Comparison of Spatial Data Types for Urban Sprawl Analysis Using Shannon's Entropy

by

Cora Hoi-Shan Chong

A Thesis Presented to the
Faculty of the USC Graduate School
University of Southern California
In Partial Fulfillment of the
Requirements for the Degree
Master of Science
(Geographic Information Science and Technology)

May 2017

Dedicated to my sister, Cindy Chong (1986-2006), who taught me courage in the face of immeasurable circumstance, compassion where shadows exist, and an insatiable desire for adventure and exploration amongst a labyrinth of challenges. You are invincible.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my advisors, Dr. Karen Kemp and Dr. Darren Ruddell, for their guidance, direction, and assistance throughout the thesis process. I would also like to thank my thesis committee members, Dr. Elisabeth Sedano and Dr. Robert Vos, for their advice and support. Additionally, I would like to thank the faculty and staff of the Spatial Sciences Institute for hosting the graduate program and providing excellent instruction through coursework and correspondence. Lastly, I want to thank my family, friends, and loved ones for their continuous care and encouragement throughout the duration of my graduate studies.

List of Abbreviations

GIS	Geographic information system
GISci	Geographic information science
LDR	Low density residential
MRLC	Multi Resolution Land Consortium
MSA	Metropolitan Statistical Area
NLCD	National Land Cover Database

Abstract

This study compares the use of cadastral land use data with remotely sensed land cover data for urban sprawl studies using the Shannon's Entropy spatial metric. Many rapidly urbanizing countries lack the technological or economic infrastructure necessary to establish and maintain digital cadastral systems, so remotely sensed land cover data may be a viable option for performing urban growth and urban sprawl studies due to its accessibility, cost, and thematic consistency. Shannon's Entropy is a commonly used metric for measuring sprawl in regions outside of the United States, where cadastral data is not available. Few studies have compared land cover and cadastral land use data using Shannon's Entropy as the main comparison metric. The study uses Model Builder in ArcGIS to perform Shannon's Entropy calculations on the metropolitan areas of Minneapolis and Chicago during the period from 2000 to 2011. The calculation uses the proportion and dispersion of low-density land development within the study area to quantify sprawl. The study cities are divided up into buffer zones, and the proportion of low density land development is measured for each zone. This study found that there was no significant difference between the Shannon's entropy results between land use and land cover. The results suggest that land cover data may be suitable for urban footprint studies in regions where cadastral data is not readily available or otherwise unavailable. This study also found that both metropolitan areas had high entropy values, but entropy did not significantly increase over the study period. These results help inform the broader literature on usable data types for urban footprint studies, as well as the use of Shannon's entropy for such studies.

Chapter 1 Introduction

A world demographic shift is occurring as cities and towns promise new opportunities for personal development, upward mobility, and greater access to social infrastructures and services. In this shift, villagers and rural inhabitants are moving towards urban areas and shape the urban environment, leaving footprints in the form of vertical cities and sprawling metropolises. The world's urban population grew from 746 million people in 1950 to 3.9 billion people in 2014 (United Nations 2014). The world's urban population surpassed the rural population for the first time in 2008, marking a widespread shift in lifestyle choice throughout the world. Among the countries leading in population growth and urbanization are India, China, and Nigeria. This rapid rate of urbanization prompts an urgent need to manage growth to account and prepare for emergency management, disaster preparedness, and public health issues that may arise.

While China is in the process of developing GIS-based cadastral land management systems (Guo et al. 2013), many more countries are also growing and urbanizing, and lack the necessary funds to maintain comprehensive digital cadastral systems. When they do exist, digital cadastral data for urban growth studies in many developing countries may be out of date, have low data quality, or may be inaccessible due to political or economic reasons.

Remotely sensed data taken with satellite imagery may be a strong alternative to cadastral systems in terms of accessibility, temporal frequency, and data processing. Satellite sensors are rapidly improving, and some countries have begun to examine the use of remote sensing data as an alternative to traditional ground-based surveying or as a supplement to existing cadastral data (Zahar 2012). This study hopes to provide insight for researchers and city planners to anticipate and plan for the changing landscape of growing urban areas in developing countries around the world.

1.1 Motivation

Measuring urban growth around the world is crucial for understanding how human populations affect the natural environment as urban areas continue to grow and expand. This thesis project compares remote sensing imagery with cadastral data using the spatial metric of Shannon's Entropy. Shannon's Entropy is a metric that measures dispersion of a variable within a system (Yeh et al. 2001).

This study examines some of the fastest growing metropolitan areas of the United States: the Minneapolis-St. Paul-Bloomington, MN Metro Area, and the Chicago-Naperville-Elgin, IL-IN-WI Metro Area (U.S. Census Bureau 2011).

This project gives insight into the growth patterns of these two metropolitan areas, noting whether these areas are growing in dispersed or compact manners. Despite the fast rates at which cities within the study areas are expanding, other cities around the world are, on average, growing at much faster rates than cities within the U.S. (U.S. Census Bureau 2016). Although this study focuses only on two metropolitan areas in the United States, the results of this study may provide insight into the data sources needed to adequately measure urban growth and expansion in other countries. The time period selected for the study includes the years between 2000-2011, an 11-year span during which significant economic events such as the Great Recession occurred. Data from the years 2000-2011, 2005-2006, and 2010-2011 are used in this study.

Several methods of quantifying changing land cover patterns from urban growth using remote sensing are widely used outside of the U.S. (Masoumi 2015). These include transition matrices, spatial metrics, spatial statistics, Shannon's entropy (Dadras et al. 2015; Bhatta 2010; Sun et al. 2007). This is because many countries outside of the U.S. lack the organizational or technological infrastructure needed for developing and maintaining cadastral data (Effat et al. 2015). While the United States uses accurate parcel-level vectors to monitor urban growth, the use of remote sensing may be a more cost- and time-effective approach to monitoring urban growth than traditional field surveying methods (Herold et al. 2005).

In the United States, individual counties keep parcel-level data on all land, including built and unbuilt parcels (National Research Council 2007). The data are thematic and commonly include parcel numbers, acreage of the parcel, property address, and property description. Though most counties keep records in digital databases, some rural counties still exclusively use paper maps to maintain parcel data. Counties that do use digital records typically store the data within a GIS database.

The results of this study may be useful in the U.S., as we as other countries that do have digital cadastral records, due to the potential lack of uniform data coverage across all regions of a country. In the U.S., historical information about individual parcels may be recorded in paper records, but are not necessarily retained in the county GIS databases. For example, several emails requesting digital copies of historical land use information were sent to the County of Riverside and the state of Massachusetts for this project. A representative from the County of Riverside had replied stating that only the most current land use records are retained (Riverside County Information Technology 2016). Additionally, a representative from the Massachusetts Office of Information Technology stated that land use data had not been maintained after 2005.

The US does not have a nationwide parcel database system (National Research Council 2007). This means that record-keeping is conducted at the county or state level, which introduces inconsistencies across administrative boundaries. One of the many applications of parcel data is for classifying how each land parcel is being used, suitably named land use data. Land use data is an added field or set of fields to parcel vector data describing what the land is being used for, such as a parcel being used for a school or for a shopping center. Land use can be determined using administrative records, aerial imagery, or a combination of both administrative records supplemented by aerial imagery.

Many different types of land uses exist, and land uses can vary regionally. Some cities may have very broad land use categories, while others may subdivide land uses into more specific categories, such as dividing educational institutes into high schools or colleges. Land use classifications may not be consistent between counties, and data release frequency may differ between counties. Thus, using parcel-level vector data to compare land use change and urban growth at regional, state, or national levels may be irregular and yield unreliable results for models (Hurtt et al. 2001).

Until a nationwide land parcel system can be developed, land use and land parcel data will remain inconsistent across administrative boundaries. The alternative for inter-county or inter-state urban growth studies, therefore, is to use nationally consistent, pre-classified satellite imagery such as land cover data from the National Land Cover Database (NLCD). Land cover data describes the physical characteristics of the earth's surface, such as water, forest, or manmade structures. Land cover data can be classified using remotely sensed satellite or aerial imagery.

Depending on the resolution of the sensor, satellite imagery can show how land cover changes from fine to coarse spatial and temporal scales. For example, high spatial resolution sensors can capture detailed imagery from areas as small as a meter. High temporal resolution sensors can capture imagery almost daily, potentially allowing for more frequent land cover change analyses. Unlike land use data, satellite imagery ignores socio-political boundaries, allowing regional, state, and national areas to be consistently compared. Land cover definitions are standardized throughout the US, and the temporal frequency of image acquisition also allows for uniform comparison between different socio-political areas (Homer et al. 2004).

Satellite remote sensing offers an efficient, cost-effective, and accessible method of gathering data on land cover (Longley et al. 2002). Analyzing the existing patterns of land cover can provide clues regarding where growth is occurring the most and can help provide a basis for studies to predict where growth is likely to occur. The spatial metric of entropy can provide insight into the dispersion pattern of urban sprawl in urbanizing regions.

Using remote sensing methods to observe urban sprawl may also supplement or replace existing parcel-level observations. The use of pre-classified satellite imagery for observing urban sprawl may be more efficient than the use of parcel-level data. This study compares the efficacy of using remote sensing data with the efficacy of using vector parcel data in urban footprint studies. By analyzing both the land use and land cover change in the study area in the context of urban sprawl, this study aims to measure and compare the two sources of data for measuring urban sprawl. This concept is tested using Shannon's Entropy, a spatial metric that quantifies urban sprawl by measuring dispersion of low density built residential areas. The secondary goal of the study is to observe and describe the sprawl patterns in the study area.

1.2 The Study Area

The Minneapolis and Chicago metropolitan areas were chosen as study areas based on national census rankings of fastest growing and most populous metropolitan areas. The Minneapolis metro area is the sixteenth most populous metropolitan statistical area (MSA) in the U.S., and the Chicago metro area is the third most populous MSA in the U.S. Selection criteria for the study area also included data availability for the period from 2000-2011 and that the city footprint be centered around a high-density urban core.

The study examined county, state, and city websites of the major metropolitan areas listed in the 2010-2013 US Census population change documentation to find historical land use data (US Census Bureau 2013). Many MSAs, states, and municipalities lacked historical land use data in any format, whether in spatial or in tabular format. The San Bernardino-Riverside-Ontario MSA had historical land use for the study period, but was too multi-nodal in terms of an urban high density core to perform Shannon's Entropy without too many confounding variables. Of the major metropolitan areas throughout the United States, only the Chicago and Minneapolis metro areas matched the criteria of population growth, data availability, and a central urban core.

The Minneapolis-St. Paul-Bloomington, MN metropolitan area is the sixteenth most populous MSA in the U.S. and has a population of 3,524,583 residents (US Census Bureau 2015). The Minneapolis-St. Paul area is known as the Twin Cities for the two cities that occupy the area, Minneapolis and St. Paul. The MSA occupies 16 different counties, 14 of which are located in Minnesota and two of which are located in Wisconsin. The most populous county in the region is Hennepin County, which includes the city of Minneapolis. The Mississippi and Minnesota rivers run through the MSA, and a number of lakes can be found throughout the region. Major industries in the state of Minnesota include bioscience, manufacturing, data centers, and renewable energy (Minnesota Department of Employment and Economic Development 2016). The population of the MSA was 2,969,000 in 2000 and 3,348,859 in 2010, marking a 12.79% overall increase in the population within the ten-year span. No physical deterrents to growth, such as large bodies of water or mountains exist in the Minneapolis MSA. This means that growth can potentially spread in all directions.

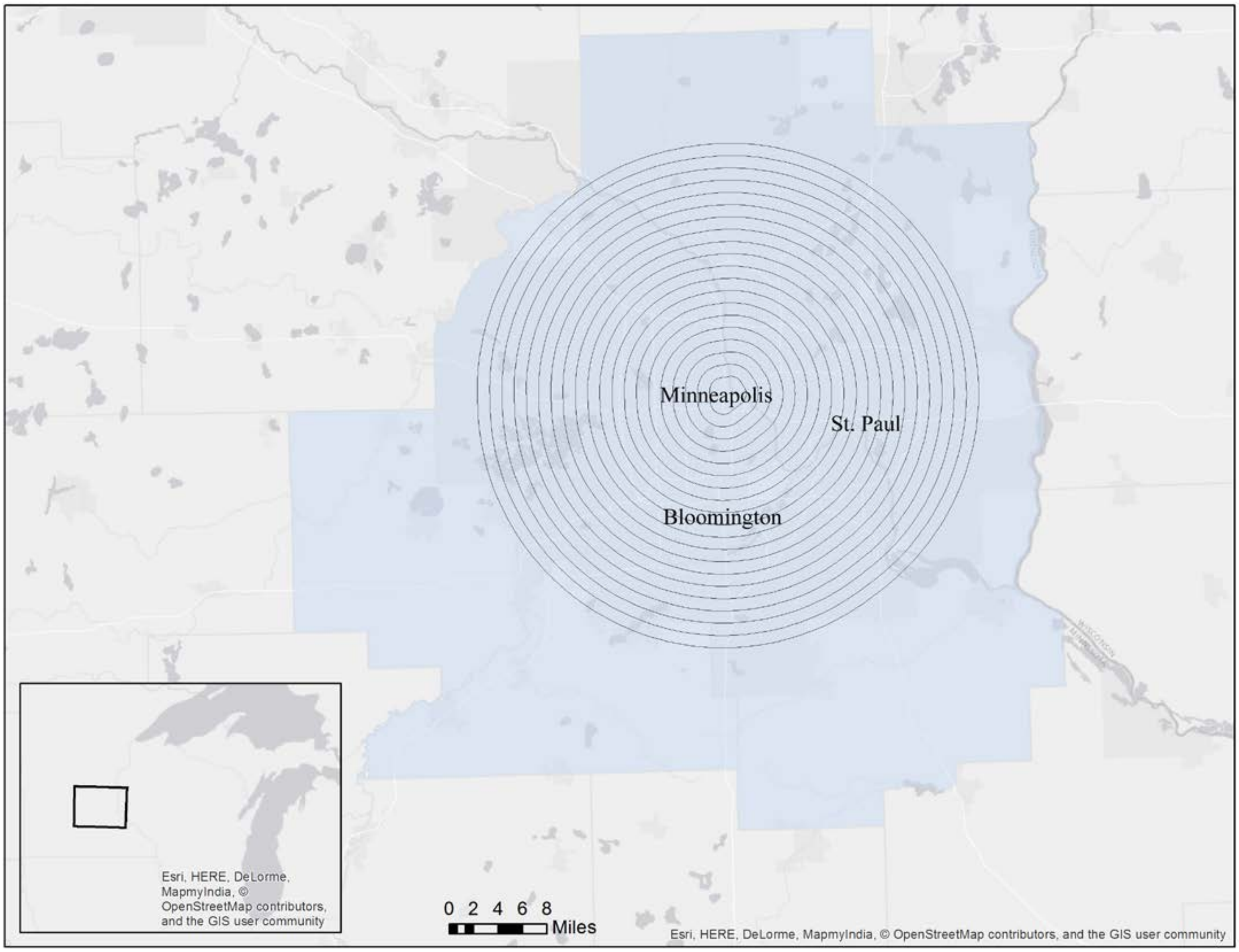

Figure 1. Minneapolis study area, with buffer zones displayed.

1.2.1. About the Chicago-Naperville-Elgin metropolitan statistical area

The Chicago-Naperville-Elgin, IL-IN-WI metropolitan area is the third most populous metropolitan statistical area in the United States and is home to 9,551,031 residents (US Census Bureau 2015). The city of Chicago is the third most populous city in the United States. The MSA is located on the southwestern end of Lake Michigan in Illinois, and partially stretches into the neighboring state of Indiana. The Chicago and Calumet rivers run through the MSA. Major industries include auto manufacturing, biotechnology, business services, and energy (Illinois

Department of Commerce & Economic Opportunity 2016). The population of the MSA was 8,182,000 in 2000 and was 9,461,105 in 2010, marking a 15.63% overall increase in the population from 2000 to 2010. Unlike the Minneapolis MSA, the growth of the Chicago MSA is physically limited by Lake Michigan to the east. Although the goal of the study does not involve comparing between the two study areas, the physical composition of the study areas provide contrast between an area that appears to be sprawled in all directions, versus an area whose growth is physically limited in one direction.

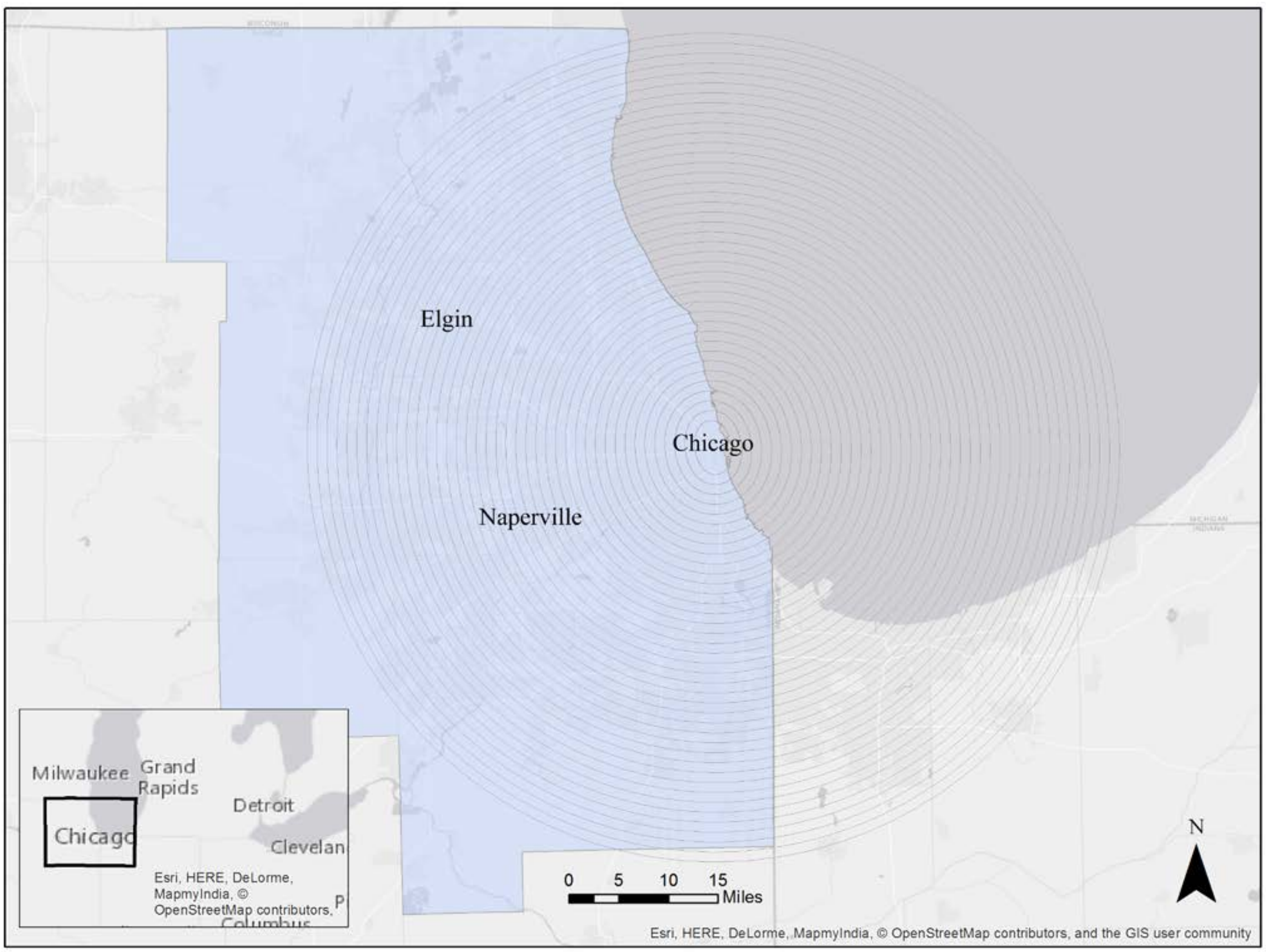

Figure 2. Chicago metropolitan area study area, with buffer zones displayed.

Despite the existence of literature related to urban growth analysis using remote sensing techniques, few or no studies have been conducted on the metropolitan areas of Chicago or Minneapolis using remote sensing. Furthermore, although several studies document the use of remote sensing in conjunction with cadastral data, few studies have compared land cover with land use data for urban growth studies in the United States (Irwin 2003; Wu et al. 2009). This study builds on past research on remote sensing for urban growth analysis to observe the growth of the two metropolitan areas in the 10-year span from 2001 to 2011.

1.3 What is Shannon's Entropy?

In physics, entropy is a measure of disorder or randomness in a system. The second law of thermodynamics states that in a closed system, the entropy of the system never decreases, but rather is inclined towards a state of thermodynamic equilibrium, where various physical and chemical properties are at equilibrium. Maximum entropy, or maximum dispersion, is at thermodynamic equilibrium.

Shannon (1948) employed the concept of entropy to measure the uncertainty of a variable. This uncertainty is expressed as the average expected value of information contained in a message. Information can be defined as the negative log of the probability distribution of outcomes. The probability distribution is the set of all discrete probabilities. The logarithmic function is used because many variables in information science have been observed to vary linearly with the logarithm of the number of possibilities (Shannon 1948). The sum of all discrete possibilities or probabilities is equal to 1.

Shannon's Entropy values ranges from the minimum value of 0 to a maximum value of log(n). A value of 0 indicates that the variable being studied is maximally concentrated in a single area, and a value of log(n) indicates that the variable is evenly dispersed across the entire study area. Dividing the value of Shannon’s Entropy by log(n) results in the relative Shannon’s Entropy, which scales the possible range of entropy values from 0 to log(n) to 0 to 1. This rescaling allows for comparison between different zone sizes and number of zones. The relative Shannon’s Entropy value is used in this study for comparison between the land cover and land use data.

Entropy in the context of spatial science measures dividedness, “the extent to which some total population is evenly distributed among its component parts” (Thomas 1981). Urban sprawl can be measured in the study using Shannon’s entropy to determine whether growth has occurred in a more concentrated or dispersed manner. This study executes Shannon’s Entropy measure on pre-classified NLCD imagery, which should allow for greater consistency of results than executing manual classifications of land cover data. This study also executes Shannon’s Entropy measure on vector land use data acquired from the Chicago Metropolitan Agency for Planning and the Open Data Minneapolis websites for comparison between the different data types. No existing studies have measured entropy using pre-classified NLCD imagery, and most existing studies that have employed the Shannon’s Entropy method focus on study areas in Central Asia, China, or Egypt. These studies use classification schemes similar to that of the NLCD.

Shannon's Entropy is the key method for measuring the magnitude and pattern of urban land use and land cover in this study. By identifying the entropy values of the variable that represents low-density urban land in the region, this research provides a starting point for decision-makers who can interpret the results and implement political, economic, and physical strategies for managing the urban footprint accordingly. The study can also serve as a basis for future studies seeking to quantify urban sprawl using Shannon’s entropy method for metropolitan regions in the U.S.

1.4 Thesis Organization

Chapter 2 discusses the definition of urban growth and urban sprawl and examines past studies of urban sprawl using the Shannon's Entropy metric. Chapter 3 details the procedures used to process the data and calculate Shannon's Entropy. Chapter 4 describes the results of the study and provides insight into the implications of the study. Chapter 5 provides a guideline for future work that can improve upon or be derived from this study.

Chapter 2 Related Work

Urban growth and urban sprawl are concepts that are inconsistently defined and measured. The purpose of this chapter is to examine and critically evaluate several studies quantifying urban growth and, specifically, sprawl. This chapter helps to situate the current study amongst the broad spectrum of urban footprint studies. The chapter first examines cadastral approaches to measuring urban growth in the United States and Europe. The chapter then investigates remote sensing approaches to measuring urban growth and sprawl. Finally, the chapter discusses studies using Shannon's Entropy as the primary metric for quantifying urban sprawl.

2.1 Urban Growth Measurement Using Cadastral Data

Urban growth can be measured using parcel-level cadastral land use data. Change in land use from rural farmlands or other undeveloped land to urban or exurban residential or commercial uses can indicate the presence of sprawl. The definitions of the cadastre, land use, land cover, and urban sprawl are discussed in the following sections.

2.1.1. About Cadastral Data and Land Use

The casdastre refers to the comprehensive database of real property ownership information in a location. Cadastral systems in the US include ownership information and precise metes and bounds, and usually parcel value or other relevant information about the parcel. In the United States, cadastral surveys involve creating associated maps, diagrams, and plats to register and update the parcel in the cadastral system (Esri 2016). This data is stored in a GIS and updated as needed. Parcel level maps, including land use maps used at the city or county level, can be created using cadastral data. Although general aspects of land use can be inferred from land cover, land use and land cover differ conceptually (Hurtt et al. 2001). Land use maps differ from land cover maps in terms of data type and concept. Land cover describes the type of land present, such as desert, open water, urban developed, or tundra, and land use describes how humans use the land. Standards for defining urban land use areas differ from country to country, and administrative boundaries may be larger or smaller than the actual built-up areas of cities (Bhatta 2010).

2.1.2. Urban Growth and Urban Sprawl

When humans build structures on existing rural or natural lands for residential or business purposes, they are changing how the land is being used. Urban growth occurs when the distribution of the human population shifts from being low-density village based to more high-density city based (Clark 1982). According to Bhatta (2010), the boundary of a city or town may be measured through morphological or physical characteristics, or functional or economic characteristics.

The subcategory of urban growth examined in this study is urban sprawl. Urban sprawl is a type of urban growth that describes the expansion of low-density built areas. While there is a lack of consensus on the official definition of urban sprawl, existing literature notes that sprawl is related to changes in land use that lead to poorly planned or uneven patterns of urban growth (Bhatta 2010). This pattern of growth exacerbates issues with traffic congestion including automobile dependency, increased water demand, and increased energy demand.

Urban sprawl is associated with the dispersion of low-density urban developments across expanses of space (Cabral et al. 2013). However, Bhatta (2010) notes that there is “conceptual ambiguity” surrounding sprawl, and as a concept it “suffers from difficulties in definition.” While sprawl is typically characterized as outward expansion from a single narrow urban center, Kreuger (2012) challenges the definition of the urban center, thereby redefining the notion of how sprawl occurs around a single, highly concentrated urban core. This kind of ambiguity affects efforts to measure sprawl using remote sensing imagery, so Bhatta (2010) compares and contrasts different research efforts conducted on urban sprawl. Bhatta's study finds that singular metrics to describe sprawl are often insufficient to determine if areas are sprawling. Despite the multitude of metrics available, the use of entropy for measuring sprawl is the most reliable and most widely used metric. Though sprawl may not seem difficult to identify with the naked eye, quantifying sprawl may be necessary to influence policy and promote sustainable urban development.

For this study, the eight dimensions of sprawl defined by Galster et al. (2001) are assumed for the definition of sprawl: density, continuity, concentration, clustering, centrality, nuclearity, mixed uses, and proximity. Density is associated with the number of residential units per given area. Continuity refers to the fragmentation of urban developed areas. Concentration is the spread of development over a certain area. Clustering is the footprint pattern of development – whether development has been bunched together in an area of developable land. Centrality refers to the measure of distance and spread of developed areas away from a central business district. Nuclearity refers to the number and characteristic of central areas. Mixed-use refers to the number of common human land uses such as business and residential. Proximity describes the distance between different land uses. Galster states that sprawl is a "condition of land use" that can be characterized by the presence of low values in one or more of these eight dimensions.

While Shannon's Entropy does not explicitly measure each of the eight dimensions of sprawl, the concept of sprawl delineated by Galster is used to alleviate the ambiguity around the term "urban sprawl." The study used several of the dimensions – density, concentration, clustering, centrality, and proximity – to determine the city center and the buffer zones that radiate outwards from the city center. The city center is an area with a high concentration of high density residential, governmental, and commercial land uses that are within close proximity of each other. These land uses are bunched together on developed land with low amounts of open or low density land uses. A city center often is characterized by multi-story buildings grouped together. The buffer zones used in this study radiate outwards from the city center, and are assumed to demonstrate low centrality, low density, low concentration, and low proximity between land uses. Shannon's Entropy helps produce an overall picture of whether the study area is sprawled based on the concentration and proximity of low density land use and land cover classes.

2.1.3. The role of cadastral data in land use classification and urban growth

Irwin (2003) investigates land use change at the rural-urban fringe using parcel data in Calvert County, Maryland. The study reviews the conversion of non-residential rural lands, such as undeveloped open space and farmland, to residential lands as a measure of urban growth at the urban-rural fringe. She notes that there is difficulty in tracking and assessing growth in many communities due to the lack of historical data resources necessary for documenting growth. Despite these limitations, she notes that GIS and remote sensing data for land use and land cover studies are becoming more widespread.

According to Wu et al. (2009), the combination of remotely sensed and cadastral data for land use is interpretation is already widespread many municipalities. However, updating cadastral land use data can require vast amounts of time and effort. Laborious manual digitization and extensive human resources are still required for creating cadastral data. Historical legacy data can create problems with inconsistent land use interpretation. Wu's study proposes using a hybrid raster and vector approach to interpreting land use change. The study uses raster field-based urban land use classification approach informed by vector parcel data to guide interpretation. This approach is best suited for countries with established cadastral data and a need for faster update and interpretation of land use, because countries with data-poor environments would be unable to utilize this approach due to the lack of cadastral data.

Epstein et al. (2002) explores techniques for analyzing suburban sprawl in Columbia County, Georgia, using both cadastral data in GIS, and Landsat 5 TM NLCD data from 1993. Epstein’s study conceptually mirrors the current study in that two different data sources are compared for a study of urban sprawl. In her study, single family dwellings of any size are classified as low density residential (LDR), which is the main indicator of a sprawling neighborhood.

Instead of using land use vector data, however, Epstein uses road network coverage to identify LDR areas. The overall accuracy of correctly identifying LDR areas with the remote sensing data was reported to be 72.6 percent, and the accuracy of the vector data was 88.8 percent. The study was conducted in 2002, so newer satellite sensors and improved classification methodologies may increase the accuracy of classifying LDR data. The study concludes that significantly more time was spent processing the vector datasets than for the raster datasets, but the time was justified due to the vector’s improved spatial contiguity and thematic accuracy. However, many developing countries lack the infrastructure to build and maintain cadastral databases. The 72.6 percent accuracy of classifying LDR data may be sufficient for urban sprawl studies in developing countries.

As satellite sensors and remote sensing classification methodologies continue to improve, urban growth and sprawl may be more effectively monitored using remotely sensed data (Longley et al. 2002). Progress in the technology and methodology is likely to benefit both data-poor countries, where cadastral data may be inaccurate or otherwise unavailable, and developed countries, where data may not be up-to-date or is inconsistently defined. For example, Zahar (2012) investigates the use of high-resolution satellite imagery for defining parcel boundaries and land use in Pakistan. The study concludes that high resolution satellite imagery may be viable for delineating parcels in place of traditional field-based surveys. While cadastral data is typically perceived to be more accurate, remotely sensed data for urban growth studies are gaining traction due to its temporality, availability, and lower processing and update times (Longley et al. 2002).

2.2 Urban Growth Measurement Using Remote Sensing

Myriad studies have been conducted on urban growth measurement using remote sensing techniques (Alsharif et al. 2015; Belal and Moghanm 2011; Bhatta 2010; Crowther 2015; Masoumi and Roque 2015). Urban growth is typically measured using change detection techniques on classified land cover imagery (Bhatta 2010). Land cover pixels are classified as natural or impermeable, and impermeable surfaces such as roads and buildings are considered to be “developed.” However, urban developed areas are more than simply impervious surfaces. Urban areas contain a wide range of spectral signatures due to a variety of different urban infrastructure materials, landscaping, and tree canopy. This wide range of spectral signatures creates a mixed pixel problem that is observed during the classification process for low to medium resolution satellites like Landsat (Pena 2012).

Despite the limitations of the mixed pixel problem, land cover classification initiatives such as the NLCD have been found to be between 84-85% accurate for2001 and 2006 (Wickham 2013). This accuracy range is acceptable for most remote sensing land cover studies and satellite data classification studies (Anderson et al. 1976, Ismail et al. 2008**)**. The primary limitation of

NLCD data for measuring urban sprawl includes the difficulty in measuring low-density residential developments in rural areas, also known as exurban development (Irwin 2007). This is because the low-density residential developments tend to have smaller distinctions between vegetative cover and impervious surfaces, thus easily confounding data that is not verified by ground-truthing.

Irwin argues that finer scale land cover data would be ideal for more accurate measures of urban sprawl, and that NLCD data is not at the appropriate spatial scale for measuring urban sprawl. However, Irwin's argument evaluates only a single urban area in Maryland, which may not be extendable to the rest of the nation or rest of the world. In contrast with Irwin's argument, Wolman et al. (2005) finds that NLCD data can be appropriate for measuring sprawl depending on sensitivity of the definition of sprawl. Wolman et al. also uses Galster's (2001) eight conceptual components of sprawl in their study. Additionally, studies of sprawl using similar Landsat data and methodologies of classification have been successful in quantifying sprawl (Sudhira et al. 2004; Deka 2010; Effat 2015).

2.2.1. Studies of urban growth and sprawl using remote sensing

Sources used in this project include Anthony Gar-On Yeh and Xia Li's article on monitoring urban sprawl using Shannon's Entropy approach, as well as R.W. Thomas' 1981 article on Shannon's Entropy in the context of spatial studies. Yeh and Li's (2001) article discusses the definition of urban sprawl, methods of measuring urban growth, and how sprawl can be analyzed using remote sensing imagery, which aligns with the goals of this project. Thomas' (1981) article provides a detailed explanation of Shannon's Entropy, and how the metric can be used in spatial studies. Other sources include thesis projects by Richard Crowther

(2015) and Eric Pena (2012). Both Crowther's and Pena's studies provide a local view of urban land use change.

Crowther (2015) uses NLCD data from 1992-2001 to review urban land cover change in the cities of Pasadena and Inglewood, California. His study explores a methodology for aggregating land cover datasets from 1992 to compare to 2001 datasets, and roughly compares the aggregated datasets. While Crowther's study does not measure urban sprawl, the study provides a framework for working with NLCD datasets, and gives insight into the changing landscape of selected Southern California cities. Crowther notes that combining Census population data with land cover change would be a useful metric for projecting future land cover change.

Pena combines land cover change and population data to measure urban growth in the Lower Rio Grande Valley in Texas. Pena uses supervised classification techniques to classify Landsat 5 imagery into three different urban land cover classes and one uninhabited land cover class. He overlays this imagery with dasymetric population maps based on Census data to explore the magnitude of growth (low, medium, or high) in the region. The study notes that the consideration of social and demographic changes can influence the built environment, so it is important to take population data into consideration when assessing land cover change.

2.2.2. International Studies of Urban Sprawl

The majority of urban sprawl studies using remote sensing techniques concentrate on developing countries such as India and China rather than the U.S. This is because using conventional surveying techniques for urban sprawl studies is often cost-prohibitive and time consuming (Effat et al. 2015). Using remotely sensed imagery for urban sprawl measurements is often more cost effective, provides coverage of large areas, and can be performed at more frequent intervals than conventional surveying (Herold et al. 2005). Improving radiometric, spatial, and spectral resolutions, as well as decreasing costs on satellite sensors further make satellite remote sensing a viable alternative to traditional surveying and mapping methodologies. The images taken from the Landsat series of satellites are commonly used in urban growth studies due to their consistent coverage, accessibility, and adequate spatial, temporal, radiometric, and spectral resolutions (Sudhira et al 2004).

2.3 Shannon's Entropy Approach to Measuring Urban Sprawl

Several urban growth studies use Shannon's entropy on urban areas to measure sprawl over time **(**Bhatta 2010; Dadras et al 2015**;** Sun et al 2007). Bhatta (2010) notes that one of the most popular methods used to measure urban sprawl using remote sensing and GIS methods is Shannon's Entropy, due to its decreased sensitivity to issues relating to the modifiable areal unit problem, where results may change drastically with the size, shape, extent, and number of regions involved (Bhatta 2010). Thomas (1981) states that unlike other, more traditional methods of measuring spatial dispersal, the value of relative Shannon's Entropy is "invariant" with the number of regions, n. The metric is still sensitive to the size and shape of the zones within the regions. The aggregation effect of coarser scale combined with few numbers of zones can have a substantial effect on the metric, as the entropy value would be very low if only one zone were used across an entire metro area. Additionally, using drastically different scales, such as state scale versus city scale, would produce incomparable results.

Yeh and Li (2001) are two of the first researchers to measure urban sprawl using Shannon's entropy on the Pearl River Delta Region of China. Their work has been highly influential in later studies of urban sprawl using remote sensing (Bhatta 2010; Singh 2014; Sudhira et al. 2004). They used principal component analysis on multi-temporal Landsat

Thematic Mapper (TM) images to identify urban areas and track changes in urban land development over time. They then defined dense areas of land development (called "town centers") and major roads in a multitude of cities and towns. Next, they created buffers radiating outward from these areas and roads. To calculate the urban footprint, they use a modified version of the Shannon's Entropy equation presented by Thomas (1981) that accounts for the density of the land development within the region. Their study found a high average value of Entropy for all areas studied, indicating that many of the cities studied were indeed sprawled. The study also found that in some cities and towns, buffer zones centered around major roads produced higher entropy values than buffer zones placed around the town center. This indicates that in some cities, the pattern of sprawl was attributed to distance from roads, while in other places, sprawl was attributed to distance from the town center. The current study draws upon Yeh and Li's (2001) modified Shannon's Entropy equation to examine the urban footprint of the Minneapolis and Chicago MSAs.

Shannon's Entropy has been used to calculate sprawl in India, Iran, China, Egypt, and other developed or developing countries, as observed by Masoumi and Roque (2015). Masoumi and Roque compare and contrast different Shannon's Entropy studies of urban sprawl around the world with their study of Ensenada, Northern Mexico. The researchers find that Ensenada has comparable sprawl speed values as several fast-growing Indian, Portuguese, and Chinese cities and greater intensity values than several Indian cities. This is significant because Masoumi and Roque (2015) are successful in comparing the different sprawl studies, whereas very few studies comparing the Shannon's Entropy metric exists.

The different studies Masoumi and Roque (2015) investigate use different land cover classifications, number of buffer zones, and satellite sensors for land cover data, but can be compared, because the authors apply a standardization equation to convert the entropy to relative entropy. Relative entropy scales the value of entropy from the range of 0 to log(n) to a range of 0 to 1. By converting entropy to relative entropy, Masoumi and Roque (2015) argue that the degree of dispersion can be compared between with different sizes and numbers of buffer zones.

Relative entropy can be calculated by dividing the entropy by log(n). The authors caution that the study is limited to developed or industrial countries, and reliable studies in South America and much of Africa are unavailable for comparison. The current study finds that differing sizes and number of buffer zones can have a slight effect on the entropy value, so Masoumi and Roque's study is used primarily as a reference for the magnitude of change between entropy values over time, with differing numbers and sizes of buffers, and across different geographical regions around the world.

While traditionally the Gini coefficient has been used to describe distribution patterns, this measure introduces the modifiable areal unit problem because it is highly sensitive to the size and shape of the study area (Yeh and Li 2001; Openshaw 1991). The Gini coefficient is a ratio between 0 and 1 that measures the inequality of a distribution (Tsai 2005). The ratio compares the distribution of a variable to a perfectly uniform distribution. In studies of urban form, the Gini coefficient is most often used with population or employment density data. Like Shannon's Entropy, the Gini coefficient divides the study area into sub-areas or zones. However, the size and number of zones has a substantial effect on the calculation because of the mathematical properties of the metric.

Another metric used to measure distribution patterns is the Moran coefficient. The Moran coefficient measures clustering of high-density zones (Tsai 2005). The range of the Moran coefficient is between -1 and 1, where -1 indicates low clustering of high density zones, and 1 indicates high clustering of high density zones. Like the Gini coefficient, the Moran coefficient is highly sensitive to the number and size of zones used due to the mathematical properties of the metric. In contrast, Shannon’s Entropy is less sensitive to the modifiable areal unit problem because the entropy value does not depend exclusively on the area of the zones, but rather how evenly observations are dispersed over space (Yeh and Li 2001).

Buffer zones are typically created around city centers and road networks to help calculate “distance-decay properties of urban sprawl” (Singh 2014). Distance decay refers to how distance affects different subjects, such as how developments are highly concentrated along roads and in the city center. By creating evenly-spaced buffers around city centers and roads, researchers can observe how distance affects the spatial placements of new and existing urban developments.

The current study uses buffer zones around the densest centralized area of the MSAs to calculate Shannon's entropy. The proportion of the variables being studied in each zone is used to measure the overall dispersion of those variables across all zones. The city center itself is not included in the calculation, because of consideration for the centrality aspect of urban sprawl. Centrality considers the spread of developed areas away from a central business district or city center, rather than within the city center itself. The next section discusses methods for performing the calculations.

Chapter 3 Methods

The primary objective of the project is to evaluate the efficacy of using remote sensing land cover data as a comparable data source to cadastral data for urban sprawl studies using the Shannon's Entropy metric. The secondary objective is to examine whether the selected cities have grown in compact or dispersed sprawl patterns over the course of the study period. This chapter is broken up into six different sections. Section 3.1 outlines how the study was designed. Section 3.2 explains the Shannon's Entropy metric. Section 3.3 discusses the data, how the data was acquired, and how the geodatabase is set up. Section 3.4 describes the initial data processing steps. Section 3.5 describes how Shannon's Entropy was calculated using ArcMap and Microsoft Excel.

3.1 Research Design

This project involves calculating Shannon's Entropy for two different metropolitan areas in the United States over three different time periods, using two different data types – raster land cover data and vector land use data. The major variable in this study is E, Entropy, as well as change in Entropy. This variable measures the dispersion of another variable (land use or land cover type) across the study area. Shannon's Entropy is calculated six times for each data type, for a total of 12 calculations.

To assist in planning, cases of urban sprawl measurement using Shannon's Entropy in India, Pearl River Delta Region, China, and Cairo, Egypt are studied for background and methodologies **(**Dadras et al 2015**,** Bhatta 2010**,** Sun et al 2007). While spatial characteristics differed between cities, the cases all of classified land cover data and quantified sprawl using Shannon's Entropy as calculated by Yeh and Li (2001). The studies divide their respective study areas into different buffer zones based on either administrative or other social or spatial boundaries, with no specified standard for the number of buffer zones to optimally execute the calculation. The geometry of the buffers also differed between studies. Some used concentric circles as buffer zones (Dadras et al. 2014), while others used concentric square buffers (Deka et al. 2012). Other studies divided the study area into quadrants to compare entropy values between quadrants. The current study uses concentric circles as buffer zones because the circular shape captured the shape of the MSAs most effectively.

Bhatta (2010) argues that because Shannon's Entropy measures the dispersion of low-density land cover using the proportion of low density land cover in relation to total land cover of each buffer zone, the size, shape, and number of the buffer zones can vary while achieving similar results. This study conducted an analysis using several different numbers of buffer zones to test and verify the sensitivity of the Shannon's Entropy metric. This study performed a sensitivity analysis using a range from 39 buffer zones to 5 buffer zones to determine if the size and number of zones affect the final Entropy result, and found that the metric is somewhat sensitive to the size and number of buffer zones.

ArcGIS split ring buffer, Model Builder, tabulate intersection, and zonal histogram functions were the primary tool sets employed in this study. ArcCatalog was used for geodatabase management.

3.2 Shannon's Entropy

The study used Shannon's entropy on the metro areas of Minneapolis-St. Paul-Bloomington, MN and Chicago-Naperville-Elgin, IL-IN-WI, to quantify the degree of urban sprawl in the regions. Shannon's entropy is a concept grounded in information science that uses entropy as a measure of information gathered from a system based on probable dispersion or concentration of the information (Li et al. 2015**,** Deka et al 2012). The value of entropy ranges from 0 to log(n), where a value of 0 represents minimal dispersion of the variable, and a value of log(n) indicates maximum dispersion of the variable (Sudhira et al. 2004). Half of log(n) is typically used as the threshold to determine whether the area can be described as more or less sprawling (Dadras 2014). The variable examined in this project is low density land cover and low density residential land use.

Shannon's entropy can be applied to urban sprawl calculations by measuring the number of pixels classified as low-density developed land in comparison with number of pixels classified as undeveloped land. A value of 0 indicates that development in the study area is concentrated, and a value of log(n) means that development in the study area is fully dispersed or spread out.

The Shannon's entropy equation for measuring urban sprawl is as follows:

$$E = \sum_{i}^{n} PDENi * \log_{10}(\frac{1}{PDENi}) / \log_{10} n$$

WHERE

$$PDENi = DENi / \sum_{i}^{n} DENi$$

Where *n* is the total number of zones the study area is divided into and *i* is the current zone being studied

Equation 1 Shannon's Entropy equation for measuring urban sprawl (Yeh and Li 2001)

PDENi describes the density of the variable, and PDENi is equal to the “amount of land development divided by the total amount of land in the ith buffer in the total of n buffers” (Yeh and Li 2001). For example, 35,274 pixels are classified as low density land cover for one of the buffer zones in Minneapolis in 2011, out of a total of 334,819 total land cover pixels in that particular buffer zone. DENi is 35,274 divided by 334,819 which equals 0.10535. PDENi is equal to DENi divided by the sum of DENi across all buffer zones for that year. The next step is to multiply PDENi by log(1/PDENi), and then take the sum of the results for all zones for a final Shannon's Entropy value. The result can then be converted to relative entropy, which ranges the results from 0 to 1, by dividing Shannon's Entropy by log(n).

The term $\log_{10}(\frac{1}{PDENi})$ can be used interchangeably with $-\log_{10}(\text{PDENi})$. The reason why a fraction or negative natural log is used is because PDENi is a fraction, and the log of a number less than 1 yields a negative number. Shannon's Entropy can be calculated in GIS using either custom scripts or calculated fields with the Model Builder tool, though the current study uses a combination of ArcGIS and Microsoft Excel to perform the calculations. Sections 3.4 and 3.5 describe the processing and Shannon's Entropy calculation steps in more detail.

3.3 About the Data

The necessary data was required to provide full spatial coverage of the study area from the three years specified: 2001, 2006, and 2011. The datasets needed were pre-classified land cover data and land use data for the Minneapolis and Chicago metropolitan areas.

Several assumptions were made for this study. The study assumes that the land use and land cover classifications are accurate and includes the variable being studied for the time period specified. The accuracy of the land cover data should be around 85% to be considered reliable for land cover studies (Anderson 1976). Thus, the land cover data is assumed to be reliable for this study. The study also assumes that land use data from the years 2000 through 2011 have greater accuracies than the land cover data for those years. Additionally, the land use classifications are assumed to be consistent throughout the study period.

3.3.1. Obtaining land cover data

The study used three different Landsat 5 Thematic Mapper and Landsat 7 Enhanced Thematic Mapper + imagery sets from three different years: 2001, 2006, and 2011. The datasets for 2001, 2006, and 2011 are part of the NLCD, which is created and maintained by the Multi-Resolution Land Characteristics Consortium (MRLC).NLCD.

The Multi Resolution Land Characteristics Consortium (MRLC) uses remotely sensed satellite imagery to create a database with land cover information for the entire United States (Homer 2007). The first NLCD was completed in 1992 using Anderson Level I classifications defined by Anderson et al (1976). The goal of developing a land cover classification system is to provide a consistent, reproducible, and reliable method of deriving land cover data from remote sensing imagery (Homer 2015). The Anderson Land Cover Classification system separates land cover types into 8 different classes, ranging from water to barren land. The classes are further broken into subclasses, such as open water and perennial ice/snow and shrub/scrub. The 2001, 2006, and 2011 NLCD data use a modified version of the original Anderson Land Cover Classification.

The MRLC modified the mapping legend between the NLCD 1992 and 2001 data releases as remote sensing technology and classification methodology improved (MRLC 2016). Land cover data from the 1992 NLCD was desired to expand the scope of this project, but could not be used due to differences in low density land cover classification definitions. Directly comparing the 1992 and 2001 datasets is not recommended, so this study used data only from 2001 and beyond.

Data from the NLCD are preprocessed and pixels have already been pre-classified into the different sub-classes of the Anderson Land Cover Classification. Data from the NLCD is not classified into different land use categories, such as industrial or residential. Instead, it is separated into categories of low, medium, and high intensity developed land (Table 1). This is because NLCD data describes land cover - the physical characteristic of the land, rather than land use – how humans use the land.

Table 1 National Land Cover Database Product Legend – Developed Land Cover Categories

Class\Value	Classification Description
21	**Developed, Open Space** Areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20% of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes.
22	**Developed, Low Intensity** Areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20% to 49% percent of total cover. These areas most commonly include single-family housing units.
23	**Developed, Medium Intensity** Areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50% to 79% of the total cover. These areas most commonly include single-family housing units.
24	**Developed High Intensity** Highly developed areas where people reside or work in high numbers. Examples include apartment complexes, row houses and commercial/industrial. Impervious surfaces account for 80% to 100% of the total cover.

Source: Multi-Resolution Land Characteristics Consortium (MRLC)

This study examines the distribution and proportion of developed low intensity land cover in comparison to all other land cover classes using the Shannon’s Entropy metric. Urban sprawl is primarily concerned with low intensity development, which is classified in the NLCD datasets with the value of 22. Developed Open Space is classified with the value of 21, and includes some large-lot single family housing, but also consists primarily of parks and golf courses. For this study, these types of lands are considered to be undeveloped land.

As Table 1 describes, low and medium intensity land cover describes surfaces that contain between 20%-49% and 50%-79% impervious surfaces. The difference between low and medium intensity categories is the percentage of impervious surfaces: the higher the percentage, the greater the development and density of built area. Both low and medium intensity developed land cover typically consists of single-family housing units.

The term “single family housing” only vaguely describes the urban footprint of this kind of development. This is because single family houses may either be built close to other single family houses on small lots, or built far from other single family homes on large, sweeping lots. Using the impervious surface percentage as a method to quantify the footprint of these developments helps to remove ambiguity surrounding the term “single family housing” when describing urban sprawl. Land cover data uses impervious surface percentage to differentiate between low, medium, and high intensity urban land surfaces. This provides an objective approach to classifying land that may be associated with single-family or multi-family housing units. In contrast, the parcel size for single-family housing units in land use data can vary within and between local jurisdictions, making meaningful comparisons between cities difficult.

The 2001, 2006, and 2011 NLCD data had been downloaded from the MRLC website using the Viewer tool, which allows the user to select a custom area for land cover data download. The land cover data is federal data that is part of the public domain and may be used at no cost.

3.3.2. Obtaining land use data

To examine land use change, the study used land use data obtained from the Chicago Metropolitan Agency for Planning and the Open Data Minneapolis websites. Historical data is available for the years 2000, 2006, and 2011 in Minneapolis, and 2000, 2005, and 2010 in

Chicago. The land use data is at the parcel level and contains fields describing land use types, both in numerical code and descriptive text forms. The land use type examined for this study is classified as “residential single family detached”. This land use refers to individual residential structures that are occupied by single families. This land use type differs from residential single family attached land uses, which are single structures that are divided into two or more units that house single families. Examples of single family attached homes include townhouses and duplexes. The land use data for both study areas include similar descriptions of the single family residential detached land use.

The Minneapolis land use metadata states that some discrepancies between single-family attached housing and low density multifamily residential land use may exist. (Metropolitan Council 2016) However, only the single family detached housing variable is examined in the current study. The Minneapolis data is generally consistent through the study period. Prior to 2000, the Minneapolis data portrayed land use as a function of ownership. The land use categories were adjusted in 2000 to describe how the land was being used. The introduction of higher resolution aerial and satellite imagery as well as supplemental administrative data in 2000 also improved the accuracy of the land use data. Additionally, land use categories were divided into subcategories, such as the different classes of residential land uses. Because of the multitude of changes that occurred in 2000, land use data for Minneapolis prior to 2000 is not recommended for comparison with newer datasets. The land use data is free of topological errors, and the data is complete in terms of geographic coverage. The horizontal and positional accuracy for the datasets are “likely better than the accuracy of the previous years, however, it is still difficult to quantify that accuracy” (Metropolitan Council 2016).

According to the metadata, the land use classifications for Minneapolis are based on custom classification schemes, and the land use designations were derived from analyses of aerial photography and field surveys. The lineage of the land use data includes digital orthophoto quarter quads, historical land use delineations, parcel data, centerline layers, and field checks, among other resources.

The Chicago land use metadata states that data from 1990 through 2005 defined land uses using polygons that extended to street centerlines, excluding highways and other large roadways (Chicago Metropolitan Agency for Planning 2016). Data after 2005 were modified to observe rights of ways. The impact of the modification meant that certain land uses, such as Wetland and Vacant Grassland land uses would no longer be in use. This modification may introduce some discrepancy between the 2000, 2005, and 2010 land use data for Chicago. The data is free of topological errors, and the land use attribute accuracy is stated to be accurate based on authoritative inspection. The metadata also notes that the accurate is likely less than 100% due to the lack of extensive field testing of the data. The lineage of the data includes orthophotos, highway data, parcel data, street centerline data, and historical land use delineations, among numerous other resources.

There is a temporal difference between the land use and land cover data in both the study areas: the land use for 2000 has been acquired, while the land cover for 2001 is used for comparison. This temporal difference may explain some of the variation in the results. The land use data is free to download and use.

3.3.3. The geodatabase

Two file geodatabases were created using ArcCatalog to organize and maintain the data. Each metropolitan area was placed into a separate geodatabase and handled with a different map document due to differences in projection. Several vector and raster datasets were generated for the project, so careful planning and strict organization of the geodatabase was a priority. All of the rasters were handled as individual raster datasets. A partial example of the geodatabase organization can be found in Figure 3 below.

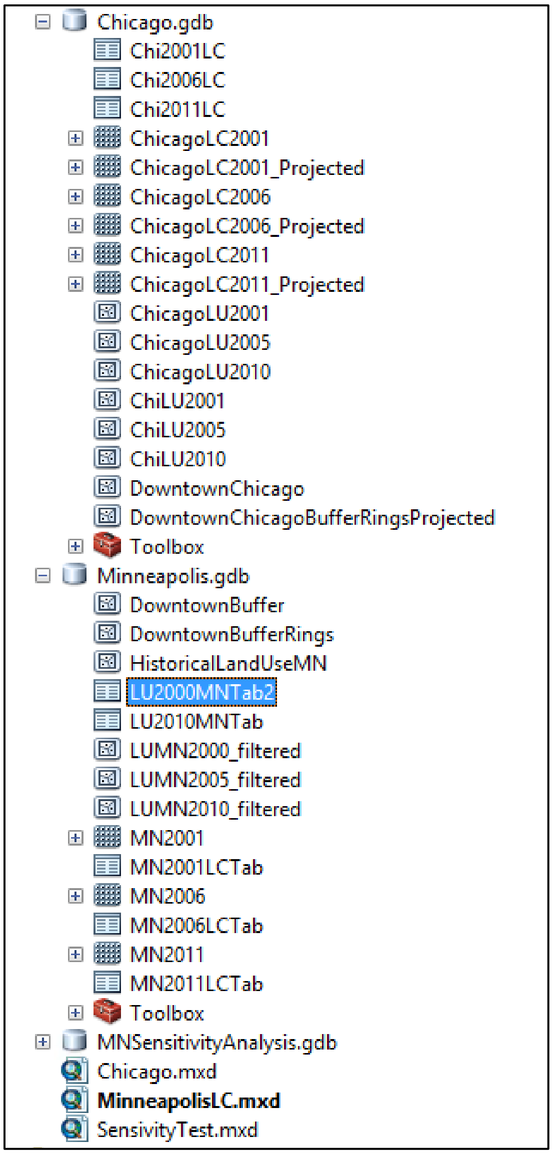

Figure 3. *Geodatabase organization*

The Minneapolis data was projected into the NAD_1983_UTM_Zone_15N projected coordinate system with a Transverse Mercator projection. The Chicago data was projected into the NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet projected coordinate system with a Transverse Mercator projection.

3.4 Processing the Data

After all the data had been gathered, both the land use and land cover datasets were examined, and a downtown area was delineated based on a visual comparison of high density land cover and land use areas. Clustered areas of both high density land use and land cover types were defined as the core downtown area. This downtown area serves as the central core from which the buffer zones will radiate out. 20 multiple ring buffers were created around the downtown polygons at 1 mile intervals in Minneapolis, and 40 multiple ring buffers were created at 1 mile intervals in Chicago.

3.4.1. Creating buffers

A series of 20 concentric buffer zones for Minneapolis, and 40 buffer zones for Chicago set at 1 mile intervals were created around the city center using ArcMap. The total number of pixels within each concentric buffer was measured, and the proportion of low density developed area within each zone was calculated as part of the Shannon's Entropy calculation.

In other studies, researchers used round or square buffers that covered the entire urban region, such as the entire city of Pune, India. Pune is surrounded by rural land, so it is possible to view the expansion of urban area without considering political and administrative forces that might shape development. The Chicago and Minneapolis metropolitan areas are a collection of multiple cities, but the buffer zones captured the sprawl phenomenon across a significant portion of the entire region.

Using the Multiple Ring Buffer tool on ArcGIS Model Builder (Figure 3), buffers for each city were created at 1 mile distances. These distances were chosen because they cover the majority of the metropolitan areas. The buffers were dissolved to eliminate any overlap between buffers.

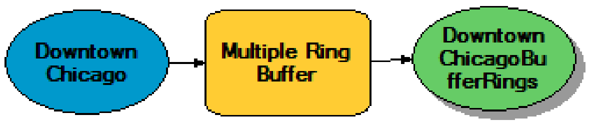

Figure 3. Model for creating buffers using city boundaries.

3.4.2. Processing the land cover data

The land cover data for Minneapolis was projected using the Project Raster tool. Because the Chicago land use data was restricted to an area comprised of several counties, the land use data had to be clipped to the land use data boundaries in order to produce a study area of equivalent size and data content. The land use data was duplicated as an additional layer, and then dissolved to produce a boundary. The dissolve function took a significant amount of time to process, as there were a large number of features to process. The land cover data was then clipped to the dissolved land use feature.

3.4.3. Processing the land use data

The land use data for Minneapolis was projected to the NAD_1983_UTM_Zone_15N projected coordinate system, and the Chicago data was projected into the NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet projected coordinate system using the Project tool. The data were queried to exclude any land uses relating to water, and the resulting selections were exported as new feature classes.

3.5 Calculating Shannon’s Entropy using ArcGIS

The steps for calculating Shannon's Entropy using GIS is partially based on methodology published by the University of North Carolina at Chapel Hill (2016). Using the Zonal Histogram tool, the number of pixels of each individual land cover type within each buffer zone was calculated. The Zonal Histogram tool in ArcGIS generates a table summarizing the pixel count of each land cover type by zone. The tool calculates the intersection of the raster land cover data’s pixel values and the vector buffer zones.

Using the Tabulate Intersection tool, the area of each land use type within each buffer zone was calculated. The Tabulate Intersection tool generates a tale summarizing the area of each land use type by zone. The tool calculates the intersection of the vector land use polygons with the vector buffer zones. While the Zonal Histogram tool is used for raster data, the Tabulate Intersection tool is used for vector data. The output land cover and land use tables were converted to an Excel table using the Table to Excel tool, and further calculation steps were completed in Microsoft Excel.

3.5.1. Calculating Shannon’s Entropy using the land cover data

Calculating Shannon’s entropy with the land cover data involves calculating the proportion of low intensity land cover pixels to total pixels in each zone, dividing the result by the total of all low intensity land cover pixels, multiplying the result by the negative log of the low intensity land cover divided by total low intensity land cover, and finally adding up the results of each zone to achieve the final Shannon’s Entropy result. These calculations were executed in Microsoft Excel for streamlined and flexible data management. The total for each land cover category in each zone, excluding water, was summed. The proportion of low intensity land cover (represented with a value of ‘22’) in relation to total land cover of the zone was calculated (DENi). Then, the DENi for each zone was divided by total low intensity land cover across all zones to calculate PDENi.

After PDENi has been created and populated, a new calculated field multiplying PDENi and log(1/PDENi) is created. The next step is to sum up the calculated values for low intensity land cover. This step adds the values of PDENi * log(1/PDENi) value for low intensity land cover of all the buffer zones. The final step to calculating Shannon's Entropy is to divide the sum of the low intensity land cover values by log(n) for this study. The result of this step gives the relative Shannon's Entropy value for the dataset.

3.5.2. Calculating Shannon's Entropy using land use data

The land use data was also processed following a similar methodology for the NLCD data. The Tabulate Intersection tool was used to calculate the area per land use type in each buffer zone. The resulting table was then exported to Excel, where the proportion of single family detached land use types relative to the total area of the zone (represented as a code of 113 in Minneapolis and 1110/1111 in Chicago for the years of 2000-2005 and 2010, respectively) was calculated (DENi). The sum of all the values of DENi for each zone were calculated and used to calculate PDENi. Then PDENi was multiplied by log(1/PDENi), and the resulting numbers for each zone were summed to produce the Shannon's Entropy number. Relative Shannon's Entropy was then calculated by dividing Shannon's Entropy by log(n)

3.5.3. Observing Entropy through time

The temporal component of the study involves calculating the change in entropy over time. A decrease in entropy over time indicates that the city is developing in a compact, concentrated manner, while an increase in entropy indicates that the city is developing in a dispersed manner.

$$\Delta E = E(t_2) - E(t_1)$$

WHERE

t = time

Equation 2. The change in Entropy values can be used to evaluate the change in the degree of urban sprawl (Yeh and Li 2001).

The expected outcome is that Shannon's Entropy value for land cover data and land use data will be similar across the study period. The value of Shannon's Entropy and resulting change in Entropy for the land use data is assumed to be the “control” with which to compare the usability of NLCD data for urban growth studies. If the values calculated from both types of datasets are determined to have a high degree of similarity, NLCD data may be a viable alternative to cadastral data for urban growth studies. This outcome would also mean that remote sensing data may be viable for use in urban sprawl studies that use Shannon's entropy as the main metric. If the values are dissimilar, land use data is recommended for future urban growth studies.

3.5.4. Sensitivity Analysis

Shannon’s Entropy is run twelve times (for two cities, three years, two data types) for this study. A sensitivity analysis of the buffer zones is conducted on Minneapolis for the year 2001. 20 buffer zones with one mile widths are used in the study. The sensitivity analysis tested the value of Shannon’s Entropy using 4 buffer zones with five mile widths and 39 buffer zones with half mile widths. The 4 buffer zones were chosen to test the result of Shannon’s Entropy if the number of buffer zones were reduced by 1/5th, while increasing the size of the buffer zones by 5 times. The 39 buffer zones were chosen to test the result if the number of buffer zones were increased by almost 2 times, while reducing the size of the buffer zones by 1/2th. The results of the sensitivity analysis show if there is a significant effect on the metric due to different levels of aggregation of the variables being tested.

The values for Shannon’s Entropy are placed in a results table and compared. If the values differ by more than 5% between land cover and land use data, caution may be advised for future studies measuring urban sprawl with land cover data. The results also inform whether the three cities are indeed sprawling, and the magnitude of how sprawl has changed over time in the study area.

Chapter 4 Results

The chapter is broken into three sections: Minneapolis metro area results, Chicago metro area results, and sensitivity analysis results. The first two sections discuss the results of the calculations on each respective study area, as well as the change in entropy over time. The last section discusses the results of the sensitivity analysis on the Minneapolis metro area.

The results of the Shannon's Entropy calculations help inform the similarity between the land use and land cover data. This study includes two findings. The primary finding is that there is no significant difference between the low density land use and land cover class Shannon's entropy results. Because the two data types show consistent results, the land cover data may be suitable for use in urban footprint studies on other metropolitan areas. The secondary finding is that the entropy values for both the study areas are high – ranging from 0.97-0.98 out of a maximum of 1 and minimum of 0. This finding suggests that the variable is evenly dispersed throughout the study area, and that there was no significant growth in the variable during the study period.

4.1 Minneapolis metro area results

The Shannon's Entropy calculation was performed using Minneapolis land use and land cover data from 2000-2010. 20 buffer zones at one mile intervals originating from a downtown center were used to observe the dispersion of the low-intensity residential land cover/land use variable in the Minneapolis metro area. The land use data throughout the study period can be found in Figure 4, with single family (detached) residential land use polygons highlighted in red. The land cover data throughout the study period can be found in Figure 5, with low intensity urban land cover pixels highlighted in red.

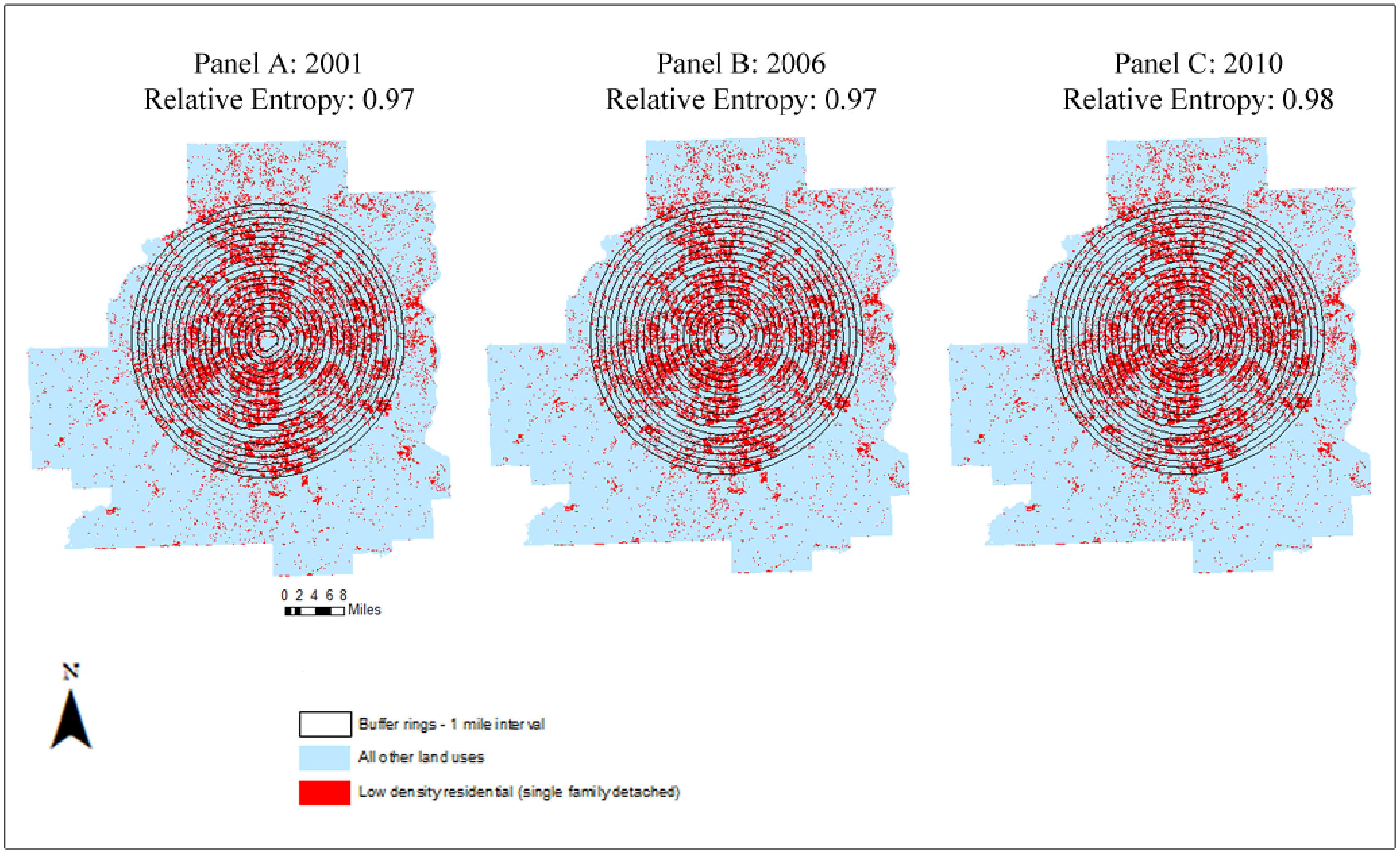

Figure 4. Single family (detached) residential land use polygons in the Minneapolis metro area. Panel A: Year 2001. Panel B: Year 2005. Panel C: Year 2010

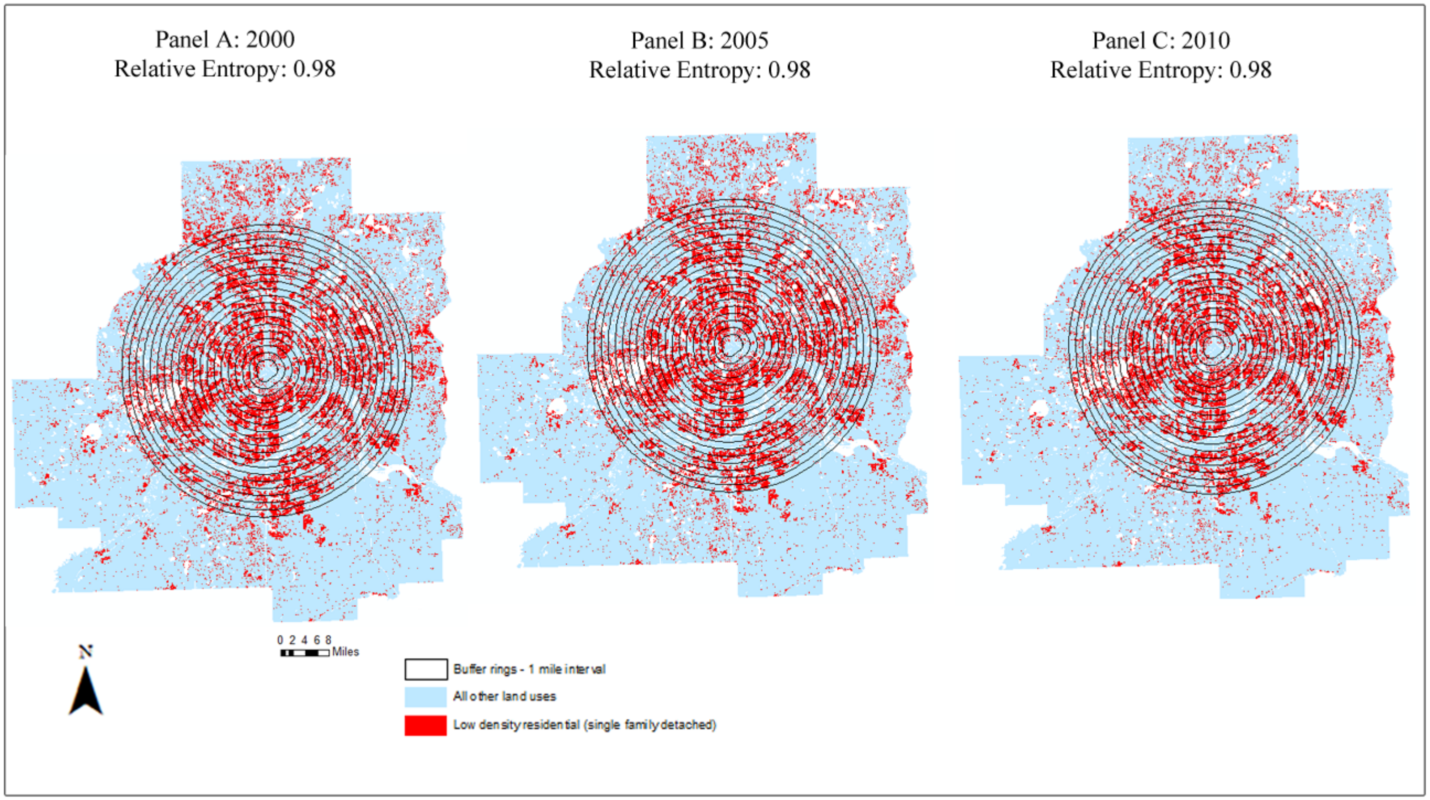

Figure 5. Low-intensity urban land cover in the Minneapolis metro area. Panel A: Year 2001. Panel B: Year 2006. Panel C: Year 2011.

The Shannon's Entropy results for the Minneapolis-St. Paul metropolitan area can found in Table 2. Both the entropy and relative entropy are reported for all three years.

Table 2. Shannon's Entropy results for the Minneapolis metro area, 2000-2010

	Land Use		Land Cover	
Year	Entropy	Relative Entropy	Entropy	Relative Entropy
2001	1.27	0.98	1.26	0.97
2006	1.28	0.98	1.20	0.97
2011	1.28	0.98	1.27	0.98

The results indicate that entropy increased from years 2001-2011. The land use and land cover results differ by 1/100ths in both 2001 and 2006, but differ by only 1/5000ths in 2011. As the results show, the difference between the land use and land cover data is not significant. This means that the land use and land cover data used in this study is consistent and may be used comparably when performing the Shannon's Entropy metric. The difference in entropy values across the years differs by less than 1%, which indicates that the entropy values did not vary significantly across the 10-year study period.

Relative entropy is scaled from a minimum value of 0 to a maximum value of 1, so a value of 0.97-0.98 as observed in the results for Minneapolis indicate that the entropy is high for all three years. This means that low density, single family (detached) land cover is evenly dispersed amongst all 20 zones of the study area.

4.1.1. Change in Entropy over time – Minneapolis

The change in Entropy was calculated using the formula from Equation 2. The results can be found in Table 3.

Table 3. Change in entropy over time, Minneapolis metro area

Year	Land Use	Land Cover
2001-2006	0.0029	0.0034
2006-2011	0.0014	0.0021

The change in entropy over time indicates that the change in entropy from 2001-2006 is greater than the change in entropy from 2006-2011 in both the land use and land cover datasets. However, the magnitude of change is small, within the 1/1000ths. This means that the dispersion of low intensity land cover and low density residential land use did not increase or decrease significantly throughout study period.

4.2 Chicago metro area results

The Shannon's Entropy calculation was performed using Chicago land use and land cover data from 2000-2010. 20 buffer zones at 1 mile intervals originating from a downtown center were used to observe the dispersion of the low-intensity residential land cover/land use variable in the Chicago metro area. The land use data throughout the study period can be found in Figure 6, with single-family (detached) residential land use polygons highlighted in red. The land cover data throughout the study period can be found in Figure 7, with low intensity urban land cover pixels highlighted in red.

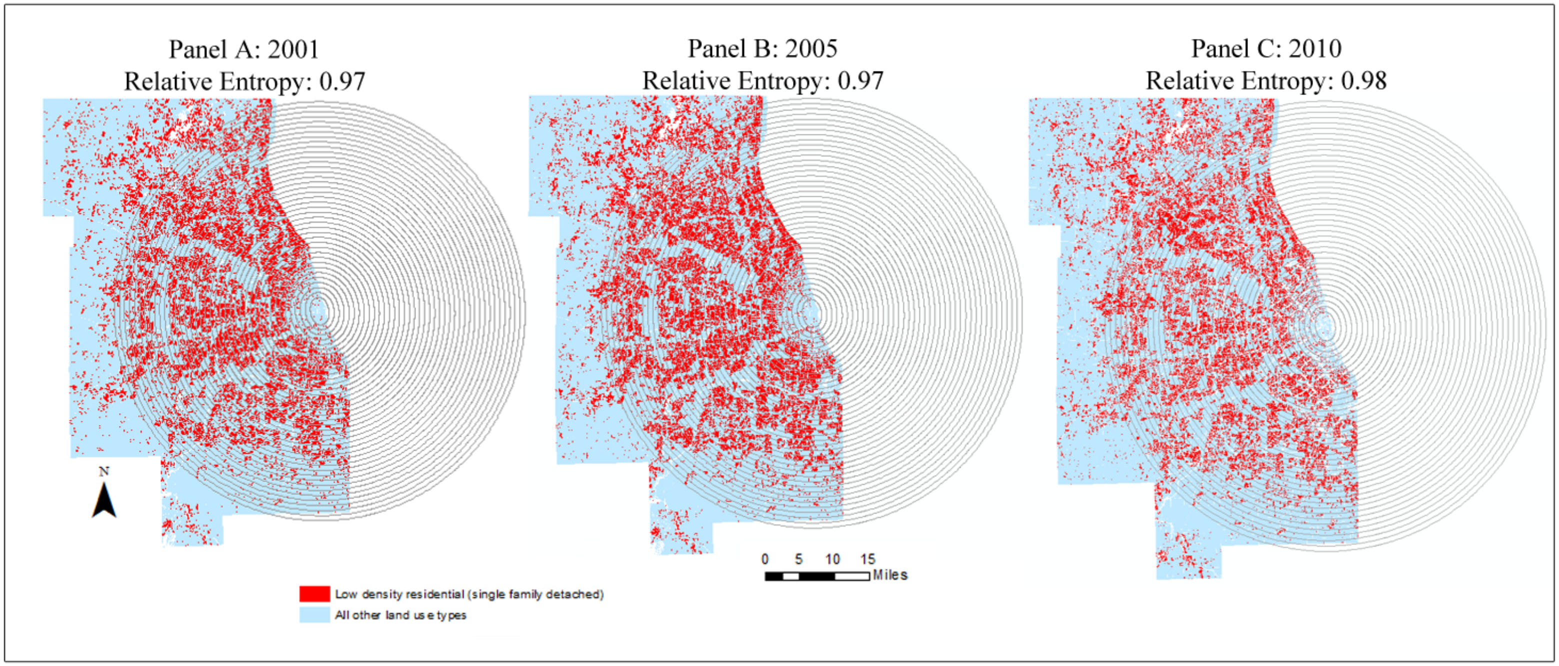

Figure 6. Single family (detached) residential land use polygons in the Chicago metro area. Panel A: Year 2001. Panel B: Year 2005. Panel C: Year 2010.

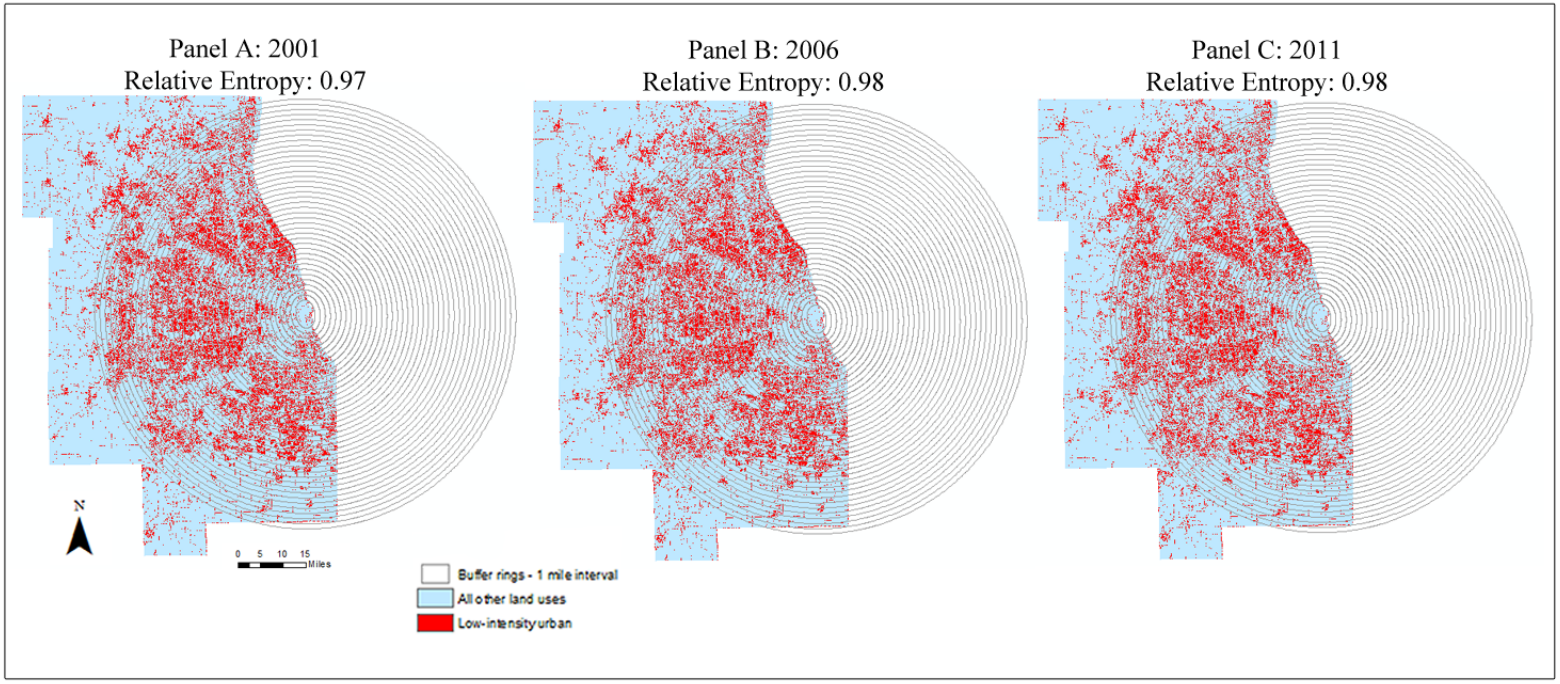

Figure 7. Low-intensity urban land cover in the Chicago metro area. Panel A: Year 2000. Panel B: Year 2005. Panel C: Year 2010.

The Shannon’s Entropy results for the Chicago-Naperville-Elgin metropolitan area can be found in Table 4. Both the entropy and relative entropy are reported for all three years.

Table 4. Shannon’s Entropy results for the Chicago metro area, 2000-2010

	Land Use		Land Cover	
Year	Entropy	Relative Entropy	Entropy	Relative Entropy
2001	1.56	0.97	1.56	0.97
2006	1.56	0.97	1.57	0.98
2011	1.57	0.98	1.57	0.98

The results indicate that entropy increased from years 2001-2011 according to the land use data, but decreased slightly from 2006-2011 according to the land cover data. The land use and land cover results differ by 1/100ths in all three years. As the results show, the difference between the land use and land cover data is not significant. This means that the land use and land cover data used in this study are comparable. A Shannon’s entropy value of 0.97-0.98 as observed in the results for Chicago indicate that the entropy is high for all three years. This high value of entropy indicates that that no single zone contains the variable disproportionately over the other zones in the study area.

The land use and land cover data differs by less than 1% for all three years, far lower than the less than 5% threshold used to establish the usability of the land cover data. Because the difference between the results of the different data types is less than 1%, land cover data may be considered as an alternative to land use data for urban footprints studies using the Shannon’s Entropy metric. The difference in entropy values across the years also differs by less than 1%, which indicates that the entropy values did not vary significantly across the ten-year study period.

4.2.1. Change in Entropy over time – Chicago

The change in Entropy was calculated using the formula from Equation 2. The results can be found in Table 5.

Table 5. Change in entropy over time, Chicago metropolitan area

Year	Land Use	Land Cover
2001-2006	0.0031	0.0026
2006-2011	0.0026	-0.0000

The change in entropy over time indicates that the change in entropy from 2001-2006 is greater than the change in entropy from 2006-2011 in both the land use and land cover datasets. The magnitude of change is small, within the 1/1000ths in the both the datasets. However, the entropy in land cover appears to have decreased from 2006-2011. The decrease in entropy means that the proportion of low intensity urban land cover decreased across the zones, either through the increase of high or medium intensity urban land cover, or conversion to a different land cover type. The increase in entropy from 2006-2011 as observed in the land use data contrasts with the decrease in entropy for the same time period observed in the land cover data. This discrepancy is likely due to differences between how land use and land cover data characterizes land, as well as precision differences between raster and vector data. This discrepancy may also be due to administrative or update errors in the land use data, or misclassification in the land cover data.

While land cover data categorizes urban land cover category data based on percent impervious surface, land use data is based on a mix of administrative records and physical characteristics of the land. The vector data are more likely to precisely delineate features, while the raster data classifies mixed pixels as singular land cover types. The 30 by 30 spatial resolution of the NLCD data is a coarse enough resolution to be subject to some classification imprecision. However, NLCD raster data may more accurately reflect how the land is being used than singular vector parcels that are larger than the 30 by 30 resolution of the NLCD data. This is because a large vector parcel that is classified as "urban low density" development may in reality be a singular house on a very large, multi-acre lot. There are disadvantages to using each type of data, so a visual comparison of the data to aerial or satellite imagery is recommended to assess the accuracy of representation.

4.3 Sensitivity analysis results

A sensitivity analysis was performed to determine the sensitivity of the Shannon's Entropy metric to the number and size of buffer zones. The sensitivity test was performed on the Minneapolis-St. Paul metropolitan area land use data for the year 2000, and land cover data for the year 2001. The test measured Shannon's Entropy using 39 buffer zones at ½ mile intervals, nearly doubling the number of zones while halving the size of the zones. The test also measured Shannon's Entropy using 5 buffer zones at 4 mile intervals, decreasing the number of buffer zones by 1/4th, while increasing the size of the zones by 4. The results of the sensitivity analysis can be found in Table 6.

Table 6. Sensitivity analysis results for the Minneapolis metro area

Minneapolis	Land Use (2000)		Land Cover (2001)	
Observed Results	Entropy	Relative Entropy	Entropy	Relative Entropy
20 buffers at 1 mile intervals	1.27	0.98	1.26	0.97
Minneapolis	Land Use (2000)		Land Cover (2001)	
Sensitivity Test	Entropy	Relative Entropy	Entropy	Relative Entropy
39 buffers at 1/2 mile intervals	1.57	0.98	1.55	0.97
5 buffers at 4 mile intervals	0.68	0.97	0.66	0.95

The sensitivity analysis results indicate that increasing the number of buffers while decreasing the size of the buffers can provide similar values to the observed in the man study. Decreasing the number of buffers while increasing the size of the buffers, provide results that differ by more than 1/1000ths, which is not a large enough difference to label the metric as extremely sensitive to the size and shape of the buffers. Using a greater number of smaller buffer zones is typically more recommended than using a small number of large buffer zones, due to potential information loss from aggregation effects when using larger buffers (Yeh and Li 2001).

Chapter 5 Conclusions

This study compared between two different data types – remotely sensed land cover imagery, and cadastral land use vector data – to determine if land cover imagery could be suitable for use in urban footprint studies. The main metric used to compare the data is a metric called Shannon's Entropy, a concept rooted in information science that measures spatial dispersion of a variable.

The results of the Shannon's Entropy calculation show that there is no significant difference between the land use and land cover data in Minneapolis and Chicago for the duration of the study period. However, there exist differences in the patterns shown by land use and land cover data in Chicago – where land use data reports a slight increase in entropy, the land cover reports a very slight decrease in entropy. This chapter summarizes overall findings, discusses considerations and limitations, and suggests future work for studies of this nature.

5.1 Major Findings

The results of the analysis helped inform whether land cover data may be suitable for use in urban footprint studies where land use data is unavailable or too cumbersome to obtain. Many past studies of urban growth have used the Shannon's Entropy metric to quantify the dispersion of an urban land cover variable using classified imagery data. These past studies have often been conducted data-poor countries, where cadastral land use data is absent, not readily available, or is of poor data quality. Few studies have compared the suitability and accuracy of the land cover data for urban footprint studies with cadastral land use data, which is assumed to be more precise because it is produced at a finer scale than land cover data. Thus, the current study performed the comparison between the two data types in order to inform future studies of the usability of land cover data, particularly the National Land Cover Dataset and other data utilizing similar classification methodologies.

The current study found that the entropy levels in both study areas was high, ranging from 0.97 to 0.98 out of a maximum value of 1.00 in the study period. This high number means that the low density residential land use or low intensity land cover is evenly dispersed amongst all zones of the study area. The study also found that the data showed a general trend of entropy increasing throughout the study period, although the differences between the years were not statistically significant.

The Minneapolis metropolitan area showed a slight increase in entropy over the study period with both data types, while the Chicago metropolitan area only showed an increase in entropy with the land use data, where the land cover data showed a slight decrease in entropy. The discrepancy examined was not at a statistically significant level, but the difference may highlight uncertainties in either of the datasets. The discrepancy may be due a difference in accuracy between land use and land cover data, a recordkeeping or update error in land use data, or a classification error in the land cover data. Further studies of other metropolitan areas over longer study periods are recommended to determine if differing entropy patterns between the data types are indeed negligible.

The low relative sensitivity to changing parameters such as buffer size and number of zones makes Shannon's Entropy a suitable metric for detecting if a significant amount of growth has occurred in an urban area while giving researchers flexibility in selecting study parameters. While no significant changes were detected in the current study, researchers such as Yeh and Li (2001) have been able to use Shannon's Entropy to detect significant changes in rapidly developing regions like the heavily populated Pearl River Delta in China.

The wide spatial and temporal coverage of medium to high resolution imagery, as well as improving satellite and sensor technology, make remotely sensed imagery invaluable in urban growth studies, where urban land can be identified by presence of impervious surfaces. With finer spatial and spectral resolution sensors, satellite imagery can also assist with identifying land uses. The percentage of impervious surface can broadly identify how land is being used, as individual land uses are typically aligned with certain spectral signatures or are composed of a percentage of impervious surfaces.

While land cover data is unable to discern between specific types of land use – such as industrial versus manufacturing land uses, the data can adequately identify the footprint of the land use, and can help users discern if land resources can be allocated more effectively. The difficulty in obtaining usable, consistent historical land use data also highlights the merits of using land cover data for urban footprint studies spanning multiple administrative boundaries. Land use data may not be necessary to measure urban growth patterns such as urban sprawl, as alternative datasets such as land cover data may be used as effectively.

5.2 Considerations and Limitations

Accuracy assessments are often an important element of land cover studies by assessing the usability of classification work. A common method of accuracy assessment involves computing an Error Matrix to determine the percentage of correctly classified pixels. This study did not perform an additional accuracy assessment, as the NLCD pre-classified land cover data has been found to meet the minimum 85% accuracy threshold recommended by Anderson et al. (1976). The land use data is assumed to exhibit greater than 85% accuracy, and did not need an accuracy assessment.

The main limitation of the study is due to the limited number of iterations of Shannon's Entropy performed. Because of the uniform methods to classify land use data within a singular county, interstate measurements of urban sprawl may yield widely differing results. This is because other states may use different land use classification standards, or may use extensively modified Anderson classifications. An extended study period could also increase the number of iterations of Shannon's Entropy performed, as well as provide more detailed insight into the patterns of entropy observed in the study areas over time. This study intended to use 1992 National Land Cover data to extend the study period to a 19-year span of time. Stretching the study period would have also covered a span of time that encompassed a large period of suburban growth in the United States. Unfortunately, the study was limited by the availability or usability of such data. The 1992 National Land Cover data is not directly comparable with the 2001-2011 land cover data due to differences in how several land cover classes - including the low intensity urban land cover class – were defined.

The study also intended to incorporate more metropolitan areas to perform more iterations of the metric. However, there was a lack of historical land use data in many of the major U.S. cities. Many cities only had the most current land use data available, and did not keep archives of historical data. Although the land use data used was sufficient for the needs of the study, the data was limited in coverage, and the land cover data needed to be clipped to the boundaries of the land use data. The land use data needed to be dissolved in order to easily clip the land cover data, and a significant amount of time was spent performing the dissolve function on the land use data.

This study performed the Shannon's Entropy calculation on only one prominent type of land use and land cover – single family detached residential housing and low intensity urban land cover. Changes in other land use or land cover types, such as single family attached residential units, multifamily units, or medium intensity land cover, would not be well detected in this study. For example, if the city of Chicago had added 5,000 new medium density units on several parcels from 2005-2010 across several of the zones, the change would not have made a significant difference to the value of Shannon's Entropy for the variables studied. Changes that may have appeared to make an impact from ground observation may not have been effectively captured by the study due to the emphasis on singular variables.

Another limitation of this study is the lack of population data used to better inform the results. This is especially true for the land use data, which may classify a single land use to a large parcel that may not be completely inhabited. This kind of model would involve using dasymetric mapping to more accurately map out developable and undevelopable areas before calculating Shannon's Entropy. This way, undevelopable areas can be excluded from the calculation. Another method of increasing accuracy could include filtering out a wider range of land use parcels that should not be included in the study. Land use parcels with low density residential units with areas greater than a threshold size can be excluded from the calculation, as it would confound the results. This would also allow for a more strict interpretation of low density residential land use classes.

Despite the limitations of the study, the study conclusively finds that there is no significant difference between the different data types. Land cover data may provide comparable results to land use data when performing urban footprint studies using the Shannon's Entropy metric. Growing cities and metropolitan areas in data-poor countries that do not or cannot support a cadastral infrastructure can benefit from using land cover data to measure urban growth patterns such as sprawl. The measurement of such patterns can provide insight into how communities are growing, and can identify areas where urban land may be used more efficiently.

5.3 Future Work

The current study may be expanded to study different areas of the United States or other countries where vector cadastral data is readily available. Future studies would also benefit from an elongated study period to include historical and future data in order to observe entropy over a longer period of time. A longer study period can give more insight into whether land use and land cover data observe similar patterns of entropy over time, as well as more insight into the changing urban landscape of the study areas. There are existing plans to produce synchronized NCLD products, including the 1992 datasets, as of 2016. A reclassified 1992 NCLD dataset that is directly comparable with NLCD datasets produced after 2001 would allow for extended observations of urban growth in the U.S.

Future studies can benefit from combining different land use and land cover types to measure entropy, or from examining different land classes than the ones used in the current study. Different entropy values may ensue from combining single family detached housing with low density multifamily housing, or from combining low intensity urban land cover with medium intensity urban land cover. Studying high intensity urban land cover exclusively can measure the concentration of densely inhabited areas throughout the study region.

REFERENCES

Abadie, Abel. 2005. ‘Semiparametric Difference-in-differences Estimators.” *The Review of Economic Studies* 72: 1-18.

Alsharif, A., Pradhan, B., Mansor, S., Shafri, H. 2015. Urban expansion assessment by using remotely sensed data and the relative Shannon Entropy model in GIS: A case study of Tripoli, Libya. *Theoretical and Empirical Researches in Urban Management* 10(1), 55-71.

Anderson, J., Hardy, E., Roach, J., Witmer, R. 1976. A land use and land cover classification system for use with remote sensor data. *USGS Professional Paper 964.* A revision of the land use classification system as presented in the USGS Circular 671.

Bhatta, B., Saraswati, S., Bandyopadhyay, D. 2010. Urban sprawl measurement from remote sensing data. *Applied Geography* 30, 731-740.

Belal, A., Moghanm, F. 2011. Detecting urban growth using remote sensing and GIS techniques in Al Gharbiya governorate, Egypt. *The Egyptian Journal of Remote Sensing and Space Sciences* 14, 73-79.

Cabral, P., Augusto, G., Tewolde, M., Araya, Y. 2013. Entropy in Urban Systems, *Entropy* 15, 5223-5236.

Crowther, Richard J., Jr., “A comparison of urban land cover change: A study of Pasadena and Inglewood, California, 1992-2011” (master’s thesis, University of Southern California, 2015).

Dadras, Mohsen, Helmi Z. M. Shafri, Noordin Ahmad, Biswajeet Pradhan, and Sahabeh Safarpour. 2015. Spatio-temporal analysis of urban growth from remote sensing data in Bandar Abbas City, Iran. *The Egyptian Journal of Remote Sensing and Space Science* 18(1), 35-52.

Deka, J., Tripathi, O., Khan, M. 2010. Urban growth trend analysis using Shannon Entropy approach – A case study in North-East India. *International Journal of Geomatics and Geosciences*, 2(4), 1062-1066.

Effat, H., Shobaky, M., 2015. Modeling and mapping of urban sprawl pattern in Cairo using multi-temporal Landsat images, and Shannon's Entropy. *Advances in Remotes Sensing* 4, 303-318.

Epstein, J., Payne, K., Kramer, E. 2002. Techniques for Mapping Suburban Sprawl. *Photogrammetric Engineering and Remote Sensing* 68(9), 913-918.

Eric N. Pena, “Using Census Data, Urban Land-Cover Classification, and Dasymetric Mapping to Measure Urban Growth of the Lower Rio Grande Valley, Texas” (master's thesis, University of Southern California 2012).

“Frequently Asked Questions”, Multi Resolution Land Characteristics Consortium, accessed 10 June, 2016, http://www.mrlc.gov/faq.php.

Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N., and Wickham, J. 2011. Completion of the 2006 National Land Cover Database for the Conterminous United States, *PE&RS*, 77(9), 858-864.

Galster, G., Hanson, R., Ratcliffe, M., Wolman, H., Coleman, S. 2010. Wrestling sprawl to the ground:defining and measuring an elusive concept. *Housing Policy Debate,* 12(4), 681-717.

“Generalized Land Use – Historical 1984, 1990, 1997, 2000. 2005, and 2010, for the Twin Cities Metropolitan Area”, Metropolitan Resource Council, accessed August 15th 2016, https://metrocouncil.org/Data-and-Maps/Data/Metadata/Landuse-Hist-Research.aspx#Entity_and_Attribute_Information.

Guo, R., Li, L., Ying, S., Luo, P., He, B., Jiang., R. Developing a 3D cadaster for the administration of urban land use: A case study of Shenzhen, China. *Computers, Environment and Urban Systems,* 40, 46-55.

Herold, M. 2005. The role of spatial metrics in the analysis and modeling of urban land use change. *Computers, Environment and Urban Systems*, 29(4), 369-399.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J.N., and Wickham, J. 2007. Completion of the 2001 National Land Cover Database for the Conterminous United States. *Photogrammetric Engineering and Remote Sensing*, 73(4), 337-341.

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K.2015. Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, 81(5), 345-354.

Hurtt, G.C., Rosentrater, L., Frolking, S., Moore, B. 2001, Linking remote-sensing estimates of land cover and census statistics on land use to produce maps of land use of the coterminous United States. *Global Biogeochem. Cycles*, 15(3), 673-685.

Illinois Department of Commerce & Economic Opportunity. “Target Industries in Illinois.” Accessed October 15th, 2016. https://www.illinois.gov/dceo/whyillinois/TargetIndustries/Pages/default.aspx.

Ismail, M., Kamaruzaman, J.2008. Satellite classification accuracy assessment based from reference dataset. *International Journal of Computer and Information Science and Engineering*, 2(2), 96-102.

Irwin, E., Bell, K. 2003. Modeling and managing urban growth at the rural-urban fringe: A parcel-level model of residential land use change. *Agricultural and Resource Economics Review* 32(1), 83-102.

Irwin, E. 2007. The evolution of urban sprawl: Evidence of spatial heterogeneity and increasing land fragmentation. *Proc Natl Acad Sci USA* 104(52), 20672-20677.

Krueger, Samuel, “Delimiting the Postmodern Urban Center: An Analysis of Urban Amenity Clustersin Los Angeles” (master's thesis, University of Southern California, 2012).

“Land Use Inventory”, Chicago Metropolitan Agency for Planning, accessed August 15 2016, http://www.cmap.illinois.gov/data/land-use/inventory.

Li, J., Qiu, R., Xiong, L., Xu, J.2015. A gravity-spatial entropy model for the measurement of urban sprawl. *Science China Earth Sciences* 59(1), 207-213.

Longley, P. 2002. Geographical Information Systems: will developments in urban remote sensing and GIS lead to 'better' urban geography? *Progress in Human Geography* 26(2), 231-239.

Masoumi, H., Roque, D. 2015. Evaluation of urban sprawl speed and intensity based on international urbanization. Example from a Mexican city. *Journal of Settlements and Spatial Planning* 6, 27-35.

Minnesota Department of Employment and Economic Development. “Major Industries and Sectors.” Accessed September 15th, 2016. https://www.mn.gov/deed/business/locating-minnesota/industries-sectors/.

National Research Council. “Background.” In *National Land Parcel Data: A Vision for the Future*, 1-15. Washington, DC. National Academies Press, 2007.

Openshaw, S., “A spatial analysis research agenda.” In *Handling Geographical Information: Methodology and Potential Applications*, 18-37. Harlow, Essex, England: Longman Scientific and Technical, 1991.

Shannon, C. 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal* 27, 379-423.

Singh, B. 2014. Urban Growth Using Shannon’s Entropy: a Case Study of Rohtak City. *International Journal of Advanced Remote Sensing and GIS* 3(1), 544-552.

“Steps for Computing the Dissimilarity, Entropy, Herfindahl-Hirschman and Accessibility (Gravity with Competition) Indices”, University of North Carolina, Chapel Hill, accessed 20 June 2016, http://planningandactivity.unc.edu/stepsdehaccessibiityindic.pdf.

Sudhira, H., Ramachandra, T., Jagadish, K. 2004. Urban sprawl: metrics, dynamics, and modelling using GIS. *International Journal of Applied Earth Observation* 5, 29-39.

Sun, H., Forsythe, W., Waters, N. 2007. Modeling urban land use change and urban sprawl: Calgary, Alberta, Canada. *Networks and Spatial Economics* 7(4), 353-76.

Thomas, R.W. 1981., "Information Statistics in Geography in Concepts and Techniques in Modern Geography" In *Concepts and Techniques in Modern Geography 31*. Norwich, England: Geo Abstracts, 1981.

Tsai, Y. 2005. Quantifying urban form: compactness versus 'sprawl'. *Urban Studies* 42 (1), 141-161.

U.S. Census Bureau. 2015. American FactFinder. Retrieved September 1, 2016, from U.S. Census Bureau: http://census.factfinder2.gov.

U.S. Census Bureau. 2013. American FactFinder. Retrieved October 1, 2016, from U.S. Census Bureau: http://census.factfinder2.gov.

U.S. Census Bureau. 2011. American FactFinder. Retrieved June 10, 2016, from U.S. Census Bureau: http://census.factfinder2.gov.

U.S. Census Bureau. 2000. American FactFinder. Retrieved July 25, 2016, from U.S. Census Bureau: http://census.factfinder2.gov.

U.S. Census Bureau. 2016. U.S. and World Population Clock. Retrieved July 25 2016, from U.S. Census Bureau: http://www.census.gov/popclock/.

Wickham, J., Stehman, S., Gass, L., Dewitz, J., Fry, J., Wade, T. 2013. Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sensing of Environment* 130, 294-304.

Wolman, H., Galster, G., Hanson, R., Ratcliffe, M., Furdell, K., Sarzynski, A. 2005. The fundamental challenge in measuring sprawl: Which land should be considered? *The Professional Geographer* 57(1), 94-105.

Wu, S., Qiu, X., Usery, E., Wang, L. 2009. Using geometrical, textural, and contextual information of land parcels for classification of detailed urban land use. *Annals of the Association of American Geographers*. 99(1), 76-98.

Yeh AG, Li X. 2001. Measurement and monitoring of urban sprawl in a rapidly growing region using entropy. *Photogrammetric Engineering & Remote Sensing* 67, 83–90.

Zahar, A. 2012. Assessing usefulness of high-resolution satellite imagery in GIS-based cadastral land information system. *Journal of Settlements and Spatial Planning* 3(2), 93-96.