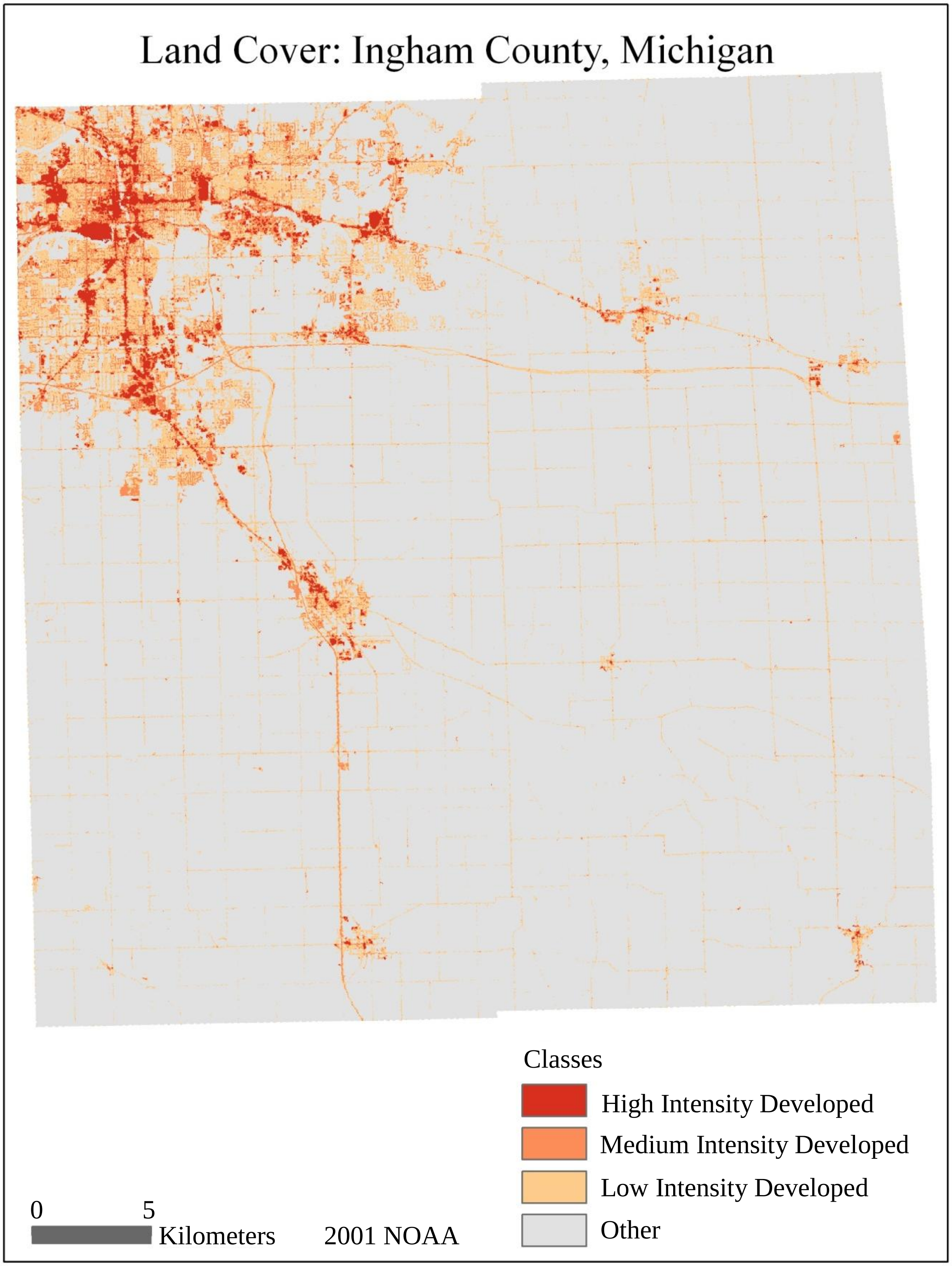

Figure 4.1 Urban land cover of Ingham County where 70%, 20%, and 10% of the census block population is weighted to high, medium, and low-intensity developed and other land cover class are assigned zero population density.

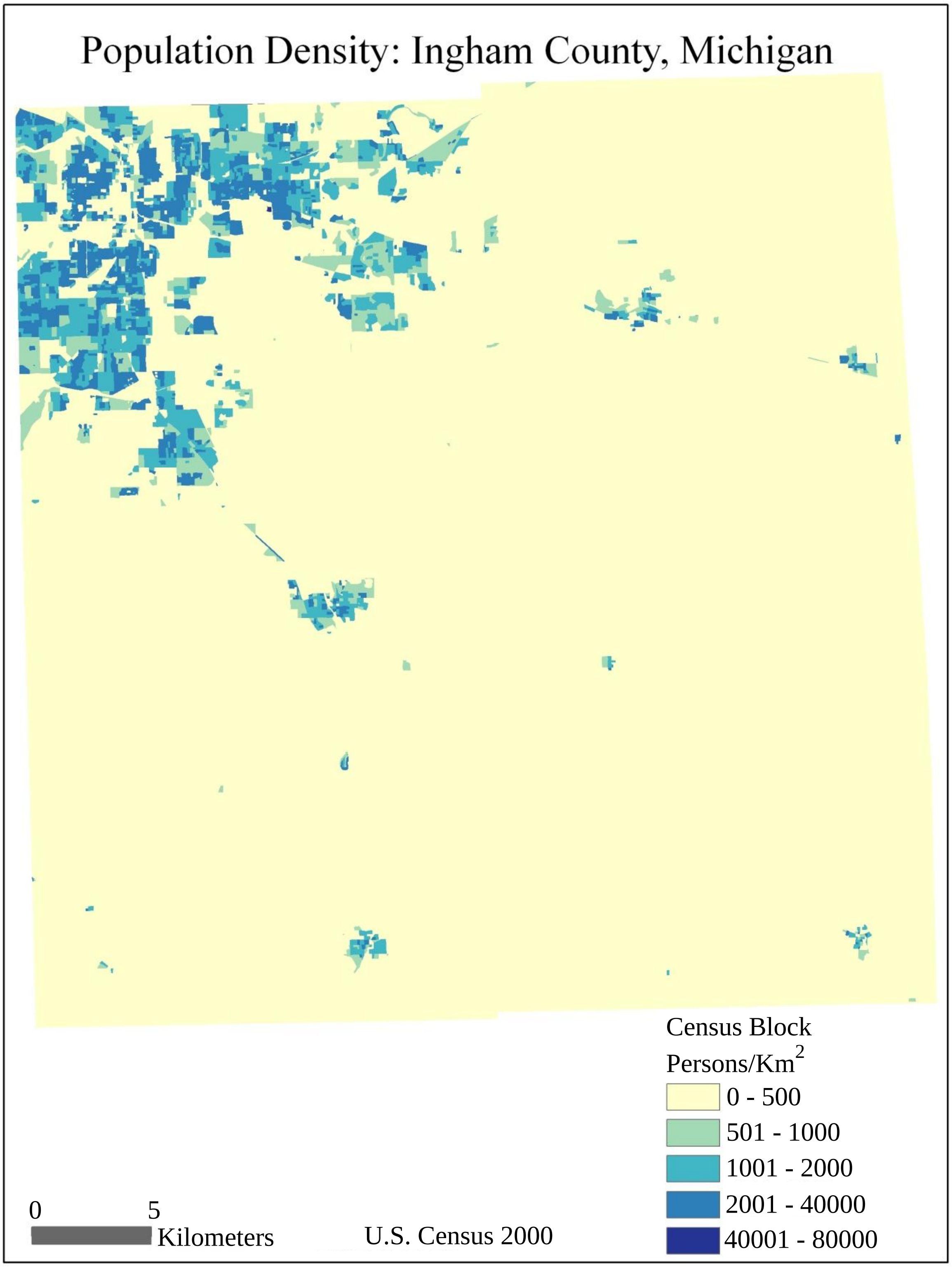

Figure 4.2 Choropleth map of 2000 U.S. Census block population of Ingham County, showing a homogenous distribution of population within each block. However, in reality the land cover and population is heterogeneous.

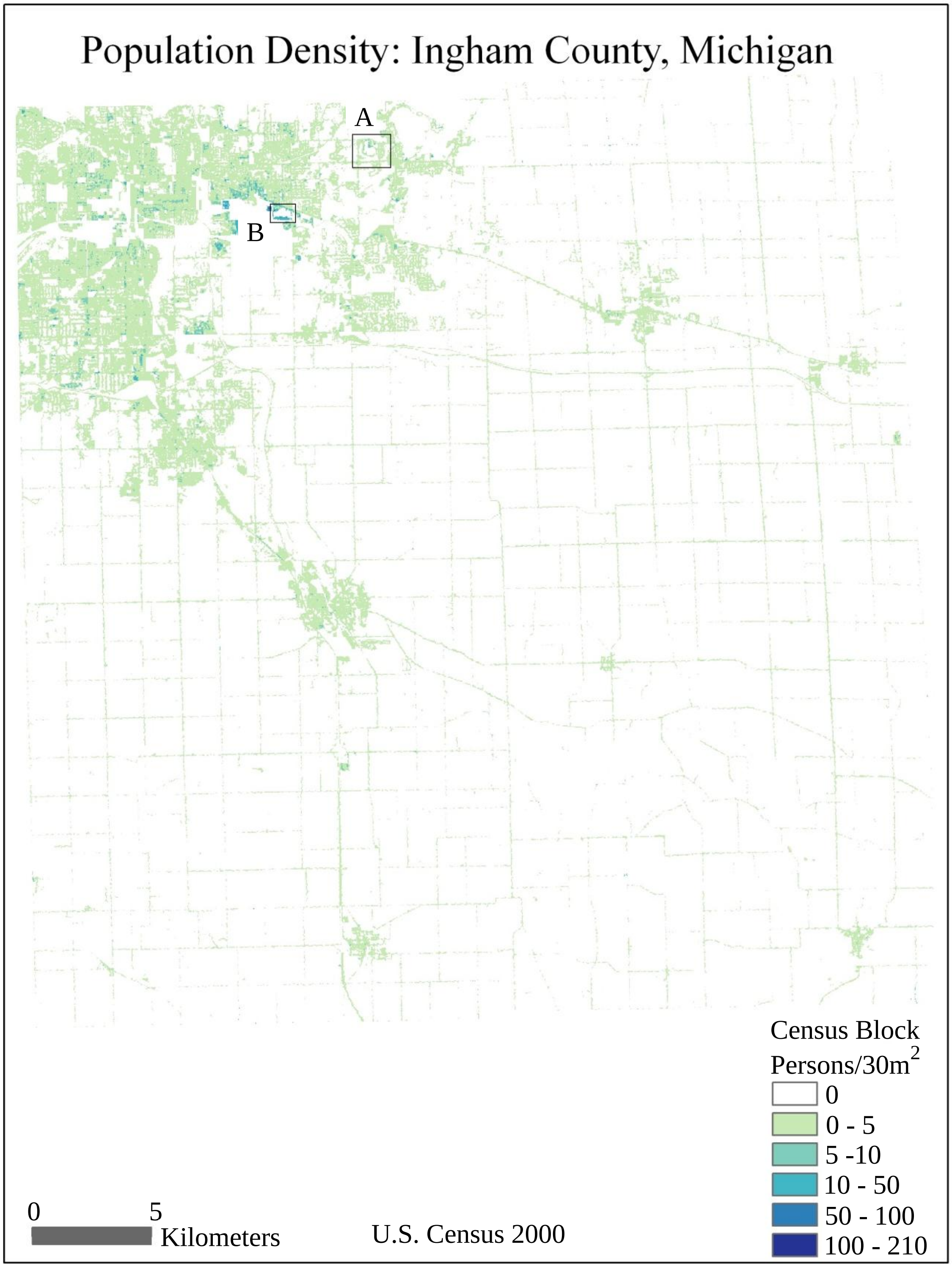

Figure 4.3 Conventional dasymetric mapping disaggregated the census block population to only urban land cover classes. Zero population is assigned to other land cover class such as water, agriculture, and forest. Box A and B indicate the area of detail shown in Figure 4.4 and Figure 4.7 respectively.

4.2.1 Separating human population from water bodies

Choropleth maps distribute population homogenously across enumeration units. Figure 4.4 presents an area of the town of Haslett in Ingham County, indicated as box A in Figure 4.3. In box A, there was one census block that includes a water body feature. The presence of a relatively large water body adjacent to densely populated urban neighborhoods caused considerable intra-block variation in both land cover and population density in that area. Figure 4.4a and Figure 4.4b compares the representation of population in raster surface with the vector census block shapefile. The dasymetric approach (Figure 4.4a) redistributed the total population of the census block to certain raster cells within that enumeration unit according to the classes of urban land cover. It was manifested around the periphery with residential houses and water body features at the center of that census block. On the other hand, the choropleth map (Figure 4.4b) indicated a relatively homogenous population distribution in which the raster surface had distributed the population only to certain sub-block regions. A photograph of this area was taken to verify these results (Figure 4.5). In addition, a Google Earth representation of the water body and the residential houses clearly indicated how much the dasymetric map had been improved compared to the choropleth map (Figure 4.6). Since people do not live in water bodies such as lakes, the dasymetric technique corrects the shortcomings of the choropleth map.

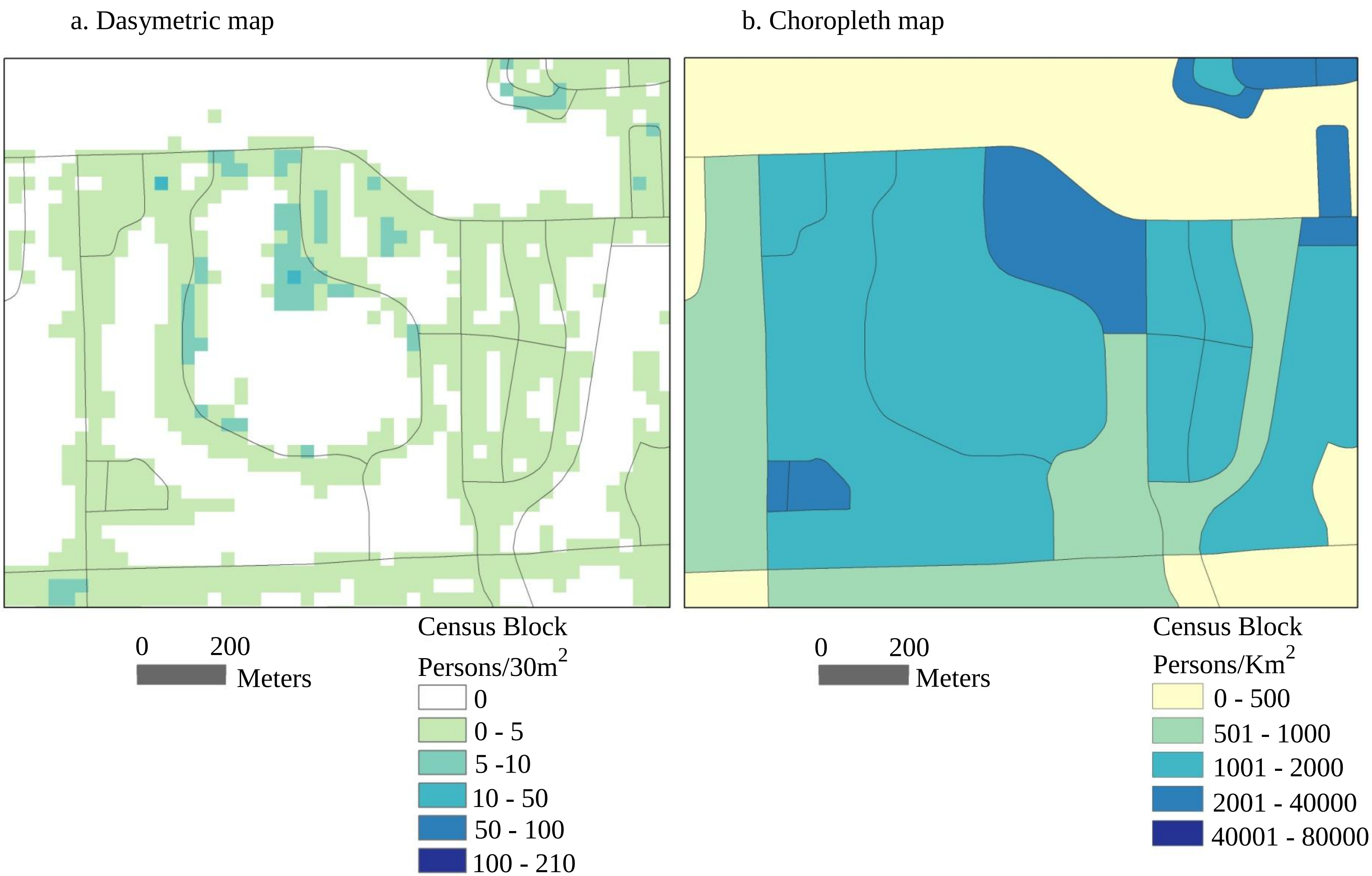

Figure 4.4 Detail of census blocks specified in Figure 4.2, as box A, showing the difference between a disaggregated residential population in a dasymetric map (a) compared to a uniformly distributed population in a choropleth map (b) in which the central enumeration unit includes a small lake.

Figure 4.5 Small lake by Lake View apartments, Haslett, Michigan. The lake and the residential building are clearly distinguished, as represented on Figure 4.4a. April 2010, photo by author.

Figure 4.6 This aerial image identifies the lake and buildings clearly. The geographic distribution of population implied by the image is best represented in Figure 4.4a. The Google Earth [*last accessed 04-20-2010*]

4.2.2 Separating human population from forest

Mapping human population in areas close to forests poses similar challenges to mapping human populations near water-bodies. Chloropleth maps show homogenously distributed populations based on census enumeration units. This problem was evident at Michigan State University, Ingham County (box B in Figure 4.3). One of the census blocks for Michigan State University had both a large portion of forested area and dormitories occupied densely by students. While the choropleth map (Figure 4.7b) failed to display this intra-unit variability, the dasymetric map (Figure 4.7a) redistributed population within each census block. Photos of these areas were taken to verify the information on the ground (Figures 4.8 and 4.9). Furthermore, a Google Earth image of the same area showed that a smaller portion of the census block was occupied by dormitories than by forest (Figure 4.10). This supports the results of the dasymetric because a smaller population density is shown in that unit area.

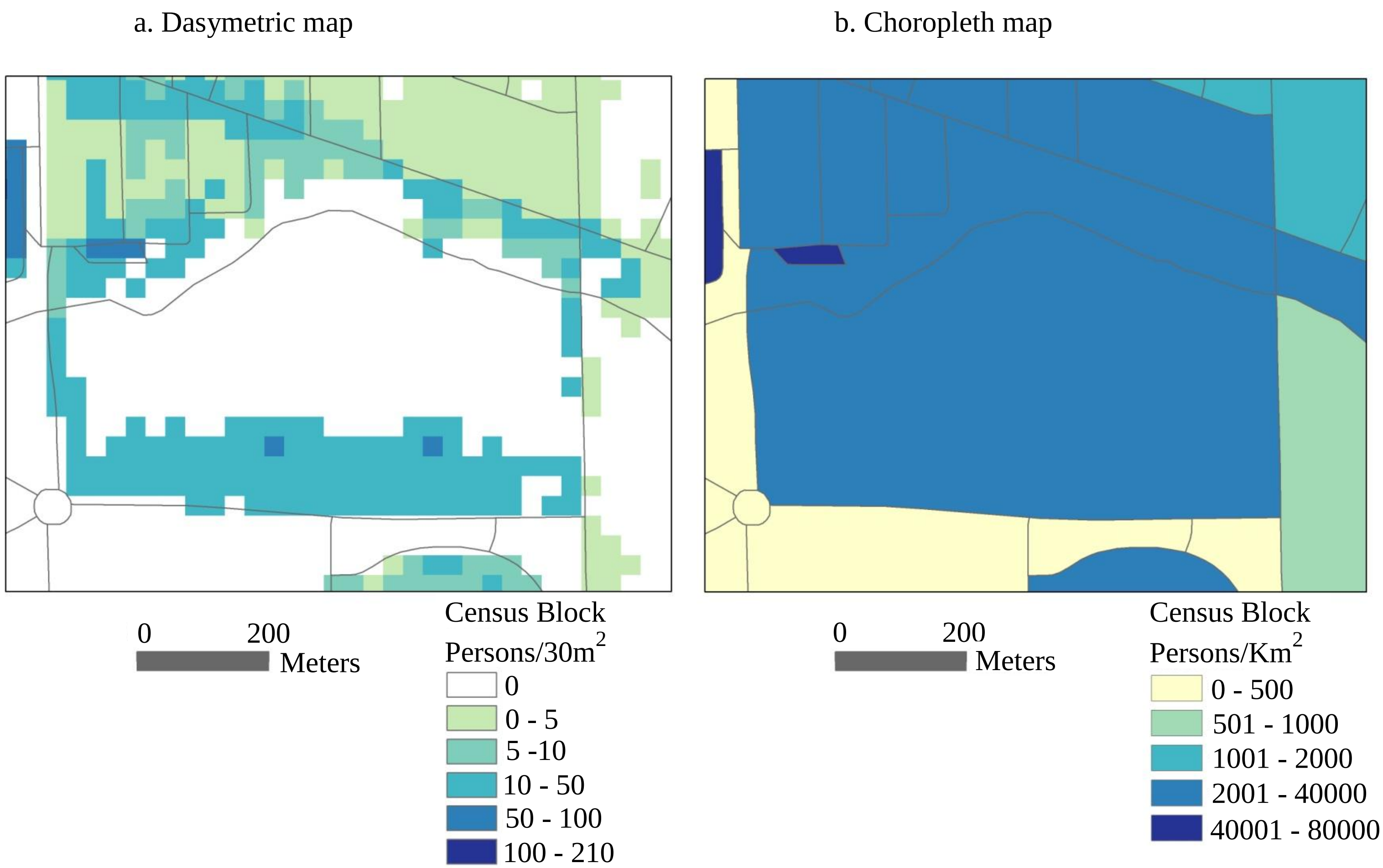

Figure 4.7 Detail of census blocks specified in Figure 4.2 as box B, showing the difference between dasymetric map (a) and choropleth map (b) representation of population density.

Figure 4.8 The choropleth map underestimates the population density in the dormitories of MSU. April 2010, photo by author.

Figure 4.9 The choropleth map overestimates the population density in forest areas of MSU. April 2010, photo by author.

Figure 4.10 Dormitories and forest aerial view at MSU. The choropleth map is misleading as large area of the census block is covered by forest. Google Earth [*last accessed 04-20-2010*]

4.3 Case study 1: Municipal zoning data as ancillary information

Integrating vector zoning data as ancillary information also helped to refine human population density when the land cover data alone were not able to classify the inhabitable area appropriately. The zoning data include all residential areas, agricultural areas, commercial districts, business districts and so on. Incorporating the zoning information into the dasymetric mapping process helped reliably distinguish residential from non-residential areas. Different zoning classes were categorized into inhabitable and uninhabitable areas and converted to raster data. The zoning data were then integrated with the land cover data to give a more accurate population density map. The resulting ancillary information was more refined than a conventional dasymetric process due to the incorporation of both zoning and land cover data.

4.3.1 Overview of zoning dasymetric map (ZDM)

The zoning data were categorized into residential and non-residential areas and then integrated with the land cover data to disaggregate all non-residential areas from the land cover information. It was observed that commercial areas connected with multiple road networks were classified as high-intensity developed. As a result, these areas would receive a higher percentage of population. However, this study assumed that population does not live in commercial and business areas. Figure 4.11 represents the population density of East Lansing by census block, Figure 4.12 shows the land cover classification, and Figure 4.13 indicates the zoning data from East Lansing. All of these input data sources were integrated into a GIS to get a more realistic population distribution. Figure 4.14 is the output ZDM that refined population distribution more thoroughly compared with conventional dasymetric maps.

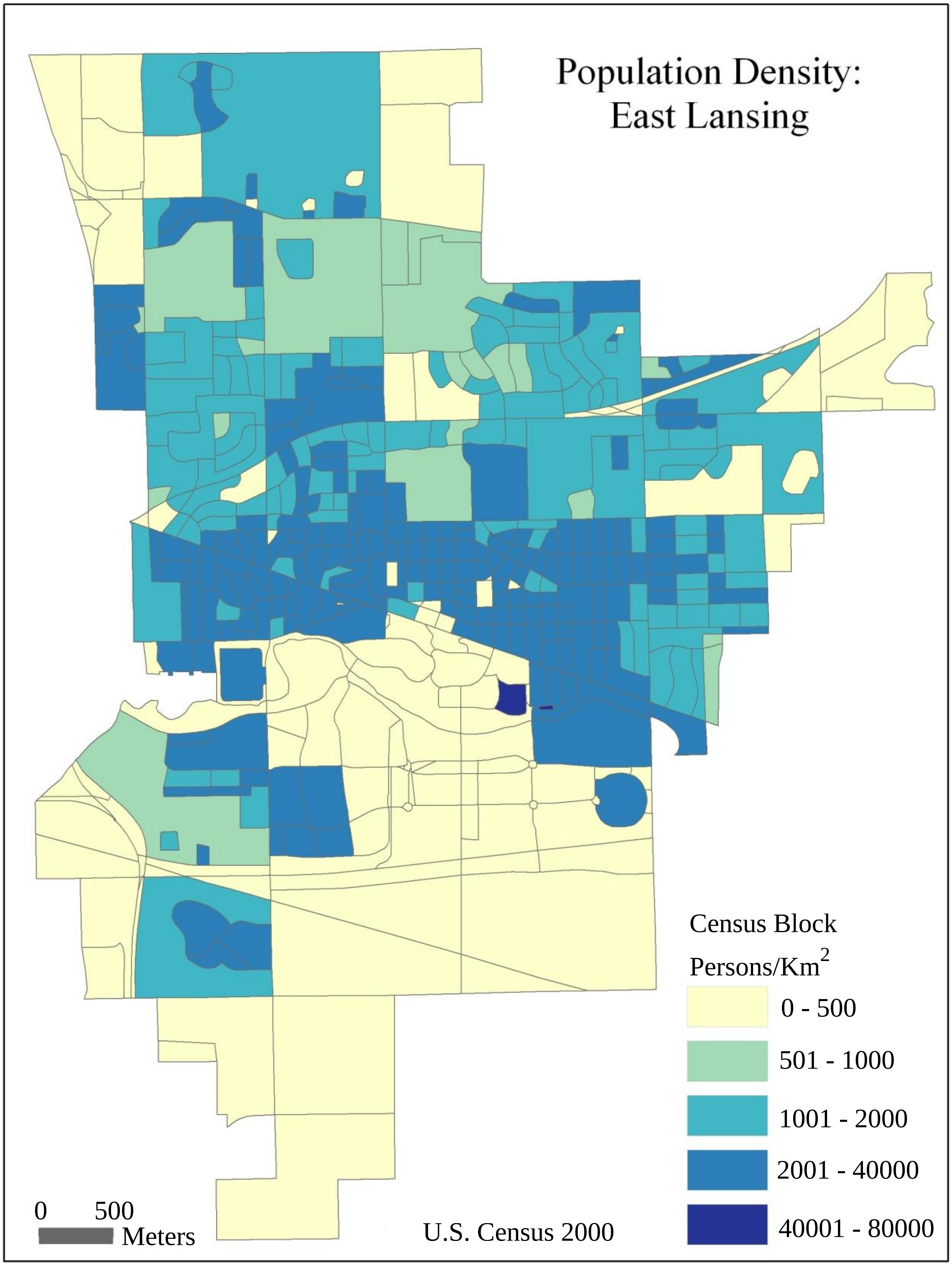

Figure 4.11 Choropleth map of East Lansing showing population density homogenously distributed within census blocks. In reality, we could not get such uniformity, as some places are non-residential, such as commercial and business districts.

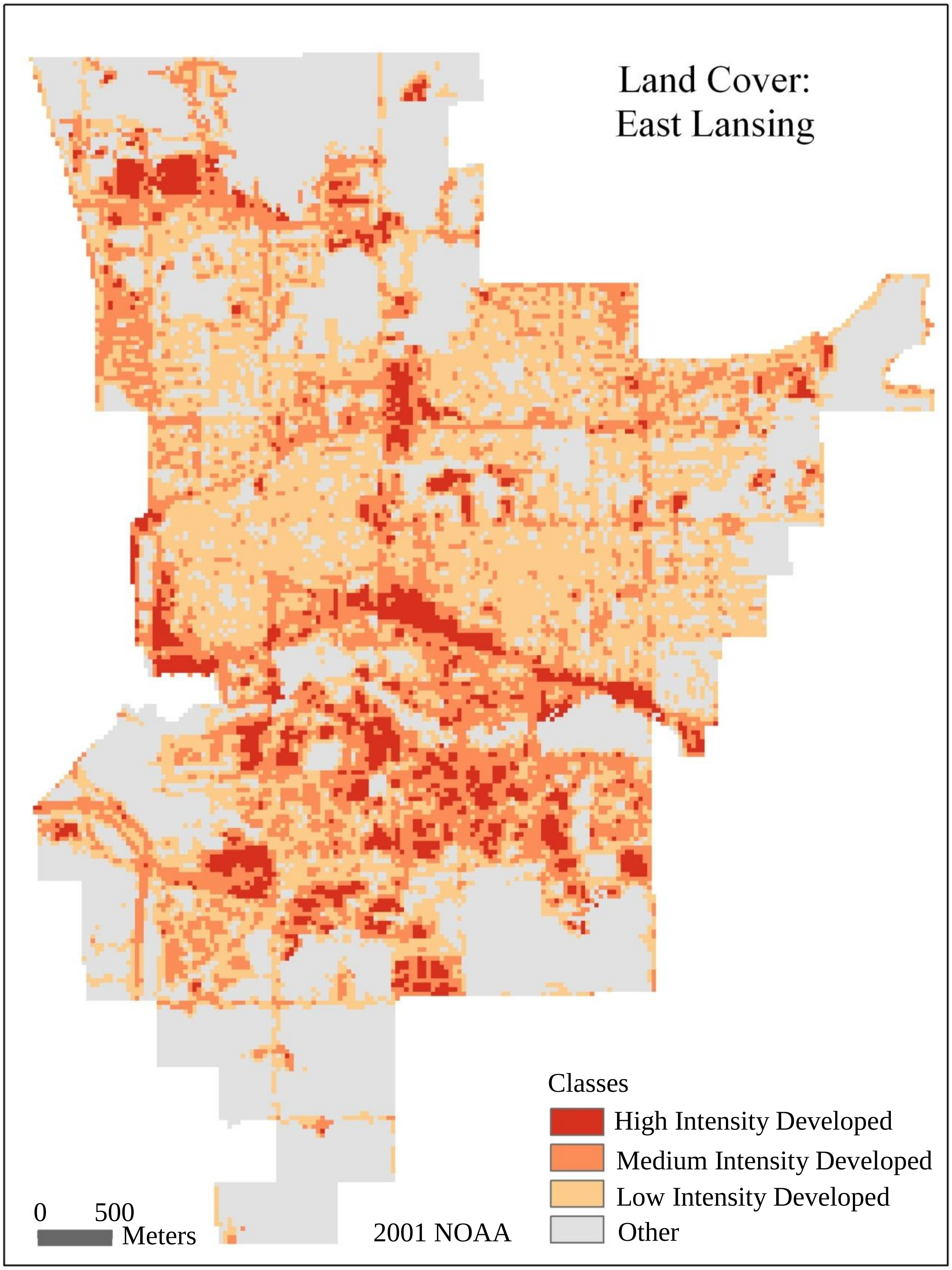

Figure 4.12 Land cover of East Lansing first integrated with zoning data where 70%, 20% and 10% of the census block population is assigned to high, medium and low-intensity developed respectively. Other types of land cover class would be assigned a population of zero.

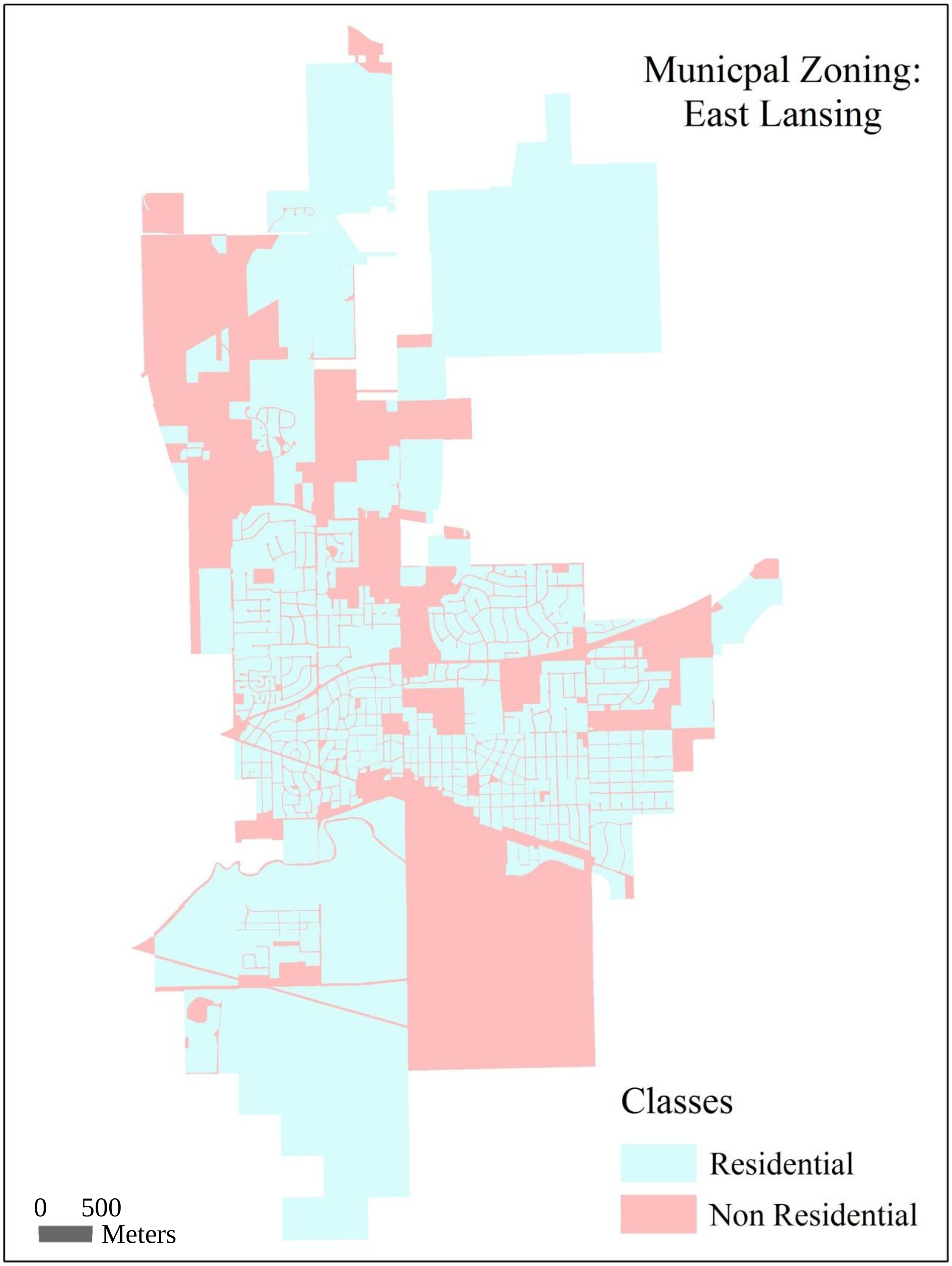

Figure 4.13 Residential and non-residential zones designated by the city of East Lansing. Residential areas have “R” in the zoning codes, and include housing types such as low density single-family and medium density multiple-family. Non-residential areas are all places without “R” of the zoning codes, such as commercial and university areas.

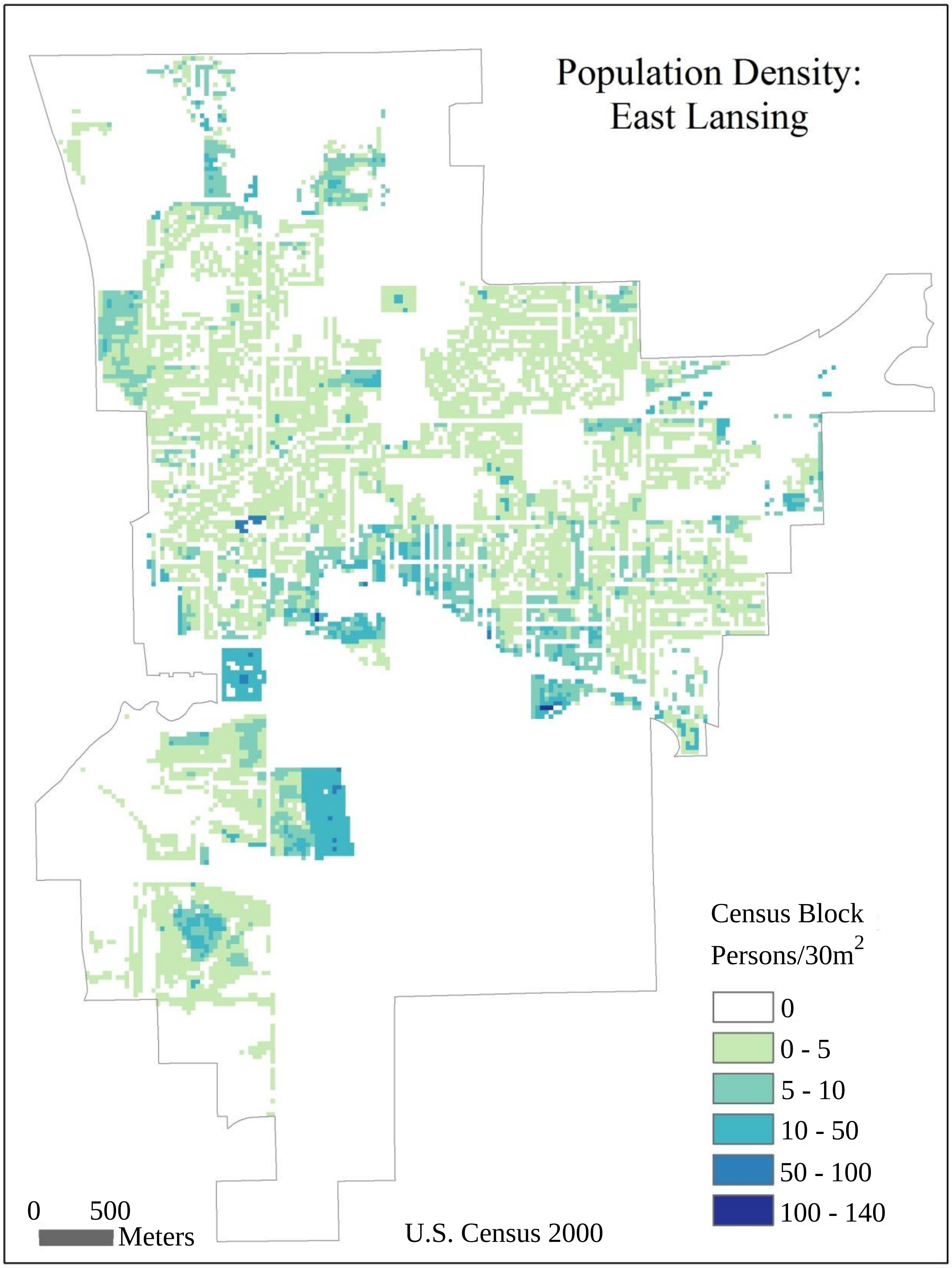

Figure 4.14 ZDM of East Lansing showing human population density more disaggregated to residential areas while all non-residential areas such as commercial or business zones and roads have been assigned zero population density.

4.4 Case study 2: Building level dasymetric mapping (BLDM)

Footprints of all buildings in MSU, in vector format, were incorporated with the aggregated census data as well as with the supplementary campus data to create dasymetric maps. First, the campus building GIS data were categorized into residential and non-residential buildings. Then, it was converted from vector to raster data in order to be in a grid format. The BLDM technique assigned the reported census block population only to residential buildings and distributed population more accurately compared with the conventional and ZDM techniques.

4.4.1 Building level population of census blocks

As mentioned previously, dasymetric mapping can be incorporated with other ancillary information to refine the distribution of population. As a result, building footprints can be a good source of ancillary information for dasymetric mapping to determine the exact location where people live (Maantay, 2007). Since land cover and zoning data were classified aggregately, errors arise from these datasets when they were applied as ancillary information in dasymetric method. For instance, a census block bounded by a residential area and some commercial areas might be classified in the land cover data as high-intensity developed. From a remote sensing perspective it could be legitimate to classify both structures in the developed urban category. However, as discussed in chapter one, distributing the population uniformly in the enumeration unit would have a misleading effect; in reality, commercial and industrial areas are not densely populated. This research had examined how zoning data would improve population mapping by identifying places of residential and non-residential data. It was found that integrating zoning data as one type of ancillary information refined the distribution of population to residential areas only and created an accurate dasymetric mapping compared to applying only land cover data.

Nevertheless, even zoning data had their own shortcomings because when the city of East Lansing categorized the zoning, it aggregated to one class. For instance, all low density-family residential areas were aggregated to one zone, although the parcels include the backyard and house building. As a result it would cause an error in the dasymetric mapping by distributing the population to all areas designated as residential.

Figure 4.15 represents the choropleth map of population density of MSU. Similarly to other choropleth maps, the distribution of population within the census block was homogenous. Applying zoning data as ancillary information improved the accuracy of dasymetric mapping; however, most parts of MSU were categorized as "university" in the zoning data, which were classified in this study as non-residential areas. It became evident that ZDM was not able to represent MSU population accurately (Figure 4.14). Figure 4.16 shows all buildings in MSU classified as residential and non-residential based on their usage as described by the MSU GIS office. Figure 4.17 presents the dasymetric output showing only building level population. Therefore, by applying building footprint data as ancillary information to the dasymetric mapping process, this study was able to demonstrate the potential of this technique for visualizing and depicting the true population density of MSU.

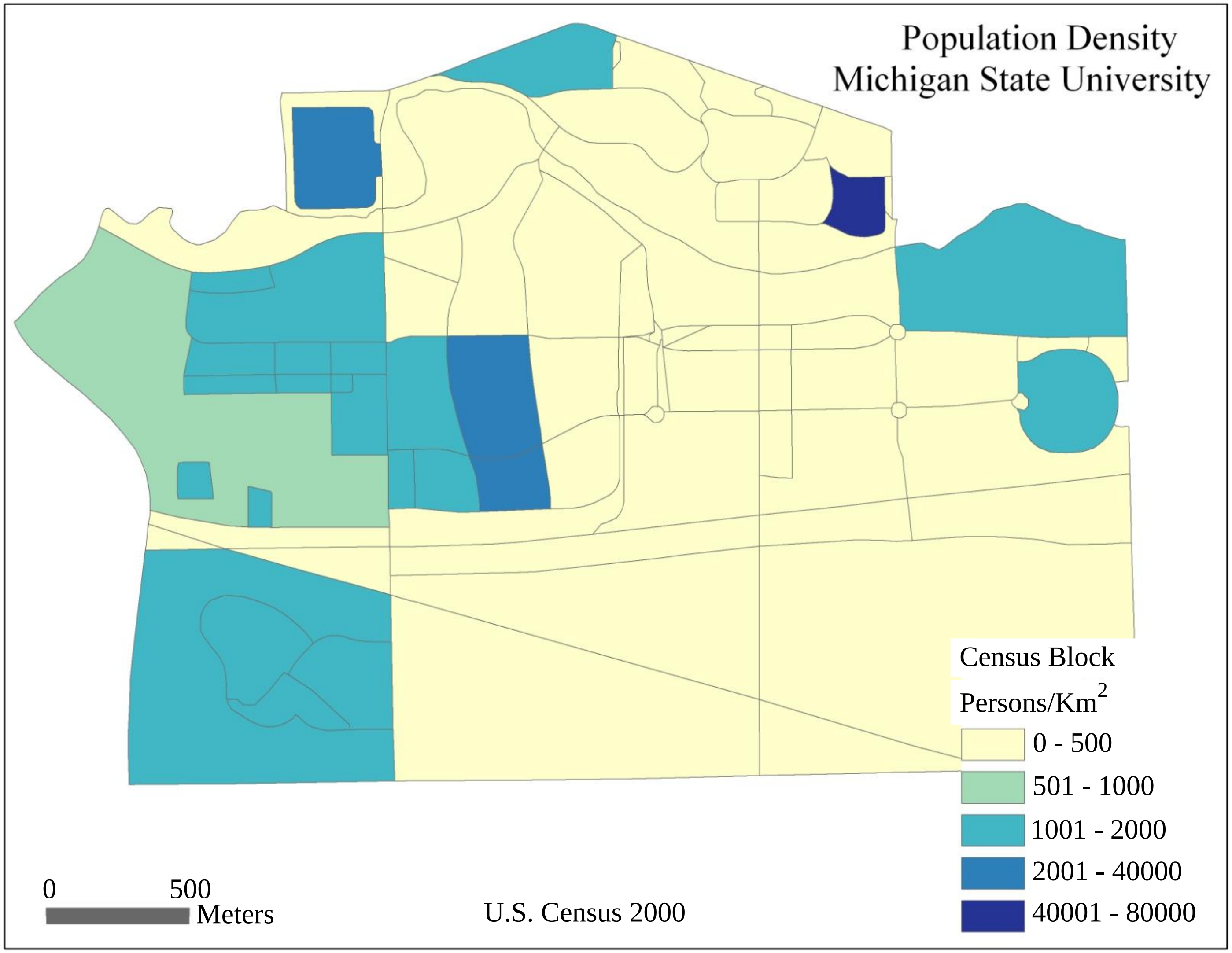

Figure 4.15 Choropleth map of MSU shows homogenous population density within census blocks. In reality, the population is only concentrated in the residential buildings of the university halls and apartments.

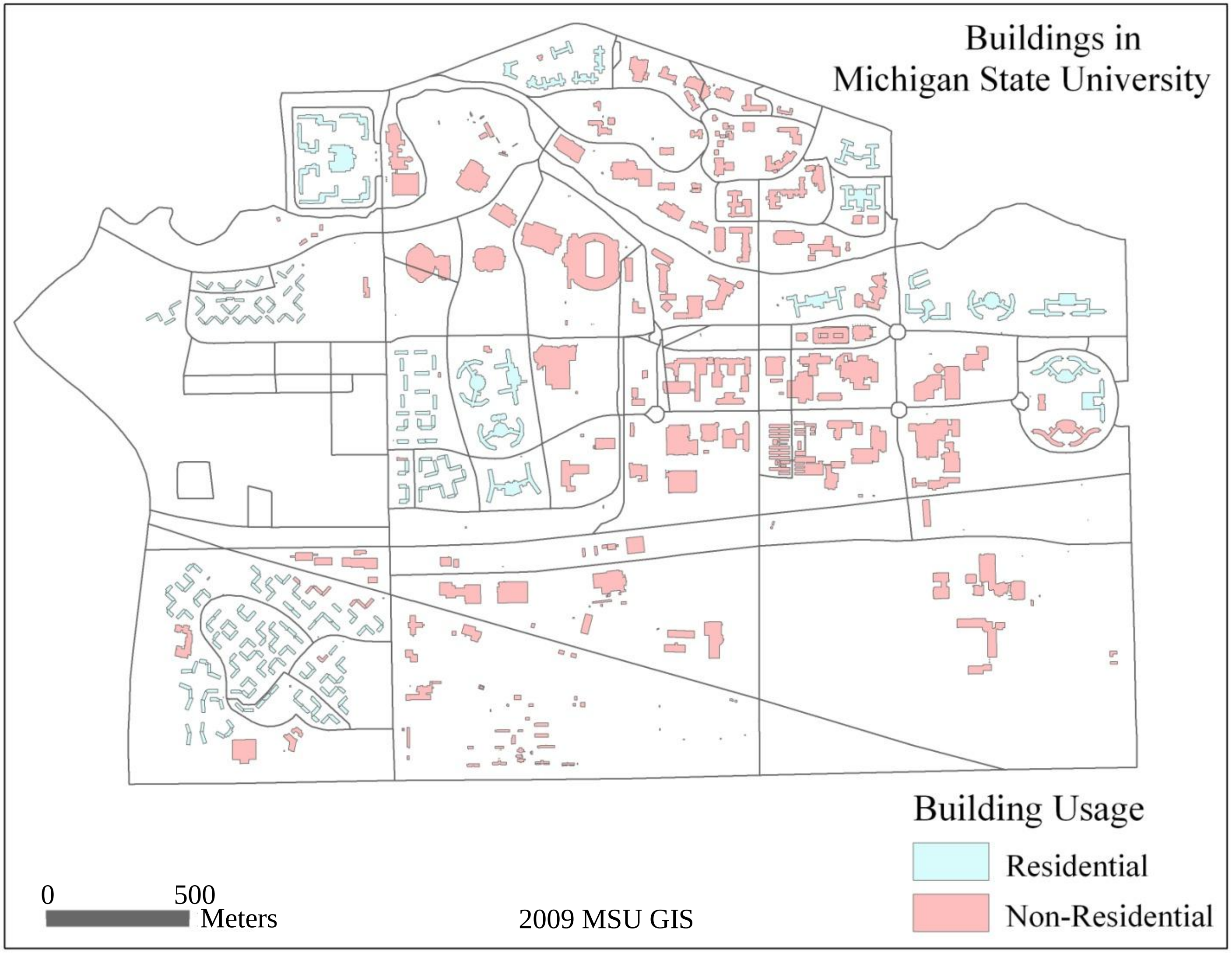

Figure 4.16 Building footprints of MSU that were used as ancillary information in the process of dasymetric mapping. Residential buildings include all university residence halls and apartments while non-residential buildings are all offices, classrooms and parking structures.

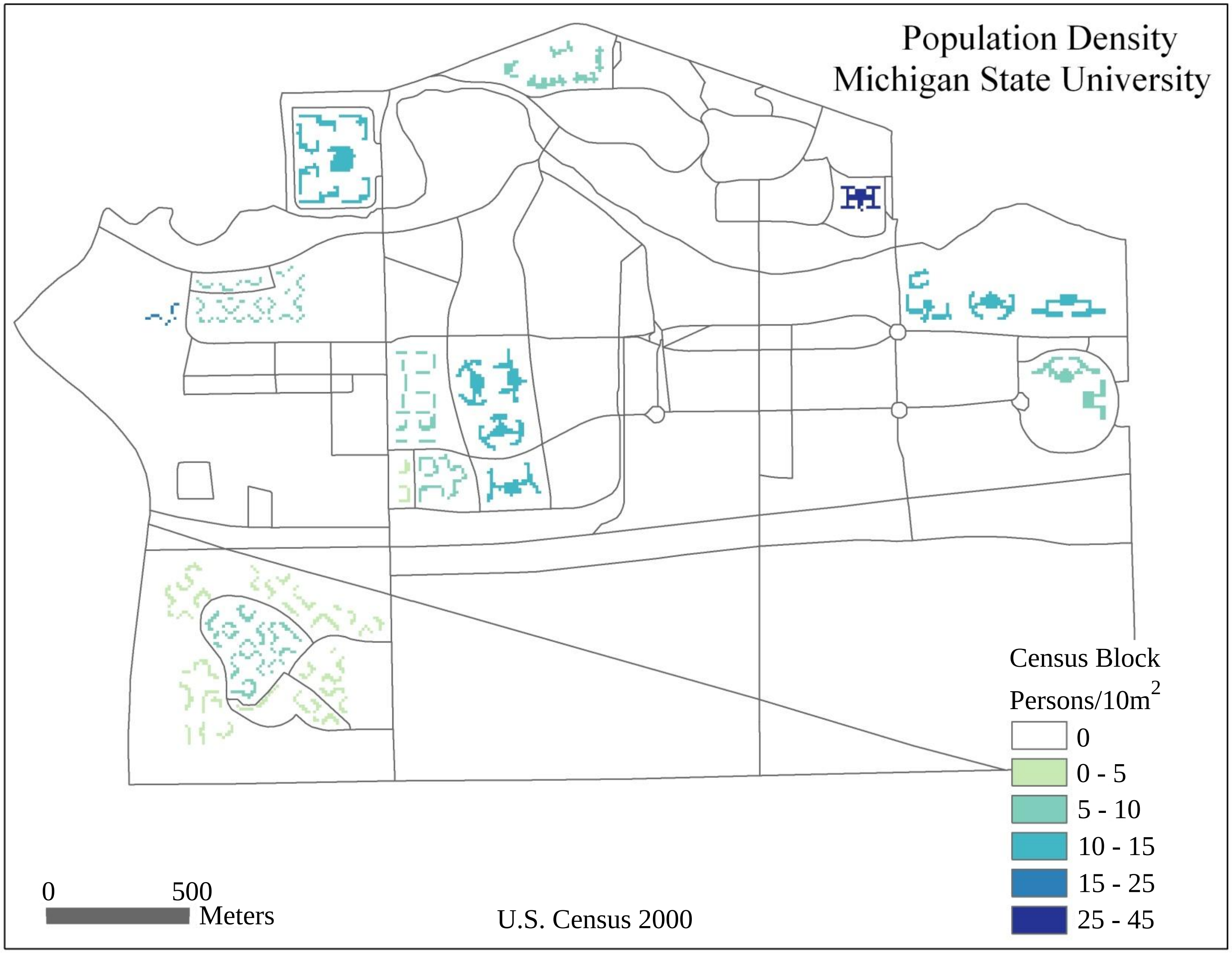

Figure 4.17 The building dasymetric map shows more refined population density, since it controls the total number of people from the census block to only residential areas of MSU.

4.4.2 Building level population of supplementary campus data

The dasymetric mapping process using supplementary campus population data was similar to the mapping process using census data. The supplementary campus data was collected for each residential building of MSU as opposed to census data that is reported aggregately at block-level. The area of each building footprint served as the enumeration unit to produce a dasymetric map based on supplementary data. These data were able to address the heterogeneity of individual buildings in population density maps. However, the supplementary data were not as accurate as the census block data because the building boundaries, which served as enumeration units, did not preserve the shape and size of the buildings when converted from vector to raster. As a result, they were not able to achieve pycnophylactic validation. Figure 4.18 shows population density of each building based on 2010 supplementary collected data.

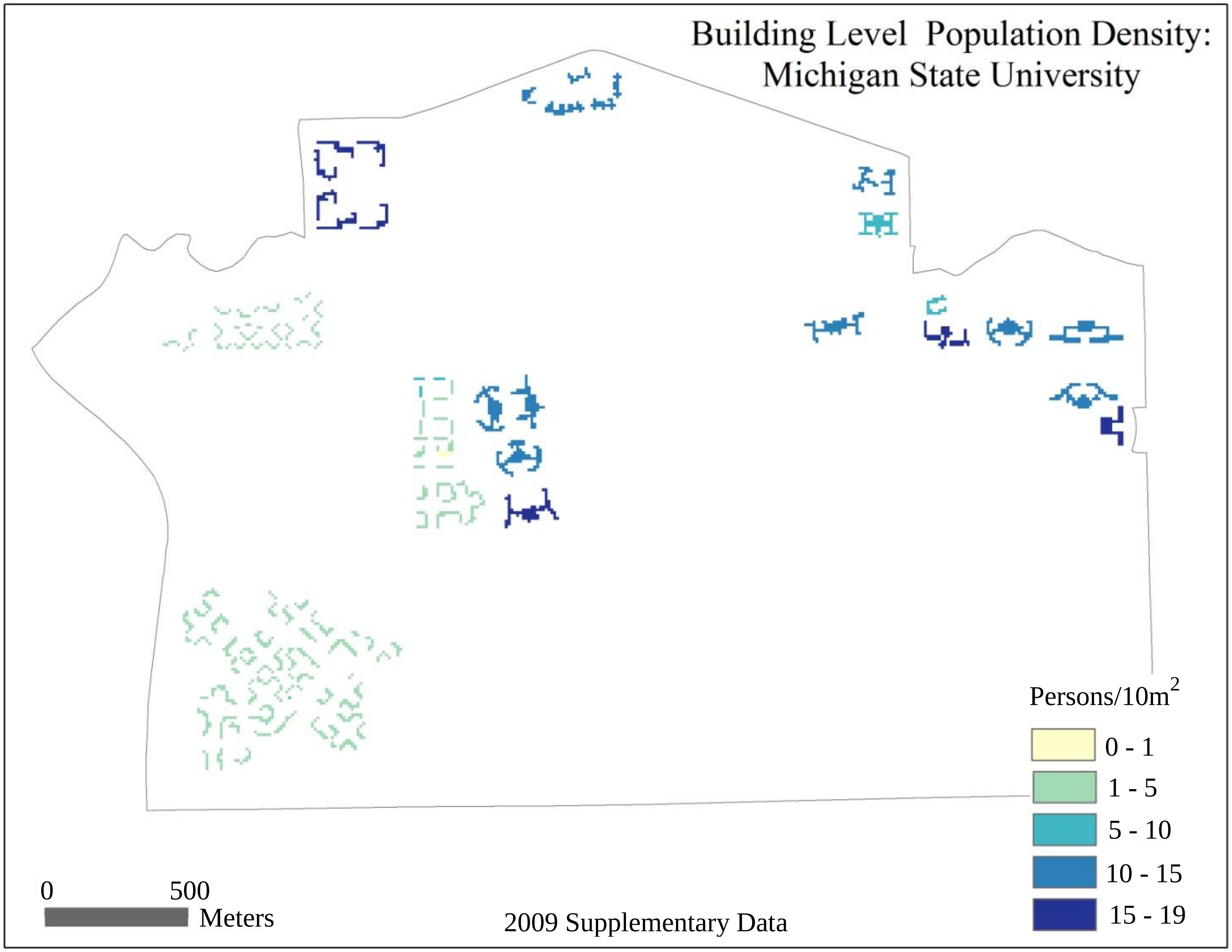

Figure 4.18 Building dasymetric map based on 2009 supplementary population data. It was able to show the population density of each individual building

4.5 Statistical and visual analysis of errors in dasymetric maps

This section explores the errors and uncertainties related to dasymetric mapping techniques. The focus of the analysis was the statistical and visual analysis of error patterns. Mennis (2009) stated that the purpose of dasymetric mapping is mainly to disaggregate data when the reporting units are not able to give finer information. As a result, some alternative means of representing the uncertainty of the data is necessary. The dasymetric mapping technique produces not only maps, but also data, by expressing the nature of the sampling and the variance that exists in the relationship between ancillary data classes and continuous surface data (Mennis & Hultgren, 2006).

4.5.1 Validation of dasymetric maps

Eicher and Brewer (2001) mentioned that a variety of methods have been used in an areal interpolation research to measure the errors. The experiment they conducted in dasymetric maps followed the Fisher and Langford (1995) technique of using root mean squared error (RMSE) and coefficient of variation to describe errors in dasymetric zones. They evaluated map accuracy using both statistical analysis and visual presentation of errors. Goodchild *et al*. (1993) summarized the areal interpolation errors by applying mean percentage error and mean absolute percentage errors. Fisher and Langford (1996), and Wu and Murray (2005) calculated the RMSE, the mean absolute error (MAE), and coefficient of variation to assess the accuracy of their experiment.

This study used the Tobler (1979) pycnophylactic property, which specifies that the total population contained within an original choropleth map zone equals the boundaries within the zones of the following dasymetric map. Further the study calculated the RMSE of the four different techniques to examine the population count estimation of each census block and building for the supplementary data. Maantay *et al.* (2007), Mennis (2003), and Mennis (2009) applied the pycnophylactic property to their research to verify whether their estimated (modeled) value of the census units remained the same when re-aggregated to the original value of the enumeration unit. The dasymetric method they followed achieved the requirement of preserving total population from original values to the transformed estimated values.

Table 4.1 shows the result of the pycnophylactic property performed in GIS by calculating the statistics and summarizing the values of the dasymetric raster data within the boundary of the original census units and, in the case of BLDM based on supplementary data, each building was used as a zone. All census blocks have been reported with their FIPS ID and can be easily joined with the original data to show the statistical values. The following table illustrated that the BLDM based on census data had preserved the total population and achieved the pycnophylactic property with an RMSE of zero. The conventional and zoning dasymetric methods had significantly different RMSE errors because the procedure for their implementation and the extent of the study area for both methods vary. The ZDM resulted in tremendous underestimation of population with an RMSE of 179 because the zoning data did not consider population from non-residential areas. On the other hand, BLDM based on supplementary data was found to have greater inaccuracy with an RMSE of 43 compared with the BLDM based on the census data.

Mapping techniques	Number of zones	Total population of choropleth zone	Total population of dasymetric zone	Difference of dasymetric zone and choropleth zone	Percent of zones correctly estimated with the range -5 to 5 people	RMSE
Conventional dasymetric mapping	5079	279,260	278,630	-630	59%	13
ZDM	465	47,211	40,712	-6,499	58%	179
BLDM based on census data	85	16,550	16,550	0	100%	0
BLDM based on supplementary data	233	16,804	13,258	-3,546	43%	43

Table 4.1 Summary result of pycnophylactic property

4.5.2 Visual analysis of spatial errors

Eicher and Brewer (2001) stated that many previous studies lacked a visual presentation of spatial errors resulting from dasymetric mapping. As a result, they presented error maps based on the percent and count errors in each of the three polygon mapping methods they experimented for their research. Percent error maps shows error relative to the total population of a polygon in the census blocks with the dasymetric output (Eicher & Brewer, 2001). For example, if the original census block had a total population of 80, after the process of transformation to a dasymetric map, that single enumeration unit might estimated to be 120 people. The percent error for that enumeration unit would be calculated from the difference of those values, which is 40 divided by the value of a polygon census block and then multiplied by 100%, (40/80 * 100 % = 50 %). The result could be a negative or positive value depending on whether the representation of the dasymetric map underestimated or overestimated the values. For census blocks to have 0% percent error means the population had been preserved, fulfilling the criteria of the pycnophylactic property. Positive and negative values of the percent error represent the overestimation and underestimation of population from the original census data, respectively. Count error was calculated using the actual population of the census block subtracted from the estimated population data to provide the absolute error presented across the study area (Eicher & Brewer, 2001; Mennis & Hultgren, 2006).

4.5.2.1 Error maps for conventional dasymetric method

Figure 4.19a shows the conventional dasymetric map of Ingham County, produced using land cover data as ancillary information, Figure 4.19b represents the percent error map, and Figure 4.19c represents the count error map. The percent error map shows that overestimation and underestimation of population occur in both urban and rural census blocks. However, it had been observed generally that rural blocks tend to be overestimated while relatively small urban blocks tend to be either correctly estimated or underestimated. The same patterns had been investigated by other researchers in the dasymetric mapping process (Eicher & Brewer, 2001). One possible weakness of percent error maps was that rural areas generally have higher percent errors due to lower totals within their zones. In addition, most of the residential buildings found in such areas are along the roads, which create error in representing the population. In most cases, the census boundary follows either man-made or natural features to delineate the enumeration unit. As a result, these areas were subject to the mixed-pixel problem. Furthermore, presenting the percent error in map form would visually exaggerated the higher errors in the study area, since most rural map zones tend to be larger than urban map zones.

Count error maps generally overestimated and underestimated population for the majority of census blocks in the study area, though the values were not extremely large. Comparing rural and urban areas, relatively large rural blocks tend to be overestimated while relatively small urban blocks tend to be underestimated. Similar pattern of errors had been found from previous research (Eicher & Brewer, 2001; Harvey, 2002; Mennis & Hultgren, 2006)

As Maantay *et al.* (2007) contended, preservation of pycnophylactic property was not always achievable for dasymetric methods on population density derived from land use/land cover data. Maps of count error would not signal high errors in rural areas with low total populations, as shown on the percent error maps. The count error map enabled the reader to visualize and understand the overall quality of the variables that have been mapped inaccurately. However, it had a major weakness; that is, to present the count error data with areal fill, since larger polygons tend to have higher values compared to small polygons with the same characteristics.

4.5.2.2 Error maps for ZDM

Figure 4.20a presents the ZDM of East Lansing produced by using zoning and land cover data as ancillary information, Figure 4.20b shows the percent error map, and Figure 4.20c shows the count error map. Although the general characteristics of percent and count error maps of ZDM resembled that of a conventional dasymetric map, there was one big difference in the zoning dasymetric error maps. One major weakness of the zoning data was that they did not consider population reported from non-residential areas. Several census blocks had been reported with population, but zoning data discarded all those reported population, causing the error to increase. For example, because MSU is part and parcel of the city of East Lansing, large portions of the campus area were designated as “university”, although these areas had a large number of people reported from the dormitories of the specified census blocks. The percent error map (Figure 4.20b) shows that large rural block were overestimated while others were underestimated as the zoning data were classified as non-residential area. The count error map provides the information to visualize the overall quality of the map by showing the overestimation and underestimation of population (Figure 4.20c).

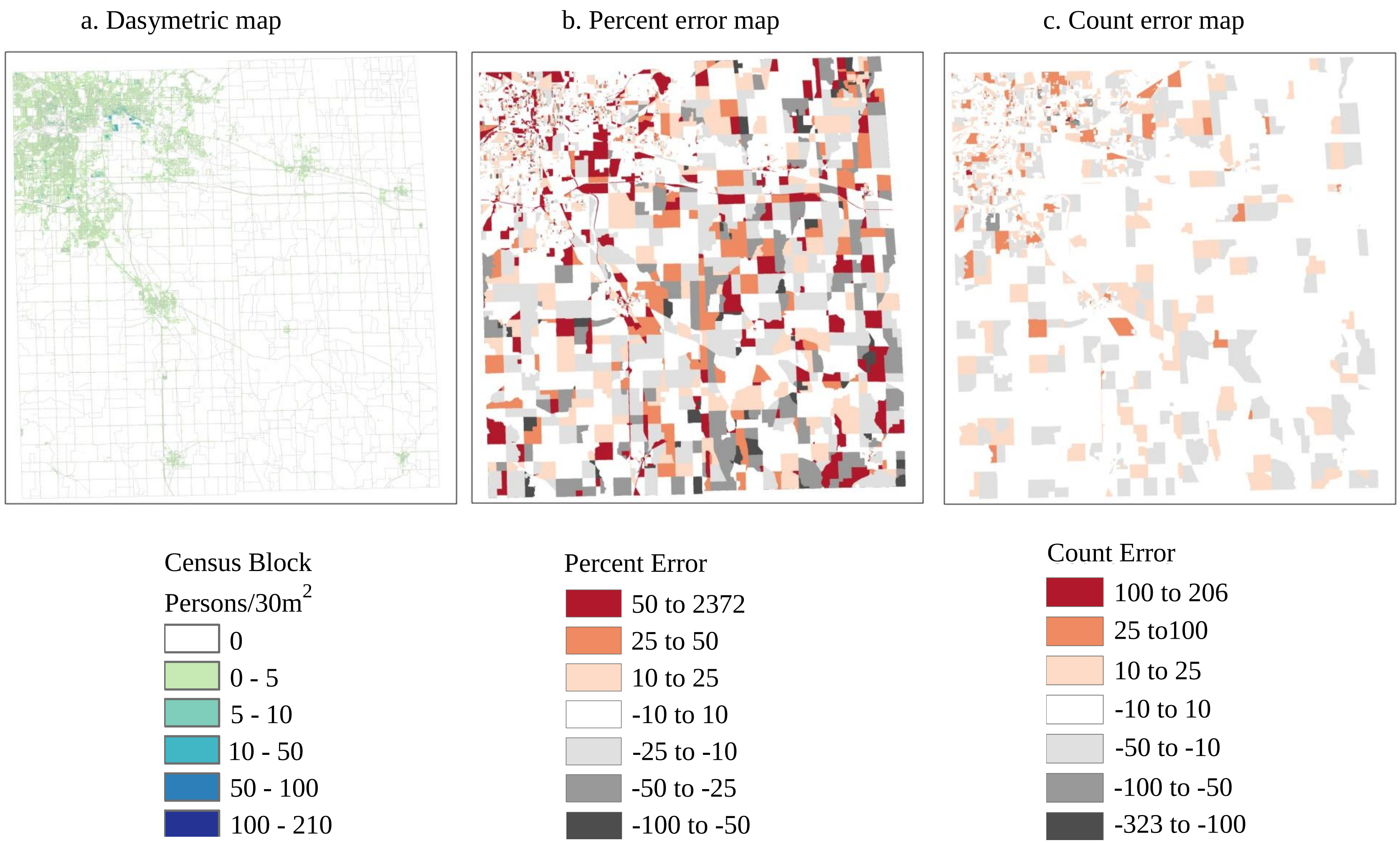

Figure 4.19 Dasymetric map (a) using only land cover as ancillary information in Ingham County showing a percent error map (b), and count error map (c). Percent error is error relative to the total population of a polygon in the census blocks while count error is a value calculated using the actual population of the census block subtracted from the estimated population data.

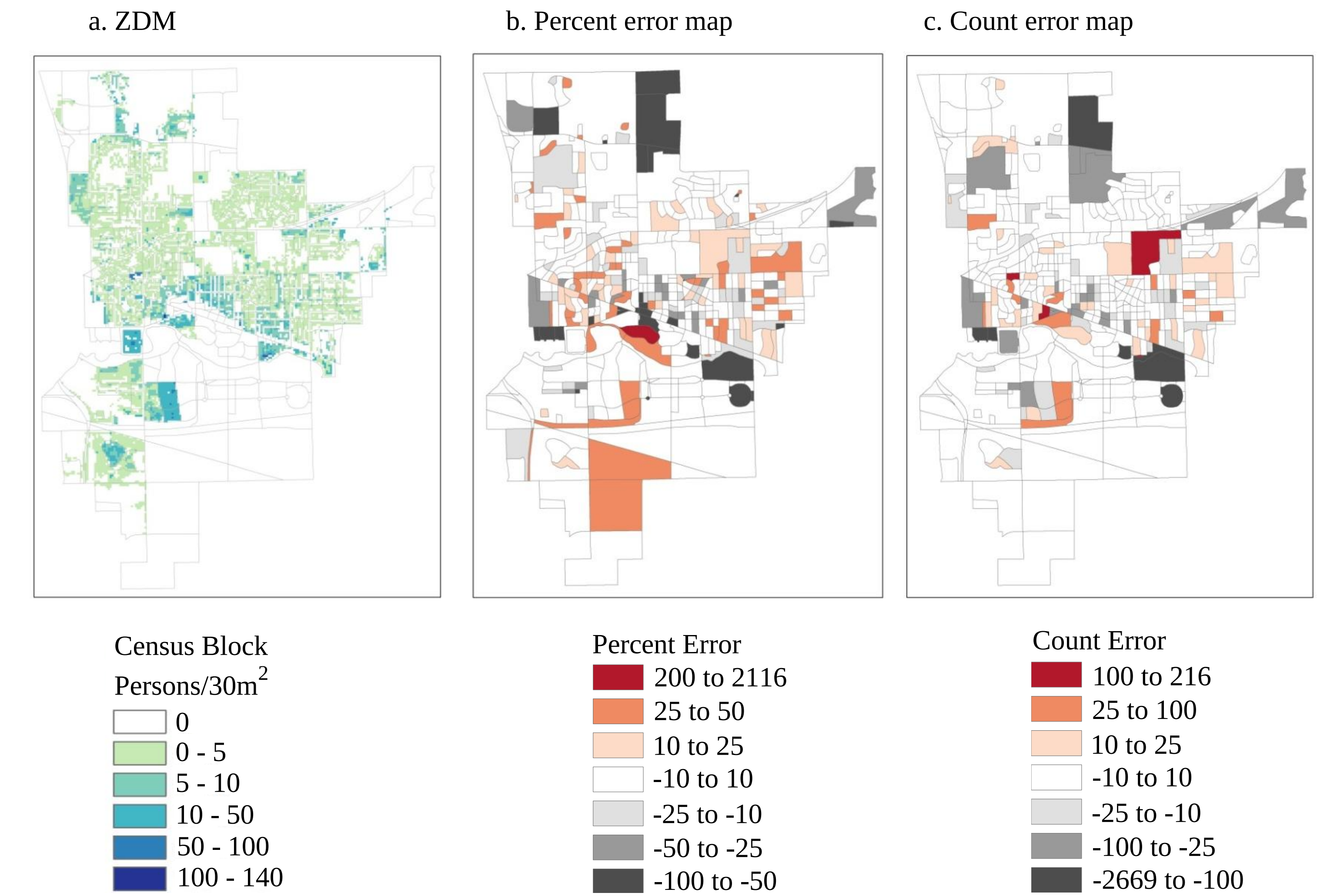

Figure 4.20 ZDM (a), percent error map (b), and count error map (c) in East Lansing. Percent error is error relative to the total population of a polygon in the census blocks while count error is a value calculated using the actual population of the census block subtracted from the estimated population data.

4.5.2.3 Error maps for BLDM

Assigning census block population based on MSU building data had been found to always preserve the population and did not require preparing percent and count error maps. The procedure for building dasymetric mapping was different because all census block population had been assigned to only residential buildings at MSU within the enumeration units. Therefore, the building dasymetric method was able to meet the requirement of pycnophylactic property fully, since the estimated (modeled) values of all census blocks remain the same when re-aggregated to the original value of the census blocks.

However, the BLDM based on the supplementary data were not able to preserve the population and acquire the same accuracy compared with the BLDM based on the census data. The main reason for this inaccuracy could be due to its data implementation process. The building shapefile data served as a polygon in which supplementary population data were assigned. Several buildings were so small that when converted from vector data to raster data, they lost their shape and size. In GIS, the sum and other statistics that summarize the value of the dasymetric raster data with the boundary of the original data (in this case, the boundary of the building) could not preserve the population, leading to a high RMSE. As a result, the majority of building population counts had been either underestimated or reported without data. Figure 4.21 illustrates the problem, showing the building data overlaid on the dasymetric output. Applying high resolution of the dasymetric mapping technique could be a possible solution for using individual buildings as ancillary information.

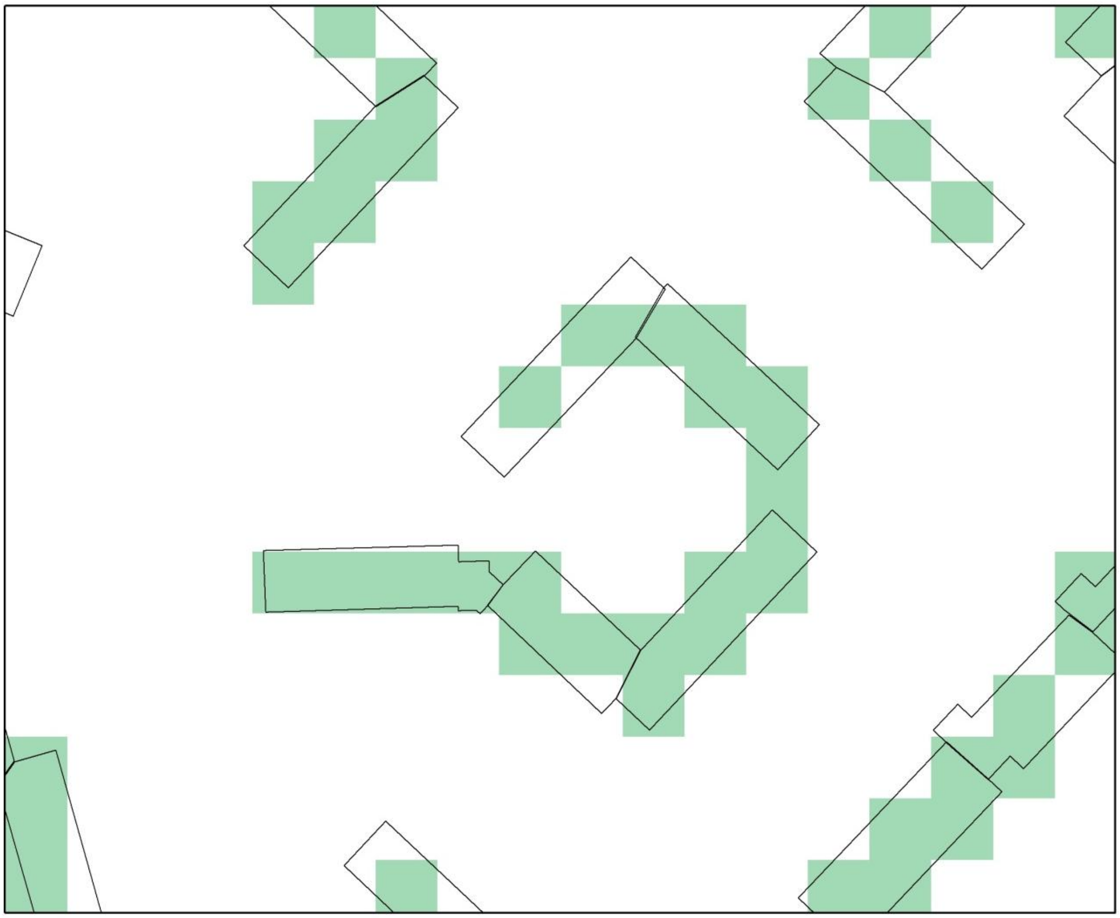

Figure 4.21 Illustration of problems related to BLDM based on the supplementary population data. The building shapefile overlaid on the BLDM shows how the shape and size of the buildings are distorted and how this method was not able to preserve the population using the assessment of the pycnophylactic property.

Chapter Five – DISCUSSION

5.1 Mapping human population

This chapter will discuss the errors and uncertainties in dasymetric mapping and presents a summary table. Conventional dasymetric map, ZDM, and BLDM have been compared. This chapter also discusses future directions of dasymetric mapping, and finally it will present a brief conclusion summarizing what has been discussed.

One way of understanding and visualizing human settlement patterns is to analyze the characteristics of population distribution over space. For example, it is good to know whether a population is concentrated in urban areas or sparsely distributed over a rural landscape. When the first U.S. Census was conducted in 1790, 3.9 million people were living mainly in rural areas (Suchan *et al.,* 2007). As social and economic characteristics of the U.S. population changed over time, the size and geographical distribution of population also changed. The uneven distribution of population continued and by the end of the twentieth century, high population densities existed in some parts of the country while many areas continued to have low population densities.

U.S. Census data provide detailed information for the U.S. population and associated socio-economic characteristics such as age, sex, and other social standing (Langford, 2003). Although these demographic datasets are useful for research projects and applications, they are inherently limited because the data are only available in the aggregate format and based on arbitrarily designed areal units (e.g., block groups or blocks). Arbitrary partitioning of areal units creates many problems for applying spatial analysis due to the incompatible spatial units (Goodchild *et al.*, 1993). Considerable research has been conducted in an effort to map population distribution more accurately (Eicher & Brewer, 2001; Holloway *et al.,* 1996; Langford & Unwin, 1994; Mennis, 2003; Wu & Murray, 2005). Dasymetric mapping is one technique that addresses these potential problems. This thesis implemented different dasymetric models by employing alternative ancillary information and evaluated the associated errors and uncertainties.

5.2 Uncertainties and errors in dasymetric mapping

Uncertainty is present in all geospatial information. In order to allow complex geospatial information to be simple enough to understand and use in analyses, abstractions of reality must be made (Roth, 2009). The two main sources of error in dasymetric mapping were in the nature of ancillary data chosen to make the map and the process of integrating those data sources into one map. No data sources, including census data at block-level, the GIS layers, and satellite imagery, are error free (Liu, 2003). The positional inaccuracy of census units with zoning data and errors related in land cover data can have an impact on the performance of areal interpolation methods. The process of dasymetric mapping also requires estimation and discretization of a geographic phenomenon, commonly population, which is conceptualized as a continuous phenomenon in nature, but is represented in discrete entities whose boundaries are estimated based on the pattern of other data (Mennis, 2009). As a result, dasymetric mapping always involves some error. Additionally, specific errors that can be introduced in each step of the dasymetric process (Yuan *et al.,* 1997) will be addressed in this chapter.

5.3 Limitations of conventional dasymetric mapping

Although conventional dasymetric mapping produces more refined population estimates than choropleth mapping, dasymetric mapping still introduces errors and uncertainties (Mennis, 2009). Better understanding the input data of the dasymetric map would help to determine what causes the error and how it affects the accuracy. Input data for conventional dasymetric maps are census data as well as land use/land cover data.

5.3.1 U.S. Census aggregation problem

U.S. Census data are collected at the household level; however, they are reported aggregately at different enumeration units (e.g., census block, census block group, and census tract). This study applied the smallest unit for reporting U.S. Census data, the census block, to minimize the error of data aggregation, but even so, the problem existed. Never the less, remote sensing data are one of the most important data sources available in providing large quantities of timely and accurate spatial information regarding locations of socioeconomic phenomena, human activities, and residential areas (Liu, 2003; Yuan *et al*., 1997). Integrating the two complementary datasets is important, but difficulties are encountered as both types of data are collected and structured differently, and they are meant for different applications (Yuan *et al.*, 1997).

5.3.2 Uncertainties in classification of remotely sensed imagery

Raw digital data of remotely sensed images has to be classified to apply for various purposes. Generally, the procedure for image classification is conducted by treating the individual pixel values with different spectral bands and by comparing pixels to one another and to pixels of known identities (Compbell, 2002). Uncertainties of remotely sensed data are a concern when using land cover in population mapping. The overall accuracy estimate of NOAA land cover data used for this research was 87.7%, indicating that 12.3% of the pixels were not classifies correctly, which in turn contributes to errors and uncertainties into the final dasymetric map. There are two examples related to the inaccurate classification of land cover. First, the pixel can be incorrectly identified as high-intensity developed, when in fact it is not, and shares 70% of the total population of the census unit. This is a problem of data error propagating on the dasymetric map. Second, if the pixel is correctly classified by the sensor as high-intensity developed, the building located there is not residential, but may be a commercial area which leads to a semantic problem. This problem arises because satellites could not differentiate residential building from a nearby non-residential building. As a result, a signal received by the remote sensor from both buildings might be the same perhaps due to similar building material. Figure 5.3 illustrates the above mentioned problems.

5.3.3 Mixed pixel problem

A problem exists in digital images when more than one land cover type is found within one pixel. As a result of the mixed pixels, pure spectral responses of specific features are mixed together. This happens mostly near the edges of large parcels, or along linear features like rivers and highways, where contrasting brightness values are immediately adjacent to one another. The resulting digital value of the mixed pixel may not resemble any of the categories and, in some cases; the values formed may resemble other categories that are not actually present with in a pixel (Compbell, 2002). In this way, mixed pixels can be a source of error and confusion in land cover classification.

Urban areas are usually heterogeneous, having different cover types, such as asphalt, concrete, grassy parks, and water. However, many of these land cover types are smaller than the resolution of a pixel cell. Therefore, radiation to the sensor from a single element on the ground could therefore be derived from different categories, creating a unique spectral signature and thus, providing a mixed response that does not represent any land cover types (Forster, 1985).

5.3.4 Conversion from vector to raster

Geographical information is stored in either a vector or raster data structure. The conversion of census blocks from vector to raster is a necessary step for the dasymetric mapping process although Fonte (2006) stated that the shape and position may be represented more accurately in vector data structure than in raster structure. In the conversion process from vector to raster there is discretization of geographical space which results in a Boolean classification, either to belong to, or, not to belong to the geographical entities. This causes a loss of information since the entities' boundaries do not follow the shape of the pixels exactly (Fonte, 2006; Mennis, 2003) and, therefore, introduces error.

The conversion from vector to raster can lead to errors in total area estimation, and also errors in individual positions. This will be more exacerbated for small (relative to the cell size) and complex shapes, both of which are likely in urban areas. Area is one parameter of the dasymetric technique; therefore, creating error in area may lead to gross population estimate errors. One example of this error is the errors associated with the BLDM based on supplementary data. The majority of its building population counts have been underestimated and that could be explained due to conversion problems from vector to raster format. It was clear from Figure 4.21

in chapter four, how the size and area of most buildings have been distorted after they were converted from vector to raster to implement the dasymetric mapping.

5.3.5 Subjective population weighting

Subjective weighting is another source of error and uncertainty in dasymetric mapping. Eicher and Brewer (2001) noted the subjective weighting of 70%, 20% and 10% to urban, agricultural/woodland and forests respectively were one of the weaknesses of the dasymetric mapping. Mennis (2003) addressed this weakness by applying empirical sampling techniques to determine the appropriate percentage of assignment value to only land use or urban land cover which has been specified as “urbanization” or the degree of urban development. This study used similar approach to Mennis (2003), but instead of using level urban land cover, it used three urban classes from land cover data. However, the predefined percentage to weight the total population of the census block still introduces errors since they are assigned somewhat arbitrarily.

By changing the weighting scheme of population, ZDM was tested to learn how sensitive the model is to this change. The new weighting assigned 80% of the population to high-intensity developed, 15% to medium-intensity developed, and 5% to low-intensity developed. Figure 5.1a and Figure 5.1b show the weighting scheme of 70-20-10 and 80-15-5 respectively. To compare the results of the two weighting schemes, the 70-20-10 dasymetric map was subtracted from 80-15-5. The result was roughly similar with the exception of a few cells that have a higher value as positive and negative numbers (Figure 5.1c). An RMSE of 2 was calculated when assessing the difference between the two dasymetric maps. Generally, it can be concluded that slightly changing the values of the population weighting scheme would not affect the result very much.

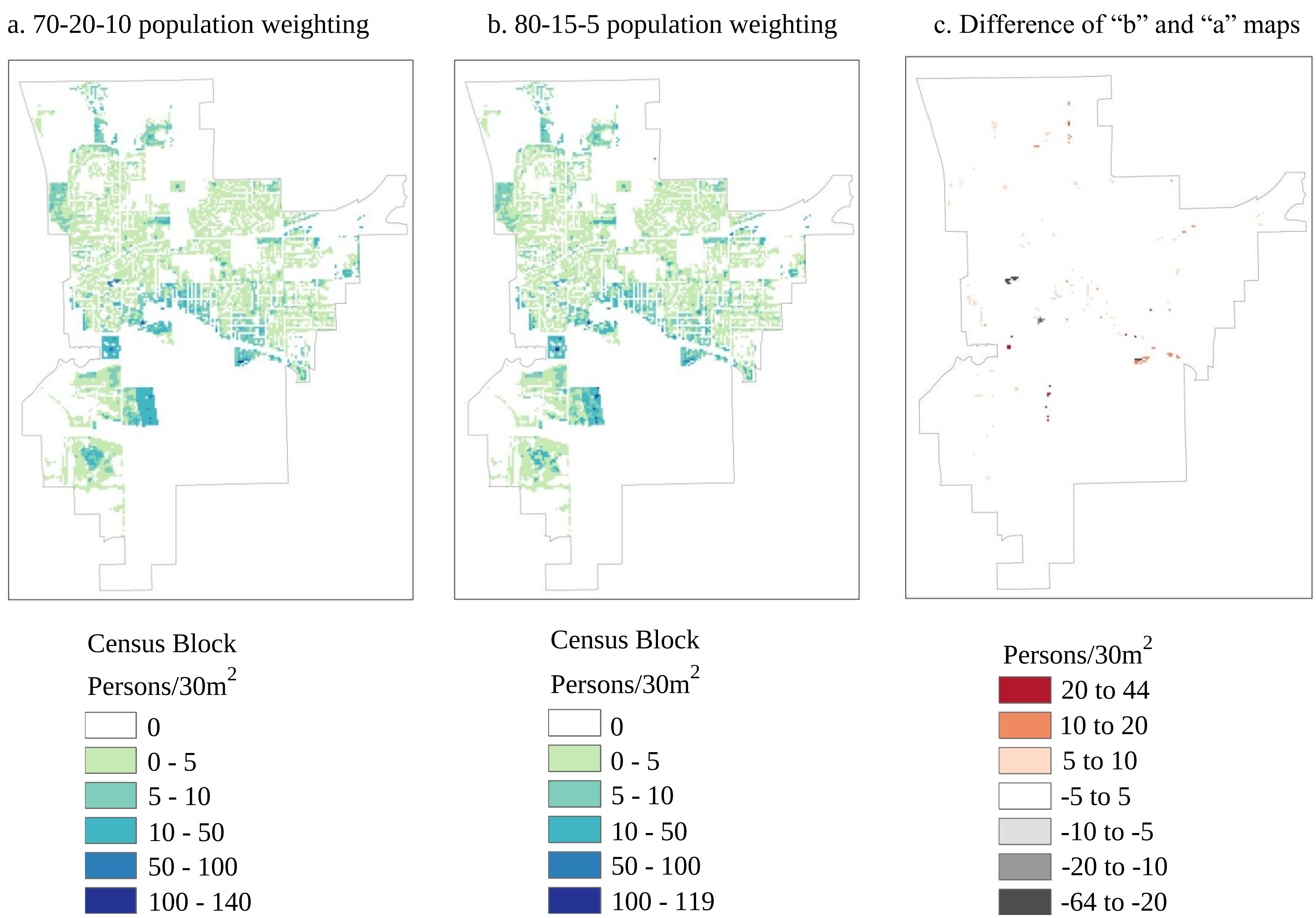

Figure 5.1 The ZDM was prepared based on population weighting of 70%, 20%, and 10% to high, medium, and low urban land cover respectively (a). The ZDM was modified to 80%, 15%, and 5% of the population assigned to high, medium, and low urban land cover respectively (b). The 70-20-10 population weighting was subtracted from 80-15-5 population weighting to show how the two dasymetric maps are different (c).

5.4 Advantage of zoning data as ancillary information

The conventional dasymetric maps in general distributed population based on urban land cover categories. However, the land cover data have several problems that hinder an accurate estimation of population. For instance, all commercial and business areas including roads were classified as high-intensity developed in the NOAA land cover data. This error could be caused due to the mixed pixel problem, the 30 meter by 30 meter spatial resolution was too large to acquire the information on the ground, or the overall errors and uncertainties related with land cover data. Integrating zoning data with the land cover data would make it easy to separate all non-residential areas such as roads, or commercial and business districts that could not share population from the corresponding census unit. Figure 5.3 shows an area around MSU, denoted as a box A in Figure 5.2. Because this area lies on the main road of Grand River Avenue, there are several places designated as commercial and business districts with multiple road networks adjacent to a densely populated neighborhood. As a result, considerable intra-block variation exists in the land cover and, thus, of population density.

Figure 5.3 compares the representation of population in a conventional dasymetric map and a ZDM within the census block. The ZDM can be easily identified as all roads, and commercial and business districts as having zero population density. This is because the zoning data helped to eliminate all non-residential areas, even if it has been classified as high-intensity developed areas in the land cover map. For example, the census block with the Seven Eleven logo in Figure 5.3, the conventional dasymetric map weights a greater amount of the population to non-residential areas, such as a business district. However, in the ZDM, the population is only assigned to residential buildings. The information for that census block has been verified on the ground and the ZDM had improved population estimation at pixel level (Figure 5.4). The same is also true for the roads, as land cover data were not able to differentiate roads from residential buildings, but with the help of zoning information, non-residential areas can be classified correctly.

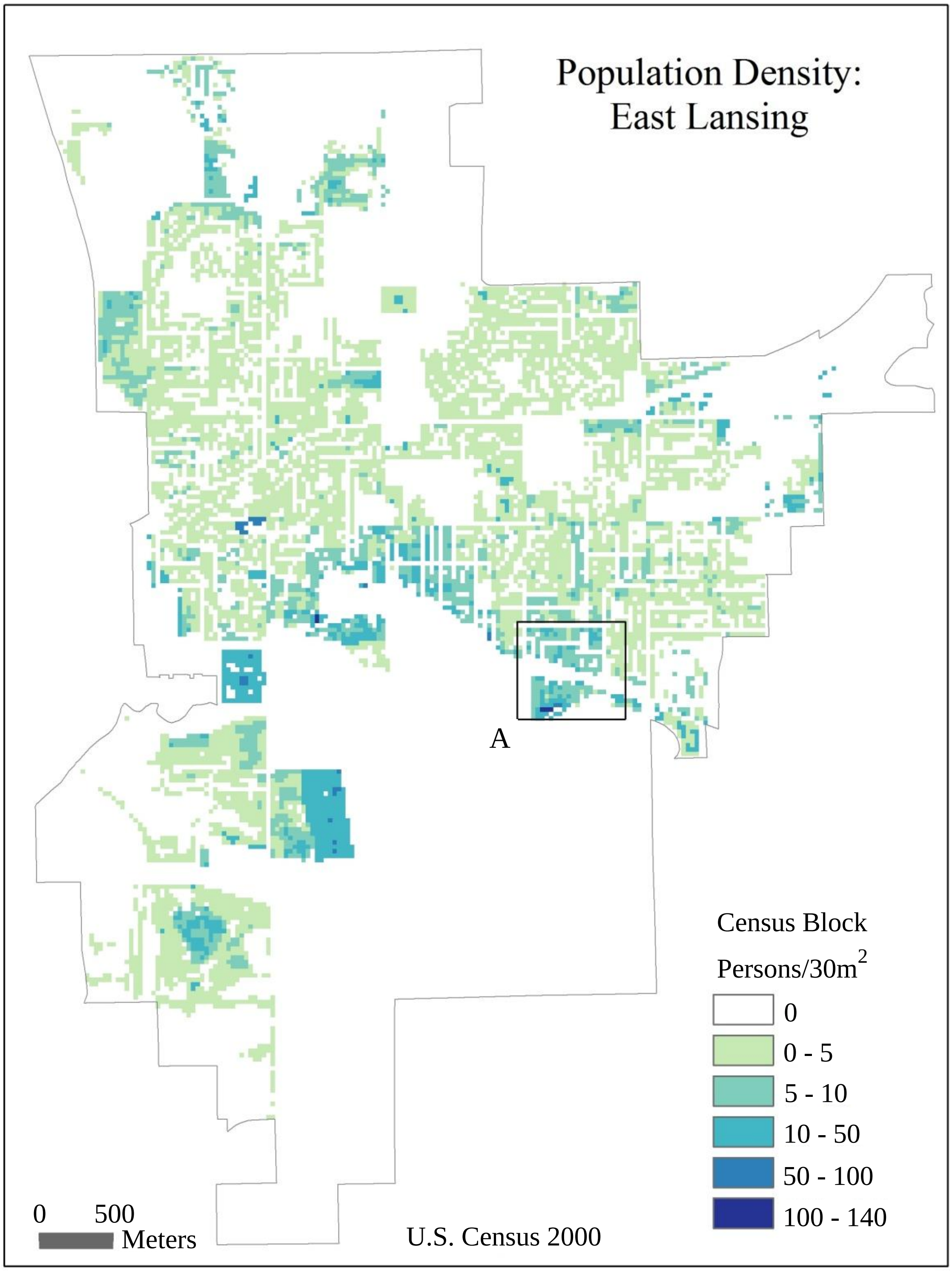

Figure 5.2 ZDM of East Lansing improved human population density compared to the conventional dasymetric map. Box A in indicates the area of detail shown in Figure 5.3.

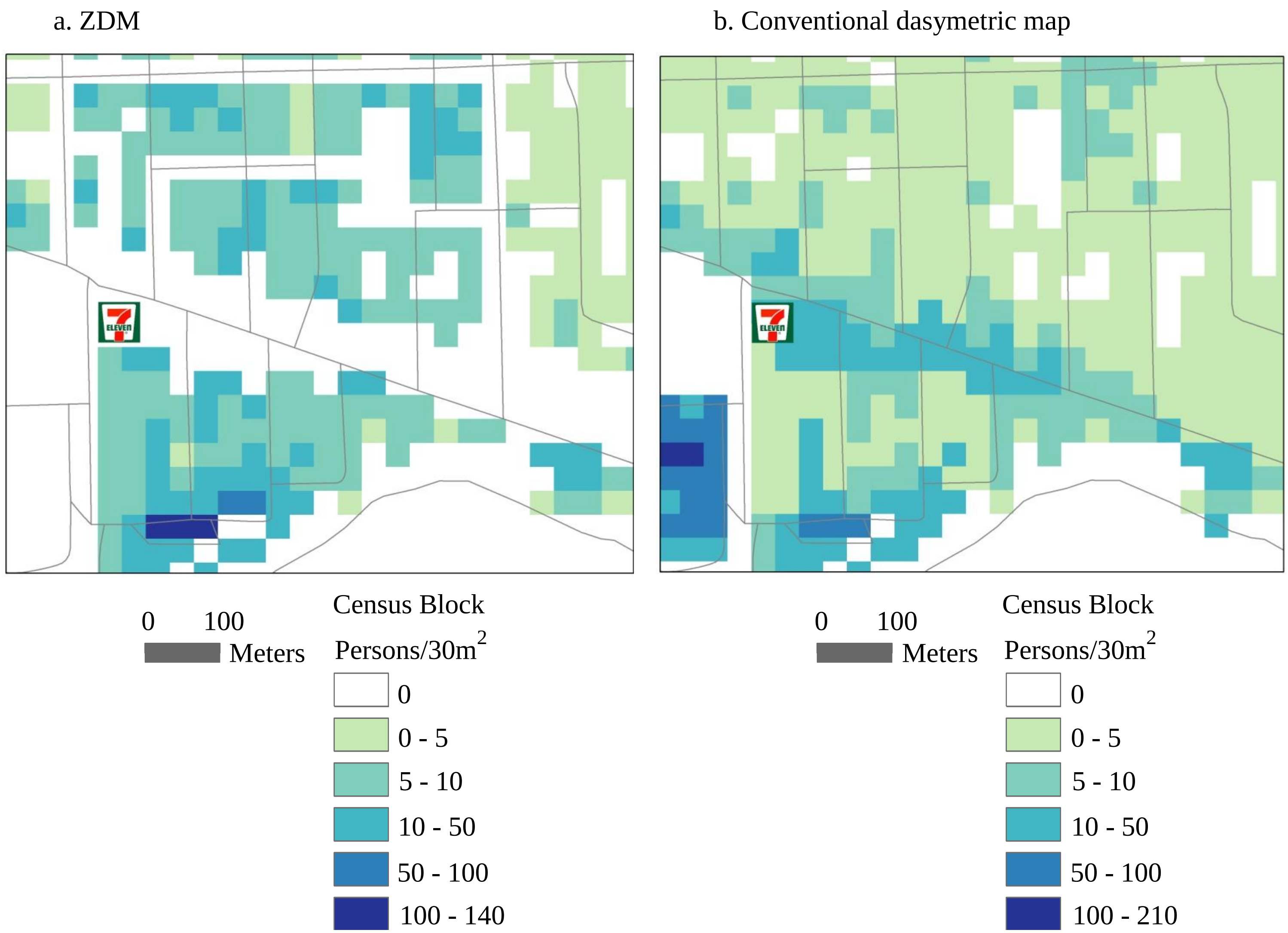

Figure 5.3 Detail of census blocks specified in Figure 5.2 as box A, showing the difference between ZDM (a) where the Seven Eleven area was designated a zero population, and conventional dasymetric map (b) in which, commercial areas and roads were assigned higher population inaccurately.

Figure 5.4 Commercial areas (Seven Eleven) and roads in East Lansing. This area was classified as high-intensity developed in land cover and the conventional dasymetric map allocated population inaccurately. April 2010, photo by author.

5.4.1 Limitations of ZDM

Although ZDM improves greatly the estimation of population density compared to conventional dasymetric mapping techniques, it still has shortcomings. First, the boundary of the census block did not match with the boundary of the zoning information of East Lansing. As it has been discussed in chapter one, the boundary of the census units were delineated to ease the enumeration process rather than to represent the approperate geographical distribution of population or any socio-economic variables (Liu, 2003; Wu & Murray, 2005). Specifically, along the border of East Lansing, there were several census blocks that did not have the same boundary as the zoning information. Second, the zoning data did not take into account population from the census block of non-residential areas. There were several blocks that have a quite number of people reported from the non-residential census block; however, the zoning data distributed population to residential area only. That people were not represented on the ZDM producing additional error. For example, a big building can be used for commercial as well as residential purposes. As a result, there is a combination of residential areas in the commercial or industrial zones. The same is also true for residential zones as some buildings may be used for commercial purposes. In reality, the boundary is not clear, creating errors in the dasymetric map. Third, the zoning data assumed that all residential areas have uniform density. However, the density of residential buildings is actually different from a single-family residential to multiple-family residential buildings.

Finally, the temporal variation of the data sources was another limitation. Although the census data were collected in 2000, land cover data were acquired in 2001, and the zoning data were from 2009. Since man-made features are dynamic, several changes may have occurred during this time, such as a new development of a residential area and movement of people within the city, which adds error to the ZDM. In order to acquire accurate population distribution the datasets should be collected from the same time frame.

5.5 Advantage of building data as ancillary information

Though zoning information improved the conventional dasymetric map, it also had limitations as zoning information was aggregated to a specific class. Since human being live in building structures, building footprint data were applied as ancillary information in the dasymetric mapping process by controlling the distribution of population to only residential building (Figure 5.5). The building footprint has the advantage of displaying the population more accurately because it allocated the population to only residential building footprints. Figure 5.6 shows the comparison of dasymetric maps produced by applying only land cover, both land cover and zoning, and only building footprint data. The conventional and zoning dasymetric maps were not able to display the population as it is found in the ground, but the building footprint was able to assign the population more accurately.

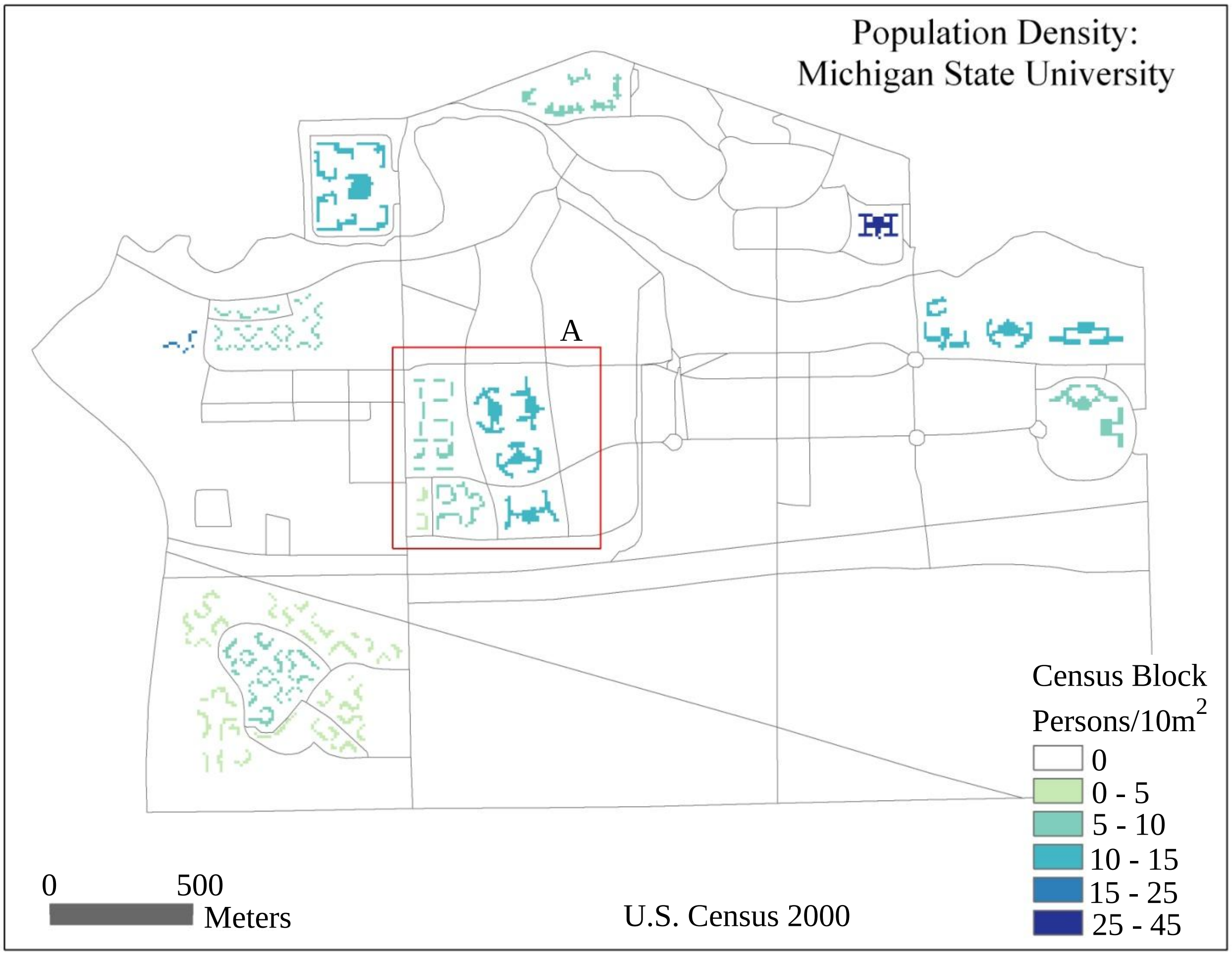

Figure 5.5 Building dasymetric map produced by distributing population to only residential building of MSU. It improved human population density compared to the zoning and conventional dasymetric maps. Box A indicates the area of detail shown in Figure 5.6.

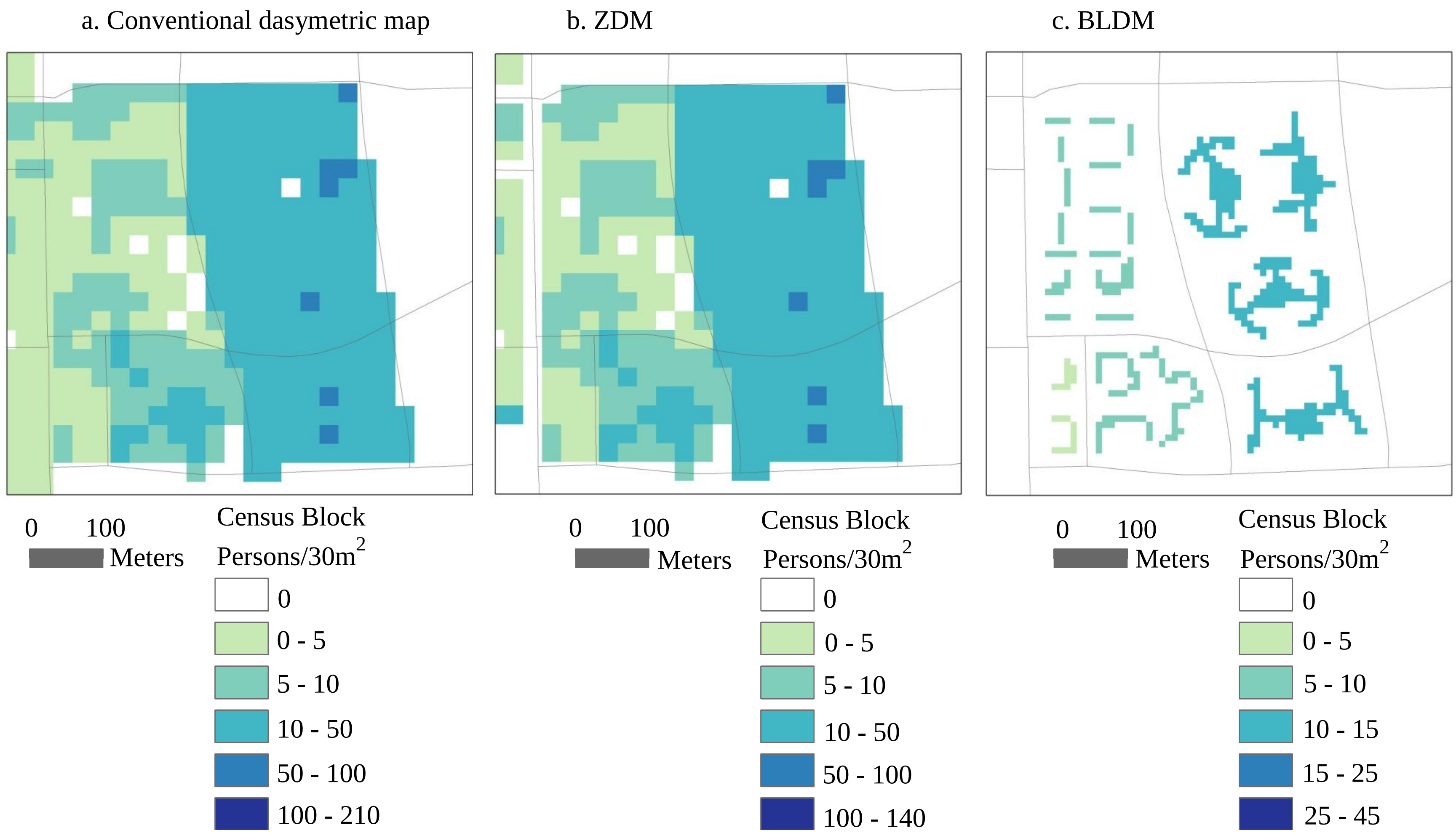

Figure 5.6 Detail of census blocks specified in Figure 5.2 as box A, showing the difference among a, conventional dasymetric map b, ZDM and c, BLDM. Both land cover and zoning ancillary information alone were not able to display the population distribution accurately. The BLDM was able to allocate the population to only residential building as it is found on the ground.

5.5.1 Limitations of building level population

The study of campus building data was limited to the educational institution of MSU and was not inclusive of larger study areas or a variety of residential patterns. Because the data were controlled to distribute the census block population to only residential buildings of the specific enumeration units, it was not able to recognize the heterogeneity of population density among the buildings. As a result, it assumed 100% of the total population of the census block had been distributed uniformly in all the residential buildings. However, in reality those buildings might have different numbers of people.

Supplementary data solved the above mentioned problem by assigning the total population of each building and acknowledging the heterogeneity of population density among them. However, both types of mapping did not take into consideration the height of every building. Because MSU is an educational institution, all residential buildings are not used strictly for residence purposes (e.g., university dormitories might have classrooms and cafeterias within them). In addition, Bhaduri *et al*. (2007) stated that the temporal resolution of the population density is complicated; it varies from daytime to nighttime population as well as weekly, monthly and seasonally. For example, the population density of the MSU campus is higher in the fall, winter and spring compared to the summer when many students leave. The overall summary of errors and uncertainties in dasymetric techniques are summarized in Table 5.1.

Data	Types of uncertainty and errors	Description of uncertainty and errors
Census data	Population boundaries and data	Data errors Population undercounts/over counts Arbitrary boundaries Snapshot problem
Ancillary data	Land cover data	Inaccurate land cover classification Mixed pixel problem Land cover data are not meant for population mapping Satellites cannot differentiate residential from non- residential Snapshot problem
	Zoning data	Positional inaccuracy Classification error Snapshot problem Variation in buildings height
	Building footprint data	Buildings can serve as residential and non-residential purposes Snapshot problem
Implementation process	Translation parameters	Conversion from vector to raster data Data disagreement Subjective population weighting No standard technique for its implementation

Table 5.1 Summary of errors and uncertainties in dasymetric mapping techniques.

5.6 Daytime vs. nighttime population distribution

Mapping population distribution is a complex issue for cartographers since populations are not static phenomena. However, most population maps are static because they are prepared on counts of people in residential areas, based on census data. Most spatial analysis depends on censuses to represent population density over space. This analysis might be true for nighttime populations; however, censuses have a constraint both in space and time and do not capture the population dynamics as functions of those variables (Bhaduri *et al.*, 2007). LandScan is a global database developed by Oak Ridge National Laboratory that provides population estimates for grid cells approximately one square kilometer and finer for the U.S. continent (Dobson *et al*., 2003). The global LandScan population was produced by distributing the best available census counts to grid cells based on coefficient probability, and the data have been further refined by integrating road proximity, slope, land cover, and nighttime lights. LandScan has the unique ability to differentiate residential and ambient populations. Normally, censuses are collected based on where people reside and the result will be a residential population. However, the resulting LandScan distribution represents an "ambient" or average population distribution over 24 hours by integrating the diurnal movements and collective commuting behavior at an instant in time (Dobson *et al*., 2000). For example, the U.S. Census 2000 indicated that the block that used to contain the World Trade Center had only fifty-five people (Dobson *et al*., 2003). In cases of manmade or natural disaster, the whereabouts of people are of much interest (Slocum *et al*., 2009). The global population distribution has several applications among which it can be used to find or evacuate people in case of natural disasters; nuclear, biological, and chemical accidents; terrorist incidents; or other threats (Dobson *et al*., 2000).

Bhaduri *et al.* (2007) noted that human population distribution is dependent on space and time. However, more attention was given to the spatial than to temporal population distribution in the various interpolation methods. Daytime population can be acquired from the temporal relocation of residents from their usual residences to businesses, educational centers, and recreational areas and so on. These types of human activities also directly relate to various demographic groups. Previous research mentioned the challenges in acquiring accurate daytime population data. Bhaduri *et al*. (2007) stated that understanding and modeling the temporal resolution of population is not an easy task as it not only varies from simple day and nighttime distribution, but also can vary on time scales as fine as hourly or as coarse as yearly. In order to acquire high spatial and temporal resolution, it would be necessary to study other impacts such as weather, climate, seasons, and other special occasions of social gathering such as church.

5.7 The future of dasymetric mapping

Accurate data on population distribution is essential for many practical applications. New development of geospatial technology would help to contribute to the progress of dasymetric mapping. If, for example, all building footprints are available and integrated with zoning information, it will be possible to differentiate residential and non-residential areas. As a result, building footprints can substitute for land use/land cover as ancillary information in the dasymetric mapping technique. U.S. Census data are collected every decade and high spatial and temporal resolution of population density can be acquired by collecting real-time data.

5.7.1 LiDAR building extraction

The development of airborne LiDAR technology started in the 1970s and 1980s in the U.S. and Canada (Irish & Lillycrop, 1999) and became well recognized in the geomatics field in the late 1990s (Ma, 2005). LiDAR functions by emitting a laser pulse to the source and precisely measuring the return time; the range can be computed using the speed of light (Miliaresis & Kokkas, 2007). LiDAR data are unique in their potential to provide a very high vertical (elevation) accuracy of the Earth's surface compared to the manual reconstruction from photogrammetric techniques. Traditionally, aerial photographs, high-resolution satellite images, and photogrammetry were the most effective data sources for extracting building footprints and acquiring three-dimensional (3D) data. However, this process was time consuming and not a cost-effective method (Ma, 2005; Miliaresis & Kokkas, 2007; Zhang *et al*., 2006). Processing LiDAR data, which is suitable for DEM generation and building extraction, begins with the separation of ground and non-ground points. After this separation, further processing is required so a DEM can be generated from ground points and objects such as buildings can be extracted from non-ground points.

Recent emerging airborne LiDAR systems use irregularly spaced 3D points to measure objects such as buildings, trees, cars, and the ground by the laser beneath the aircraft. LiDAR measurements are not influenced by sun shadow or relief displacement as opposed to aerial photographs and satellite images (Zhang *et al*., 2006). Extracted building footprint and 3D data are being used in increasing numbers of applications to estimate energy demand, town planning, urban population, cartographic mapping, and civilian and military emergency response (Sohn & Dowman, 2004; Zhang *et al*., 2006; Zhou *et al*., 2004).

Building footprint and height information from LiDAR (Sohn & Dowman, 2004; Zhang *et al.*, 2006) will be a great source of ancillary data for the study of population distribution. This research has shown that using MSU building footprints as ancillary data were more accurate than the other ancillary information, such as land cover and zoning data since the total population of the census block was distributed to only building level residential areas. Further, by using the phycnophylactic property (Tobler, 1979) to validate the output, the total population has been preserved with 100% accuracy. This means that the original total population of the choropleth zones was similar to the sum of the estimated population of the dasymetric zones.

As volume of the buildings can be generated from LiDAR data, it can be possible to weight the population of census data depending on the area and height of every building. Land cover data provide only the 2D features of the earth's surface compared to LiDAR data. So far, several studies (Eicher & Brewer, 2001; Mennis, 2003; Mennis & Hultgren, 2006; Holt *et al.*, 2004; Langford & Unwin, 1994; Sleeter, 2004) have been conducted on the 2D raster based dasymetric method. However, more research needs to be done regarding the use of 3D models to analyze the area and height of buildings and assign them population.

5.7.2 Digital Elevation Model (DEM) as ancillary information

Slope and elevation strongly influence population distribution over the surface of the earth. It can be argued that, given similar weather and climatic conditions, people would prefer to live on a smooth area rather than mountainous or rugged places. For example, Holloway *et al*. (1996) used DEM data as ancillary information to disaggregate population and they restricted for open and forested lands with slope less than or equal to 15%.

Slope is one source of ancillary information for population distribution and can be calculated using DEM for each 1-kilometer grid cell (Slocum *et al*., 2009). Furthermore, the global LandScan population data include a calculated slope gradient using Digital Terrain Elevation Data (DTED) equivalent to the LandScan cell size. Dobson *et al*. (2000) found that slope in the LandScan population probability coefficient was helpful as low slope is highly correlated with larger human settlements. In other words, a grid cell with flat or gentle slope will get a higher probability of population compared to a steep slope.

5.7.3 Population weighting based on zoning data

Assigning the percentage of the population based on land cover classes is fairly subjective and is not supported by statistical evidence (Maantay, 2007). Weighting population based on residential types of the zoning data seems more logical than the subjective weighting used by Eicher and Brewer (2001) in the three-class method. Maantay (2007) explained that zoning and other parcel-based information provide precise land use information that may be richer in content than spaced-based observations and offers ancillary information that is strongly related to population distribution.

The population of the census block can be assigned based on zoning data. For example, if there are three types of residential buildings in a census block: single, double, and multiple family homes, and if we assume that two persons, three persons, and four persons live in single, double, and multiple family homes, respectively, then the total population would be nine people. Then, the corresponding proportions of the population (i.e., 2/9 =0.22 or 22%, 3/9= 0.33 or 33%, and 4/9= 0.45 or 45%), can be used to weight the population. There are limitations with this type of population weighting as the number of people living in each type of residential home could vary. However, this method has a logical basis instead of arbitrarily weighting the population by 70%, 20%, and 10% to high, medium, and low-intensity developed of urban land cover, respectively.

5.7.4 Real-time population

Currently, increasing wireless communication devices and geospatial technologies such as GPS are able to provide real-time information about the interaction of people and space (Goodchild, 2009). With increased access to cell phones, companies are making great progress in the incorporation of GPS to cell phone devices. Therefore, it has been possible to track with accuracy mobile phones, locations of vehicles including public transit, and the state of congestion everywhere in real-time (Goodchild, 2009). The Real Time Rome is a project developed by Massachusetts Institute of Technology using aggregated data from cell phones obtained using Telecom Italia's innovative Lochness platform. This project integrated data from various real-time networks to visualize and understand the pattern of daily life in Rome. They were able to map the distribution of the population by interpolating the aggregate number of people based on their cell phone usage in real-time. This includes the distribution of buses and taxis and how different social groups interact in the city (Massachusetts Institute of Technology, 2006).

Real-time population provides not only spatial, but also the temporal resolution, which is of great interest for emergency mangers for fast and accurate responses to man-made or natural disasters. Even Goodchild (2009) discussed with increased geospatial technologies it will be possible to monitor the state of human health everywhere and provide real-time maps of disease outbreaks.

5.8 Summary of limitations

Like many other methods, the dasymetric mapping used in this research utilizes census data. However, census data have constraints because they only represent residential populations as you might see at night when people are sleeping. The census data do not provide any information about the daytime population distribution. For example, the MSU building dasymetric map would be significantly different during the daytime versus nighttime. Classrooms and offices are more populated during the daytime than the residential areas such as dormitories of the university. Therefore, fine temporal resolution is important to accurately depict the population density. Additionally, census data used were from the 2000 U.S. Census block and too old to provide an accurate estimation given 10 years of time difference.

Pycnophylactic property (Tobler, 1979), which states the total population of the original areal units in the census block should be preserved after the areal transformation into dasymetric output, is generally considered to be a validation for areal interpolation methods. However, even census data are not without error. For instance, from the MSU building dasymetric map, one block was reported as zero population from 2000 U.S. Census data, while, based on the supplementary collected data, it was found that Mason and Abbot Halls of MSU dormitories have a total population of 600. It suggested that if the census data are not accurate, preserving the result means preserving the error. Therefore, in order to achieve the goal of a dasymetric map that accurately estimates population, accuracy measurements based on error-free ground truth data are necessary rather than depending on the census enumeration (Liu, 2003).

5.9 Conclusions

This study produced conventional dasymetric map using land cover data and investigated the errors and uncertainties that can affect the accuracy of population distribution. The contribution of the errors were analyzed and it was discovered that the error could be from census data, uncertainties in classification of remotely sensed data, mixed pixel problem, conversion from vector to raster, and subjective population weighting. This study also produced dasymetric maps using municipal zoning and building GIS ancillary data. These two new ancillary data sources improved the accuracy of dasymetric mapping though using them also involved errors and uncertainties. The municipal zoning data produced errors mainly because it did not take into consideration population reports from non-residential areas. In addition, MSU building data assumed the population is distributed uniformly in the residential buildings of campus. However, in reality, those buildings might have different number of people.

Additionally, MSU is an educational institution, and therefore, all residential buildings are not strictly for residence purposes, and that temporal resolution can also vary significantly. Supplementary data were able to solve the problem of uniform distribution to MSU residential building by assigning population to individual buildings.

Each technique was validated using the pycnophylactic property (Tobler, 1979), which preserves the population count of each census unit. In addition, RMSE was calculated for each model and it was found that the BLDM based on the census data had an RMSE of zero and it also accurately preserved the pycnophylactic property. However, the ZDM underestimated the population greatly and had a large RMSE of 179. The main reason for this error could be explained because the ZDM assumed that people do not live in non-residential areas although the census data had quite number of populations reports from some of those blocks. The conventional dasymetric map had a lower RMSE, but a close investigation of the results showed that it was distributing population mainly to high-intensity developed although, these areas might have been commercial or industrial areas. On the other hand, the BLDM, based on supplementary data, significantly underestimated the population and did not achieve pycnophylactic property and has an RMSE of 43. Further visual representation of percent error and count error were applied to conventional dasymetric mapping and ZDM, the general pattern of errors of both mapping were similar. Accurate population estimation is useful for public health studies, crime mapping, and risk assessment (Maantay, 2007).

This study also discussed the future development of dasymetric mapping by incorporating other ancillary information such as LiDAR and DEM data. By acquiring building footprints from LiDAR and by integrating zoning information, it will be possible to differentiae residential and non-residential areas. As a result, building footprint can substitute land use/land cover data as ancillary information in the dasymetric mapping technique. In addition, volume of the buildings can be generated from LiDAR data, which can help to weight the population of census data depending on the area and height of every building. Land cover data provide only the 2D features on the earth surface compared to LiDAR data. So far, several studies have been conducted on the 2D raster-based dasymetric method. However, more research needs to be conducted on applying 3D models to assign population, based on the area and height of every building.

REFERENCES

REFERENCES

Anderson, J.R., Hardy, E.E., Roach, J.T., & Witmer, R.E. (1976). *A land use and land cover classification system for use with remote sensor data*. (USGS Professional Paper No. 964). Washington, DC: U.S. Government Printing Office.

Anderson, M. J. (2000). *Encyclopedia of the U.S. Census* . Washington, DC: CQ Inc.

Bhaduri, B., Bright, E., Coleman, P., & Urban, M. (2007). Landscan USA: A high-resolution geospatial and temporal modeling approach for population distribution and dynamics. *GeoJournal*, *69*(1-2), 103-117.

Bolstad, P. (2005). *GIS Fundamentals: A First Text on Geographic Information Systems* (2nd ed). White Bear Lake, MN: Eider Press.

Briggs, D. J., Gulliver, J., Fecht, D., & Vienneau, D. M. (2007). Dasymetric modelling of small-area population distribution using land cover and light emissions data. *Remote Sensing of Environment, 108*(4), 451-466.

Compbell, J. B. (2002). *Introduction to Remote Sensing.* (3rd ed.). New York: Guilford.

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., & Worley, B. A. (2000). LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing, 66*(7), 849-857.

Dobson, J. E., Bright, E. A., Coleman, P.R., & Bhaduri, B. L. (2003). A global population database for estimating population at risk. In V.Mesev (Ed.), *Remotely Sensed Cities* (pp. 267-279). London: Taylor & Francis.

Eicher, C. L., & Brewer, C. A. (2001). Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science, 28*(2), 125-138.

Fisher, P. F., & Langford, M. (1995). Modeling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A, 27*(2), 211-224.

Fisher, P. F., & Langford, M. (1996). Modeling sensitivity to accuracy in classified imagery: A study of areal interpolation by dasymetric mapping. *The Professional Geographer, 48*(3), 299-309.

Fonte, C. (2006). Conversion between the vector and raster data structures using Fuzzy Geographical Entities: 7th International Symposium on Spatial Accuracy Assessment in Natural and Environmental Sciences. Lisbon, Portugal.

Forster, B. C. (1985). An examination of some problems and solutions in monitoring urban areas from satellite platforms. *International Journal of Remote Sensing, 6*(1), 139-151.

Goodchild, M. F. (1992). Geographical data modeling. *Computers & Geosciences, 18*(4), 401-408.

Goodchild, M. F. (2009). Geographic information systems and science: Today and tomorrow. *Procedia Earth and Planetary Science, 1*(1), 1037-1043.

Goodchild, M. & Lam, N. (1980). Areal interpolation: A variant of the traditional spatial problem. *Geo- Processing, 1*, 297-312.

Goodchild, M. F., Anselin, L., & Deichmann, U. (1993). A framework for the areal interpolation of socioeconomic data. *Environment and Planning A, 25*(3), 383-397.

Harvey, J. T. (2002). Estimating census district populations from satellite imagery: Some approaches and limitations. *International Journal of Remote Sensing*, *23*(10), 2071-2095.

Holloway, S. R., Schumacher, J., & Redmond, R. L. (1996). People and Space: Dasymetric mapping using ARC/INFO. Missoula: Wildlife Spatial Analysis Lab, University of Montana.

Holt, J. B., Lo, C. P., & Hodler, T. W. (2004). Dasymetric estimation of population density and areal interpolation of census data. *Cartography and Geographic Information Science, 31*(2), 103-121.

Irish, J. L., & Lillycrop, W. J. (1999). Scanning laser mapping of the coastal zone: The SHOALS system. *ISPRS Journal of Photogrammetry and Remote Sensing, 54*(2-3), 123-129.

Jensen, J. R., & Cowen, D. C. (1999). Remote sensing of urban suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering and Remote Sensing, 65*(5), 611-622.

Ji, M., & Jensen, J. R. (1999). Effectiveness of subpixel analysis in detecting and quantifying urban imperviousness from Landsat thematic mapper imagery. *Geocarto International, 14*(4), 33-41.

Langford, M. (2003). Refining methods for dasymetric mapping using satellite remote sensing. In V.Mesev (Ed.), *Remotely Sensed Cities* (pp. 137-156). London: Taylor & Francis.

Langford, M. (2006). Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems, 30*(2), 161-180.

Langford, M. (2007). Rapid facilitation of dasymetric-based population interpolation by means of raster pixel maps. *Computers Environment and Urban Systems, 31*(1), 19-32.

Langford, M., & Unwin, D. J. (1994). Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal, 31*, 21-26.

Lavin, M.R. (1996). *Understanding the Census: A guide for marketers, planners, grant writers and other data users*. Kenmore, NY. Epoch Books, Inc.

Liu, X. H. (2003). Estimation of the spatial distribution of urban population using high spatial resolution satellite imagery. Ph.D. dissertation, University of California, Santa Barbara, United States - California. Retrieved from Dissertations & Theses. (AAT 3103443).

Liu, X. H., Clarke, K., & Herold, M. (2006). Population density and image texture: A comparison study. *Photogrammetric Engineering and Remote Sensing, 72*(2), 187-196.

Liu, X. H., Kyriakidis, P. C., & Goodchild, M. F. (2008). Population-density estimation using regression and area-to-point residual kriging. *International Journal of Geographical Information Science, 22*(4), 431-447.

Ma, R. J. (2005). Dem generation and building detection from lidar data. *Photogrammetric Engineering And Remote Sensing, 71*(7), 847-854.

Maantay, J. A., Maroko, A. R., & Herrmann, C. (2007). Mapping population distribution in the urban environment: The cadastral-based expert dasymetric system (CEDS). (Author abstract). *Cartography and Geographic Information Science, 34*(2), 77-102.

Maantay, J. A., Maroko, A.R., & Porter-Morgan, H. (2008). Research Note-A new method or mapping population and understanding the spatial dynamics of disease in urban areas: Asthma in the Bronx, New York . *Urban Geography, 29*(7), 724- 738.

MacDonald, H. I., & Peters, A. H. (2004). *Unlocking The Census With GIS*. Redlands, California: ESRI Press.

MacEachren, A. M. (1979). The evolution of thematic cartography: A research methodology and historical review. *Canadian Cartographer*, *16*, 17-33.

MacEachren, A.M. (1992). Visualizing uncertain information. *Cartographic Perspective*, *13*, 10-19.

Massachusetts Institute Technology. (2006). Real time Rome. Retrieved May, 20, 2010 from http://senseable.mit.edu/realtimerome/

McCleary, G. F. (1969). The dasymetric method in thematic cartography. Ph.D. dissertation, The University of Wisconsin - Madison, United States - Wisconsin. Retrieved from Dissertations & Theses. (AAT 6922430).

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

Mennis, J. (2009). Dasymetric mapping for estimating population in small areas. *Geography Compass, 3*(2), 727-745.

Mennis, J., & Hultgren, T. (2006). Intelligent dasymetric mapping and its application to areal interpolation. *Cartography and Geographic Information Science, 33*(3), 179-194.

Miliaresis, G., & Kokkas, N. (2007). Segmentation and object-based classification for the extraction of the building class from LIDAR DEMs. *Computers & Geosciences, 33*(8), 1076-1087.

Moon, Z. K., & Farmer, F. L. (2001). Population density surface: A new approach to an old problem. *Society & Natural Resources: An International Journal, 14*(1), 39-51.

Municode Library. (2009). Code of ordinances city of East Lansing, Michigan. Retrieved April 20, 2010, from http://library1.municode.com/default-test/home.htm?infobase=13690&doc_action=whatsnew

Okabe, A., & Sadahiro, Y. (1997). Variation in count data transferred from a set of irregular zones to a set of regular zones through the point-in-polygon method. *International Journal of Geographical Information Science, 11*(1), 93-106.

Openshaw, S. (1984). Ecological fallacies and the analysis of areal census data. *Environment and Planning A, 16*(1), 17-31.

Poulsen, E., & Kennedy, L. W. (2004). Using dasymetric mapping for spatially aggregated crime data. *Journal of Quantitative Criminology, 20*(3), 243-262.

Robinson, A.H., Morrison, J., Muehrcke, P.C., Kimerling, J., & Guptill, S. C. (1995). *Elements of Cartography* (6th ed.). New York: Wiley.

Roth, R.E. (2009). The nature of geospatial uncertainty. *Cartography 2.0.* Retrieved May, 2, 2010 from http://cartography2.org/Chapters/page15/GeospatialUncertainty.html

Sleeter, R. (2004). Dasymetric mapping techniques for the San Francisco Bay region, California: Urban and Regional Information Systems Association, Annual Conference Proceedings. Reno, NV.

Sleeter, R., & Wood, N. (2006). Estimating daytime and nighttime population density for coastal communities in Oregon: Urban and Regional Information Systems Association, Annual Conference Proceedings, Vancouver, BC, September 26-29, 2006.

Slocum, T.A., McMaster, R.B., Kessler, F.C., & Howard, H.H. (2009). *Thematic Cartography and Geovisualization* (3rd ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.

Sohn, G., & Dowman, I. (2007). Data fusion of high-resolution satellite imagery and LIDAR data for automatic building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, *62*(1), 43-63.

Suchan, T. A., Perry, M.J., Fitzsimmons, J.D., Juhn, A.E., Tait, A.M., & Brewer, C. A. (2007). *Census Atlas of the United States: Census 2000 Special Reports*. Washington, DC: U.S. Census Bureau.

Sutton, P. C., Elvidge, C., & Obremski, T. (2003). Building and evaluating models to estimate ambient population density. *Photogrammetric Engineering and Remote Sensing, 69*(5), 545-553.

Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association*, *74*(367), 519-530.

U.S. Census Bureau. (n.d). Glossary.Retrieved April 6, 2010, from http://factfinder.census.gov/home/en/epss/glossary_a.html

Wright, J. K. (1936). A method of mapping densities of population: With Cape Cod as an example. *Geographical Review, 26*(1), 103-110.

Wu, C., & Murray, A. T. (2005). A cokriging method for estimating population density in urban areas. *Computers, Environment and Urban Systems*, *29*(5), 558-579.

Yuan, Y., Smith, R. M., & Limp, W. F. (1997). Remodeling census population with spatial information from Landsat TM imagery. *Computers, Environment and Urban Systems, 21*(3-4), 245.

Zhang, K., Yan, J. H., & Chen, S. C. (2006). Automatic construction of building footprint's from airborne LIDAR data. *IEEE Transactions on Geoscience and Remote Sensing, 44*(9), 2523-2533.

Zhou, G. Q., Song, C., Simmers, J., & Cheng, P. (2004). Urban 3D GIS from LIDAR and digital aerial images. *Computers & Geosciences*, *30*(4), 345-353.