2014

Landscape and Impervious Surface Mapping in the Twin Cities Metropolitan Area using Feature Recognition and Decision Tree techniques

Philipp Nagel
Minnesota State University - Mankato

Landscape and impervious surface mapping in the Twin Cities Metropolitan Area using Feature Recognition and Decision Tree techniques

By

Philipp Nagel

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

Master of Science

In

Geography

Minnesota State University, Mankato

Mankato, Minnesota

May 2014

Landscape and impervious surface mapping in the Twin Cities Metropolitan Area using Feature Recognition and Decision Tree techniques

Philipp Nagel

This thesis has been examined and approved by the following members of the student's committee.

_____
Fei Yuan, Ph. D.

_____
Martin Mitchell, Ph. D.

_____
Cynthia Miller, Ph. D.

Acknowledgments

I would like to thank my advisor, Dr. Fei Yuan, for her continuous support, encouragement, and dedication. She persistently pushed me to do my best work. Her guidance in not only my thesis research and writing, but also my previous endeavors at the Minnesota State University (MSU) was essential in my academic success. I would also like to thank her for helping fund this project through her faculty research grant. She is truly a great teacher and advisor.

I would also like to thank the remaining members of my thesis committee, Dr. Cynthia Miller and Dr. Martin Mitchell. Both of you have taught me a tremendous amount during my time at MSU. I am very lucky to have the opportunity to work with and have on my committee such excellent and interesting scholars.

Great thanks to the Department of Geography and all of its faculty and staff for creating an inviting and collegial working atmosphere.

I would like to thank my wife for her encouragement, and pushing me always to my best work.

Finally, my heartfelt thanks go to my parents for their financial support and for being supportive and encouraging even when I moved to a different continent.

Landscape and impervious surface mapping in the Twin Cities Metropolitan Area using Feature Recognition and Decision Tree techniques

Philipp Nagel

Master of Science Thesis in Geography

Minnesota State University, Mankato

May 2014

Abstract

Land Use and Land Cover (LULC) and Impervious Surface Area (ISA) are important parameters for many environmental studies, and serve as an essential tool for decision makers and stakeholders in Urban & Regional planning. Newly available high spatial resolution aerial ortho-imagery and LiDAR data, in combination with specialized, object-oriented and decision-tree classification techniques, allow for accurate mapping of these features. In this study, a method was developed to first classify LULC using an object-based classifier, and then use the resulting map as input for a decision-tree model to classify ISA in the Twin Cities Metropolitan Area in Minnesota.

It was found that vegetation cover classes were the most prevalent in the study area, making up over half of the land area. Water was the smallest class, followed by urban land cover, which made up 11%. Impervious surface was determined to make up 14% of the TCMA area.Overall classification accuracy for LULC cover was estimated to be 74%, and 95% for the ISA classification.

Table of Contents

List of Figures

List of Tables

List of Equations

1. Introduction

Accurate landscape maps, such as land use and land cover (LULC) and impervious surface area (ISA) maps are essential inputs for local decision makers as well as many researchers. While a product with high spatial resolution and great accuracy is desired, the available data today is often only of low to medium spatial resolution, and varying degrees of accuracy.

Extracting LULC and ISA information using high-resolution remote sensing imagery, LIDAR-derived elevation data, and other ancillary data is difficult. This difficulty is in part due to the fact that the input data required to generate land cover maps are often available only at low or medium spatial resolution, whereas high resolution imagery is often priced too high to allow for its efficient use. Further, there is a lack of well-established techniques to process high-resolution spatial data. Many established datasets are also relatively old. In particular, for the study area of the Twin Cities Metropolitan Area (TCMA), the most recent dataset is based on 2006 Landsat imagery with spatial resolution of 30 meters (Yuan 2010) . While this LULC dataset for the study area possesses good accuracy, it is now eight years old. Urban development is fast-paced, hence, an updated product would be desirable to provide more recent data and enable change analysis. New LULC and ISA data for the study area are necessary to help stakeholders assess the effects of urbanization and other LULC changes. Traditional classification techniques may not deliver the best results possible when applied to the newly available, high spatial resolution imagery. In this context, I developed an analytical

method to extract the desired high resolution LULC and impervious surface information using advanced techniques such as object-oriented classification and decision tree modeling.

In the past, methods developed for land features identification have been focused on using medium-resolution satellite images. Since high-resolution aerial imagery and LIDAR data have become more readily available, and computing power has increased, new techniques are becoming more promising. In particular, in this study, I used an object-based classifier to map high resolution land cover types from 1 m digital orthoimagery for the TCMA, Minnesota. I also developed a decision-tree model to extract impervious surface data from a combination of data sources. The resulting data and developed method provide important decision-making inputs and tool for local governments and other agencies and organizations in the area.

1.1. Study Area

This study will focus on the Twin Cities (Minneapolis and St. Paul) Metropolitan Area of Minnesota. In particular, the study area consists of seven counties: Anoka, Carver, Dakota, Hennepin, Ramsey, Scott, and Washington counties (Figure 1.1). These counties have a combined area of about 2,939 square miles. From 2000 to 2010, the population increased from 2.6 million to 2.8 million, which comprised approximately 54% of the total population of Minnesota (Metropolitan Council 2000; US Census 2013). At the center of the TCMA are the cities of Minneapolis and St. Paul. St. Paul is the capital of Minnesota, which has a distinct culture from Minneapolis. There are many populous suburbs, as well as highly commercialized areas. The Mall of America, one of the largest indoor shopping centers in the United States (US) is located in Bloomington, south of Minneapolis. Another large shopping mall, the Southdale Center in Edina, is considered the oldest mall in the US. Several major corporations and Fortune 500 companies, such as Target Corporation in Minneapolis, The Toro Company in Bloomington, Dairy Queen in Edina, 3M in Maplewood, and General Mills in Golden Valley, are headquartered in the Twin Cities.

A large part of the early economy of the Twin Cities was influenced by the presence of the Mississippi River and its Saint Anthony Falls, providing hydropower to sawmills and later flourmills. These facilities at Saint Anthony Falls were some of the first to use hydropower in the US (Anfinson 1995). The area was also a major transportation hub for rail and water cargo and passenger services. Grain was a common good to be shipped into the Twin Cities via river or rail, and consequently, flour and other

milling products were then exported. Timber harvested in Minnesota was also an important shipping good. Manufacturing followed to be a major part of the Twin Cities economy. Today, the economy is dominated by tertiary sector businesses, high-tech research and production, and financial services.

The landscape of the study area is relatively diverse, with a large number of lakes. High density urban development is mostly located in the central part while vegetated land cover and agricultural land are found in the outer perimeter. The maps and data produced in this study will elaborate specific patterns of these land cover classes. The climate in the Twin Cities is typical of the Midwestern US with extreme cold temperatures in the winter and extreme heat in the summer. Precipitation peaks in the summer months. The area is prone to many types of natural disasters, such as tornadoes and other wind storms, flash flooding, extreme temperatures, and winter storms.

The Twin Cities have been found by a variety of surveys over the last years to be one of the most attractive metropolitan areas in the US, and one of the best places to live. These ratings were in large parts due to the proximity to natural features such as lakes, the extensive parks and trails system, and the robust economy.

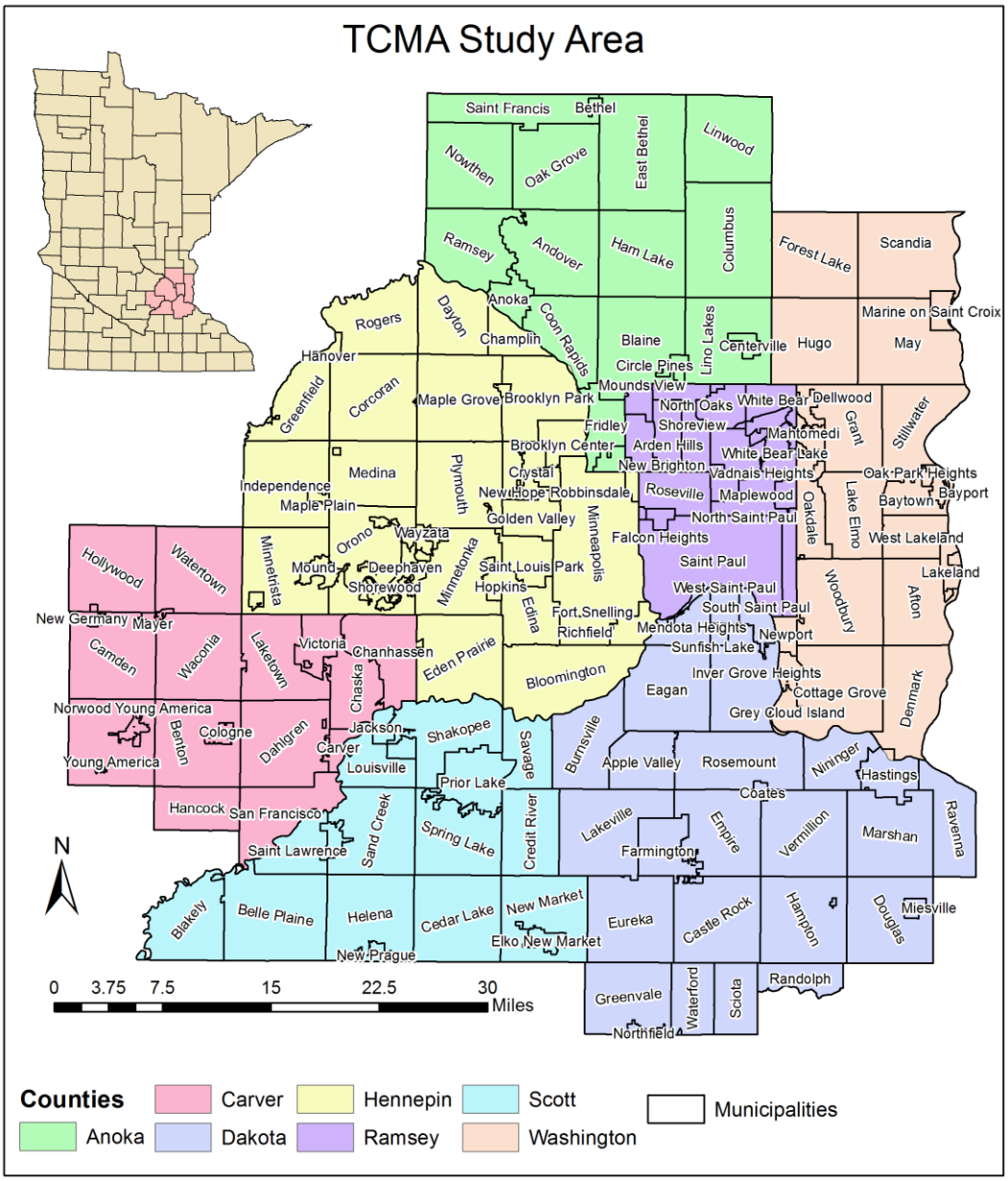

Figure 1.1: Study area overview map.

2. Literature Review

This study is concerned with the accurate extraction of impervious surface data and LULC classes from high-resolution aerial imagery and other data sources. The following will give an overview of the existing literature that is pertinent to the various techniques used here. I will first discuss the importance of imperviousness and some techniques that have been used to identify impervious surface areas. I will further give an overview of the research that has been done in relation to the various components of this study, such as data integration, decision tree modeling, and object-based classification.

2.1. Land use and land cover

Land use and land cover data are essential inputs and tools for local and regional decision makers. While urban growth indicates economic growth, it is also a major environmental concern. Urban growth not only leads to increased impervious surfaces and associated problems (see section 2.2), it also significantly degrades air quality at local to global scales, as well as increases energy consumption, and consumes agricultural and forestry resources (Squires 2002). Collinge (1996) conducted a thorough literature review concerning effects of urban sprawl on biodiversity. He concluded that urban sprawl, due to its segmented growth patterns, is a major contributor to habitat fragmentation and therefore reduction of biodiversity. Accurate LULC data are important in order to help decision makers at local, regional, and global scales improve policy

regarding future development, wetland and habitat conservation, and climate protection (Anderson et al. 1976).

2.2. Imperviousness

The concept and question of imperviousness has received a lot of interest in many fields recently, particularly in geographical and urban studies. Impervious surface restricts the flow of water through the surface. It is often considered to be comprised of rooftops of buildings and transportation features (Schueler 1994); however, it should also be noted that bare, compacted soil or exposed bedrock, at extraction sites for example, may have impervious qualities. In the case of rain events, snow melt, or flooding, water cannot penetrate the ground, but would rather be carried on the surface, picking up many surface pollutants along the way (Chen et al. 2007). Nonetheless, with the increasing dependence on the automobile in the US and other developed and developing countries, the amounts of impervious surfaces and their inherent problems are increasing.

Impervious surfaces accelerate the movement of runoff and pollutants collected over large area, which attributes to many of today's water pollution problems. As early as 1994, the US Environmental Protection Agency determined that non-point source pollution (of which impervious surface runoff pollution is an example) is the largest contributor and threat to the quality of water in the US (1994). Since non-point source pollutants are carried into surface and ground water, they impact both the natural and human ecosystems.

Impervious surface is not only a major contributor to non-point source pollution, but also a very good indicator of water quality. Impervious surfaces have been found to be a "key environmental indicator" to estimate many other factors. Arnold Jr and Gibbons (1996) found that knowledge of the amount of impervious surface can serve as a framework for solving problems of natural resource planning. This is particularly advantageous for local planning agencies that may not have the resources to commit more complex studies of particular problematic areas. Impervious surface is not only a general environmental indicator by itself, it is also strongly related to, and can be considered a proxy measure for other indicators that are much harder to measure (Schueler 1994). Impervious surface can be used as a measure of environmental impact not only locally, but also globally, as pointed out by Sutton et al. (2009), who constructed a global model based on satellite imagery and calibrated the model with high-resolution aerial imagery in an effort to show that impervious surface can be a proxy measure for the overall ecological footprint of societies. Although imperviousness has strong impacts on the environment across different scales, it is most powerful at the local scale. As Schueler (1994) noted that impervious surface data are relatively easy to obtain compared to other environmental indicators, and the amount of impervious surface can be managed by local policies.

2.3. Determining Impervious Surface Area

There are many approaches to estimating the amount of ISA in a study area. Field mapping can be used to achieve accurate results, but is often time-consuming and expensive. Remote sensing techniques offer a more efficient method. Traditional per-pixel approaches classify remote sensing data by assigning land cover classes to each pixel in an image, often based on an algorithm that makes statistical assumptions about the data. To extract impervious surfaces, the image is firstly classified into categories that will allow the researcher to aggregate them into impervious and pervious covers in the next step. For example, urban, transportation, bare soil (such as gravel pits or construction sites), and mining/extraction classes would be considered impervious, while open water, cropland, and wetland classes would be considered pervious. Dougherty et al. (2004) compared this approach to a sub-pixel method. They found that the traditional per-pixel method yielded slightly better results than the sub-pixel method, but the accuracy of both methods depended strongly on the types of classified land cover (Dougherty et al. 2004). Lua et al. (2011) described a method that uses the traditional classification method in combination with a segmentation-based method and manual editing to eliminate the drawbacks of each individual method.

Another technique that is relatively prevalent is the use of sub-pixel classifiers to estimate the percentage of impervious surface per area unit, or pixel. This method is based on the use of remote sensing images that have low to medium spatial resolution, which means that a pixel represents a fairly large area on the earth's surface, and likely comprises many different types of land cover. This method was used by Civco and Hurd

(1997) to map the amount of impervious surface areas of Connecticut. Their approach involved the use of an artificial neural network, which can be calibrated with high spatial resolution training data, but applied to medium spatial resolution imagery to deliver more accurate results for larger areas. Similar methods were also used by Stocker (1998). Van De Voorde et al. (2009) used two different sub-pixel classification models to extract impervious surface percentages in a comparative study. Similar to Civco and Hurd (1997), they used high spatial resolution images to calibrate their model, and then applied the model to lower spatial resolution images of large areas. They found that the multilayer perceptron model, which is relatively complex to use, performed relatively better than the spectral mixture analysis model. Taking the sub pixel classification approach further, Jennings et al. (2004) developed a model to estimate impervious surface areas, in which various data sources such as the National Land Cover Dataset (NLCD) and municipal transportation layers were used to generate sub-pixel impervious surface maps. These maps were then classified further into conceptual classes describing the amount of impervious surface areas.

A different approach was taken by Ridd (1995) with a "Vegetation – Impervious – Soil" (VIS) model to differentiate urban land cover classes. The model was initially developed for visual interpretation of aerial imagery, but was adapted by Ridd (1995) to be used with digital remote sensing data. The VIS model describes the composition of land based on the three classes it is named for, and can be used with the addition of a water class to determine the amount of impervious surface.

A modeling method that was not based strictly on remote sensing data was developed by Chabaeva et al. (2004). The authors created a model that can determine the ISA based on population parameters derived from US Census data. They built the model using NLCD shapefiles and created a regression model using inductive learning software. They calibrated the model for different localities and were able to determine the percentage of ISA fairly accurately, however only to the Census tract level (Chabaeva et al. 2004).

After reviewing the literature regarding the extraction of impervious surface, it becomes clear that this topic still has many open questions in terms of which method delivers the most accurate results. Every method described has its own advantages. The method used should be chosen based on the desired results and the available data.

2.4. Data integration

Most of the previously discussed methods of impervious surface extraction mostly only used one specific set of data as their input, such as satellite or aerial imagery, or census and parcel data. Some studies, however, used more than one type of data to extract land cover or impervious surface information for an area. More specialized methods are required, however, to classify using a combination of imagery and more abstract ancillary data types.

For example, Kontoes et al. (1993) described a method using SPOT imagery and ancillary map data that was manually digitized and edited. The authors than employed data derived from both the imagery and the ancillary data in a knowledge-based system

that allowed them to classify the data and extract agricultural crop coverage with relatively high accuracy. In another approach, McNairn et al. (2005) combined several types of imagery to extract the desired data. They employed and compared the maximum likelihood and decision tree methods. They reported that a decision tree approach allows for the integration of additional data that is not imagery.

Mesev (1998) described a method to extract urban land cover information by combining imagery with census data. However, unlike Kontoes et al. (1993), who employed a knowledge-based model, he was able to integrate the additional data in a maximum likelihood classifier (MLC).

An approach that integrates reflectance data with surface temperature data derived from Landsat data was taken by Lu and Weng (2006). In this case, the researchers used an imagery product and a derivative of the additional infrared emission layer that is delivered with Landsat TM data. Researchers have also combined optical remote sensing with active remote sensing products such as Radar imagery to improve results of classifications. Rignot et al. (1997) compared classifications of a site in Brazil rainforest obtained from the SIR-C radar data and the optical Landsat TM, SPOT, and JERS-1 sensors. They found that each sensor had specific strengths and weaknesses. They were able to combine these results to obtain an overall more accurate final map to identify biomass in their study area. Saatchi et al. (1997) also used radar data to map deforestation in the Brazilian rain forest. They used Landsat TM data to verify their results and also combined their results derived from both data to improve the overall accuracy of their classification. Optical and radar remote sensing data complement each other and

therefore can improve accuracy, and that radar data can be used where optical data shows weaknesses due to cloud cover or layered vegetation.

2.5. Decision Trees

Decision tree modeling is an artificial intelligence and machine learning technique, as demonstrated by Breiman (1984) and Wu and Kumar (2009). In this study, a combination of the object-based classification, integration of various data sources and types, and decision tree classifier was used. The decision tree software is a machine learning program that analyzes existing data and builds a decision tree model that fits the data best into predetermined classes. Decision trees are used not only for image classification, but also for many other applications in various fields. In general, they are useful for analyzing case data based on specific attributes and assigning discrete output values to each case (Mitchell 1997). There are many medical studies that use decision tree models: Granzow et al. (2001) used decision trees to find relationships between types of tumors and genetic properties. In a different application in cancer research, Kuo et al. (2002) built a decision tree model that could be calibrated to find patterns of breast tumors in different types of ultrasound data. Silva et al. (2003) used decision tree models to classify large amounts of data found in Intensive Care Unit databases to help doctors predict the likelihood of organ failure for patients.

Decision trees have also been used in economics studies to help in making decisions for the creation of stock portfolios (Tseng (2003). Sen and Hernandez (2000) created a decision tree model that helped apartment buyers analyze the various data about

the apartments and real estate markets that is publicly available, and make better buying decisions based on these data. Arditi and Pulket (2005) were able to use decision tree models to predict the outcome of construction litigation. Another interesting example of decision tree models for real-world applications was presented by Copeck et al. (2002) with their machine learning process to summarize documents.

2.5.1. Decision Trees in Geography

In Geographical studies, decision trees are most often used for image classification, but have also found some other applications: Lang and Blaschke (2006) used a decision tree model to identify the best suited locations for wildlife bridges to protect brown bear habitat in Slovenia. Hansen et al. (1996) described decision trees as an alternative to traditional land cover classifiers and found that they have similar accuracy to Maximum Likelihood Classification, while offering more flexibility for the requirements of input data. Gahegan (2000) examined the particular advantages and disadvantages of using machine learning algorithms to analyze geographical data, as compared to the more traditional statistical tools used in many studies. He also suggested that machine learning tools are often better suited to cope with the very large datasets now used in Geography (Gahegan 2000, 2003). A general comparison of traditional classifiers, artificial neural networks (ANN), and decision tree classifiers was presented by Pal and Mather (2003). The researchers found that decision tree classifiers have advantages over the traditional classifiers since they can handle various types of data on different scales and units, and do not depend on statistical assumptions about the data. In comparison to artificial neural networks, they found that decision trees are advantageous

because they are easier to use, require less training and parameters to be setup, process large sets of data quickly, and are widely available on the internet. They also found that the decision tree delivered acceptable results compared to other classifiers in most cases (Pal and Mather 2003). In contrast, Rogan et al. (2008) and Rogan et al. (2003) found that ANN can achieve better accuracies for land cover change mapping.

Some good examples of decision tree applications for very large datasets are presented in the publications regarding several US nationwide land cover datasets. Decision tree classifiers were used in building a database of 22 land cover classes with remote sensing data from 2000 and 16 classes with data from 2001 for the entire United States (Homer et al. 2007; Homer et al. 2002). Furthering the use of these datasets, Fry et al. (2009) used decision tree models to map the differences between the 1992 and 2001 National Land Cover Database products efficiently. Another nationwide product that was developed with decision tree models is the 2009 Cropland Data Layer (Johnson and Mueller 2010).

An additional advantage of decision tree classification is that it is able to handle many attributes, or sets of data, and identify the most important ones. This is exemplified by Bricklemyer et al. (2007) to verify the association of agricultural practices with soil carbon sequestration. Similarly, Ban et al. (2010) used decision trees to combine Quickbird and Radarsat data to aid in urban land cover classification. Zhang and Wang (2003) also used decision tree models to classify urban land cover types from high-resolution multispectral imagery. Another study used two types of imagery (medium spatial resolution Landsat and high spatial resolution aerial imagery) to estimate the

density of tree cover for large areas (Huang et al. 2001). Instead of utilizing two sources of imagery, Harris and Ventura (1995) used Landsat imagery and more abstract geographical data, such as parcel and census data, to classify urban land cover types. In contrast, Griffin et al. (2011) used decision trees to include environmental factors in a classification of various vegetation types in a specific ecosystem. For a study to assess animal habitats and agricultural land cover, Lucas et al. (2007) employed a rule-based decision tree to map the habitats and cover classes based on multi-temporal satellite imagery, various derivatives of the imagery, and data retrieved from an agricultural management system. A similar approach was taken by Wright and Gallant (2007) to increase the accuracy of wetland mapping in Yellowstone National Park.

In addition to all of the previously mentioned advantages of decision tree modeling, another benefit of this technique is its ability to deal with errors very well (Mitchell 1997). Two major error sources in remote sensing are uncertainties already present in the imagery due to acquisition and processing issues, and errors introduced by the analyst while generating training data (Foody et al. 2002). Decision tree models are particularly tolerant towards both of these error types, and can even handle cases where some of the attributes are missing very well (Mitchell 1997).

It is evident that decision tree classification systems can deliver accurate results for many different applications in geographic research, particularly when dealing with datasets that are either very large, contain different data scales or units, or are problematic for traditional or statistical models. Decision trees are found to be flexible, user-friendly, and efficient.

2.5.2. Decision Trees for Impervious Surface Extraction

Studies that used decision trees to identify ISA only have emerged in the past ten to fifteen years. Smith (2000) employed a decision tree model to estimate the sub-pixel level ISA from Landsat imagery in the diversely urbanized area of Santa Barbara in Southern California. Similarly, Yang, Huang, et al. (2003) and Yang, Xian, et al. (2003) used decision trees to extract sub-pixel ISA from Landsat TM and ETM+ and high-resolution aerial imagery, and to detect urban land cover changes, respectively. High spatial resolution aerial imagery was used by Cutter et al. (2002) to extract ISA. Goetz et al. (2003) used decision trees to extract not only impervious surfaces, but also tree covers from IKONOS imagery.

While decision tree classifiers have been used occasionally to extract impervious surface from medium spatial resolution imagery by means of sub-pixel classification, they seem to be most efficient for use with high spatial resolution imagery. This is noted by Cutter et al. (2002), who found that traditional classifiers are often unable to handle the challenges posed by high spatial resolution imagery. The fact that this high spatial resolution imagery is becoming more widely available may also explain the fact that there has very little work been done for impervious cover extraction with decision trees until recently.

2.6. Object-based Classification

While a decision tree approach was used in this study to classify ISA, the remaining LULC classification was completed using object-based classification. Object-based classification is a relatively new concept compared to pixel-based classification. Its development began when per-pixel classification was found to be lacking in some aspects, and when computing power increased which allows for the development of more advanced techniques. Tobler (1970) defined the first law of Geography as: "Everything is related to everything else, but near things are more related than distant things." Therefore, many researchers have criticized the focus on the pixel as a unit in image classifications. They have found that it makes more geographical sense to include not only the information that is present in one pixel, but also what surrounds that pixel. Considering this, one should not only focus on individual pixels in a study, but also should consider the data in their surroundings (Fisher (1997) and Cracknell (1998). Haralick et al. (1973), Haralick and Shapiro (1985), and Myint (2001) all suggested to integrate contextual information by calculating textures based on surrounding pixel values in order to implement this principle in remote sensing applications. In this study, the texture-band approach was followed for the impervious surface classification.

Object-based classification was employed as an additional method of incorporating contextual information. Instead of looking at each pixel individually, this method attempts to find patterns in the pixel values and group pixels according to these patterns. This process is also referred to as image segmentation (Blaschke and Strobl 2001). This approach has been found to be advantageous particularly when classifying

imagery that has very limited spectral resolution, such as grayscale imagery (Blaschke and Strobl 2001), and for imagery that is problematic to classify because of its high spatial resolution (Miller et al. 2009). While object-based classification is mostly suitable for extraction of certain objects (such as trees, buildings, water bodies), it can be adapted to be used on the extraction of land cover classes based on multi-scale objects (Baatz and Schäpe 2000). In this study, object-based classification was used for the general LULC classification part.

2.7. Feature Analyst

Feature Analyst is the software chosen here to implement the object-based LULC classification. Feature Analyst is a third-party extension for ESRI ArcGIS, and is considered an object-based, inductive learning classification system.

In fact, Feature Analyst is a combination of various classification algorithms. It is object-based because it makes use of image segmentation, and is capable of identifying individual objects in an image, compared to many other systems which can only perform so called “wall-to-wall” classifications where every object within an image has to be included in the classification.

Aside from image segmentation, Feature Analyst makes use of several other classification models. These include: (1) decision trees, (2) variants of ANN, which are designed to assess information in a similar way to the human brain (Opitz and Blundell 2008; Rumelhart et al. 1986), (3) Bayesian learning, which is similar to ANNs, but additionally makes use of probability assumptions about the data, and (4) K-nearest

neighbor, which attempts to assign a class to a case simply based on how "close" its attributes are to those of known cases (Mitchell 1997). Feature Analyst automatically includes one or more of these approaches in its classification models, depending on which approach is best suited for the data to be classified (Opitz and Blundell 2008).

In addition to selecting one or several classification approaches, Feature Analyst will also make use of ensembles, a concept very similar to boosting in decision trees (see 3.3.2). Ensembles are sets of classification models, which are trained using the same data, and whose results are combined to produce a final result. While there are several options to combine the results, the most common, and the one used in Feature Analyst, is a weighted average of all results (Opitz 1999). Several studies have found that ensemble predictions become more accurate if the individual predictors disagree as much as possible (Breiman 1996; Freund and Schapire 1996; Opitz and Shavlik 1996a, 1996b). Therefore, Feature Analyst actively attempts to build several models that produce diverse results, thus increasing the possible accuracy of the entire ensemble (Opitz and Blundell 2008). This approach is claimed to be more accurate than similar techniques, such as decision tree boosting.

In summary, Feature Analyst makes use of several innovative and advanced image classification techniques that promise improved accuracy for high spatial resolution imagery compared to other, pixel-based classification techniques.

3. Methods

3.1. Data Acquisition and Preprocessing

The data used in this study were obtained from the United States Geological Survey (USGS) Seamless Data Warehouse website (http://seamless.usgs.gov/), and the Minnesota Geospatial Information Office (MNGEO) Data Clearinghouse (http://www.mngeo.state.mn.us/). The specific datasets were 2010 high-resolution digital ortho-imagery, Light Detection and Ranging (LIDAR) elevation data, and road centerline data.

3.1.1. Aerial Imagery

The aerial imagery was part of the National Agriculture Inventory program (NAIP) funded by the U.S. Farm Services Agency (FSA), which makes these data available for public use at no charge. The 2010 NAIP ortho-images were flown during the agricultural growing season, specifically during the months of July and August. The images have a spatial resolution of 1 meter per pixel, and radiometric resolution of 8 bit pixel depth. The NAIP imagery is ortho-rectified by the data vendor to a horizontal accuracy of +/- 5 meter using ground control points. The vendor further mosaicked the individual images to produce county mosaics using a last-in-last-on-top strategy, and color-balanced the mosaics using the Impho Orthovista software.

The aerial images were acquired as full county mosaics for each of the seven metropolitan counties. Instead of mosaicking these individual files into a new raster, the ArcGIS mosaic dataset functionality was used. A mosaic dataset is a dynamic mosaicking

and processing tool that allows for images to be mosaicked on the fly, rather than statically. The dataset itself only contains references to the individual files, and is therefore very disk-space efficient. In addition to applying various mosaicking functions, it is also possible to add other raster functions on the fly, for example NDVI, pan sharpening, or hillshade processing for elevation rasters. The major advantages of the mosaic dataset include: it is very fast to apply functions; it reduces the required storage space by avoiding duplication of data; and it is compatible with any ArcGIS raster tools. In this study, the mosaic dataset was used first to mosaic the NAIP images to achieve coverage of the entire study area. A clip function was applied to exclude areas outside of the seven county metropolitan area of interest.

In order to provide additional classification parameters for both the LULC and the impervious surface classification, the near-infrared band of the NAIP imagery was also used to calculate texture values. Texture, when thought of as variance in specific localized parts of the data, has previously been seen as undesirable as it could make classification with per-pixel methods more difficult (Ryherd and Woodcock 1996; Gong and Howarth 1990; Herold et al. 2003; Zhang 1999). This study uses an object-based and a pixel-based decision tree classifier. Texture information is known to be valuable for use in object-based systems (Ryherd and Woodcock 1996), and due to the winnowing function of C5, it can also be used in this pixel-based approach. When texture is specifically used as an input for image classification, especially in segmentation-based processes, it can be beneficial. The texture layer was calculated as variance of a 3x3 pixel window (Yuan 2008). The equation used to calculate variance is as follows:

$$Variance = \frac{\sum \left(x_{ij} - \frac{\sum x_{ij}}{n} \right)^2}{n-1}$$

Equation 3.1: Variance calculation where x_{ij} is the DN value of the pixel at i,j, and n is the number of pixels in the window. Adopted from (Yuan 2008).

Unfortunately, because the functions that can be used with mosaic datasets are predefined, this step could not be performed directly with one function. Chaining of functions in a mosaic dataset allows for several functions to be applied to the data in a specified order. Therefore, variance was implemented by first applying a standard deviation function, and then squaring the result of that function in a separate function. These functions were applied with a 3 x 3 pixel window.

3.1.2. Elevation Data

In addition to aerial imagery, elevation data derived from LiDAR data were utilized. The elevation data used in this study were acquired in 2011 and 2012 by private vendors in cooperation with the Minnesota Department of Natural Resources (MNDNR). These data are part of Minnesota's statewide Elevation Mapping Project. The Twin Cities Metro Region dataset used here has three different point densities, depending on the area covered: Anoka, Carver, Hennepin, Ramsey, Scott, and Washington counties were sampled at 1.5 points per square meter, Dakota County at 2 points per square meter, and the cities of Minneapolis, St. Paul, and Maple Grove at 8 points per square meter. The MNDNR determined the vertical accuracy to be 5 cm, 10.8 cm, and 8.3 cm for the entire

metro block area, Dakota county, and Maple Grove, respectively. The data vendor produced a Digital Elevation Model (DEM) from these LiDAR point data based on the terrain-only points. Since the DEM files are provided at the same spatial resolution as the aerial imagery (1 m pixel size), no further processing was necessary. DEM rasters were mosaicked using a mosaic dataset and masked to the same extent as the aerial imagery using a mosaic dataset function.

3.1.3. Road centerline data and Road Density

As an additional dataset, the roads basemap was downloaded from the MNDNR Data Deli website, a statewide geospatial data portal. The purpose of the roads layer in this study was the production of a road density layer. The roads layer is a digitized map of roads, produced by the Minnesota Department of Transportation (MNDOT). Roads were digitized based on 7.5 minute USGS quadrangle topographic maps, and was updated in 2001. The horizontal accuracy of these roads is less than +/- 12 m. Roads are represented as centerlines, and detailed road class information is given.

3.1.4. Additional supporting data

In addition to the aforementioned data that were used directly in creating the classification maps, ancillary datasets were acquired to support the analysis and interpretation of the results. First, political boundary polygonal shapefiles were acquired for the county and municipality levels. County boundaries were retrieved from the MNGeo Geographic Data Clearinghouse online (http://www.mngeo.state.mn.us/). These boundaries were current as of June 2013. They are represented at a nominal scale of

1:24,000, and have a spatial accuracy of +/- 12 meters (40 feet). Similarly, boundaries of all municipalities were also retrieved from MNGeo. These boundaries have the same temporal and spatial characteristics as the county boundaries, and include cities, townships, and unorganized territories (CTUs). It should be noted that the township boundaries refer to political entities, which do not necessarily coincide with public land survey entities.

In addition, the US Census Factfinder website (http://factfinder2.census.gov) was used to acquire population data for the area and the political entities represented in the previously mentioned boundary files.

3.2. Land use and land cover classification

The LULC classification was performed using Feature Analyst, an extension for ArcGIS that employs a proprietary, object-based, inductive learning classification algorithm. Before conducting the classification, the class scheme was determined. The extracted LULC classes include water (rivers, lakes, pools, and other open bodies of water), urban impervious infrastructure (roads, buildings), cropland (non-vegetated and vegetated fields and pasture), forested areas (deciduous, evergreen, and mixed forests), other vegetated areas (shrub, herbaceous plants, non-forested wetlands), and bare soil and rock (mining operations such as gravel pits, bedrock). An important note to be made is that the cropland class was classified in two steps: first, it was separated into the vegetated and non-vegetated parts of cropland, and then these two classes were merged to form one class. These steps were taken due to the bi-modal distribution of spectral values

within this class. While object-based classification methods should be better at coping with this issue than traditional pixel-based approaches, initial testing showed that this step increases classification accuracy for cropland class by avoiding its confusion with other LULC classes.

Object-based classification in Feature Analyst makes use of the inductive learning approach. This means that the algorithm relies on an expert, or teacher, to identify examples of the desired outcome in the imagery, therefore providing a set of training data to the classification system. The algorithm then uses the information contained in these samples to build the model which is used to perform the classification. Training samples had to be determined manually based on a set of criteria for each class. The classifier not only relies on spectral information contained in each pixel, but also considers information derived from pixel groups. This is based on a process known as image segmentation, which divides the individual pixels into groups, representing objects. The segmentation is generally based on homogenous pixel values, edges, and shapes. After defining these objects, additional information from training data is used to determine which class an object belongs to. Spectral values, along with additional data such as the size and shape of the object, and its patterns, texture, and neighboring objects are used to assign class values. Accordingly, training samples had to represent all of these characteristics for each class. Training samples were created as polygons, and were drawn as close to the edge of each feature as possible to allow the sample to represent the shape and edge type of the object. For example, most buildings are relatively simple, rectangular features, while rivers or lakes have more complex, curved outlines which were indicated by the shape of

their training polygons. Care was also taken to include the full variety of each feature found in the study area. For example, small outbuildings, residential homes, and large commercial or industrial facilities were all included in the urban training set to appropriately represent the range of sizes and shapes of this object type within the study area.

After creating the training polygon set, Feature Analyst presents several other settings that are used during the classification process to adjust the algorithm for the specific situation and improve its accuracy. First, input data are defined. Feature Analyst allows the use of multiple sets and types of data. Therefore, the LULC classification was based on the four band NAIP imagery from 2010, along with the LiDAR-derived elevation data and a texture layer. This is possible in Feature Analyst without previously stacking or otherwise altering the input layers; they are simply added to a list of inputs, and their type (i. e. optical, elevation, texture) is defined. Since Feature Analyst is capable of extracting individual features or perform complete, exhaustive and inclusive classifications of land cover features, there are several other options available. First, the "wall-to-wall" option was selected to force Feature Analyst to classify all areas of the imagery, rather than just extracting some features. Feature Analyst is also able to automatically stretch the image data, however, this options was disabled because the data were already stretched on-the-fly by the mosaic dataset functions. In order to make use of contextual information (i. e. analyze objects based on their neighboring objects), Feature Analyst uses a system called input representation. These representations are essentially local windows which let the classifier see more than one pixel at a time during

processing. This allows the classifier to build its model based on not only one pixel, but also its neighboring pixels. However, this approach increases the amount of processing time required to build and implement the model, and it still only takes into account a relatively small areas surrounding each pixel. Feature Analyst allows for more complex window patterns, which are known as foveal representations, as they are designed to mimic the way the human eye sees things (Opitz and Bain 1999). Figure 3.1 shows an example of these foveal representations used in Feature Analyst, the Manhattan pattern. Colored pixels are those that would be visible to the classifier while analyzing the center pixel, while uncolored pixels would be ignored. Compared to simple local windows, this approach allows the classifier to give more importance to pixels nearer to the pixel being processed, yet take into consideration information found in pixels further away as well. This approach should increase the amount of information available to the classifier, while at the same time not increasing the amount of data to be processed as much as the simple window approach. This approach makes the integration of contextual information in a classification model more efficient, and again is a testament to Tobler's First Law of Geography, which states that near things are more related than distance things (Tobler 1970).

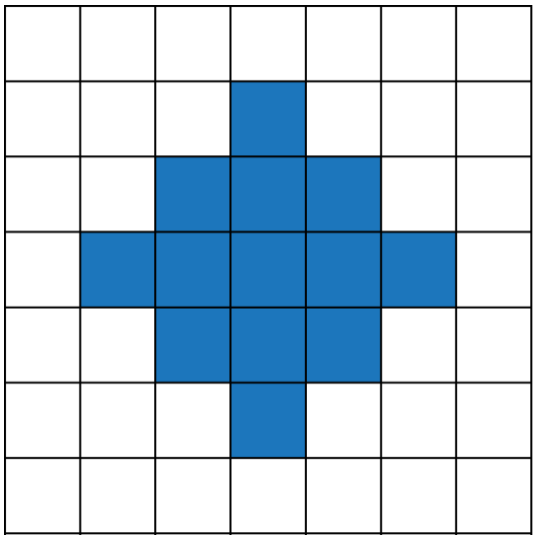

Figure 3.1: Example of Feature Analyst foveal representation (Manhattan pattern).

3.3. Impervious surface classification

3.3.1. Road Density

Road density is often considered an important measure of urbanization (Schueler 1994). Therefore, road density was computed as a raster surface to be used as one of the input parameters for the impervious surface decision tree modeler. Road density is generally defined as kilometers of road per 100 square kilometers of area, or miles of road per square miles of area. To make the data processing more manageable and to ensure that road density was considered locally rather than globally for the entire study area, it was calculated for an area of 1 square kilometer around each pixel. The road density calculation was conducted based on the road centerline layer. The density was calculated as meters of roads per square kilometer of area, and stored in a floating point raster with 1 m pixel size. The resulting data were used as input parameters for the decision tree model.

3.3.2. Decision Tree Modeling

Impervious surface was classified using a decision tree classifier. Specifically, the See5 software was used to generate the decision tree based on the C5 algorithm (Quinlan 2013a). C5 is very similar to the C4.5 algorithm, with the addition of several features that have the potential to increase the classification accuracy (Quinlan 1993, 2013b).

In general, C4.5 and C5 are expert-knowledge systems that require human inputs in the form of training data. The general purpose is to use the training data to identify to which class each case should belong, and then find an accurate model representation to assign a class to each case in the general population based. To achieve this, C4.5 and C5 algorithms begin by dividing the cases based on their attributes and then identify a natural break-point in the attribute based on the class value. This approach produces a set of decision rules that can be combined to build a tree consisting of branches and leaves. Each branch represents a test that is performed on the data, and leads to either a further branch, or a leaf, where a decision is reached and a class assigned. These steps are performed automatically by the C5 algorithm. In addition, C5 can perform winnowing (decide to exclude attributes if they do not significantly contribute to the model) and pruning (removing branches that do not significantly contribute to the model) (Quinlan 1993, 2013a, 2013b).

In order to create training data for the decision tree model, a set of 300 randomly distributed points was created. A second set of points following the same principle was created to be used for accuracy assessment (Congalton 1988). Both sets were created at this time because C5 is able to conduct the accuracy assessment automatically after

generating the model based on a separate set of points. Therefore, the accuracy of the decision tree model could be evaluated immediately, and the algorithm configuration could be adjusted if necessary. For each set of points, the ArcGIS "Value to points" tool was used to write raster values to the attribute table for each point. The values of each band of the aerial imagery, the calculated texture layer, elevation raster, LULC raster, and road density map were included in this process. A field containing binary impervious surface data was added and populated by visually determining whether each point is located on impervious surface or not. The resulting tables were then adjusted in Microsoft Excel and exported for use in See5.

In See5, the data were analyzed several times using different options to identify the settings that can deliver the best accuracy and efficiency. Specifically, the winnowing, pruning, and boosting options were evaluated. Winnowing prompts See5 to evaluate the impact of each attribute on the final model and decide whether it should be used or not. This aids in faster processing and makes the resulting decision tree smaller and less complex, often leading to better accuracy (Quinlan 1993; Foody et al. 2002). Pruning is also a method of making large decision trees smaller and less complex. When pruning is used, the tree is first produced normally, and then pruned. Pruning works by dividing the tree into several subtrees and estimating the likelihood of misclassification for each subtree. This estimation is then compared to the case where the subtree would simply be replaced by a leaf. If this change does not change the likelihood of misclassification by more than a certain threshold, the change is committed to the tree; otherwise, the subtree is left the way it is (Quinlan 1993). Pruning is useful to improve

trees that suffer from over fitting, a condition where the decision tree fits the training data almost perfectly, but is biased to these data, and therefore fails to accurately model the test or accuracy assessment data (Foody et al. 2002).

Boosting is a method that is used only to increase the accuracy of decision trees, at the cost of making the model more complex and computationally expensive. Boosting works by creating more than one model to solve the same problem and using the results from each model to "vote" for a final result (Freund et al. 1999; Quinlan 2013b; Schapire 1999). For example, if ten models are created to classify pixels, and six of them determine a pixel to be in class a, while four assign class b, the final result would be the majority vote, class a.

The See5 output is a text file representing the decision tree in a pseudo-graphical way. This model was manually "translated" to be used in a Python script to carry out the actual data processing. The script makes use of the ArcGIS ArcPy module, which allows the use of ArcGIS tools within a Python script.

The decision tree was implemented in the script by using a series of nested "Con" conditional statements from ArcGIS Spatial Analyst. The Con function evaluates a condition on a per-pixel basis, and can either output a constant, another raster value, or initiate another con statement nested within it. Some of the advantages of implementing this function in a Python script is that it is easily possible to save and adjust the script at any time, and that operations can be performed in memory rather than from the hard drive, therefore improving performance. The use of ArcPy also makes it possible to utilize ArcGIS mosaic datasets as inputs, rather than statically mosaicked raster files.

While it would be possible to process the decision tree model in Python without using ArcPy components, this would only work with fully mosaicked raster files. Therefore, by using mosaic datasets instead, an additional processing and data intensive step is cut from the workflow. After processing the decision tree based on the defined input layers, the result was written to disk as a 1-bit raster file. Impervious surface was classified using this newly developed decision tree model. Specifically, to build the model, the algorithm had access to the following attributes for each case (sample pixel): the four-band NAIP aerial imagery, the LIDAR DEM, the texture layer generated from the near infrared band of the aerial imagery, the road density layer, and the LULC raster map. While building the decision tree model, several advanced options were evaluated in terms of their ability to increase the classification accuracy of the model. These options are specifically winnowing, pruning, and boosting (see 3.3.2). It was found that the use of winnowing did not make a difference in the model. See5 used the same attributes whether winnowing was used or not. Therefore, winnowing was not used for the final model. Further, the decision tree model created without pruning was already relatively small and had very good accuracy. Pruning the tree did not reduce its size enough to justify the loss in accuracy caused by the use of pruning. For the boosting options, a manageable amount of ten trials was evaluated. It was found that using boosting with ten models did not increase accuracy enough to justify the additional complexity and computational expense of the resulting model. In fact, while accuracy increased for the training dataset, boosting decreased the accuracy attained for the test dataset. Based on these evaluation results, the basic decision tree without advanced options was chosen as the final model.

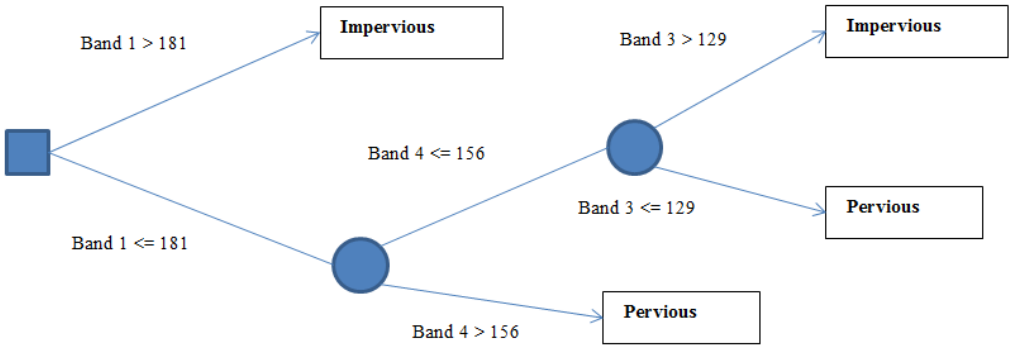

Figure 3.2: Decision tree model.

The final decision tree model only includes bands one, three, and four from the aerial image. The decision tree consists of two branches and four leaves (see Figure 3.2).

3.4. Accuracy Assessment

An accuracy assessment was conducted on both the LULC and impervious surface maps independently. For each map, a set of 300 random sample points was created (Congalton 1988, 1991b). Reference values assumed to be "ground true" were assigned to these points based on visual inspection of the imagery. To determine the classified values for the LULC map, the ArcGIS "Value to points" tool was used to automatically write raster values from the LULC map to the sample point table. For the impervious surface, this task was achieved by See5, which allows for the input of a separate set of accuracy assessment data and automatically evaluates the model against these data.

Standard accuracy matrices were generated for both maps, and per-class user's and producer's accuracy, overall accuracy, and estimated Kappa accuracy were calculated based on the matrices (Cohen 1960; Bishop et al. 1975; Congalton 1991a; Congalton and Mead 1983). User's accuracy is also referred to as error of commission, which describes classification errors where a pixel that belongs to one class was falsely assigned to a different class. In contrast, producer's accuracy or error of omission is an error where a pixel that should have been assigned a certain class value, but was not included in that class (Campbell 2002). The Kappa statistics were estimated with an equation given by Cohen (1960) (see Equation 3.2). Cohen's Kappa is also referred to as inter-observer agreement, and originated in Psychological studies (Cohen 1960). The Kappa coefficient was first proposed by Congalton and Mead (1983). estimates "the difference between the observed agreement between two maps […] and the agreement that might be attained solely by [chance]" (Campbell 2002).

$$\hat{K} = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} x_{i+}\, x_{i+}}{N^2 - \sum_{i=1}^{r} x_{i+}\, x_{i+}}$$

Equation 3.2: Cohen's Kappa Estimation (Cohen 1960; Congalton and Mead 1983).

4. Results

4.1. Land use and land cover classification

4.1.1. LULC: Entire study area

Land use and land cover distribution was first determined and analyzed for the entire study area without further subdividing into political entities. Table 4.1 shows the area of each LULC class by county as well as total values. The largest LULC class in the study area was other vegetation, which had a total area of 1136 sq. miles, or 38.2% of the entire study area. This was followed by forest (772 sq. mi, 26%), cropland (527 sq. mi, 17.7%), urban (335 sq. mi, 11.3%), and water (170 sq. mi, 5.7%). The smallest class was bare soil, which only made up 34.6 sq. miles or 1.1% of the entire study area.

County	Water		Forest		Cropland		Other Veg		Urban		Bare Soil		TOTAL Sq. Mi.
	Sq. Mi.	% of total county	Sq. Mi.	% of total county	Sq. Mi.	% of total county	Sq. Mi.	% of total county	Sq. Mi.	% of total county	Sq. Mi.	% of total county	
Anoka	18.80	4.22%	157.95	35.46%	33.27	7.47%	186.55	41.88%	44.42	9.97%	4.43	0.99%	445.41
Carver	18.88	5.02%	60.28	16.04%	85.72	22.82%	190.16	50.62%	17.92	4.77%	2.72	0.72%	375.68
Dakota	21.56	3.68%	135.80	23.18%	166.80	28.47%	199.68	34.08%	55.44	9.46%	6.62	1.13%	585.90
Hennepin	47.86	7.89%	172.93	28.52%	44.42	7.33%	220.73	36.40%	111.89	18.45%	8.54	1.41%	606.37
Ramsey	14.28	8.41%	58.64	34.52%	2.58	1.52%	40.13	23.62%	53.33	31.39%	0.94	0.55%	169.90
Scott	14.31	3.89%	73.77	20.05%	120.01	32.62%	132.22	35.93%	21.47	5.84%	6.18	1.68%	367.96
Washington	33.89	8.01%	112.33	26.56%	74.52	17.62%	166.64	39.41%	30.35	7.18%	5.15	1.22%	422.89
TOTAL	169.58		771.70		527.32		1136.11		334.83		34.58		2974.11
MEAN	24.23	5.88%	110.24	26.33%	75.33	16.83%	162.30	37.42%	47.83	12.44%	4.94	1.10%	
STANDARD DEVIATION	11.43	1.97%	43.67	6.66%	51.39	10.91%	56.09	7.59%	29.52	8.78%	2.35	0.36%	

Table 4.1: LULC area by county.

The final 1 m LULC classification map is shown in Figure 4.1. Because of the high spatial resolution of the classified map and the relatively large size of the study area, some of the more detailed analysis is lost when the map is viewed in its resampled form showing the entire study area. In order to better analyze the patterns of each LULC class, the data were aggregated into a regular, hexagonal grid (see Figure 4.2). Each hexagon has a side length of 1 km, giving it an area of approximately 1 sq. mile. The major roads shown are Interstate and Minnesota State highways, and are included for reference purposes. In order to identify spatial patterns of LULC classes, individual maps were created for each class. In addition, hot and cold spots and clustering of the features were evaluated using Getis-Ord G* and Moran's I statistics, respectively. Getis-Ord G* generates a z-score that indicates hot spots (clustering of high values) and cold spots (clustering of low values). The equation is indicated in Appendix B. Getis-Ord G* outputs a map that further indicates areas of high or low concentration of each LULC type and aids in identifying patterns of spatial distribution (see Figure 4.3). As it is assumed that most of the LULC features are clustered in certain areas, the Moran's I statistics was used to verify this assumption. The equation is given in Appendix A. Global Moran's I statistics are given for each class in Figure 4.3. The Moran's I statistics for all LULC classes have p-values of 0 and z-scores much larger than the critical value of 2.58 (for a confidence interval of 0.99). Therefore, it can be assumed that the LULC patterns are not randomly distributed. Further, all classes exhibit a positive Moran's I index value, therefore indicating a tendency towards clustered distribution.

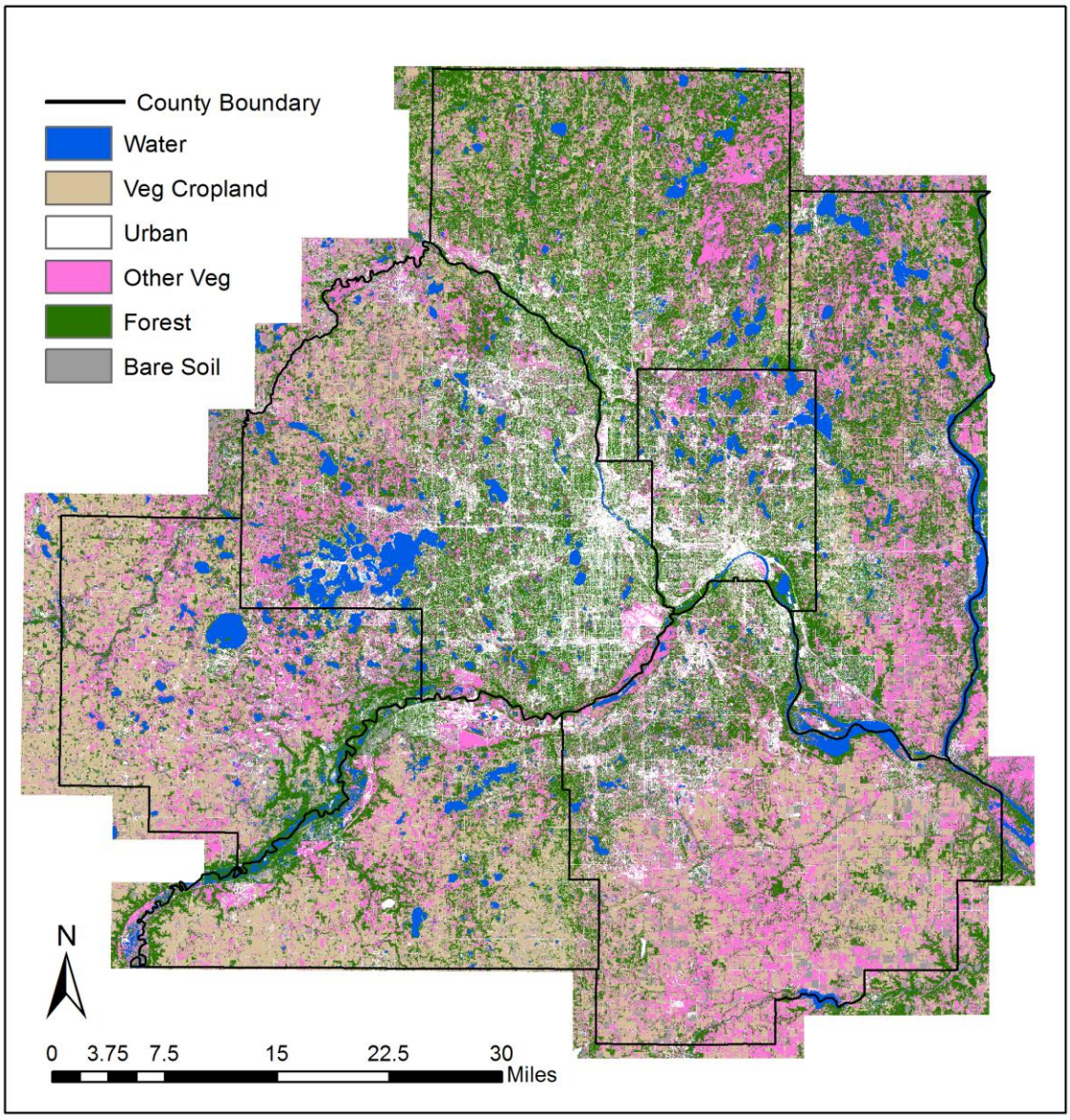

Figure 4.1: LULC map.

The urban LULC class exhibits a pattern of high density towards the center of the study area, and gradually decreasing density towards the outer edges. There are, however, several smaller, outlying clusters of urban cover found in the outer perimeter. These clusters likely show locations of smaller towns surrounding the major metropolitan area

in the center of the study area. To some extent, urban cover also seems to follow the path of the major roads indicated in the map. Analysis of the Getis-Ord G* maps (Figure 4.3) confirms this distribution pattern of urban land cover. The outlying clusters of urban cover are not as clearly identified here, but the large hot spot in the center of the study area is very evident. The p-value and z-score (0 and 76.4, respectively) of the Moran's I statistics indicate that the spatial distribution is not random, and the index value itself (0.8) points toward clustering of the features.

Bare soil exhibits a less obvious pattern. There is relatively little bare soil found in the western and north-eastern part of the study area. Some small clusters are indicated just to the north-west, south, and south-west of the center. This distribution is more clearly visible in the hot spot map for bare soil. While the Moran's I values for the bare soil class also show a non-random, clustered distribution, the index value itself is lower than that of all other classes (0.37). This indicates a less clustered distribution of this class, as evidenced in the maps.

Other vegetation shows a pattern of relatively high density around the perimeter of the study area, and lower density towards the center of the study area. There is also an area of low density along the south-central edge of the study area. The distribution of this class is interesting when compared with the distribution of urban land cover. Relatively little vegetation was found in the area of the highest urban cover density. However, other urban areas, where there are likely more residential buildings found, still have some amount of other vegetation (i.e. vegetation in backyards, neighborhood parks, etc.). The hot spot map shows hot spots of other vegetation particularly in the western, north-

eastern, eastern, and south-eastern parts of the study area. Cold spots are found in the center and south-central part, as was indicated earlier. Moran's I values also confirm the clustered distribution of this class (Moran's Index: 0.63).

The spatial distribution of cropland is essentially opposite to that of other vegetation. While both have low density in the central, urbanized part of the study area, cropland is concentrated along the southern edge of the study area, where there was relatively little other vegetation found. There are additional clusters of cropland in the western and south-eastern corners of the study area. Small densities of cropland are found all along the outer perimeter of the study area, with the exception of the north-eastern part, where there is a high concentration of other vegetation. Moran's I statistics indicate a highly clustered distribution of cropland (Moran's Index: 0.79).

The distribution of forest cover is interesting because a lot of forest is found within urbanized areas. While there is a relatively low concentration of forest in the most urbanized area in the center of the study area, the suburban areas surrounding this central location are identified as forest hot spots. Additionally, the northern part of the study area also has high concentrations of forest cover. Cold spots are mostly found along the western and southern perimeter of the study area, where there are high densities of cropland. Forest is also identified as clustered by the Moran's I test.

The distribution of the water class also follows a clustered pattern. The location of rivers and large lakes can be identified in the hexagonal map. A notable feature is Lake Minnetonka, in the western part of the study area. This very large lake has some of the highest concentrations of water in the study area. Compared to other LULC classes, water

appears to have lower concentrations in most locations throughout the study area. Moran's I indicates that water is clustered (Moran's Index: 0.5).

In general, it was found that urban land cover is mostly concentrated in the central part of the study area. Water was found in some very dense pockets throughout the study area. The periphery of the study area was made up mostly by vegetated land covers (forest, cropland, other vegetation), which are clustered in specific parts of the periphery and are somewhat exclusive of each other.

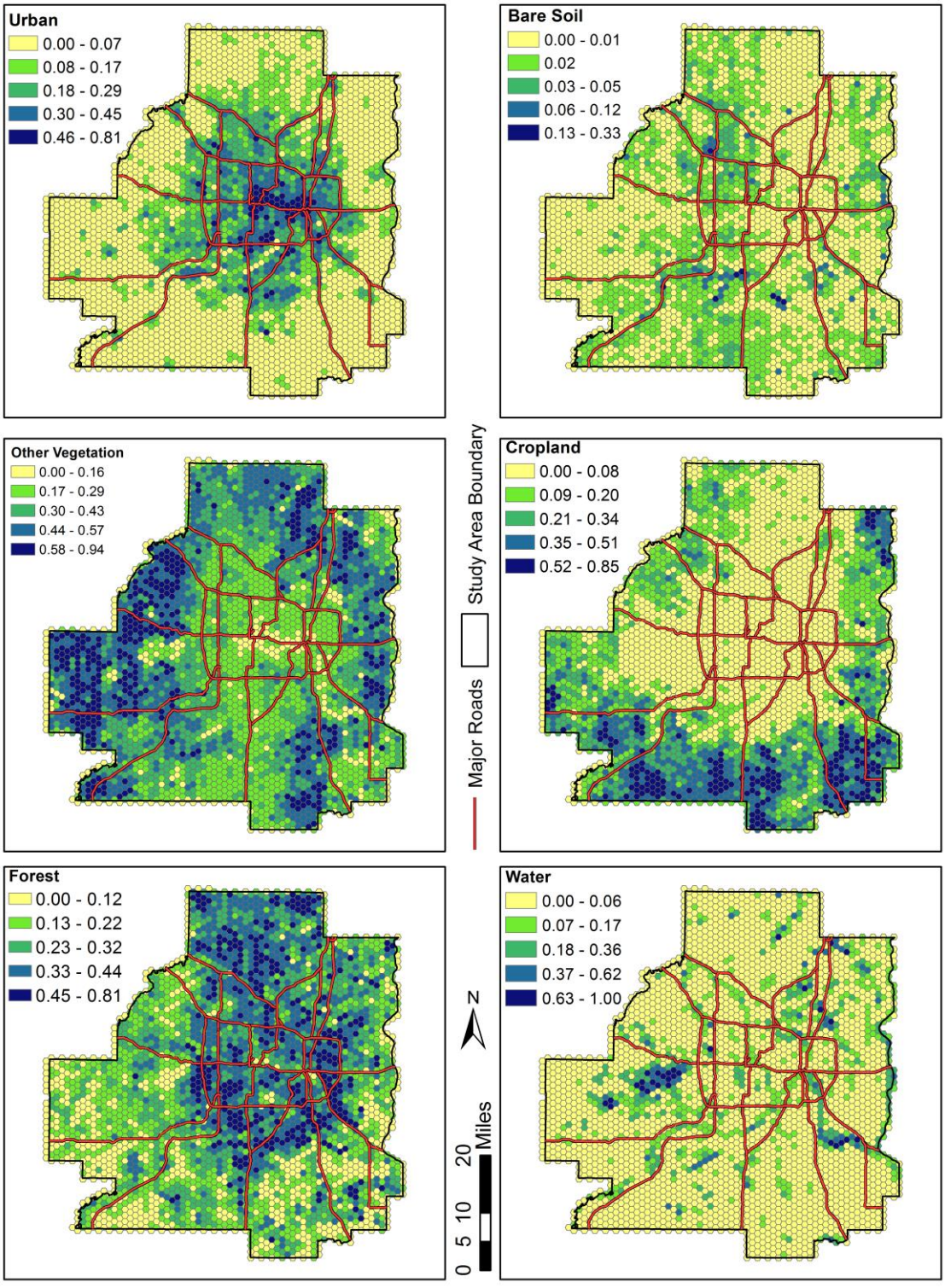

Figure 4.2: LULC distribution over study area (in square miles).

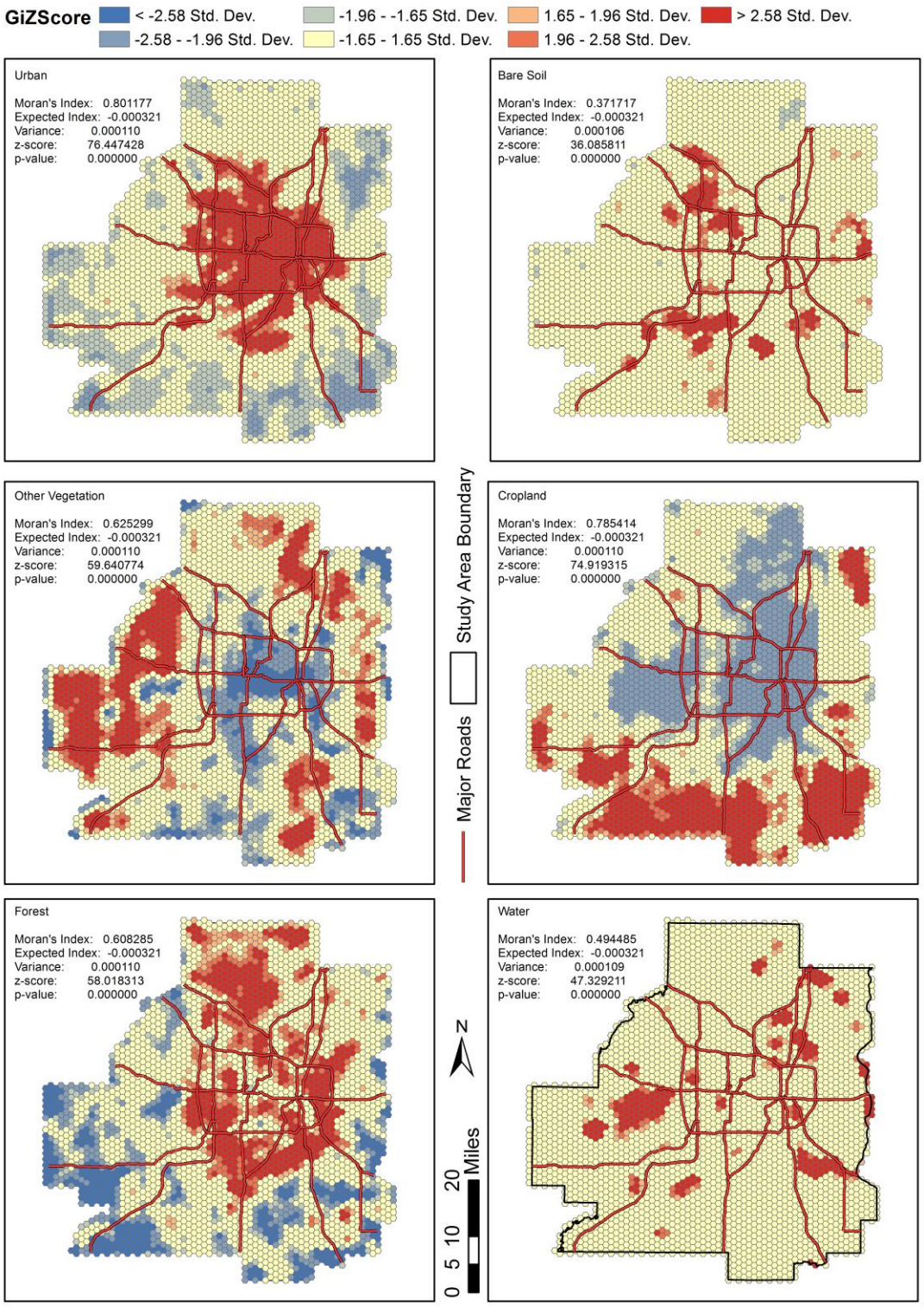

Figure 4.3: LULC hot and cold spots and Moran's I statistics.

4.1.2. LULC: by county

In addition to observing the distribution of LULC classes over the entire study area, similar comparisons were made for each county. The total area and percentage of each LULC class within each of the seven counties are shown in Table 4.1. As a supplement, Figure 4.4 represents the percentage values as a map.

The county with the highest percentage of urban land cover was Ramsey County (53 sq. mi, 31.4% of county), and the county with the largest overall area of urban cover was Hennepin County (112 sq. mi, 18.5%). In comparison, the smallest percentage of urban cover was found in Carver County (18 sq. mi, 4.8%); this is also the county with the smallest overall urban cover area. The average area of urban cover among all counties was 47.8 sq. miles, with a standard deviation of 29.5 sq. miles.

Bare soil had a slightly different distribution: the highest percentage was found in Scott County (1.68%, 6.18 sq. mi), while the highest total area was found in Hennepin County (8.54 sq. mi, 1.41%). The lowest percentage and lowest overall area was in Ramsey County (0.94 sq. mi, 0.55 sq. mi). The mean area of bare soil in all counties was 4.94 sq. miles, and the standard deviation was 2.35 sq. miles.

Other vegetation was most prevalent in Carver County by percentage (50.6%, 190 sq. mi) and Hennepin County by total area (220.7 sq. mi, 36.4%). The smallest amount of other vegetation in overall and by percentage was found in Ramsey County (40.1 sq. mi, 23.6%), which is also the smallest county. The mean area of other vegetation among all counties was 162.3 sq. miles, with a standard deviation of 56 sq. miles.

The most cropland cover was identified in Dakota County (166.8 sq. mi, 28.35%), and the highest percentage of cropland was in Scott County (32.6%, 120 sq. mi). Ramsey County again had the smallest total area and percentage, with only 2.58 sq. miles and 1.5% of total area. On average, the cropland cover area was 75 sq. miles (Standard deviation: 51.4 sq. mi).

Forest was the second-most prevalent class in the study area. The county with the largest total area of forest was Hennepin County (173 sq. mi, 25.5%), while the highest percentage of forest was in Anoka County (35.5%, 158 sq. mi), closely followed by Ramsey County (34.5%, 58.36 sq. mi). Ramsey was also the county with the smallest total forest area, while Carver County had the smallest percentage of forest cover (16%, 60.3 sq. mi). The average area of forest among all counties was 110 sq. miles, with a standard deviation of 43.7 sq. miles.

Hennepin was the county with the largest overall area of water cover (47.9 sq. mi, 7.9%), while the highest percentage was found in Ramsey County (8.4%, 14.3 sq. mi). Ramsey County also had the smallest total area of water, closely followed by Scott County (14.31 sq. mi, 3.9%). The smallest percentage of water cover was in Dakota County (3.68%, 21.6 sq. mi). The mean area of water among the seven counties was 24.2 sq. miles, with a standard deviation of 11.4 sq. miles.

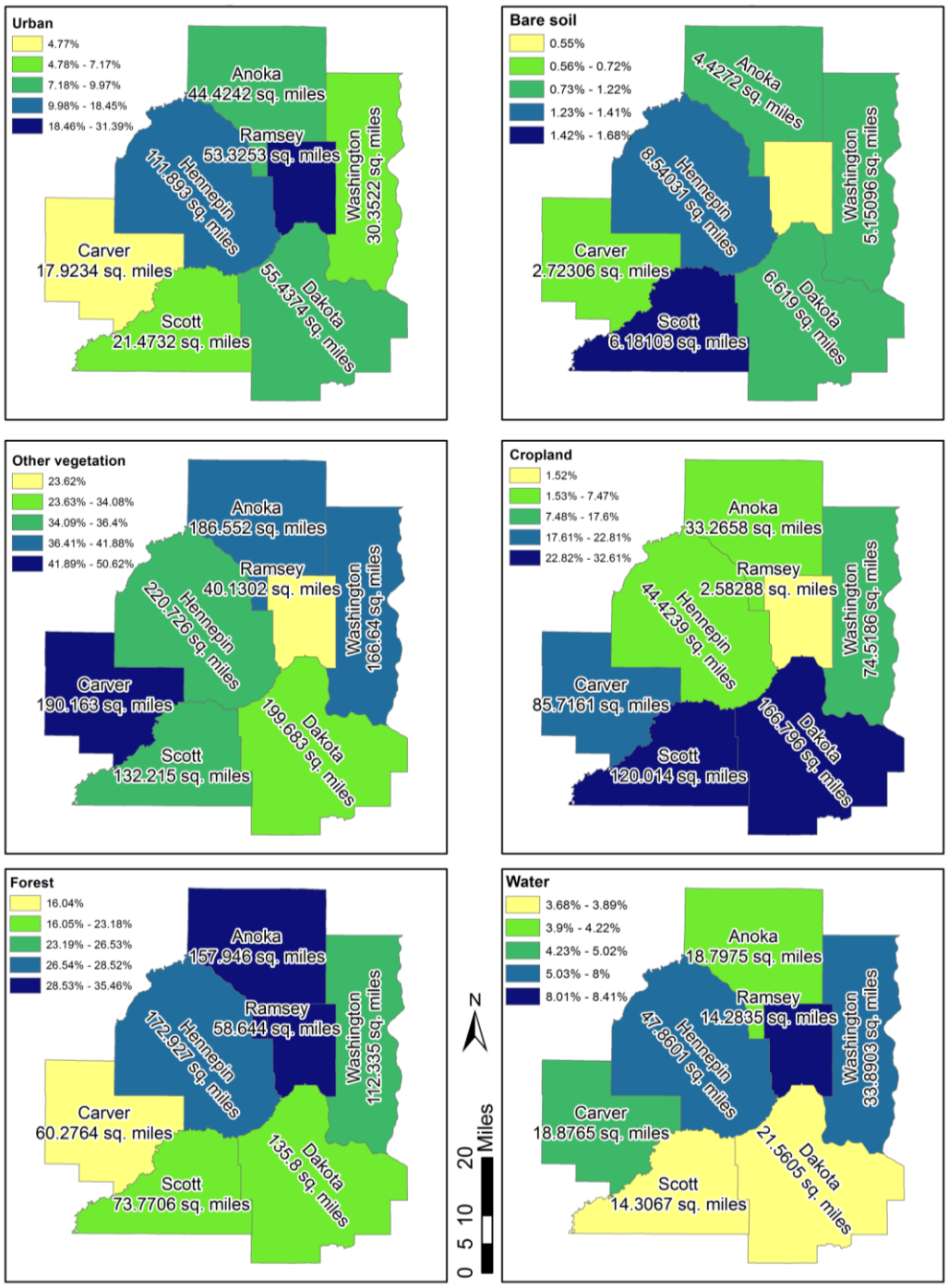

Figure 4.4: LULC cover by county. Map annotation shows area of indicated LULC class for each county.

4.1.3. LULC: by city and township

In addition to extracting impervious surface area for each county, the same procedure was conducted for each city and township within the TCMA. There are a total of 179 cities and townships in the seven county TCMA, many of which are rural townships with small population and large land area, or independent municipalities located within urbanized area, typically with small land area but dense population. The smallest entity is the City of Landfall (less than 0.1 sq. mi), and the largest entity the City of Minneapolis (58 sq. mi). The overall mean area of the 179 political entities analyzed here was 17 sq. miles with a standard deviation of 14.7 sq. miles. Complete data for all municipalities observed in the study area are found in Appendix C.

The spatial patterns of LULC class distribution are very similar to those by county (see Figure 4.4 Figure 4.5), but more detailed due to the smaller size of the cities and townships compared to counties. Urban land cover is mostly concentrated in the central part of the study area, with some outlying pockets in the perimeter. Interestingly, when sorted by percentage of urban cover, the highest numbers are found in some of the smallest entities. For example, Hilltop and Osseo are both less than 1 sq. mile in total size, but have 55.8% and 49.5% urban cover, respectively. In fact, Hilltop is the only municipality with more than 50% urban land cover. In comparison, both Minneapolis and St. Paul have a total size of about 55 sq. miles, and urban cover of 40.6% and 41.7%, respectively. These two cities have the largest overall area of urban cover of all the municipalities in the study area (23.66 sq. mi and 23.36 sq. mi). This is followed by Eagan (10.18 sq. mi, 30.4%), Bloomington (9 sq. mi, 23.5%), and Plymouth (8.44 sq. mi,

23.8%, which are some of the most urbanized and commercial suburbs in the study area. The lower percentages of urban land cover are found in small municipalities, such as Douglas, with 0.5% (0.17 sq. mi), Miesville (0.02 sq. mi urban, 1.1%; 1.74 sq. mi total) and New Trier (0.003 sq. mi urban, 1.5%; 0.18 sq. mi total area. New Trier is also the municipality with the smallest overall amount of impervious surface, followed by Miesville. However, some other municipalities that have very small overall areas of urban cover have relatively high percentages. Examples of this are Willernie (0.03 sq. mi, 22.9%) and Landfall (0.04 sq. mi, 44.7%). This appears to be mostly due to their location within the study area. Less urbanized municipalities are generally found on the perimeter of the area, while there are some very small, still independent municipalities in the center of the study area. The mean area of bare soil among all municipalities was 1.87 sq. miles, with a standard deviation of 3 sq. miles. The average percentage was 16.1%, with a standard deviation of 12.6%.

Bare soil was most prevalent by percentage in smaller municipalities, such as Rockford (7.5%, 0.02 sq. mi bare soil; 0.25 sq. mi total) or Grey Cloud (4.65%, 0.14 sq. mi bare soil; 3.07 sq. mi total), which could be attributed to extraction activities or a higher amount of unpaved roads found in these areas. Savage was a larger municipality with relatively high percentage of bare soil (4.25%, 0.7 sq. mi bare soil; 16.42 sq. mi total), which is more likely caused by construction activities at the time of the image capture. When sorted by total area, bare soil was most prevalent in Maple Grove (1.32 sq. mi, 3.76%), followed by neighboring Shakopee (0.82 sq. mi total, 2.8%). The smallest total area of bare soil was in Landfall (0.00012 sq. mi, 0.08%), followed by Willernie

(0.00013 sq. mi, 0.13%), and Gem Lake (0.05 sq. mi, 1.1%). Most of the entities with small areas of bare soil have a very small total size (less than 1 sq. mi). The mean area of bare soil among the municipalities was much smaller than that for urban cover at only 0.19 sq. miles (standard deviation: 0.21). The average percentage of bare soil was 1.31%, with a standard deviation of 1%.

Other vegetation was the overall most prevalent class. This is likely due to large areas of conservation lands in the TCMA perimeter, along with many extensive parks in the central part of the study area. Among the municipalities, Hollywood had the highest percentage of other vegetation (63.2%, 22.57 sq. mi). There were a total of 24 municipalities that had more than 50% of other vegetation cover. The lowest percentage of other vegetation was found in Landfall, which is relatively highly urbanized due to its central location (7.3%, 0.01 sq. mi). Landfall also had the lowest total area of other vegetation. Hilltop (0.03 sq. mi, 21.3%) and Willernie (0.02 sq. mi, 16%) follow in this ranking. Columbus was the municipality with the largest area of other vegetation cover (27 sq. mi, 56.8%), followed by East Bethel (23 sq. mi, 48.2%). On average, municipalities had 6.35 sq. miles of other vegetation, with the same standard deviation. The mean percentage was 35.1%, with a standard deviation of 11.8%.

For the cropland class, the largest amount was found in Eureka Township both by percentage and total area (19.3sq. mi, 54%). Ranked by percentage, this was followed by Douglas Township (53.3%, 18.1 sq. mi), while Belle Plaine Township was second if ranked by total area (19 sq. mi, 43.3%). Douglas Township was ranked third by total area. Marshan Township was fourth both by percentage and total area (52.3%, 18 sq. mi).

These townships and cities are all located in the agriculturally dominated southern part of the TCMA. The smallest amount of cropland is in the small urban mobile home community of Landfall which had the lowest total area (0.0004 sq. mi, 0.46%), while it was third to last by percentage. The lowest percentage was found in Gem Lake (0.41%, 0.0045 sq. mi). The municipalities with the lowest percentage of cropland tend to have small overall areas as well. The mean area of cropland was 2.95 sq. miles (standard deviation: 4.46 sq. mi), while the mean percentage was 13.45% (standard deviation: 14.54%).

Forest cover by total area was highest in some of the largest and most urbanized entities in the study area. Saint Paul had the largest area (17.9 sq. mi, 32%; 56 sq. mi total), followed by Minneapolis (17.5 sq. mi, 30%; 58.3 sq. mi total) and Bloomington (16.5 sq. mi, 42.9%; 38.3 sq. mi total). Therefore, some of the most densely urbanized communities also have large amounts of forest cover. Of the highest ranked entities by percentage, the majority had a small total area (less than 5 sq. miles): Birchwood Village was ranked first (0.25 sq. mi, 72.3%; 0.35 sq. mi total), followed by Willernie (0.08 sq. mi, 56%; 0.13 sq. mi total). The largest of the top ten cities and townships ranked by percentage of forest was Edina (7.42 sq. mi, 46.4%; 16 sq. mi total area). In contrast, the municipalities with the lowest percentage of forest cover were relatively diverse in terms of total size: The lowest percentage was found in Coates (5.5%, 0.08 sq. mi; 1.38 sq. mi total size), followed by Miesville (7.12%, 0.12 sq. mi; 1.74 sq. mi total). When ranked by area of forest cover rather than percentage, the lowest ranked entities also generally had small overall sizes (less than 1 sq. mi). The smallest total area of forest was found in

Landfall (0.02 sq. mi, 26.1%; 0.08 sq. mi total), followed by Hilltop (0.02 sq. mi, 17.4%; 0.13 sq. mi total). Among all cities and townships, the mean area of forest cover was 4.3 sq. miles, with the same standard deviation. The mean percentage of forest cover was 27.4%, with a standard deviation of 11.3%.

The largest total area of water was found in Orono (8.8 sq. mi, 34.6%), which was ranked fourth by percentage, followed by Shorewood (7.4 sq. mi, 57%), which was also second when ranked by percentage. These municipalities are all located around the largest lake in the area, Lake Minnetonka. Lakeland Shores was ranked first by percentage of water (0.41 sq. mi, 57.7%). In terms of lowest amount of water, the lowest ranked municipalities were Maple Plain and Hamburg, both with no water found at all. The smallest percentage of water for municipalities that had water present was Marine on Saint Croix (0.05%, 0.002 sq. mi). Ranked by total area of water, Rockford had the smallest area (0.001 sq. mi, 0.4%), followed by Loretto (0.001 sq. mi, 0.4%). The mean area of water in the cities and townships was 0.94 sq. miles, with a standard deviation of 1.3 sq. miles, and 6.63%, with a standard deviation of 8.64%.

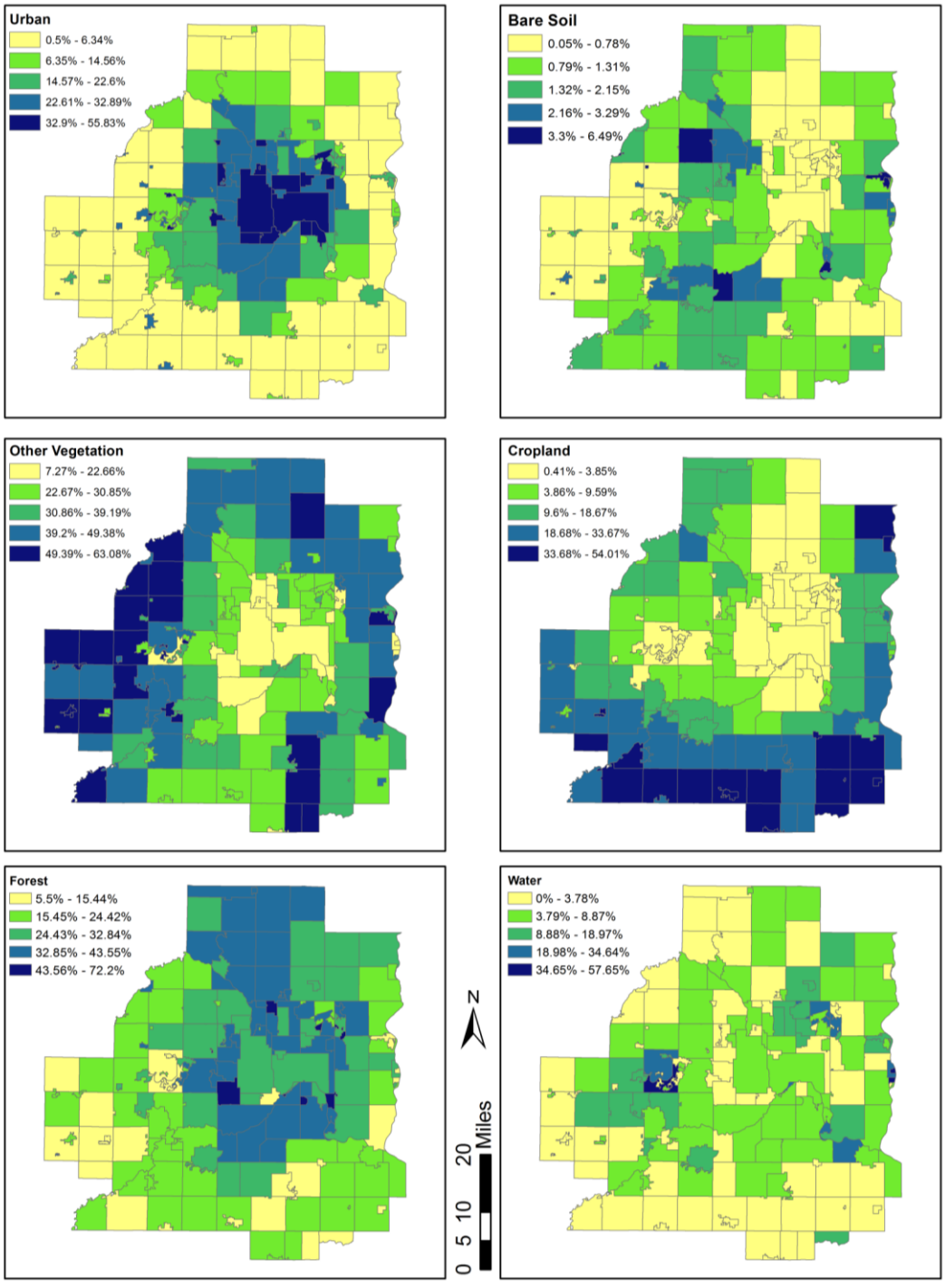

Figure 4.5: LULC cover by city or township.

4.2. Impervious surface classification

4.2.1. Impervious surface: Entire study area

The total impervious area found in the TCMA was 422 square miles (1,093 square kilometers), of the total area of 2,939 square miles (7,612 square kilometers). Therefore, approximately 14% of the area of the seven county TCMA is made up of impervious surface (Figure 4.6). The impervious surface area generally follows a pattern of highest density in the center of the study area, and decrease towards the perimeter of the study area. It further appears that impervious surface is concentrated along major transportation corridors. ISA also appears to closely follow the distribution of urban land cover, which it is mainly comprised of. The map in Figure 4.7 shows the amount of impervious surface based on a regular hexagonal grid. Each hexagon has a side length of 1 km, giving it an area of approximately 1 sq. mile. The hexagonal grid was used to show spatial patterns of impervious surface density based on a regular pattern, rather than political or other arbitrary entities. The major roads shown are Interstate and Minnesota State highways. In order to quantitatively confirm that the spatial distribution of impervious surface in the study area follows a pattern of clustering, spatial autocorrelation was calculated for these hexagons using the Global Moran's I index (Equation used: see Appendix A; Complete results: see Figure 4.11). Moran's I gives an indication of whether there is a spatial pattern found in the data, and whether the data are clustered, dispersed, or randomly arranged. The resulting Moran's I index for these hexagons was 0.65, which indicates a tendency towards clustered spatial distribution of the analyzed features. The z-score of 62 and a p-value of 0 indicate that the Moran's I results are statistically significant at a

confidence interval of 99%. In order to confirm the assumption that the clustering of impervious surface is stronger in the central area of the study area, around the Twin Cities of Minneapolis and St. Paul, a hot spot map was also generated using the Getis-Ord G* statistic (see Appendix B for equation). The resulting map (see Figure 4.11) confirms that there is a clustering of high impervious surface values towards the center of the study area and adjacent to major transportation corridors. Additionally, there are several cold spots towards the perimeter of the study area.

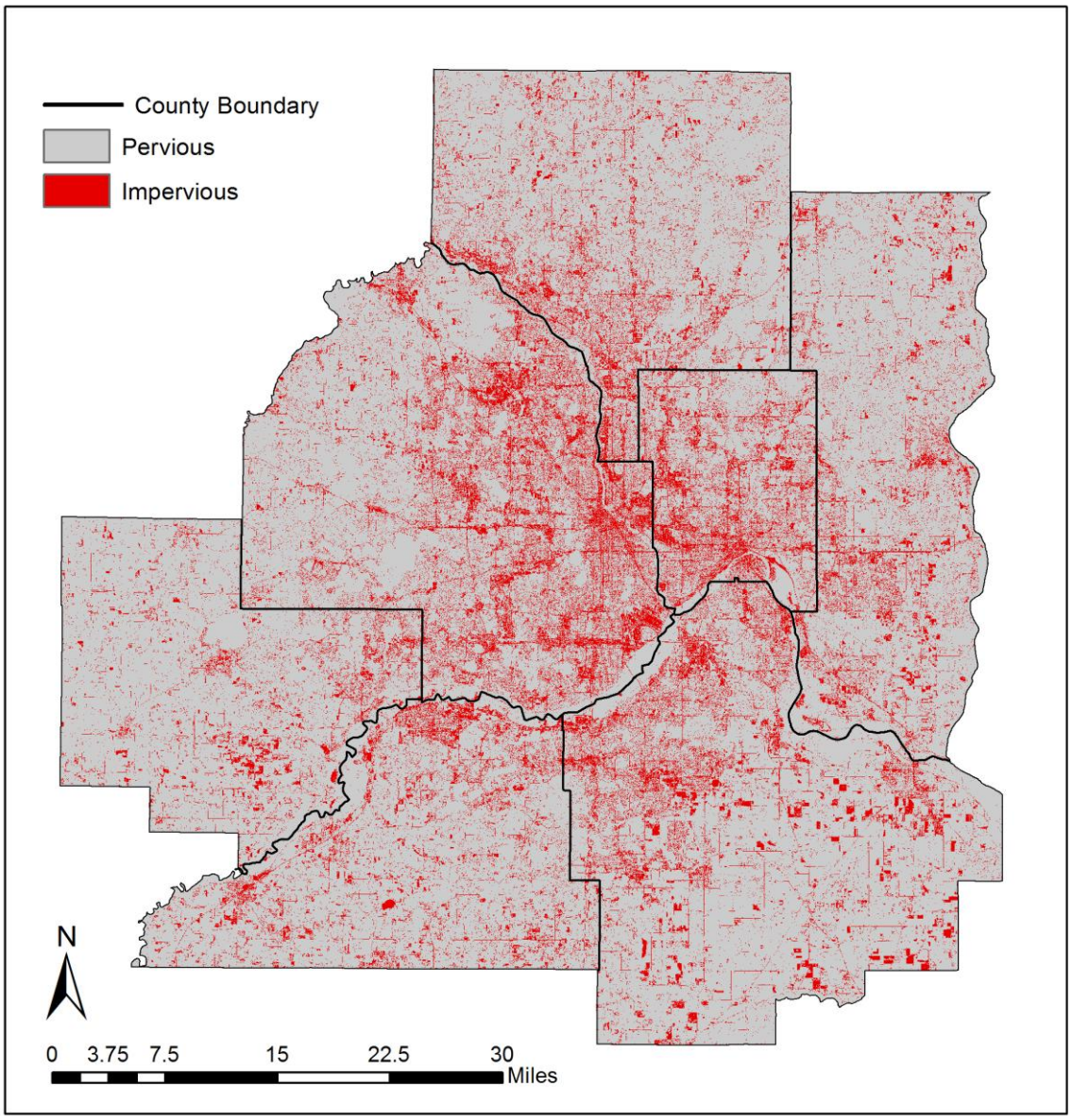

Figure 4.6: Impervious surface area map.

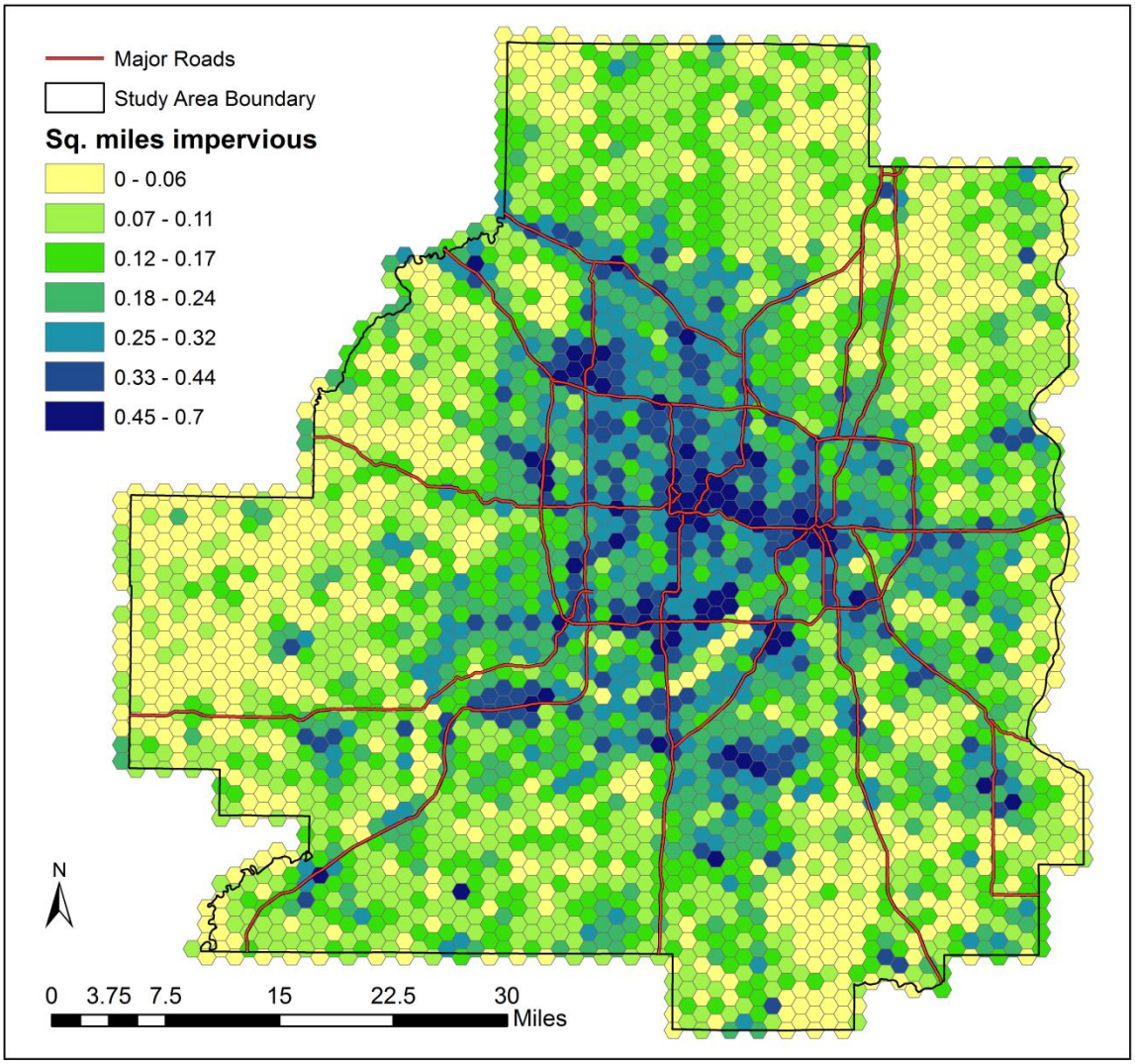

Figure 4.7: Impervious surface density (1 sq. mile hexagon area).

4.2.2. Impervious surface: by county

Impervious surface area was also evaluated based on political entities. The total ISA (422 sq. miles) in the TCMA study area is divided among the seven counties as follows: 32 sq. mi (8% of TCMA total) are found in Carver County, 41 sq. mi (10%) in

Ramsey County, 48 sq. mi (13%) in Scott County, 49 sq. mi (12%) in Washington County, 53 sq. mi (13%) in Anoka County, 57 sq. mi (20%) in Dakota County, and 116 sq. mi (28%) in Hennepin County (see Figure 4.9). Therefore, in terms of the total area of impervious surface within each county, Hennepin County leads with more than twice the area of the second-ranked Dakota County, while Carver County has the smallest amount of impervious surface. This ranking coincides with the total area of each county, except for Carver County, which is larger than both Ramsey and Scott counties (see Figure 4.8). Complete results are given in Table 4.2.

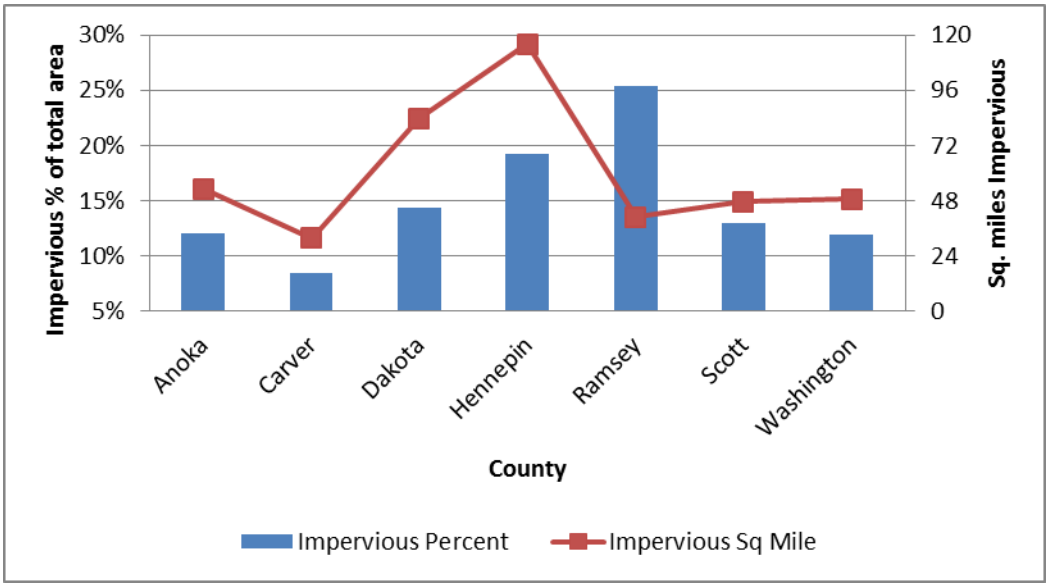

Figure 4.8: Total area and impervious percentage for each county.

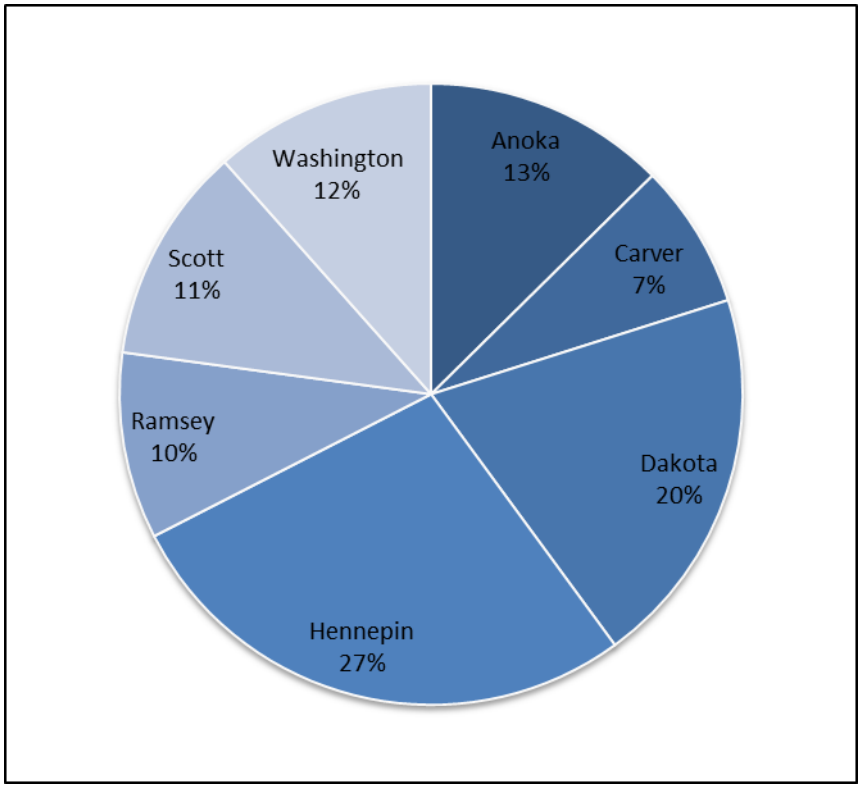

Figure 4.9: Percentage of total TCMA impervious surface by county.

County	Pervious Sq Mile	Impervious Sq Mile	Total Sq Mile	Pervious Percent	Impervious Percent	Impervious % of total
Anoka	386.90	53.09	439.98	88%	12%	13%
Carver	343.98	31.72	375.70	92%	8%	8%
Dakota	499.14	83.63	582.77	86%	14%	20%
Hennepin	487.59	116.02	603.61	81%	19%	28%
Ramsey	120.15	40.90	161.05	75%	25%	10%
Scott	320.23	47.59	367.83	87%	13%	11%
Washington	359.31	48.74	408.05	88%	12%	12%
Total:	2517.30	421.69	2938.99	n/a		100%
Mean:	359.61	60.24	419.86	85%	15%	14%
Standard Deviation:	116.94	27.23	137.58	5%	5%	6%

Table 4.2: Impervious surface by county.

When comparing the percentage of ISA within each county, this ranking is slightly different. Carver County is still at the bottom of the ranking (8% of its total area is impervious). This is followed by Washington and Anoka counties (both 12%), Scott County (13%), Dakota County (14%), Hennepin County (19%), and Ramsey County with the highest percentage (25%). These numbers are illustrated in Figure 4.8. Therefore, Hennepin County has the largest absolute impervious surface area, and is also the largest county by total area. In comparison, Ramsey County has a larger percentage of impervious surface area than Hennepin County, but it is the smallest county by total area. Hennepin and Ramsey are the counties where the two major municipalities of the TCMA, Minneapolis and St. Paul (the Twin Cities), are located. Figure 4.10 gives an overview of the location of each county, and the amount of impervious surface found within it.

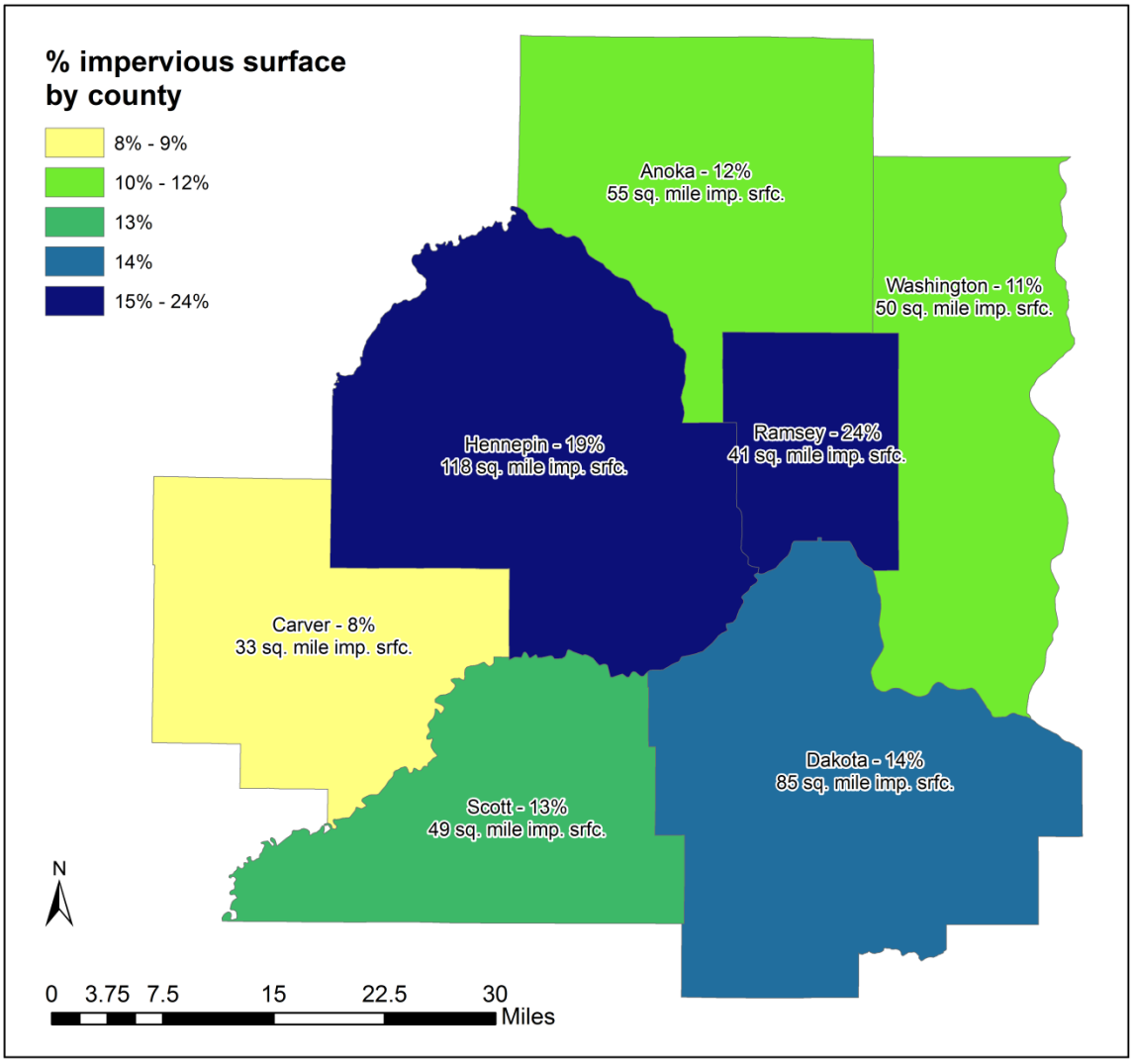

Figure 4.10: Impervious surface by county (percentage and total area).

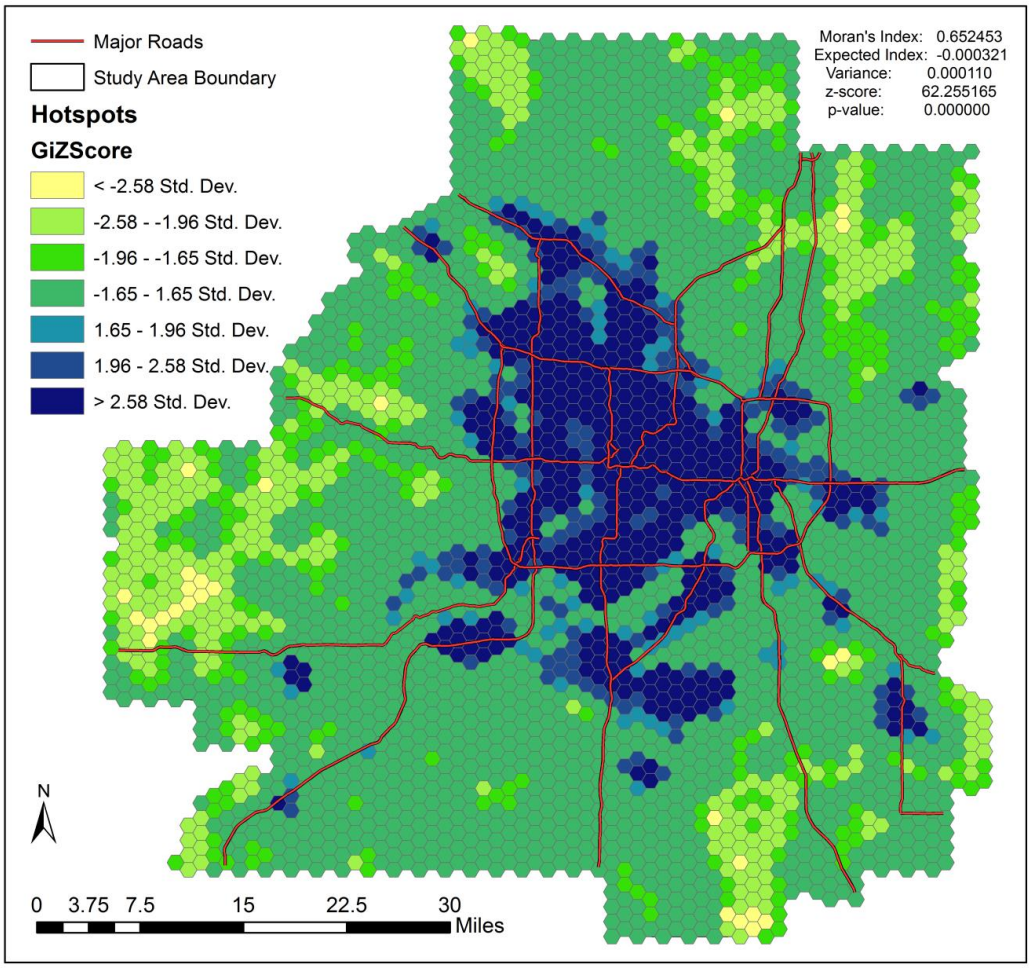

Figure 4.11: Hotspot map of impervious surface hexagons.

4.2.3. Impervious surface: by city or township

Similar to the LULC section, ISA was also analyzed at a city and township level. The complete results are listed in Appendix E. In general, most political entities that are relatively small in size also had relatively small impervious surface areas. Some of the very small cities, however, had relatively large areas of impervious surface. For example,

Hilltop had a total area of 0.13 sq. mi, of which 0.06 sq. mi were impervious (50%). In 2010, Hilltop had a population of 744 (US Census 2013). Another small city is Osseo, which has a total area of 0.72 sq. miles, of which 48% were impervious surface. Osseo's 2010 population was 2,430 (US Census 2013). This is not typical however, as the majority of these small cities have small areas of impervious surface, and low population numbers. In general, most of the small cities (less than 1 to 2 sq. mi) also have a small percentage of impervious surface area (less than the mean of 18%).

The map in Figure 4.12 indicates the percentage of impervious surface area for each city and township within the seven county TCMA. It is evident that municipalities in the center of the study area, close to the Twin Cities of Minneapolis and St. Paul, generally have greater percentage of impervious surface compared to those municipalities on the outer perimeter of the study area. This pattern can also be observed in Figure 4.7. An additional interesting pattern is that many small municipalities in the outer, more rural perimeter of the study area also have relatively high impervious surface ratios. This could be an indication that these municipalities act as small, regional centers for the surrounding, less densely populated area.

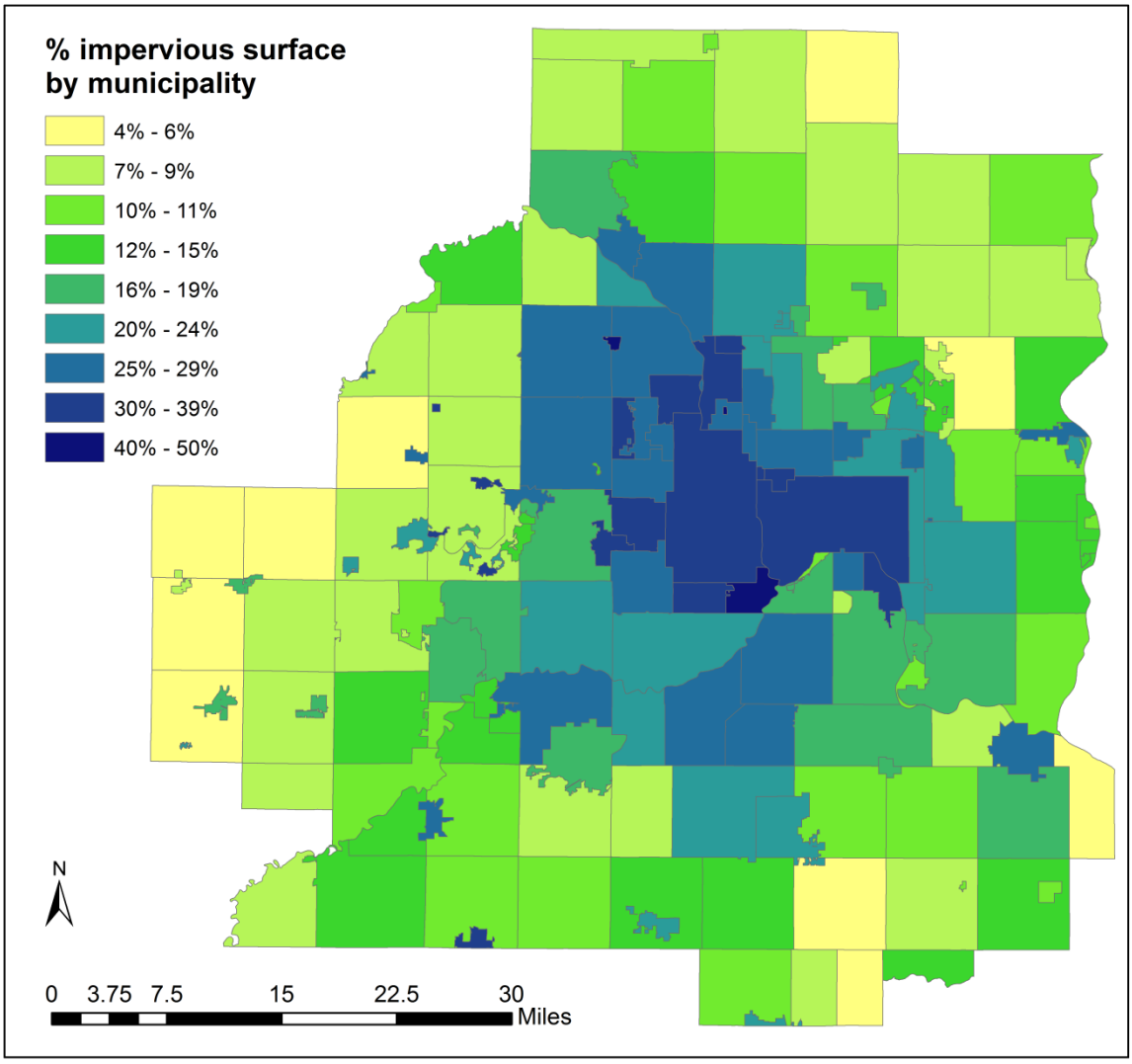

Figure 4.12: Percentage of impervious surface by city or township.

4.3. Accuracy Assessment

4.3.1. Land Use and Land Cover Accuracy

Accuracy of the LULC map was estimated based on a set of 300 randomly distributed sample points. Each point was assigned a reference value. These reference class values were cross-tabulated with the classified value at the point location, and recorded in a standard accuracy matrix (see Table 4.3). The overall classification accuracy for the LULC map was 74%. Kappa accuracy was estimated to be 66%. The highest user's accuracy (error of commission) was achieved for the water class (100%), and the lowest for the other vegetation class (63%). Water also had the highest producer's accuracy (error of omission; 89%), whereas cropland had the lowest producer's accuracy at 57%.

The greatest amount of confusion appears to be between relatively similar classes, such as cropland and other vegetation, and forest and other vegetation. Urban and water were classified with the highest accuracy, likely because both of these classes are very unique in terms of their spectral attributes and shape and texture factors. In comparison, cropland and other vegetation have relatively low classification accuracy. This is likely due to the fact that these classes are very similar both spectrally, but also in their shape and texture.

Classified Data	Reference Data								
		Water	Forest	Cropland	Other Veg.	Urban	Bare Soil	Grand Total	User's (Commission)
	Water	16						16	100%
	Forest	1	56	4	11	2		74	76%
	Cropland		1	40	8	2		51	78%
	Other veg.		15	26	72	1	1	115	63%
	Urban	1	1		2	33		37	89%
	Bare Soil					2	5	7	71%
	Grand Total	18	73	70	93	40	6	300	
	Producer's (Omission)	89%	77%	57%	77%	83%	83%		
	Overall Accuracy	74%							
	Kappa Accuracy	66%							

Table 4.3: Accuracy matrix for LULC map.

4.3.2. Impervious Surface Accuracy

The accuracy of the impervious surface classification was based on the reference matrix generated by See5 when the decision tree model was built. An independent set of 300 random points was provided to See5 for accuracy assessment purposes (test cases). Based on the matrix, overall, per class user's and producer's accuracy and the Kappa coefficient were calculated. The overall classification accuracy was 95%. User's and producer's accuracy for the pervious class were both above 95%, while they were only 90% and 76% respectively for impervious surface. The Kappa coefficient was estimated to be 79%. The complete accuracy matrix is shown in Table 4.4.

In general, these values can be considered very good. The impervious surface model appears to have a slight tendency to over-estimate the amount of impervious surface. However, this seems to have a relatively low impact on the overall accuracy. The relatively high omission error for impervious surface is likely due to the timing of the image acquisition for this study: NAIP images are meant for agricultural crop inventories and are therefore flown during the peak growing season. It is common for tree canopies to cover parts of ISA features such as roads and buildings. The high spatial resolution of the data used here could enhance the impact of this.

Classified data	Reference Data			
	Pervious	**Impervious**	**Total**	**User's (Commission)**
Pervious	250	11	261	95.79%
Impervious	4	35	39	89.74%
Total	254	46	300	
Producer's (Omission)	98.43%	76.09%		
Overall Accuracy:	95.00%			
Kappa:	79.46%			

Table 4.4: Accuracy matrix for impervious surface classification.

5. Discussion

5.1. LULC and ISA Classification

5.1.1. LULC and ISA Distribution and Patterns

Minneapolis and St. Paul, the "Twin Cities", are the core cities of the seven county metropolitan area. Therefore, these two cities comprise the high-density urban core of the study area. Centrally located, their suburbs are spread around these core cities relatively evenly. Notable suburbs are those to the south of the core of the study area, particularly Bloomington, which is considered a third major city to some extent, rather than a suburb of the Twin Cities. In general, the suburbs surrounding Bloomington are also some of the most densely urbanized areas in the TCMA.

Settlement in the Twin cities first began in this area, at Fort Snelling, which is located between Bloomington and St. Paul. Major corporate and retail as well as transportation infrastructures, such as the international airport, are located in this part of the study area. This pattern is reflected in both the ISA classification and the urban cover found in the LULC classification. The pattern of impervious surface is similar to the distribution of urban and bare soil land cover. Urban surface is the main constituent of ISA, with compacted soil and other components making up much smaller portions. Therefore, the close relationship between ISA and urban cover is to be expected. When adding the 12.4% urban cover and 1.1% bare soil found in the study area, the resulting percentage of 13.5% comes very close to the 14% of ISA identified. This close agreement of the two classifiers is an indicator of the quality of the ISA classification. Figure 5.1 shows the population density based on 2010 Census data at the block level for the entire

study area. This map is a clear reflection of the patterns of urban LULC and ISA seen in previous maps. High population densities, like urban land cover and ISA, are concentrated in the center of the study area, along a corridor of major highways to the northwest of the core Twin Cities, and in the suburbs to the south. This pattern agrees with findings by Baerwald (1978), who predicted a shift from Central Business District (CBD) oriented cities to Suburban Freeway Corridor (SFC) oriented cities. His study used the southern suburbs of the TCMA as a model for this concept, which is confirmed by both the urban and ISA patterns found here. The perimeter, particularly the agriculturally dominated southern and western parts of the study area, is much less densely populated.

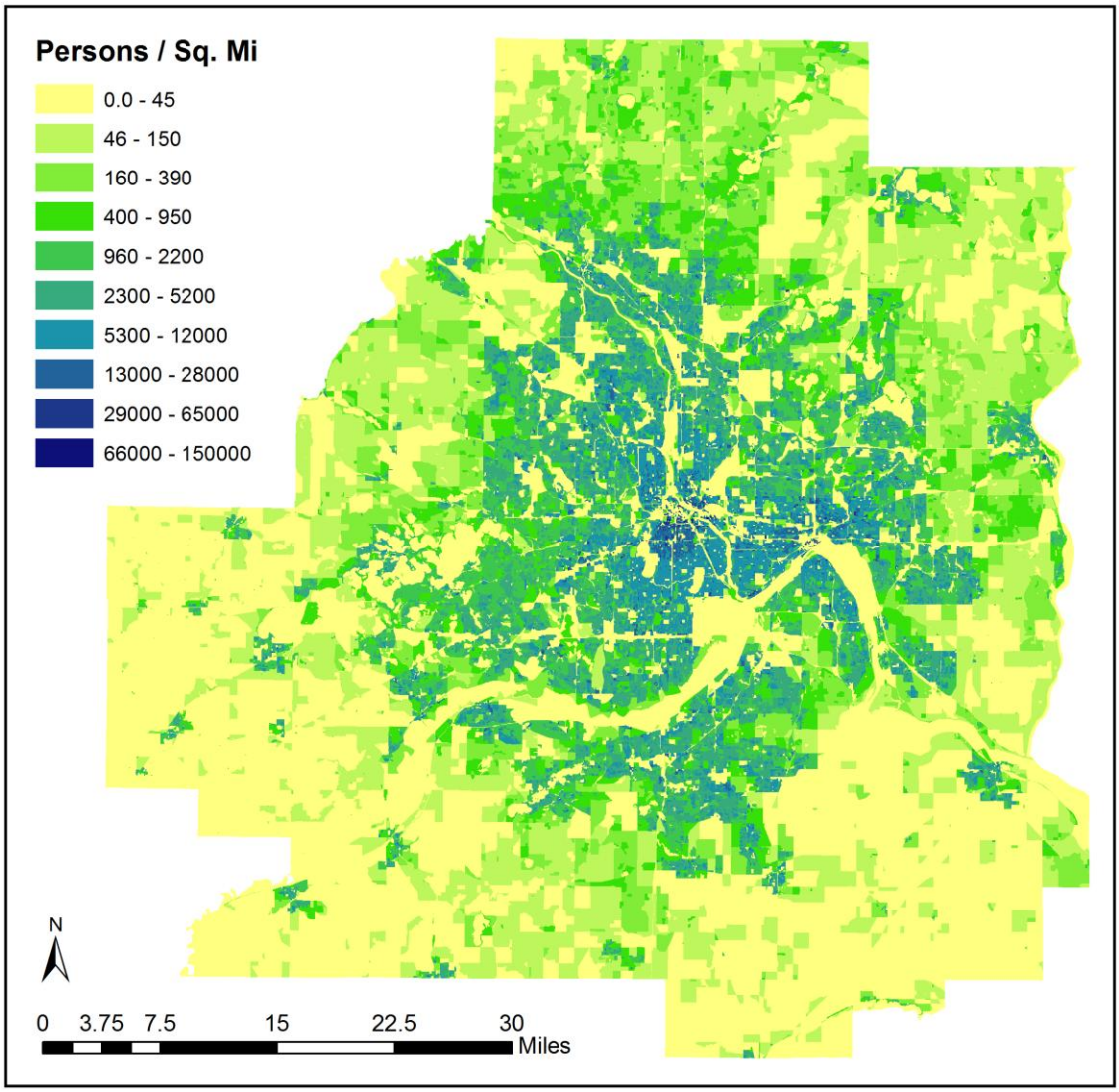

Figure 5.1: Population Density in TCMA in 2010.

The TCMA has an extensive system of public parks and trails, especially surrounding the numerous lakes. This partially explains why some of the most urbanized cities in the study area also have large areas of forest: many of these parks are located centrally and are dominated by woody vegetation. Temperate deciduous forests comprise much of the native vegetation type in most of the study area. However, during periods of

intense urbanization and agricultural development, much of this native vegetation was lost. Many of the current parks, however, were established early, and were therefore able to preserve a somewhat natural state.

Large parts of the other vegetation class are likely green spaces in public and private places, such as yards and lawns, sports fields and public parks, but also natural lands such as grassland, wetlands, and conservation lands. The two major conservation programs in Minnesota are the Conservation Reserve Program (CRP), which puts agricultural land out of production and plants native vegetation, and Reinvest in Minnesota Reserve Program (RIM), which works similarly to CRP, but focuses on riparian lands and wetlands.

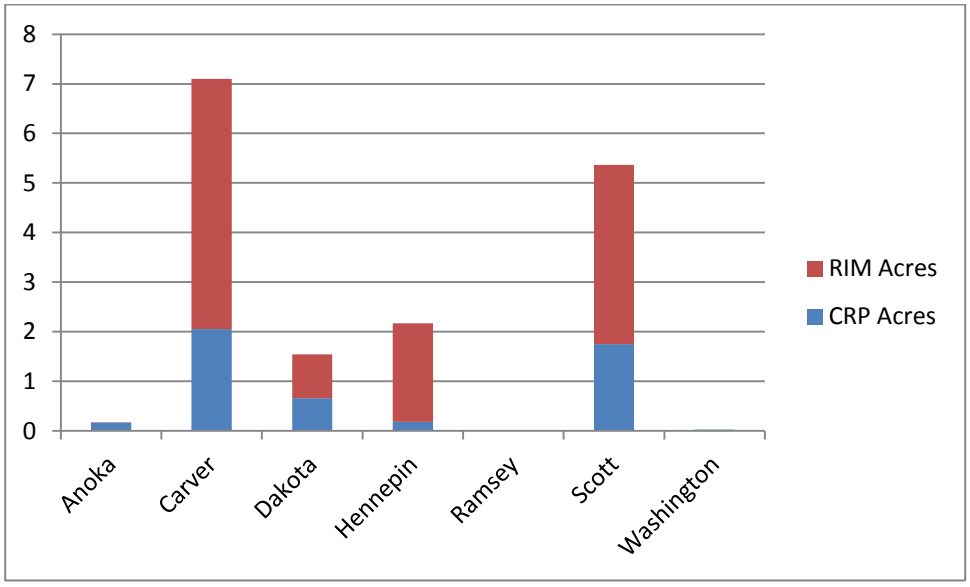

Figure 5.2: Conservation lands by county (UMN 2013).

Overall, the county with the greatest acreage of conservation lands is Carver County, followed by Scott County. This ranking coincides with the percentage of other vegetation land in the LULC classification. The majority of CRP and RIM lands should be captured by the other vegetation class, while a small part of the RIM land may also contain forest.

The dominance of agricultural land use in the southern and western parts of the study area generally falls in line with a pattern observed on a state-wide scale: Agriculture is one of the major land uses to the south and west of the Twin Cities in Minnesota. These areas are actually some of the most productive growing regions in the US. Throughout the northeast of the state, land use shifts towards forestry, beginning at the Twin Cities. The beginnings of this pattern can be observed in the LULC distribution found here, which shows that a large part of the land in the north-central part of the study area is covered by forest. The TCMA lies at the southern extent of Minnesota's vast expanse of aspen.

Even though this is not evident when comparing this classification with previous studies, as discussed in more detail in 5.2, the TCMA has been growing relatively steadily in terms of population and urban infrastructure. This trend can be observed, for example, in population figures derived from Census datasets. However, the study area has large amounts of other LULC cover types. Much more water is found in the TCMA than many other urban areas due to the large number of lakes, resulting from the last glacsine episode. Further, the TCMA has many nature parks, even close to the urban core of the metropolitan area. Agriculture is also very strong in Minnesota in general, and also

in parts of the TCMA. These are things local decision makers have to consider when moving forward with urban planning initiatives. The TCMA has historically been less fond of automobile transportation than other metro areas. This is in part due to the fact that many residents are able to live close to work and activities they pursue outside of work. Urban planning in the TCMA moves towards supporting more public transit again to support this notion. A light rail line is already in place, with many others being planned. The historic St. Paul Union Depot has recently been remodeled and will be used for TCMA transit and long distance rail travel in the future. However, as was evident in the LULC and ISA maps, much of the urban development in the area currently clusters around major automotive transportation routes. While the TCMA is the economic and population center of Minnesota, planners and decision makers have to keep in mind that agriculture is considered the backbone of Minnesota's economy. Even in the highly urbanized TCMA, agriculture still plays a somewhat important role and should not be neglected.

5.1.2. LULC and ISA Methodology

As is shown in the accuracy assessment performed for both classification parts of this study (Chapter 4.3), the ISA extraction methodology delivers better results than the LULC method. This is likely due to several factors. Both techniques used similar input data, and samples were collected by the same technique. However, the ISA classification only included two classes, while the LULC classification extracted six classes. Often, more classes in a classification model results in more chances for confusion between classes and therefore a higher degree of misclassification. In addition, while Feature

Analyst appears to make use of advanced techniques, much of its processing happens behind the scenes, and cannot be influences by the analyst. In comparison, the methods See5 uses to reach its resulting model is published in the literature and well described, and the resulting model can be read and interpreted by humans very easily. Feature Analyst gives no insight into its classification algorithm. This can lead to situations where the software possibly chooses a model that is not necessarily suitable for the data being evaluated. Another issue could be that specific training samples cause issues, which would go unnoticed in Feature Analyst, but could be noticed when using See5. Therefore, some confusion could likely be avoided by using See5 because its process is more transparent and the analyst has more influence over the software. Therefore, it is easier to fix errors in the classification. Finally, it is also possible that the relatively higher complexity of Feature Analyst models could introduce additional error compared to See5.

5.2. Previous LULC Classification Study

The most recent previous LULC classification in the TCMA was conducted with 2006 Landsat imagery with 30 meter spatial resolution by Yuan (2009). The resulting map can be found in Figure 5.3. The 2006 study used a slightly different classification scheme, which was adopted to match the one used here as closely as possible, and summarized, by county in Appendix D. In general, the distribution of LULC classes across the study area is similar in both studies. However, there are some differences in the area statistics. Urban cover was estimated higher in the 2006 map, which could be due to several reasons. It is possible that the use of lower spatial resolution Landsat TM imagery slightly overestimated urban land cover. The TCMA has a high amount of urban

and suburban land cover, which includes many features that are smaller than the 30 meter Landsat TM pixels. Therefore, in this situation, the mixed-pixel problem exhibited in all remote sensing data likely has a higher impact in Landsat images than in the 1 meter aerial images. Further, due to differences in the classification scheme, it is possible that some features that were not considered urban in this study were included in Yuan's urban class. This could apply to suburban areas, where there is a mixture of urban features and vegetated areas. It is also possible that some areas that were included in bare soil here could be considered urban in Yuan's 2006 study.

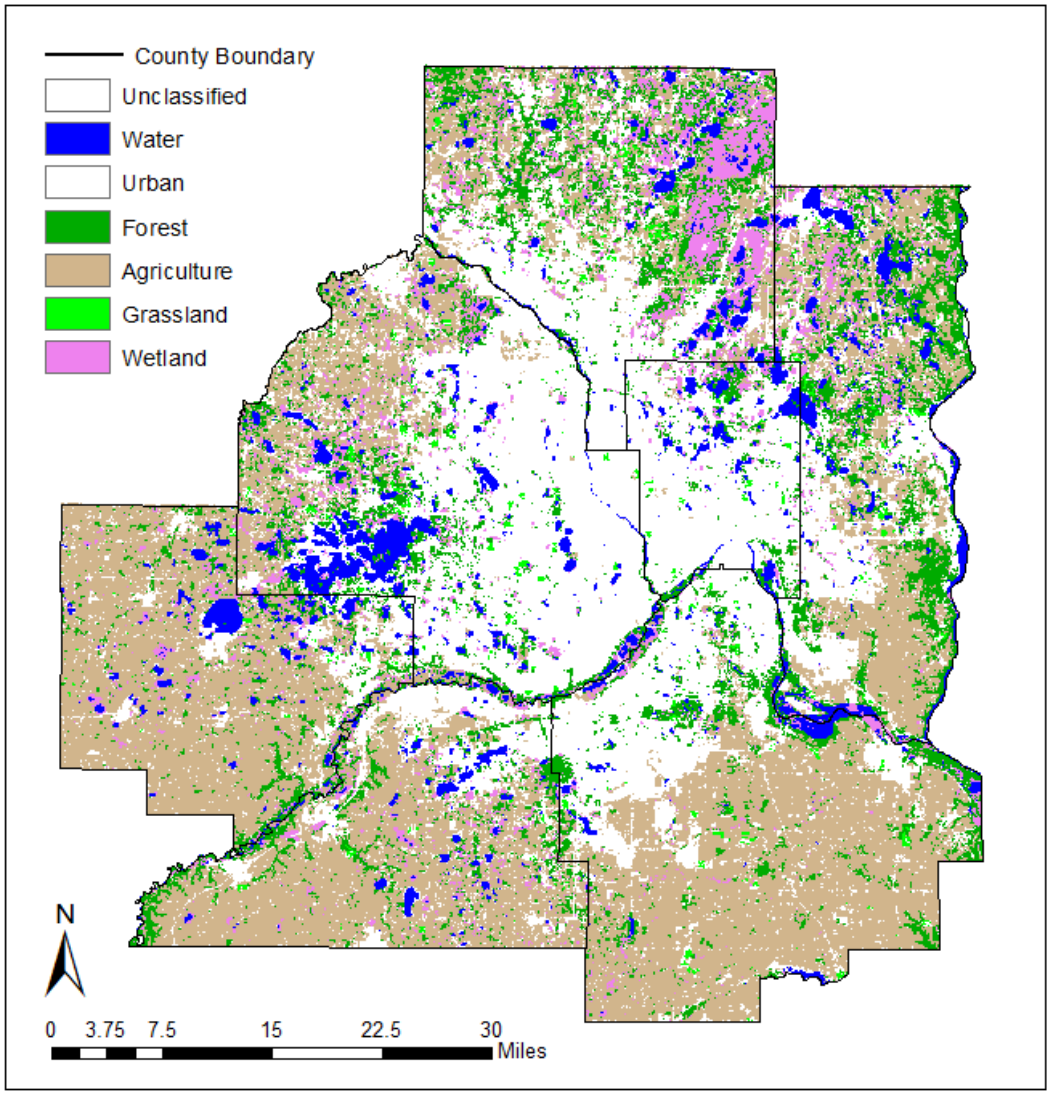

Figure 5.3: 2006 LULC Classification Map (Yuan 2009).

There was also some discrepancy in the vegetation classes. For example, this study estimated the forest cover at about twice as large as the 2006 study, while cropland was estimated at less than half the area shown in Yuan's 2006 classification. This study further estimated the amount of other vegetation that is much higher than Yuan. These discrepancies could be due to the mixed pixel problem discussed earlier. Many large,

contiguous areas of vegetation, like forest and cropland, are better captured by the lower spatial resolution Landsat TM imagery. The additional spectral information in the Landsat imagery further enhances the classification of these vegetation types. At the same time, however, some vegetation in suburban areas, where there is a mixture of LULC classes of relatively small patch sizes, may be underestimated in the 2006 classification.

Water was the only class that was nearly identical in both classifications. This is likely due to the very unique spectral signature of water compared to all other LULC classes in the study area, and the large number of relatively large bodies of water.

Due to the assumed discrepancies caused by differences in classification methods, it is not possible to assess the changes in LULC that have occurred in the study area between 2006 and 2010. While the TCMA is growing relatively fast and LULC changes can happen relatively quickly, this time period is still relatively short. It may be appropriate to conduct another study similar to this one in the near future using available aerial imagery to be able to assess and model LULC change in the TCMA.

In general, Landsat TM imagery has advantages and disadvantages compared to the aerial imagery used in this study. Both types of imagery are available free of charge, however, that is not always the case with aerial imagery. Landsat imagery is typically acquired with much higher temporal frequency than aerial imagery. Due to its lower spatial resolution, it has smaller file sizes for any given study area, but also gives less spatial detail. At the same time, Landsat TM imagery has a higher spectral resolution, as it includes several infrared bands, rather than just one as was the case for the aerial

images used here. These characteristics make Landsat images suitable for studies that involve change detection or multi-temporal classifications, studies of regional or larger areas of interest, or study areas that are not very diverse in terms of the distribution of LULC classes. In the case of the TCMA LULC classifications presented here, it appears that both the aerial imagery used in this study and the Landsat TM imagery used in Yuan's 2006 classification, have advantages and disadvantages. When the main focus of a study is a relatively small area, or urban and suburban areas, the use of aerial imagery may be the better choice, while Landsat TM could be more suitable when the main interest lies in the classification of vegetative LULC. Additionally, this comparison also shows that it could be beneficial to conduct classification with both types of imagery, and merge their results, in order to achieve a better overall classification.

5.3. Data Visualization

Due to the high spatial resolution of the data used here, resulting maps are somewhat cluttered and hard to interpret. This makes them less usable. In order to present a map that is easier to interpret, results were aggregated into hexagonal grid maps as shown in Chapter 4.1.1. Additionally, an interactive web mapping application was developed to present the results to users. This mapping application was based on ArcGIS Server Flex API, with the data being served by ArcGIS Server. The application can be used in web browser that supports Adobe Flash.

The web mapping application allows user to view the data at various scales, dependent on their area of interest. It is also possible to extract area statistics for specific

areas defined by the user. Therefore, rather than using units of area like counties and cities and townships, as was done in this study to illustrate the results, users would be able to use the results of this analysis for their own area of interest. This mapping application was developed as a prototype to demonstrate the possibilities involved with interactive web maps. While the prototype allows users to explore and analyze the data generated here, it was not yet made public. This was due to constraints in the setup of Minnesota State University's ArcGIS Server system.

6. Conclusions

6.1. Summary of methods and results

Two sets of data were produced in this study: a general LULC map and an impervious surface map. These maps were summarized to reflect LULC and impervious surface statistics for the entire study area, and for each county and municipality within the study area. A combination of aerial imagery, LiDAR-derived elevation data, road centerline data, and other post-processing products were used to generate these maps. LULC maps were produced using Feature Analyst, an object-based classification tool. The impervious surface maps were produced with a decision tree model generated with the See5 software. Finally, an accuracy assessment was conducted for both maps using 300 simple random points for each map. Reference values obtained by manual assignment were tabulated in an accuracy matrix.

The LULC map indicated that, across the entire study area, the most prevalent LULC class was other vegetation, making up about 38.2% or 1136.1 sq. miles. This was followed by forest, which comprised 26% (771.7 sq. miles) of the entire area. The third-largest class was cropland, with 527.3 sq. miles or 17.7%, urban was ranked fourth with 11.3% or 334.8 sq. miles, and water ranked fifth at 170 sq. miles or 5.7%. The smallest class was bare soil, with 34.6 sq. miles or 1.16%. The spatial distribution of these classes indicates that the most urbanized area is the central part of the study area, while there are also some small urban areas in the perimeter of the study area. Forest was found to be most dense in the south-western and northern parts of the study area, and surrounding the

central urbanized area. Vegetated LULC classes (forest, cropland, other vegetation) were mostly found in the perimeter of the study area. Further, other vegetation was concentrated in the western and eastern parts, while forest was mostly found in the north and south-west, and cropland was mostly identified along the southern edge of the study area. Water and bare soil were generally clustered together in smaller areas all throughout the study area.

It was found that about 14% of the entire study area, or about 421.7 sq. miles, are impervious surface. The spatial distribution of impervious surface generally follows the expected pattern of higher concentration in the most urbanized, central part of the study area, with some small, outlying pockets, and lower concentrations of impervious surface in the perimeter of the study area. When considering that impervious surface is mostly made up of urban land cover and some bare soil, the impervious surface numbers confirm the amount of urban and bare soil found in the LULC maps.

The overall accuracy of the LULC map was 74%, with the best per-class accuracy found in water and urban, while cropland and other vegetation were classified the least accurate. Impervious surface was classified with 95% overall accuracy.

6.2. Limitations of this study and implications for future studies

Any study that relies on remotely sensed data has inherent limitations regarding the accuracy of the final product. This is dependent on the characteristics and quality of the data used. In this study, two separate sets of data were produced. The LULC map had lower overall accuracy compared to the impervious surface map. This may be in part due to the higher number of classes, and likelihood of confusion between similar classes, and in part due to the methodology used. In the future, it may be worthwhile to assess how well the decision tree technique would be able to classify the general LULC classes. When this study was conducted, the decision on which classifier to use for which part of the study was based largely on literature describing experiences with the classifiers and review of descriptions of the individual products. This review led to the conclusion that Feature Analyst would be a good choice for classification of LULC features. Further, decision trees tend to get very complex when they are used with many classes, which is why the binary ISA map was conducted with this classifier. Additionally, the aerial imagery used for this study is part of the National Agriculture Inventory Program, and although provided free of charge, the images for this program are always recorded during the agricultural growing season. It is likely that the use of "leaf off" imagery would have benefited this study because it could increase the accuracy of specifically urban and impervious surface classification.

A further limitation of this study was that neither the training data nor the accuracy samples could be collected in the field. While the highest possible diligence was

used in correctly identifying these samples, there is still some amount of uncertainty introduced.

In terms of change analysis of LULC, while there are some similarities between the classification scheme used here for LULC and that used in previous studies conducted by other researchers, some classes are defined differently, making a direct change analysis of LULC difficult. In this case, using class definitions more similar to previous studies would have likely resulted in lower classification accuracy due to the technique used. Further, the general differences in technique used and particularly spatial resolution of the imagery used would have led to major differences in the results and increased the difficulty of LULC change analysis. The spatial resolution of the data produced in this study is higher than that found for any previous studies in the TCMA. In fact, imagery at this resolution covering large study areas has only recently become available. This type of data presents an advantage when estimating the amoung of ISA, but it appears difficult to utilize the LULC data from this study for change analysis with previous studies, or even to produce similar datasets for earlier years. It should be noted that, even though using high spatial resolution data for LULC and ISA studies for past years is not possible because of lacking data, future studies should be able to make use of similar imagery and then be able to compare their results to those obtained here.

Finally, even though computational resources and speed are ever increasing, handling the large amounts of data needed for this study was a challenging task. Particularly the use of multiple input datasets of high spatial resolution required much preprocessing. The actual LULC classification process in Feature Analyst took up to one

week per attempt. Further, the visualization of these high resolution datasets in easy-to-read maps is challenging. Maps displaying the full dataset for the entire study area are difficult to read and interpret. An attempt was made to improve visualization by using the hexagonal aggregation of results as shown in Chapter 4.1.1. Further, the presentation of the data in an interactive web map allows for better visualization of the data depending on the user's needs. However, these methods are not as clear cut as it would be to use the unaltered resulting map from the analysis.

Bibliography

Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer. 1976. A Land Use And Land Cover Classification System For Use With Remote Sensor Data. In *USGS Professional Paper*, 28.

Anfinson, J. A. 1995. The Secret History of the Mississippi's Earliest Locks and Dams. *Minnesota History* 54:254-267.

Arditi, D., and T. Pulket. 2005. *Predicting the Outcome of Construction Litigation Using Boosted Decision Trees*: American Society of Civil Engineers.

Arnold Jr, C. L., and C. J. Gibbons. 1996. Impervious surface coverage. *Journal of the American Planning Association* 62 (2):243.

Baatz, M., and A. Schäpe. 2000. Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. *International Journal of Remote Sensing* 58 (3-4):12-23.

Baerwald, T. J. 1978. The Emergence of a New" Downtown". *Geographical review*:308-318.

Ban, Y., H. Hu, and R. I. M. 2010. Fusion of Quickbird MS and RADARSAT SAR data for urban land-cover mapping: object-based and knowledge-based approach. *International Journal of Remote Sensing* 31 (6):1391-1410.

Bishop, Y. M. M., P. W. Holland, and S. E. Fienberg. 1975. *Discrete multivariate analysis: theory and practice*. Cambridge Mass.: MIT Press.

Blaschke, T., and J. Strobl. 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GIS-Zeitschrift für Geoinformationssysteme* 14 (6):12-17.

Breiman, L. 1984. *Classification and regression trees*. Belmont, Calif.: Wadsworth International Group.

———. 1996. Bagging predictors. *Machine Learning* 24 (2):123-140.

Bricklemyer, R. S., R. L. Lawrence, P. R. Miller, and N. Battogtokh. 2007. Monitoring and verifying agricultural practices related to soil carbon sequestration with satellite imagery. *Agriculture, Ecosystems & Environment* 118 (1-4):201-210.

Campbell, J. B. 2002. *Introduction to remote sensing*. 3rd ed. ed. London, UK: Taylor & Francis.

Chabaeva, A., D. Civco, and S. Prisloe. 2004. Development Of A Population Density And Land Use Based Regression Model To Calculate The Amount Of Imperviousness, University of Connecticut.

Chen, Y., P. Shi, T. Fung, J. Wang, and X. Li. 2007. Object-oriented classification for urban land cover mapping with ASTER imagery. *International Journal of Remote Sensing* 28 (20):4645-4651.

Civco, D. L., and J. D. Hurd. 1997. Impervious surface mapping for the state of Connecticut. Paper read at American Society for Photogrammetry & Remote Sensing 1997 annual conference, at Seattle, Washington.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20 (1):37-46.

Collinge, S. K. 1996. Ecological consequences of habitat fragmentation: implications for landscape architecture and planning. *Landscape and urban planning* 36 (1):59-77.

Congalton, R. 1991a. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment* 37:35-46.

Congalton, R., and R. A. Mead. 1983. A quantitative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering and Remote Sensing* 49 (1):69-74.

Congalton, R. G. 1988. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing* 54.

———. 1991b. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37 (1):35-46.

Copeck, T., S. Szpakowicz, and N. Japkowicz. 2002. Learning How Best to Summarize. Paper read at Proceedings of the Workshop on Multi-Document Summarization Evaluation of the 2nd Document Understanding Conference at the 4Oth Meeting of the Association for Computational Linguistics, Philadelphia, PA.

Cracknell, A. P. 1998. Synergy in remote sensing - what's in a pixel? *International Journal of Geographical Information Science* 19 (11):2025-2047.

Cutter, S. L., R. Golledge, and W. L. Graf. 2002. The Big Questions in Geography. *The Professional Geographer* 54 (3):305-317.

Dougherty, M., R. L. Dymond, S. J. Goetz, C. A. Jantz, and N. Goulet. 2004. Evaluation of Impervious Surface Estimates in a Rapidly Urbanizing Watershed. *Photogrammetric Engineering and Remote Sensing* 70 (11):1275-1284.

Environmental Protecion Agency. 1994. The Quality of Our Nation's Water: 1992, ed. U. S. E. P. Agency. Washington DC: USEPA Office of Water.

ESRI. 2013. *How Hot Spot Analysis (Getis-Ord Gi*) works* 2013a [cited December 19 2013]. Available from http://resources.arcgis.com/en/help/main/10.1/index.html#/How_Hot_Spot_Analysis_Getis_Ord_Gi_works/005p00000011000000/.

———. 2013. *How Spatial Autocorrelation: Moran's I (Spatial Statistics) works* 2013b [cited December 19 2013]. Available from http://resources.arcgis.com/en/help/main/10.1/index.html#/How_Spatial_Autocorrelation_Global_Moran_s_I_works/005p0000000t000000/.

Fisher, P. 1997. The pixel: a snare and a delusion. *International Journal of Geographical Information Science* 18 (3):679-685.

Foody, G. M., P. M. Atkinson, and J. Wiley. 2002. *Uncertainty in remote sensing and GIS*. Hoboken, NJ.: Wiley Online Library.

Freund, Y., R. Schapire, and N. Abe. 1999. A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence* 14 (771-780):1612.

Freund, Y., and R. E. Schapire. 1996. Experiments with a new boosting algorithm. Paper read at International Conference on Machine Learning.

Fry, J. A., M. J. Coan, C. G. Homer, D. K. Meyer, and J. D. Wickham. 2009. Completion of the National Land Cover Database (NLCD) 1992–2001 Land Cover Change Retrofit Product. In *U.S. Geological Survey Open-File Report 2008–1379*: USGS.

Gahegan, M. 2000. On the Application of Inductive Machine Learning Tools to Geographical Analysis. *Geographical Analysis* 32 (2):113-139.

———. 2003. Is inductive machine learning just another wild goose (or might it lay the golden egg)? *International Journal of Geographical Information Science* 17 (1):69.

Goetz, S. J., R. K. Wright, A. J. Smith, E. Zinecker, and E. Schaub. 2003. IKONOS imagery for resource management: Tree cover, impervious surfaces, and riparian buffer analyses in the mid-Atlantic region. *Remote Sensing of Environment* 88 (1-2):195-208.

Gong, P., and P. J. Howarth. 1990. The use of structural information for improving land-cover classification accuracies at the rural-urban fringe. *Photogrammetric Engineering and Remote Sensing* 56 (1):67-73.

Granzow, M., D. Berrar, W. Dubitzky, A. Schuster, F. J. Azuaje, and R. Eils. 2001. Tumor classification by gene expression profiling: comparison and validation of five clustering methods. *Association for Computing Machinery, Special Interest Group SIGBIO Newsletter* 21 (1):16-22.

Griffin, S., J. Rogan, and D. Runfola. 2011. Application of Spectral and Environmental Variables to Map the Kissimmee Prairie Ecosystem Using Classification Trees. *GIScience & Remote Sensing* 48 (3):299-323.

Hansen, M., R. Dubayah, and R. DeFries. 1996. Classification trees: an alternative to traditional land cover classifiers. *International Journal of Remote Sensing* 17 (5):1075-1081.

Haralick, R. M., K. Shanmugam, and I. H. Dinstein. 1973. Textural Features for Image Classification. *Systems, Man and Cybernetics, IEEE Transactions on* 3 (6):610-621.

Haralick, R. M., and L. G. Shapiro. 1985. Image segmentation techniques. *Computer Vision, Graphics, and Image Processing* 29 (1):100-132.

Harris, P. M., and S. J. Ventura. 1995. The integration of geographic data with remotely sensed imagery to improve classification in an urban area. *Photogrammetric Engineering and Remote Sensing* 61 (8):993-998.

Herold, M., X. Liu, and K. C. Clarke. 2003. Spatial metrics and image texture for mapping urban land use. *Photogrammetric Engineering and Remote Sensing* 69 (9):991-1002.

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

Homer, C., C. Huang, L. Yang, and B. Wylie. 2002. Development Of A Circa 2000 Landcover Database For The United States: US. Geologic Survey.

Huang, C., L. Yang, B. Wylie, and C. Homer. 2001. A Strategy For Estimating Tree Canopy Density Using Landsat 7 Etm+ And High Resolution Images Over Large Areas: US. Geologic Survey.

Jennings, D. B., S. T. Jarnagin, and D. W. Ebert. 2004. A Modeling Approach for Estimating Watershed Impervious Surface Area from National Land Cover Data 92. *Photogrammetric Engineering and Remote Sensing* 70 (11):1295-1307.

Johnson, D. M., and R. Mueller. 2010. The 2009 Cropland Data Layer. *Photogrammetric Engineering & Remote Sensing* (November 2010):1202 - 1205.

Kontoes, C., G. G. Wilkinson, A. Burrill, S. Goffredo, and J. Mégier. 1993. An experimental system for the integration of GIS data in knowledge-based image analysis for remote sensing of agriculture. *International journal of geographical information systems* 7 (3):247-262.

Kuo, W.-J., R.-F. Chang, W. K. Moon, C. C. Lee, and D.-R. Chen. 2002. Computer-Aided Diagnosis of Breast Tumors with Different US Systems. *Academic Radiology* 9 (7):793-799.

Lang, S., and T. Blaschke. 2006. Bridging remote sensing and GIS–What are the main supportive pillars? Paper read at Proceedings of the 1st International Conference on Object-based Image Analysis.

Lu, D., and Q. Weng. 2006. Use of impervious surface in urban land-use classification. *Remote Sensing of Environment* 102 (1-2):146-160.

Lua, D., S. Hetricka, and E. Moran. 2011. Impervious surface mapping with Quickbird imagery. *International Journal of Remote Sensing* 32 (9):2519-2533.

Lucas, R., A. Rowlands, A. Brown, S. Keyworth, and P. Bunting. 2007. Rule-based classification of multi-temporal satellite imagery for habitat and agricultural land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing* 62 (3):165-185.

McNairn, H., J. Shang, C. Champagne, E. Huffman, A. Smith, and T. Fisette. 2005. A multi-sensor Approach to inventorying Agricultural land use. Paper read at Proceedings of the 31st International Symposium on Remote Sensing of the Environment, Petersburg, Russia.

Mesev, V. 1998. The Use of Census Data in Urban Image Classification. *Photogrammetric Engineering and Remote Sensing* 64 (5):431-438.

Metropolitan Council. 2013. *Twin Cities Metropolitan Area Geographic Definitions* 2000 [cited February 19 2013]. Available from http://www.metrocouncil.org/census/KeyFacts/MetropolitanAreaDefinitions.pdf.

Miller, J. E., S. A. C. Nelson, and G. R. Hess. 2009. An Object Extraction Approach for Impervious Surface Classification with Very-High-Resolution Imagery. *Professional Geographer* 61 (2):250-264.

Mitchell, T. M. 1997. Machine learning. 1997. *Burr Ridge, IL: McGraw Hill* 45.

Myint, S. W. 2001. A Robust Texture Analysis and Classification Approach for Urban Land-Use and Land-Cover Feature Discrimination. *Geocarto International* 16 (4):29-40.

Opitz, D., and W. Bain. 1999. Experiments on learning to extract features from digital images. *International Association of Science and Technology for Development: Signal and Image Processing Conference*.

Opitz, D., and S. Blundell. 2008. Object recognition and image segmentation: the Feature Analyst® approach. In *Object-Based Image Analysis*, eds. T. Blaschke, S. Lang and G. Hay, 153-167: Springer Berlin Heidelberg.

Opitz, D. W. 1999. Feature selection for ensembles. Paper read at Innovative Applications of Artificial Intelligence Conference.

Opitz, D. W., and J. W. Shavlik. 1996a. Actively searching for an effective neural network ensemble. *Connection Science* 8 (3-4):337-354.

———. 1996b. Generating accurate and diverse members of a neural-network ensemble. *Advances in neural information processing systems*:535-541.

Pal, M., and P. M. Mather. 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment* 86 (4):554-565.

Quinlan, J. R. 1993. *C4.5: programs for machine learning*. San Mateo, CA: Morgan Kaufmann.

———. 2013. *Information on See5/C5.0*. Rulequest 2013a [cited November 11 2013]. Available from http://www.rulequest.com/see5-info.html.

———. 2013. *Is C5.0 Better Than C4.5?* 2013b [cited November 11 2013]. Available from http://rulequest.com/see5-comparison.html.

Ridd, M. K. 1995. Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities. *International Journal of Remote Sensing* 16 (12):2165-2185.

Rignot, E., W. A. Salas, and D. L. Skole. 1997. Mapping deforestation and secondary growth in Rondonia, Brazil, using imaging radar and thematic mapper data. *Remote Sensing of Environment* 59 (2):167-179.

Rogan, J., J. Franklin, D. Stow, J. Miller, C. Woodcock, and D. Roberts. 2008. Mapping land-cover modifications over large areas: A comparison of machine learning algorithms. *Remote Sensing of Environment* 112 (5):2272-2283.

Rogan, J., J. Miller, D. Stow, J. Franklin, L. Levien, and C. Fischer. 2003. Land-Cover Change Monitoring with Classification Trees Using Landsat TM and Ancillary Data. *Photogrammetric Engineering and Remote Sensing* 69 (7):793-804.

Rumelhart, D. E., G. E. Hintont, and R. J. Williams. 1986. Learning representations by back-propagating errors. *Nature* 323 (6088):533-536.

Ryherd, S., and C. Woodcock. 1996. Combining spectral and texture data in the segmentation of remotely sensed images. *Photogrammetric Engineering and Remote Sensing* 62 (2):181-194.

Saatchi, S. S., J. V. Soares, and D. S. Alves. 1997. Mapping deforestation and land use in amazon rainforest by using SIR-C imagery. *Remote Sensing of Environment* 59 (2):191-202.

Schapire, R. E. 1999. A brief introduction to boosting. Paper read at International Joint Conferences on Artificial Intelligence, at Stockholm, Sweden.

Schueler, T. R. 1994. The importance of imperviousness. *Watershed Protection Techniques* 1 (3).

Sen, S., and K. Hernandez. 2000. A buyer's agent. In *Proceedings of the fourth international conference on Autonomous agents*, 156-162. Barcelona, Spain: ACM.

Silva, Á., J. Pereira, M. Santos, L. Gomes, and J. Neves. 2003. Organ Failure Prediction based on Clinical Adverse Events: A Cluster Model Approach. Paper read at Artificial Intelligence and Applications (AIA 2003), at Benalmádena, Spain

Smith, A. J. 2000. Subpixel Estimates of Impervious Surface Cover Using Landsat TM Imagery, Geography Department, University of Maryland, College Park, College Park.

Squires, G. D. 2002. *Urban sprawl: Causes, consequences, and policy responses*: The Urban Insitute.

Stocker, J. 1998. Methods for Measuring and Estimating Impervious Surface Coverage, 3: Nonpoint Education for Municipal Officials (NEMO).

Sutton, P. C., S. J. Anderson, C. D. Elvidge, B. T. Tuttle, and T. Ghosh. 2009. Paving the planet: impervious surface as proxy measure of the human ecological footprint. *Progress in Physical Geography* 33 (4):510-527.

Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography* 46:234-240.

Tseng, C.-C. 2003. Comparing Artificial Intelligence Systems for Stock Portfolio Selection: Society for Computational Economics.

UMN. 2014. *Minnesota Land Economics* 2013 [cited February 17 2014]. Available from http://www.landeconomics.umn.edu/mle/default.aspx.

US Census. 2013. *American Factfinder* 2013 [cited February 19 2013]. Available from http://factfinder2.census.gov.

Van De Voorde, T., T. De Roeck, and F. Canters. 2009. A comparison of two spectral mixture modelling approaches for impervious surface mapping in urban areas. *International Journal of Remote Sensing* 30 (18):4785-4806.

Wright, C., and A. Gallant. 2007. Improved wetland remote sensing in Yellowstone National Park using classification trees to combine TM imagery and ancillary environmental data. *Remote Sensing of Environment* 107 (4):582-605.

Wu, X., and V. Kumar. 2009. *The top ten algorithms in data mining*: CRC Press.

Yang, L., C. Huang, C. G. Homer, B. K. Wylie, and M. J. Coan. 2003. An approach for mapping large-area impervious surfaces: synergistic use of Landsat-7 ETM+ and high spatial resolution imagery. *Canadian Journal of Remote Sensing* 29 (2):230-240.

Yang, L., G. Xian, J. M. Klaver, and B. Deal. 2003. Urban Land-Cover Change Detection through Sub-Pixel Imperviousness Mapping Using Remotely Sensed Data. *Photogrammetric Engineering and Remote Sensing* 69 (9):1003 - 1010.

Yuan, F. 2008. Land-cover change and environmental impact analysis in the Greater Mankato area of Minnesota using remote sensing and GIS modelling. *International Journal of Remote Sensing* 29 (4):1169-1184.

Yuan, F. 2009. Urban growth monitoring and projection using remote sensing and geographic information systems: a case study in the Twin Cities Metropolitan Area, Minnesota. *Geocarto International* 25 (3):213-230.

———. 2010. Urban growth monitoring and projection using remote sensing and geographic information systems: a case study in the Twin Cities metropolitan area, Minnesota. *Geocarto International* 25 (3):213-230.

Zhang, Q., and J. Wang. 2003. A rule-based urban land use inferring method for fine-resolution multispectral imagery. *Canadian Journal of Remote Sensing* 29 (1):1-13.

Zhang, Y. 1999. Optimisation of building detection in satellite images by combining multispectral classification and texture filtering. *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (1):50-60.

Appendices

Appendix A. Global Moran's I Equation

The Moran's I statistic for spatial autocorrelation is given as:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2} \quad (1)$$

where z_i is the deviation of an attribute for feature i from its mean $(x_i - \bar{X})$, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features, and S_0 is the aggregate of all the spatial weights:

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \quad (2)$$

The z_I-score for the statistic is computed as:

$$z_I = \frac{I - \mathrm{E}[I]}{\sqrt{\mathrm{V}[I]}} \quad (3)$$

where:

$$\mathrm{E}[I] = -1/(n-1) \quad (4)$$

$$\mathrm{V}[I] = \mathrm{E}[I^2] - \mathrm{E}[I]^2 \quad (5)$$

Adopted from ESRI (2013b).

Appendix B. Getis-Ord G* Equation

The Getis-Ord local statistic is given as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}} \quad (1)$$

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \quad (2)$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \left(\bar{X}\right)^2} \quad (3)$$

The G_i^* statistic is a z-score so no further calculations are required.

Adopted from ESRI (2013a).

Appendix C. LULC by city or township

City or Township	Water		Forest		Cropland		Other Vegetation		Urban		Bare Soil		TOTAL Sq. Mi.
	Sq. Mi.	% of total mun.	Sq. Mi.	% of total mun.	Sq. Mi.	% of total mun.	Sq. Mi.	% of total mun.	Sq. Mi.	% of total mun.	Sq. Mi.	% of total mun.	
Afton	1.24	4.71%	4.89	18.60%	5.63	21.40%	12.99	49.38%	1.33	5.06%	0.22	0.85%	26.31
Andover	0.65	1.86%	14.51	41.65%	3.34	9.59%	12.63	36.26%	3.29	9.44%	0.42	1.21%	34.83
Anoka	0.40	5.67%	2.30	32.40%	0.56	7.85%	1.93	27.17%	1.68	23.62%	0.23	3.29%	7.11
Apple Valley	0.79	4.51%	6.18	35.13%	0.92	5.21%	4.24	24.10%	5.03	28.55%	0.44	2.50%	17.60
Arden Hills	0.92	9.53%	2.63	27.21%	0.31	3.25%	3.72	38.50%	1.98	20.50%	0.10	1.01%	9.66
Bayport	0.16	8.66%	0.21	11.70%	0.38	20.98%	0.71	39.19%	0.29	15.82%	0.07	3.65%	1.81
Baytown	1.03	11.26%	1.67	18.35%	1.45	15.92%	4.55	49.82%	0.32	3.48%	0.11	1.18%	9.12
Belle Plaine	0.62	1.42%	3.91	8.92%	18.96	43.26%	17.97	41.00%	1.93	4.40%	0.44	1.00%	43.83
Benton	0.63	1.86%	3.78	11.12%	10.16	29.92%	18.37	54.11%	0.79	2.34%	0.22	0.65%	33.95
Bethel	0.04	4.23%	0.39	37.78%	0.05	5.26%	0.46	43.90%	0.07	7.06%	0.02	1.77%	1.04
Birchwood Village	0.00	1.27%	0.25	72.26%	0.00	0.89%	0.04	12.90%	0.04	12.21%	0.00	0.48%	0.35
Blaine	1.09	3.20%	12.13	35.65%	1.16	3.42%	12.12	35.64%	7.28	21.41%	0.23	0.68%	34.02
Blakely	0.67	2.51%	4.34	16.16%	5.93	22.08%	15.15	56.38%	0.33	1.22%	0.44	1.65%	26.87
Bloomington	1.63	4.25%	16.45	42.91%	2.12	5.52%	8.69	22.66%	9.02	23.51%	0.44	1.14%	38.34
Brooklyn Center	0.39	4.70%	2.66	31.75%	0.43	5.13%	2.17	25.87%	2.50	29.87%	0.22	2.68%	8.38
Brooklyn Park	0.71	2.66%	7.82	29.39%	3.12	11.73%	7.71	28.96%	6.49	24.39%	0.76	2.86%	26.61
Burnsville	2.14	7.97%	10.54	39.18%	1.45	5.38%	5.64	20.98%	6.35	23.63%	0.77	2.85%	26.89
Camden	0.44	1.29%	5.68	16.53%	10.83	31.51%	16.54	48.14%	0.73	2.12%	0.14	0.41%	34.36
Carver	0.44	10.11%	0.90	20.82%	0.53	12.35%	2.09	48.25%	0.34	7.86%	0.03	0.62%	4.32
Castle Rock	0.41	1.15%	7.36	20.84%	8.64	24.46%	18.08	51.18%	0.53	1.50%	0.31	0.88%	35.33
Cedar Lake	1.37	3.78%	7.95	21.97%	15.69	43.32%	9.95	27.49%	0.75	2.06%	0.50	1.38%	36.21
Centerville	0.32	13.01%	0.79	32.52%	0.08	3.19%	0.74	30.51%	0.50	20.53%	0.01	0.25%	2.43
Champlin	0.59	6.78%	3.52	40.16%	0.74	8.49%	2.11	24.08%	1.64	18.75%	0.15	1.75%	8.75
Chanhassen	2.37	10.33%	4.96	21.66%	0.93	4.06%	10.75	46.91%	3.64	15.90%	0.26	1.13%	22.91

Chaska	0.81	4.59%	3.64	20.63%	1.86	10.52%	8.53	48.29%	2.57	14.56%	0.25	1.41%	17.66
Circle Pines	0.17	8.76%	0.76	39.19%	0.02	1.20%	0.59	30.33%	0.39	19.97%	0.01	0.55%	1.94
Coates	0.00	0.10%	0.08	5.50%	0.54	39.00%	0.59	42.83%	0.17	12.04%	0.01	0.54%	1.38
Cologne	0.16	8.86%	0.18	9.79%	0.72	40.15%	0.52	28.85%	0.19	10.88%	0.03	1.46%	1.79
Columbia Heights	0.12	3.34%	1.39	39.89%	0.10	2.91%	0.65	18.61%	1.20	34.23%	0.04	1.03%	3.50
Columbus	1.77	3.72%	14.28	30.00%	1.52	3.19%	27.05	56.80%	2.83	5.95%	0.17	0.35%	47.62
Coon Rapids	0.99	4.25%	8.93	38.37%	1.09	4.66%	6.41	27.52%	5.43	23.32%	0.44	1.88%	23.29
Corcoran	0.20	0.56%	5.90	16.36%	6.73	18.67%	21.81	60.50%	1.02	2.82%	0.39	1.08%	36.05
Cottage Grove	4.12	10.99%	10.08	26.88%	6.24	16.64%	12.34	32.93%	4.12	11.00%	0.59	1.57%	37.49
Credit River	0.56	2.36%	6.35	26.61%	7.76	32.53%	8.09	33.89%	0.77	3.22%	0.33	1.40%	23.86
Crystal	0.12	2.08%	2.10	35.84%	0.44	7.52%	1.66	28.31%	1.41	24.10%	0.13	2.15%	5.85
Dahlgren	0.56	1.58%	4.99	14.03%	11.99	33.67%	16.08	45.18%	1.64	4.62%	0.33	0.93%	35.60
Dayton	1.65	6.56%	6.15	24.43%	6.20	24.62%	10.03	39.87%	0.83	3.29%	0.31	1.24%	25.16
Deephaven	0.03	1.17%	0.98	41.36%	0.02	0.67%	0.93	39.07%	0.41	17.10%	0.01	0.63%	2.37
Dellwood	0.11	3.72%	1.10	38.88%	0.02	0.78%	1.20	42.44%	0.39	13.66%	0.01	0.52%	2.82
Denmark	1.96	6.48%	3.80	12.59%	7.61	25.19%	15.91	52.66%	0.59	1.97%	0.33	1.10%	30.21
Douglas	0.11	0.31%	6.40	18.81%	18.12	53.25%	8.72	25.63%	0.17	0.50%	0.51	1.50%	34.02
Eagan	1.47	4.39%	11.94	35.68%	0.69	2.07%	8.92	26.65%	10.18	30.43%	0.26	0.78%	33.47
East Bethel	2.35	4.92%	16.27	34.13%	3.00	6.30%	22.97	48.18%	2.71	5.68%	0.38	0.79%	47.68
Eden Prairie	2.71	7.70%	11.50	32.71%	1.45	4.13%	11.45	32.56%	7.55	21.48%	0.50	1.42%	35.16
Edina	0.48	2.98%	7.42	46.43%	0.81	5.08%	3.33	20.87%	3.70	23.13%	0.24	1.51%	15.98
Elko New Market	0.08	2.52%	0.47	14.25%	1.31	39.97%	0.94	28.83%	0.41	12.45%	0.06	1.98%	3.27
Empire	0.69	2.20%	4.64	14.71%	8.04	25.50%	16.64	52.78%	1.13	3.58%	0.39	1.23%	31.52
Eureka	0.45	1.25%	5.44	15.26%	19.25	54.01%	9.52	26.72%	0.61	1.71%	0.37	1.04%	35.63
Excelsior	0.04	6.09%	0.08	12.30%	0.01	1.72%	0.23	34.67%	0.30	44.23%	0.01	0.99%	0.68
Falcon Heights	0.02	1.10%	0.68	30.38%	0.18	8.11%	0.61	27.49%	0.71	31.89%	0.02	1.02%	2.24
Farmington	0.85	5.74%	1.95	13.13%	4.90	33.02%	5.38	36.23%	1.67	11.24%	0.09	0.64%	14.85
Forest Lake	4.62	13.02%	10.70	30.11%	1.90	5.34%	15.09	42.48%	2.93	8.26%	0.28	0.79%	35.52
Fort Snelling	0.14	2.03%	0.90	13.49%	0.12	1.86%	2.28	34.04%	3.19	47.64%	0.06	0.94%	6.69
Fridley	0.76	7.05%	3.65	33.82%	0.59	5.50%	2.14	19.77%	3.37	31.22%	0.29	2.64%	10.81
Gem Lake	0.04	3.36%	0.55	49.80%	0.00	0.41%	0.35	32.29%	0.15	14.09%	0.00	0.05%	1.10
Golden Valley	0.37	3.48%	4.21	39.93%	0.53	5.02%	2.67	25.36%	2.57	24.36%	0.19	1.85%	10.55

Grant	0.83	3.12%	9.38	35.45%	3.49	13.20%	11.87	44.87%	0.64	2.42%	0.25	0.95%	26.46
Greenfield	0.73	3.40%	4.77	22.27%	2.86	13.34%	11.89	55.55%	0.84	3.91%	0.33	1.53%	21.41
Greenvale	0.09	0.30%	5.64	19.79%	13.62	47.74%	8.22	28.83%	0.58	2.05%	0.37	1.30%	28.52
Greenwood	0.01	1.65%	0.13	37.14%	0.00	0.55%	0.12	34.83%	0.09	25.17%	0.00	0.67%	0.35
Grey Cloud Island	0.76	24.62%	1.30	42.32%	0.14	4.51%	0.58	19.02%	0.15	4.87%	0.14	4.65%	3.07
Ham Lake	1.47	4.10%	14.88	41.52%	1.33	3.71%	14.52	40.51%	3.44	9.59%	0.20	0.57%	35.84
Hamburg	0.00	0.00%	0.07	35.27%	0.03	14.55%	0.06	27.55%	0.04	18.72%	0.01	3.91%	0.20
Hampton	0.18	0.51%	7.03	19.69%	16.07	45.04%	11.58	32.45%	0.47	1.33%	0.35	0.99%	35.69
Hancock	0.05	0.27%	1.55	8.74%	7.29	41.05%	8.51	47.93%	0.19	1.08%	0.16	0.93%	17.75
Hanover	0.02	0.86%	0.82	38.72%	0.13	5.94%	0.98	46.57%	0.13	6.38%	0.03	1.54%	2.11
Hastings	0.92	8.19%	2.20	19.54%	1.49	13.24%	4.52	40.21%	2.01	17.86%	0.11	0.95%	11.25
Helena	1.24	3.65%	6.79	19.94%	16.24	47.68%	8.70	25.56%	0.79	2.33%	0.28	0.83%	34.05
Hilltop	0.00	1.83%	0.02	17.40%	0.00	2.35%	0.03	21.28%	0.07	55.83%	0.00	1.30%	0.13
Hollywood	0.13	0.37%	4.73	13.25%	7.57	21.20%	22.57	63.18%	0.59	1.65%	0.13	0.36%	35.72
Hopkins	0.03	0.82%	1.47	35.87%	0.18	4.43%	0.92	22.39%	1.47	35.95%	0.02	0.54%	4.10
Hugo	1.97	5.48%	10.95	30.43%	3.41	9.48%	17.33	48.16%	2.07	5.76%	0.25	0.69%	35.98
Independence	1.79	5.17%	8.39	24.28%	3.01	8.72%	19.85	57.46%	1.25	3.62%	0.26	0.75%	34.55
Inver Grove Heights	1.45	4.82%	13.11	43.55%	0.67	2.22%	8.93	29.65%	5.70	18.94%	0.25	0.82%	30.12
Jackson	0.19	3.03%	1.03	16.32%	1.04	16.36%	3.49	55.04%	0.48	7.53%	0.11	1.72%	6.34
Jordan	0.14	4.33%	0.57	17.77%	0.75	23.29%	0.92	28.81%	0.77	24.02%	0.06	1.79%	3.20
Lake Elmo	1.64	6.75%	6.98	28.78%	4.01	16.56%	10.24	42.26%	0.99	4.08%	0.38	1.57%	24.24
Lake Saint Croix Beach	0.43	44.17%	0.11	11.45%	0.05	4.68%	0.21	21.28%	0.16	16.12%	0.02	2.31%	0.98
Lakeland	0.89	30.27%	0.74	25.08%	0.32	10.89%	0.65	22.02%	0.25	8.46%	0.10	3.27%	2.94
Lakeland Shores	0.41	57.65%	0.09	12.37%	0.04	5.91%	0.11	15.70%	0.05	7.39%	0.01	0.99%	0.72
Laketown	2.67	9.56%	4.99	17.89%	3.71	13.31%	15.48	55.46%	0.87	3.11%	0.19	0.68%	27.91
Lakeville	1.90	5.03%	9.97	26.34%	8.93	23.58%	10.48	27.68%	5.93	15.66%	0.64	1.70%	37.86
Landfall	0.02	21.25%	0.02	26.13%	0.00	0.46%	0.01	7.27%	0.04	44.74%	0.00	0.15%	0.08
Lauderdale	0.01	1.82%	0.14	33.94%	0.01	2.60%	0.10	22.60%	0.16	36.97%	0.01	2.07%	0.42
Lexington	0.00	0.71%	0.28	40.97%	0.01	1.35%	0.19	27.52%	0.20	28.97%	0.00	0.48%	0.69
Lilydale	0.20	23.51%	0.28	32.84%	0.00	0.56%	0.21	24.74%	0.15	18.27%	0.00	0.09%	0.84
Lino Lakes	3.77	11.36%	9.35	28.17%	1.20	3.63%	14.73	44.40%	3.96	11.94%	0.16	0.50%	33.18
Linwood	1.85	5.13%	14.26	39.51%	0.97	2.70%	17.16	47.56%	1.71	4.74%	0.13	0.36%	36.09

Little Canada	0.45	9.96%	1.26	28.03%	0.06	1.29%	1.12	24.80%	1.59	35.29%	0.03	0.62%	4.50
Long Lake	0.09	9.38%	0.12	12.62%	0.02	1.83%	0.35	37.34%	0.35	37.20%	0.02	1.63%	0.94
Loretto	0.00	0.40%	0.06	24.35%	0.01	2.71%	0.09	35.89%	0.09	32.89%	0.01	3.75%	0.26
Louisville	1.18	8.10%	2.67	18.37%	3.16	21.76%	6.48	44.64%	0.57	3.92%	0.46	3.20%	14.52
Mahtomedi	1.96	34.46%	1.78	31.35%	0.07	1.15%	0.80	14.12%	1.03	18.16%	0.04	0.76%	5.69
Maple Grove	2.33	6.66%	10.02	28.62%	2.78	7.94%	11.16	31.86%	7.41	21.15%	1.32	3.76%	35.02
Maple Plain	0.00	0.00%	0.36	33.64%	0.03	2.43%	0.38	35.89%	0.27	25.04%	0.03	3.00%	1.07
Maplewood	0.60	3.31%	7.71	42.91%	0.30	1.66%	4.32	24.07%	4.94	27.51%	0.10	0.54%	17.97
Marine on Saint Croix	0.00	0.05%	0.78	20.35%	2.01	52.56%	0.94	24.50%	0.07	1.91%	0.02	0.63%	3.82
Marshan	0.11	0.33%	6.01	17.49%	17.96	52.27%	9.58	27.89%	0.44	1.28%	0.26	0.75%	34.37
May	2.78	7.28%	8.81	23.10%	9.95	26.07%	15.75	41.26%	0.46	1.20%	0.42	1.09%	38.16
Mayer	0.02	1.07%	0.27	18.62%	0.03	2.18%	0.86	60.10%	0.24	16.42%	0.02	1.61%	1.43
Medicine Lake	0.00	1.87%	0.08	47.22%	0.01	3.37%	0.06	34.86%	0.02	12.17%	0.00	0.50%	0.18
Medina	1.44	5.30%	7.53	27.75%	2.54	9.37%	13.95	51.42%	1.47	5.43%	0.20	0.73%	27.13
Mendota	0.01	5.43%	0.10	45.73%	0.00	1.14%	0.06	25.03%	0.05	22.35%	0.00	0.32%	0.22
Mendota Heights	0.49	4.82%	3.70	36.55%	0.17	1.67%	3.35	33.03%	2.38	23.45%	0.05	0.49%	10.13
Miesville	0.02	1.09%	0.12	7.12%	0.73	42.11%	0.83	47.41%	0.02	1.16%	0.02	1.11%	1.74
Minneapolis	3.90	6.70%	17.47	29.98%	1.50	2.58%	11.00	18.88%	23.66	40.61%	0.73	1.26%	58.27
Minnetonka	1.95	6.70%	12.46	42.87%	0.60	2.06%	8.13	27.97%	5.62	19.31%	0.32	1.09%	29.08
Minnetonka Beach	0.01	2.53%	0.11	22.94%	0.01	1.51%	0.24	50.71%	0.10	21.14%	0.01	1.18%	0.47
Minnetrista	5.34	16.72%	6.75	21.12%	1.78	5.58%	15.97	49.99%	1.94	6.06%	0.17	0.53%	31.95
Mound	0.68	18.97%	0.91	25.45%	0.03	0.77%	0.95	26.56%	0.97	27.06%	0.04	1.18%	3.59
Mounds View	0.09	2.30%	1.98	48.39%	0.11	2.68%	0.89	21.86%	1.00	24.49%	0.01	0.28%	4.08
New Brighton	0.58	8.19%	2.75	39.09%	0.21	2.98%	1.45	20.58%	2.02	28.68%	0.03	0.49%	7.03
New Germany	0.04	4.00%	0.18	17.39%	0.18	17.33%	0.55	54.78%	0.06	5.60%	0.01	0.90%	1.01
New Hope	0.06	1.22%	1.59	31.07%	0.16	3.18%	1.50	29.36%	1.71	33.48%	0.09	1.68%	5.12
New Market	0.42	1.28%	6.87	21.09%	15.76	48.35%	8.17	25.07%	0.78	2.40%	0.59	1.80%	32.59
New Prague	0.04	1.67%	0.21	9.56%	0.70	31.46%	0.61	27.17%	0.62	27.77%	0.05	2.38%	2.24
New Trier	0.00	1.66%	0.04	20.35%	0.07	36.17%	0.07	38.46%	0.00	1.51%	0.00	1.86%	0.18
Newport	0.25	6.36%	1.88	48.64%	0.07	1.84%	0.75	19.49%	0.86	22.24%	0.06	1.44%	3.86
Nininger	3.82	22.37%	3.41	19.93%	3.98	23.26%	5.36	31.35%	0.40	2.36%	0.12	0.73%	17.10
North Oaks	1.26	14.59%	3.66	42.49%	0.05	0.56%	2.56	29.71%	1.07	12.43%	0.02	0.22%	8.61

North Saint Paul	0.16	5.25%	1.24	40.94%	0.04	1.33%	0.53	17.50%	1.03	34.12%	0.03	0.85%	3.03
Northfield	0.03	1.72%	0.27	18.38%	0.76	51.62%	0.29	19.85%	0.12	7.89%	0.01	0.54%	1.48
Norwood Young America	0.00	0.20%	0.52	22.92%	0.12	5.15%	1.17	51.26%	0.41	18.11%	0.05	2.37%	2.28
Nowthen	1.09	3.09%	11.05	31.43%	6.43	18.31%	15.33	43.61%	0.77	2.19%	0.48	1.38%	35.15
Oak Grove	1.16	3.30%	12.25	34.80%	4.79	13.60%	15.27	43.37%	1.33	3.77%	0.41	1.16%	35.20
Oak Park Heights	0.60	18.28%	0.30	9.19%	0.56	17.09%	1.18	35.74%	0.52	15.63%	0.13	4.07%	3.30
Oakdale	0.36	3.15%	4.32	38.21%	0.33	2.95%	3.18	28.12%	3.04	26.87%	0.08	0.70%	11.30
Orono	8.77	34.64%	3.42	13.49%	0.29	1.13%	10.56	41.70%	2.13	8.40%	0.16	0.63%	25.32
Osseo	0.00	0.27%	0.17	23.72%	0.02	2.48%	0.15	21.24%	0.36	49.52%	0.02	2.77%	0.72
Pine Springs	0.12	12.74%	0.42	46.04%	0.01	1.33%	0.27	29.01%	0.10	10.36%	0.00	0.52%	0.92
Plymouth	2.60	7.34%	11.14	31.41%	1.45	4.09%	11.22	31.63%	8.44	23.79%	0.61	1.73%	35.47
Prior Lake	2.64	14.36%	5.33	28.96%	2.03	11.04%	5.50	29.87%	2.52	13.70%	0.38	2.07%	18.40
Ramsey	0.61	2.06%	10.23	34.38%	3.39	11.40%	12.27	41.23%	2.74	9.20%	0.52	1.74%	29.75
Randolph	1.19	10.40%	1.18	10.29%	4.58	39.99%	4.04	35.26%	0.32	2.82%	0.14	1.24%	11.46
Ravenna	1.47	6.77%	5.23	23.98%	7.35	33.69%	7.40	33.96%	0.21	0.97%	0.14	0.63%	21.80
Richfield	0.17	2.34%	2.32	32.58%	0.20	2.78%	1.78	25.04%	2.56	36.00%	0.09	1.28%	7.12
Robbinsdale	0.20	6.62%	1.05	35.04%	0.16	5.46%	0.72	23.98%	0.78	26.17%	0.08	2.73%	2.98
Rockford	0.00	0.39%	0.06	22.43%	0.01	2.85%	0.10	37.80%	0.07	29.03%	0.02	7.51%	0.25
Rogers	0.57	2.16%	4.50	17.19%	3.12	11.90%	14.95	57.12%	2.78	10.63%	0.26	1.01%	26.18
Rosemount	1.43	4.06%	8.38	23.79%	5.14	14.59%	15.05	42.73%	4.77	13.55%	0.45	1.29%	35.22
Roseville	0.61	4.37%	5.08	36.64%	0.20	1.45%	2.99	21.62%	4.89	35.27%	0.09	0.65%	13.85
Saint Anthony	0.13	5.67%	0.92	39.31%	0.05	1.96%	0.52	22.06%	0.72	30.59%	0.01	0.40%	2.34
Saint Bonifacius	0.00	0.18%	0.27	25.88%	0.01	1.11%	0.45	43.11%	0.29	28.29%	0.01	1.42%	1.04
Saint Francis	0.22	0.96%	9.53	40.82%	3.52	15.06%	8.97	38.42%	0.84	3.61%	0.27	1.14%	23.36
Saint Lawrence	0.65	4.50%	2.53	17.37%	5.70	39.22%	5.01	34.47%	0.42	2.86%	0.23	1.57%	14.54
Saint Louis Park	0.29	2.64%	3.81	35.22%	0.61	5.64%	2.73	25.20%	3.25	30.01%	0.14	1.28%	10.83
Saint Marys Point	0.01	3.67%	0.11	28.35%	0.02	5.66%	0.16	39.90%	0.08	19.56%	0.01	2.85%	0.39
Saint Paul	3.52	6.28%	17.92	31.99%	0.69	1.24%	10.26	18.31%	23.36	41.68%	0.28	0.50%	56.03
Saint Paul Park	0.32	8.85%	1.30	36.36%	0.18	4.96%	0.89	24.99%	0.81	22.60%	0.08	2.25%	3.57
San Francisco	1.33	5.51%	5.40	22.42%	7.73	32.06%	8.86	36.77%	0.49	2.02%	0.29	1.21%	24.09
Sand Creek	1.02	3.14%	6.41	19.75%	9.53	29.37%	14.28	44.00%	0.91	2.82%	0.30	0.92%	32.46

Savage	0.56	3.39%	5.72	34.83%	1.72	10.49%	5.06	30.85%	2.66	16.20%	0.70	4.25%	16.42
Scandia	2.97	7.46%	10.17	25.52%	15.23	38.23%	10.19	25.56%	0.76	1.92%	0.52	1.31%	39.85
Sciota	0.19	1.29%	1.68	11.29%	4.46	30.02%	8.18	55.06%	0.22	1.51%	0.12	0.83%	14.86
Shakopee	1.16	3.95%	6.94	23.61%	4.12	14.01%	10.33	35.13%	6.02	20.49%	0.82	2.80%	29.39
Shoreview	1.42	11.24%	4.52	35.70%	0.15	1.17%	3.53	27.93%	2.97	23.48%	0.06	0.49%	12.65
Shorewood	7.38	56.94%	1.53	11.77%	0.07	0.55%	2.75	21.21%	1.18	9.09%	0.06	0.45%	12.96
South Saint Paul	0.38	6.22%	1.87	30.63%	0.08	1.39%	1.25	20.45%	2.47	40.51%	0.05	0.79%	6.09
Spring Lake	1.73	5.85%	5.62	19.01%	9.60	32.46%	11.48	38.83%	0.72	2.44%	0.42	1.41%	29.56
Spring Lake Park	0.07	3.59%	0.71	33.91%	0.10	4.59%	0.49	23.58%	0.69	33.34%	0.02	0.99%	2.08
Spring Park	0.05	12.76%	0.06	15.45%	0.00	1.20%	0.11	27.30%	0.18	42.61%	0.00	0.68%	0.41
Stillwater	2.08	8.49%	5.57	22.74%	3.71	15.16%	11.29	46.14%	1.46	5.97%	0.37	1.50%	24.48
Sunfish Lake	0.10	5.75%	0.89	52.56%	0.01	0.48%	0.50	29.42%	0.20	11.59%	0.00	0.20%	1.69
Tonka Bay	0.01	1.06%	0.11	11.09%	0.02	2.17%	0.54	55.69%	0.28	29.07%	0.01	0.92%	0.97
Vadnais Heights	0.92	11.19%	2.61	31.67%	0.07	0.87%	2.31	28.08%	2.28	27.65%	0.04	0.54%	8.24
Vermillion	0.22	0.62%	5.72	16.26%	14.73	41.85%	13.48	38.30%	0.78	2.21%	0.27	0.76%	35.21
Victoria	1.62	15.38%	2.75	26.18%	0.19	1.85%	4.82	45.86%	1.00	9.50%	0.13	1.24%	10.52
Waconia	5.29	14.98%	4.85	13.73%	5.90	16.71%	17.00	48.11%	2.11	5.96%	0.18	0.51%	35.33
Waterford	0.51	3.44%	2.85	19.34%	3.38	22.94%	7.44	50.51%	0.45	3.06%	0.10	0.71%	14.73
Watertown	1.85	5.20%	6.62	18.58%	5.28	14.82%	20.40	57.28%	1.32	3.70%	0.15	0.42%	35.61
Wayzata	0.07	2.35%	0.83	26.50%	0.05	1.57%	1.25	39.87%	0.87	27.89%	0.06	1.81%	3.14
West Lakeland	0.38	2.99%	3.52	27.81%	2.06	16.31%	5.89	46.51%	0.52	4.14%	0.28	2.24%	12.66
West Saint Paul	0.04	0.73%	1.75	34.92%	0.06	1.20%	1.20	24.02%	1.94	38.69%	0.02	0.45%	5.01
White Bear	3.01	27.66%	2.61	24.03%	0.10	0.89%	3.11	28.61%	1.97	18.11%	0.07	0.69%	10.87
White Bear Lake	0.53	6.05%	3.11	35.57%	0.08	0.95%	2.07	23.73%	2.91	33.27%	0.04	0.43%	8.74
Willernie	0.00	1.51%	0.08	59.02%	0.00	0.44%	0.02	16.00%	0.03	22.93%	0.00	0.10%	0.13
Woodbury	1.29	3.63%	10.84	30.40%	5.58	15.65%	11.39	31.95%	6.20	17.40%	0.34	0.97%	35.65
Woodland	0.03	4.66%	0.14	22.36%	0.02	3.85%	0.34	55.02%	0.08	13.06%	0.01	1.05%	0.61
Young America	0.48	1.45%	4.21	12.61%	10.66	31.97%	17.01	50.99%	0.85	2.55%	0.14	0.43%	33.36
TOTAL	168.87		771.56		527.25		1135.97		334.79		34.56		2973.00
MEAN	0.94	6.63%	4.31	27.41%	2.95	13.45%	6.35	35.09%	1.87	16.10%	0.19	1.31%	16.61
STANDARD DEVIATION	1.30	8.64%	4.26	11.31%	4.46	14.54%	6.35	11.83%	3.02	12.63%	0.21	1.00%	14.73

Appendix D. LULC area by county, 2006

	Unclassified		Water		Forest		Cropland		Other Vegetation		Urban		Total
County	Sq. Mi.	% of total county	Sq. Mi.	% of total county	Sq. Mi.	% of total county	Sq. Mi.	% of total county	Sq. Mi.	% of total county	Sq. Mi.	% of total county	Sq. Mi.
Anoka	0.16	0.04%	20.56	4.62%	85.09	19.10%	95.59	21.46%	75.42	16.93%	168.65	37.86%	445.47
Carver	0.08	0.02%	19.73	5.25%	29.46	7.84%	256.80	68.35%	16.48	4.39%	53.16	14.15%	375.71
Dakota	0.06	0.01%	14.24	2.43%	55.11	9.41%	323.07	55.14%	17.09	2.92%	176.37	30.10%	585.93
Hennepin	0.23	0.04%	48.35	7.97%	59.64	9.83%	111.47	18.38%	41.01	6.76%	345.71	57.01%	606.42
Ramsey	0.00	0.00%	13.89	8.17%	13.88	8.17%	2.93	1.72%	9.03	5.32%	130.20	76.62%	169.93
Scott	0.09	0.02%	13.48	3.66%	42.70	11.60%	211.66	57.52%	18.41	5.00%	81.63	22.18%	367.97
Washingtor	0.59	0.14%	33.92	8.01%	92.39	21.82%	142.75	33.72%	29.51	6.97%	124.20	29.34%	423.36
TOTAL	1.21		164.18		378.26		1144.28		206.95		1079.91		2974.79
MEAN	0.17	0.04%	23.45	5.73%	54.04	12.54%	163.47	36.61%	29.56	6.90%	154.27	38.18%	424.97
STANDARD DEVIATION	0.20	0.05%	13.09	2.34%	28.32	5.60%	108.03	24.41%	22.75	4.64%	95.16	21.64%	147.18

Appendix E. Impervious surface by city or township.

City or Township	Pervious Sq mile	Impervious Sq mile	Total Sq mile	Pervious Percent	Impervious Percent	Impervious % of total
Afton	23.08	3.23	26.31	88%	12%	0.37%
Andover	30.21	4.62	34.83	87%	13%	0.54%
Anoka	5.12	2.00	7.11	72%	28%	0.23%
Apple Valley	12.58	5.02	17.60	71%	29%	0.58%
Arden Hills	7.76	1.90	9.66	80%	20%	0.22%
Bayport	1.39	0.42	1.81	77%	23%	0.05%
Baytown	8.14	0.99	9.13	89%	11%	0.11%
Belle Plaine	38.02	5.84	43.86	87%	13%	0.68%
Benton	31.32	2.63	33.95	92%	8%	0.30%
Bethel	0.94	0.10	1.04	90%	10%	0.01%
Birchwood Village	0.32	0.02	0.35	93%	7%	0.00%
Blaine	27.07	6.95	34.03	80%	20%	0.81%
Blakely	24.86	2.09	26.95	92%	8%	0.24%
Bloomington	29.38	8.96	38.34	77%	23%	1.04%
Brooklyn Center	5.68	2.70	8.38	68%	32%	0.31%
Brooklyn Park	18.94	7.68	26.62	71%	29%	0.89%
Burnsville	20.13	6.76	26.89	75%	25%	0.78%
Camden	32.94	1.46	34.40	96%	4%	0.17%
Carver	3.87	0.45	4.32	89%	11%	0.05%
Castle Rock	33.63	1.70	35.33	95%	5%	0.20%
Cedar Lake	32.65	3.57	36.22	90%	10%	0.41%
Centerville	2.03	0.40	2.43	84%	16%	0.05%
Champlin	6.84	1.91	8.75	78%	22%	0.22%
Chanhassen	18.97	3.94	22.91	83%	17%	0.46%
Chaska	14.49	3.17	17.66	82%	18%	0.37%
Circle Pines	1.64	0.30	1.94	85%	15%	0.03%
Coates	1.17	0.21	1.38	85%	15%	0.02%
Cologne	1.46	0.33	1.79	81%	19%	0.04%
Columbia Heights	2.47	1.03	3.50	71%	29%	0.12%
Columbus	44.42	3.20	47.62	93%	7%	0.37%
Coon Rapids	17.48	5.81	23.29	75%	25%	0.67%
Corcoran	33.05	3.00	36.05	92%	8%	0.35%
Cottage Grove	31.58	5.91	37.49	84%	16%	0.69%

Credit River	21.73	2.13	23.86	91%	9%	0.25%
Crystal	4.26	1.59	5.85	73%	27%	0.18%
Dahlgren	30.31	5.28	35.60	85%	15%	0.61%
Dayton	23.18	1.99	25.17	92%	8%	0.23%
Deephaven	2.04	0.33	2.37	86%	14%	0.04%
Dellwood	2.60	0.23	2.82	92%	8%	0.03%
Denmark	26.94	3.27	30.21	89%	11%	0.38%
Douglas	29.81	4.21	34.02	88%	12%	0.49%
Eagan	24.84	8.63	33.47	74%	26%	1.00%
East Bethel	43.36	4.32	47.68	91%	9%	0.50%
Eden Prairie	27.22	7.95	35.17	77%	23%	0.92%
Edina	12.05	3.93	15.98	75%	25%	0.46%
Elko New Market	2.55	0.72	3.27	78%	22%	0.08%
Empire	28.21	3.31	31.52	90%	10%	0.38%
Eureka	31.42	4.22	35.63	88%	12%	0.49%
Excelsior	0.45	0.23	0.68	66%	34%	0.03%
Falcon Heights	1.53	0.70	2.23	69%	31%	0.08%
Farmington	11.84	3.01	14.85	80%	20%	0.35%
Forest Lake	32.70	2.82	35.52	92%	8%	0.33%
Fort Snelling	3.63	3.06	6.69	54%	46%	0.35%
Fridley	7.20	3.60	10.81	67%	33%	0.42%
Gem Lake	0.98	0.12	1.10	89%	11%	0.01%
Golden Valley	7.78	2.77	10.55	74%	26%	0.32%
Grant	24.83	1.63	26.46	94%	6%	0.19%
Greenfield	19.60	1.90	21.50	91%	9%	0.22%
Greenvale	25.73	2.81	28.54	90%	10%	0.33%
Greenwood	0.28	0.07	0.35	79%	21%	0.01%
Grey Cloud Island	2.78	0.29	3.07	91%	9%	0.03%
Ham Lake	32.31	3.53	35.84	90%	10%	0.41%
Hamburg	0.15	0.05	0.20	76%	24%	0.01%
Hampton	32.56	3.13	35.69	91%	9%	0.36%
Hancock	16.62	1.23	17.85	93%	7%	0.14%
Hanover	1.92	0.20	2.12	90%	10%	0.02%
Hastings	8.48	2.78	11.25	75%	25%	0.32%
Helena	30.66	3.39	34.05	90%	10%	0.39%
Hilltop	0.06	0.06	0.13	50%	50%	0.01%
Hollywood	34.03	1.74	35.77	95%	5%	0.20%
Hopkins	2.64	1.46	4.10	64%	36%	0.17%
Hugo	33.28	2.70	35.98	93%	7%	0.31%

Independence	32.54	2.04	34.58	94%	6%	0.24%
Inver Grove Heights	25.24	4.88	30.12	84%	16%	0.57%
Jackson	5.53	0.81	6.34	87%	13%	0.09%
Jordan	2.31	0.90	3.20	72%	28%	0.10%
Lake Elmo	21.67	2.57	24.24	89%	11%	0.30%
Lake Saint Croix Beach	0.86	0.11	0.98	88%	12%	0.01%
Lakeland	2.57	0.36	2.94	88%	12%	0.04%
Lakeland Shores	0.63	0.08	0.72	89%	11%	0.01%
Laketown	25.97	1.94	27.91	93%	7%	0.22%
Lakeville	30.19	7.67	37.86	80%	20%	0.89%
Landfall	0.06	0.02	0.08	72%	28%	0.00%
Lauderdale	0.28	0.14	0.42	67%	33%	0.02%
Lexington	0.53	0.15	0.69	78%	22%	0.02%
Lilydale	0.76	0.08	0.84	90%	10%	0.01%
Lino Lakes	29.73	3.45	33.18	90%	10%	0.40%
Linwood	34.02	2.09	36.11	94%	6%	0.24%
Little Canada	3.28	1.22	4.50	73%	27%	0.14%
Long Lake	0.64	0.30	0.94	68%	32%	0.03%
Loretto	0.17	0.09	0.26	67%	33%	0.01%
Louisville	12.48	2.04	14.52	86%	14%	0.24%
Mahtomedi	4.98	0.71	5.69	87%	13%	0.08%
Maple Grove	26.02	8.99	35.02	74%	26%	1.04%
Maple Plain	0.80	0.26	1.07	75%	25%	0.03%
Maplewood	14.03	3.94	17.97	78%	22%	0.46%
Marine on Saint Croix	3.55	0.26	3.82	93%	7%	0.03%
Marshan	29.19	5.18	34.37	85%	15%	0.60%
May	35.23	3.05	38.28	92%	8%	0.35%
Mayer	1.22	0.22	1.43	85%	15%	0.03%
Medicine Lake	0.15	0.02	0.18	87%	13%	0.00%
Medina	25.01	2.12	27.13	92%	8%	0.25%
Mendota	0.19	0.03	0.22	84%	16%	0.00%
Mendota Heights	8.23	1.90	10.13	81%	19%	0.22%
Miesville	1.55	0.19	1.74	89%	11%	0.02%
Minneapolis	38.29	19.98	58.27	66%	34%	2.32%
Minnetonka	23.65	5.43	29.08	81%	19%	0.63%
Minnetonka Beach	0.39	0.08	0.47	82%	18%	0.01%

Minnetrista	29.61	2.34	31.95	93%	7%	0.27%
Mound	2.79	0.80	3.59	78%	22%	0.09%
Mounds View	3.18	0.91	4.08	78%	22%	0.11%
New Brighton	5.24	1.78	7.03	75%	25%	0.21%
New Germany	0.93	0.08	1.01	92%	8%	0.01%
New Hope	3.44	1.68	5.12	67%	33%	0.20%
New Market	28.42	4.18	32.60	87%	13%	0.49%
New Prague	1.50	0.74	2.24	67%	33%	0.09%
New Trier	0.16	0.02	0.18	90%	10%	0.00%
Newport	3.09	0.77	3.86	80%	20%	0.09%
Nininger	15.77	1.33	17.10	92%	8%	0.15%
North Oaks	7.93	0.69	8.61	92%	8%	0.08%
North Saint Paul	2.25	0.77	3.03	74%	26%	0.09%
Northfield	1.16	0.32	1.48	78%	22%	0.04%
Norwood Young America	1.86	0.42	2.28	82%	18%	0.05%
Nowthen	32.51	2.67	35.18	92%	8%	0.31%
Oak Grove	31.71	3.49	35.20	90%	10%	0.40%
Oak Park Heights	2.38	0.95	3.32	72%	28%	0.11%
Oakdale	8.80	2.50	11.30	78%	22%	0.29%
Orono	23.41	1.91	25.32	92%	8%	0.22%
Osseo	0.37	0.35	0.72	52%	48%	0.04%
Pine Springs	0.84	0.08	0.92	92%	8%	0.01%
Plymouth	26.78	8.69	35.47	75%	25%	1.01%
Prior Lake	15.19	3.21	18.40	83%	17%	0.37%
Ramsey	25.10	4.66	29.76	84%	16%	0.54%
Randolph	10.02	1.47	11.50	87%	13%	0.17%
Ravenna	20.51	1.31	21.83	94%	6%	0.15%
Richfield	4.80	2.32	7.12	67%	33%	0.27%
Robbinsdale	2.13	0.85	2.98	71%	29%	0.10%
Rockford	0.21	0.09	0.29	71%	29%	0.01%
Rogers	22.75	3.50	26.25	87%	13%	0.41%
Rosemount	29.41	5.82	35.22	83%	17%	0.67%
Roseville	9.82	4.04	13.85	71%	29%	0.47%
Saint Anthony	1.74	0.60	2.34	74%	26%	0.07%
Saint Bonifacius	0.79	0.25	1.04	76%	24%	0.03%
Saint Francis	21.42	1.95	23.38	92%	8%	0.23%
Saint Lawrence	12.75	1.79	14.54	88%	12%	0.21%
Saint Louis Park	7.48	3.35	10.83	69%	31%	0.39%

Saint Marys Point	0.34	0.05	0.39	88%	12%	0.01%
Saint Paul	38.39	17.64	56.03	69%	31%	2.05%
Saint Paul Park	2.90	0.68	3.57	81%	19%	0.08%
San Francisco	21.72	2.38	24.10	90%	10%	0.28%
Sand Creek	29.15	3.31	32.46	90%	10%	0.38%
Savage	12.91	3.51	16.42	79%	21%	0.41%
Scandia	35.68	4.18	39.86	90%	10%	0.48%
Sciota	14.27	0.59	14.86	96%	4%	0.07%
Shakopee	21.69	7.71	29.40	74%	26%	0.89%
Shoreview	10.36	2.30	12.65	82%	18%	0.27%
Shorewood	11.97	1.00	12.96	92%	8%	0.12%
South Saint Paul	4.12	1.97	6.09	68%	32%	0.23%
Spring Lake	26.99	2.57	29.56	91%	9%	0.30%
Spring Lake Park	1.42	0.66	2.08	68%	32%	0.08%
Spring Park	0.25	0.16	0.41	61%	39%	0.02%
Stillwater	21.61	3.31	24.93	87%	13%	0.38%
Sunfish Lake	1.54	0.15	1.69	91%	9%	0.02%
Tonka Bay	0.75	0.22	0.97	78%	22%	0.03%
Vadnais Heights	6.65	1.59	8.24	81%	19%	0.18%
Vermillion	31.70	3.51	35.21	90%	10%	0.41%
Victoria	9.41	1.11	10.52	89%	11%	0.13%
Waconia	32.53	2.80	35.33	92%	8%	0.32%
Waterford	13.43	1.30	14.73	91%	9%	0.15%
Watertown	33.43	2.20	35.63	94%	6%	0.25%
Wayzata	2.36	0.78	3.14	75%	25%	0.09%
West Lakeland	11.05	1.60	12.66	87%	13%	0.19%
West Saint Paul	3.54	1.48	5.01	71%	29%	0.17%
White Bear	9.54	1.33	10.87	88%	12%	0.15%
White Bear Lake	6.69	2.06	8.74	76%	24%	0.24%
Willernie	0.12	0.02	0.13	88%	12%	0.00%
Woodbury	28.27	7.38	35.65	79%	21%	0.86%
Woodland	0.53	0.08	0.61	87%	13%	0.01%
Young America	31.66	1.72	33.38	95%	5%	0.20%
Sum:	2543.39	431.01	2974.40	n/a		100%
Mean:	14.21	2.41	16.62	82%	18%	0.56%
Standard Deviation:	12.84	2.73	14.74	9%	9%	0.63%