USING REMOTE SENSING AND GIS TO MONITOR AND PREDICT URBAN GROWTH—CASE STUDY IN ALACHUA COUNTY, FLORIDA

By

YONG HONG GUO

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2012

To my beloved wife Zheng Song and my parents Guo Zhizheng and Xu Jie
In memory of my father Mr. Guo Zhizheng

ACKNOWLEDGMENTS

I would like to thank my wife, Zheng Song, for her full consideration and wholeheartedness in supporting my Ph.D. study. Because of her vigorous and cooperative support, my Ph.D. study was conducted smoothly and efficiently. Also, I would like to thank my parents, Mr. Guo Zhizheng and Mrs. Xu Jie, for their moral and financial support throughout my entire Ph.D. I would not be able to finish this Ph.D. without their timely and enthusiastic encouragement and assistance.

In addition, I would like to thank Dr. Paul Zwick, my committee chair, for his support and advice during my study. Dr. Zwick provided valuable comments and guidance for this research; he also financially assisted me for the research. Dr. Zwick provided me with heuristic discussions, effective technical comments and input for the writing of this dissertation; without his technical support and financial assistance, this dissertation could not have been fulfilled. I am also thankful for my committee members, Dr. Michael Binford, Dr. Ilir Bejleri, and Prof. Margaret Carr for their valuable comments and guidance for this dissertation. Finally, I am also grateful for Mr. Chad Riding at Caltrans, Calif., Mr. Thomas Cole and Ms. Lyndsay Brown at the Reading and Writing Center at the University of Florida who helped me edit the manuscripts of this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABM	Agent-Based Model
AHP	Analytic Hierarchy Process
AI	Artificial Intelligence
APA	American Planning Association
BAU	Business As Usual
BEBR	Bureau of Economic and Business Research
CA	Cellular Automata
CART	Classification and Regression Tree
CBD	Central Business District
C-I-T	Commercial-Institutional-Transportation
CLUE	Conversion of Land Use and Its Effects
CLUE-S	Conversion of Land Use and its Effects at Small Regional Extent
CPU	Central Processing Unit
CUF	California Urban Future
CV	Cross Validation
DN	Digital Number
DOQQ	Digital Orthophoto Quarter Quadrangle
EML	ERDAS Macro Language
ETM+	Enhanced Thematic Mapper Plus
GA	Genetic Algorithm
GIS	Geographic Information System
GRD	Gross Residential Density
GUD	Gross Urban Density
LBCS	Land Based Classification Standards

LUCIS	Land Use Conflict Identification Strategy
LULC	Land Use and Land Cover
MCM	Markov-Chain Model
MLR	Multinomial Logistic Regression
MSS	Multispectral Scanner System
MUA	Multiple Utility Assignment
NDVI	Normalized Difference Vegetation Index
NLCD	National Land Cover Database
NN	Neural Networks
OLS	Ordinary Least Squares
PC	Principal Component
QUEST	Quick, Unbiased, and Efficient Statistical Tree
ROC	Relative Operating Characteristic
ROI	Region of Interest
SE	Standard Error
TIGER	Topologically Integrated Geographic Encoding and Referencing
TM	Thematic Mapper
UGB	Urban Growth Boundary
USGS	U.S. Geological Survey
V-I-S	Vegetation, Impervious Surface, Soil

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

USING REMOTE SENSING AND GIS TO MONITOR AND PREDICT
URBAN GROWTH—CASE STUDY IN ALACHUA
COUNTY, FLORIDA
By

Yong Hong Guo

May 2012

Chair: Paul D. Zwick
Major: Design, Construction and Planning

Alachua County experienced remarkable urban growth in the extent of the urban area during the past three decades, largely due to population growth. This dissertation simulates and predicts future urban growth in the county based on population growth. First, this dissertation classifies urban land uses and land covers (LULCs) based on eleven classes equivalent to the U.S. Geological Survey (USGS) Level III standard using remote sensing Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) data, as well as the state-of-the-art classification and regression tree (CART) method. The CART method is then compared to the Vegetation-Impervious Surface-Soil (V-I-S) and conventional supervised classification methods based on a pilot area. Next, this dissertation classifies 1982, 1994, and 2003 LULC classes for Alachua County as a whole. Second, a multinomial logistic regression (MLR) model is employed to simulate urban LULC development in 2003 based on the urban LULC classified for 1982, 1994, and 2003, respectively. Accordingly, a number of probability maps are created for the four classified urban uses, which include single-family, multi-family, commercial-institutional-transportation, and industrial-warehouses. Third, an urban LULC allocation process is undertaken for Alachua County, based on five scenarios for 2020 and 2030. The five scenarios include the business as usual (BAU) scenario, infill development, increased density

development, redevelopment, and the conservation scenario. This research has wide application in both the public and the private sectors of urban planning and additional areas in which the ability to describe the spatial extent of urban areas and their future expansion is of interest.

CHAPTER 1
INTRODUCTION

The State of Florida experienced an average annual population growth of 3.5 percent over the past four decades (Dewey and Denslow, 2001). Following this trend, Alachua County had remarkable urban growth in the extent of the urban area during the same period, expanding 130 percent between 1990 and 2000.[1] This dissertation investigates urban growth in Alachua County over the past three decades and simulates urban development for the future 20 years using remote sensing and Geographic Information System (GIS) technologies. Specifically, this dissertation illustrates how urban growth proceeded in Alachua County in 1982, 1994, 2003, and 2010, respectively, and projects how it might proceed in 2020 and 2030. The study first discusses three urban land use and land cover (LULC) classification methods in remote sensing, which include the classification and regression tree (CART) method, the Vegetation-Impervious Surface-Soil (V-I-S) method, and the supervised method. The goal of this research is to compare the state-of-the-art CART method with the other two in order to determine the best method for the research. Thus, a pilot area is chosen in this regard. Based on the pilot area, this dissertation classifies eleven urban and natural LULC classes in Alachua County by using the best method obtained. Next, LULC for Alachua County as a whole is classified. In addition, this research predicts urban growth in the county in 2020 and 2030, respectively, by using the multinomial logistic regression (MLR) model. As a result, five development scenarios are proposed for future LULC simulations in the county: the business as usual (BAU) scenario, the infill scenario, the increased density scenario, the redevelopment scenario, and the conservation scenario. The study provides a framework for researchers to apply the best and/or newest methods introduced in this research to other counties or jurisdictions in the State of Florida as well as the rest of the nation.

---

[1] Alachua County had approximately 62 square miles of urbanized areas in 1990, and approximately 81 square miles in 2000 (Data source: http://www.fgdl.org/).

Brief Review of Urban Growth in the U.S.

Urban growth patterns in the U.S. have experienced a dramatic change over the past 150 years. The Industrial Revolution, which began in the early nineteenth century, had a huge impact on American cities, which not only grew rapidly in population, but also in areas. U.S. urban growth in the industrial era was concentrated in cities because a large number of people from rural areas, along with international immigrants, poured into cities to find jobs (Jackson, 1985). In Manhattan Island, New York City, for example, there were about 3.3 million inhabitants in a 22-square mile area during 1900, an average density of 100,000 persons per square mile. In the lower East Side of New York, the density in some wards was several times higher (Levy, 2003). When enough people were concentrated inside cities, a residential development market was created, and large numbers of private houses were built near major urban centers (Jackson, 1985).

Decentralization of the central cities began simultaneously with this time of urban concentration: the Industrial Revolution produced faster transportation tools, such as trains, electric trolley cars, and subways, which provided great impetus for affluent central city residents to flee to the suburbs so as to avoid overcrowding and dirtiness in central cities (Jackson, 1985; Mumford, 1961). In the U.S., a number of suburban towns were built between 1850 and 1920 such as Riverside, Illinois, and Forest Hill Gardens, New York. However, the number and size of railroad suburban towns remained small because people preferred a walking environment, and neighborhoods were planned based on walking distance (Mumford, 1961).

If concentration and decentralization were major themes of urban growth in the nineteenth century America, decentralization generally characterized urban America in the twentieth century, and this process continues today (Levy, 2003). The technology that fostered this decentralized urban growth was the private automobile. With the mass production of automobiles beginning in earnest in the 1920s, a large number of urban dwellers spread from the central city out to the suburbs, which gave rise to the first great wave of suburbanization in the U.S. (Levy, 2003).

American automobile ownership has increased steadily over the years. By 1930, there were about 25 million automobiles; by 1950, the number had risen to 40 million; in 1960, the number stood at 62 million (Levy, 2003); and as of the end of 1999, the number reached 132.4 million (Bureau of Transportation Statistics, 2002). Today, automobile ownership has outnumbered population increase at national, regional, and local levels; as a result, car ownership increased 383 percent between 1950 and 2000, compared to an 80 percent increase in population growth (The Boston Indicators Project, no date).

The construction of inter-state freeways after World War II further contributed to post-war suburbanization (Jackson, 1985; Levy, 2003). Along with cheap gas and the mass production of automobiles, inter-state freeways provided lower transport costs that greatly stimulated decentralized urban growth (Jackson, 1985). In addition, the wide usage of telephones made long-distance communication without face-to-face contact possible, and this further contributed to decentralized urban economic activities and growth (Levy, 2003). Today, the World Wide Web, email, mobile phones, and fax all make long-distance communication easier and more economical (Levy, 2003).

National demographic trends echoed the above phenomenon. On the one hand, the metropolitan population was on the rise. Between 1950 and 2000, the population in America’s metropolitan areas increased from 55 percent to 80 percent (Lucy and Phillips, 2006). On the other hand, the urban population declined significantly in metropolitan areas, and suburban population increased steadily. Between 1950 and 1970, the suburban population grew approximately two-fold, from 36 million to 74 million (Jackson, 1985). In 1970, suburbanites for the first time outnumbered urban dwellers in the U.S. (Jackson, 1985). During the 1980s, the suburban population grew ten times faster than the central city population in the nation's largest metropolitan areas, and the suburban population contributed nearly 60 percent of the nation's metropolitan population as of 1990 (Benfield et al., 1999). Baltimore and St. Louis were two extreme examples: each lost 31 and 59 percent, respectively, of their central city populations between 1950 and 2000 (Lucy and Phillips, 2006).

From the perspective of urban land consumption, statistics show that between 1950 and 1990, suburbanization led to drastic growth in the size of metropolitan areas, from 208,000 square miles to 585,000 square miles (Squires, 2002). In Los Angeles, for example, the urban area grew an astonishing 300 percent between 1970 and 1990, while the urban population increased by only approximately 45 percent (Benfield et al., 1999). The trend of growth in urban land area surpassing urban population growth is unlikely to change in the near future.

Alternatives to Decentralized Urban Development

In light of the recent trend for decentralized urban development, researchers are contemplating alternatives to decentralized urban growth because of its negative externalities. Although outward urban growth possesses both good and bad effects (Cieslewicz, 2002; Benfield et al., 1999; Lucy and Phillips, 2006; Jargowsky, 2002; Squires, 2002; Kahn, 2006; Wassmer, 2001), a number of remedies, as well as new visions, have been outlined by several analysts (Lucy and Phillips, 2006; Squires, 2002). Their proposed remedies to decentralized urban growth are discussed below.

Squires (2002) emphasized that "smart growth" should be proposed as an alternative to decentralized urban sprawl in local jurisdictions. Smart growth principles seek to use existing resources efficiently, and also to allocate them more equitably. Based on this notion, Squires

(2002) outlined eight remedies to offset negative externalities of the decentralized outward urban growth:

- Reusing existing land and infrastructure resources;
- Restricting development in outlying suburban and exurban areas;
- Relying less on the automobile by developing a number of transportation alternatives;
- Concentrating residential and commercial development centrally and along mass transit lines;
- Devoting more money to area-wide revenue sharing and regional investment pools;
- Constructing more affordable housing and distributing it throughout metropolitan areas;
- Enforcing more vigorously fair housing laws; and
- Increasing public and private investments in central cities for more balanced development throughout the region.

Downs (1994) proposed five components of visions to address the negative outcomes of decentralized outward urban growth, which include ownership of detached, single-family homes; ownership of private automotive vehicles; employment in scattered, low-density workplaces, themselves in landscaped settings with free parking; living in small communities with strong local governments; and provision of housing to the urban poor through a "trickle-down" [2] (p.10) neighborhood change process (Downs, 1994). In opposition to nearly universal single-family detached housing, Downs (1994) suggested the construction of high-density residential units in new growth areas, or mixing higher density housing with low-density single-family homes. He proposed a confinement mechanism that was similar to the urban growth boundary (UGB) concept. To reduce near-universal private car ownership, Downs (1994) recommended the use of public transit, bicycles, walking, or ride-sharing. Further, to address scattered jobs among low-

[2] Federal government subsidies low-income households directly. Will be narrated below.

density areas, Downs (1994) proposed that new jobs should be located in large employment clusters. He thought job density ought to be intensified even though new jobs could be widely scattered. For local governments, Downs (1994) put forward “a fully-functioned elected metropolitan government” (Downs, 1994, p.132) by merging the central city government with one or more adjacent counties, or using the state government to coordinate local planning actions. Downs (1994) also suggested changing the existing affordable housing standards and providing federal low-income housing subsidies directly to the urban poor through the “trickle down” process noted above.

Duany et al. (2000) proposed a traditional neighborhood concept as a solution to outward urban growth. The traditional neighborhood concept consisted of six rules: a clear center within a neighborhood; no more than a five-minute walk for ordinary daily life; the connection of one location to another via a continuous web and numerous paths; small streets; mixed-use blocks; and providing unique sites for civic buildings (Duany et al., 2000). In fact, the above six rules are New Urbanism principles proposed to combat decentralized outward urban growth in the twenty-first century. One example includes Alexandra, Virginia, a town built based on the six traditional neighborhood design principles mentioned above (Duany et al., 2000).

Duany et al. (2000) proposed a number of strategies to implement their traditional neighborhood concept. They suggested that development take place within a comprehensive regional plan designed to limit the use of automobiles and preserve open space, rather than focusing on individual buildings. Based on this notion, they believed that new development ought to be placed adjacent to infrastructure, e.g., a transit stop. Because existing and future rail lines are often addressed in regional plans, new neighborhoods and town centers could be planned based upon those transit lines. In addition, mixed-uses were suggested based on their traditional neighborhood concept. Duany et al. (2000) recommended that every residential neighborhood should have a corner store to provide residents with daily needs. The proposed corner stores could limit automobile dependence because of their location within walking distance. In addition to corner stores, neighborhood-scaled shopping centers could be considered for a large population, and they can be next to through-traffic. Mixed-uses should also include office spaces, which ought to be developed simultaneously with residential development. In addition, civic buildings, such as city halls, libraries, neighborhood elementary schools, etc., need to be constructed inside new communities, and land ought to be reserved for those buildings, which enhance community identity. Moreover, Duany et al. (2000) recommended road network connectivity in a residential neighborhood, in which roads ought to be fully connected in all directions around a residential unit in order to avoid using the same few collector roads (Duany et al., 2000).

In general, the remedies to decentralized outward urban growth rest upon reshaping government structure to be regional, strengthening state government's oversight, new neighborhood planning and design concepts, increasing housing density, promoting mixed-use development, and developing new transportation alternatives. In addition, despite the fact that decentralized outward urban growth is a national trend, its degrees vary locally. Remedies should be undertaken based on a particular political boundary, with an accurate assessment of the degree of decentralized outward urban growth in a jurisdiction. As a result, monitoring and/or detecting decentralized outward urban growth at a local level becomes a very important task, and the question of how to monitor and predict it, along with measuring the degree in a jurisdiction, is a serious research question. Currently, there are no universal methods that define and simulate decentralized outward urban growth at the national as well as local level (Broos and Day, no date). This dissertation uses remote sensing technologies, combined with the GIS tool, to research decentralized urban growth.

Research Questions

The research questions of this dissertation aim to model urban growth in Alachua County in the past, at present, and in the future and to develop and test a methodology for predicting decentralized urban growth using satellite imagery and property parcel data. Specifically, this dissertation looks into three LULC classification methods using remote sensing technologies—the CART method, the V-I-S method, and the supervised—and compares them in a pilot area. The dissertation explores a new method—the CART method—for urban LULC classifications, and investigates these three methods in detail. The newly obtained method, i.e., the CART method, is later used to create a LULC classification for Alachua County. This study uses the Landsat data Thematic Mapper (TM) and the Enhanced Thematic Mapper Plus (ETM+) imageries for analysis because satellite data are less costly, easier to obtain than aerial photographic data, and are capable of supporting sub-parcel LULC analysis. LULC data are helpful for the analysis of physical urban form as compared to vector parcel data. Also, this study uses GIS technology as a supplemental tool for the satellite data in order to generate both raster and ancillary data.

This dissertation compares Alachua County land use of 1982, 1994, and 2003, respectively, using the most optimal LULC classification method identified herein. Alachua County has experienced remarkable urban growth over the past decades. Based on the identification of the current urban growth in 2010, future land use patterns are predicted for 2020 and 2030, respectively.

This research contributes to the academic field by identifying a new method for LULC classification and predicting future land use based on the new LULC classifications. Few studies have been conducted on urban growth that use Landsat images to accurately classify LULC equivalent to U.S. Geological Survey (USGS) Level III. This study identifies a methodology of utilizing 30 × 30 meters Landsat TM and ETM+ imageries for LULC classifications.

Research Applications

This research has wide applications in urban planning and supplementary areas including land use planning; urban design; transportation planning; parks, tourism, and recreation facilities planning; population forecasting; urban economics; water and soil conservation; crime prevention; environmental and ecosystem protection; natural resources conservation; and real estate and industrial development. It will benefit both the public and the private sectors.

Specifically, since the major intent of this research is to provide accurate predictions of future land-use, it supports evaluation of local and state land-use policy for curbing urban sprawl by allowing for comparisons of land use change over time. The intent of this research is to make future land-use patterns tangible. The research also provides additional intuitive technical applications, such as population forecasts, traffic counts, and a methodology for site selection for real estate development. In sum, the applications of the research are monitoring urban growth, heuristic state and local land-use policy making, and intuitive technical applications.

CHAPTER 2
URBAN LAND USE AND LAND COVER CLASSIFICATION

Chapter Overview

The difficulty of classifying urban LULC has been noticed by researchers for many years. Because of the spectral characteristics of urban LULC classes, it is often difficult to classify urban uses and achieve high accuracy levels using the conventional approaches such as the supervised and unsupervised classification methods. Alternative approaches such as the CART method is being discussed (Paul, 2007). This chapter explores a way to apply the CART method in urban LULC classifications to Alachua County by comparing the CART method with two other classification methods. This chapter explores the CART method, the V-I-S method, and the supervised classification method in a pilot area in Alachua County, and uses the CART method to classify urban LULC classes for the whole county. This new CART method that is applied makes use of the advantages of the method and yields the highest accuracy level. The overall accuracy of using the CART method reaches more than 90 percent for 1982, 1994, and 2003, respectively, in the whole county.

Literature Review

Supervised and Unsupervised Classification Methods

Currently, there are several ways that have been adopted by researchers to classify urban LULC classes using satellite imageries applying the conventional supervised and unsupervised classification methods. Yang (2000) used an unsupervised classification method to classify urban LULC classes for the region of Atlanta, Georgia, based on the Multispectral Scanner System (MSS) and TM images. In his research, Yang (2000) classified urban LULC classes into six types, which were high density urban, low density urban, cultivated/exposed land, cropland/grassland, forest, and water. For his research, Yang (2000) utilized the ISODATA iteration to classify urban LULC classes into two types, namely, high-density urban and low-density urban, and sorted out the mixed pixels by using the modal filter so as to identify the pixels with the highest frequency in the histogram of a Digital Number (DN) value, i.e., the pixel value, as the representatives of the most commonly occurring class in a patch. To address the issue of urban LULC classes that have similar spectral characteristics, Yang (2000) used an image interpretation approach to identify four major pairs of urban LULC classes that have similar spectral signatures, which were low-density urban and forest; low-density urban and cropland, forests; and high-density urban and exposed land. After this identification was completed, Yang (2000) started a reclassification process, in which the leftover erroneous land covers were manually recoded into the correct land covers. The overall accuracy level was elevated from the initial, about 60 percent, to 87 percent. Finally, Yang (2000) used aerial photography as training data for accuracy assessment. However, Yang (2000) classified urban landscapes based on only two categories, which were the high-density urban and low-density urban. He did not separate urban LULC classes further into several detailed categories such as residential, commercial, and industrial.

Similar to Yang (2000), Lo and Choi (2004) employed a hybrid supervised method combined with unsupervised approach to mapping urban LULC classes by using Landsat ETM+ imagery. Their study area was also Atlanta, Georgia. Similar to Yang (2000), they initially utilized the ISODATA to cluster all likely homogeneous pixels. They subsequently used a fuzzy supervised classification method to address the mixed pixels based on the created clusters. Their LULC classes were also high-density urban use, low-density urban use, cultivated/exposed land, cropland or grassland, forest, and water.

Using the unsupervised classification method, Lo and Choi (2004) generated 60 classes as an optimal number for ISODATA experienced from their previous Atlanta-related projects and then labeled the homogenous pixels and left out those unlabeled mixed pixels to supervised fuzzy classification. After this was completed, they created a fuzzy subset for those heterogeneous pixels based on a membership function, in which membership grade was assigned to each class. The membership grade was between 0 and 1. This was a fuzzy signature assigning process. To determine the initial membership grade for each LULC class, they utilized DOQQ images to collect training samples, for which three to four training sites were selected. For example, for a typical site the fuzzy membership grade that was assigned for high-density urban land use was 0.92 and 0.08 for low-density urban land use while 0 was used for other LULC classes. Then, based on those fuzzy membership grades, Lo and Choi (2004) calculated fuzzy mean and fuzzy covariance matrix parameters for each class, which was significant for supervised fuzzy classification. When the supervised fuzzy classification was finished, Lo and Choi (2004) used an overlay to create a union of the pervious ISODATA classified LULC map with the supervised fuzzy classified LULC map in order to formulate a complete LULC map for the study area. The overall accuracy for the hybrid classification method was 91.5 percent (Lo and Choi, 2004).

From Lo and Choi (2004), the hybrid classification requires a fine assignment of initial membership grade matrix for each urban LULC class. Although they obtained a high accuracy level for urban classes in their research, like Yang (2000), they did not separate each urban LULC further to residential, commercial, and industrial. Whether a high accuracy level can be obtained by applying their method to more detailed urban LULC classes is unknown.

Jensen and Toll (1982) used the Landsat MSS imagery to detect residential land-use development at the urban fringe in Denver, Colorado. For this purpose, Jensen and Toll (1982) proposed a change detection model to identify stages of the residential development with respect to the residential development cycle from parcel clearing to complete landscaping. Jensen and Toll (1982) first used panchromatic aerial photography to manually identify 10 stages of single-family development, which namely were (1) original land cover; (2) area cleared; (3) area cleared, subdivided, and paved roads; (4) area cleared, subdivided, paved roads, and building; (5) subdivided, paved roads, buildings, partially landscaped; (6) cleared, subdivided, dirt roads; (7) cleared, subdivided, dirt roads, and buildings; (8) subdivided, dirt roads, buildings, and partially landscaped; (9) subdivided, dirt roads, buildings, and landscaped; and (10) subdivided, paved roads, buildings, and landscaped (p.630). Then, they utilized a gray-tone spatial dependency matrix to extract textural information from the MSS Band 5 imagery because they thought Band 5 (0.6 – 0.7 μm) enhanced the contrast between vegetated and non-vegetated surfaces (Jensen and Toll, 1982). Next, they chose the image differencing method to do the change analysis based on the previously extracted texture images because they found that the image differencing method produced lower change detection errors (Jensen and Toll, 1982). After overlapping the Band 5 change image with the panchromatic aerial photography, they found that Band 5 was good at detecting land use changes from original, natural vegetation to partially or fully landscaped residential development while unable to detect the changes between partially developed land and un-irrigated rangeland. Therefore, they made a texture change map and overlaid the newly created texture change map onto the previously generated spectral change map and formed a single change map. The absolute accuracy of the new change map hence reached 81 percent.

From Jensen and Toll’s (1982) research, it is found that texture difference provides useful information for the identification of different residential development stages at urban fringes because each stage of development presents different texture characteristics. However, textural information must be used in complement with the spectral information such as Band 5. In addition, the identification of the residential development cycle is only one of the aspects in urban LULC identification. Whether Landsat imageries can be applied to identify and classify additional urban LULC classes and simultaneously yield high accuracy has not been reported by researchers so far.

V-I-S Model

Ridd (1995) proposed a V-I-S model in 1995 for the first time to classify LULC classes for Salt Lake City. Borrowing ideas from soil texture compositions, he put vegetation, impervious surface, and soil on vertexes of a triangle (Figure 2-1). Urban land features are determined based on the component percentages to be represented as a percentage on the triangle sides. For example, high density residential use is located on the impervious surface side, which is close to the central business district (CBD), occupying about 60 percent of the impervious surfaces while low-density residential occupies about 30 percent of the impervious surfaces that are close to vegetation (Figure 2-1).

For practice, Ridd (1995) applied the V-I-S model to a pilot area in Salt Lake City with data collected from two linear transects radiating from the city center. The linear transect Ridd (1995) applied was a line stretching from the city center to the suburbs, on which equal frames were selected that forms a transect. Then, sample points were collected in each sample frame of a transect. The sample data were tabulated and marked as seven zones, in which homogeneous data were found. These zones included CBD, near-town old residential, stable mature residential, country club, postwar mid-income residential, Mt. Olympic Cove, and natural woodland. Each zone has its specific V-I-S compositions, and the V-I-S compositions for an urban environment are identified from interpreting the diagrams of the compositions of each zone.

Ward et al. (2000) applied the V-I-S model in Queensland, Australia, which was an example after Ridd's Salt Lake City project. The basic elements of their research were similar to Ridd's (1995). To distinguish vegetation, Ward et al. (2000) used the normalized difference vegetation index (NDVI), Band 5, and Band 3. Ward et al. (2000) also applied mineral indexes to distinguish rural and agricultural land uses from urban land uses. For urban impervious surfaces, Ward et al. (2000) used Band 3 to extract the texture image.

Phinn et al. (2002) explored the V-I-S model in the City of Brisbane, Australia, on a sub-pixel basis, mimicking the approach of Ward et al. (2000). They first conducted per-pixel image classification based on the NDVI, Band 5, and Band 3 to produce 20 classes, which separated vegetated and non-vegetated classes. Then, similar to Ward et al. (2000), they classified the unclassified non-vegetated surfaces into the soil related classes using mineral indexes. After this is completed, they employed a texture layer to extract urban areas. The five classes output from their V-I-S process were forest/woodland-vegetation; grass/sparse-vegetation; cleared areas-soil; developed areas-impervious; and water bodies.

Similar to Ridd (1995), Phinn et al. (2002) also extracted the V-I-S components from aerial photography along transects radiating from Brisbane's city center to the limits of its metropolitan area and found 1-by-1 km sample frames. Within each sample frame, thirteen random sets were generated, which determined whether a set belong to a certain component: vegetation, impervious surfaces, and soil. In this case, the V-I-S components of a certain point can be calculated so that various component percentages can be calculated based on the V-I-S diagram for each location. Therefore, an area can be identified for their specific use based on the extracted V-I-S components.

Hung (2003) explored the V-I-S model in order to analyze urban LULC classes on a sub-pixel basis for the Salt Lake Valley area. According to Hung (2003), most V-I-S researchers still used a per-pixel-based analytic method to explore urban LULC change, in which one pixel represented one LULC type and other types were ignored (Hung, 2003). Therefore, Hung (2003) proposed a sub-pixel method by using the V-I-S model, which he believed could better deal with mixed and confused pixels than a conventional pixel-by-pixel method.

Hung (2003) used three different types of satellite images for his research, which were MSS, TM, and ETM+, covering the timeframe from 1972 through 1999 with each sensor covering a particular timeframe. He used a supervised sub-pixel classifier to classify the ground LULC classes. The resulting images included multi-channels with each channel indicating a certain proportion of a typical ground cover. According to Hung (2003), the channel with the proportion of a ground cover contained more ground cover information than a single pixel classifier image did. Hung's (2003) V-I-S model is presented in Figure 2-2, which is similar to Ridd's (1995). For Point T in the figure, it has 55 percent soil, 10 percent vegetation, and 35 percent impervious surfaces so that urban land use types such as residential, commercial, industrial, and recreational can be determined based on the component percentages of each type.

Hung (2003) also categorized urban LULC classes into six types based on the V-I-S diagram above, which were Vgg (green grass vegetation), Vts (tree/shrub vegetation), Ibr (bright impervious surface), Imd (medium impervious surface), Idk (dark impervious surface), and Sdv (soil/dry vegetation). After conducting proper geometric rectification and co-registration processes, Hung (2003) defined training sites from remotely sensed images and used them as end-members with each pixel softly classified into six membership grades, which were later translated into component percentages for ground covers and checked by a linear spectral mixture model (Hung, 2003). Then, for accuracy assessment purposes, the resulting component percentages of ground covers were compared to another set of component percentages of ground covers derived from aerial photography visual interpretation (Hung, 2003).

Hung (2003) applied Baye's algorithm to estimate the component percentages of ground covers in which the likelihood between the candidate pixel and the six predefined ground covers, such as Vgg, Vts, and so on, was calculated. Each pixel had six likelihood values and these likelihoods were then converted to percentages (which equaled to 100 percent) based on a conversion algorithm. Then, a linear spectral mixture model was applied. The linear spectral mixture model is the algorithm that spectral reflectance of a typical pixel was the sum of all the spectral reflectance of its ground components weighted by the proportion of each component on the ground (Hung, 2003). Hung (2003) chose Band 4 as the checking band because he thought Band 4 was sensitive to vegetation. An expert system rule was also applied, which was designed to determine the dominant component in a candidate pixel as well as to determine which component(s) to adjust.

An accuracy assessment was processed later in his study. Because MSS, TM, and ETM+ each covered a specific timeframe, accuracy assessments were processed based on different satellite images, which specifically were the 1979 MSS image, the 1987 and 1990 TM images, and the 1999 ETM+ image. Hung (2003) selected two sample areas, namely, Salt Lake City and West Jordan, for accuracy assessments. The component percentages of ground covers derived from the above images were compared to another set of the component percentages of ground cover derived from the visual interpretation from aerial photography (Hung, 2003). The comparison was processed based on regression analysis, in which correlation coefficients were used to measure accuracy level. In terms of the Salt Lake City sample area, the overall correlation coefficient for MSS 1979 was 0.883 with some specific correlation coefficient values for each component; the overall correlation coefficient for TM 1990 was 0.75; and the overall correlation coefficient for ETM+ 1999 was 0.755. Similar results were obtained for the West Jordan area. Hung (2003) concluded that overall the estimated percentages showed a significant relationship with surveyed percentages but had an underestimate of percentages of soil and an overestimate of impervious surfaces for foothills, mountains, and farmlands due to the seasonable agricultural landscape change and soil types (Hung, 2003).

Based on the ground component percentages characteristics according to the V-I-S model, Hung (2003) categorized urban land uses into several general urban features, namely, downtown, light industry, heavy industry, city park, golf course, low density residential, medium density residential, high density residential, the University of Utah campus, foothill, ranch, crop field with vegetation, crop field with mixed vegetation and soil, and crop field with soil. For example, downtown had 7 percent vegetation, 3 percent soil, and 90 percent impervious surface. Low density residential had 68 percent vegetation, 6 percent soil, and 26 percent impervious surface, and so on for other urban features.

From Ridd (1995), Ward et al. (2000), Phinn et al. (2000), and Hung’s (2003) research, it is evident that the V-I-S model is capable of identifying various urban LULC types in detail, such as residential in various densities, light and heavy industrial, recreation, agricultural, and the CBD. The V-I-S model can successfully deal with urban LULC issues such as mixed pixels coupled with approaches such as Baye’s algorithm and the linear spectral mixture. However, Ridd (1995), Ward et al. (2000), Phinn et al. (2002), and Hung (2003) did not classify the commercial, institutional, and transportation uses further from their urban uses; for example, they just classified the commercial uses based on the CBD or downtown. The classifications of commercial, institutional, and transportation uses in other parts of the study area are missing. Their simple classifications of the CBD are not satisfactory to address the complicated nature of urban LULC classes because the commercial use, the institutional use, and the transportation use are important urban LULC classes, which may have similar spectral characteristics to industrial and be intermingled with residential or other uses. Furthermore, wetlands were not addressed by the V-I-S researchers either. This is due to the fact that the V-I-S model does not take water into consideration.

To make up the inadequacy of the V-I-S model, Lu and Weng (2004) proposed a model named “Lu-Weng Urban Landscape Model,” which is an alternative from the V-I-S model. The Lu-Weng Model divides urban landscapes into three components, which are shade, green vegetation, and soil/impervious surface rather than vegetation, impervious surface, and soil. Lu and Weng (2004) argued that the V-I-S model was not adequate to address water and wetlands because it did not have a water component. The Lu-Weng model had a shade element, however, which could be used to classify water and wetland. Lu and Weng (2004) utilized their model in Marion County, Indiana. Their final product included the classification of the urban LULC classes into six categories, which were forest, grass, pasture and agricultural lands, residential, urban, and water. Specifically, they combined commercial and industrial into urban as well as high-intensity residential and low-intensity residential into residential. The overall accuracy reached 80 percent.

Lu and Weng’s (2004) research is closer to the comprehensive classifications of urban LULC classes, which covers not only the vegetation covers such as forests, grasses, and pastures and agricultures, but also non-vegetation covers such as water, as well as some urban uses and features such as residential and urban. However, their research is still unable to sort out commercial and industrial from urban uses because commercial and industrial are in similar impervious percentages in their model. In addition, their overall accuracy is still low because the urban landscape components are different in different geographic locations and hence one fixed model cannot fit all the landscape scenarios. For example, in an area that is predominantly covered by vegetation, it is inappropriate to use the shape, green vegetation, and soil/impervious surface model because of less impervious surfaces in the area, while in an urban area that is predominantly covered by impervious surfaces, this model is the best to represent each component on the ground, and usually the study area incorporates these two areas together at the same time. In this case, researchers need to adjust their model for each geographic landscape scenario and propose different models for each sub-area. This creates technical complexity for the research.

CART Method

Although efforts have been made by using the conventional classification methods as well as the V-I-S method for urban LULC classes classifications, research finds that the conventional supervised and unsupervised classification methods using spectral data alone have a 5 to 10 percent lower overall accuracy than methods using ancillary data (Rogan et al., 2003). Therefore, alternative approaches such as the rule-based classification methods have received increasing attention as of late (Zambon et al., 2006; Lawrence and Wright, 2001). A knowledge-based decision tree model named CART has emerged to classify urban LULC classes, which showed promise for improving classification accuracy in a number of studies (Zambon et al., 2006).

Lewis (2000) compared CART with the multivariate logistic regression model in his clinical research and found that CART had several advantages over the multivariate regression model. First, CART is non-parametric, which can handle highly skewed or multi-modal data; compared to multivariate regression models, CART does not require transformation if data are not normally distributed. This is concurred by Huang and Jensen (1997) in remote sensing, in which their research showed clearly the maximum likelihood classifier would be significantly influenced by the distribution of data and could not effectively handle a bi- or multi-modal distribution; because of that, spectral data alone were not able to discriminate among mixed classes. Second, CART can deal with missing data better than multivariate regression models. For missing data, CART uses "surrogate" variables instead of dropping off missing data by multivariate regression models (Lewis, 2000, p.7). Third, CART has an automatic "machine learning" method compared to multivariate regression models (Lewis, 2000, p.6). CART is different from other expert systems because it has an automate knowledge base building program without extensive a priori expert knowledge (Lawrence and Wright, 2001, p.1,141), which eliminates or reduces the occurrence of a "knowledge acquisition bottleneck" (Huang and Jensen, 1997, p.1,185). As a result, CART requires relatively little input while the multivariate regression model requires extensive input from the analyst, analysis of interim results, and later modification of the method (Lewis, 2000, p.6).

CART allows ancillary data to go with the classification, which has advantages over other parametric methods for the elevation of accuracy levels (Lawrence and Wright, 2001). It can incorporate either continuous/categorical or raster/vector ancillary data. Usually GIS data, as well as texture information, can be included into ancillary data (Lawrence and Wright, 2001). Other data such as elevation, slope, soils, road networks, and so on can also be included (Jensen, 2005).

Herold et al. (2003) applied the CART to map impervious surfaces and forest canopies on a sub-pixel basis for a study area in the eastern portion of the Chesapeake Bay watershed in Maryland, where the study area was covered by clouds. Like Lewis (2000), Herold et al. (2003) used a recursive binary partition process for the training data, which were sampled from the high resolution imagery sources. These training data were representatives of the entire dataset and used in the production rule sets, which enabled the software to build a "knowledge base" (Herold et al., 2003, p.1). Because the CART method was able to provide low-cost, high-quality knowledge bases, the CART software was able to approximate the human learning process without a large amount of users' input.

The research of Herold et al. (2003) utilized 30-meter ETM+ data from three different seasons between 1999 and 2001. After ETM+ images were geometrically and radiometrically corrected, Herold et al. (2003) selected training samples from the DOQQ pictures of 1-meter resolution and classified them based on spatial and spectral characteristics of the ETM+ imagery. The sampling process was accomplished by applying a stratified random sampling method within ERDAS Imagine's Statistical Sample Selection tool in order to increase the possibility of accuracy of the dataset as a whole (Herold, et al., 2003).

After the high resolution classification was completed, they used CART software to determine a classification rule based on the percent surface values that were identified by the sample sites as well as the relationship between various data layers. The CART software was developed with the C language and the ERDAS Macro Language (EML). During the process, validity data were also made in a random selection manner. As a result, each of the target bands had 50,000 training pixels and 50,000 validity pixels (Herold et al., 2003). Then, the training data and validity data were input into the CART software, in which regression trees and a production rule set were formulated. The output predicted values were evaluated based on the training data and validity data. The final outputs were a series of maps showing the impervious surfaces and tree canopies free of clouds. The accuracy level for impervious surfaces was 0.90 while the accuracy level for the tree canopy was 0.93. Although the research of Herold et al. (2003) achieved high accuracy for mapping impervious surfaces and forest canopies, their research is urban impervious surfaces-based only. Their research did not include classifications of several detailed urban LULC categories such as residential, commercial, and industrial.

From the above literature review, systematic classifications of urban LULC classes into USGS Level II and the USGS Level III have not been tested so far. Although Lu and Weng (2004) utilized an altered V-I-S model, i.e., the "Lu-Weng" model, to a certain level, their overall accuracy is not very high. In addition, the research of Herold et al. (2003) did not have a comprehensive classification of urban LULC classes, although they utilized the CART method to classify urban impervious areas. This chapter explores a method to classify urban LULC classes into the USGS Level II and the USGS Level III and simultaneously yields high accuracy. The CART method is explored, and ancillary data are included in the method so as to yield high accuracy.

Spectral Characteristics of Urban LULC

The difficulty of classifying urban LULC classes with high accuracy based on satellite imageries relies on three factors. First, an urban environment consists of a combination of different land covers that may have similar spectral characteristics (Paul, 2007). For example, a suburban residential development may consist of open space, which may be in the form of golf courses, parks, etc., that have similar spectral signatures to the natural, non-urban grasslands or forest covers. Also, some exposed bare land and agricultural land may present similar spectral values to urban impervious areas (Kim, 2007). Second, urban LULC classes are rarely homogenous (Myint et al., 2007) and can be mixed with each other in a satellite image pixel. For example, on a Landsat TM image (30 × 30 meters), an urban road with a vegetation covered median may have vegetation covers mixed with impervious covers such as concrete or asphalt in the same pixel. For single-family residential units, it may have residential rooftops mixed with hard surfaces, swimming pools, exposed soil, or vegetation (Myint et al., 2007). In addition, urban features consist of various types of materials such as glass, concrete, metals, plastics, asphalt, grass, shingles, shrubs, trees, soil, water, etc., which are very complex and have completely different spectral characteristics (Myint, 2001). Third, urban growth in the suburbs is often characterized as low-density single-family development, in which single-family housing units may be set apart from each other at certain distances and where their texture information is beyond the detection of satellite remote sensors. Based on the research of Cowen et al. (1995), for clear identification of an object-of-interest, the minimum resolution of high quality imagery is one-half of the diameter of the smallest object (Cowen et al., 1995; Jensen and Cowen, 1999), which means for a single-family house with 10 meters width, the minimum spatial resolution to identify that single-family house is 5 × 5 meters. This creates a serious obstacle since satellite imageries are not able to provide some important spectral bands for LULC classification based on this resolution. When high resolution imagery is used, e.g., QuickBird and IKONOS, mid-infrared and thermal bands are absent because these high resolution sensors are not equipped with the above specified spectral bands that are critical for urban LULC identifications. According to Jensen and Cowen (1999), certain electromagnetic spectra such as visible color (0.4 – 0.7 μm), near-infrared (0.7 – 1.1 μm), and middle-infrared (1.5 – 2.5 μm) are useful to extract USGS Level III land cover in an urban environment, and the thermal infrared portion of the spectrum (3 – 12 μm) is useful to detect urban temperature (Jensen and Cowen, 1999).

According to Toll (1984), Landsat 4 and 5 TM Band 5 (1.55 – 1.75 μm), Band 6 (10.4 – 12.5 μm), and Band 7 (2.08 – 2.35 μm) can improve the overall discrimination level of urban LULC classes over Landsat MSS because of added spectral bands (Toll, 1984).

Research Question

The leading research question in this chapter is how to develop a method that uses the Landsat TM and ETM+ imageries to classify urban LULC classes into USGS Level II and USGS Level III, and simultaneously yields high accuracy. Based on the current research, it is difficult to classify urban LULC classes into USGS Level II and USGS Level III by reaching an accuracy level of more than 85 percent when using Landsat sensors because of the limitations of the conventional supervised and unsupervised classification methods. How to explore a way that better addresses the spectral characteristics of urban LULC classes, while simultaneously delivering high accuracy results, is a serious research question. This chapter compares and explores the CART method, the V-I-S method, and the conventional supervised methods in order to classify urban LULC classes to the level similar to USGS Level II and USGS Level III.

Methodologies

Classification System

To classify urban LULC classes, a workable classification system should be adopted. Currently, there are two classification systems available to remote sensing fields: one is the USGS Classification System and the other is American Planning Association's (APA) Land Based Classification Standards (LBCS) (Jensen and Cowen, 1999). For urban LULC classes, the USGS Classification System does not specify single-family and multi-family residential classes at Level II, which are the two most important land use categories to measure urban growth in this research, and depends upon the user to define them at Level III. In addition, the USGS Classification System does not define golf courses, zoos, and urban parks at Level II. However, commercial and industrial uses are both defined at Level II. These make the classification of urban LULC classes difficult without having a standard level to incorporate important urban LULCs. On the other hand, the APA's LBCS does not provide land cover and vegetation information in an urban area (Jensen and Cowen, 1999). Therefore, this research adopts a revised USGS Classification System, referenced from the National Land Cover Database (NLCD) classification system. The detailed classes are as follows (LCI, 2007, p.1; Anderson et al., 1976, p.10-21):

SINGLE-FAMILY. This includes areas with a mixture of constructed materials and vegetation. Constructed materials account for 30-80 percent of the cover. Vegetation may account for 20 to 70 percent of the cover. These areas commonly include single-family attached and detached housing units and mobile homes. Population densities will be lower than in the multi-family residential areas.

MULTI-FAMILY. This includes highly developed areas where people reside in high numbers. Examples include apartment complexes, condominiums, and row houses. Vegetation accounts for less than 20 percent of the cover. Constructed materials account for 80 to100 percent of the cover.

COMMERCIAL-INSTITUTIONAL-TRANSPORTATION. This includes areas dominated by retails, sales, services, and offices; various educational, religious, health, correctional, and military facilities; and utility and transportation infrastructures, e.g. power plants, roads, railroads, parking lots, etc.

INDUSTRIAL-WAREHOUSES. This includes a wide range of land uses from light manufacturing to heavy manufacturing plants, which also include mineral mines and warehouses.

GRASSLANDS. These are areas dominated by upland grasses, forbs, and shrubs. In rare cases, herbaceous cover is less than 25 percent but exceeds the combined cover of the woody species present. These areas are not subject to intensive management, but they are often utilized for grazing.

FORESTS. These are areas characterized by tree cover (natural or semi-natural woody vegetation, generally greater than 6 meters tall); tree canopy accounts for 25-100 percent of the cover.

AGRICULTURAL. This is land used primarily for production of food and fiber.

RECREATIONAL AND OTHERS. This includes golf courses, parks, zoos, conservations and reserves, and cemeteries.

WETLANDS. These are areas where the soil or substrate is periodically saturated with or covered with water.

WATER. These are all areas of open water, generally with less than 25 percent cover of vegetation/land cover.

BARREN. These are areas basically characterized as landfills or characterized by bare rock, gravel, sand, silt, clay, or other earthen material, with little or no "green" vegetation present, regardless of its inherent ability to support life. Vegetation, if present, is more widely spaced and scrubby than that in the "green" vegetated categories; lichen cover may be extensive.

The above classification system crosses USGS Classification Level I, Level II, and Level III. According to Jensen and Cowen (1999), the resolution for a remote sensor for USGS Level I is spatial 20 – 100 meters and spectral V-NIR-MIR-Radar; for Level II it is spatial 5 – 20 meters and spectral V-NIR-MIR-Radar; and for Level III it is spatial 1 – 5 meters and Pan-V-NIR-MIR. The Landsat ETM+ imagery is not able to provide the spatial resolution at Level III as suggested by Jensen and Cowen (1999). However, according to Toll (1984), as spectral resolution is more meaningful than spatial resolution in detecting urban LULC classes, the 30 × 30 meters Landsat spatial resolution, for example, can still be used for the urban growth analysis.

Data Inventory and Pilot Area

The satellite data source involved in this section for a pilot area is the 2003 Landsat ETM+ with imagery data acquired on February 11, 2003, determined based on cloud conditions with 0 percent cloud coverage. The training data are acquired from the 2004 digital ortho quarter quads (DOQQs), which is the most recent available for download. A pilot area in Alachua County is chosen to test the CART method, the V-I-S method, and the supervised method. The pilot area, coded as Q4619SW in DOQQ images, is located in central Gainesville, covering an area of 20,700 acres (Figure 2-3). A specific method that accurately classifies urban LULC classes and yields the highest level is selected to classify the urban LULC classes for the entire county. The selection of the pilot area is based on the factors that urban LULC classes and natural LULC classes are intermingled in an area where the sensitivity of a classification method can be tested as to whether it is able to yield high accuracy for all eleven LULC classes. This quadrant was selected because the area is a typical urban area where it covers a variety of important urban and natural landscapes such as the University of Florida campus, adjacent commercial and single-family and multi-family uses, forests, wetlands, industrial uses, shopping centers, agricultural and nurseries, golf courses, parks, ranchland and grassland, and water. As a result, the pilot area comprises all the urban uses and the most natural LULC classes. This is different from the pilot area that is either occupied by all urban LULC classes or by all natural LULC classes. As mentioned above, successful classifications of urban and natural LULC classes in this quadrant help extend the method to the entire county.

Based on the above quadrant, the 2003 ETM+ imagery is masked so that the extent of the ETM+ image is the same as the DOQQ's. The software for this research is ERDAS Imagine 9.1, IDRISI Andes, ENVI 4.4, and ArcGIS 9.3.

CART Method

This chapter tests the CART method based on these two software packages, IDRISI Andes and ENVI 4.4. Huang and Jensen (1997) described the knowledge-based building procedure in remote sensing into three steps: training, decision tree generation, and from decision trees to production rules (p.1,186 – p.1,187). First, for training, it is important to have the learning process begun based on the induction from the training data. Second, a decision tree is generated based on the training data. In this step, a recursive procedure is processed so that no more pixels can be divided in the end. An optimal size of the tree is also determined in this step. Several splitting rules such as the Gini, the Entropy, and the Towing can be applied in order to generate a tree with the least errors in each node (Huang and Jensen, 1997; Breiman et al., 1984; Zambon et al., 2006). Third, a decision tree is transformed to production rules. For example, in a child node, i.e., a leaf, the classified LULC, say, grasslands, can be represented by the form of "ISODATA = 33," which can be transformed from the 'Yes' class equivalent to grasslands.

Similar processes to Huang and Jensen's (1997) approach are undertaken to test the CART method in this research. First, various relevant layers such as the ETM+ layer and ancillary layers are stacked; second, sample regions of interest (ROIs) are collected, which can be collected from the DOQQ training datasets; third, a splitting rule is selected as to whether it is Gini, Entropy, Ratio Gain, or something else; fourth, trees are built based on the sample ROIs as well as the selected splitting rule; fifth, tree is pruned; sixth, classification maps are created; and seventh, accuracies are assessed.

Because of the different splitting rules offered by the two different software packages, in which the ENVI 4.4 RuleGen 1.02 offers the splitting rule called Quick, Unbiased, and Efficient Statistical Tree (QUEST) and IDRISI Andes applies three splitting rules, namely, gain ratio, entropy, and Gini. Splitting rules in the different software packages produce different classification maps with different results. This study finds a splitting rule that best represents the ground truth by yielding the highest accuracy level.

The ENVI 4.4 RuleGen 1.02 has the QUEST module which allows numeric and categorical variables to be stacked together. This is different from IDRSI Andes, in which categorical layers are treated as numeric layers. The QUEST module also provides options to input standard errors (SEs) and cross validations (CVs) for tree pruning. For the splitting rules, the QUEST module provides the choices between univariate and linear. Other options are also provided such as the input of Alpha value, minimum node size, variable selection method, output Pstricks tree, and so on.

For IDRISI Andes, it provides three splitting rules. The gain of a single classification is defined as the entropy after classification X and Gain (X) tests the maximization of the information gain. The entropy algorithm is given to oversplitting since every split can potentially contribute to information gain. The gain ratio algorithm attempts to overcome this potential bias through a normalization process. The entropy rule aims to identify splits where as many groups as possible are divided as precisely as possible and forms groups by minimizing the within group diversity. The Gini splitting rule is a measure of impurity at a given internode that is at the maximum when all pixels are equally distributed among all classes (IDRISI Andes Help, no date).

ENVI 4.4 RuleGen 1.02

For the CART method, based on the above descriptions, an NDVI layer is first developed from the ETM+ image by using the ERDAS Imagine 9.1 software. A Principal Components (PCs) layer and a Tasseled Cap layer are also developed by using the ERDAS Imagine 9.1 software. In addition, an unsupervised classification layer, namely, the ISODATA layer, is created in ERDAS Imagine 9.1, based on 60 iterations following Yang (2000) and Lo and Choi's (2004) research, in which Yang (2000) and Lo and Choi (2004) found that an iteration of 60-80 is optimal for a successful unsupervised classification. These layers are stacked together with three additional layers, which are mentioned below for decision tree building in ENVI 4.4 RuleGen 1.02.

The reason for stacking various layers is to find various variables for decision tree building. It is optimal to find as many variables and/or layers as possible that are related to the research purposes. The above variables and/or layers are used to represent vegetation as well as other variables by simplifying band combinations. For this end, three additional variables and/or layers are found, which are the parcel layer, the parcel assessed value layer, and the parcel residential unit layer. The vector parcel data are sorted based on the parcel land use code and categorized according to the previously defined eleven LULC classes and then rasterized in

ArcGIS 9.3. After this is completed, the land use codes in the parcel data are reclassified into the eleven classes. The reason to include the parcel data into the ancillary data is to reduce the occurrence of pixel confusion given the mixed or spectral similarity for urban LULC classes. This is a very important step since without parcel signatures urban pixels can be easily mixed. For example, the University of Florida campus is a single parcel containing various LULC classes. Without parcel signatures, the LULC classes on the campus such as nurseries, student apartments, recreational facilities, and so on cannot be discriminated accurately based on the 30-meter ETM+ imageries. Therefore, additional parcel signatures must be sought in addition to the existing parcels on the campus. In order to differentiate the spectral similarity of the industrial and commercial uses, a parcel assessed value layer is introduced in addition to the parcel layer. Because industrial and commercial uses share different assessed values, in which industrial uses are assessed much lower in value than commercial uses, the addition of the assessed values into the ancillary data helps identify correctly the industrial and commercial uses. To further differentiate the single-family and multi-family uses, a parcel number of unit layer is included because single-family and multi-family uses have different numbers of units in a parcel. The inclusion of this layer helps clarify the single-family use from the multi-family use because of their spectral similarities. After this is accomplished, the two layers are stacked in ENVI 4.4 RuleGen 1.02 along with the above mentioned NDVI, PCs, Tasseled Cap, ISODATA, and parcel layers.

At the time of layer stacking, sample LULCs are collected from the DOQQ image by using the ROI tool in ENVI 4.4. ROIs are polygons identified by the user to define a sample boundary for each LULC class. The sampling training sites are selected based on the stratified random samples that are generated from the ERDAS Imagine 9.1 software. As a result, seven samples

(257 pixels) are collected for water, five samples (257 pixels) for wetlands, nine samples (1,141 pixels) for urban recreational and others, thirty-four samples (810 pixels) for commercial-institutional-transportation, seventeen samples (1,251 pixels) for multi-family, twenty-nine samples (1,072 pixels) for single-family, nine samples (422 pixels) for industrial-warehouses, forty-nine samples (442 pixels) for forests, and thirty-three samples (210 pixels) for grasslands. These sample polygons are spread throughout the pilot area. Because the barren use is not found on the DOQQ image, there are no ROIs for the barren class. When sample ROIs are collected, they are reconciled to the stacked image because the 1-meter resolution of sample ROIs are different from the 30-meter stacked image generated from the ETM+ images.

A decision tree is built in ENVI 4.4 thereafter based on the reconciled ROIs and the stacked image. The computer picks up DN values from the stacked image and splits these DN values into binary groups based on the rules the computer generates from the training samples. It proceeds as a recursive process until it is not possible to continue further. The algorithm selected in the calculation is the QUEST with other parameters set as default. (Trees are not provided in this research because its size has exceeded the page limits.) No manual pruning is conducted and the pruning process is executed automatically by the computer. After the decision tree is built, it is input into the ENVI 4.4 to map the ten LULC classes (barren not included). When the map is created, 196 random points are selected by using stratified random sampling on the classified image to do an accuracy assessment. These 196 random points are the outcome with an expected accuracy of 85 percent, a Z score of 1.96, and an allowable uncertainty of 0.05. The LULC of random points are verified from the DOQQ image and field trips.

IDRISI Andes

A similar process is undertaken based on the IDRISI Andes software. Because the stacked layer has already been created in ENVI 4.4, it is imported into IDRISI Andes by containing a bunch of sub-layers. After the sub-layers are imported, a vector digitized training sites file is created based on the previously generated ROIs in the ENVI 4.4. Then, a splitting rule is selected so that a tree can be built. The tree pruning process is also conducted automatically. After the tree is built, a classification map is produced, and the map is imported into the ERDAS Imagine 9.1 software for accuracy assessment. The accuracy assessment is processed also based on 196 stratified sampling points. This research tests three splitting rules in the IDRISI Andes software package. They are gain ratio, entropy, and Gini. Because the three trees produce the maps that have apparent accuracy results, the map with the most obviously correct result is used to do an accuracy assessment, which can save research time.

V-I-S Method

This study uses an ETM+ image to extract five components based on the V-I-S model and uses the maximum likelihood method to extract these five components (Hung and Ridd, 2002). Procedurally, first, end members are collected from DOQQ images and later reconciled to the ETM+ images for the five classes as mentioned above using the ROI tool in ENVI 4.4. When enough end members are collected, five rule maps are produced with each rule representing each component. Then, the rule maps are transformed to be exponential and are further adjusted according to the five-class compositions collected from the training sample points on the DOQQ image. As a result, nine quadrant fishnets in 30 by 30 meters dimension are created randomly in ArcGIS 9.3 on the DOQQ images so that sub-pixel spectra can be unmixed and different component percentages within a pixel can be determined. A linear regression model is applied in this regard. After rule maps are adjusted, they are classified into ten classes, namely single-family, multi-family, commercial-institutional-transportation, industrial-warehouses, grasslands, forests, agricultural, recreational and others, wetlands, and water (no barren is found in the pilot area), using the CART method. Finally, 196 stratified random sampling points are found, and accuracy is assessed.

Supervised Method

The supervised method is implemented in the ENVI 4.4 software. The classification is based on the classifier algorithms, which namely are parallelepiped, minimum distance, Mahalanobis distance, and maximum likelihood. Because of the previously created ROIs, the process of the supervised method is fairly straightforward. The Landsat ETM+ imagery with ROIs is input into the computer, and the results using different classifier algorithms are produced. As introduced in the *Literature Review*, because of the spectral characteristics of urban LULC classes, the overall accuracy level using this method is expected to be lower than using the CART method. In fact, after producing maps applying the above four types of classifier algorithms, it is found that the supervised method is unable to distinguish the urban LULC classes for their spectral characteristic for mixed pixels and similar pixels. It is also unable to handle these many urban classes as proposed in this research with respect to high accuracy based on the Landsat ETM+ imageries. Therefore, it is meaningless to measure the individual accuracy level of each supervised classifier algorithm. This research does not go into that direction; instead, this research uses the supervised classification maps as a reference to the maps created by the CART method to see whether the CART method can reveal every detail of urban LULC classes. More details are discussed further in the next part of the chapter.

Pilot Area Results

CART Method

The results of using the CART method to classify LULC classes are quite encouraging. As Figures 2-4, 2-5, 2-6 and 2-7 show, the ENVI 4.4 RuleGen 1.02 QUEST algorithm, the IDRISI Andes gain ratio, entropy, and Gini splitting rules all have good turnouts in classifying urban

LULC classes. On these maps, single-family, multi-family, commercial-institutional-transportation, industrial-warehouses, and recreational-others have excellent classifications with clear boundaries that match soundly with the parcel boundaries and ground truths. In addition, wetlands, water, forests, agricultural, and grasslands—the land covers that are within the parcels—are all nicely presented on these maps. These illustrate that the CART method is an effective tool to handle the data at both the parcel level and at the sub-parcel level.

Comparing the maps produced by the different splitting rules, it is found that there are still slight differences for each rule to map urban LULC classes. This research finds that the ENVI 4.4 QUEST algorithm performs soundly in classifying urban LULC classes with high accuracy. This is due to the fact that the QUEST algorithm takes categorical variables into consideration. For the IDRISI Andes software package, although some studies suggest that the Gini splitting rule yields higher accuracy than other splitting rules, such as the entropy (Zambon et al., 2006) and gain ratio, this study finds that the gain ratio splitting rule outperforms the Gini or the entropy rule from direct visual observations. For example, in the Gini rule, the natural curve of Banias Arm Lake is flattened out by nearby wetlands as well as a nearby recreational use while the gain ratio rule keeps the lake's natural contours in a fairly good shape. This may be due to the fact that the gain ratio rule normalized split gains during node splitting processes.

The overall accuracy of the RuleGen 1.02 QUEST algorithm reaches 87.24 percent based on 196 stratified random sampling points. The specific accuracy breakdowns for each class are presented in Tables 2-1, 2-2, and 2-3. From the tables, it is evident that water and agricultural have the highest accuracy level while forests, single-family, commercial-institutional-transportation, recreational and others, industrial-warehouses, grasslands, and wetlands also have a fairly good accuracy level. The multi-family use has a relatively lower accuracy level, however. This is because the multi-family use can easily be confused with single-family and commercial-institutional-transportation uses. From the accuracy assessment, specifically, single-family is slightly mixed with multi-family, commercial-institutional-transportation, and forests. Commercial-institutional-transportation is slightly mixed with single-family, forests, recreational and others, multi-family, and grassland. Multi-family is slightly mixed with single-family, and commercial-institutional-transportation. Industrial-warehouses is slightly mixed with commercial-institutional-transportation.

The overall accuracy of the IDRISI Andes gain ratio splitting rule was 85.20 percent based on 196 stratified random sampling points. The specific breakdowns for each class are listed in Tables 2-4, 2-5, and 2-6. From the tables, urban LULC classes such as industrial-warehouses, commercial-institutional-transportation, single-family, multi-family, forests, water, and agricultural are nicely classified. Different from the results of the ENVI 4.4 QUEST module, recreational and others, grasslands, and wetlands in IDRISI Andes have a relatively lower accuracy level. Specifically, grasslands are slightly mixed with commercial-institutional-transportation; wetlands are slightly mixed with single-family; and recreational and others are slightly mixed with grasslands, wetlands, and commercial-institutional-transportation.

Overall, the CART method is proficient in classifying urban LULC classes at detailed levels. It is capable of handling both the parcel-level and sub-parcel level data and can well address sub-pixel, spectral similarity, and mixed-pixel issues for urban LULC classifications. In particular, in this study, a large parcel, the University of Florida campus, is found together with the surrounding smaller parcels, but urban LULCs at the sub-campus level are well presented. In many cases, by using the parcel ancillary data, the CART method can mask out the mixed and similar pixels and represent them by a single class. This is especially useful when researchers want to identify residential areas that are covered by forests and grasslands, in which low density single-family units can be easily overlooked by satellite sensors. The identification of the residential areas at the parcel level could avoid underestimates of urban built areas due to vegetation cover.

The overall accuracy levels for using the two software packages are both above the 85 percent threshold. As what is found in this study, the CART method is able to identify sub-parcel land covers. However, the identification of the sub-parcel land covers is mostly suitable for large parcels such as the University of Florida campus. For small parcels, the sub-parcel covers are masked out, however. As compared to the maps produced by the supervised classification method, the CART method is more capable of illustrating urban land uses rather than urban land covers. Contrarily, the conventional supervised method is more able to present urban land covers, however.

V-I-S Method

The results for the V-I-S model are based on the rule maps that are generated from the ENVI 4.4 software. Because there are five classes, namely, bright impervious surface, dark impervious surface, forests, grasslands, and soil, for the V-I-S model, there will be five rule maps generated for these five classes. The original five rule maps are illustrated in Figure 2-8 with the brightest presenting the highest DN values.

Because the rule maps are in a discriminant function format, they will be converted into the exponential function format and then normalized. Therefore, the above Figure 2-8 maps are converted to the maps in Figure 2-9, using the Band Math tool in ENVI 4.4. Based on the nine quadrant fishnets of 30 by 30 meters in dimension that are randomly collected in ArcGIS 9.3 on the DOQQ images, sub-pixel spectra on Figure 2-9 are further unmixed, and different component percentages within a pixel are determined for the rule maps in Figure 2-9. As a result, a linear regression model is applied in this case. The adjusted rule maps based on the linear regression model are extracted in ENVI 4.4 and presented in Figure 2-10; the highest DN values in these maps show each of the five land covers, which are presented by the brightest areas in Figure 2-10. These five land covers are further illustrated in Figure 2-11.

As mentioned above, the final V-I-S classification map is presented in Figure 2-11. From 2-11, it is evident that the single-family use is missing while the multi-family use, the commercial-institutional-transportation use, and the industrial-warehouses use are represented by bright impervious surfaces and dark impervious surfaces. The final classification map should use the CART method in order to yield the eleven LULC classes described in the research question session. The result is presented in Figure 2-12, in which the above four urban uses such as the single-family use, the multi-family use, the commercial-institutional-transportation use, and the industrial-warehouses use are well represented. The accuracy assessment that uses 196 stratified sampling points is summarized in Tables 2-7, 2-8, and 2-9. The overall accuracy level for the V-I-S model using the CART method reaches 74.88 percent.

Supervised Method

A supervised classification method is also conducted based on the maximum likelihood algorithm. Comparing the outcomes of the different classifier algorithms, it is found that the maximum likelihood algorithm produces classification maps that are closer to the ground truths than other algorithms from visual interpretations whereas it still has a disparity from accurate classifications of ground LULC classes (Figure 2-13). From the map, it is clear that some urban features are mixed with each other. For example, agricultural is mixed with grasslands; multi-family is mixed with commercial-institutional-transportation; multi-family is mixed with single-family; industrial-warehouses is mixed with commercial-institutional-transportation; single-family is mixed with forests; and recreational-others is mixed with grasslands, single-family, and agricultural land. However, a forest corridor on the northwestern part of the quadrant can be clearly identified, where creeks are flowing in between. This is beyond the scope of the CART method that uses parcel layers as ancillary data because forests and other LULC features within the residential parcels are masked out.

Pilot Area Summaries of CART Method, V-I-S Method, and Supervised Method

From the above comparisons, the strengths and weaknesses of each method are summarized in Table 2-10. These descriptions provide guidelines of urban LULC classifications for the entire county, which is explained in detail in the next part of the research. Based on the results in Table 2-10, this study uses the CART method to map urban LULC classifications for the entire county.

County-Wide Urban LULC Classifications Using the Preferred CART Methodology

The county-wide urban LULC classifications apply the CART method. The ENVI 4.4 RuleGen 1.02 QUEST module is employed in this regard. The classifications are based on 1982, 1994, and 2003, respectively, using the Landsat TM and ETM+ imageries. The training data use the county 1982 aerial photography image, 1995 and 2004 DOQQ images. Different from the above pilot area study, the layer stacking for the county-wide classifications applies thirteen layers, which is the same number of layers for all three timeframes. These layers include (1) the Landsat TM or ETM+ layer, (2) the Tasseled Cap layer, (3) the PCs layer, (4) the NDVI layer, (5) the ISODATA layer, (6) the parcel layer, (7) the residential versus commercial-institutional-transportation parcel layer, (8) the single-family versus multi-family parcel layer, (9) the single-family versus natural land parcel layer; (10) the commercial-institutional-transportation layer versus natural land parcel layer, (11) the forests versus agricultural land parcel layer, (12) the recreational-others versus agricultural land parcel layer, and (13) the grasslands versus agricultural land parcel layer. The residential versus commercial-institutional-transportation parcel layer is the one that uses the codes to differentiate between the residential use and the commercial-institutional-transportation use in the parcel data. It is the same for the layers of (8) to (13) in order to differentiate these uses by the parcel codes. The above layers are empirically tested in order to compare the accuracy levels for each combination of layer stacking before and after a certain parcel layer is added. This effort is made in order to reduce the pixel confusions given by the raw TM and ETM+ data because pixel confusions can happen between residential uses and the commercial-institutional-transportation use, the single-family use and the multi-family use, commercial-institutional-transportation use and the natural land, forests and the agricultural land, the recreational-others land and the agricultural land, and grasslands and the agricultural land. The ISODATA layer is operated based on 60 iterations that are the same as the classification method tested for the pilot area based on the relevant research results. In addition, the parcel layers from (6) to (13) are all categorical datasets. The layers from (1) to (5) above are continuous datasets, however. In particular, for the layers from the above (6) to (13), 1 is coded for one category, and 10 is coded for the other category. The coding is properly adjusted before these layers are added into the layer stacking, and they are selected as categorical datasets in the ENVI 4.4 RuleGen 1.02 QUEST module. The remaining settings use default ones in the ENVI RuleGen 1.02 QUEST module.

Because the initial ROI signatures create a large quantity of data to run in the computer, which overburden ENVI 4.4, reduced ROI signatures are sought based on the random sampling technique. As a result, for the 2003 classification maps, after the adoption of the random sampling techniques, there are 10,952 pixels for the single-family use, 10,247 pixels for the multi-family use, 11,238 pixels for the commercial-institutional-transportation use, 5,878 pixels for the industrial-warehouses use, 35,013 pixels for the grasslands use, 37,751 pixels for the forests use, 12,827 pixels for the agricultural use, 12,340 pixels for the recreational-others use, 40,611 pixels for the wetlands use, 11,346 pixels for water, and 1,067 pixels for the barren use. Similarly, for the 1994 classification maps, there are 21,224 pixels for the single-family use, 6,891 pixels for the multi-family use, 11,195 pixels for the commercial-institutional-transportation use, 4,122 pixels for the industrial-warehouses use, 38,043 pixels for the grasslands use, 30,350 pixels for the forests use, 16,072 pixels for the agricultural use, 12,303 pixels for the recreational-others use, 41,953 pixels for the wetlands use, 12,514 pixels for water, and 695 pixels for the barren use. For the 1982 classification maps, there are 12,812 pixels for the single-family use, 4,365 pixels for the multi-family use, 8,889 pixels for the commercial-institutional-transportation use, 2,656 pixels for the industrial-warehouses use, 58,960 pixels for the grasslands use, 48,045 pixels for the forests use, 14,922 pixels for the agricultural use, 11,902 pixels for the recreational-others use, 15,690 pixels for the wetlands use, 10,080 pixels for water, and 695 pixels for the barren use. The ROI pixels of LULC classes in different timeframes are illustrated in Table 2-20. The Central Processing Unit (CPU) time for producing each classification map for the three timeframes is 15 to 20 hours to finish a CART tree, which is later converted into classification maps using the trees generated. The final 1982 classification map is illustrated in Figure 2-14. The final 1994 classification map is illustrated in Figure 2-15. The final 2003 classification map is illustrated in Figure 2-16. The accuracy assessments for each of the three timeframes are based on 1,100 stratified random sample points, which are 100 points for each class, and they are generated in ERDAS Imagine 9.1. These accuracy assessments are illustrated in Tables 2-11, 2-12, 2-13, 2-14, 2-15, 2-16, 2-17, 2-18, and 2-19, respectively.

Table 2-1. Error matrix (ENVI 4.4 RuleGen 1.02 QUEST algorithm)

	Reference Data										
Classified Data	Single-Family	Commercial-Institutional-Transportation	Forests	Recreational-Others	Multi-Family	Grasslands	Water	Industrial-Warehouses	Wetlands	Agricultural	Row Total
Single-Family	33	1	1	0	1	0	0	0	0	0	36
Commercial-Institutional-Transportation	1	32	2	1	1	1	0	1	1	0	40
Forests	1	0	46	0	1	1	0	1	0	0	50
Recreational-Others	0	0	1	8	0	0	0	0	0	0	9
Multi-Family	2	3	0	0	13	0	0	0	0	0	18
Grasslands	0	1	0	0	0	8	0	0	0	0	9
Water	0	0	0	0	0	1	18	0	1	0	20
Industrial-Warehouses	0	1	0	0	0	0	0	10	0	0	11
Wetlands	0	0	0	0	0	0	0	0	2	0	2
Agricultural	0	0	0	0	0	0	0	0	0	1	1
Column Total	37	38	50	9	16	25	18	12	4	1	196

Table 2-2. Accuracy totals (ENVI 4.4 RuleGen 1.02 QUEST algorithm)

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Single-Family	37	36	33	89.19%	91.67%
Commercial-Institutional-Transportation	38	40	32	84.21%	80.00%
Forests	50	50	46	92.00%	92.00%
Recreational-Others	9	9	8	88.89%	88.89%
Multi-Family	16	18	13	81.25%	72.22%
Grasslands	11	9	8	72.73%	88.89%
Water	18	20	18	100.00%	90.00%
Industrial-Warehouses	12	11	10	83.33%	90.91%
Wetlands	4	2	2	50.00%	100.00%
Agricultural	1	1	1	100.00%	100.00%
Totals	196	196	171		

Overall classification accuracy = 87.24%

Table 2-3. KAPPA (K^) statistics (ENVI 4.4 RuleGen 1.02 QUEST algorithm)

Class Name Conditional Kappa for each category	Kappa
Single-Family	0.8973
Commercial-Institutional-Transportation	0.7519
Forests	0.8926
Recreational-Others	0.8835
Multi-Family	0.6975
Grasslands	0.8823
Water	0.8899
Industrial-Warehouses	0.9032
Wetlands	1.0000
Agricultural	1.0000
Overall Kappa Statistics	0.8473

Table 2-4. Error matrix (IDRISI Andes Ratio Gain rule)

	Reference Data										
Classified Data	Industrial-Warehouses	Single-Family	Recreational-Others	Forests	Commercial-Institutional-Transportation	Multi-Family	Water	Grasslands	Wetlands	Agricultural	Row Total
Industrial-Warehouses	12	0	0	0	0	0	0	1	0	0	13
Single-Family	0	33	0	0	0	2	0	0	0	0	35
Recreational-Others	0	0	6	0	1	0	0	2	1	0	10
Forests	1	1	0	42	1	0	0	1	2	0	48
Commercial-Institutional-Transportation	0	0	0	1	30	0	0	0	0	0	31
Multi-Family	0	0	0	0	6	13	0	0	0	0	19
Water	0	0	0	0	0	0	19	0	0	0	19
Grasslands	0	0	0	0	2	0	0	7	0	0	9
Wetlands	0	1	1	5	0	0	0	0	4	0	11
Agricultural	0	0	0	0	0	0	0	0	0	1	1
Column Total	13	35	7	48	40	15	19	11	7	1	196

Table 2-5. Accuracy totals (IDRISI Andes Ratio Gain rule)

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Industrial-Warehouses	13	13	12	92.31%	92.31%
Single-Family	35	35	33	94.29%	94.29%
Recreational-Others	7	10	6	85.71%	60.00%
Forests	48	48	42	87.50%	87.50%
Commercial-Institutional-Transportation	40	31	30	75.00%	96.77%
Multi-Family	15	19	13	86.67%	68.42%
Water	19	19	19	100.00%	100.00%
Grasslands	11	9	7	63.64%	77.78%
Wetlands	7	11	4	57.14%	36.36%
Agricultural	1	1	1	100.00%	100.00%
Totals	196	196	167		

Overall classification accuracy = 85.20%

Table 2-6. KAPPA (K^) statistics (IDRISI Andes Ratio Gain rule)

Class Name Conditional Kappa for each category	Kappa
Industrial-Warehouses	0.9176
Single-Family	0.9304
Recreational-Others	0.5852
Forests	0.8345
Commercial-Institutional-Transportation	0.9595
Multi-Family	0.6580
Water	1.0000
Grasslands	0.7646
Wetlands	0.3401
Agricultural	1.0000
Overall Kappa Statistics	0.8256

Table 2-7. Error matrix (CART method V-I-S model)

Classified Data	Reference Data Agricultural	Water	Wetlands	Recreational-Others	Commercial-Institutional-Transportation	Multi-Family	Single-Family	Industrial-Warehouses	Forests	Grasslands	Row Total
Water	0	7	0	0	0	0	0	0	0	0	7
Wetlands	0	0	3	0	0	0	0	0	0	0	3
Recreational-Others	0	0	0	5	0	0	0	0	1	1	7
Commercial-Institutional-Transportation	1	0	0	0	41	1	0	0	5	3	51
Multi-Family	0	0	0	0	5	12	3	0	4	0	24
Single-Family	0	0	0	0	3	7	63	1	9	0	83
Industrial-Warehouses	0	0	0	0	0	0	0	5	0	0	5
Forests	0	0	0	0	3	1	3	0	22	1	30
Grassland	0	0	0	0	0	1	0	0	0	0	1
Agricultural	0	0	0	0	0	0	0	0	0	0	0
Column Total	1	7	3	5	52	22	69	6	41	5	211

Table 2-8. Accuracy totals (CART method V-I-S model)

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Water	7	7	7	100.00%	100.00%
Wetlands	3	3	3	100.00%	100.00%
Recreational-Others	5	7	5	100.00%	71.43%
Commercial-Institutional-Transportation	52	51	41	78.85%	80.39%
Multi-Family	22	24	12	54.55%	50.00%
Single-Family	69	83	63	91.30%	75.90%
Industrial-Warehouses	6	5	5	100.00%	71.43%
Forests	41	30	22	91.30%	75.90%
Grasslands	5	1	0	0.00%	0.00%
Agricultural	1	0	0	0.00%	0.00%
Totals	211	211	158		

Overall classification accuracy = 74.88%

Table 2-9. KAPPA (K^) statistics (CART method V-I-S model)

Class Name Conditional Kappa for each category	Kappa
Single-Family	0.6419
Commercial-Institutional-Transportation	0.7398
Forests	0.6690
Recreational-Others	0.7074
Multi-Family	0.4418
Grasslands	-0.0243
Industrial-Warehouses	1.0000
Agricultural	0.0000
Water	1.0000
Wetlands	1.0000
Overall Kappa Statistics	0.6735

Table 2-10. Comparisons of the strengths and weaknesses of urban LULC classification methods

LULC Classification Method	Purpose	Strengths	Weaknesses
CART Method	A knowledge based method. Some researchers apply it to classify urban LULC classes that may have mixed pixels in Landsat TM and ETM+ images. It is used to yield high accuracy for land use classifications. For urban research, current research has only applied it to classify urban impervious surfaces as opposed to natural areas. However, it is capable of classifying LULC to USGS Level III concerning Landsat TM and ETM+ sensors. The author's research includes classifying urban LULC using Landsat TM and ETM+ sensors equivalent to USGS Level III.	Is able to deal with mixed-pixels because it is not a statistically based approach such as minimum distance; Performs well at the sub-pixel level because it circumvents the urban sub-pixel issue by applying ancillary data; Yields high accuracy; Can incorporate either numerical or categorical data to be included as ancillary data; Is capable of classifying urban LULC classes equivalent to USGS Level III; Raw data can be Landsat data and does not necessarily have to have high resolution data; Best in classifying urban LULC classes at the county level, which is beyond the capability of higher resolution sensors to handle because of excessive amount of data;	Relies on ancillary data; Needs to test different ancillary data and finds the most optimal combinations for layer stacking; Still needs conventional methods to help create ancillary data, such as ISODATA, NDVI, PCs, Tasseled Cap, and so on; Needs to combine with other methods such as GIS for image processing; and Makes it easier to find accurate classifications of land uses rather than land covers if parcel data are used as ancillary data (this is because the parcel layer will mask out land cover details).

Table 2-10. Continued

LULC Classification Method	Purpose	Strengths	Weaknesses
		Is able to utilize maximally the advantages of Landsat data (band 1 through band 7) for urban research in terms of spectral characteristics of urban LULC classes; Has default settings that are good enough; does not need to know complicated algorithms; can perform nicely based on fairly user-friendly interface; and Does not need to trim trees, and automatic tree pruning performs well to yield high accuracy.	
V-I-S Method	A sub-pixel LULC method designed to address urban mixed pixels in terms of Landsat TM and ETM+ sensors. It classifies urban LULC classes based on three elements: vegetation, impervious surface, and soil. An urban LULC is determined based on the percentage of each of three elements in a pixel. As a result, end members of vegetation, impervious surface, and soil need to be extracted from satellite images.	Attempts to deal with urban mixed pixels; Simplifies multiple LULC categories into three; Is capable of dealing with mixed pixels for urban LULC classes simply based on impervious surfaces and yield high accuracy; and Has the potential to classify multiple vegetation covers.	Needs to combine with other methods, such as Artificial Intelligence (AI) to determine actual classes in light of mixed pixels; Is not an easy method to retrieve end members in an accurate fashion because of the technical difficulty; Has weak performance to classify different urban uses such as single-family, multi-family, commercial, and industrial;

Table 2-10. Continued

LULC Classification Method	Purpose	Strengths	Weaknesses
			Has relatively low accuracy to classify multiple urban uses; May still need to use the CART method to classify multiple urban classes; Simplifies use classes, which causes loss of pixel abundance in terms of V-I-S elements; Cannot deal with water or wetland. Has difficulty in defining the urban uses considering their percentages of three elements (there is technical difficulty in classifying multiple urban LULC classes based on the referencing diagram); Has difficulty in obtaining soil end members in urban environment (using mineral indexes to sort out soil class does not guarantee to yield high accuracy in urban LULC classification).

Table 2-10. Continued

LULC Classification Method	Purpose	Strengths	Weaknesses
Supervised and Unsupervised (Or Hybrid) Method	The conventional supervised and unsupervised method are applied in multiple fields, including the areas of urban and non-urban research. The supervised and unsupervised methods are the bases of all LULC classification methods, which is the most flexible method for all circumstances. The hybrid method utilizes the advantages of supervised and unsupervised methods. For urban research, it is still difficult to handle mixed pixels. As a result, it is a per-pixel based method. Research finds that the hybrid method combining with other method such as fuzzy logic works better.	Supervised method usually yields higher accuracy than unsupervised method (Short, no date); Is not required to be familiar with spectral characteristics of ground truths (Kramber and Morse, 1994); the unsupervised method can help; Addresses spectral similarities to some degree; has classes that are beyond researchers' knowledge, which can be identified (Liou and Yang, no date) using hybrid method; Uses ISODATA that is able to differentiate needed classes if based on optimal numbers of classes and iterations; Is more flexible in including fuzzy (soft) classification method into LULC classes; and Works better if combined with band ratio or vegetation index (or other index) images (Banman, no date).	It is not clear whether the hybrid method yields high accuracy in terms of urban research; It is less capable of dealing with urban mixed pixels and spectral similarity issues; It still needs to be familiar with the study area for supervised signatures; otherwise, supervised signatures may contain other uses and covers; and It needs to be a professional who is an expert on supervised and unsupervised methods.

Table 2-11. Error matrix (1982 classification map)

	Reference Data											
Classified Data	Single-Family	Multi-Family	Commercial-Institutional-Transportation	Industrial-Warehouses	Grasslands	Forests	Agricultural	Recreational-Others	Wetlands	Water	Barren	Row Total
Single-Family	33	0	0	0	2	5	0	0	0	0	0	40
Multi-Family	0	4	0	0	0	0	0	0	0	0	0	4
Commercial-Institutional-Transportation	0	0	30	0	4	2	0	0	0	0	0	36
Industrial-Warehouses	0	0	0	3	0	0	0	0	0	0	0	3
Grasslands	3	0	2	0	343	16	0	0	2	0	0	366
Forests	2	0	2	0	11	405	0	0	12	0	0	432
Agricultural	0	0	1	0	0	1	55	0	0	0	0	57
Recreational-Others	0	0	0	0	0	1	0	64	3	0	0	68
Wetlands	0	0	0	0	0	1	0	0	40	0	0	41
Water	0	0	0	0	0	0	0	0	0	52	0	52
Barren	0	0	0	0	0	0	0	0	0	0	1	1
Column Total	38	4	35	3	360	431	55	64	57	52	1	1,100

Table 2-12. Accuracy totals (1982 classification map)

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Single-Family	38	40	33	86.84%	82.50%
Multi-Family	4	4	4	100.00%	100.00%
Commercial- Institutional-Transportation	35	36	30	85.71%	83.33%
Industrial-Warehouses	3	3	3	100.00%	100.00%
Grasslands	360	366	343	95.28%	93.72%
Forests	431	432	405	93.97%	93.75%
Agricultural	55	57	55	100.00%	96.49%
Recreational-Others	64	68	64	100.00%	94.12%
Wetlands	57	41	40	70.18%	97.56%
Water	52	52	52	100.00%	100.00%
Barren	1	1	1	100.00%	100.00%
Totals	1,100	1,100	1,030		

Overall classification accuracy = 93.64%

Table 2-13. KAPPA (K^) statistics (1982 classification map)

Class Name Conditional Kappa for each category	Kappa
Single-Family	0.8187
Multi-Family	1.0000
Commercial-Institutional-Transportation	0.8279
Industrial-Warehouses	1.0000
Grasslands	0.9066
Forests	0.8972
Agricultural	0.9631
Recreational-Others	0.9375
Wetlands	0.9743
Water	1.0000
Barren	1.0000
Overall Kappa Statistics	0.9122

Table 2-14. Error matrix (1994 classification map)

	Reference Data											
Classified Data	Single-Family	Multi-Family	Commercial-Institutional-Transportation	Industrial-Warehouses	Grasslands	Forests	Agricultural	Recreational-Others	Wetlands	Water	Barren	Row Total
Single-Family	69	1	5	0	1	4	0	0	1	0	0	81
Multi-Family	0	5	0	0	0	0	0	0	0	0	0	5
Commercial-Institutional-Transportation	0	0	13	0	1	0	0	0	0	0	0	14
Industrial-Warehouses	0	0	0	4	1	0	0	0	0	0	0	5
Grasslands	1	0	4	0	212	15	1	0	1	0	0	234
Forests	5	0	1	0	19	394	2	0	13	0	0	434
Agricultural	0	0	0	0	0	3	50	0	0	0	0	53
Recreational-Others	0	0	0	0	0	0	0	53	2	0	0	55
Wetlands	0	0	1	0	8	10	0	0	143	2	0	164
Water	0	0	0	0	0	0	0	0	0	54	0	54
Barren	0	0	0	0	0	0	0	0	0	0	1	1
Column Total	75	6	24	4	242	426	55	53	160	56	1	1,100

Table 2-15. Accuracy totals (1994 classification map)

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Single-Family	75	81	69	92.00%	85.19%
Multi-Family	6	5	5	83.33%	100.00%
Commercial- Institutional-Transportation	24	14	13	54.17%	92.86%
Industrial-Warehouses	4	5	4	100.00%	80.00%
Grasslands	242	234	212	87.60%	90.60%
Forests	426	434	394	92.49%	90.78%
Agricultural	53	53	50	94.34%	94.34%
Recreational-Others	53	55	53	100.00%	96.36%
Wetlands	160	164	143	89.38%	87.20%
Water	56	54	54	96.43%	100.00%
Barren	1	1	1	100.00%	100.00%
Totals	1,100	1,100	998		

Overall classification accuracy = 90.73%

Table 2-16. KAPPA (K^) statistics (1994 classification map)

Class Name Conditional Kappa for each category	Kappa
Single-Family	0.8410
Multi-Family	1.0000
Commercial-Institutional-Transportation	0.9270
Industrial-Warehouses	0.7993
Grasslands	0.8795
Forests	0.8496
Agricultural	0.9405
Recreational-Others	0.9618
Wetlands	0.8502
Water	1.0000
Barren	1.0000
Overall Kappa Statistics	0.8790

Table 2-17. Error matrix (2003 classification map)

Classified Data	Reference Data: Single-Family	Multi-Family	Commercial-Institutional-Transportation	Industrial-Warehouses	Grasslands	Forests	Agricultural	Recreational-Others	Wetlands	Water	Barren	Row Total
Unclassified	0	0	0	0	0	1	0	0	0	0	0	1
Single-Family	80	1	2	0	3	7	0	0	0	0	0	93
Multi-Family	1	9	3	0	0	1	0	0	0	0	1	15
Commercial-Institutional-Transportation	0	0	15	0	0	0	0	1	0	0	0	16
Industrial-Warehouses	0	0	0	6	0	0	0	0	0	0	0	6
Grasslands	1	0	0	0	203	18	0	0	8	0	0	230
Forests	5	0	1	0	19	378	1	0	15	0	0	419
Agricultural	0	0	0	0	1	0	56	0	0	0	0	57
Recreational-Others	0	0	0	0	0	0	0	51	2	0	0	53
Wetlands	0	1	0	0	2	7	0	2	152	2	0	166
Water	0	0	0	0	0	0	0	0	1	42	0	43
Barren	0	0	0	0	0	0	0	0	0	0	1	1
Column Total	87	11	21	6	228	412	57	54	178	44	2	1,100

Table 2-18. Accuracy totals (2003 classification map)

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Single-Family	87	93	80	91.95%	86.02%
Multi-Family	11	15	9	81.82%	60.00%
Commercial- Institutional-Transportation	21	16	15	71.43%	93.75%
Industrial-Warehouses	6	6	6	100.00%	100.00%
Grasslands	228	230	203	89.04%	88.26%
Forests	412	419	378	91.75%	90.21%
Agricultural	57	57	56	98.25%	98.25%
Recreational-Others	54	53	51	94.44%	96.23%
Wetlands	178	166	152	85.39%	91.57%
Water	44	43	42	95.45%	97.67%
Barren	2	1	1	50.00%	100.00%
Totals	1,100	1,100	993		

Overall classification accuracy = 90.27%

Table 2-19. KAPPA (K^) statistics (2003 classification map)

Class Name Conditional Kappa for each category	Kappa
Single-Family	0.8482
Multi-Family	0.5960
Commercial-Institutional-Transportation	0.9363
Industrial-Warehouses	1.0000
Grasslands	0.8519
Forests	0.8436
Agricultural	0.9815
Recreational-Others	0.9603
Wetlands	0.8994
Water	0.9758
Barren	1.0000
Overall Kappa Statistics	0.8746

Table 2-20. ROI pixels of LULC classes in 1982, 1994, and 2003 (pixels)

LULC	2003	1994	1982
Single-Family	10,952	21,224	12,812
Multi-Family	10,247	6,891	4,365
Commercial-Institutional-Transportation	11,238	11,195	8,889
Industrial-Warehouses	5,878	4,122	2,656
Grasslands	35,013	38,043	58,960
Forests	37,751	30,350	48,045
Agricultural	12,827	16,072	14,922
Recreational-Others	12,340	12,303	11,902
Wetlands	40,611	41,953	15,690
Water	11,346	12,514	10,080
Barren	1,067	695	695

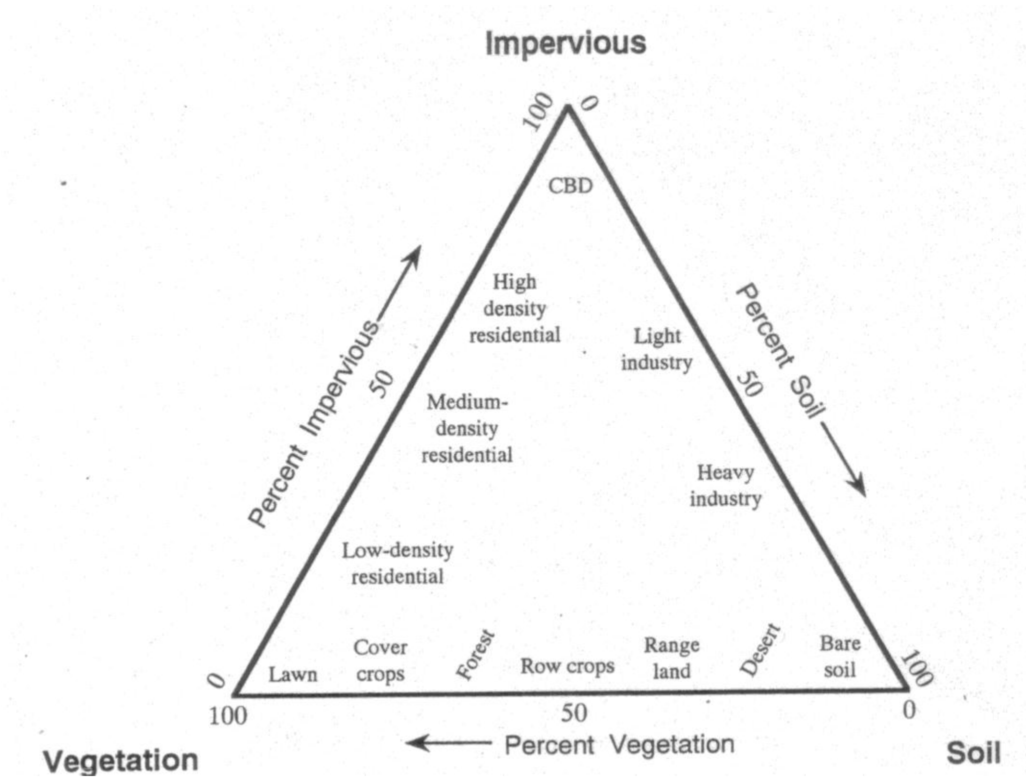

Figure 2-1. Ridd's (1995) V-I-S model (Ridd, 1995. p.2,173)

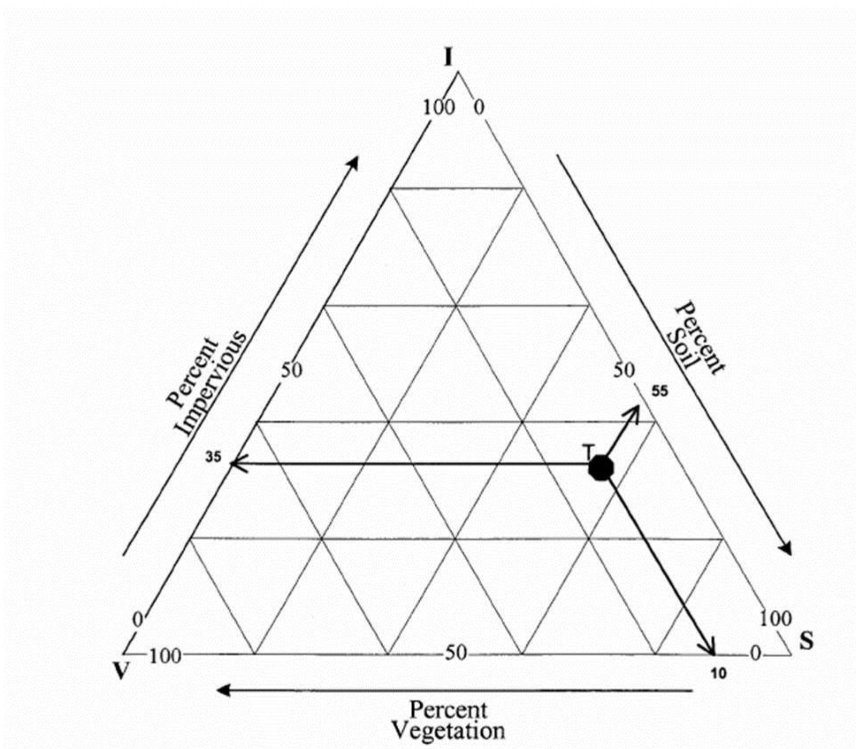

Figure 2-2. V-I-S model and point T (Hung, 2003, p.13; Hung and Ridd, 2002, p.1,174)

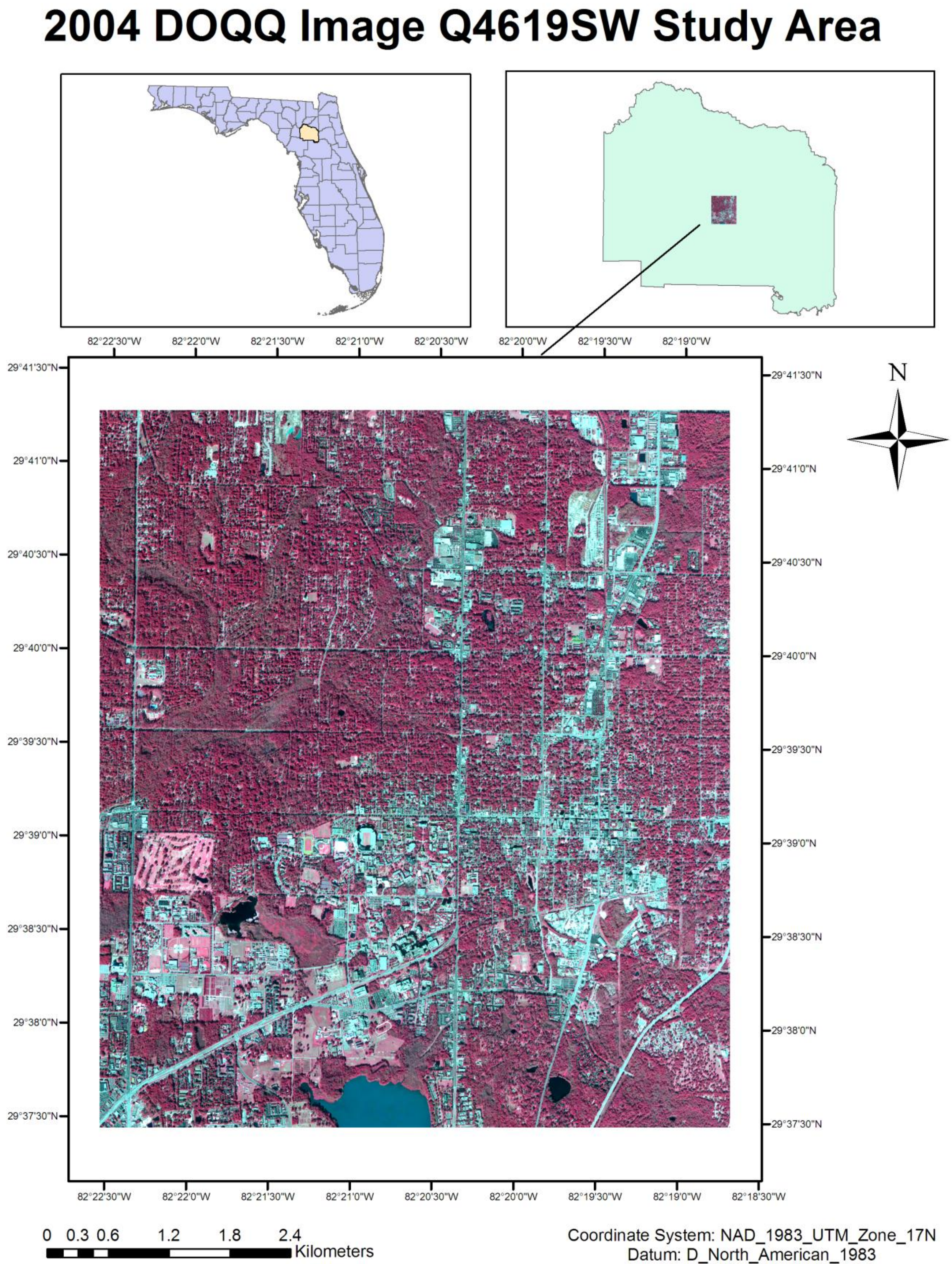

Figure 2-3. Pilot area location for LULC classification

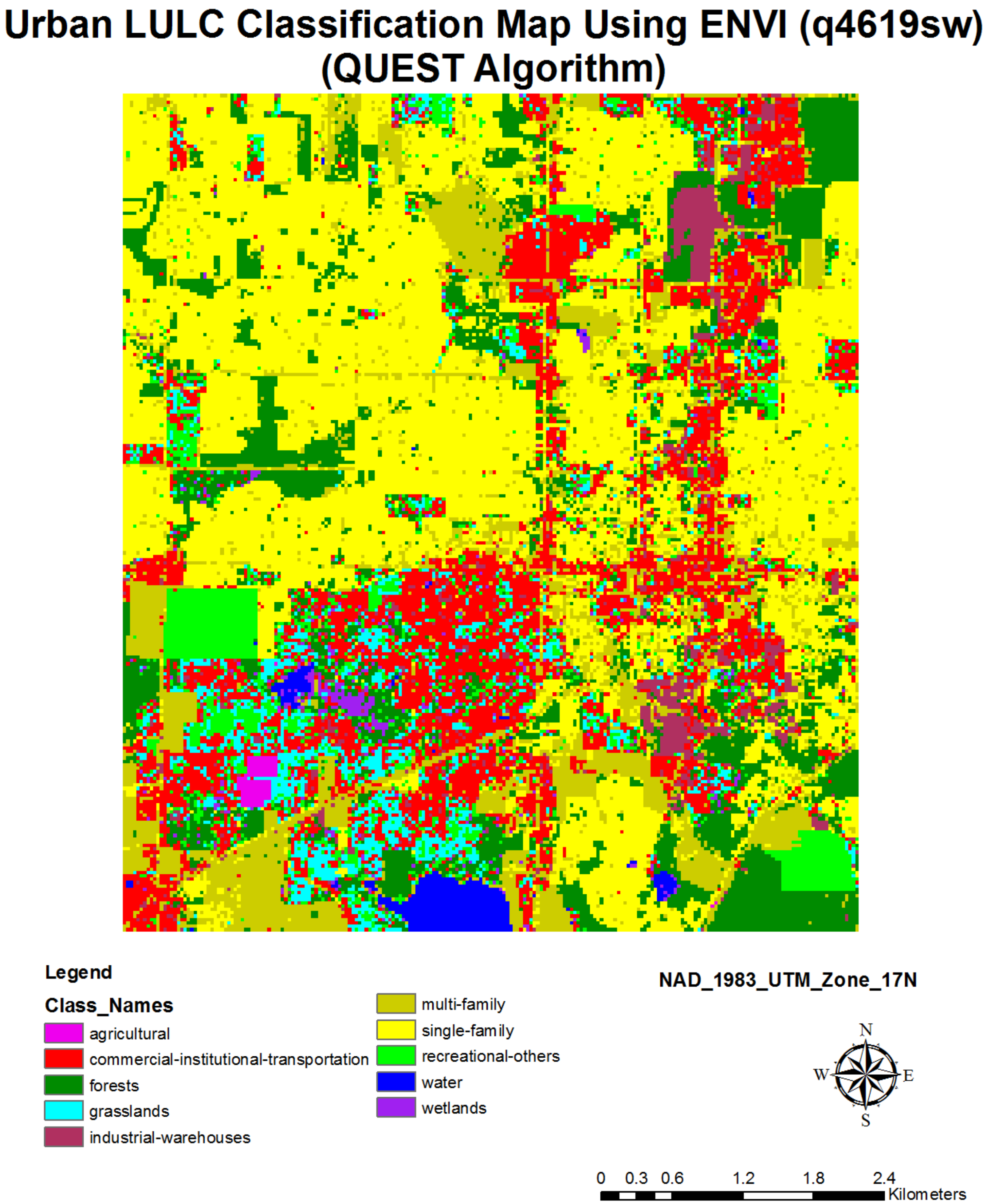

Figure 2-4. Urban LULC classifications using ENVI 4.4 RuleGen 1.02 QUEST module

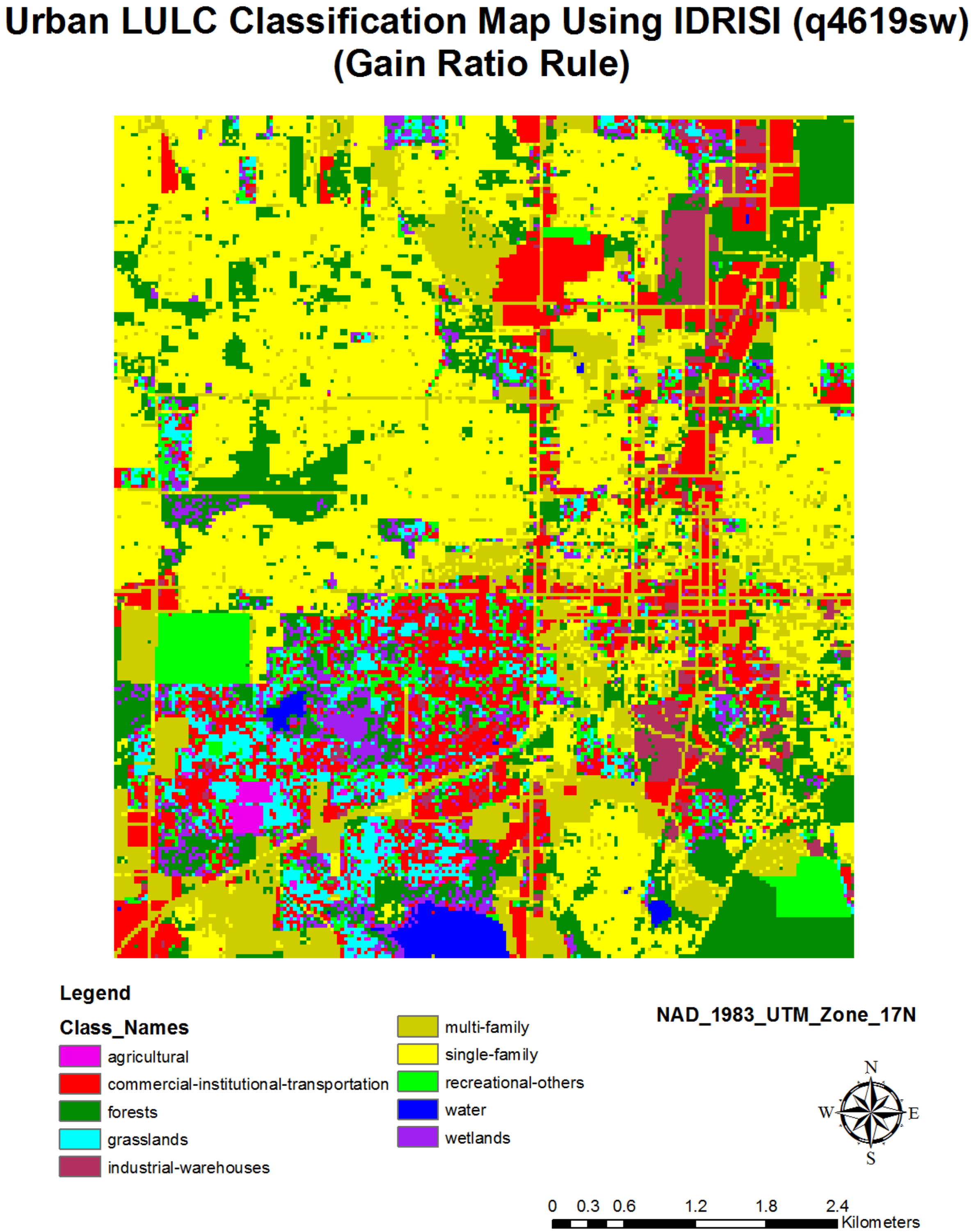

Figure 2-5. Urban LULC classifications using IDRISI Andes gain ratio rule

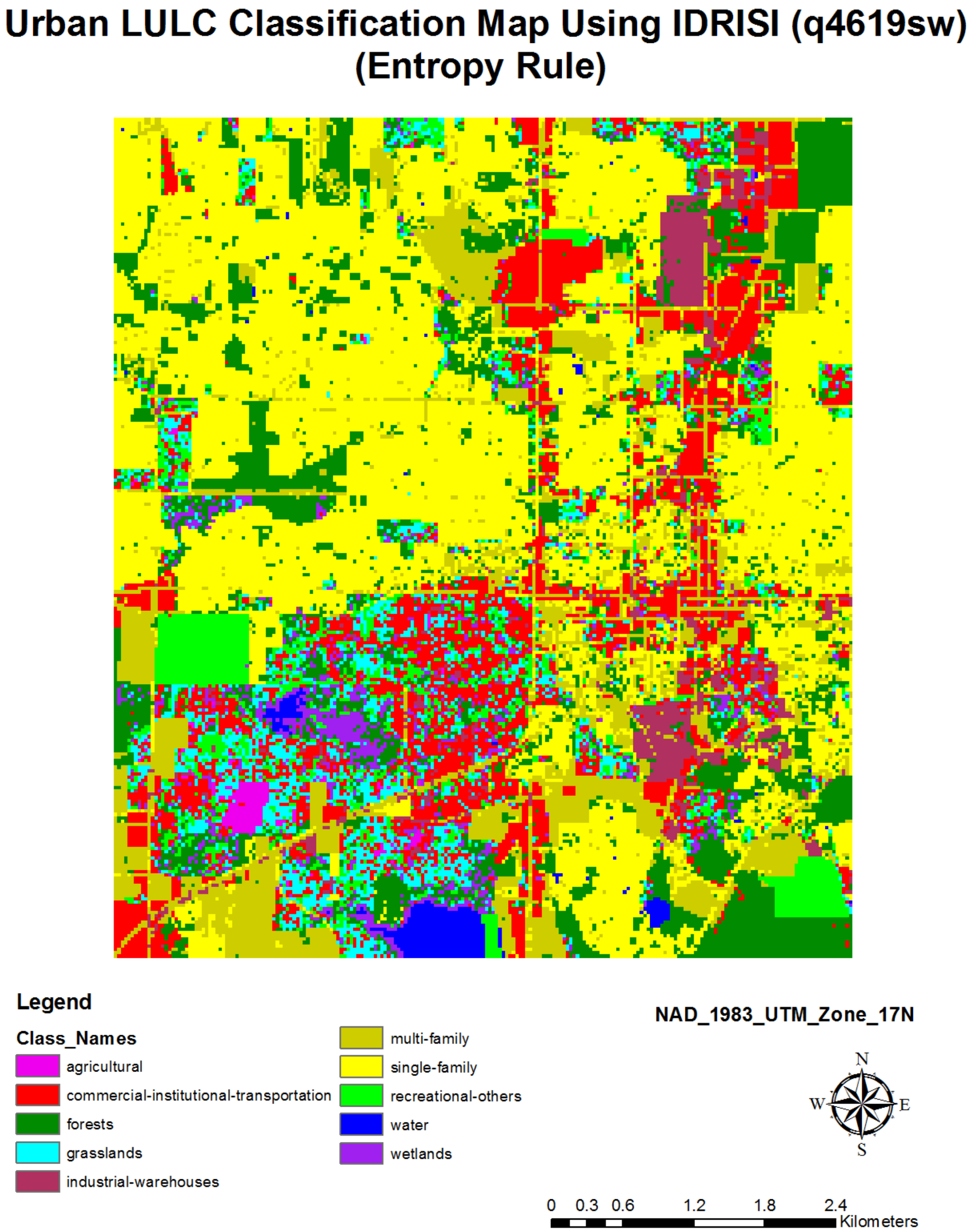

Figure 2-6. Urban LULC classifications using IDRISI Andes entropy rule

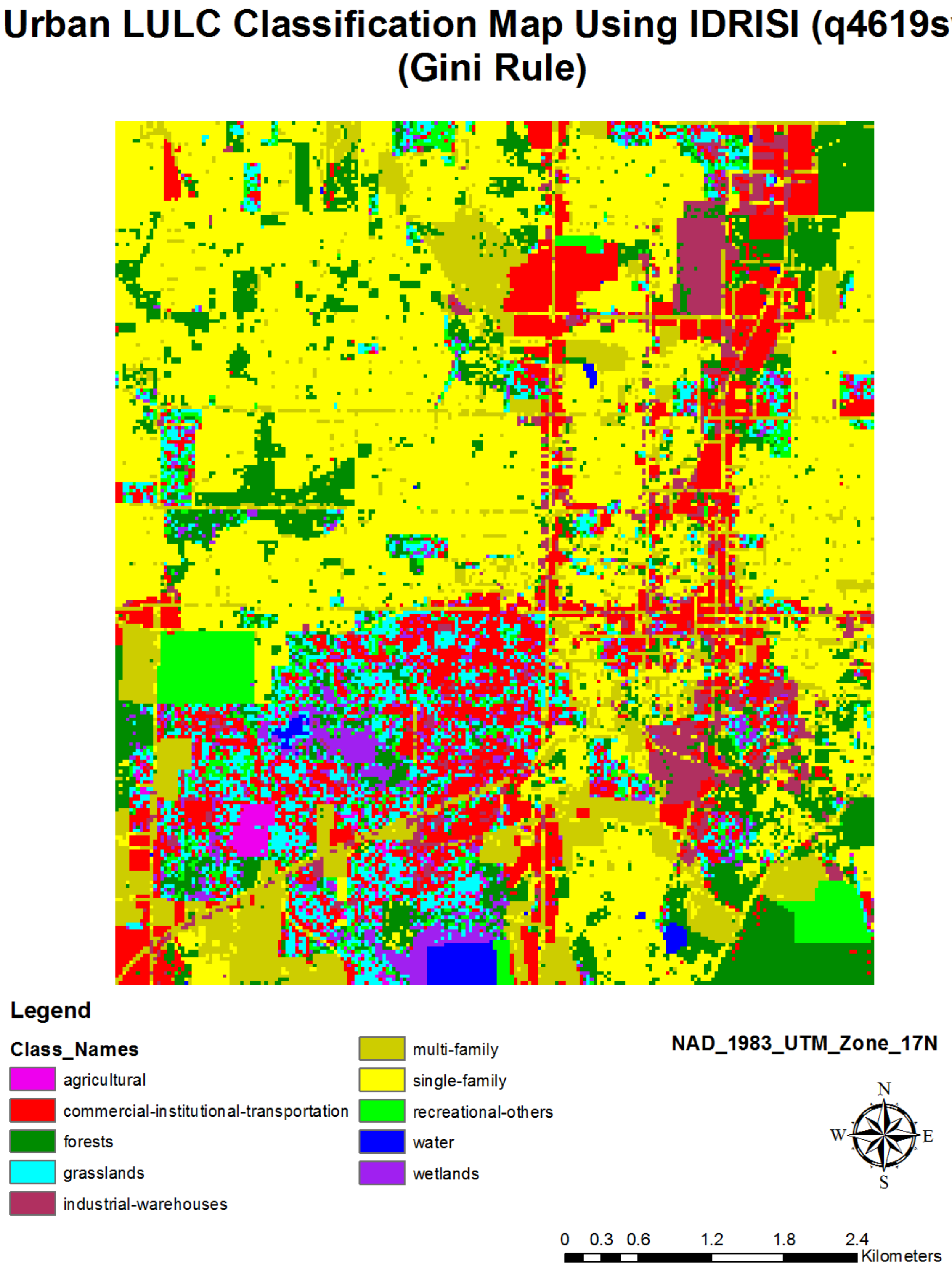

Figure 2-7. Urban LULC classifications using IDRISI Andes Gini rule

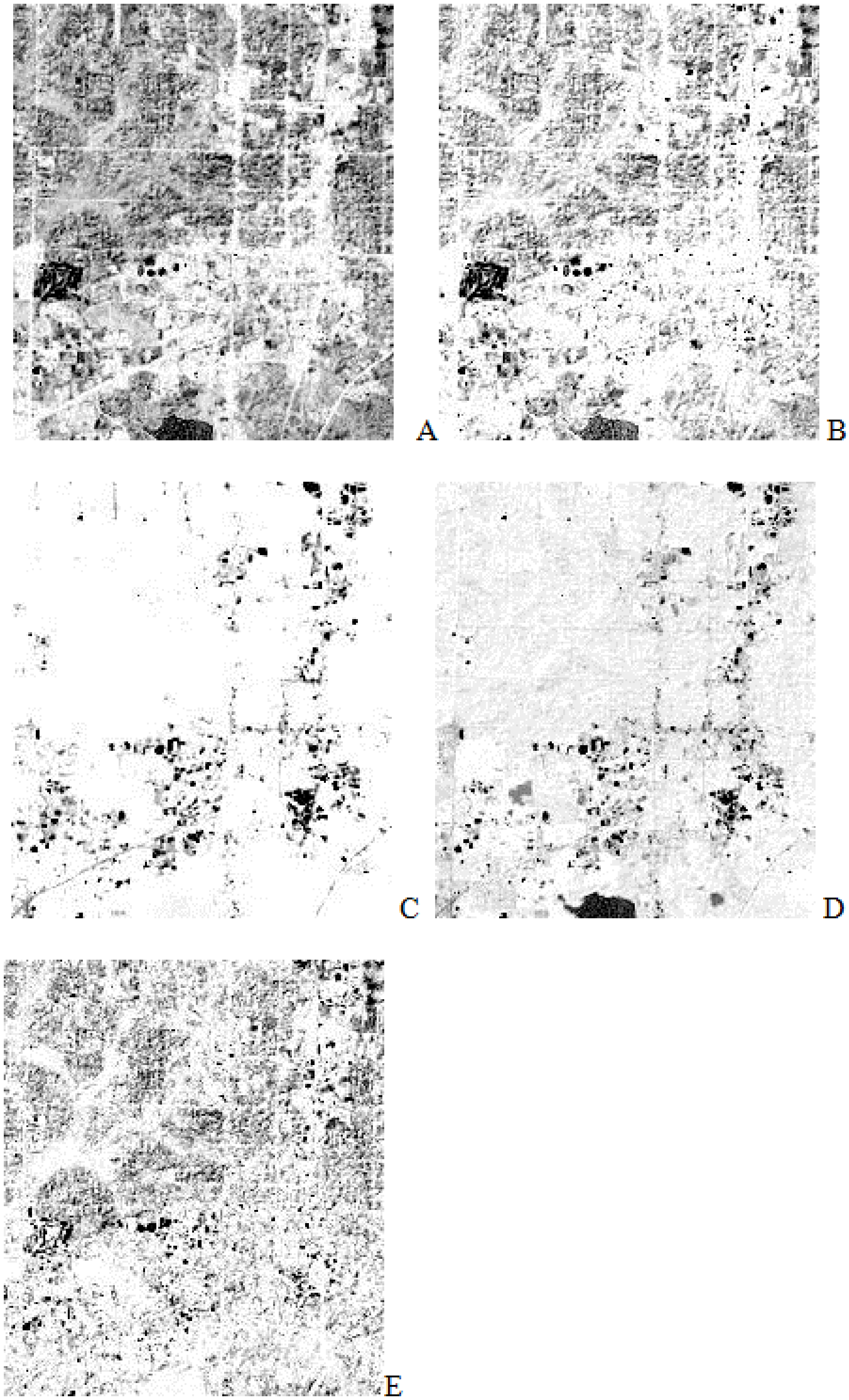

Figure 2-8. Original maps for five components. A) Bright Impervious Surface, B) Dark Impervious Surface, C) Forests, D) Grasslands, and E) Soil.

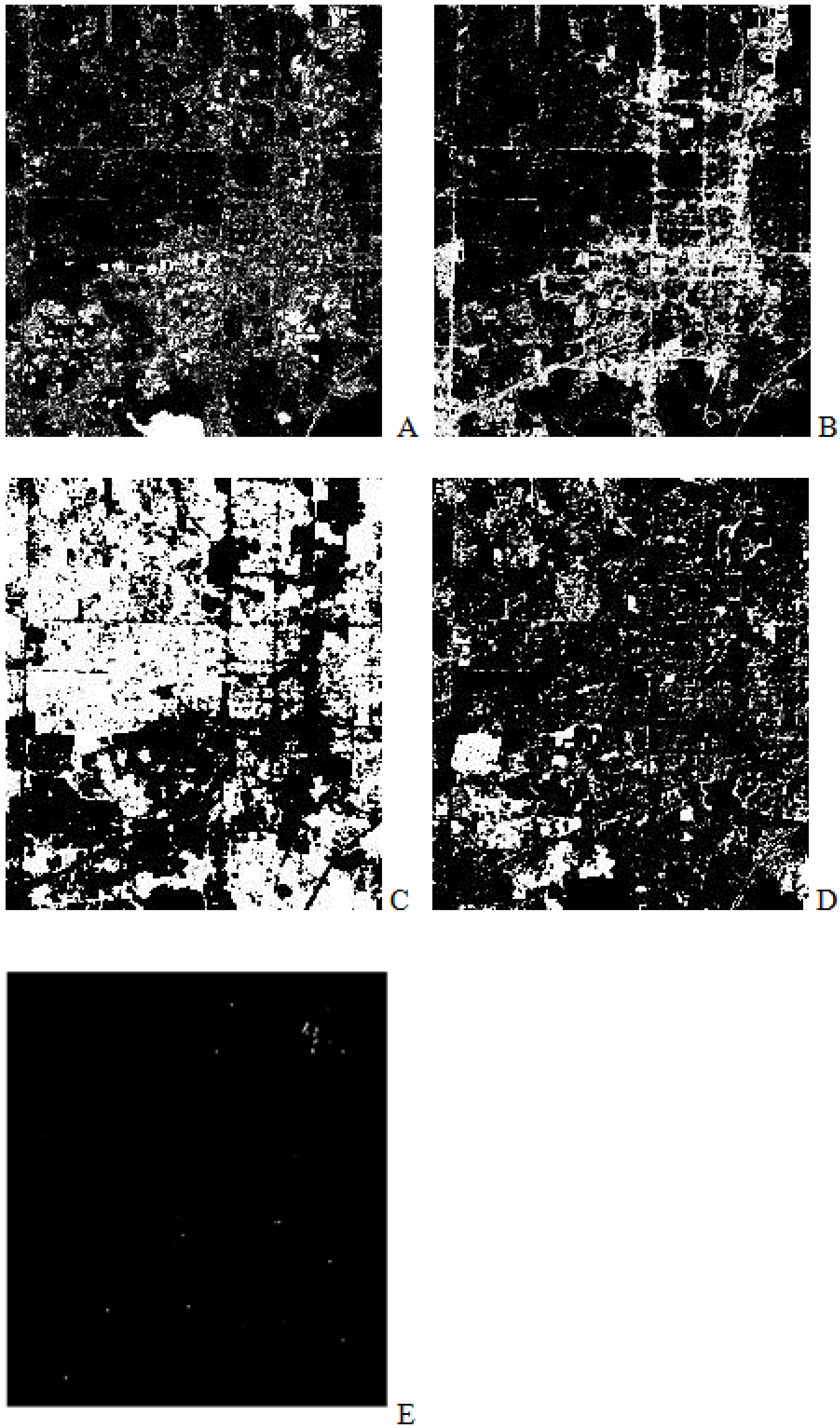

Figure 2-9. Exponential transformed maps for five components. A) Bright Impervious Surface, B) Dark Impervious Surface, C) Forests, D) Grasslands, and E) Soil.

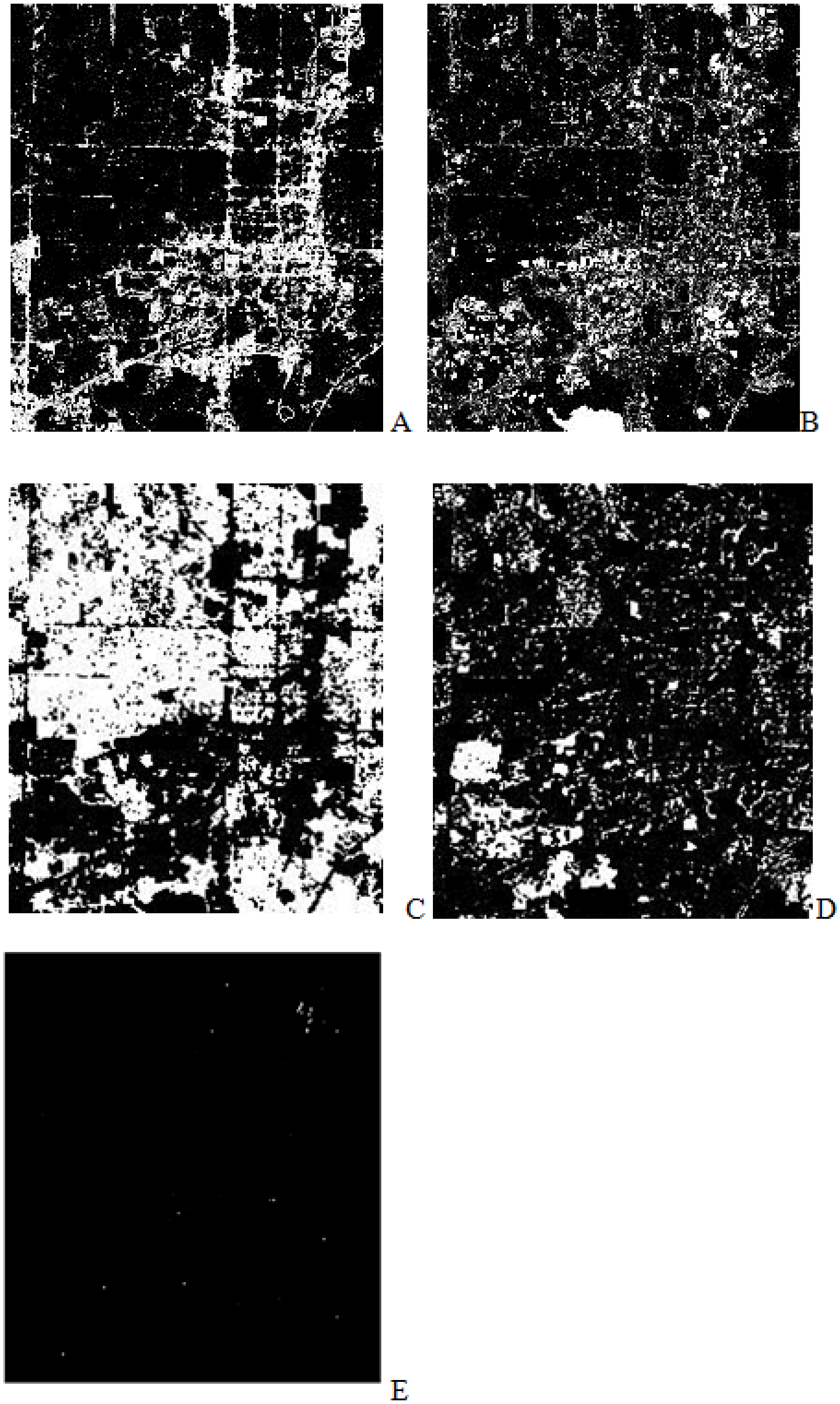

Figure 2-10. Adjusted rule maps for five components. A) Bright Impervious Surface, B) Dark Impervious Surface, C) Forests, D) Grasslands, and E) Soil.

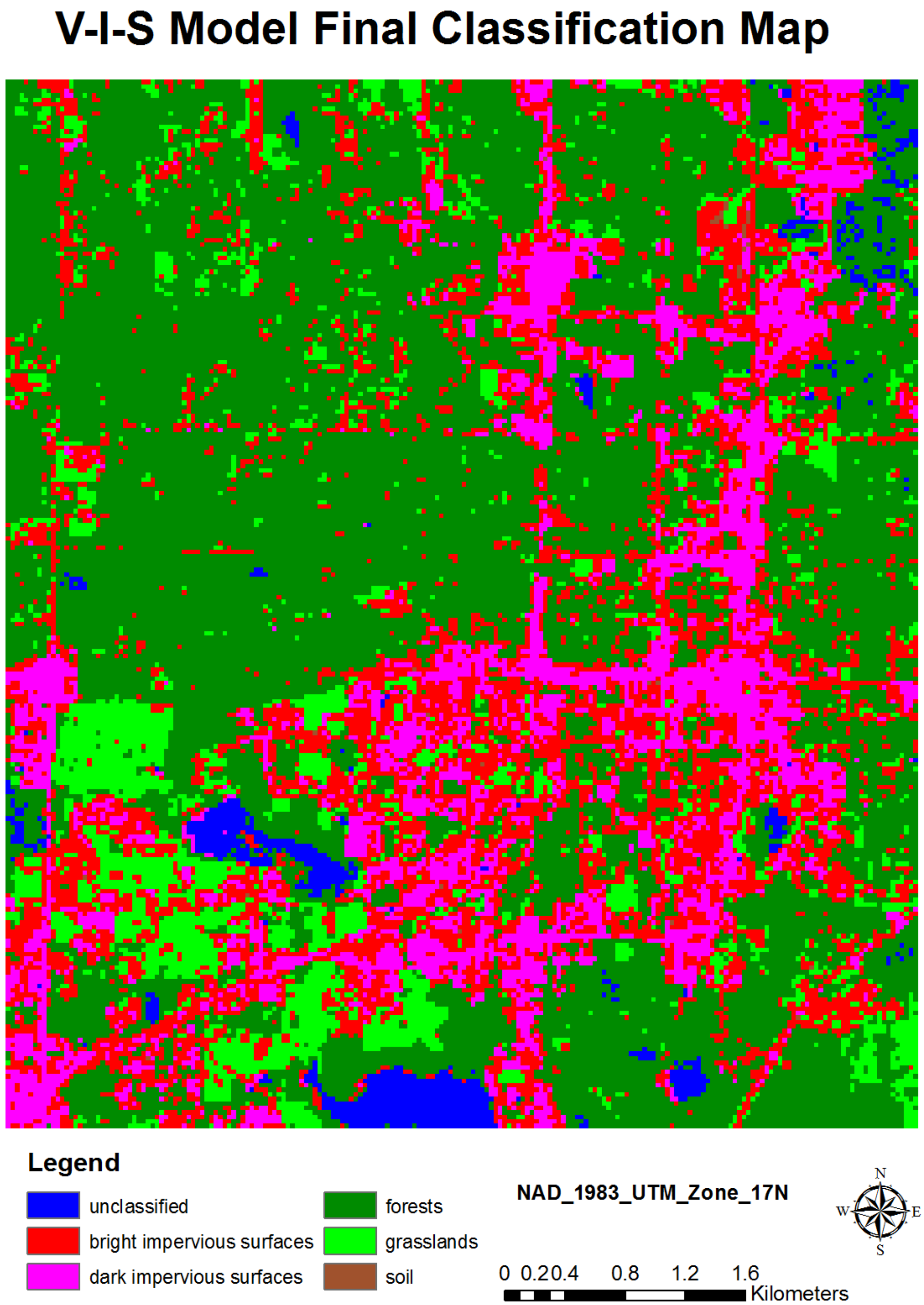

Figure 2-11. V-I-S model final classification map

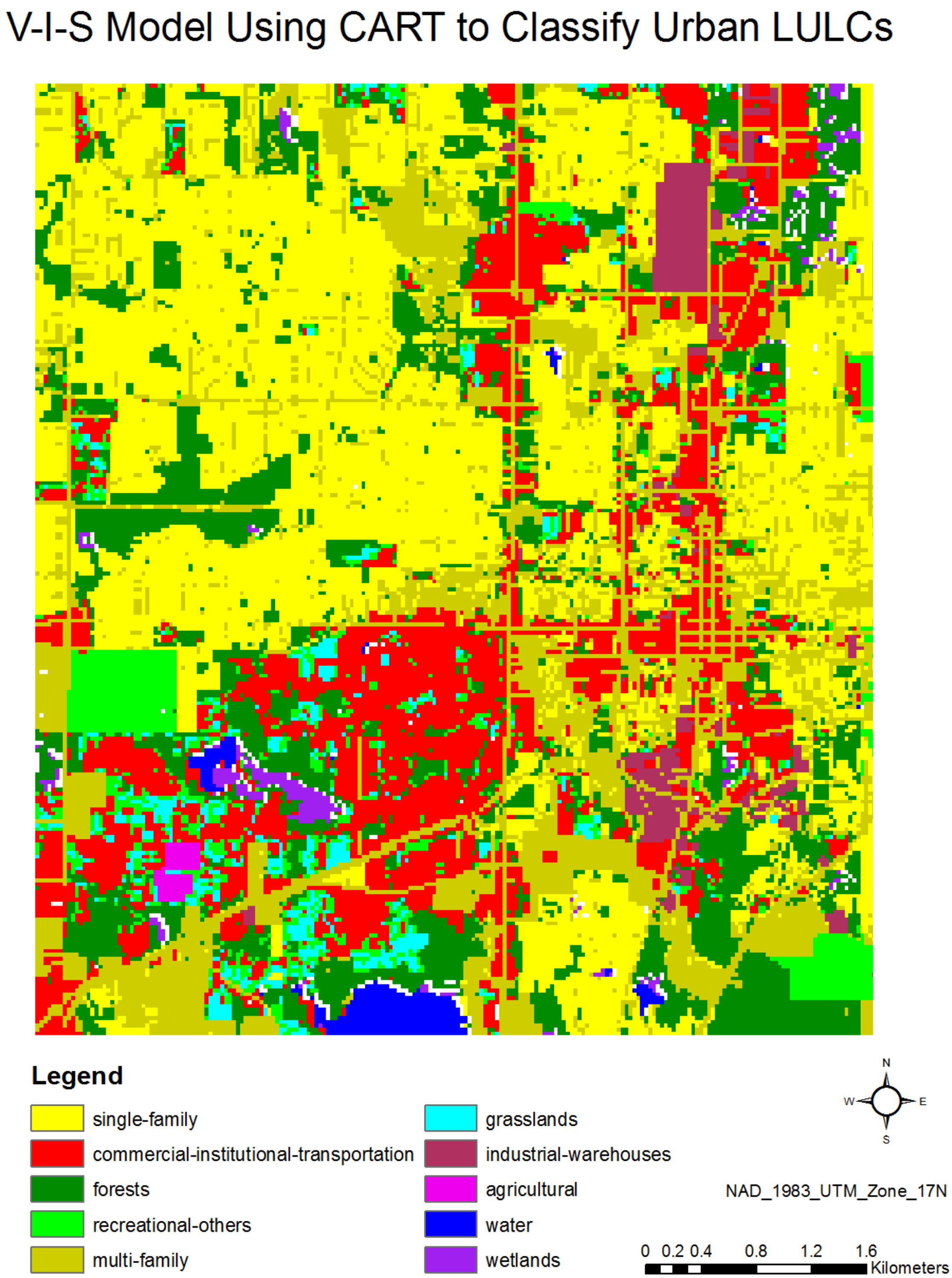

Figure 2-12. V-I-S model using CART method to classify LULC

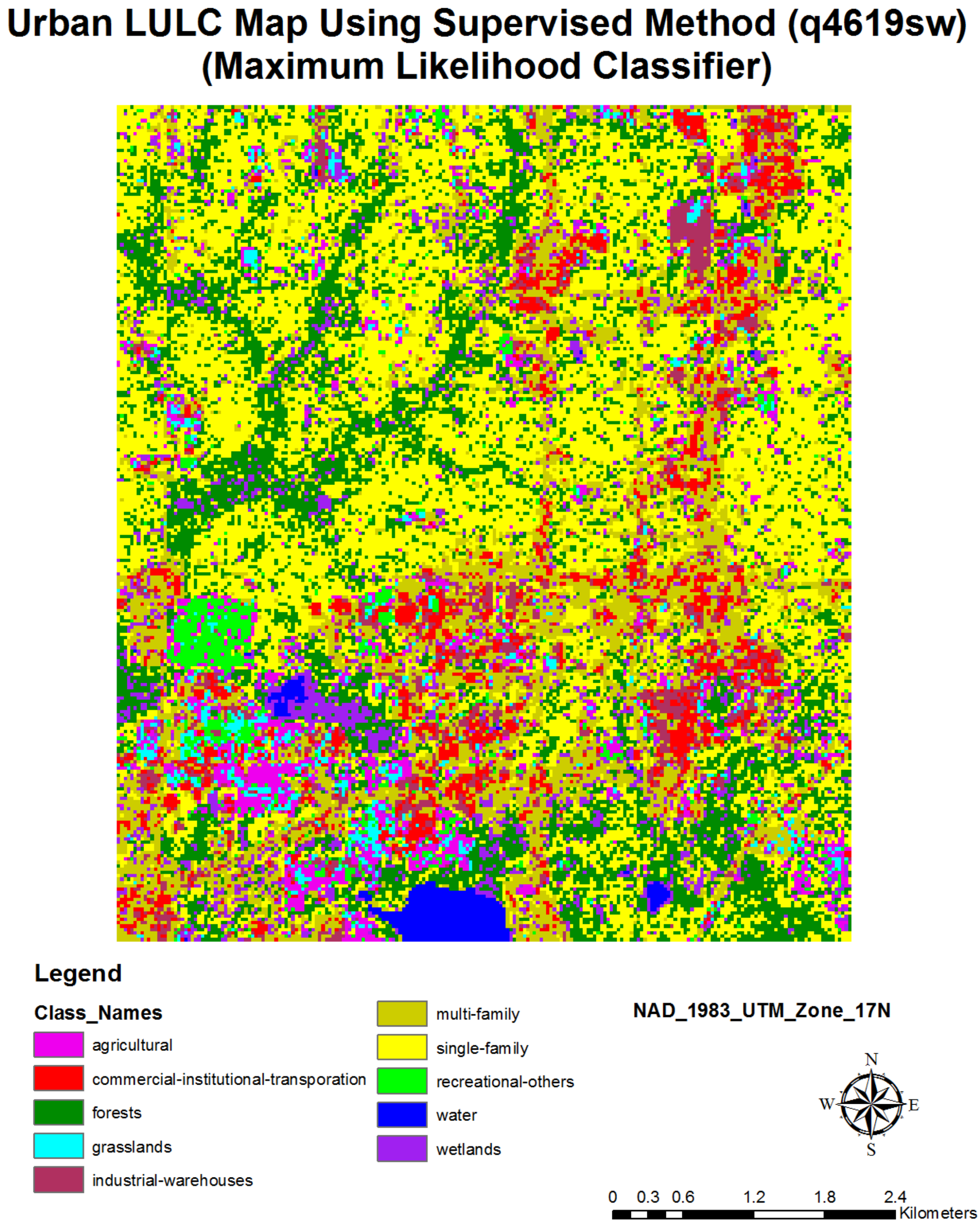

Figure 2-13. Urban LULC classifications using the supervised method

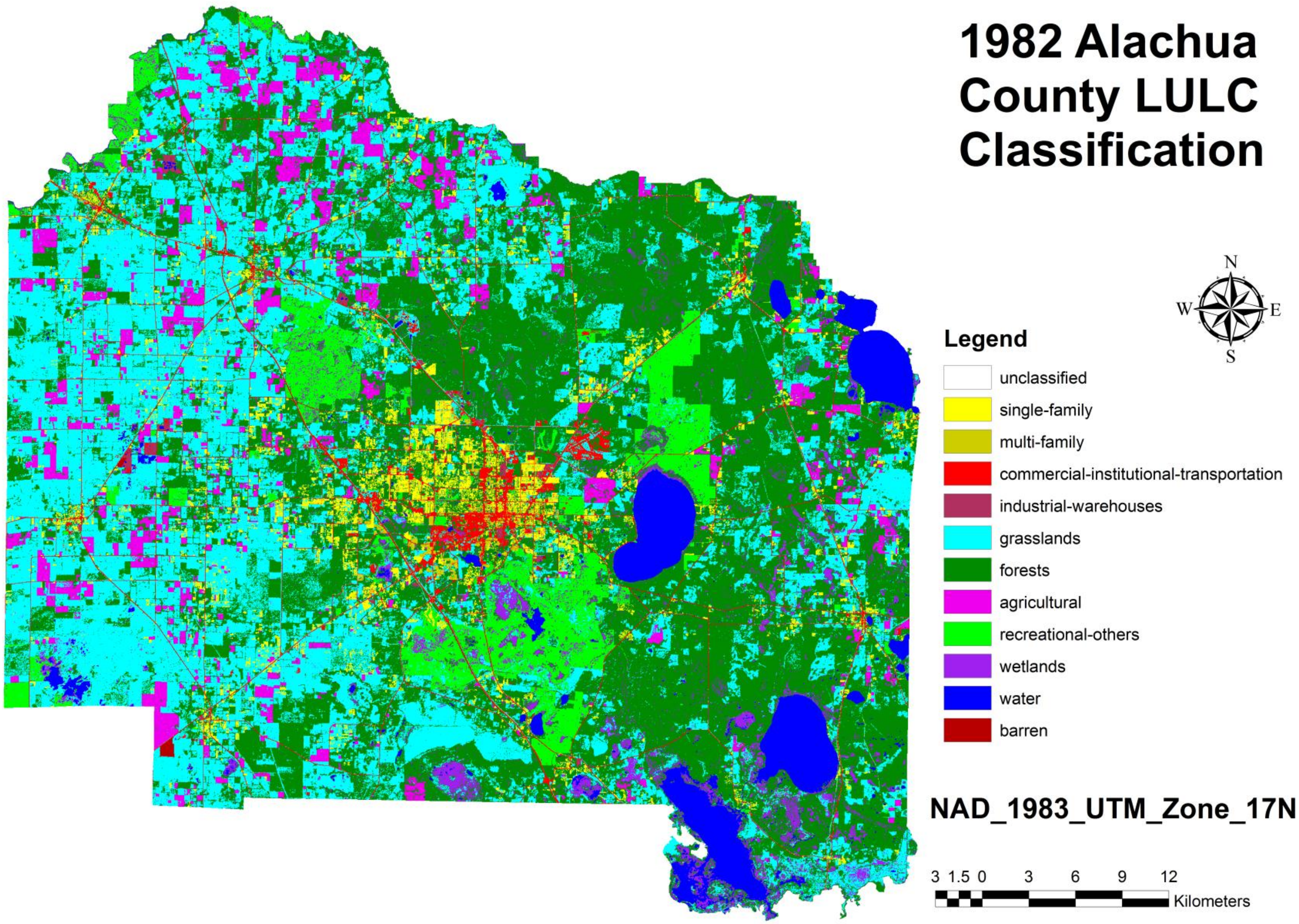

Figure 2-14. 1982 classification map for Alachua County

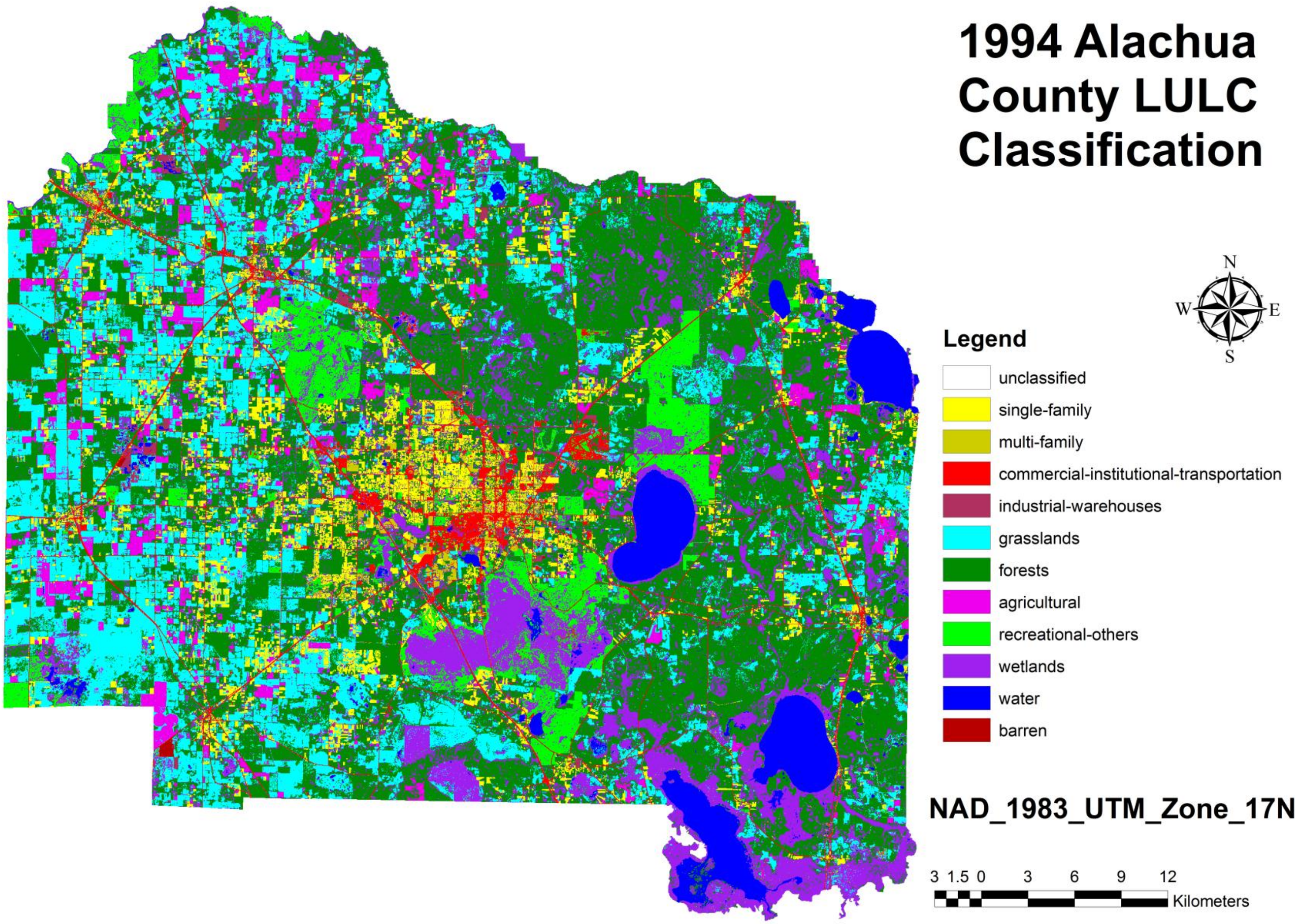

Figure 2-15. 1994 classification map for Alachua County

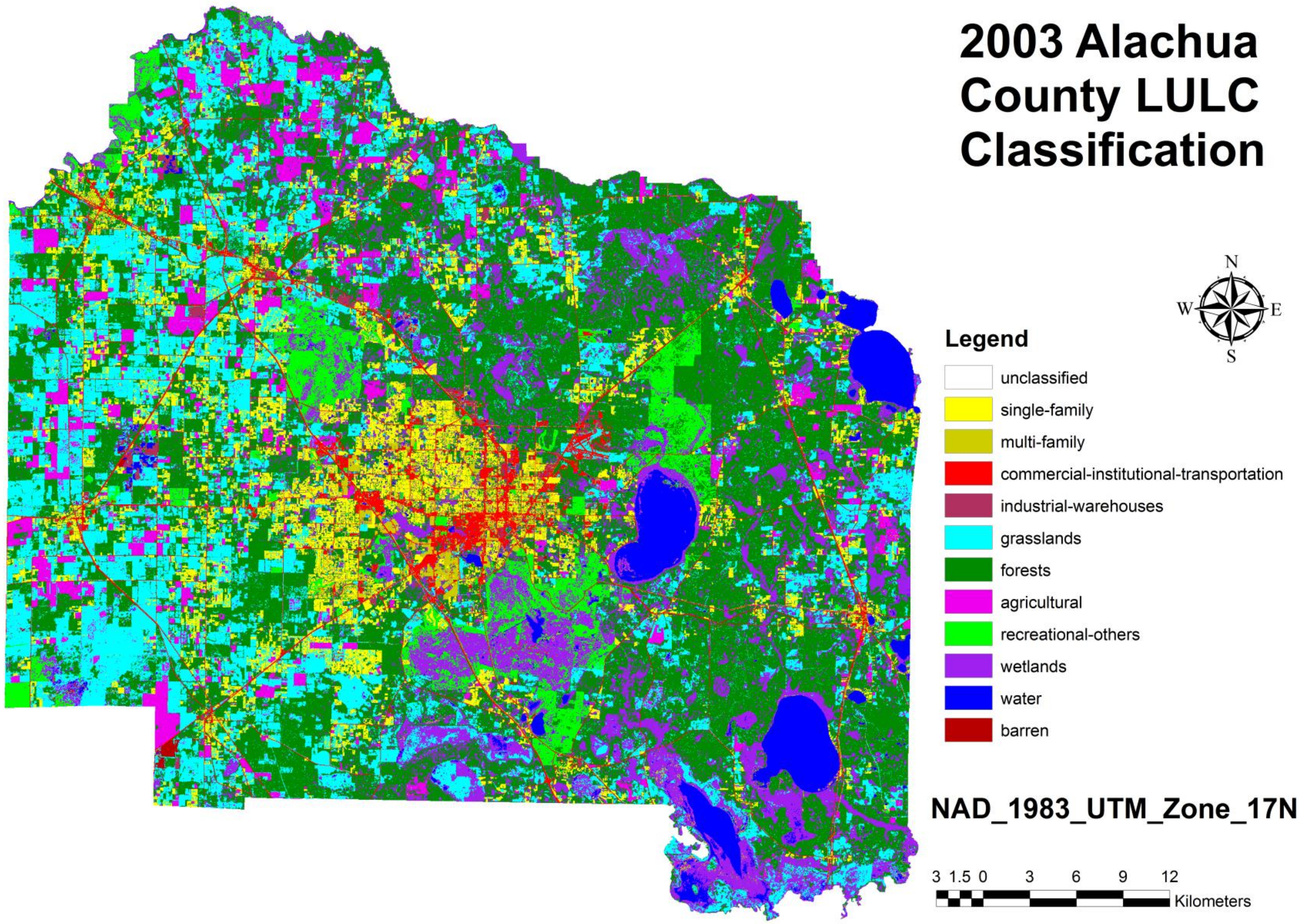

Figure 2-16. 2003 classification map for Alachua County

CHAPTER 3
URBAN MODEL BUILDING—THE MULTINOMIAL LOGISTIC REGRESSION MODEL

Chapter Overview

This chapter uses the MLR model to simulate urban growth in Alachua County through modeling the historical urban development in the county for the past 30 years. This chapter simulates the urban growth in the county for 2003, based on eleven LULC classes identified in Chapter 2 for 1982, 1994, and 2003, respectively, according to the Landsat TM and ETM+ data. These eleven LULC classes are single-family, multi-family, commercial-institutional-transportation, industrial-warehouses, recreational-others, agricultural, forests, grasslands, wetlands, water, and barren. The study uses the images that are classified for 1982, 1994, and 2003, respectively, to simulate the urban LULC classes for 2003.

Logistic Regression Model. Logistic Regression is a model to predict membership of a dependent variable based on a series of independent variables, which can be continuous, discrete, dichotomous, or some combination of several data types. The output of the modeling results is categorical or dichotomous—belonging to a specific membership or not—or a continuous probability value, between 0 and 1, for the membership. Logistic Regression can be applied to an urban study to predict the occurrence of urban development in a probabilistic manner, based on a series of independent variables such as demographic, social-economic, ecologic-physical, and spatial factors that are identified as significant (Hu and Lo, 2007). The Logistic Regression model considers historical factors towards urban development. Unlike the Cellular Automata (CA) model, which simulates urban spatial development without delving into the reasons behind the development, Logistic Regression is capable of informing those reasons (Hu and Lo, 2007).

The Logistic Regression model is applied to deforestation, agricultural, and urban growth (Hu and Lo, 2007). For urban growth, practically, Hu and Lo (2007) used it to simulate urban growth in Atlanta, Georgia. They found that the Logistic Regression model is able to include a variety of variables into analysis from physical, spatial, demographic, social, and economic aspects as well as land use development policies and environmental protection. This is advantageous over non-statistic models such as the CA model. More importantly, the Logistic Regression model requires much less CPU time for data processing. This is a big advantage over the CA model. However, Logistic Regression is a stationary model that treats land use development as if it occurs at the same time. It tells where the development takes place rather than when and hence is less temporally dynamic (Hu and Lo, 2007). In addition, the Logistic Regression model does not consider personal or household preferences for land use development nor personal behavior in land use development, which is unlike the Agent-Based Model (ABM) that considers personal behaviors and choices. The logistic regression model will be explained in detail later in this chapter, which will provide reasoning as to why the logistic regression model is chosen in this study for urban LULC modeling.

The Accuracy of Urban Growth Modeling

The models to simulate urban land use changes have been developed since the 1950s (Batty and Longley, 1994). The development of models for urban land use change attempts to simulate urban land uses from various angles by studying the inter-relationship between human behaviors and land use development. The development of urban simulation models is much related to humans' understandings of cities as well as urban land use development. As a result, the accuracy level of an urban development simulation model reflects how well humans understand a city. However, due to the fact that humans' understandings of cities are still limited, it is still difficult for humans to simulate urban development with great accuracy. The only way humans can do this is to characterize the nature of urban development as much as possible and work as accurately as possible to reflect the real-world urban development patterns. On the other hand, almost all models for predicting an event have errors ranging from sample errors to random errors[1] and to systematic errors, or bias.[2] The urban models are no exceptions. As a result, it is always necessary to conduct accuracy assessments for urban simulation models.

Moreover, urban development is much more historically related. The prediction of urban land use changes typically starts from analyzing the land use historical development. Since most cities originate from organic, unplanned states (Batty and Longley, 1994), some may still be developed in an organic way; they make urban simulation difficult and, as a result, may introduce additional errors compared to a planned city. In addition, people's behaviors are unpredictable and uncertain in nature. For land use development, land use decisions may be beyond the prediction of a scientific model. For example, different owners may deal with their land in a different way. Considering the same geophysical and economic conditions of their land, some may sell their land more quickly than others simply because of personal financial reasons. This may be beyond the prediction of an urban land use model. To deal with land use development uncertainties, consequently, it is common to use probability to address the issue. However, the introduction of probability to urban land use change models will unavoidably bring errors because, given a typical sample size, it is not possible to realize the predicted land uses that 100 percent match the observed land uses: there is always a difference between the observed and the predicted. Lastly, there is a policy uncertainty for urban land use development, in which it is not possible for an urban land use change model to predict whether there will be more conservative or more liberal policies for urban land use development and whether more strict environmental laws will be enacted in the future given the quantitative nature of the urban land use change models.

---

[1] The difference between observed values and predicted values (Mowrer and Congalton, 2000).
[2] The consistent difference between the central tendency, the mean, of observations and the "true" value (Mowrer and Congalton, 2000).

Uncertainty of Future Land Use Development. Because the urban models are rule based, there are orders or rationales behind them. As a result, predictions of future land use development to some extent carry researchers' "subjective" views of urban development. In real-world circumstances, however, a city may still be developed in its own way, being presented in a relatively organic fashion. As mentioned previously, because there are uncertainties related to people and their land use decisions, at the micro level, the actual urban land use changes may vary in accordance with developers' preferences and wills or reasons such as availability of funds, even though there are land use control tools available such as the future land use plan and zoning. For example, a well-predicted land for single-family use may actually be developed as multi-family, and it is common that the land that is suitable for development, which is predicted as "developed," are actually not developed or vice versa. These can make the overall land use patterns "softer" and less "planned" and thus introduce more errors in contrast to our "subjective," "planned," and "hard" forecasts.

Uncertainty of Future Land Use Policies. There is a land use policy uncertainty in the development of urban land use models. In reality, the future land use policies may vary from the current ones at the local, state, and federal levels. For example, a local government may change its attitude towards land use development and may become more liberal or more conservative towards land use development. The same is true for the state and federal government. For example, whether the state or federal government will pass the growth control laws to discourage land use development or whether the Congress will enact stricter environmental laws will be determined by the political dynamics at the executive and legislative levels, which are beyond researchers' ability to know in advance. These will unavoidably influence the accuracy of the land use modeling. In addition, natural disasters, such as fires, earthquakes (NCGIA, no date), and hurricanes, which are unpredictable in nature will inevitably influence land use development and thus influence the accuracy of urban models. Moreover, economic downturn and upturn (NCGIA, no date) will affect land use development as well and hence influence the accuracy of urban land use simulations. As a result, it is not likely that a model 100 percent matches the actual development; it is expected that the created models reflect the actual land use patterns to the maximum possible extent and best estimate the occurrence of future land use development.

Relevant Literature Review for Logistic Regression

The Literature Review is preceded based on the contexts that are related to this chapter's discussions, which is the model building related to the MLR model for urban growth simulation. Poelmans and Rompaey (2010) utilized the Landsat MSS, TM, and ETM+ images to simulate urban growth based on 1976, 1988, and 2000, respectively, employing a hybrid model of the logistic regression and the CA models. Learned urban simulation studies that were mostly based on the logistic regression, CA, or combination of them, they proposed a hybrid model to simulate urban growth for the Flanders-Brussels region of Belgium. Their land classification included urban land, arable land, grassland, forest, and water, and their independent variables included distance to cities, slope, employment potentials, distance to roads, and zoning status. These land categories as well as identified variables allowed the authors to produce thirty-one models to simulate urban growth based on probabilities. They found that the CA model contributed less to the accuracy of the model than the logistic regression model. They also found that zoning status was the most important factor determining the accuracy of their model by comparing the odd ratios for each variable, and also that the combination of the logistic regression and the CA model dramatically enhanced the accuracy of the model. However, the drawback of their research is that their raw data, derived from the Landsat images, is not of high accuracy. The accuracy levels for 1976, 1988, and 2000 in their research were 77.6 percent, 82.8 percent, and

82.3 percent, respectively. The 1988 and 2000 data had just reached the benchmark of the 80 percent accuracy level, whereas the 1976 data was lower than the necessary level of 80 percent. Poelmans and Rompaey’s (2010) research shows that the spatial independent variables are sought to simulate urban growth. Also, the logistic regression model is adopted to improve the accuracy of the model. These elements are inspirational for this research, in which various spatial independent variables can be useful to simulate urban uses which can lead to the destinations of this research.

Almeida et al. (2003) also applied a hybrid CA and logistic regression model to simulate urban uses based on five categories: residential use, industrial use, services, and mixed use, for the Town of Bauru, Brazil. They used the hybrid model to select possible explanatory (independent) variables that were associated with each land use identified above through looking up the Cramer’s statistics, the Joint Information Uncertainty, and the Correlation Indices. These values were selected when they were greater than a benchmark, which created a total of nineteen explanatory variables. Then, these nineteen variables were input into the software, known as DINAMICA. During the process of calculating the probability values in the logistic regression model, the Wald Chi-square test and the G-statistics were assessed in order for them to exclude the variables having the least significance level. As a result, MINITAB was used in this regard so as to obtain the *P*-value for each variable based on the maximum likelihood method. Finally, DINAMICA was employed again to compute the probabilities for each of the five urban uses mentioned above. Calibration was also conducted at this step so as to fine tune the parameters such as the number of iterations, average size and variance of patches, and so on, in the CA model. The final map includes a predicted land use map for 1988 based on the five land use categories.

There are three merits in the research of Almeida et al. (2003). First, they simulate urban uses for more than two categories, in which five land use categories are included, not merely one, e.g., urban vs. non-urban. Second, they employ the hybrid model to simulate urban growth based on the probabilities for each land use category. As a result, five different urban uses can be allocated through comparing the ordinal values of each use. Third, they use the *P*-value to select independent variables that are significant to their research. These three merits of the research of Almeida et al. (2003) are also the direction that this study is going to take.

Landis (1994) proposed a California Urban Future (CUF) model to simulate urban growth in the northern bay area of California which comprised fourteen counties. His model has four modules: population growth, spatial database, spatial allocation, and annexation, which is essentially a vector-based urban growth analysis with the simulation based on the Census Bureau’s Topologically Integrated Geographic Encoding and Referencing (TIGER) data. On the other hand, Landis’ (1994) model discusses urban growth based on the urban category only, i.e., the residential use (Landis and Zhang, 1997). It does not address detailed urban uses for the CUF model.

To remedy the inadequacies of the CUF model, Landis and Zhang (1997) developed a CUF II model to simulate urban growth. The reason that the CUF II model is closely related to this research is that it proposed an MLR technique to model urban development and redevelopment. In addition, the CUF II model added four additional uses by expanding the original single urban use to residential use, commercial use, industrial use, public use, and transportation use. The CUF II model is inspirational for urban growth modeling in this study as it suggests using the MLR model, for the first time, to simulate urban development and redevelopment. It also carries some heuristic elements such as inclusions of various urban uses, e.g., residential use, commercial use, industrial use, public use, and transportation use, as mentioned above, considering urban development and redevelopment. The CUF and CUF II models that are inspirational for this study are also represented in additional elements such as the adoption of population forecasts, the proposal of different scenarios, and the bid of competitive urban uses for a typical site, which will be discussed in detail in Chapter 4.

Hu and Lo (2007) developed a method to simulate urban growth in the Atlanta region in Georgia, where it incorporated 13 counties, using the Landsat images. The land use categories of their research are based on six LULC categories: high-density urban, low-density urban, bare land, crop or grassland, forest, and water. For the reason that their study area comprised 13 counties, they modeled urban growth based on urban use only, along with some natural land. Their remote sensing data covered the period of 1987 and 1997, respectively, for the study area. They proposed twenty independent variables, which include spatial and social data such as income per capita, poverty rate, distance to nearest urban cluster, distance to CBD, distance to active economic center, distance to the nearest major road, and so on. Hu and Lo (2007) used the Relative Operating Characteristic (ROC) curve to assess the accuracy of the model. Hu and Lo's (2007) research is inspirational for this study in terms of model building, model interpretation, and sensitivity analysis, for which this research will also follow their research sequence and offer the similar contents as what they provided. They will be described in detail in later of this chapter.

Methodologies

Currently, there are three data sources available for county-based urban research: parcel data, aerial photographic images, and satellite images. Parcel data is a commonly used data source, which is provided by the property appraiser's office of a county. The most important advantage of the parcel data is its coverage, which covers an entire county. Unlike some data sources, e.g., zoning data and future land use data, which are based on political boundaries, such as incorporated cities and towns, the parcel data is based on the entire county, which facilitates urban studies to be conducted without combining data from piecemeal sources.

The parcel data can be used directly based on urban research needs, which usually need to be aggregated into urban uses through combinations of various property use codes. As a result, parcel data is able to provide land use information for important urban use types such as single-family, multi-family, commercial, industrial, institutional, and recreational. It also provides the information for natural land uses, such as forests, grasslands, agricultural land, and so on. Moreover, the parcel data provides information on vacant land, some of which are categorized based on their usages. For example, parcel data categorizes vacant residential uses, vacant commercial uses, vacant industrial uses, and vacant institutional uses, which can be used for infill development purposes.

Because the parcel data is based on property value assessments, some land-use types are absent from it, such as transportation corridors, residential right of ways, and wetlands. In addition, parcel data reflects land uses rather than land covers. For example, the University of Florida campus is a large parcel in Alachua County's parcel data without having detailed land cover information such as buildings, lawns, parking lots, forests, agricultural land, and so on within the parcel. Moreover, parcel data with its land categories do not reflect their actual uses.

Because of the shortcomings with the parcel data, it is necessary to find alternative data sources. Currently, there are remote sensing data available for urban studies based on aerial photographic images and satellite images. The challenge of using satellite images to classify urban LULC classes rests on three factors as introduced in the previous chapter. First, urban LULC classes usually have similar spectral characteristics, which preclude them from being distinguished accurately (Paul, 2007). Second, urban LULC classes are barely homogeneous, in which different types of LULC classes mix together in a pixel, which makes the pixel hard to differentiate based on its pixel value. Third, satellite sensors may not be sharp enough to detect low-density urban development at the urban fringe. These three factors make urban LULC classification difficult to classify that are equivalent to the USGS Level III standard (Anderson et al., 1976), in which residential uses such as single-family and multi-family are differentiated, which is critical to urban growth research. Also, these three factors prevent researchers from obtaining accurate ground truth information for urban LULC classes. Because of this, county-level studies, which require urban uses to be differentiated further into USGS Level III, are seldom conducted.

In addition, based on the recent trend, many urban researchers used AI models to simulate urban growth. Because some AI models, e.g., the CA model and the ABM, consider people and their choices in land use decisions, as a result, the utilization of other model types, e.g., the logistic regression model, to mimic detailed urban development has been neglected. In fact, the logistic regression model is a heuristic platform for urban growth modeling to build upon as long as urban LULC classes are accurately classified and important urban use categories are provided.

The logistic regression model offers a number of advantages compared to some AI models such as the CA model. First, the logistic regression model considers multiple factors such as demographic, social-economic, ecologic-physical, and spatial factors (Hu and Lo, 2007) that are significant to urban land use changes. It can also include governmental land use policies and environmental protection factors into the model (Landis, 1994; Landis 1995). In addition, the logistic regression model takes historical factors into consideration (Landis and Zhang, 1997). Consequently, longitudinal data can be imported into the model to test whether historical development of land use patterns influence the current land use patterns and whether historical factors can be considered to predict the future land use patterns. Generally speaking, the logistic regression model is capable of revealing the reasons behind urban land use patterns and changes (Hu and Lo, 2007). Most importantly, the logistic regression model requires less CPU time to process the data, which is time-saving for simulating urban land use development especially in large areas, with respect to large quantities of data processing needs. For the county-based study, the logistic regression works well with research needs, which is capable of predicting urban uses and allocating urban uses based on multi-use categories.

Logistic Regression Model Construction. This study adopts the MLR model to simulate urban growth. The MLR model is a model evolved from the binary logistic regression model (IDRISI Help, no date), in which multi-variables instead of two variables are included into the model to simulate urban growth based on the probability of a dependent variable it receives. It is mainly applied to predict categorical or dichotomous variables, or continuous variables, which are in probability between 0 and 1, given one or more predictors or explanatory variables.

Logistic regression is similar to the Ordinary Least Squares (OLS) regression model (Field, 2000). Like the OLS regression model, the logistic regression model reflects the relationships between a dependent variable and one or more independent variables (Hosmer Jr and Lemeshow, 1989). It has coefficients for each independent variable and also a constant intercept (Field, 2000). It can incorporate categorical and continuous data for independent variables into the model. The logistic regression model expresses the probability of whether a dependent variable belongs to a certain category or is a member in a logit format (Menard, 1995). A logit is also called the natural logarithm of the odds, which represents the natural logarithm of the probability ratio of the dependent variable that belongs to a member or the probability that does not belong to a member (Menard, 1995). For the value of the dependent variable, if a value is closer to 1, it indicates that it is more likely for that variable to belong to a member in a study; conversely, if the value is closer to 0, it indicates that it is more likely that it does not belong to the member in question (Field, 2000). As a result, 0 often refers to not belonging to a member and 1 refers to belonging to the member. The logit can be written as:

$$\log it(Y) = \ln(Odds) = \ln\{p(Y=1)/[1 - P(Y=1)]\} \qquad (3\text{-}9)^{3}$$

Based on the above description, the logit can also be written as the lineal form of independent variables and the dependent variable for an MLR model:

$$\log it(Y) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k \qquad (3\text{-}10)^{4}$$

Because $\log it(Y) = \ln(Odds)$, it is obvious that $Odds(Y=1) = e^{\log it(Y)}$; as a result, in the form of exponentiation, Equation 3-10 can be re-written as:

$$Odds(Y=1) = e^{\ln[Odds(Y=1)]} = e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k} \qquad (3\text{-}11)^{5}$$

And

$$P(Y=1) = \frac{e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k}}{1 + e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k}} \qquad (3\text{-}12)^{6}$$

And

$$P(Y=1) = \frac{1}{1 + e^{-z}} \qquad (3\text{-}13)^{7}$$

Where

$$z = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \qquad (3\text{-}14)$$

[3] Source: Menard, 1995
[4] Source: Menard, 1995
[5] Source: Menard, 1995
[6] Source: Menard, 1995
[7] Source: Field, 2000

In order to obtain the coefficients as well as the constant intercept, the logistic regression model adopts a maximum likelihood approach to determine the coefficients and the constant, which is different from the OLS regression model, in which the coefficients and the constant are calculated directly. Based on this maximum likelihood approach, the logistic regression model tries tentative solutions and then slightly revises those tentative numbers until the changes, given by different trials, are the closest so that they can be neglected. This is actually an iterative process, which can only be performed by computers (Menard, 1995).

Three techniques can be adopted for the goodness-of-fit analysis for an MLR model: the Chi-square test, the Taylor-Russell Diagram analysis, and the ROC curve. The Chi-square test is used to analyze nominal or ordinal data to see whether they are statistically significant. In the logistic regression model, Chi-square is used to examine whether the observed and the expected are statistically significant. The Taylor-Russell Diagram is used to analyze how good the prediction of a model is, given the observation and prediction interactions based on four possible outcomes such as true positive, true negative, false positive, and false negative. The ROC curve is used to measure the degrees of agreements between the simulated use and the actual use occurred (Pontius and Schneider, 2001). It answers the question as to how well an LULC can change given high suitability scores provided by the model for an area (Hu and Lo, 2007). Statistically, the ROC value is between 0.5 and 1. If a ROC value is 1, it means that high suitability scores in an area result in a perfect agreement for the area to change its LULC classes. Conversely, if a ROC value is 0.5, it means that high suitability scores are assigned more randomly at locations across an area (Pontius and Schneider, 2001). This chapter will use the Chi-square test as well as the ROC curve to conduct sensitivity analyses. The Taylor-Russell Diagram is not used because the Taylor-Russell Diagram addresses the sensitivity of modeling results such as true positive, true negative, false positive, and false negative that are redundant with the Chi-square test and the ROC curve test. The ROC curve also addresses these values nevertheless with more heuristic statistics not available in the Taylor-Russell Diagram.

Logistic Regression Results

This study uses both the LOGISTICREG and MULTILOGISTICREG modules in IDRISI so as to conduct urban growth simulations for Alachua County. The simulations are preceded based on four urban uses identified in the LULC classification process in Chapter 2: single-family, multi-family, commercial-institutional-transportation, and industrial-warehouses. Accordingly, four MLR models are established based on the four urban LULC classes identified, with each LULC class having a corresponding MLR model.

First, relevant independent variables for each of the four LULC classes are found. They are input into the model to test whether they are statistically significant to the dependent variable and also whether they yield high accuracy. When an independent variable is found to be statistically significant, it will be input into the MLR model for refining purposes. As a result, non-statistically-significant independent variables are not included in the refining models. Benchmarks are also set for increasing the overall accuracy level of a dependent variable. For residential uses, including the single-family use and the multi-family use, at least 80 percent of the overall accuracy should be reached. For commercial-institutional-transportation and industrial-warehouses development, at least 90 percent accuracy should be achieved. As a result, an incremental process is undertaken for each of the four urban LULC classes in order to find the relevant independent variables that yield the highest overall accuracy. As a result, prospective independent variables are added into the model one-by-one to test whether they increase the general accuracy level. This is an empirical process, which assesses spatial characteristics of the four LULC classes, such as the proximity to existing roads and existing residential development, and so on, rather than social economic factors such as population, income, and poverty. The reason for evaluating spatial characteristics of each dependent variable is that the social economic factors are in fact reflected and implied in the land use development, which do not need to be assessed specifically. For example, the increased single-family use in the study area reflects the population growth, and the increased commercial-institutional-transportation use in the area reflects the economic development of the study area. Therefore, this study does not simulate these background factors in particular. Generally speaking, this study tests fourteen independent variables to simulate the single-family use, eight variables for the multi-family use, fourteen for the commercial-institutional-transportation use, and eight for the industrial-warehouses use. For refined models, this study tests five independent variables for the single-family use, six independent variables for the multi-family use, eight independent variables for the commercial-institutional-transportation use, and six independent variables for the industrial-warehouses use.

Second, when probability maps are produced for four land uses of interest (usually following the refined models), they are tested in light of each use's sensitivity. In this case, Chi-square test, Pseudo R-Square and Cramer's V tests, and the ROC test are conducted. When the sensitivity tests generate positive results such as good agreement between the predicted and the observed for goodness-of-fit, the four simulated uses are mosaiced to create a 2003 LULC simulation map, along with a number of natural land such as recreational-others, agricultural, wetlands, forests, grasslands, water, and barren.

Finally, the four prediction maps will be used to allocate future land uses based on the eleven LULC classes identified. This step will be illustrated in detail in Chapter 4. In Chapter 4, four urban uses together with seven natural land are allocated for 2020 and 2030, respectively. This chapter does not cover the allocation process, however.

Model Calibration Single Family

To predict the single-family use, fourteen independent variables are found. They are (1) existing single-family use in 1982 based on remote sensing classifications; (2) existing single-family use in 1994 based on remote sensing classifications; (3) existing single-family use in 2003 based on the parcel data; (4) proximity to industrial-warehouses in 1994; (5) proximity to existing 1982 single-family use based on remote sensing classifications; (6) proximity to existing 1994 single-family use based on remote sensing classifications; (7) proximity to 1982 road network; (8) proximity to 1994 road network; (9) future land use for the single-family use; (10) 1994 single-family density; (11) proximity to schools; (12) located within urban cluster areas; (13) zoning for the single-family use; and (14) vacant land for the single-family use in 2003. They are listed in Table 3-1. The dependent variable is the 2003 single-family use based on the classified use from the remote sensing data. It is noted that these independent variables are all in thematic maps where 1 and 0 numbers are assigned to represent a cell within a specific target area of interest or outside a target area. In addition, zonal statistics are applied to proximity analyses, in which proximity to industrial, proximity to existing single-family uses, proximity to roads, and proximity to schools are partitioned by mean values of the zonal statistics.

To avoid autocorrelation provided by the independent variables, 0.2 percent of stratified sampling is applied. This is because 0.2 percent of random sampling is tested against the simulated single-family use for its autocorrelation characteristics, which results in a random distribution for the single-family cells. As a result, the Moran's I value and z-score are -0.04 and -0.62 (Table 3-5 and Table 3-6), respectively, obtained from extracting 0.2 percent stratified random points in ENVI 4.4 from the 2003 single-family prediction map, and they are tested in

ArcGIS 9.3. The outputs from operations in LOGISTICREG and MULTILOGISTICREG in IDRISI include a probability map (Figure 3-1), a residual map, and a statistical sheet, in which a bunch of statistics are calculated such as coefficient, mean, and standard deviation for each independent variable, -2 Log Likelihood, Pseudo R-Square value, Chi-square value, probability cutting threshold, 2 × 2 contingency tables for true positive and false positive percentage values, and ROC values based on 100 intervals. When examining the calculated *P*-values of each independent variable (Table 3-1), it is found that five independent variables are statistically significant to the dependent variable of the 2003 predicted single-family use, given that *P*-values of these independent variables are less than 0.1 for two tailed *P*-values (0.05 for one tailed *P*-values) (Table 3-1). These five independent variables, named as refined variables, are (1) existing single-family use in 1982, (2) existing single-family use in 2003 based on the parcel data, (3) proximity to 1982 roads, (4) future land use for the single-family use, and (5) 2003 vacant land for the single-family use. The probability map, after refinement, creates a raw dataset for urban use allocations, which will essentially be introduced in Chapter 4. The independent variables that are statistically significant will be thrown into the model to test the goodness-of-fit and the overall accuracy. The cutting threshold before refinement is set at 0.6813, in which probability values that are greater than the cutting threshold are identified as the single-family use. The actual predicted 2003 single-family use employs the cutting threshold following the refined models. The 2003 single-family use prediction map (before refinement) is presented in Figure 3-2.

Model Calibration Multi-family

The process to produce a multi-family prediction map is essentially the same as the single-family use. During the process, eight independent variables are found. They are (1) existing multi-family land use in 1982 based on remote sensing classifications; (2) existing multi-family land use in 1994 based on remote sensing classifications; (3) existing multi-family land use in 2003 based on the parcel data; (4) vacant land for the multi-family land use in 1994 with ecosystems, parks, and conservations masked; (5) proximity to the existing 1994 multi-family land use based on remote sensing classifications; (6) proximity to major roads; (7) future land use for the multi-family use; and (8) zoning for the multi-family land use. The dependent variable is the 2003 multi-family use derived from the 2003 remote sensing urban LULC classification map. They are input into the LOGISTICREG and the MULTILOGISTICREG modules in IDRISI. The eight independent variables are all thematic data, in which 1 and 0 are categorized for the cells within an area of interest or outside the area, respectively. In addition, 1.5 percent stratified sampling is applied because the random sampling based on this percentage has random distribution of the multi-family use, with autocorrelation excluded. The output maps are a probability map (Figure 3-3) and a residual map for the multi-family use. Specifically, the cutting threshold is 0.0470 before refinement because of the sampling applied. Based on this number, a 2003 multi-family prediction map is produced (Figure 3-4). Moreover, the calculated *P*-values indicate that six of the eight independent variables are statistically significant in terms of the 0.10 two-tailed *P*-values (Table 3-2). They are (1) existing multi-family land use in 1982 based on remote sensing classifications; (2) existing multi-family land use in 1994 based on remote sensing classifications; (3) existing multi-family land use in 2003 based on parcel data; (4) proximity to existing 1994 multi-family land use based on remote sensing classifications; (5) future land use for the multi-family use; and (6) zoning for the multi-family land use. This means these six variables contribute significantly to the dependent variable.

Model Calibration Commercial-Institutional-Transportation

There are sixteen independent variables for the commercial-institutional-transportation use. They are (1) existing commercial-institutional-transportation use in 2003 based on the parcel data; (2) existing commercial-institutional-transportation use in 1982 based on remote sensing classifications; (3) existing commercial-institutional-transportation use in 1994 based on remote sensing classifications; (4) vacant land for the commercial-institutional-transportation use in 2003 with ecosystem, parks, and conservations masked; (5) vacant land for the commercial-institutional-transportation use in 1994 with ecosystem, parks, and conservations masked; (6) 1994 vacant land proximity to 2003 single-family and multi-family uses; (7) 1994 vacant land proximity to 1994 single-family and multi-family uses; (8) proximity to 2003 commercial-institutional-transportation use; (9) proximity to 1994 commercial-institutional-transportation use; (10) proximity to 2003 TIGER roads; (11) proximity to major roads; (12) future land use for the commercial-institutional-transportation use; (13) proximity to road intersection; (14) zoning for the commercial-institutional-transportation use; (15) 2003 vacant land proximity to 2003 single-family and multi-family uses; and (16) 2003 vacant land proximity to the 2003 multi-family use. The dependent variable is the 2003 commercial-institutional-transportation use derived from the remote sensing data. The same as the single-family use and the multi-family use, the independent variables for the commercial-institutional-transportation use are all thematic, in which 1 and 0 are categorized according to the cell locations within or outside a target area of interest. Like the above two uses, stratified random sampling is tested on the commercial-institutional-transportation use, in which 8.5 percent sampling is applied, which outputs the highest accuracy rate, while it is still randomly distributed. The output files are a probability map, a residual map, and a statistical sheet, which is the same as the above single-family use and the multi-family use. The probability map is presented in Figure 3-5. The cutting threshold is 0.1440 before refinement. Because the program becomes over-burdened with the sixteen independent variables operated in the IDRISI software, two independent variables are not included into the data running. These two variables include proximity to major roads and 2003 vacant land proximity to 2003 single-family and multi-family uses. Based on this programming running, i.e., fourteen variables instead of the sixteen, a commercial-institutional-transportation prediction map is produced, which is presented in Figure 3-6. In particular, eight independent variables are proved to be statistically significant. These variables include (1) existing commercial-institutional-transportation use in 2003 based on the parcel data; (2) existing commercial-institutional-transportation use in 1982 based on remote sensing classifications; (3) existing commercial-institutional-transportation use in 1994 based on remote sensing classifications; (4) vacant land for the commercial-institutional-transportation use in 2003 with ecosystem, parks, and conservations masked; (5) vacant land for the commercial-institutional-transportation use in 1994 with ecosystem, parks, and conservations masked; (6) proximity to 2003 commercial-institutional-transportation use; (7) proximity to 1994 commercial-institutional-transportation use; and (8) future land use for the commercial-institutional-transportation use. Their two-tailed *P*-values are all less than the 0.10 level (Table 3-3). They will be input into the model to test the goodness-of-fit for model refining purposes.

Model Calibration Industrial-Warehouses

The industrial-warehouses dependent variable has eight independent variables. They are (1) 2003 industrial-warehouses land use proximity to major roads; (2) existing industrial-warehouses land use in 2003 based on the parcel data; (3) existing industrial-warehouses land use in 1982 based on remote sensing classifications; (4) proximity to existing residential uses; (5) existing industrial-warehouses land use in 1994 based on remote sensing classifications; (6) 1994 vacant land proximity to 1994 industrial-warehouses land use; (7) future land use for the industrial-warehouses land use; and (8) zoning for the industrial-warehouses land use. The dependent variable is the 2003 industrial-warehouses use given by the remote sensing data. Like the single-family use, the multi-family use, and the commercial-institutional-transportation use, the independent variables for the industrial-warehouses use are all thematic, with 1 categorized as located within the specific area of interest and 0 as out of the specific area of interest. In addition, 15 percent stratified random sampling is adopted because it creates a random distribution for random points located within the simulated 2003 industrial-warehouses areas at this level and also generates high accuracy. The productions include an industrial-warehouses use probability map (Figure 3-7), a residual map, and a statistical sheet. The cutting threshold is 0.1281 before refinement. In particular, among the eight independent variables, six independent variables are proved to be statistically significant to the dependent variables because their *P*-values are all less than the 0.10 level (Table 3-4). These six independent variables are (1) existing industrial-warehouses land use in 2003 based on the parcel data; (2) existing industrial-warehouses land use in 1982 based on remote sensing classifications; (3) existing industrial-warehouses land use in 1994 based on remote sensing classifications; (4) 1994 vacant land proximity to 1994 industrial-warehouses land use; (5) future land use for the industrial-warehouses land use; and (6) zoning for the industrial-warehouses land use. The 2003 predicted industrial-warehouses use is illustrated in Figure 3-8. The autocorrelation Moran's I values as well as the z-score test for each of the above four dependent variables considering the situation of with and without stratified sampling are listed in Tables 3-5 and 3-6, respectively.

Refined LULC Logistic Regression Model

A refined model is the one based on refined independent variables. Because the above independent variables for each dependent variable have *P*-values representing the correlations with the dependent variables, a refined model takes the independent variables that are statistically significant to the dependent variable. However, compared to the original models, generally speaking, refined models provide relatively the same results as the original models in terms of the overall accuracy and the goodness-of-fit based on the four parameters examined: the overall ROC value, the overall accuracy level, the McFadden value, and the Cramer's V value. From Table 3-7, the refined single-family model has almost the same values for the overall ROC value and the McFadden value, while having higher values for the overall accuracy value and the Cramer's V value. For the multi-family use, in Table 3-7, the overall accuracy level is the same as the original model, while the overall ROC value, McFadden value, and Cramer's V value are slightly less than the original model. In terms of the commercial-institutional-transportation use, it presents the same pattern as the multi-family use, in which the overall accuracy value is the same as the original model, while the refined model has slightly lower values for the overall ROC value, McFadden value, and Cramer's V value. The industrial-warehouses use for the refined model has the same value as the original model for the overall accuracy value while it has a slightly higher value for the Cramer's V value and a slightly lower value for the overall ROC value and the McFadden value.

Overall, the refined model has advantages for the overall accuracy level as the single-family has a slightly higher overall accuracy value for the refined model while the other three uses have the same overall accuracy level. These values are explicable because the non-statistical-significance variables are excluded leaving the statistically significant variables in the model that raises the overall accuracy level. Because the difference between the original models and refined models are minor, this study uses the refined models in model building, which will input into the next step for the sensitivity analyses as well as the simulation of 2003 urban uses.

Single-family

The refined single-family model has five independent variables, which are (1) existing single-family land use in 1982 based on remote sensing classifications; (2) existing single-family land use in 2003 based on the parcel data; (3) proximity to 1982 road; (4) future land use for the single-family land use ; and (5) vacant land for the single-family land use in 2003. The cutting threshold for the refined single-family model is 0.6871. The overall accuracy for refined single-family is 97.99 percent. The refined 2003 predicted single-family probability map is shown in Figure 3-9. The refined 2003 single-family prediction map is shown in Figure 3-10. The independent variables for refined single-family model are shown in Table 3-8.

Multi-family

The refined multi-family model has six independent variables. They are (1) existing multi-family land use in 1982 based on remote sensing classifications; (2) existing multi-family land use in 1994 based on remote sensing classifications; (3) existing multi-family land use in 2003 based on the parcel data; (4) proximity to existing 1994 multi-family land use based on remote sensing classifications; (5) future land use for the multi-family land use; and (6) zoning for the multi-family land use. The cutting threshold for the refined multi-family model is 0.0334. The overall accuracy for refined multi-family is 98.83 percent. The refined 2003 predicted multi-family probability map is shown in Figure 3-11. The refined 2003 multi-family prediction map is shown in Figure 3-12. The independent variables for the refined multi-family model are shown in Table 3-9.

Commercial-institutional-transportation

The refined commercial-institutional-transportation model has eight independent variables. They are (1) existing commercial-institutional-transportation land use in 2003 based on the parcel data; (2) existing commercial-institutional-transportation land use in 1982 based on remote sensing classifications; (3) existing commercial-institutional-transportation land use in 1994 based on remote sensing classifications; (4) vacant land for the commercial-institutional-transportation land use in 2003 with ecosystem, parks, and conservations masked; (5) vacant land for the commercial-institutional-transportation land use in 1994 with ecosystem, parks, and conservations masked; (6) proximity to 2003 commercial-institutional-transportation land use; (7) proximity to 1994 commercial-institutional-transportation land use; and (8) future land use for the commercial-institutional-transportation land use. The cutting threshold for the refined commercial-institutional-transportation model is 0.1408. The overall accuracy for the refined multi-family is 99.53 percent. The refined 2003 predicted commercial-institutional-transportation probability map is shown in Figure 3-13. The refined 2003 commercial-institutional-transportation prediction map is shown in Figure 3-14. The independent variables for the refined commercial-institutional-transportation model are shown in Table 3-10.

Industrial-warehouses

The refined industrial-warehouses model has six independent variables, which are (1) existing industrial-warehouses use in 2003 based on the parcel data; (2) existing industrial-warehouses use in 1982 based on remote sensing classifications; (3) existing industrial-warehouses use in 1994 based on remote sensing classifications; (4) 1994 vacant land proximity to 1994 industrial-warehouses use; (5) future land use for the industrial-warehouses use; and (6) zoning for the industrial-warehouses use. The cutting threshold for the refined industrial-warehouses model is 0.1609.The overall accuracy for refined industrial-warehouses is 99.71 percent. The refined 2003 predicted industrial-warehouses probability map is shown in Figure 3-15. The refined 2003 industrial-warehouses prediction map is shown in Figure 3-16. The independent variables for the refined industrial-warehouses model are shown in Table 3-11.

Sensitivity Analysis

The sensitivity analysis is conducted based on the number of the dependent variables since the number of dependent variables determines the number of the sensitivity analysis. Because there are four dependent variables in this study, four sensitivity analyses are conducted based on three groups of parameters: the Chi-square value, the Pseudo R-Square values and the Cramer's

V value, and the ROC curve. The Chi-square value is from the Chi-square test, which is to calculate the difference between the predicted land use and the existing land use based on observed values and expected values. In the case of this study, the Chi-square test is conducted for the predicted 2003 land use changes and the observed 2003 land use changes based on the following algorithm:

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad (3\text{-}15)^{8}$$

Where O is the observed value and E is the expected value. E can be written as:

$$E = \frac{row\ marginal \times column\ marginal}{grand\ total} \quad (3\text{-}16)^{9}$$

The goodness-of-fit Pseudo R-Square values includes the McFadden value (McFadden, 1973), which is written as:

$$\rho^2 = 1 - \frac{LL(FITTED)}{LL(INTERCEPT\ ONLY)} \quad (3\text{-}17)^{10}$$

And:

$$LL(FITTED) = -2\sum_{i=1}^{n}\left[Y_i \ln\left(\hat{Y}_i\right) + (1-Y_i)\ln\left(1-\hat{Y}_i\right)\right] \quad (3\text{-}18)^{11}$$

Where Y_i refers to the observed value Y at location i ; $\hat{Y}_i$ refers to the predicted value Y at location i .

And:

$$LL(INTERCEPTONLY) = -2\left\{\left(n_{y=1}\right)\ln\left[P(Y=1)\right] + \left(n_{y=0}\right)\ln\left[P(Y=0)\right]\right\} \quad (3\text{-}19)^{12}$$

Where n refers to the number of cases, for which $y=1$ or $y=0$.

[8] Source: Sirkin, 1999
[9] Source: Sirkin, 1999
[10] Source: McFadden, 1973 and Ainsworth, no date
[11] Source: Ainsworth, no date
[12] Source: Ainsworth, no date

The optimal goodness-of-fit McFadden value is between 0.2 and 0.4 (Ainsworth, no date and Hu and Lo, 2007); the higher the value, the better the goodness-of-fit for a simulation model to fit the actual use.

The Cramer's V value is the most popular Chi-square-based test for nominal based correlation analysis (Garson, 2009). It has a value between -1 and 1, in which 1 and -1 refer to a perfect relationship between the independent variables and the dependent variable and 0 indicates no relations (Seaman, 2001). The algorithm for the Cramer's V test is:

$$V = \sqrt{\chi^2 / nm} \qquad (3\text{-}20)$$ [13]

where χ^2 is the Chi-square value; n is the number of sample size; and m is the smaller number of (rows - 1) or (columns - 1) (Cramér, 1999; Garson, 2009).

The ROC curve is to test the validation of a model of how actual changes interact with the predicted changes in a quantifiable manner through different scenarios. It is based on a 2 × 2 contingency table for an observed use (or actual change) versus a predicted use (or predicted change) in light of a dependent variable (Table 3-12). It calculates the true positive rate, A / (A+C), as well as false positive rate, B / (B+D), for each scenario and adds them up cumulatively to generate an overall ROC value, a single index, for a dependent variable (Pontius and Schneider, 2001). The optimal goodness-of-fit ROC value is from 0.5 to 1, with 1 indicating the perfect match between the predicted use and the actual use and 0.5 indicating a random distribution of the suitability and/or probability values across the landscape (Pontius and Schneider, 2001). The scenarios are a slice of the overall suitability or probability map into several equal interval groups, in which the highest probability group is usually assigned as group 1 and so on for the rest until the smallest probability is assigned. These scenarios in different

[13] Source: Cramér, 1999 and Garson, 2009

groups show a ROC curve to be calculated according to a typical dependent variable in light of each group's agreement with the actual land type. Because the ROC contingency table is used to represent two land categories, observed versus predicted, simulation of one land class will have one contingency table. As a result, the validation of more than one set of land types, e.g., four dependent variables in this study, needs four contingency tables and accordingly four ROC curves (Pontius and Schneider, 2001). In addition, the larger the number of groups, the higher the accuracy of the ROC curves (Pontius and Schneider, 2001). The ROC curve is generated in IDRISI. This study slices the probability map for each dependent variable into 20 groups. The 2 × 2 contingency table is presented in Table 3-12 and the overall ROC value is calculated as:

$$\sum_{i=1}^{n} \left[x_{i+1} - x_i\right] \left[y_i + y_{i+1} - \frac{y_i}{2}\right] \qquad (3\text{-}21)$$[14]

where x_i is the rate of false positives for scenario *i* and y_i is the rate of true positives for scenario *i*. The symbol *n* is the number of suitability groups.

Single-family

Chi-Square Test. Based on the predicted 2003 single-family use, compared to the existing 1994 single-family use, a prediction change map is produced. Likewise, compared to the existing 2003 single-family use and the 1994 single-family use, an observed change map is created. The changes versus non-changes in both the predicted and observed maps are summarized in Table 3-13.

Subsequently, the above values are transformed into a contingency table for expected values (Table 3-14) based on Equation 3-7 above. The calculation of expected values for the non-change category of single-family is:

---

[14] Source: Pontius and Schneider, 2001

$$E_0 = \frac{2{,}642{,}052 \times 2{,}637{,}476}{2{,}789{,}031} = \frac{6.968E12}{2{,}789{,}031} = 2{,}498{,}484$$

As a result, the expected value E_0 is presented in the contingency table as Table 3-14 illustrates.

At this time, a null hypothesis H_0 is assumed:

H_0: There is no relationship between the Observed and the Predicted.

The Chi-square value is tested to see if it can reject the null hypothesis. Therefore, the Chi-square value is calculated based on the observed table (Table 3-13) and the expected table (Table 3-14).

$O =$	$E =$	$O - E =$	$(O-E)^2 =$	$\frac{(O-E)^2}{E} =$
2,611,776	2,498,484	113,292	1.2935E10	5,137
30,276	143,568	-113,292	1.2935E10	89,401
25,700	138,992	-113,292	1.2935E10	92,344
121,279	7,987	113,292	1.2935E10	1,606,996

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 1{,}793{,}878$$

Because Table 3-13 and Table 3-14 are 2 × 2 tables, the df (degrees of freedom) value is 1. As a result, the critical value for 1 df is:

	$x^2_{critical}$		$x^2_{obtained}$	
$x^2_{critical}$	0.05 = 3.84	<	1,793,878	reject H_0
$x^2_{critical}$	0.01 = 6.64	<	1,793,878	
$x^2_{critical}$	0.001 = 10.83	<	1,793,878	

$x^2_{critical}$ 0.00000001 = 32.84 < 1,793,878 so P < 0.00000001 = 0.0000

From the above, it is obvious that the calculated Chi-square value is greater than the critical value at significance level of 0.05, which is 3.84; therefore, the null hypothesis is rejected. Also, there is a strong correlation between the predicted and observed values because the P value is close to 0. Thus, the results of the MLR model are reliable and can be used to predict the future single-family development for the county.

Pseudo R-Square and Cramer's V Tests. The McFadden Pseudo R-Square value as well as the Cramer's V value is illustrated in Table 3-15. The McFadden is 0.8204, which shows a perfect goodness-of-fit for the entire single-family use model to represent the actual use. The Cramer's V value is 0.8804, which is very close to 0.71 for a 2 × 2 table (Sirkin, 1999), showing a strong correlation between the dependent variable and the independent variables. The overall accuracy value is 97.99 percent, which is presented in Table 3-16.

ROC Curve. The ROC curve is generated based on 0.2 percent stratified random sampling with 20 equal interval thresholds. The number of cells that are selected for the ROC analysis is illustrated in the contingency table as Table 3-17 illustrates. The ROC table generated in IDRISI for the ROC analysis is further projected onto a chart in EXCEL, which is presented in Figure 3-17. The overall ROC value for 20 thresholds is 0.979, showing a fairly good agreement between the predicted single-family use based on five refined independent variables and the actual single-family use in the county.

Multi-family

Chi-Square Test. Similar to the single-family use, the Chi-square test for the multi-family use is based on the predicted 2003 multi-family use and the existing 1994 multi-family use. As a result, a prediction change map is produced. Likewise, compared to the existing 2003 multi-family use and the 1994 multi-family use, an observed change map is also generated. The predicted and observed maps for the changes versus non-changes are summarized in Table 3-18.

Based on Equation 3-7, an expected value for the non-change category is calculated as follows:

$$E_0 = \frac{2{,}750{,}841 \times 2{,}779{,}433}{2{,}789{,}031} = \frac{7.6458E12}{2{,}789{,}031} = 2{,}741{,}374$$

The above expected values can be listed in the contingency table as Table 3-19 illustrated. At this stage, a null hypothesis H_0 is assumed:

H_0: There is no relationship between the Observed and the Predicted.

The Chi-square value is calculated based on the above observed table as well as the expected table as follows to see if it can reject the null hypothesis:

$O =$	$E =$	$O - E =$	$(O-E)^2 =$	$\frac{(O-E)^2}{E} =$
2,748,811	2,741,374	7,437	55,308,969	20.18
2,030	9,467	-7,437	55,308,969	5842.29
30,622	38,059	-7,437	55,308,969	1453.24
7,568	131	7,437	55,308,969	422,205.87

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 429{,}522$$

Because Table 3-18 and Table 3-19 are 2 × 2 tables, the df (degrees of freedom) value is 1. As a result, the critical value for 1 df is:

	$x^2_{critical}$		$x^2_{obtained}$	
$x^2_{critical}$	0.05 = 3.84	<	429,522	reject H_0

$x^2_{critical}$ 0.01 = 6.64 < 429,522

$x^2_{critical}$ 0.001 = 10.83 < 429,522

$x^2_{critical}$ 0.00000001 = 32.84 < 429,522 so $P < 0.00000001 = 0.0000$

Based on the above, it is obvious that the calculated Chi-square value is greater than the critical value at significance level of 0.05, which is 3.84, and the null hypothesis is rejected. In addition, there is a strong correlation between the predicted and observed values because the calculated P value is close to 0. As a result, the logistic regression model is dependable to predict the future multi-family use in the county.

Pseudo R-Square and Cramer's V Tests. The IDRISI MULTILOGISTICREG module provides the McFadden Pseudo R-Square value in conjunction with the Cramer's V value. The McFadden is 0.3260, which is perfect goodness-of-fit for the multi-family model to simulate the actual use. The Cramer's V value is 0.5188, close to 0.71 for a 2 × 2 table (Sirkin, 1999). This shows a relatively strong correlation between the dependent variable and the independent variables. The Pseudo R-Square and Cramer's V values are presented in Table 3-20. The overall accuracy value for the multi-family use is 98.83 percent, which is presented in Table 3-21.

ROC Curve. The ROC analysis is conducted based on 1.5 percent stratified random sampling with 20 equal interval thresholds. The number of cells that are selected for the ROC analysis is presented in Table 3-22. The ROC table generated in IDRISI for the ROC analysis is projected onto a chart in EXCEL, presented in Figure 3-18. The overall ROC value for 20 thresholds is 0.588, which shows a not very strong goodness-of-fit between the predicted values and the actual values for the multi-family use. Because the ROC is more than 0.5, the goodness-of-fit for the multi-family use is still acceptable.

Commercial-institutional-transportation

Chi-Square Test. Similar to the single-family use and the multi-family use, the Chi-square test for the commercial-institutional-transportation use is based on the predicted 2003 commercial-institutional-transportation use and the existing 1994 commercial-institutional-transportation use. Consequently, a prediction change map is produced. Similarly, compared to the existing 2003 commercial-institutional-transportation use and the 1994 commercial-institutional-transportation use, an observed change map is also prepared. The predicted and observed maps for the changes versus non-changes are summarized in Table 3-23.

Similarly, an expected value for the non-change category is calculated based on Equation 3-7, as follows:

$$E_0 = \frac{1{,}324{,}203 \times 2{,}754{,}676}{2{,}789{,}031} = \frac{3.6478E12}{2{,}789{,}031} = 1{,}307{,}892$$

The above expected values can be listed in a contingency table (Table 3-24) for expected values as follows:

At this point, the null hypothesis H_0 is assumed as:

H_0: There is no relationship between the Observed and the Predicted.

The Chi-square value is calculated based on the above observed matrix as well as the expected matrix to test if it can reject the null hypothesis:

$O=$	$E=$	$O-E=$	$(O-E)^2=$	$\frac{(O-E)^2}{E}=$
1,302,828	1,307,892	-5,064	25,644,096	19.6072
21,375	16,311	5,064	25,644,096	1,572.1964
1,451,848	1,446,784	5,064	25,644,096	17.7249

12,980	18,044	-5,064	25,644,096	1421.1980

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 14{,}390{,}821$$

The df (degrees of freedom) value is 1 because Table 3-23 and Table 3-24 are 2 × 2 tables. In this case, the critical value for 1 df is:

	$x^2_{critical}$		$x^2_{obtained}$	
$x^2_{critical}$	0.05 = 3.84	<	14,390,821	reject H_0
$x^2_{critical}$	0.01 = 6.64	<	14,390,821	
$x^2_{critical}$	0.001 = 10.83	<	14,390,821	
$x^2_{critical}$	0.00000001 = 32.84	<	14,390,821	so P < 0.00000001 = 0.0000

Based on the above, it is obvious that the null hypothesis is rejected because the calculated Chi-square value is greater than the critical value at the significance level of 0.05. Also, there is a strong correlation between the predicted and observed values because the P value is close to 0. Thus, the logistic regression model is dependable to predict the commercial-institutional-transportation use in the study area.

Pseudo R-Square and Cramer's V Tests. The IDRISI MULTILOGISTICREG module provides the McFadden Pseudo R-Square value along with the Cramer's V value. The McFadden is calculated as 0.8014, which is perfect goodness-of-fit for the predicted commercial-institutional-transportation use to fit the actual use. The Cramer's V value is 0.8395, which shows a strong correlation between the dependent variable and the independent variables. These values are presented in Table 3-25. The overall accuracy value for the commercial-institutional-transportation use is 99.53 percent, which is presented in Table 3-26.

ROC Curve. Similar to the above single-family use and the multi-family use, the ROC analysis for the commercial-institutional-transportation use is preceded based on 8.5 percent stratified random sampling with 20 equal interval thresholds. The number of cells that are selected for the ROC analysis is presented in Table 3-27. The ROC table generated in IDRISI for the ROC analysis is projected in Figure 3-19. The overall ROC value for 20 thresholds is 0.853, a fairly good match between the predicted use based on eight refined independent variables and the actual commercial-institutional-transportation use.

Industrial-warehouses

Chi-Square Test. The Chi-square test for the industrial-warehouses use is based on the predicted 2003 industrial use and the existing 1994 industrial use. In this case, a prediction change map is produced. Also, compared to the existing 2003 industrial-warehouses and the 1994 industrial-warehouses use, an observed change map is created. The predicted and observed maps for the changes versus non-changes are summarized in Table 3-28.

An expected value for the non-change category is calculated based on Equation 3-7, as follows:

$$E_0 = \frac{2{,}778{,}612 \times 2{,}783{,}408}{2{,}789{,}031} = \frac{7.734E12}{2{,}789{,}031} = 2{,}773{,}010$$

The expected values are input into the contingency table (Table 3-29).

The null hypothesis H_0 is assumed at this point:

H_0: There is no relationship between the observed and the predicted.

The Chi-square value is calculated to test if it can reject the null hypothesis based on the above observed table as well as the expected table as follows:

$O =$	$E =$	$O - E =$	$(O - E)^2 =$	$\frac{(O - E)^2}{E} =$
2,777,009	2,773,010	3,999	15,992,001	5.7670
1,603	5,602	-3,999	15,992,001	2,855
6,399	10,398	-3,999	15,992,001	1,538
4,020	21	3,999	15,992,001	761,524

$$\chi^2 = \sum \frac{(O - E)^2}{E} = 764{,}385$$

Because Tables 3-28 and 3-29 are 2 × 2 tables, the df (degrees of freedom) value is 1. The critical value for 1 df is:

	$x^2_{critical}$		$x^2_{obtained}$	
$x^2_{critical}$	0.05 = 3.84	<	764,385	reject H_0
$x^2_{critical}$	0.01 = 6.64	<	764,385	
$x^2_{critical}$	0.001 = 10.83	<	764,385	
$x^2_{critical}$	0.00000001 = 32.84	<	764,385	so P < 0.00000001 = 0.0000

It is evident that the null hypothesis can be rejected because the Chi-square value is greater than the critical value at the significance level of 0.05. In addition, there is a strong correlation between the predicted and observed values because the P is close to 0. Therefore, the logistic regression model is dependable to predict the industrial-warehouses use in the study area.

Pseudo R-Square and Cramer's V Tests. The IDRISI MULTILOGISTICREG module provides the McFadden Pseudo R-Square value in addition to the Cramer's V value. The McFadden is 0.6259, which is perfect goodness-of-fit between the simulated industrial use and the observed use. The Cramer's V value is 0.6488, which is close to 0.71 for a 2 × 2 table

(Sirkin, 1999) and shows a strong correlation between the dependent variable and the independent variables. The above values are presented in Table 3-30. The overall accuracy value for the industrial-warehouses use is 99.71 percent, which is presented in Table 3-31.

ROC Curve. Similar to the above three uses, the ROC analysis for the industrial-warehouses use is preceded based on 15 percent stratified random sampling with 20 equal interval thresholds. The number of cells that are selected for ROC analysis is presented in Table 3-32. The ROC table generated in IDRISI for the ROC analysis is illustrated in Figure 3-20. The overall ROC value for 20 thresholds is 0.821, a fairly good match between the predicted industrial-warehouses use based on refined six independent variables and the actual industrial-warehouses use.

2003 Urban LULC Simulation

Because of the validities of the above models to simulate the four dependent variables, a final 2003 urban LULC prediction map is produced. The final 2003 prediction map essentially mosaics all the above four predicted urban land uses based on the cutting thresholds for each dependent variable provided by the refined models, along with mosaicing the remaining natural land. However, because of the data transfer from the IDRISI software, which causes the four dependent maps to be shifted in ArcGIS 9.3, i.e., those four predicted maps create gaps in IDRISI compared to their equivalent maps in ArcGIS 9.3, a geo-reference process is employed, which is operated in the ERDAS Imagine 9.1 software in order to shift the IDRISI maps back to the ArcGIS 9.3 geo-referencing. Then, these geo-referenced maps are mosaiced to create a final 2003 LULC prediction map, which is illustrated in Figure 3-21.

This study also conducts overall accuracy assessment for the simulated 2003 classification map. The overall accuracy level reaches 97.30 percent. The statistics regarding the confusion matrix of the 2003 simulation map is presented in Tables 3-33 through 3-37.

Table 3-1. Independent variables for the single-family use

Independent Variables	Data Type	Note	Coefficient	Odds Ratio	Standard Deviation	*Z*-Score	*P*-Value (Two-Tailed)
Existing SF in 1982 based on remote sensing classifications	Thematic data	1: SF; 0: Non-SF	0.869271	2.385172	0.187856	4.627328	<0.0001***
Existing SF in 1994 based on remote sensing classifications	Thematic data	1: SF; 0: Non-SF	0.13758	1.147494	0.253376	0.542988	0.5871
Existing SF in 2003 based on parcel data	Thematic data	1: SF; 0: Non-SF	8.463601	4739.091	0.29786	28.41469	<0.0001***
Proximity to industrial-warehouses in 1994	Thematic data	1: Out of 1,000 meter radius; 0: Within 1,000 meter radius*	-0.091319	0.912727	0.500023	-0.18263	0.8551
Proximity to existing 1982 SF based on remote sensing classifications	Thematic data	1: Within 150 meter radius; 0: Out of 150 meter radius**	0.089720	1.093868	0.4031	0.222574	0.8239

Table 3-1. Continued

Independent Variables	Data Type	Note	Coefficient	Odds Ratio	Standard Deviation	*Z*-Score	*P*-Value (Two-Tailed)
Proximity to existing 1994 SF based on remote sensing classifications	Thematic Data	1: Within 50 Meter Radius; 0: Out Of 50 Meter Radius*	0.522662	1.686511	0.342529	1.525891	0.127
Proximity to 1982 roads	Thematic Data	1: Within 530 Meter Radius; 0: Out Of 530 Meter Radius**	0.859717	2.362492	0.4885	1.759912	0.0784***
Proximity to 1994 road	Thematic Data	1: Within 520 Meter Radius; 0: Out Of 520 Meter Radius**	-0.743280	0.475552	0.488734	-1.52083	0.1283
Future land use for SF	Thematic Data	1: Future SF Use; 0: Future Non-SF Use	1.638371	5.146778	0.295178	5.550451	<0.0001***
1994 single-family density	Thematic Data	1: Areas Surrounded By High Density SF Houses; 0: Other Areas	-0.329253	0.719461	0.473763	-0.69498	0.4871
Proximity to schools	Thematic Data	1: Areas Within 5,000 Meter Radius; 0: Areas Out Of 5,000 Meter Radius**	0.264259	1.302465	0.482561	0.547617	0.584

Table 3-1. Continued

Independent Variables	Data Type	Note	Coefficient	Odds Ratio	Standard Deviation	*Z*-Score	*P*-Value (Two-Tailed)
Located within urban cluster areas	Thematic Data	1: Within Urban Cluster Areas; 0: Out Of Urban Cluster Areas	-0.595032	0.551545	0.440117	-1.35199	0.1764
Zoning for SF	Thematic Data	1: Zoned For SF; 0: Zoned For Non-SF	-0.02532	0.975002	0.31321	-0.08083	0.9356
Vacant land for SF in 2003	Thematic Data	1: Vacant Land In 2003; 2: Occupied Land In 2003	2.191825	8.951535	0.258796	8.469316	<0.0001***

Note: *: Based on zonal statistical mean value

**: Based on arbitrary partition

*** Less than 0.10 for two-tailed *P*-values, which is statistical significance for the independent variable

Table 3-2. Independent variables for the multi-family use

Independent Variables	Data Type	Note	Coefficient	Odds Ratio	Standard Deviation	*Z*-Score	*P*-Value (Two-Tailed)
Existing MF in 1982 based on remote sensing classifications	Thematic Data	1: MF; 0: Non-MF	2.280251	9.779132	0.071839	31.741126	<0.0001***
Existing MF in 1994 based on remote sensing classifications	Thematic Data	1: MF; 0: Non-MF	1.851125	6.366975	0.071839	25.767683	<0.0001***
Existing MF in 2003 based on parcel data	Thematic Data	1: MF; 0: Non-MF	7.202940	1343.374633	0.064695	111.336890	<0.0001***
Vacant land for MF in 1994 with ecosystems, parks, and conservations masked	Thematic Data	1: Vacant For MF; 0: Occupied	0.360957	1.434702	0.498237	0.724468	0.4688
Proximity to existing 1994 MF based on remote sensing classifications	Thematic Data	1: Within 240 Meter Radius; 0: Out Of 240 Meter Radius*	1.676713	5.347947	0.37419	4.480913	0.0101

Table 3-2. Continued

Independent Variables	Data Type	Note	Coefficient	Odds Ratio	Standard Deviation	*Z*-Score	*P*-Value (Two-Tailed)
Proximity to major roads	Thematic Data	1: Within 500 Meter Radius; 0: Out Of 500 Meter Radius**	0.671131	1.956448	0.485658	1.381900	0.167
Future land use for MF	Thematic Data	1: Future MF Use; 0: Future Non-MF Use	0.340167	1.405182	0.074757	4.550300	<0.0001***
Zoning for MF	Thematic Data	1: Zoned For MF; 0: Zoned For Non-MF	1.089908	2.973400	0.073797	14.769000	<0.0001***

Note: *: Based on zonal statistical mean value
**: Based on arbitrary partition
*** Less than 0.10 for two-tailed *P*-values, which is statistical significance for the independent variable.

Table 3-3. Independent variables for the commercial-institutional-transportation use

Independent Variables	Data Type	Note	Coefficient	Odds Ratio	Standard Deviation	*Z*-Score	*P*-Value (Two-Tailed)
Existing COM-INST-TRANS in 2003 based on the parcel data	Thematic Data	1: COM-INST-TRANS; 0: Non-COM-INST-TRANS	5.005301	149.202003	0.134753	37.144265	<0.0001****
Existing COM-INST-TRANS in 1982 based on remote sensing classifications	Thematic Data	1: COM-INST-TRANS; 0: Non-COM-INST-TRANS	1.403852	4.070849	0.116184	12.083003	<0.0001****
Existing COM-INST-TRANS in 1994 based on remote sensing classifications	Thematic Data	1: COM-INST-TRANS; 0: Non-COM-INST-TRANS	3.146571	23.256183	0.130084	24.188763	<0.0001****
Vacant land for COM-INST-TRANS in 2003 with ecosystem, parks, and conservations masked	Thematic Data	1: Vacant For COM-INST-TRANS; 0: Occupied	-25.076209	1.28689E-11	0.498138	-50.339883	<0.0001****

Table 3-3. Continued

Independent Variables	Data Type	Note	Coefficient	Odds Ratio	Standard Deviation	*Z*-Score	*P*-Value (Two-Tailed)
Vacant land for COM-INST-TRANS in 1994 with ecosystem, parks, and conservations masked	Thematic Data	1: Vacant For COM-INST-TRANS; 0: Occupied	1.853776	6.383880	0.496737	3.731907	0.0002****
1994 vacant land proximity to 2003 SF and MF uses	Thematic Data	1: Within 250 Meter Radius; 0: Out Of 250 Meter Radius*	1.53E-01	1.164872	4.95E-01	0.308204	0.7579
1994 vacant land proximity to 1994 SF and MF uses	Thematic Data	1: Within 300 Meter Radius; 0: Out Of 300 Meter Radius*	-0.265254	0.767011	0.496888	-0.533831	0.5935
Proximity to 2003 COM-INST-TRANS	Thematic Data	1: Within 630 Meter Radius; 0: Out Of 630 Meter Radius*	20.095309	533680983.9	0.498258	40.331131	<0.0001****
Proximity to 1994 COM-INST-TRANS	Thematic Data	1: Within 550 Meter Radius; 0: Out Of 550 Meter Radius*	1.240891	3.458692	0.499053	2.486491	0.0129****
Proximity to 2003 TIGER roads	Thematic Data	1: Within 300 Meter Radius; 0: Out Of 300 Meter Radius**	0.354306	1.425191	0.470957	0.752311	0.4519

Table -3-3. Continued

Independent Variables	Data Type	Note	Coefficient	Odds Ratio	Standard Deviation	*Z*-Score	*P*-Value (Two-Tailed)
Proximity to major roads	Thematic Data	1: Within 1,000 Meter Radius; 0: Out Of 1,000 Meter Radius**	N/A	N/A	N/A	N/A	N/A
Future land use for COM-INST-TRANS	Thematic Data	1: Future COM-INST-TRANS Use; 0: Future Non-COM-INST-TRANS Use	0.429676	1.536760	0.192272	2.234730	0.0254****
Proximity to road intersection	Thematic Data	1: Within2,000 Meter Radius; 0: Out Of 2,000 Meter Radius***	0.046365	1.047457	0.499109	0.09290	0.926
Zoning for COM-INST-TRANS	Thematic Data	1: Zoned For COM-INST-TRANS; 0: Zoned For Non-COM-INST-TRANS	-0.055038	0.946450	0.4603	0.147405	0.7089
2003 vacant land proximity to 2003 SF and MF uses	Thematic Data	1: Within 250 Meter Radius; 0: Out Of 250 Meter Radius*	N/A	N/A	N/A	N/A	N/A
2003 vacant land proximity to 2003 MF uses	Thematic Data	1: Within 400 Meter Radius; 0: Out Of 400 Meter Radius*	0.184621	1.202763	0.489345	0.377282	0.706

Note: *: Based on zonal statistical mean value.
**: Based on arbitrary partition.
***: Average value.
****: Less than 0.10 for two-tailed *P*-values, which is statistical significance for the independent variable.

Table 3-4. Independent variables for the industrial-warehouses use

Independent Variables	Data Type	Note	Coefficient	Odds Ratio	Standard Deviation	*Z*-Score	*P*-Value (Two-Tailed)
2003 industrial proximity to major roads	Thematic Data	1: Within 400 Meter Radius; 0: Out Of 400 Meter Radius*	0.131403	1.140423	0.467009	0.281372	0.7784
Existing industrial in 2003 based on parcel	Thematic Data	1: Industrial; 0: Non-Industrial	4.010653	55.182906	0.0194	206.734703	<0.0001****
Existing industrial in 1982 based on remote sensing classification	Thematic Data	1: Industrial; 0: Non-Industrial	2.398089	11.002126	0.053974	44.430439	<0.0001****
Proximity to existing residential uses	Thematic Data	1: Out Of 160 Meter Radius; 0: Within 160 Meter Radius*	-0.481350	0.617948	0.493858	-0.974674	0.3297
Existing industrial in 1994 based on remote sensing classification	Thematic Data	1: Industrial; 0: Non-Industrial	5.16036964	174.228846	0.069397	74.360126	<0.0001****

Table 3-4. Continued

Independent Variables	Data Type	Note	Coefficient	Odds Ratio	Standard Deviation	*Z*-Score	*P*-Value (Two-Tailed)
1994 vacant land proximity to 1994 industrial	Thematic Data	1: Within 1,050 Meter Radius; 0: Out Of 1,050 Meter Radius*	2.719848	15.178008	0.454024	5.990537	<0.0001****
Future land use for industrial	Thematic Data	1: Industrial; 0: Non-Industrial	1.336143	3.804342	0.123114	10.852894	<0.0001****
Zoning for industrial	Thematic Data	1: Industrial; 0: Non-Industrial	2.984637	19.779312	0.111794	26.697645	<0.0001****

Note: *: Based on zonal statistical mean value.
****: Less than 0.10 for two-tailed *P*-values, which is statistical significance for the independent variable.

Table 3-5. Autocorrelation and four dependent variables for the Moran's I test

Dependent Variables	With Stratified Sampling			Without Stratified Sampling	
	Percent Of Sampling	Moran's I Value	Character	Moran's I Value	Character
2003 SF	0.2%	-0.04	Random	0.8064	Clustered
2003 MF	1.5%	0.06	Random	0.7937	Clustered
2003 COM-INST-TRANS	8.5%	0.05	Random	0.6498	Clustered
2003 INDUS-WARE	15%	-0.04	Random	0.7617	Clustered

Table 3-6. Autocorrelation and four dependent variables for the Z-score test

Dependent Variables	With Stratified Sampling	Without Stratified Sampling (Normality Assumption)
2003 SF	-0.62	2,241.4285
2003 MF	0.3	2,206.2896
2003 COM-INST-TRANS	0.85	1,806.1605
2003 INDUS-WARE	-0.69	2,117.3811

Table 3-7. Comparison between the original models and the refined models

	SF		MF		COM-INST-TRANS		INDUS-WARE	
Parameters	Original Model	Refined Model	Original Model	Refined Model	Original Model	Refined Model	Original Model	Refined Model
Overall ROC*	0.9922	0.9913	0.8420	0.8140	0.9933	0.9895	0.9628	0.9505
Overall Accuracy	97.95%	97.99%	98.83%	98.83%	99.53%	99.53%	99.71%	99.71%
McFadden	0.8243	0.8204	0.3344	0.3260	0.8018	0.8014	0.6284	0.6259
Cramer's V	0.8761	0.8804	0.5190	0.5188	0.8403	0.8395	0.6465	0.6488

*: Based on 100 intervals.

Table 3-8. Independent variables for refined single-family use

Independent Variables	Data Type	Note	Coefficient	Odds Ratio	Standard Deviation	*Z*-Score	*P*-Value (Two-Tailed)
Existing SF in 1982 based on remote sensing classifications	Thematic Data	1: SF; 0: Non-SF	1.168234	3.216307	0.187856	6.218773	<0.0001***
Existing SF in 2003 based on the parcel data	Thematic Data	1: SF; 0: Non-SF	8.709365	6,059.393491	0.297860	29.239794	<0.0001***
Proximity to 1982 road	Thematic Data	1: Within 530 Meter Radius; 0: Out Of 530 Meter Radius**	0.158141	1.171332	0.488500	0.323728	0.7461
Future land use for SF	Thematic Data	1: Future SF Use; 0: Future Non-SF Use	1.161380	3.194339	0.295178	3.934508	<0.0001***
Vacant land for SF in 2003	Thematic Data	1: Vacant Land In 2003; 2: Occupied Land In 2003	2.293864	9.913165	0.258796	8.863598	<0.0001***

Note: **: Based on zonal statistical mean value
*** Less than 0.10 for two-tailed *P*-values, which is statistical significance for the independent variable.

Table 3-9. Independent variables for refined multi-family use

Independent Variables	Data Type	Note	Coefficient	Odds Ratio	Standard Deviation	*Z*-Score	*P*-Value (Two-Tailed)
Existing MF in 1982 based on remote sensing classifications	Thematic Data	1: MF; 0: Non-MF	2.369196	10.688797	0.071839	32.979248	<0.0001***
Existing MF in 1994 based on remote sensing classifications	Thematic Data	1: MF; 0: Non-MF	1.700038	5.474156	0.071839	23.664557	<0.0001***
Existing MF in 2003 based on the parcel data	Thematic Data	1: MF; 0: Non-MF	7.414149	1,659.296689	0.064695	114.601578	<0.0001***
Proximity to existing 1994 MF based on remote sensing classifications	Thematic Data	1: Within 240 Meter Radius; 0: Out Of 240 Meter Radius*	1.896014	6.659295	0.374190	5.066981	<0.0001***
Future land use for MF	Thematic Data	1: Future MF Use; 0: Future Non-MF Use	0.382844	1.466449	0.074757	5.121180	<0.0001***
Zoning for MF	Thematic Data	1: Zoned For MF; 0: Zoned For Non-MF	1.249038	3.486986	0.073797	16.925321	<0.0001***

Note: *: Based on zonal statistical mean value
*** Less than 0.10 for two-tailed P-values, which is statistical significance for the independent variable

Table 3-10. Independent variables for refined commercial-institutional-transportation use

Independent Variables	Data Type	Note	Coefficient	Odds Ratio	Standard Deviation	*Z*-Score	*P*-Value (Two-Tailed)
Existing COM-INST-TRANS in 2003 based on the parcel data	Thematic Data	1: COM-INST-TRANS; 0: Non-COM-INST-TRANS	5.007813	149.577323	0.134753	37.162909	<0.0001****
Existing COM-INST-TRANS in 1982 based on remote sensing classifications	Thematic Data	1: COM-INST-TRANS; 0: Non-COM-INST-TRANS	1.429635	4.177176	0.116184	12.304925	<0.0001****
Existing COM-INST-TRANS in 1994 based on remote sensing classifications	Thematic Data	1: COM-INST-TRANS; 0: Non-COM-INST-TRANS	3.188094	24.242178	0.130084	24.507964	<0.0001****
Vacant land for COM-INST-TRANS in 2003 with ecosystem, parks, and conservations masked	Thematic Data	1: Vacant For COM-INST-TRANS; 0: Occupied	-25.137195	0.000000	0.498138	-50.462313	<0.0001****

Table 3-10. Continued

Independent Variables	Data Type	Note	Coefficient	Odds Ratio	Standard Deviation	*Z*-Score	*P*-Value (Two-Tailed)
Vacant land for COM-INST-TRANS in 1994 with ecosystem, parks, and conservations masked	Thematic Data	1: Vacant For COM-INST-TRANS; 0: Occupied	1.874564	6.517974	0.496737	3.773755	0.0002****
Proximity to 2003 COM-INST-TRANS	Thematic Data	1: Within 630 Meter Radius; 0: Out Of 630 Meter Radius*	20.159299	568,947,659	0.498258	40.459559	<0.0001****
Proximity to 1994 COM-INST-TRANS	Thematic Data	1: Within 550 Meter Radius; 0: Out Of 550 Meter Radius*	1.308222	3.699589	0.499053	2.621408	0.0088****
Future land use for COM-INST-TRANS	Thematic Data	1: Future COM-INST-TRANS Use; 0: Future Non-COM-INST-TRANS Use	0.419155	1.520676	0.192272	2.180010	0.0293****

Note: *: Based on zonal statistical mean value.
****: Less than 0.10 for two-tailed P-values, which is statistical significance for the independent variable

Table 3-11. Independent variables for refined industrial-warehouses use

Independent Variables	Data Type	Note	Coefficient	Odds Ratio	Standard Deviation	*Z*-Score	*P*-Value (Two-Tailed)
Existing industrial-warehouses in 2003 based on the parcel data	Thematic Data	1: Industrial; 0: Non-Industrial	4.209021	67.290613	0.019400	216.959832	<0.0001****
Existing industrial-warehouses in 1982 based on remote sensing classifications	Thematic Data	1: Industrial; 0: Non-Industrial	2.408243	11.114410	0.053974	44.618566	<0.0001****
Existing industrial-warehouses in 1994 based on remote sensing classifications	Thematic Data	1: Industrial; 0: Non-Industrial	5.122834	167.810220	0.069397	73.819239	<0.0001****
1994 vacant land proximity to 1994 industrial-warehouses	Thematic Data	1: Within 1,050 Meter Radius; 0: Out Of 1,050 Meter Radius*	2.811813	16.640055	0.454024	6.193093	<0.0001****
Future land use for industrial-warehouses	Thematic Data	1: Industrial; 0: Non-Industrial	1.304183	3.684679	0.123114	10.593299	<0.0001****
Zoning for industrial-warehouses	Thematic Data	1: Industrial; 0: Non-Industrial	3.106883	22.351276	0.111794	27.791146	<0.0001****

Note: *: Based on zonal statistical mean value.
****: Less than 0.10 for two-tailed P-values, which is statistical significance for the independent variable.

Table 3-12. 2 × 2 contingency table for the ROC curve

Model	Reality		
	Urban Land Class (1)	Non-Urban Land Class (0)	Total
Urban Land Class (1)	A	B	A+B
Non-Urban Land Class (0)	C	D	C+D
Total	A+C	B+D	A+B+C+D

Table 3-13. Contingency table of change versus non-change for observed values for the single-family use (unit: cells)

Observed	Prediction		
	Non-Change	Change	Marginal
Non-Change	2,611,776	30,276	2,642,052
Change	25,700	121,279	146,979
Marginal	2,637,476	151,555	2,789,031

Table 3-14. Contingency table of change versus non-change for expected values for the single-family use (unit: cells)

Expected	Prediction		
	Non-Change	Change	Marginal
Non-Change	2,498,484	143,568	2,642,052
Change	138,992	7,987	146,979
Marginal	2,637,476	151,555	2,789,031

Table 3-15. Pseudo R-Square and Cramer's V values for the single-family use

Goodness of Fit (Parameters)	Results
McFadden	0.8204
P-Level	0.0000
Cramer's V	0.8804

Table 3-16. Overall accuracy for the single-family use

Observed	Prediction		
	Cat 0 (non-SF)	Cat 1 (SF)	% Correct
Cat 0 (non-SF)	2,516,623	53,545	97.92%
Cat 1 (SF)	2,431	216,432	98.89%
Overall %	90.32%	9.68 %	97.99%

Table 3-17. ROC analysis and 2 × 2 contingency table based on 0.2 percent stratified sampling for the single-family use

Simulated by Threshold	Reality (Reference Image)	
	1 (SF)	0 (Non-SF)
1 (Sf)	A (number of cells)	B (number of cells)
0 (Non-SF)	C	D
For the Given Reference Image:	A+C= 454	B+D= 7,467

Table 3-18. Contingency table of change versus non-change for observed values for the multi-family use (unit: cells)

Observed	Prediction		
	Non-Change	Change	Marginal
Non-Change	2,748,811	2,030	2,750,841
Change	30,622	7,568	38,190
Marginal	2,779,433	9,598	2,789,031

Table 3-19. Contingency table of change versus non-change for expected values for the multi-family use (unit: cells)

Expected	Prediction		
	Non-Change	Change	Marginal
Non-Change	2,741,374	9,467	2,750,841
Change	38,059	131	38,190
Marginal	2,779,433	9,598	2,789,031

Table 3-20. Pseudo R-Square and Cramer’s V values for the multi-family use

Goodness of Fit (Parameters)	Results
McFadden	0.3260
P-Level	0.0000
Cramer's V	0.5188

Table 3-21. Overall accuracy for the multi-family use

Observed	Prediction		
	Cat 0 (Non-SF)	Cat 1 (SF)	% Correct
Cat 0 (Non-SF)	2,743,136	1,107	99.96%
Cat 1 (SF)	31,545	13,243	29.57%
Overall %	99.49%	0.51 %	98.83%

Table 3-22. ROC analysis and 2 × 2 contingency table based on 1.5 percent stratified sampling for the multi-family use

Simulated by Threshold	Reality (Reference Image)	
	1 (MF)	0 (Non-MF)
1 (MF)	A (number of cells)	B (number of cells)
0 (Non-MF)	C	D
For the Given Reference Image:	A+C= 425	B+D= 38,187

Table 3-23. Contingency table of change versus non-change for observed values for the commercial-institutional-transportation use (unit: cells)

Observed	Prediction		
	Non-Change	Change	Marginal
Non-Change	1,302,828	21,375	1,324,203
Change	1,451,848	12,980	1,464,828
Marginal	2,754,676	34,355	2,789,031

Table 3-24. Contingency table of change versus non-change for expected values for the commercial-institutional-transportation use (unit: cells)

Expected	Prediction		
	Non-Change	Change	Marginal
Non-Change	1,307,892	16,311	1,324,203
Change	1,446,784	18,044	1,464,828
Marginal	2,754,676	34,355	2,789,031

Table 3-25. Pseudo R-Square and Cramer's V Values for the commercial-institutional-transportation use

Goodness of Fit (Parameters)	Results
McFadden	0.8014
P-Level	0.0000
Cramer's V	0.8395

Table 3-26. Overall accuracy for the commercial-institutional-transportation use

Observed	Prediction		
	Cat 0 (Non-COM)	Cat 1 (COM)	% Correct
Cat 0 (Non-COM)	2,741,496	2,823	99.90%
Cat 1 (COM)	10,420	34,292	76.70%
Overall %	98.67%	1.33 %	99.53%

Table 3-27. ROC analysis and 2 × 2 contingency table based on 8.5 percent stratified sampling for the commercial-institutional-transportation use

Simulated By Threshold	Reality (Reference Image)	
	1 (COM)	0 (Non-COM)
1 (COM)	A (number of cells)	B (number of cells)
0 (Non-COM)	C	D
For the Given Reference Image:	A+C=3,603	B+D=305,688

Table 3-28. Contingency table of change versus non-change for observed values for the industrial-warehouses use (unit: cells)

Observed		Prediction	
	Non-Change	Change	Marginal
Non-Change	2,777,009	1,603	2,778,612
Change	6,399	4,020	10,419
Marginal	2,783,408	5,623	2,789,031

Table 3-29. Contingency table of change versus non-change for expected values for the industrial-warehouses use (unit: cells)

Expected		Prediction	
	Non-Change	Change	Marginal
Non-Change	2,773,010	5,602	2,778,612
Change	10,398	21	10,419
Marginal	2,783,408	5,623	2,789,031

Table 3-30. Pseudo R-Square and Cramer's V Values for the industrial-warehouses use

Goodness of Fit (Parameters)	Results
McFadden	0.6259
P-Level	0.0000
Cramer's V	0.6488

Table 3-31. Overall accuracy for the industrial-warehouses use

Observed	Prediction		
	Cat 0 (Non-COM)	Cat 1 (COM)	% Correct
Cat 0 (Non-COM)	2,774,034	1,738	99.94%
Cat 1 (COM)	6,264	6,995	52.76%
Overall %	99.69%	0.31 %	99.71%

Table 3-32. ROC analysis and 2 × 2 contingency table based on 15 percent stratified sampling for the industrial-warehouses use

Simulated By Threshold	Reality (Reference Image)	
	1 (INDUS)	0 (Non-INDUS)
1 (INDUS)	A (number of cells)	B (number of cells)
0 (Non-INDUS)	C	D
For the Given Reference Image:	A+C=1,999	B+D=577,774

Table 3-33. Accuracy assessment for 2003 urban LULC simulation map based on predicted pixels (1)

Class	Single-Family	Multi-Family	Commercial-Institutional-Transportation	Industrial-Warehouses	Grasslands	Forests	Agricultural	Recreational-Others	Wetlands	Water	Barren	Total
Single-Family	218,674	716	0	0	33,765	9,098	88	82	6,137	827	0	269,387
Multi-Family	38	39,533	676	36	161	816	52	72	698	50	1	42,133
Commercial-Institutional-Transportation	114	3,766	96,836	1,276	4,785	9,259	1,331	2,243	3,670	531	43	123,854
Industrial-Warehouses	37	773	1,460	11,362	3,338	10,050	4,748	92	2,472	1,020	204	35,556
Grasslands	0	0	0	0	522,074	0	0	0	0	0	0	522,074
Forests	0	0	0	0	0	1,021,887	0	0	0	0	0	1,021,887
Agricultural	0	0	0	0	0	0	140,794	0	0	0	0	140,794
Recreational-Others	0	0	0	0	0	0	0	128,462	0	0	0	128,462
Wetlands	0	0	0	0	0	0	0	0	390,722	0	0	390,722
Water	0	0	0	0	0	0	0	0	0	111,787	0	111,787
Barren	0	0	0	0	0	0	0	0	0	0	2,375	2,375
Total	218,863	44,788	98,972	12,674	564,123	1,051,110	147,013	130,951	403,699	114,215	2,623	2,789,031

Overall Accuracy = 3,760,630/3,865,155 = 97.2957%

Kappa Coefficient = 0.9667

Table 3-34. Accuracy assessment for 2003 urban LULC simulation map based on predicted percent (2)

Class	Single-Family	Multi-Family	Commercial-Institutional-Transportation	Industrial-Warehouses	Grasslands	Forests	Agricultural	Recreational-Others	Wetlands	Water	Barren	Total
Single-Family	99.91	1.60	0	0	5.99	0.87	0.06	0.06	1.52	0.72	0	9.66
Multi-Family	0.02	88.27	0.68	0.28	0.03	0.08	0.04	0.05	0.17	0.04	0.04	1.51
Commercial-Institutional-Transportation	0.05	8.41	97.84	10.07	0.85	0.88	0.91	1.71	0.91	0.46	1.64	4.44
Industrial-Warehouses	0.02	1.73	1.48	89.65	0.59	0.96	3.23	0.07	0.61	0.89	7.78	1.27
Grasslands	0	0	0	0	92.55	0	0	0	0	0	0	18.72
Forests	0	0	0	0	0	97.22	0	0	0	0	0	36.64
Agricultural	0	0	0	0	0	0	95.77	0	0	0	0	5.05
Recreational-Others	0	0	0	0	0	0	0	98.10	0	0	0	4.61
Wetlands	0	0	0	0	0	0	0	0	96.79	0	0	14.01
Water	0	0	0	0	0	0	0	0	0	97.87	0	4.01
Barren	0	0	0	0	0	0	0	0	0	0	90.55	0.09
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 3-35. Accuracy assessment for 2003 urban LULC simulation map (3)

Class	Predicted (Pixels)	Predicted (Percent)	Observed (Pixels)	Observed (Percent)
Single-Family	218,674	9.66%	218,863	7.85%
Multi-Family	39,533	1.51%	44,788	1.61%
Commercial-Institutional-Transportation	96,836	4.44%	98,972	3.55%
Industrial-Warehouses	11,362	1.27%	12,674	0.45%
Grasslands	522,074	18.72%	564,123	20.23%
Forests	1,021,887	36.64%	1,051,110	37.69%
Agricultural	140,794	5.05%	147,013	5.27%
Recreational-Others	128,462	4.61%	130,951	4.70%
Wetlands	390,722	14.01%	403,699	14.47%
Water	111,787	4.01%	114,215	4.10%
Barren	2,375	0.09%	2,623	0.09%
Total	2,684,506	100.00%	2,789,031	100.00%

Table 3-36. Accuracy assessment for 2003 urban LULC simulation map (4)

Class	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Single-Family	18.83%	0.09%	50,713/269,387	189/218,863
Multi-Family	6.17%	11.73%	2,600/42,133	5,255/44,788
Commercial-Institutional-Transportation	21.81%	2.16%	27,018/123,854	2,136/98,972
Industrial-Warehouses	68.04%	10.35%	24,194/35,556	1,312/12,674
Grasslands	0.00%	7.45%	0/522,074	42,049/564,123
Forests	0.00%	2.78%	0/1,021,887	29,223/1,051,110
Agricultural	0.00%	4.23%	0/140,794	6,219/147,013
Recreational-Others	0.00%	1.90%	0/128,462	2,489/130,951
Wetlands	0.00%	3.21%	0/390,722	12,977/403,699
Water	0.00%	2.13%	0/111,787	2,428/114,215
Barren	0.00%	9.45%	0/2,375	248/2,623

Table 3-37. Accuracy assessment for 2003 urban LULC simulation map (5)

Class	Prod. Acc.(Percent)	User Acc.(Percent)	Prod. Acc.(Pixels)	User Acc.(Pixels)
Single-Family	99.91%	81.17%	218,674/218,863	218,674/269,387
Multi-Family	88.27%	93.83%	39,533/44,788	39,533/42,133
Commercial-Institutional-Transportation	97.84%	78.19%	96,836/98,972	96,836/123,854
Industrial-Warehouses	89.65%	31.96%	11,362/12,674	11,362/35,556
Grasslands	92.55%	100.00%	522,074/564,123	522,074/522,074
Forests	97.22%	100.00%	1,021,887/1,051,110	1,021,887/1,021,887
Agricultural	95.77%	100.00%	140,794/147,013	140,794/140,794
Recreational-Others	98.10%	100.00%	128,462/130,951	128,462/128,462
Wetlands	96.79%	100.00%	390,722/403,699	390,722/390,722
Water	97.87%	100.00%	111,787/114,215	111,787/111,787
Barren	90.55%	100.00%	2,375/2,623	2,375/2,375

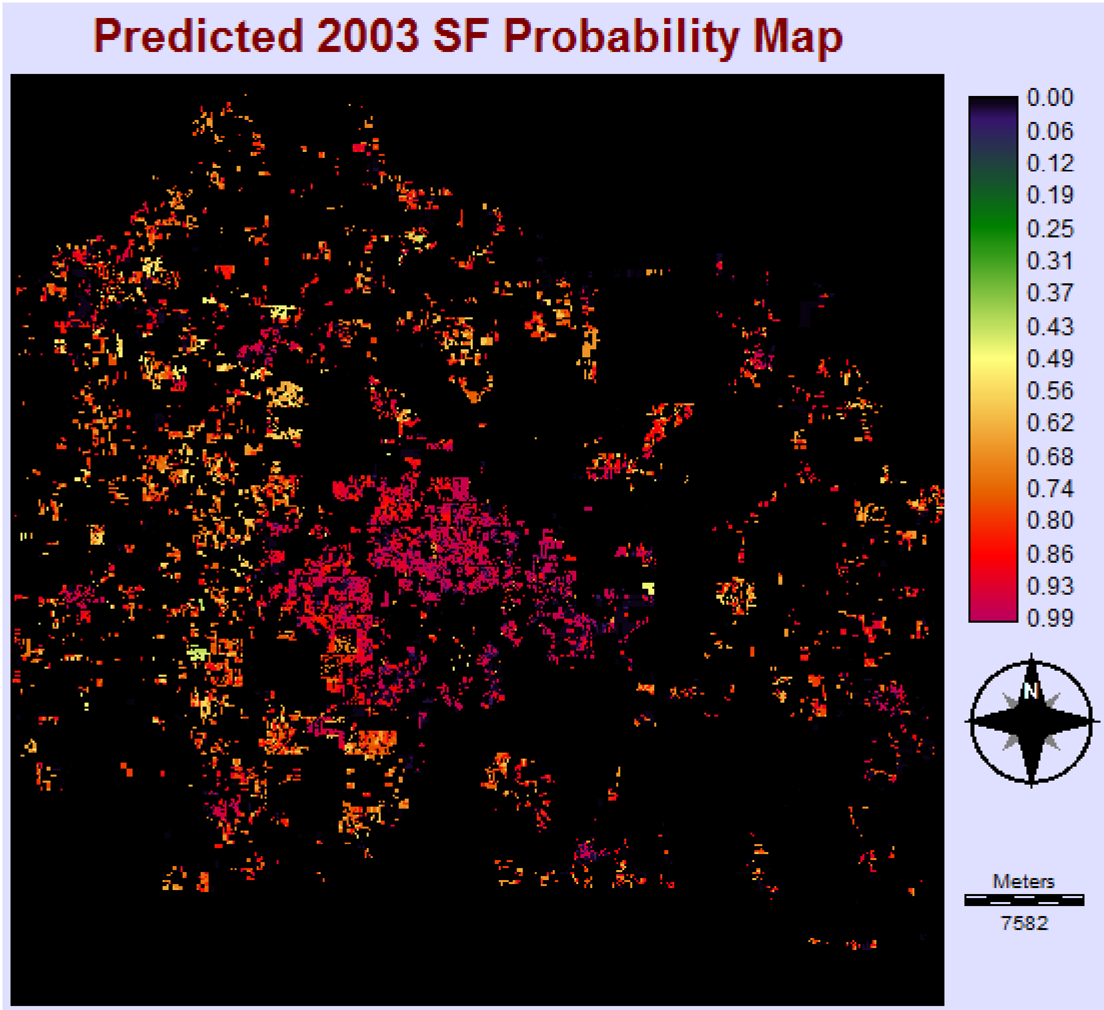

Figure 3-1. Predicted 2003 single-family probability map

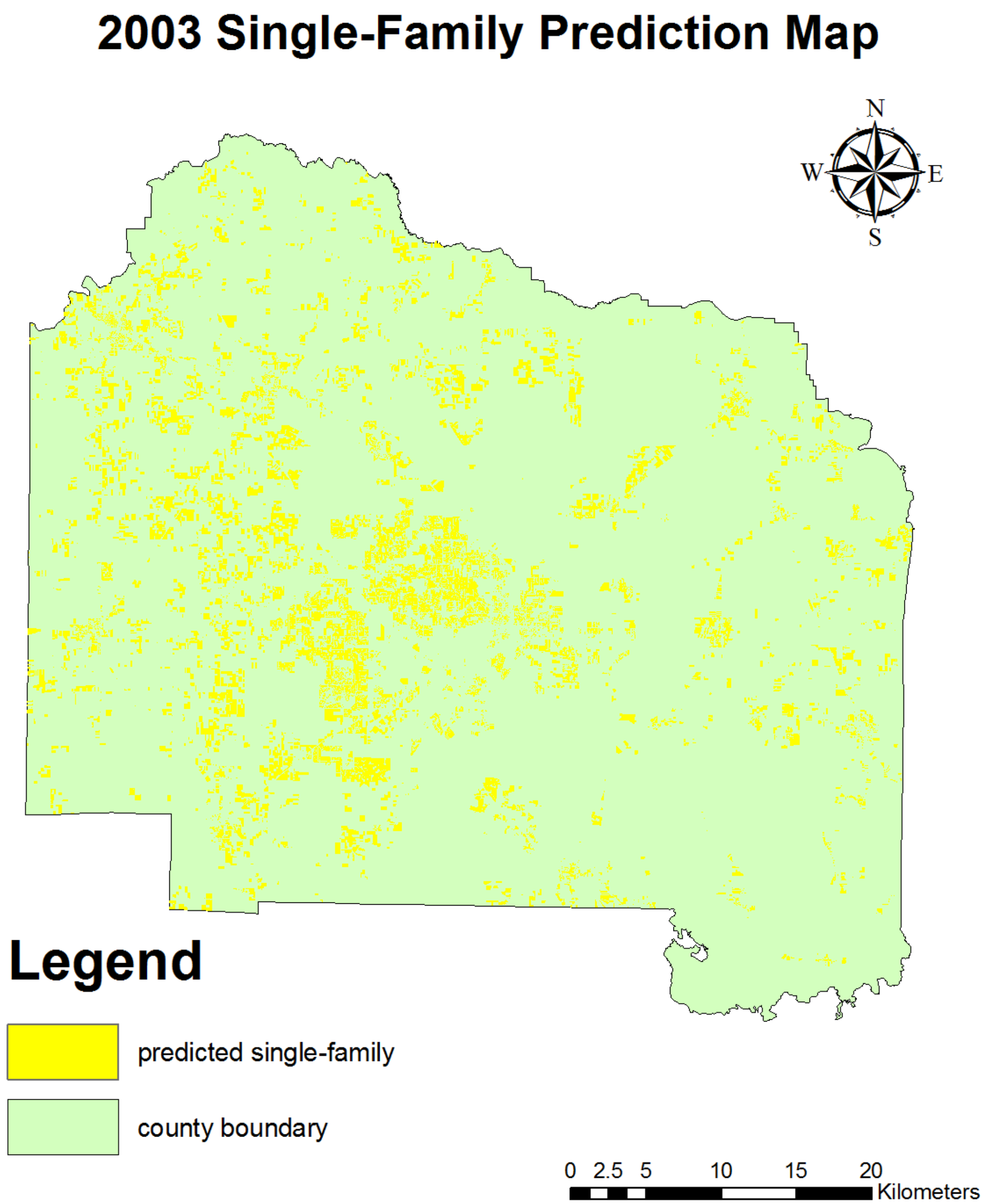

Figure 3-2. 2003 single-family prediction map

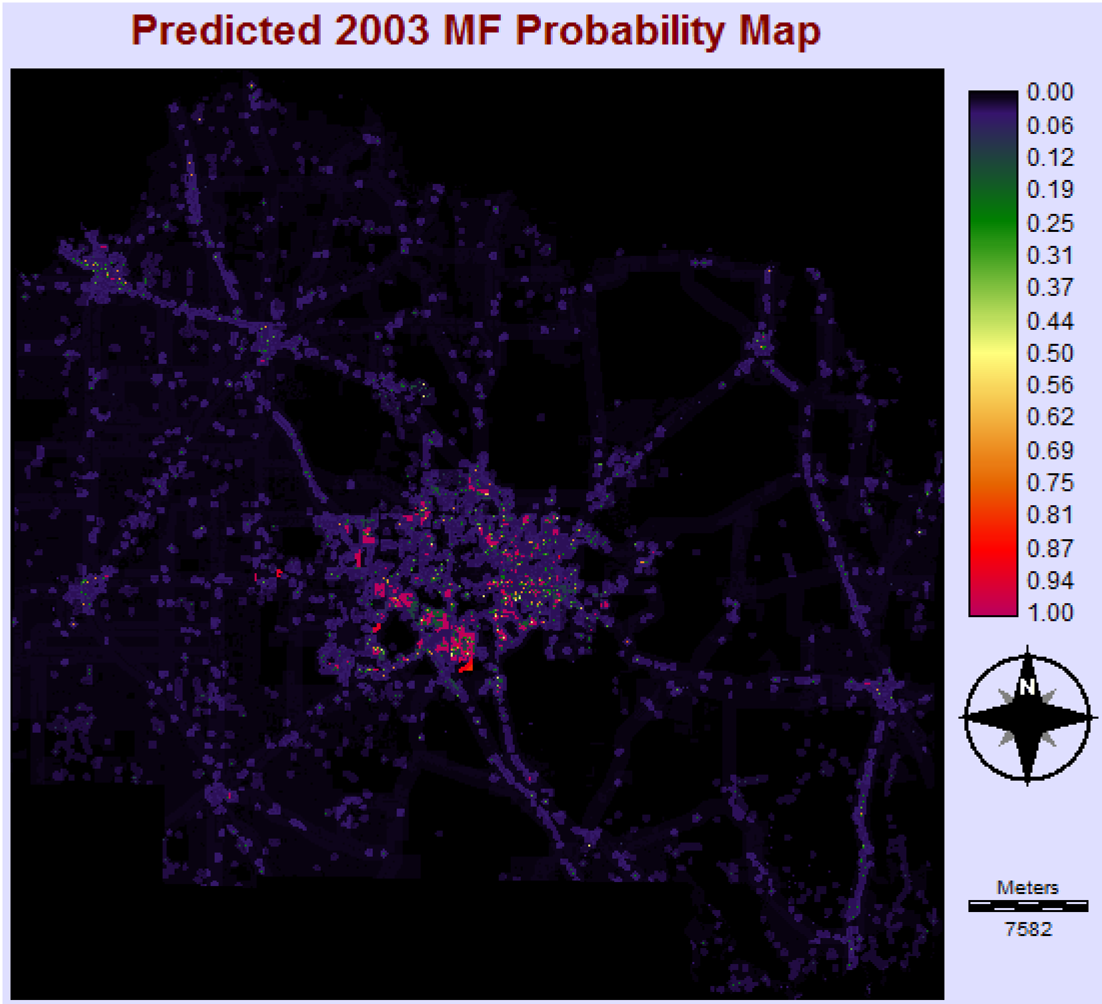

Figure 3-3. Predicted 2003 multi-family probability map

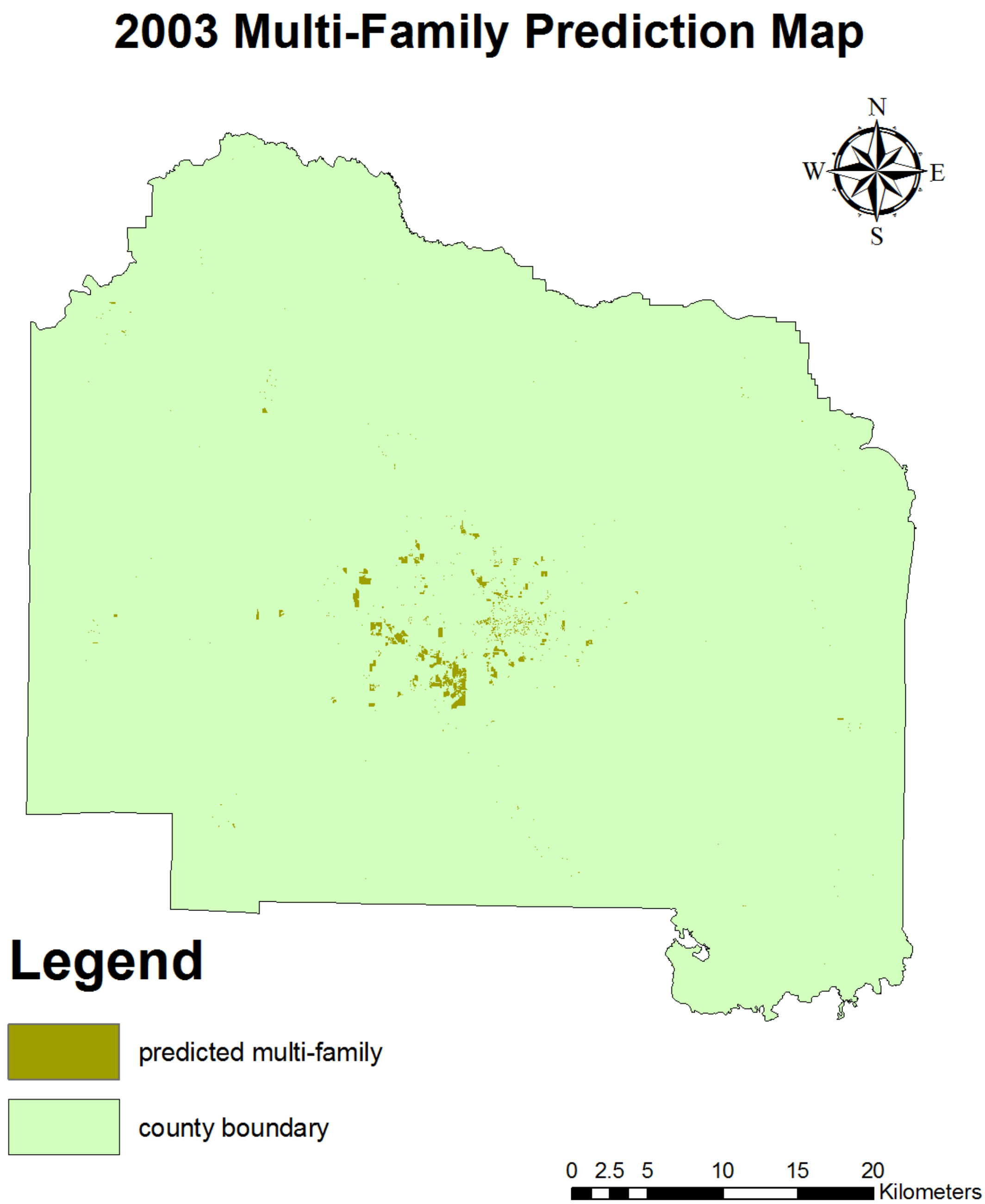

Figure 3-4. 2003 multi-family prediction map

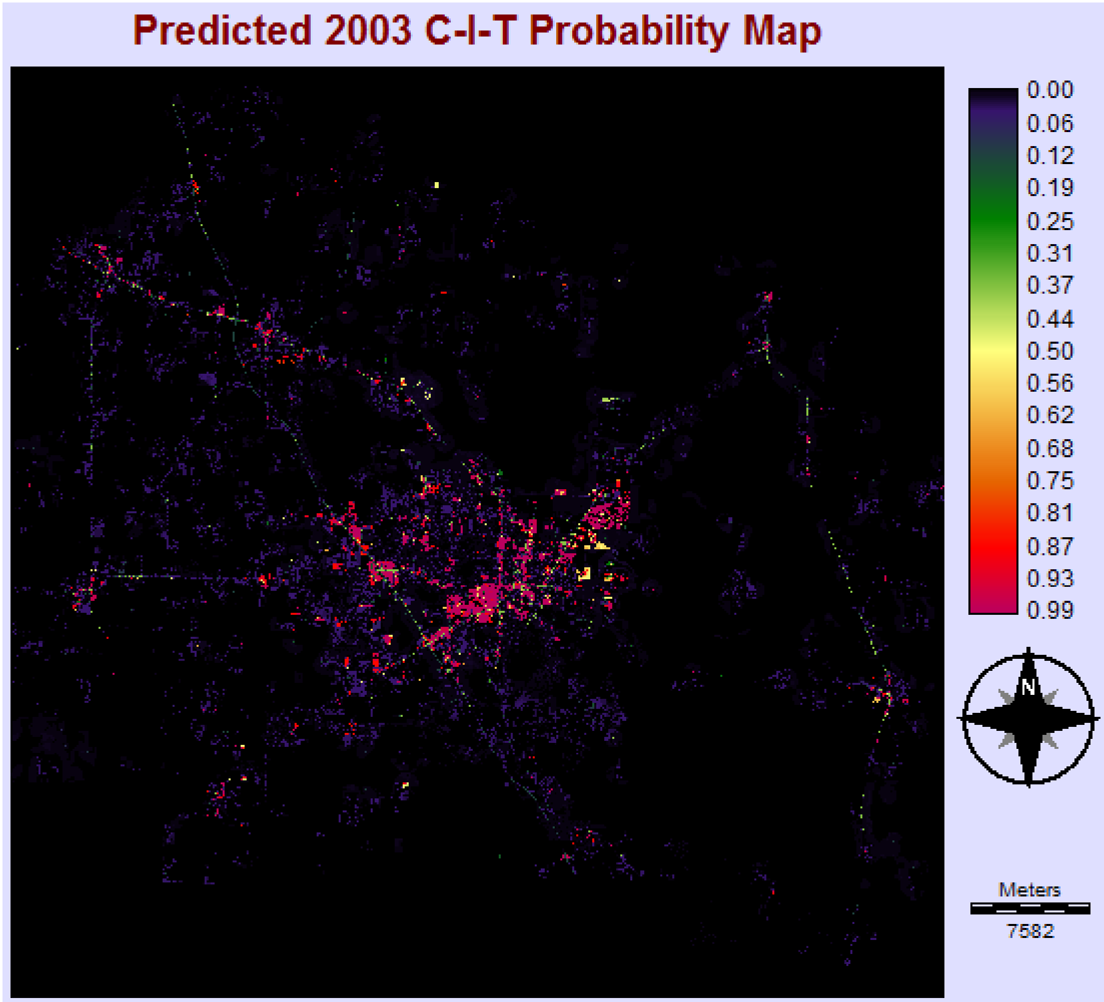

Figure 3-5. Predicted 2003 commercial-institutional-transportation probability map

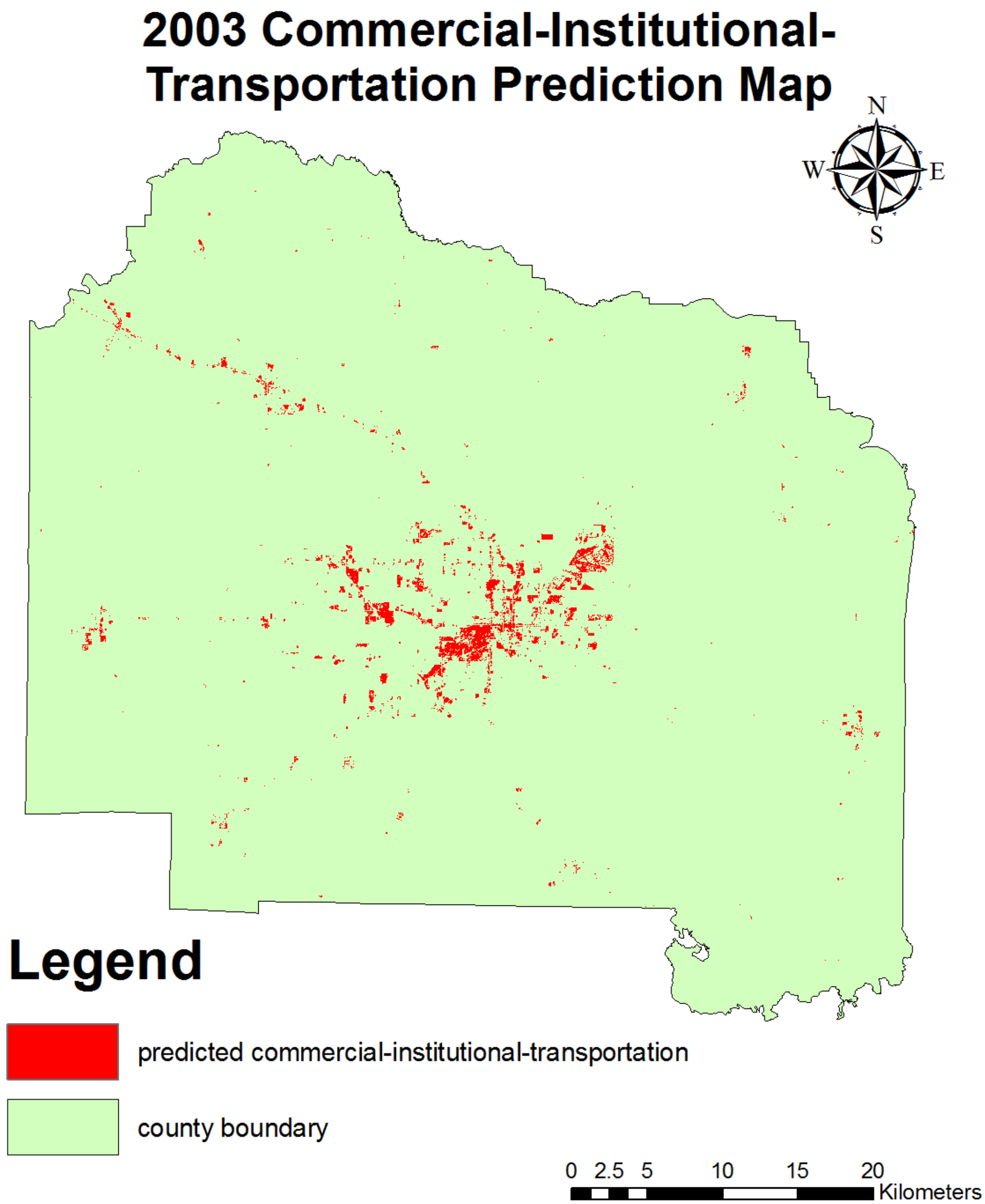

Figure 3-6. 2003 commercial-institutional-transportation prediction map

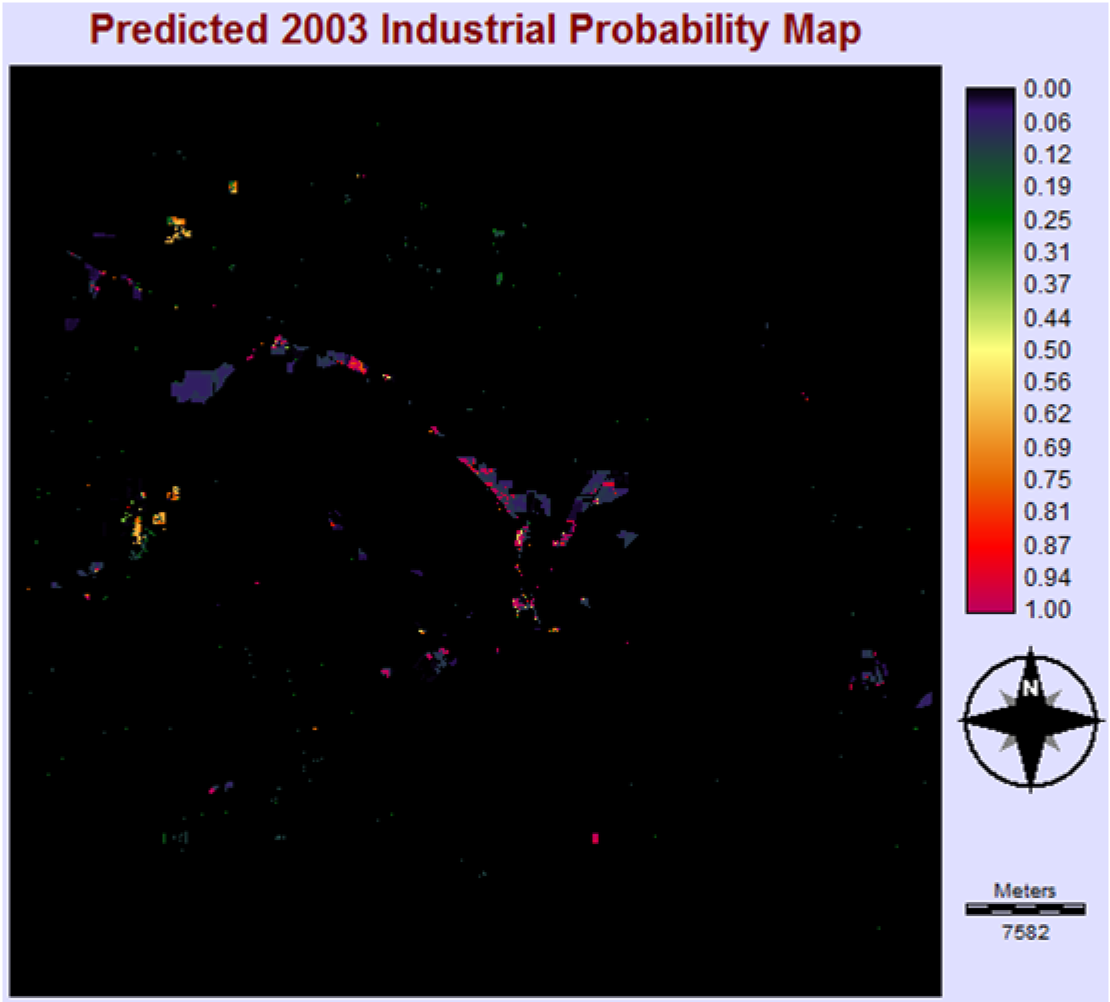

Figure 3-7. Predicted 2003 industrial-warehouses probability map

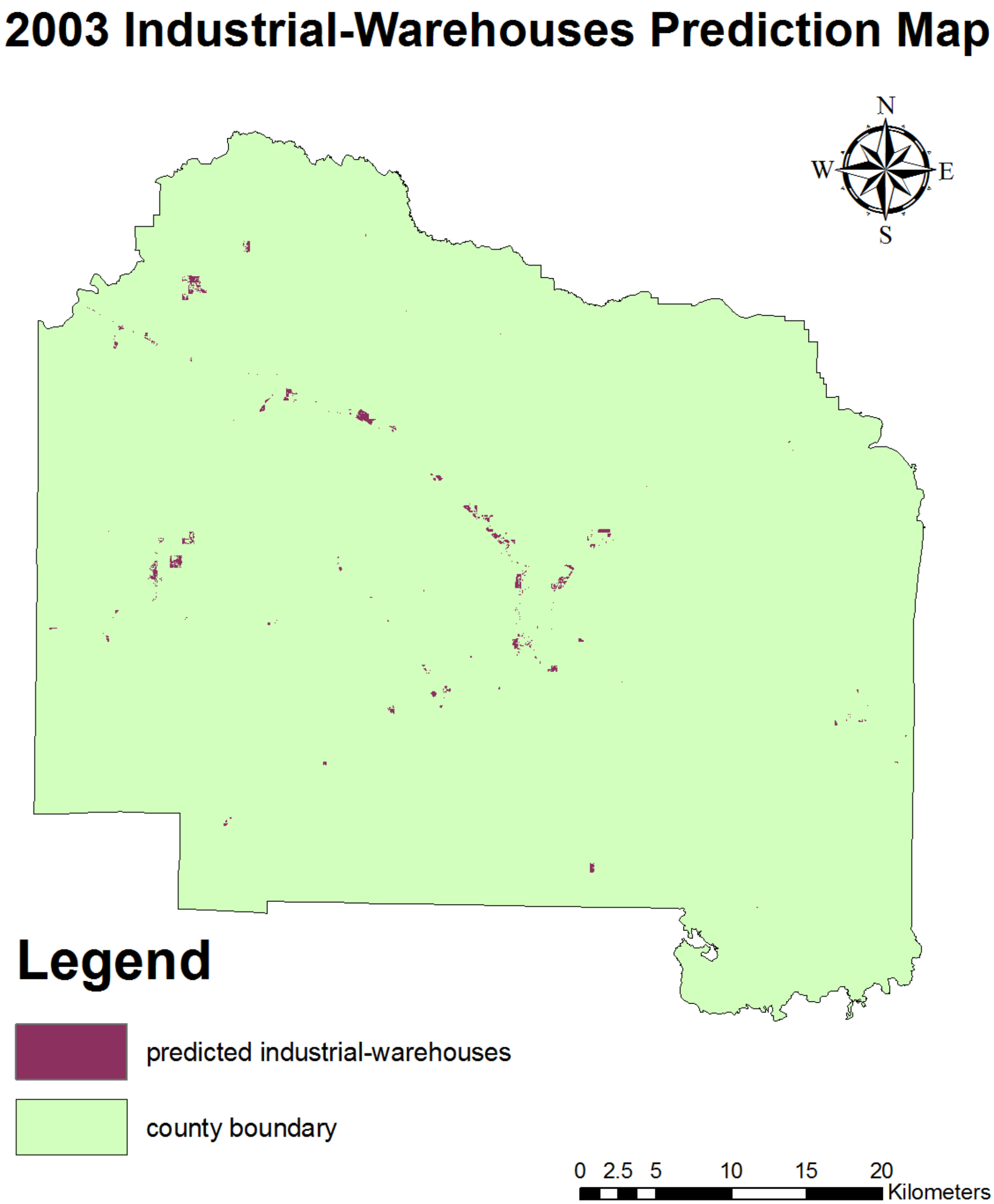

Figure 3-8. 2003 industrial-warehouses prediction map

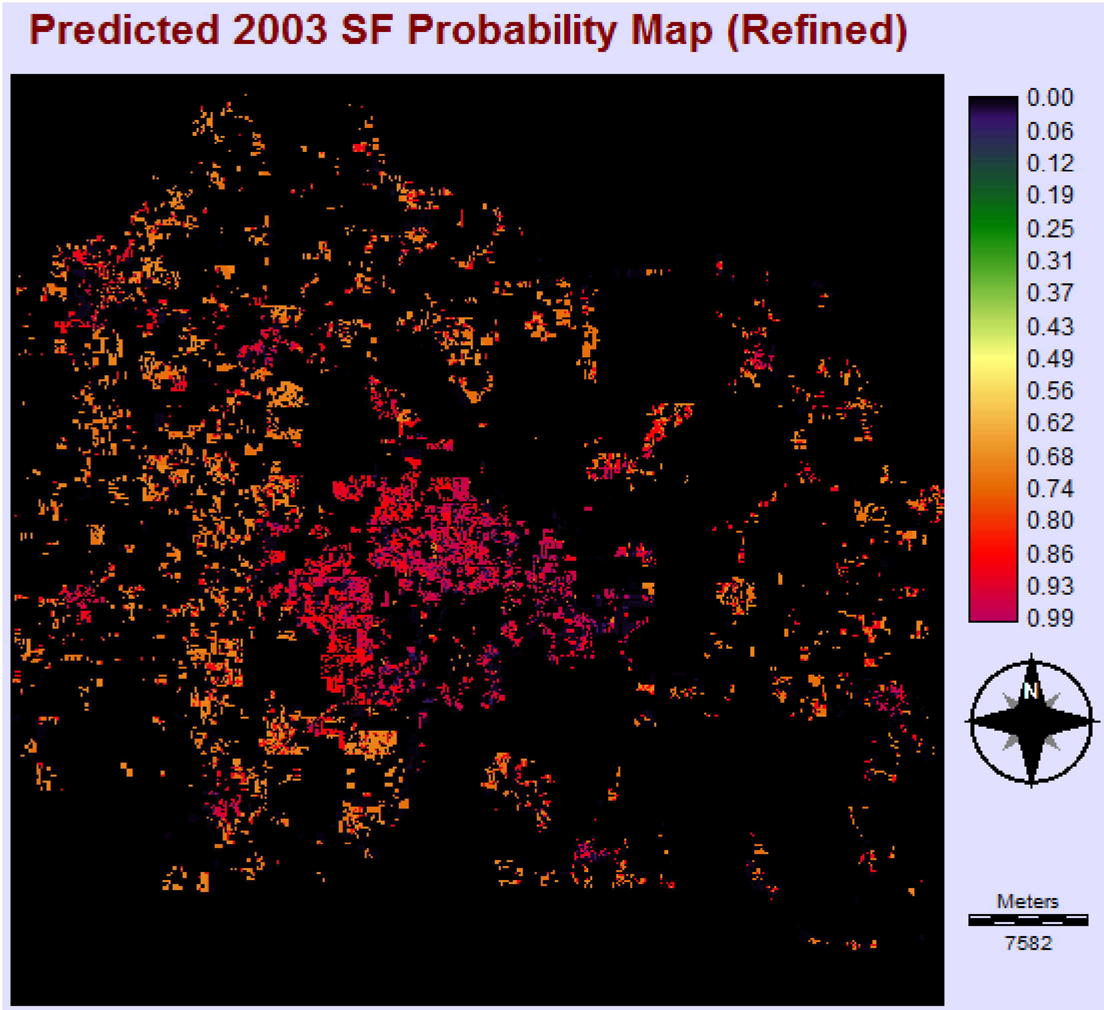

Figure 3-9. Predicted 2003 single-family probability map (refined)

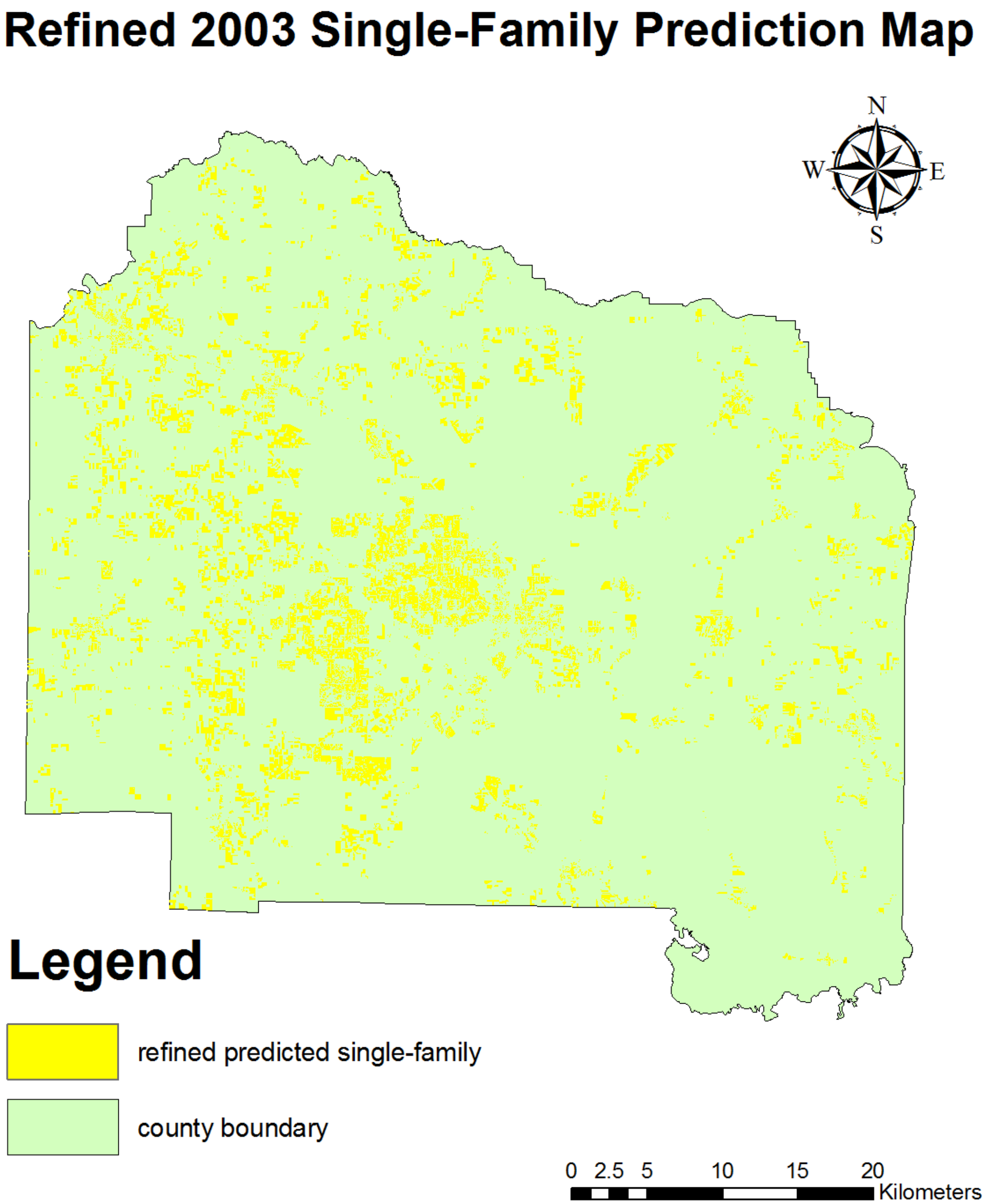

Figure 3-10. Refined 2003 single-family prediction map

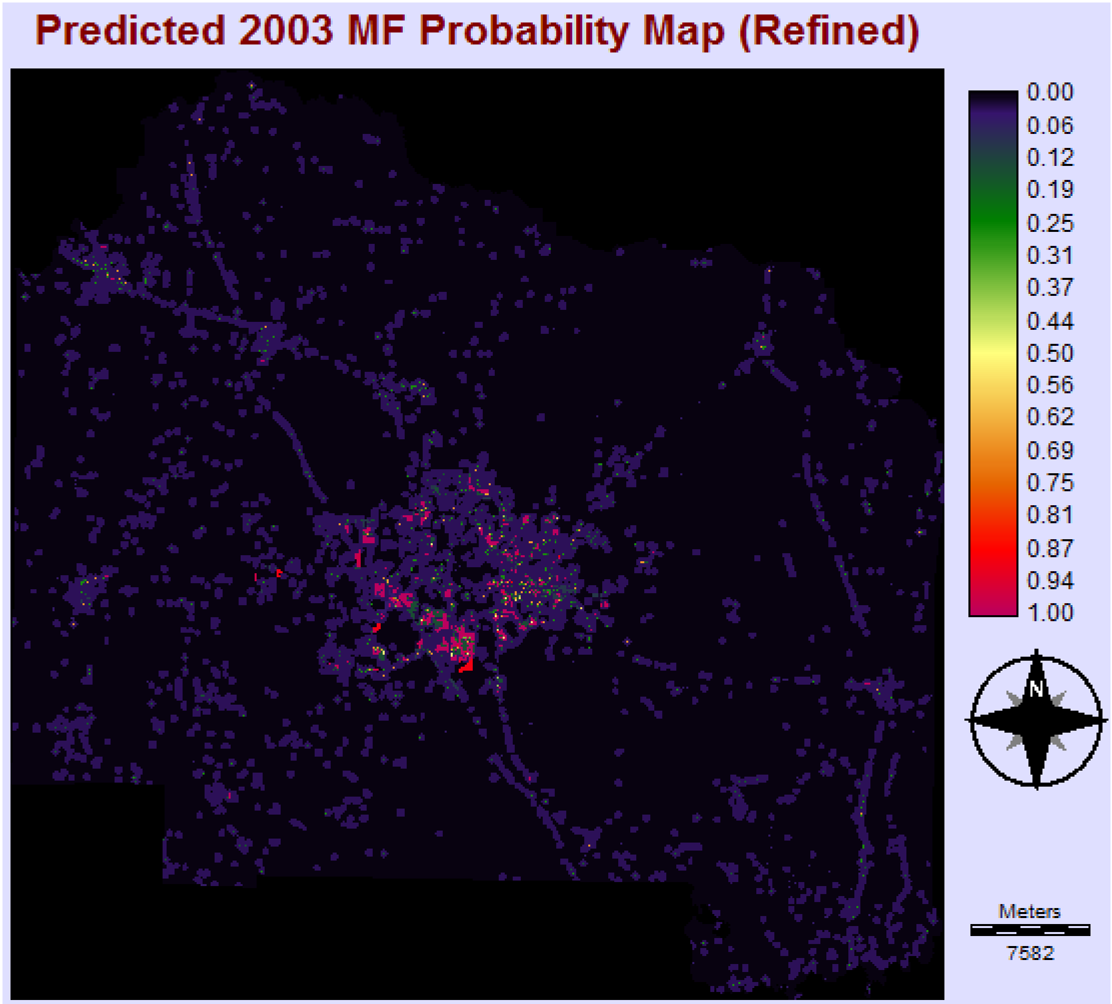

Figure 3-11. Predicted 2003 multi-family probability map (refined)

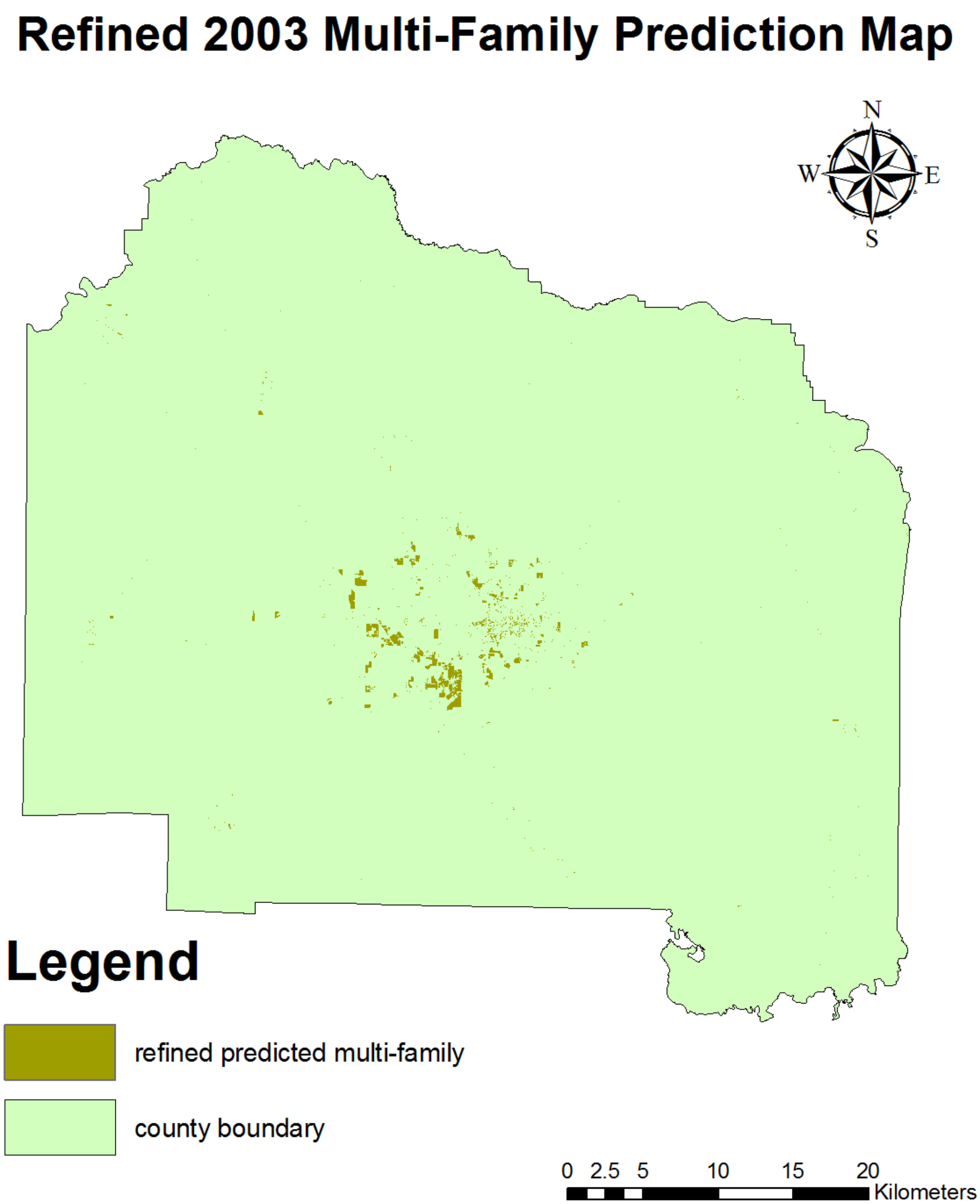

Figure 3-12. Refined predicted 2003 multi-family prediction map

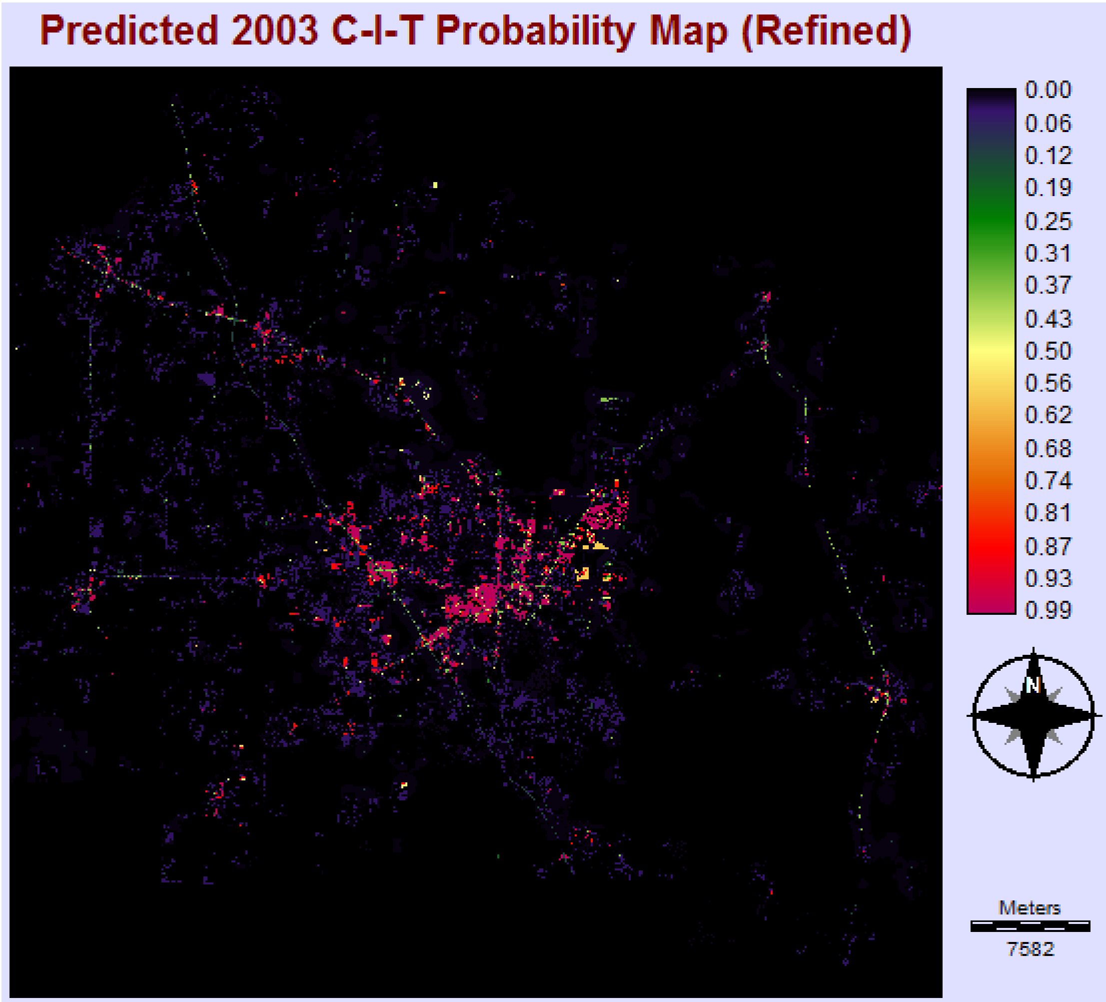

Figure 3-13. Predicted 2003 commercial-institutional-transportation probability map (refined)

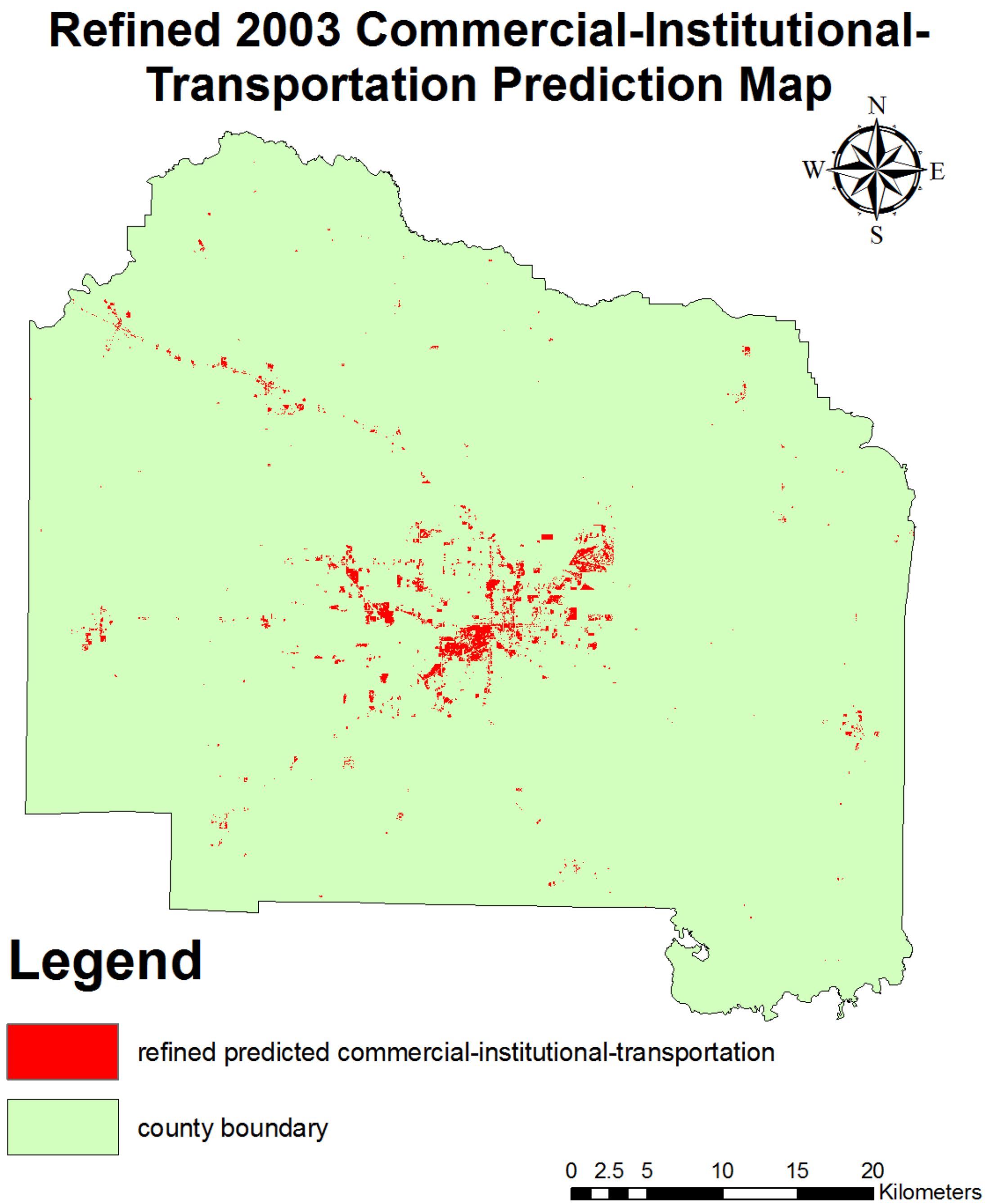

Figure 3-14. Refined 2003 commercial-institutional-transportation prediction map

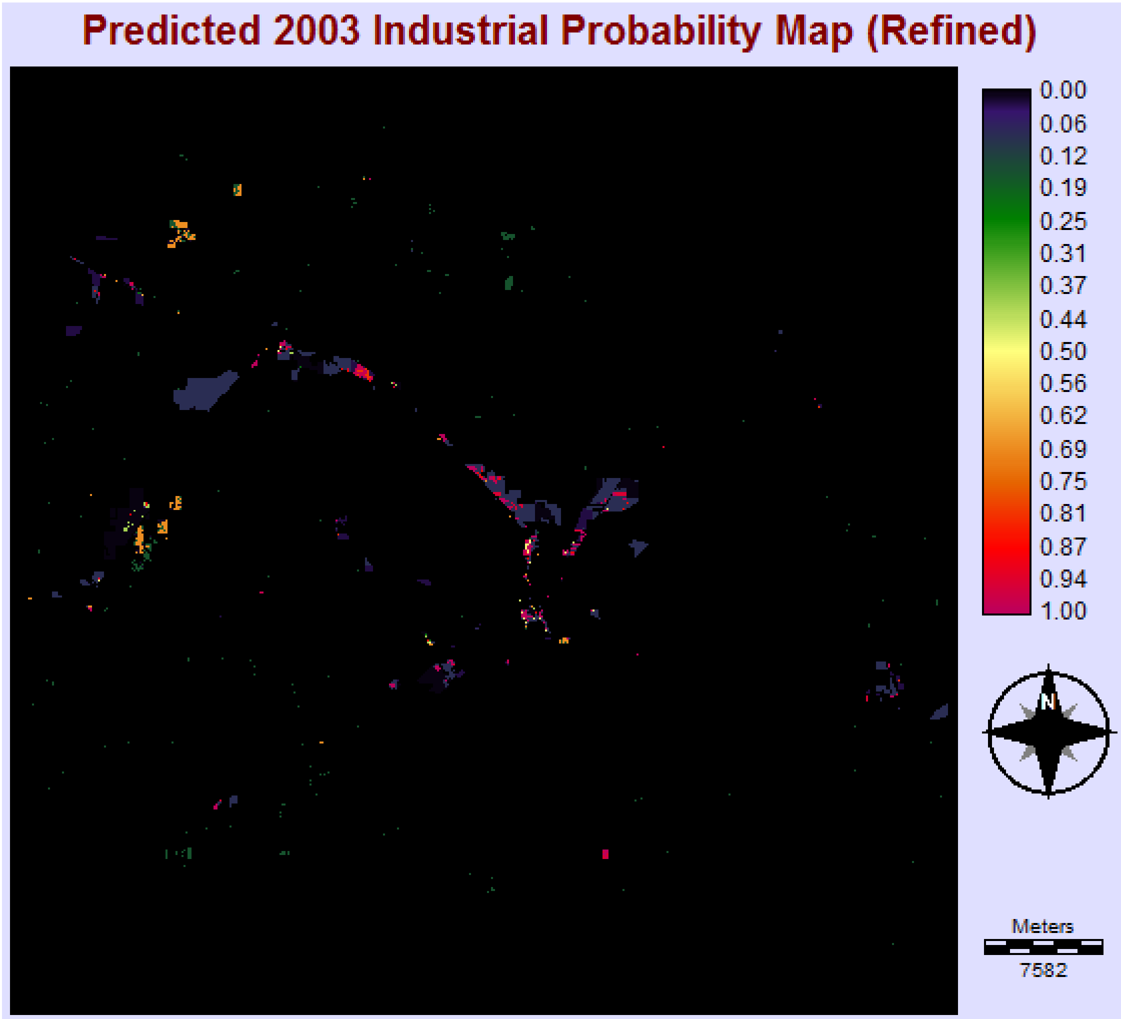

Figure 3-15. Predicted 2003 industrial-warehouses probability map (refined)

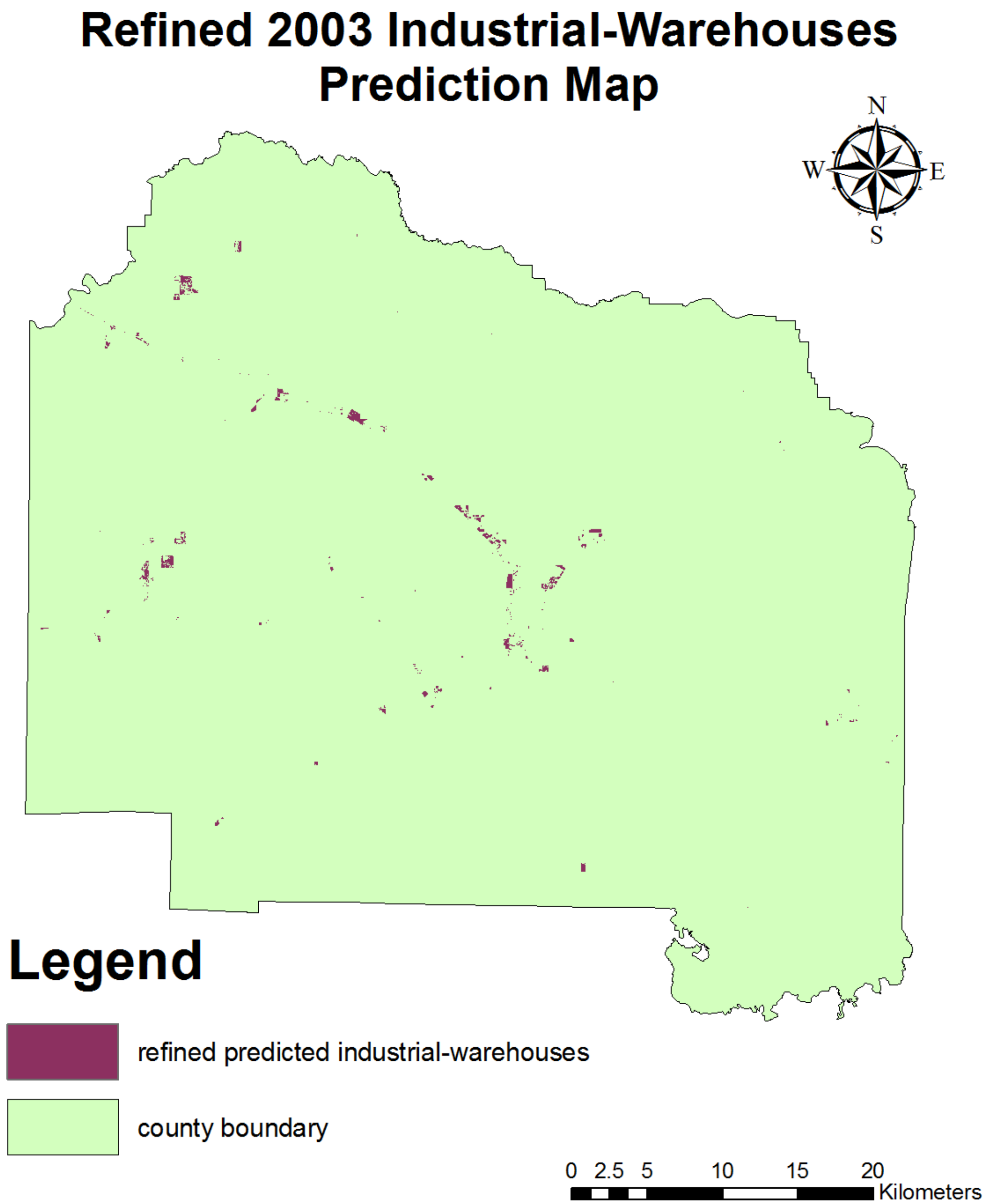

Figure 3-16. Refined 2003 industrial-warehouses prediction map

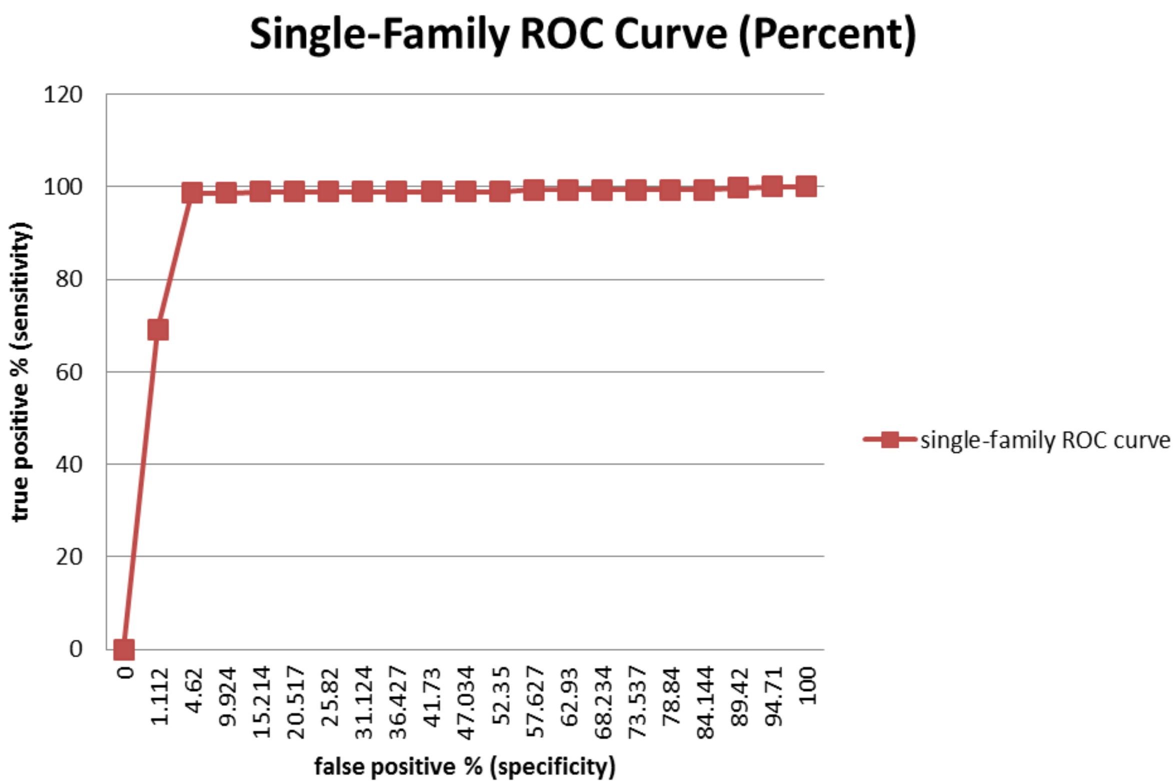

Figure 3-17. ROC curve for the single-family use

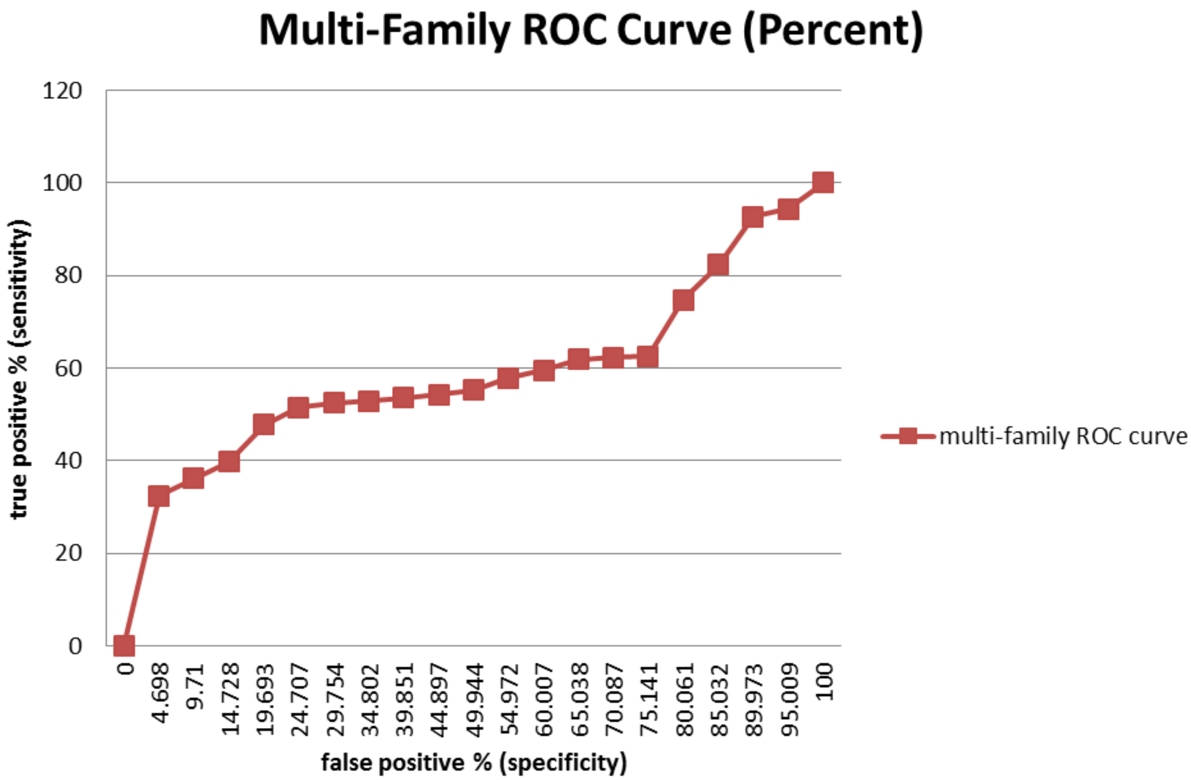

Figure 3-18. ROC curve for the multi-family use

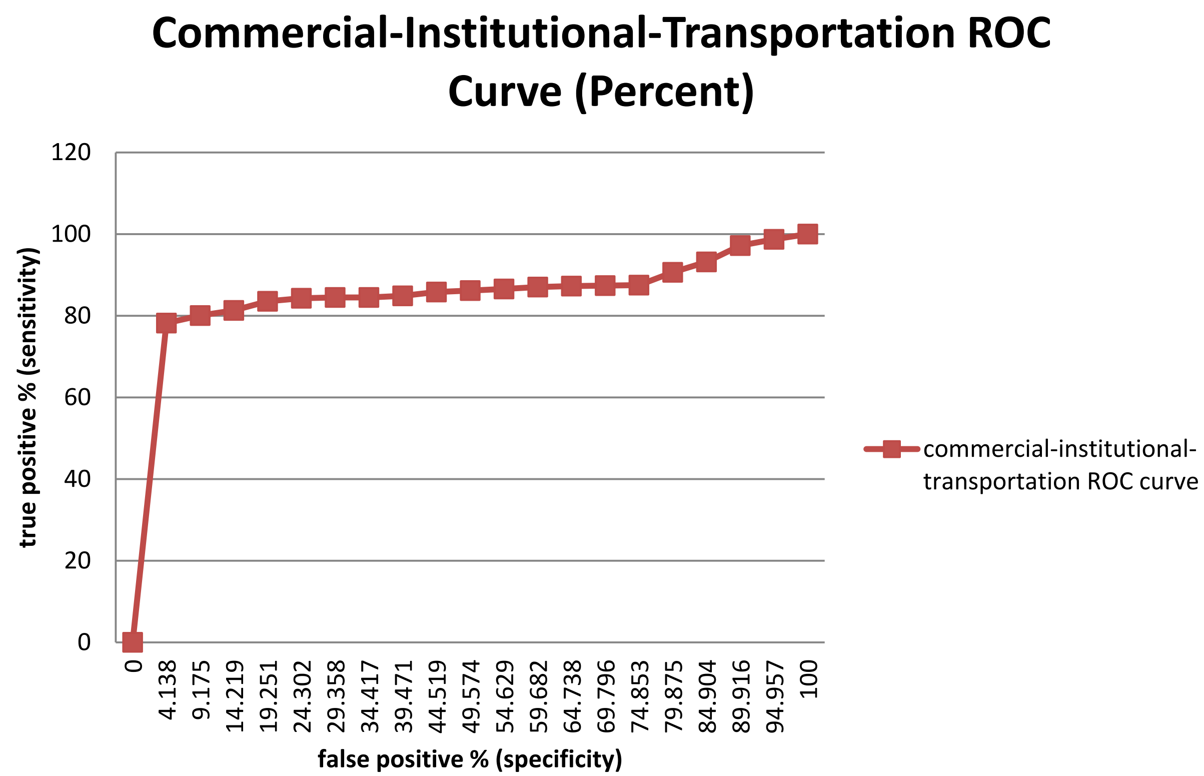

Figure 3-19. ROC curve for the commercial-institutional-transportation use

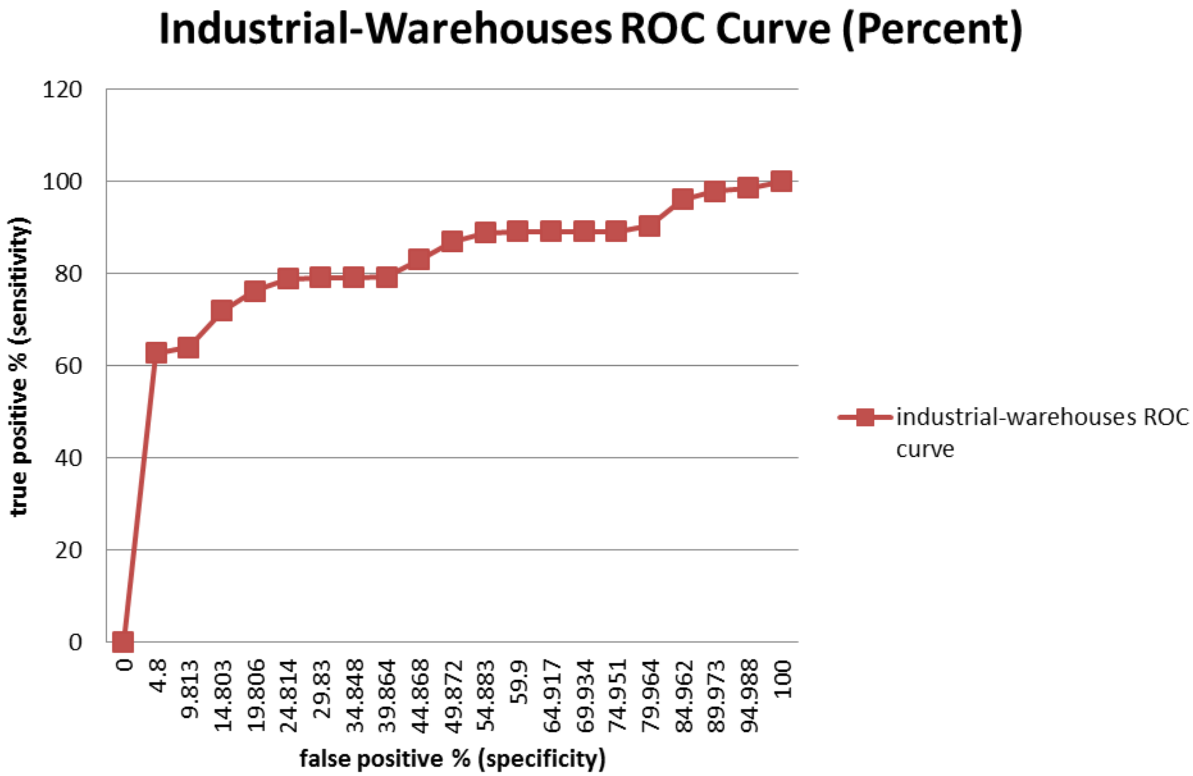

Figure 3-20. ROC curve for the industrial-warehouses use

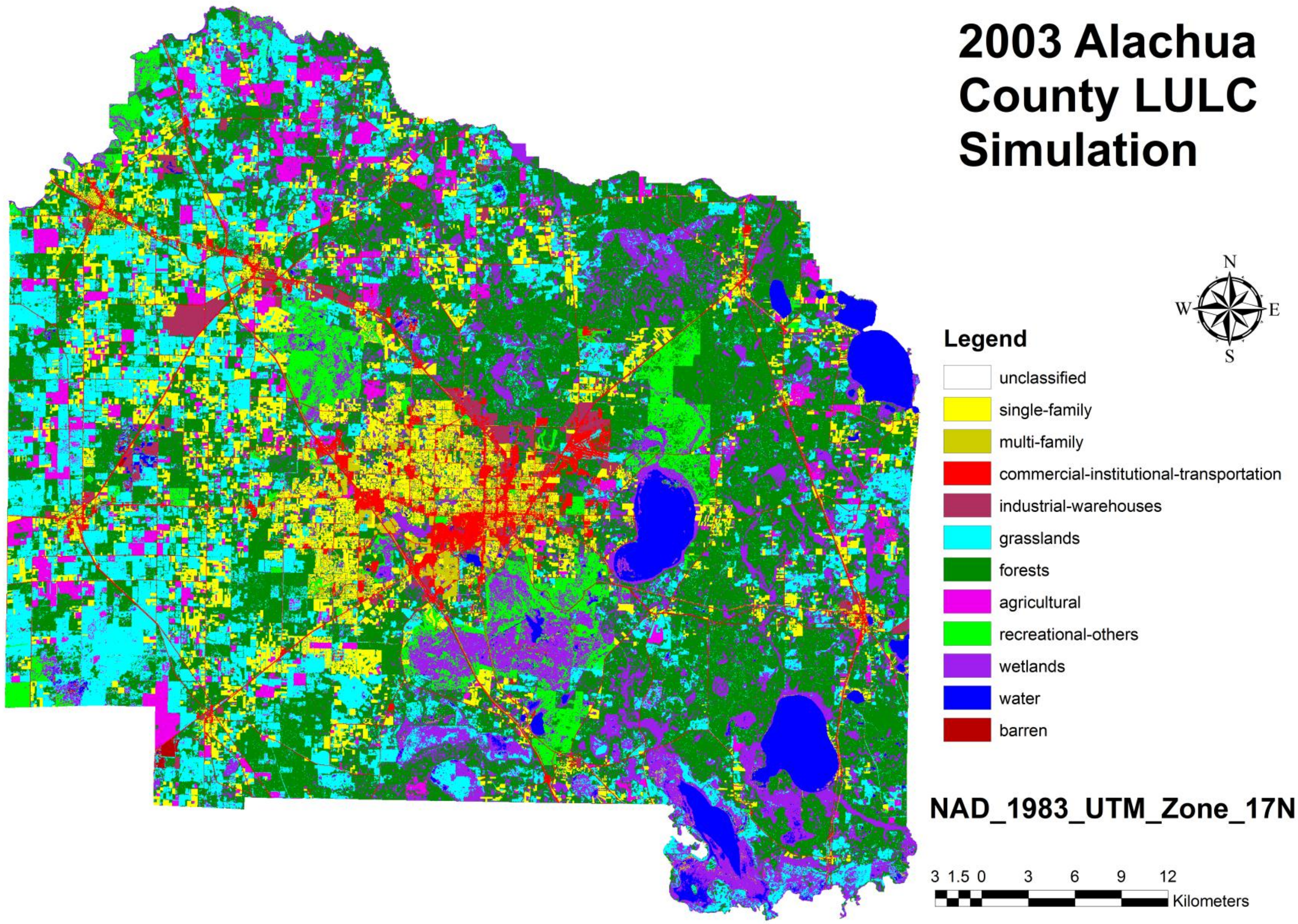

Figure 3-21. 2003 Alachua County LULC simulation

CHAPTER 4
URBAN LAND USE ALLOCATIONS

Chapter Overview

The urban land use allocation is the next step following the MLR model building in Chapter 3. The allocation is essentially a conversion process that converts the continuous probability maps into categorical discrete maps based on the different urban uses. The urban allocation process provides tangible maps derived from the abstract probability maps and fulfills the urban use allocations through a series of processes such as urban development scenario identification, urban development and land forecasts, conflict analyses, and final allocations. The fundamental goal of the allocation is to produce urban allocation maps based on different development scenarios to simulate different urban development arrangements. These will be introduced in detail in this chapter.

Urban allocation narration sequence. The narrations of overall urban use allocation are organized in three parts: (1) literature review and summaries of research that has been conducted so far; (2) narrations of allocation methods; and (3) narrations of urban use allocation results. The literature review will introduce the research currently underway to learn about methods and results. The urban allocation methods will introduce the techniques that will be applied in this study. The results are the outcomes of the applied allocation methodology.

Urban allocation scale. The study area is Alachua County, which comprises agricultural land plus a number of small cities and towns, with the largest city being the City of Gainesville. As predominantly a rural county, Alachua County has consumed significant amounts of natural land for urban development and suffered serious urban sprawl problems over the years. This rural county is a typical example that can be dissected and examined for its loss of precious natural land over the years, which can give the reasons of how natural land is consumed and converted into urban uses and why urban sprawl was triggered in such a way. The examination can inductively lead to understandings of how urban sprawl happens in Alachua County as exemplified and how the rule can be extracted and used, deductively, for other Florida counties and cities for its heuristic character. Thus, the research that finds rules from this rural county for urban growth is expected to encourage research for other Florida cities and counties. As a result, this study starts not from a big county with a large urban population, such as the City of Jacksonville or the City of Orlando, but from a relatively smaller county in population—still rural in character.

The above narration is the case that can horizontally induce urban growth rules across the political boundaries of different counties. It can also be induced vertically across different timeframes for urban growth research in the county, in which strategies dealing with future urban growth, i.e., in 2020 and 2030, can be put forth based on the analysis of the current situation in 2010 as well as the past situations in 1982, 1994, and 2003, respectively. As a result, the quantitative counts of land and spatial distribution of the current and past urban development are assessed. The future urban growth in the quantitative amount of land and spatial spreading of development will also be assessed as this urban allocation process provides clues to the quantity and quality of urban development for the next twenty years. The allocation process will also offer statistics of how many acres of urban development will be and where they will be. In fact, this urban allocation establishes foundations for policy formation of urban growth in the future.

Urban allocation method. This study incorporates five scenarios in order to simulate urban development based on five conditions. These five scenarios include the BAU scenario, the infill scenario, the increased density scenario, the redevelopment scenario, and the conservation scenario. They will be illustrated in detail in the Methodology part of the chapter. These five scenarios almost comprise all development possibilities, prospects, and visions for the future urban growth of the county. With these five scenarios, the future urban growth in Alachua County will basically follow along these development paths and fulfill these development manners.

Aside from the above five scenarios, the urban allocation method also includes three forecasting methods, which correspond to the above five scenarios, because the different development scenarios entail different forecasting methods. These three forecasting methods include the BAU (baseline) forecasting method, the increased density forecasting method, and the redevelopment forecasting method. They will be explained in detail in the Methodology part of this chapter.

Urban uses that will be allocated. The urban allocation in this study deals with four urban uses, which are single-family, multi-family, commercial-institutional-transportation, and industrial-warehouses. These four urban uses will be allocated based on the five scenarios as mentioned above. In addition, mixed-use will be simulated in the redevelopment scenario as well as in the conservation scenario. The natural LULC classes such as grasslands, forests, agricultural, recreational-others, wetlands, water, and barren will not be allocated in this study, however.

Urban allocation data. This study entails the combination of parcel data, remote sensing data, and probability data that are derived from the remote sensing data to analyze urban growth. On the one hand, the continuous probability data, derived from the MLR model, will be applied to create the collapsed maps based on each urban use. These are the maps that will be converted from the continuous data to the discrete data so that they can be later input into the conflict analysis to analyze. These probability data from the MLR model essentially come from the classification of urban and natural LULC classes, which originated from the remote sensing Landsat TM and ETM+ data. On the other hand, parcel data are utilized based on parcel counts for different urban uses from the Alachua County Property Appraiser's Office. These parcel counts of different urban uses will be extracted for their acreages and will be used in the infill scenario, increased density scenario, and the redevelopment scenario because these three scenarios call for infill acreages for development, and infill acreages come directly from the parcel counts from the parcel data. Furthermore, these parcel data are extracted to create the LULC Current Plan map, depicting the urban uses and natural LULC classes in 2010, which is the base map for the future allocation maps for 2020 and 2030, respectively.

Allocation Literature Review

The literature review will find equivalent research that has been conducted so far for urban use allocations. Because some of the descriptions in this literature review part will be related to the urban growth model building, i.e., the MLR model, which is the major contents in Chapter 3, this literature review will carry over some of the descriptions of the urban growth model building to this chapter. However, because this chapter mainly discusses the urban use allocations, the major theme of this chapter will still cover the urban use allocations as the most important part in the chapter. The literature that is reviewed has much to do with the application of the MLR model. Other models, such as the CA model, are not included in this review.

Verburg et al. (2007) used a Conversion of Land Use and its Effects at Small Regional Extent (CLUE-S) model that is derived from the MLR model to simulate agricultural land abandonment and conversions in Europe based on 1-km resolution satellite images. Their research was conducted based on two timeframes, 2000 and 2030, respectively, in which the agricultural land was analyzed as opposed to other uses such as built-up areas and grasslands. They found that agricultural land conversions varied according to geographic locations, in which agricultural land was converted in the fastest way in dense urban areas and needed vacant land the most to accommodate urban development and also in marginal areas such as mountainous regions. They also conducted a landscape analysis based on eight land classes identified as built-up, arable land, pastures, forests/nature, inland wetlands, irrigated agriculture, abandoned land, and others. Since their research was primarily on the agricultural land, their land use classes for urban development was solely based on urban built-up, along with several other natural land categories such as pastures, forests, wetlands, agricultural land, and so on. The research of Verburg et al. (2007) is inspirational as they used the overall non-spatial land consumptions to forecast the land demands and then translated them into spatial distributions for urban built-up areas and various other natural land. This study uses population forecast to translate it into land consumptions for urban uses as well.

Landis (1994) developed a CUF model to simulate urban growth in the northern San Francisco Bay Area of California, which consists of fourteen counties. The CUF model had four modules: population growth, spatial database, spatial allocation, and annexation. The CUF model first forecasted population in each city and county in the region through a “bottom-up” approach (p.403), in which population in each individual city and county was estimated using the OLS regression algorithm, rather than projecting the regional population first and then subsequently splitting them into smaller units based on each individual city and county. Second, the CUF model merged various layers related to urban growth into a single layer in a polygon format. As a result, vector based analysis was conducted at this point, for which data sources were from the Census Bureau TIGER databases such as TIGER roads, TIGER census tracts, TIGER city boundaries, TIGER hydrology, and TIGER railroads and airports; sphere of influence; slope polygons; highway buffers; earthquake faults; prime agricultural land; marsh and wetlands; and sewer and water utility service costs. Third, the CUF model allocated the forecasted population based on the undeveloped land available for development derived from the above vector spatial layers. In this module, private developers' land development behaviors, such as seeking the maximum profits, were simulated. As a result, undeveloped land was weighted as scores and sorted according to their profit potentials calculated from the model. In this step, non-developable land due to environmental constraints, ownership, and public policy issues was excluded from allocation. Finally, the forecasted population was allocated to the areas within the sphere of influence based on the scores of each undeveloped land received from the weighting factors as mentioned above. Municipal annexation and urban incorporation factors were also considered as the last module in the CUF model. As a result, Landis (1994) used a simple regression model to simulate the expansions of urban boundaries within the area.

In addition to the northern Bay Area of California, Landis (1995) also applied the CUF model to the San Francisco Bay Area based on three scenarios: business as usual, maximum environmental protection, and compact city. The business as usual scenario followed the current land development policies at the local level, in which the current local land use practices would be continued in the future. The maximum environmental protection scenario proscribed development from occurring in steeper slope areas, wetlands, and agricultural land. Compact city allowed development to only occur in the area having at least 18 persons per acre, setting aside at least 20 percent of population to infill development, and within 1,000 meters of existing urban boundaries (Landis, 1995).

The CUF model is close to the model of this study in terms of the model structure since the CUF model incorporates population forecast and population allocation,[1] the two important steps for urban growth modeling. The CUF model is based on population forecasts for each individual

[1] Can be translated into land use forecast and land use allocation.

city and county as well as the principles that either the forecasted population is completely allocated within the undeveloped areas or the undeveloped land is totally consumed up, leaving the additional population to be allocated to adjacent communities. In addition, the CUF model allocates population based on a score for potential profit each land unit receives, which is quite similar to the suitability analysis (Carr and Zwick, 2007), in which each land unit obtains a weighted score to represent the site's suitability for a typical development based on a variety of spatial, social, and economic variables. Moreover, CUF's simulation of land use development based on three scenarios is inspirational as it incorporated three different land use policies to urban development, in which this study will adopt the concept of this approach to simulate the urban growth for Alachua County in the future as well and also adopt the concept of 1,000 meters of existing urban boundaries as urban buffer areas. In addition, the infill development is also inspirational to this study. However, the CUF model addresses urban growth solely based on the urban category, or more precisely, the residential use (Landis and Zhang, 1997). It did not address additional urban land categories such as residential, commercial, and industrial, together with other non-urban natural land, which will also influence urban land use development. As a result, it is a basic urban growth model without considering various uses involved for urban growth. At this point, conflict analysis (Carr and Zwick, 2007) such as the competition among various urban uses are absent from the model.

To remedy these shortcomings, Landis and Zhang (1997) proposed a CUF II model to simulate urban growth. The major overhaul of the CUF II model is its inclusion of the MLR algorithm, in which probabilities are used to measure the probability of a site to be developed and/or redeveloped. Compared to the old CUF model, the CUF II model added four additional elements: expansion of the original single urban use to residential use, commercial use, industrial use, public use, and transportation use; inclusion of urban development together with urban redevelopment; bid of different urban uses to be located in a typical site; and inclusion of historical data to calibrate the model. Compared to the old CUF model, the CUF II model provided the results of the spatial locations of the land uses for their correlations with different variables the model specified, which resulted in the current urban growth in different counties in the San Francisco Bay Area, based on the development and redevelopment categories. For example, for vacant land development, cities with larger population size would be more likely developed in terms of residential use than cities with smaller population size in the area. For the redevelopment, for example, however, sites that were closer to transportation corridors would less likely be redeveloped into residential use. In this case, the CUF II model did not allocate land to meet the land demands given various different urban use categories the model identified, in which the CUF I model had. In addition, the CUF I and CUF II models adopt a regional approach for urban growth, in which, e.g. cities with larger population will be more likely to be developed for residential use. This is quite different from the scope of this study in which only a local county is considered for urban use allocations. Nevertheless, the CUF II is inspirational for urban growth modeling because it, for the first time, proposes to use the MLR model to simulate urban development and/or redevelopment. It also carries some heuristic elements such as inclusions of various urban uses, bid of different competitive uses for a site, and inclusion of historic data to calibrate the model, in which these elements can be used in this study (some parts are illustrated in Chapter 3).

Hu and Lo (2007) utilized the Landsat images to simulate urban growth in the Atlanta region in Georgia, which included 13 counties, based on six LULC categories: high-density urban, low-density urban, bare land, crop or grassland, forest, and water. Because their study area covered 13 counties, they modeled urban growth based on urban use only, along with other natural land. Their remote sensing data covered the period of 1987 and 1997, respectively, for the study area. During the analyzing process, they combined high-density urban and low-density urban into one variable, i.e., urban, which was used as a dependent variable for model building. They found twenty independent variables such as population density, income per capita, poverty rate, percentage of white people, employment rate, slope, distance to nearest urban cluster, distance to CBD, distance to active economic center, distance to the nearest major road, number of urban cells within a 7 by 7 window, existing/planned conservation, existing high density urban, existing low density urban, existing bare land, existing cropland/grassland, existing forest, and easting and northing coordinates. Based on these variables, Hu and Lo (2007) produced a number of maps to predict urban growth with certain percentages, e.g., 20 percent, 25 percent, or 30 percent of urban areas, within the study area. These percentages can be translated into the specific timeframes that meet the thresholds specified above. For example, the urban areas in Atlanta were 16.4 percent in 1997, in which 16.4 percent of urban development referred to 1997; in 2005, the urban areas in Atlanta reached 25 percent of the total area, in which 25 percent of the urban areas referred to 2005, and so on for the rest of urban area percentages mentioned above. Although Hu and Lo (2007) successfully modeled urban growth based on different periods translated from different percentages of urban areas, their research is not very convenient for readers to interpret the modeling results, given the percentage of urban development. Research that addresses the urban growth based on specific years will be more meaningful than the percentage development. This is what this study is going to achieve.

Carr and Zwick (2007) developed a Land-Use Conflict Identification Strategy (LUCIS) model to allocate urban, agricultural, and conservation uses, the three uses upon which their research is primarily focused. The “conflict” in the LUCIS model is essentially an analysis that deals with these three uses based on a conflict nature that one typical use, e.g., urban, is in conflict with the other two uses, i.e., agricultural and conservation, when they are allocated to a specific site. Although Carr and Zwick (2007) does not name “conflict” as “competitive,” their model essentially calls for the thesis that these three uses are competing with each other to bid for a typical site during the allocation process.

The LUCIS model is fundamentally a suitability analysis model. Carr and Zwick (2007) constructed the LUCIS model based on five steps: (1) goals and objectives identification; (2) preparation of data inventory; (3) suitability analysis; (4) preference identification; and (5) conflict analysis. Goals and objectives are the intended setting for a specific scheme that acts as an input into the suitability analysis. Each of the three uses has its equivalent goals and objectives which are hierarchically dispersed across its compositional structures so that one use, e.g., urban, has an overall goal and is composed of several sub-categorical goals and objectives as input into the suitability analysis. It is the same as for the agricultural and conservation uses. This identification of goals and objectives establishes foundations for the suitability analysis, as mentioned above, as well as the conflict analysis because the suitability analysis and the conflict analysis utilize the goals and objectives to create suitability maps for each land use category.

The data inventory for the LUCIS model is basically parcel data as well as pertinent GIS shapefiles. The three land uses have their existing datasets retrieved from the county parcel data for urban, agricultural, and conservation, respectively, in which detailed subgroups of the three uses can be extracted when needed. For example, the urban use includes retail, commercial, industrial, warehouses, and so on, which are retrieved from the county parcel data when they are required. As a result, some non-existing GIS shapefiles can be created by employing the existing

GIS vector datasets. Thus, raster datasets can be created for various analyses through manipulating the parcel data and pertinent existing GIS shapefiles.

The suitability analysis is the crux of the LUCIS model. Carr and Zwick's (2007) suitability analysis essentially tells how well a piece of land fits a land use arrangement—whether it will be best developed as urban, or preserved as agricultural or conservation. The suitability analysis primarily translates the goals and objectives in the above Step 1 into tangible maps so that uses that have specific, equivalent goals and objectives are depicted on maps to illustrate where the suitable land is in terms of a specific goal and objective. To achieve this purpose, the goals and objectives are hierarchically established and categorized as overall goals and objectives, general goals and objectives, and sub-goals and sub-objectives. A typical map is created for each sub-objective, from the bottom up, which depicts high suitability or low suitability for each sub-goal and sub-objective. They are then combined to create a suitability map that illustrates a more generalized goal and objective for a typical use (which is at the middle hierarchy in the goals and objectives organizational structure). This newly created map for the suitability of a more generalized objective, as mentioned above, will be the input files for the next step, i.e., the preference identification. Because there are a number of sub-objectives involved in the suitability analysis, a multiple utility assignment (MUA) technique, which weighs different sub-objectives, is applied by combining all the sub-objectives into this more generalized objective (the middle hierarchy of the goals and objectives). It is noted, however, that the suitability analysis only combines the sub-objective to the more generalized objective level (the middle objective as mentioned above) in the hierarchical structure of the goals and objectives. A higher level, which is the overall goals and objectives, will be combined in the next step, Step 4, the preference identification. In addition, the highest suitability of a typical use has the score of 9 while the lowest score is 1.

The preference identification is basically the step that combines the sub-categorized goals and objectives into the overall goals and objectives. To do this work, an Analytic Hierarchy Process (AHP) is employed. Throughout the AHP process, different goals at this step are combined based on the weight of each goal, which is to form a map for the overall goal. As a result, three uses to be allocated have three maps with each map corresponding with each use. These three maps are further input into the next step—Step 5—the conflict analysis with data manipulations.

Conflict analysis is the last step in the LUCIS model, which allocates the three uses based on the conflict nature that these three uses are competing with each other for a specific site. In this process, areas that cannot be developed or have already been developed will be excluded from the analysis process. Also, the preference maps will be normalized by making the values from 1 to 9 to between 0 and 1. Then, the normalized maps will be collapsed into three categories, which are low, medium, and high preferences represented by 1, 2, and 3, respectively, with the highest preferences coded as 3. The collapsed methods used in the LUCIS model comprise four collapsing methods, namely, the natural breaks method, the manual method, the equal interval method, and the standard deviation method (Carr and Zwick, 2007). When all the processes are completed, the collapsed maps will be combined into a conflict map, and final allocation will be conducted based on the conflict maps that have been created.

In the LUCIS model, future populations and future land use acreage demands are forecasted for urban use as a final allocation process. Gross urban density (GUD) is proposed for future land demands given the forecasts of future population. As a result, baseline future urban acreage demands are calculated using the above urban gross density. Carr and Zwick (2007) provided six steps for the final allocation, which are (1) allocating urban use if urban use is not in conflict with the other two uses; (2) allocating urban use for additional cells when urban use is in conflict with the other two uses; (3) creating "2050 remaining land" for urban use as their model is based on forecasts for the future 50 to 60 years; (4) allocating remaining cells to agricultural use when agricultural use is not in conflict with the other two uses; (5) allocating remaining cells to conservation land when conservation does not conflict with the other two uses; and (6) allocating remaining cells to agricultural or conservation uses when they are in conflict but with higher normalized preference values (Carr and Zwick, 2007, p. 167).

Carr and Zwick's (2007) model is inspirational to this study for its conflict analysis and final allocation elements. As a result, this study adopted some of the techniques that are utilized by Carr and Zwick (2007) in their models. Similar to Carr and Zwick (2007), this study will also apply a population forecast and the GUD to forecast future urban development demands in the next 20 years (2020-2030). In addition, a concept of conflict analysis is implemented, and four urban uses in this study, i.e., single-family, multi-family, commercial-institutional-transportation, and industrial-warehouses, are put forward into the analysis so that they compete with each other for a specific site. However, because this study adopts four uses, instead of three, the conflict analysis is somewhat more complex than the LUCIS model. As a result, more conflict scores (conflict combinations) are received from this study than the one in the LUCIS model (Table 4-23). In addition, this study does not deal with urban development from the angle of suitability but from the angle of probability as it extracts datasets from the MLR model. Therefore, this study does not need normalization for each use because the raw data from the MLR model, which has a value between 0 and 1, has already been normalized. The remaining conflict analysis procedures in this study are essentially the same as the LUCIS model, in which the collapsing process using the natural breaks method and the creation of conflict maps are recommended because they are essentially in the same process as the LUCIS model.

For the population forecasts, this study does not employ a regional population forecast method; rather it only utilizes the population for Alachua County because this study is not a region-based study. In addition, this study does not have goals and objectives that will be input into the conflict analysis because the model building suggests probability for a specific use rather than suitability. As a result, the model building approach from the MLR model in this study is fundamentally different from the LUCIS model, and suitability analysis is not conducted in this study, either. Therefore, this study does not have goals and objectives identification, suitability analysis, and preference identification: no MUA and AHP are involved in the allocation process. In addition, this study does not have a population forecast for the future 50-60 years either, as the LUCIS model proposes; but rather, the population in the future 20 years (to 2030) is projected in this research.

In terms of the preparation of data inventory, this study utilizes remote sensing data as well as the county parcel data for allocation. Because this study calls for several scenarios other than the baseline forecasting, which are introduced in the later part of this chapter, infill development is simulated by directly counting the parcel acreages for a specific urban use from the county’s parcel data. As a result, unlike the LUCIS model in which GIS vector datasets are used and converted to raster datasets, this study uses both the parcel data and remote sensing data, both the vector data and the raster data for study. In addition, because the county parcel data does not provide wetlands information, which this study does require and include, remote sensing data are utilized in this regard. Also, because the remote sensing data do not provide infill information, parcel data are employed at this point. Therefore, the advantages of both the parcel data and the remote sensing data are utilized in this study.

The literature review describes the most recent research that has been done so far across the academic field. In addition, from the above literature, many of the current urban studies are conducted based on the urban use that Hu and Lo (2007) suggest. Detailed urban categories such as single-family, multi-family, commercial, institutional, transportation, industrial, and warehouses are often not included. In addition, there is one study that attempts to use the MLR model to simulate urban development and redevelopment (Landis and Zhang, 1997). Although some studies address urban uses in detailed categories (Landis and Zhang, 1997), they did not predict future land uses through allocating land uses (Landis and Zhang, 1997). Some studies, although allocating future land uses, were neither based on detailed urban use categories nor in actual years (Hu and Lo, 2007). This research will address urban growth issues based on three perspectives: simulations involving various urban uses, conflict analysis adoptions, and various scenario predictions with clear timelines, borrowing the ideas from the LUCIS model and the CUF I and the CUF II models. Also, mixed-use will be simulated in this research, which further increases the broadness and breath of the research.

Methodology

Five Scenarios

The methodology of urban land use allocation is based on five scenarios: namely, the BAU scenario, the infill scenario, the increased density development scenario, the redevelopment scenario, and the conservation scenario. These five scenarios allocate land use and land cover based on the probability maps generated by the MLR model that is illustrated in Chapter 3. The probability maps are the maps that contain the probability numbers for each of the four urban land uses, namely, single-family, multi-family, commercial-institutional-transportation, and industrial-warehouses. These probability maps are extracted from the MLR model using two different masks based on the different scenario types: one is for the BAU scenario and the other is for the other four scenarios because the other four scenarios share the same geographic boundaries for development. Specifically, masks are the geographic boundaries that exclude existing areas and occupied urban uses, which are already developed, not suitable for development, or are forbidden for development. This study allocates four urban uses as introduced above based on these two mask types which are introduced in detail later in this chapter. According to the above five scenarios, the natural LULC classes such as grasslands, forests, agricultural, recreational-others, wetlands, water, and barren are emphatically not included in the allocation process. They are to remain the same as they were in 2003.

BAU scenario

The BAU development refers to the scenario that growth occurs according to the current development arrangement, in which the BAU scenario imitates generalized development but without many development restrictions. Development restrictions are development patterns that are imposed by the state or local authorities, in which the outward development is curbed by imposing environmental regulations on the land so that the outward development can move inward, suggested by cases such as in the patterns of infill, increased density, redevelopment, and conservation. In the BAU development scenario, however, outward development is allowed, and development on wetlands (Landis, 1995) as well as prime farmlands is also allowed. As a result, the wetlands and prime farmlands are not masked out in the scenario's mask setting, and the county's future conservation land is not masked out either. In this development pattern, existing urban uses, existing conservation land, water and creek buffers, utilities, high-voltage power lines, stormwater ponds, and parks and cemeteries are masked out, however. Besides, the BAU scenario does not restrict development within urban buffer areas, i.e., within 1,000 meters of the municipal and the Urban Cluster areas (urban area) in Alachua County, that is delineated by the Alachua County Growth Management Department (Figures 4-1 and 4-2) or outside the urban buffer areas. It can be developed, as mentioned above, anywhere in the county as long as there is vacant land available for development.

Figure 4-3 illustrates the mask of the BAU scenario. In this figure, the shaded areas are the areas that are masked out, and the remaining areas—the white areas—are the land that is available for development. From the figure, it is clear that the central city, the City of Gainesville, is almost all masked out. The lower-right portions of the map are also masked out because these belong to Payne's Prairie, which is a state park and state preserve.

Infill scenario

The infill development occurs in the county's vacant land that is set aside for a typical urban use by the county parcel data provided by the Alachua County Property Appraiser's Office. This vacant land is set aside not for present-day use but for future use. The vacant land that is set aside for future use can be any of the four urban uses: single-family, multi-family, commercial-institutional-transportation, or industrial-warehouses. For example, vacant residential uses, coded as 0000, are set aside as residential use—mostly single family use—in Alachua County's parcel data. Similarly, vacant commercial, vacant institutional, and vacant industrial are also set aside in the county's parcel data. They are coded as 1000, 7000, and 4000, respectively. In this case, when urban land uses are allocated, it will be the first priority to satisfy the vacant infill urban land needs that correspond to a typical urban use before they are allocated to other vacant, non-infill land.

The infill development mask differs from the one in the BAU scenario in that the infill development excludes future conservation for development (Figure 4-4). From Figure 4-4, most of the areas in the eastern and southeastern Alachua County in infill development are masked out, which means development in these parts of the areas are restricted or forbidden. These restricted and forbidden areas include wetlands, prime farmlands, and future conservation areas such as the Florida Forever land and the Alachua County Forever land in addition to the masked out areas in the BAU scenario. However, because of the priority of the vacant parcel data, it is imperative to satisfy the vacant, infill parcels before they go to the vacant non-infill land, as they will in the BAU scenario. Consequently, there will be overlap between the vacant infill land and the vacant non-infill land because vacant infill land can be located in vacant non-infill land, i.e., the conflict areas, which will be introduced in-depth in the later part of the chapter. Because there is no difference between the vacant infill land within the 1,000 meters urban buffer areas (urban area) and those outside 1,000 meters (greenfield area) (Figure 4-1), development can occur either inside or outside the 1,000 meters of urban buffer areas (Figure 4-5). Vacant land that is inside or outside the 1,000 meters of urban buffer areas has essentially the same opportunities for development when the infill land is actually allocated.

Increased density development scenario

The increased density development scenario presents the same mask as the infill development does. The increased density development takes place inside the 1,000 meters of urban buffer areas (Figure 4-5). Increased density occurs when the current density—the BAU scenario or the infill scenario—results in too much outward urban development. As a result, increased density is called for in order to reduce the development outside urban buffer areas. Specifically, the increased density development is the increase of GUD and correspondently increases all other densities such as the gross single-family density, the gross multi-family density, the gross residential density (GRD), the gross commercial-institutional-transportation density, and the gross industrial-warehouses density. The increased density development aims to reduce the occurrence of urban sprawl by significantly reducing the development acreages that happened outside 1,000 meters of the urban buffer area. This study adopts a reasonably increased GUD value compared to the ones in other cities and/or other counties in the State of Florida. The increased density development scenario is a continuous scenario evolved from the infill development. As a result, the increased density development incorporates infill development in the allocation process. When a certain land use is to be allocated, the first step will be the adoption of the infill process to see if a development can be accommodated in vacant infill lots. Because the infill process can be either located inside 1,000 meters of the urban buffer areas or outside the urban buffer areas (Figure 4-5), when the infill process is complete, it will then be allocated inside the increased density development areas where they are typically located within 1,000 meters of urban buffer areas because this study imitates increased density that is regulated to occur within 1,000 meters of the urban buffer area. As a result, the increased density development can also be written as the infill/increased density scenario.

Redevelopment scenario

The redevelopment scenario adopts a method to further reduce the occurrence of urban sprawl by allocating certain amounts of urban uses by a process of redevelopment in the central urban areas. When a typical urban use is allocated based on this scenario, a certain portion of an urban use will be set aside within the existing urban areas where the existing urban structures will be demolished and new structures will hence be established there on the existing urban land. After this process is complete, the urban use will be allocated to see if it can be developed inside the vacant urban infill areas. When this test is completed, the urban uses will be allocated upon the increased density areas within the 1,000 meters of the urban buffer areas. When the increased density allocation is accomplished in the increased density development areas, finally, mixed-use will be allocated within the 1,000 meters of urban buffer areas. The mixed-use is based on the locations of multi-family and commercial-institutional-transportation uses combined because mixed-use is the combination of multi-family and commercial retail, service, and office. As a result, areas containing multi-family and commercial-institutional-transportation will be scrutinized to see if they are suitable for mixed-use development. This process is undertaken by examining whether the areas are along the 200 meters of major roads because these road front areas are typically the areas apt to mixed-use development.

Generally speaking, based on the above, the redevelopment scenario is a combination of multiple development scenarios such as infill, increased density development, and mixed-use as introduced above, which can also be written as redevelopment/infill/increased density/mixed-use. As a result, the mask of the redevelopment scenario is essentially the same as the infill scenario as well as the increased density development scenario as introduced previously.

Conservation scenario

The conservation scenario allocates urban land based on the mask of all environmental conservation land, existing restricted urban areas, existing occupied land, as well as existing urban forbidden areas. This scenario is used to test if vacant land within urban conflict areas is adequate to accommodate the development by considering the exclusion of the environmental land in a maximum manner. As a result, it precludes the development not only in the existing conservation areas but also in the future conservation areas such as the Florida Forever Conservation land and the Alachua County Forever Conservation land. For urban restricted and forbidden areas, contrary to the BAU scenario, the conservation scenario does not allow development to occur on wetlands and prime farmlands. Consequently, the mask of the conservation scenario shares the same mask as the infill development, the increased density development, and the redevelopment.

The conservation scenario follows a similar allocation sequence for a typical land use as introduced above. First, the land use will be allocated within the existing urban areas for certain portions of redevelopment. Second, the land use will be allocated to see if it is within the increased density development areas. Third, the land use will be allocated to see if it accommodates mixed-use. The conservation scenario does not include the infill process. This is the major difference between the conservation scenario and the other three development scenarios such as the infill scenario, the increased density development scenario, and the redevelopment scenario. Based on the above scenario descriptions, the conservation scenario can also be written as conservation/ redevelopment/increased density/mixed use. Because of the infill development, the increased density development, and the redevelopment utilize the same mask as the conservation scenario, the mask of the conservation is the same as the above three scenarios, which is illustrated in Figure 4-4. Different from the above three scenarios (i.e., infill, increased density development, and redevelopment), however, the conservation scenario takes land directly from urban conflict areas. The urban conflict areas refer to the places that urban uses compete with each other to be allocated, which can take place either within 1,000 meters of urban buffer areas or outside. They will be introduced later in this chapter.

Forecasting Development Acreages for Five Scenarios

Baseline forecasting method (BAU forecasting and infill forecasting methods)

The development forecasts are preceded based on the timeframes of 2020, and 2030, respectively. The baseline forecasting method is also known as the BAU forecasting method as well as the infill forecasting method because the BAU scenario and the infill scenario use this technique to forecast future urban growth. In particular, the baseline forecasting method and the follow-up other three development scenarios employ a linear regression model to forecast urban growth acreages. This is because the population growth in Alachua County presents a linear growth pattern in 1982, 1994, 2003, and 2010, as well as in 2020 and 2030 in the future (Figure 4-6). During the forecasting procedure, GUD is adopted (Carr and Zwick, 2007).

The GUD refers to the total population occupying total urban land (Carr and Zwick, 2007) in the study area, where the total urban land comprises single-family, multi-family, commercial-institutional-transportation, and industrial-warehouses combined. When doing the forecasting, the first stage is to calculate the GUD for 2010. Because this study proposes that the GUD in 2010 remain constant (Carr and Zwick, 2007) as compared to those in 2020 and 2030, respectively, it will be easy to calculate the total urban acreages in 2020 and 2030, given that the value of the GUD in 2010 and the populations that have been forecasted are based on the Census Bureau's as well as the Bureau of Economic and Business Research's (BEBR) population data (Table 4-6). When the total urban acreages in 2020 and 2030 are attained, the next goal is to calculate the acreages of single-family use, multi-family use, commercial-institutional-transportation use, and industrial-warehouses use in 2020 and 2030, respectively; the acreage of each urban use in 2010 is directly counted from the 2010 parcel maps because these data are available in the county's 2010 parcel data. To do this work, which is to calculate the urban acreages based on different urban uses in 2020 and 2030, respectively, it is necessary to estimate the single-family and the multi-family population. This study proposes that the ratio of the single-family population to the multi-family population in different years, i.e., the SF/MF value, presents a linear decline format. As Figure 4-7 shows, the ratio of the single-family population to the multi-family population in 1982 was 1.53, while this ratio was 1.48 in 1994, 1.45 in 2003, 1.30 in 2010, and will be 1.26 in 2020, and 1.19 in 2030, which demonstrates a continual linear trend of decreasing throughout the years (Figure 4-7). The above SF/MF values are calculated using the SPSS V.19 software. Based on these numbers, it is easy to calculate the single-family population and the multi-family population in 2020 and 2030, respectively, which are illustrated in Tables 4-1 and 4-2. After this process is complete and because the gross single-family density and the gross multi-family density (the ratio of single-family population to the single-family acreage and the ratio of multi-family population to the multi-family acreage) remain constant throughout the years in 2010, 2020, and 2030, respectively, for the BAU scenario, the single-family and the multi-family acreages are designated simply by single-family or multi-family population being divided by the gross single-family density or the gross multi-family density in the years 2020 and 2030, respectively. Then, urban acreage changes based on different urban use categories are calculated simply by subtracting the previous years' urban acreages from the current year values for each of the four urban use types. Thus, it will be the same case for the future years in which the current years' urban acreages are subtracted from the future year values. The results are presented in Tables 4-1 and 4-2, respectively.

The baseline forecasting method for the commercial-institutional-transportation use and the industrial-warehouses use advocates a concept of employment population. The employment population is the count of the employment population based on different industrial and/or business types extracted from the *Florida Statistical Abstract 2003* (BEBR, 2003). Once the employment populations for commercial-institutional-transportation and industrial-warehouses are determined for a typical year, i.e., 2003, the gross commercial-institutional-transportation density and the gross industrial-warehouses density can also be determined given their acreages in 2003. As soon as their densities are established for 2003, their future densities, i.e., 2010, 2020 and 2030, can also be established following the same density as it was in the 2003 densities because the densities will remain constant for the BAU scenario, similar to the gross single-family density as well as the gross multi-family density. When the densities are resolved, the next stage is to calculate the commercial-institutional-transportation and the industrial-warehouses acreages. Because the ratio of the acreages of the commercial-institutional-transportation use to the acreages of the industrial-warehouses use in 2003 is a constant, i.e., 7.8, in this study, given that their total acreages derived from the total urban acreages minus total residential acreages, the final acreages for the commercial-institutional-transportation use and the industrial-warehouses use are computed. The results are presented in Tables 4-4 and 4-5, respectively.

In addition to the concept of the GUD, the GRD is also applied (Table 4-3). The GRD is the ratio of the total residential population to the total acreage of residential use, including both single-family use and multi-family use. The total residential population is the sum of single-family population and multi-family population. From Table 4-3, the GRD values increase from 2003 through 2030, which means that more people will reside on the same acreages of land compared to the values in the previous years.

Finally, urban acreage changes based on the four urban land uses are computed and illustrated in Table 4-7. It represents the acreages and cells for each of the four urban uses that will be developed in 2010, 2020, and 2030, respectively. In the table, cell numbers are utilized for the allocation of each of the four urban uses and meanwhile for projecting them on maps, which is explained in detail in the later part of the chapter. The values that are critical to the baseline forecasting method are illustrated in the Tables 4-1 through 4-8.

Increased density forecasting method

The increased density forecasting method applies an increased GUD compared to the baseline forecasting method. The increased density forecasting method is used by the increased density scenario in order to "shrink" the development acreage bases to a certain level so as to curb urban sprawl by accommodating more people in the same acreage of land compared to the non-shrinkage baseline forecasting method. This study adopts an increased density that is 15 percent more than the baseline GUD. As a result, the original baseline GUD, which is 2.61

people/acre, is elevated to 3.0 people/acre in the increased density forecasting method (Table 4-14). This new GUD is in reference with the one in Orange County, Fla., where the City of Orlando is located. The new GUD is a reasonably increased value—not too high—because Alachua County is a rural county, compared to the values in other Florida counties and cities.

Once the increased GUD is determined, similar to the baseline forecasting method, the total urban acreages based on 2020 and 2030, respectively, are computed simply by the total population being divided by the increased new GUD. The next step is to determine the ratio of the total residential acreage to the total urban acreage. These numbers are from the baseline forecasting method, which are illustrated in Table 4-8. From Table 4-8, the percentage of the residential acreage out of the total urban acreage in 2020 is 67.92 percent while it is 54.78 percent in 2030. Correspondingly, the percentage of the residential acreage change out of the total urban change in the increase density forecasting method is also 67.92 percent in 2020 and 54.78 percent in 2030, and they are the same as the percentages for the baseline forecasting method. Based on the calculated percentages of the residential acreage changes in 2020 and 2030 as well as the total urban acreages in 2020 and 2030, respectively, in the increased density forecasting method, it is not difficult to calculate the residential acreage changes in 2020 and 2030. The results are illustrated in Table 4-11. The calculation of the commercial-institutional-transportation acreage as well as the industrial-warehouses acreage also comes from the percentages of commercial-institutional-transportation use as well as industrial-warehouses occupying the total urban acreages for the baseline forecast in Table 4-8.

Finally, given that the percentages of the single-family makes up out of the total residential acreages in 2020 and 2030, respectively, in the baseline forecasting method (Table 4-8), the single-family and the multi-family acreages in 2020 and 2030 in the increased density forecasting method can be obtained simply from the total residential acreages in 2020 and 2030, respectively, multiplying the occupancies of the single-family and multi-family uses by the total urban acreages (Table 4-8). The results are illustrated in Tables 4-9 and 4-10. As mentioned previously, given the percentages of the commercial-institutional-transportation use as well as the ratio of the industrial-warehouses use to the total urban acreages, it is not difficult to estimate the acreages of the commercial-institutional-transportation use as well as the industrial-warehouses use (Tables 4-12 and 4-13). The total change for each of the four urban land uses is presented in Table 4-15. From the above tables, with the increase of the GUD in 2020 and 2030, the gross single-family density, the gross of multi-family density, the gross residential density, the gross commercial-institutional-transportation density, and the gross industrial-warehouses density are all increased compared to the ones in the baseline forecast.

Redevelopment forecasting method

The last forecasting method is the redevelopment forecasting method, in which the redevelopment scenario and the conservation scenario accept this forecasting method to predict future urban growth. Similar to the increased density method, the redevelopment forecasting method also attempts to reduce the development that goes outside the urban buffer areas by allocating certain acres inwards for redevelopment; because of that, urban sprawl in the greenfield areas will be controlled. This study suggests a 15 percent redevelopment ratio in the urban buffer areas. Based on this ratio, significant amounts of new urban development can be curtailed and reduced.

The calculation of the redevelopment forecasting method is essentially similar as the one in the increased density forecasting method, in which the changes of the total urban acreage are calculated first before the redevelopment process takes place. During this forecasting process, the same values in the total urban acreage changes in the increased density forecasting method

(Table 4-14) are adopted, which are shown in Table 4-21; so are the values for other acreages such as the single-family acreage, the multi-family acreage, the total residential acreage, the commercial-institutional-transportation acreage, and the industrial-warehouses acreage, which are the same as the increased density forecasting method proposed and are adopted by the redevelopment forecasting method (Tables 4-9 through 4-14). Then, redevelopment values are calculated based on the ratio of 15 percent as mentioned above, which leaves 85 percent of urban development for generalized development. Finally, the GUD as well as the associated densities for different urban land uses are estimated based on the 15 percent redevelopment ratio as well as the population forecasts given by the BEBR data and the population estimates for single-family and multi-family for 2020 and 2030, respectively, as illustrated in the baseline forecasting method. From Tables 4-16 through 4-22, it is obvious that the densities increase, considering the 15 percent redevelopment ratio, which means that the redevelopment scenario increases densities inside the urban buffer areas. The values of the redevelopment forecasting method are presented in Tables 4-16 through 4-22.

The Conflict Analysis

The conflict analysis is the one type of analysis that offers ways to allocate land for a typical urban use in order to find out whether it belongs to the single-family use, the multi-family use, the commercial-institutional-transportation use, or the industrial-warehouses use. The conflict analysis is based on the thesis that the four urban uses compete with each other for a typical site (Landis, 1995). As a result, the four urban uses are put in a parallel manner (Carr and Zwick, 2007) so that each urban use can be compared with others side by side so as to determine which use the site is best suited for. For example, for a specific site, the conflict analysis will result in a score in the format of XXXX, e.g., 2341 (or any other combinations of the four digits), in which the thousands digit, 2, refers to the single-family use, the hundreds digit, 3, refers to the multi-family use, the tens digit, 4, refers to the commercial-institutional-transportation use, and the units digit, 1, refers to the industrial-warehouses use. Continued from the above example, because 4 is the largest number in the combination in which it is the highest score, the commercial-institutional-transportation use is the preferred use for this site; thus, this site belongs to the commercial-institutional-transportation use. When the same two or three highest scores occur in the combination, e.g., 4423 or 3331, it means that these two or three of urban uses are in minor conflict (Carr and Zwick, 2007) in that each of the two or three uses have the same opportunities to be allocated to a specific use. In the above 4423 example, the single-family use and the multi-family use are in minor conflict, and these two uses have the same chances to be allocated to their specific uses. Similarly, as the case of 3331, the single-family use, the multi-family use, and the commercial-institutional-transportation use all have the same chances to be allocated to their specific uses. When all the four scores are the highest, e.g., 4444 or 3333, it means that the four urban uses are all in major conflict (Carr and Zwick, 2007) with regard to a high preference status, in that all the four urban uses have the same chances to be developed according to a certain use on a given site. However, when all the four scores are the lowest, e.g., 1111, it means all the four urban uses are all in major conflict with regard to a low preference position, which refers to the situation that none of the uses are suitable for their certain uses to be allocated on a given site.

Carr and Zwick (2007) proposed a four-step method for the conflict analysis to allocate land uses. They include: (1) creation of a mask that excludes the areas that are developed, not suitable for development, or restricted for development; (2) normalization of the suitability maps; (3) collapse of the normalized preference maps; and (4) combination of the collapsed maps. Carr and Zwick's (2007) method is based on the suitability analysis in which the values are between 1 and 9 (Carr and Zwick, 2007). Therefore, their method needs normalization so as to adjust the values between 0 and 1, as mentioned previously. Because, in this study, the probability maps extracted from the logistic regression model have continuous values between 0 and 1, they are already normalized, and no normalization is needed in this regard. As a result, this study adopts a three-step method for the conflict analysis: (1) creation of a mask to exclude the areas that are already developed, not suitable for development, or restricted or forbidden for development; (2) collapse of the preference maps; and (3) creation of the conflict maps. Table 4-23 illustrates the conflict scores and their equivalent descriptions in the conflict analysis in that all the combinations of the conflict scores are put forward. The conflict analysis is the selection of the conflict uses based on different scores from Table 4-23, in which masks are applied that conform to different development scenarios. In this study, two conflict masks are proposed: one is the BAU mask; the other is the infill development mask. The infill development mask is suitable for the infill scenario, the increased density development scenario, the redevelopment scenario, and the conservation scenario, respectively, because all these scenarios utilize this same mask for the conflict analysis.

Creation of the masks

As mentioned above, the mask creation incorporates two mask types: one is for the BAU scenario; the other is for the infill scenario, which is also the mask for the increased density development scenario, the redevelopment scenario, and the conservation scenario, because the last four scenarios use the same mask. These two masks are illustrated in Figures 4-3 and 4-4, respectively. The function of a mask is to exclude the areas that are already developed, not suitable for development, or forbidden for development; the leftover areas, after masking, are those areas that can be developed.

From the BAU mask, it is evident that many areas in the eastern and western parts of the county are vacant and available for development because the future conservation land is not included for masking (Figure 4-3). Because of the large areas that are not excluded for development, development choices in the county are ample in the BAU scenario. As a result, sprawled urban development that is scattered around the county outside 1,000 meters of urban buffer areas is evident based on the recent urban development practices from 2003 to 2010. Paradoxically, the opportunity for urban development in the infill development is much less than such opportunity in the BAU scenario. Because large areas of future conservation land are included in the masking process for the infill development, they mostly are located in the eastern and southeastern parts of the county; many development opportunities in the eastern part of the county are rigorously restricted (Figure 4-4). The developable land is mostly concentrated in the western part of the county, however.

The purpose of the creation of the two types of masks is to extract the preference maps that include developable land only because the raw preference maps that are derived from the logistic regression model contain both the developable and non-developable/developed land. This is an important step because the conflict maps cannot be properly produced without it, and this masking technique successfully leads to the creation of the conflict maps by leaving out undevelopable areas so that the remaining parts can be chosen for development. When mask creation is completed, the next job is to collapse the preference maps that have been masked into four categories in accordance with the four urban uses so that the conflict maps can be created thereupon.

Collapse of the preference maps

When the preference maps originating from the logistic regression model are masked, the next phase is to collapse the preference maps. The collapsed maps use the natural breaks scheme to divide the preference numbers, which are given the probability values between 0 and 1, into four categories. This study has tested different collapsing methods: the natural breaks method, the manual method, the equal interval method, and the standard deviation method; the natural breaks method handles a bimodal or multimodal distribution (Carr and Zwick, 2007) and creates four collapsed groups. The other three collapsing methods yield only three or even two collapsed groups. Consequently, the natural breaks method is adopted, which brings forth the four groups symbolized namely by 1, 2, 3, and 4, in which 4 is the highest score representing the highest probability for development and 1 is the lowest score suggesting the lowest probability. Thus, the continuous numbers in the preference maps become discrete in the collapsing maps. This collapsing process establishes a basis for the creation of the conflict maps which will be introduced in the next section. Figure 4-8 illustrates the example of a preference map, and Figure 4-9 illustrates a sample of a collapsed map for the BAU scenario.

Creation of the conflict maps

Once the collapsed maps are prepared for the five scenarios, the follow-up effort is to create conflict maps. Because this study has proposed two masks for different development scenarios, the conflict maps will have the same two equivalent mask types also: one is for the BAU scenario; the other is for the infill scenario as well as for the other three scenarios. When making a conflict map, collapsed maps for each urban use are combined by taking the score of the single-family use multiplied by 1,000 plus the score of the multi-family use multiplied by 100 plus the score of the commercial-institutional-transportation use multiplied by 10 and plus the score of the industrial-warehouses use multiplied by 1. For example, for the BAU scenario, this study has created four collapsed maps for each urban use. The single-family use has its own collapsed map while the other three urban uses all have their own collapsed maps. When they are combined to create a conflict map, they are input into the Single Output Map Algebra function in ArcGIS 9.3 to fulfill this task by making:

single-family × *1,000* + *multi-family* × *100* + *commercial-institutional-transportation* × *10* + *industrial-warehouses* × *1*

The final scores for the BAU scenario are presented in Table 4-24, and the final scores for the infill scenario as well as for the other three scenarios are presented in Table 4-25 because these three scenarios have a same conflict score. From the two tables, it is clear that the BAU scenario has a lot more vacant land for development compared to the infill scenario as well as the other three development scenarios because the BAU scenario has more land that comes to participate in the allocation process. Specifically, for the BAU scenario, the allocation directly goes to the conflict areas in order to seek vacant land. For the infill scenario, however, the infill development will first fill out the infill lots that are already set aside as vacant land for a typical urban use in the county's parcel data, either inside 1,000 meters of urban buffer areas or outside. Finally, when the infill lots are filled up, the development will go to the conflict areas so as to allocate land, according to the conflict scores a development will receive. Therefore, different from the BAU scenario, for the infill development, the sequence begins with infill allocation and end with the conflict allocation. For the increased density development, the development will allocate land in the infill lots that are either located within 1,000 meters of the urban buffer areas or outside. When this land is filled up, then it will go to the conflict areas to seek vacant land, which are typically located inside the urban buffer areas. This is the same as for the redevelopment scenario in which the land within the infill lots as well as the 1,000 meters of urban buffer areas is filled up first, before going to the conflict areas to find available land. For the conservation scenario, however, it does not go to the infill vacant lots as a first step but goes to the conflict areas directly in order to seek available land. These are the major differences when these five scenarios are compared and allocate urban land.

When the conflict areas are allocated for a specific scenario, it will first go to the specific urban use that dominates in the conflict map. For example, when a score of 4211 is suggested (Table 4-24), it is a single-family dominated area. As a result, the development pattern will be the single-family development. When all the areas dominated by single-family are filled up, then the allocation process will go and seek the single-family conflict areas with other urban uses. When these amounts of land are consumed also, the allocation process will finally utilize all of the conflict areas for all four urban uses. The commercial-institutional-transportation development in this study is the case in which the development finally finds the scores of 1111 in both the two conflict types because the commercial-institutional-transportation dominant areas and the areas of the commercial-institutional-transportation use in minor conflict with other uses cannot satisfy the development needs. As a result, it goes to the all-low-preference-all-in-conflict areas to look for needed cells.

Generally speaking, for a proper allocation of urban uses, it is imperative to follow the development sequence that single-family dominates over multi-family, multi-family dominates over commercial-institutional-transportation, and commercial-institutional-transportation dominates over industrial-warehouses. These allocations all put single-family as the first priority and the remaining urban uses in secondary status. For example, based on this rule, if a single-family use is in conflict with other urban uses, the cells belong to the single-family use. Once single-family acreage demands are satisfied, then, the cells in conflict can be allocated to other uses.

From Tables 4-24 and 4-25, it is obvious that, generally speaking, the BAU scenario has adequate land for development while the infill scenario and the other three scenarios have reduced approximately 50 percent of the total acres of land. In particular, the commercial-institutional-transportation use dominating areas (i.e., 1231, 1141, 1132, 1131, and 1121) are not sufficiently large enough to meet the development needs. As a result, they will have to be allocated from 1111, the all-low-preference-all-in-conflict areas. Figures 4-10 and 4-11 are the conflict maps for the BAU scenario as well as the infill development scenario.

Final Allocation

Timeframes of the final allocation

The timeframes of the final allocation follow the timeframes of the foresting methods, in which 2020 and 2030 are proposed. Because of the two timeframes, the final allocation will offer two maps for each scenario: namely, 2020 allocation, and 2030 allocation. These allocations are basically the results of the conflict maps as well as the infill parcels that are derived from the above conflict analysis as well as the county's parcel data. The allocation is the designation of the urban uses being allocated based on the five scenarios in accordance with the forecasts of the urban development acreages for each urban use and for each scenario. Because the 2010 allocation map depicts the LULC classes in 2010 that have already been passed, the 2010 allocation map shares the same map, and it is counted from the county's 2010 parcel data directly. Therefore, a total of twenty-two maps are created based on the five scenarios, with each scenario having two maps for a year, i.e., 2020 or 2030, and two maps for the year 2010, which is the same for all five scenarios in that year (see Figures 4-12 through 4-33).

Urban uses to be allocated in final allocation

The urban uses that will be allocated are the four uses mentioned previously: namely, the single-family use, the multi-family use, the commercial-institutional-transportation use, and the industrial-warehouses use. Specifically, mixed-use is allocated in the redevelopment scenario as well as in the conservation scenario. Consequently, these two scenarios contain the mixed-use in their final allocation maps in addition to the four urban uses mentioned above. In addition to the urban uses, natural LULC classes such as grasslands, forests, agricultural, recreational-others, wetlands, water, and barren will be mapped also. Specifically for the barren land, landfill is suggested by this natural LULC. In addition to the mixed use, existing and future conservation land is marked out as a result in the conservation scenario maps.

Final allocation sequence

The allocation sequence follows the scenario sequence described in the forecasting section. The first scenario that will be allocated is the BAU scenario; the second to be allocated is the infill development scenario; the third is the increased density development scenario; the fourth is the redevelopment scenario; and the fifth is the conservation scenario. Based on the allocation forecasting method for each scenario, the overall land consumptions that are forecasted for the five scenarios are actually decreasing, and they are mostly located within the urban buffer areas. This development pattern presents an optimistic form for curbing urban sprawl in Alachua County with respect to urban land acreage consumptions as well as urban development locations.

Final allocation cell size and spatial reference

The raster maps adopt a 30 × 30 meters resolution, in accordance with the resolution of the remote sensing datasets. As a result, the computation of cell counts and cell acreages are based on this resolution. One cell, i.e., 30 meter × 30 meter, is 900 square meters, which is equivalent to 0.22 acres. Therefore, the acreage computation is based on this conversion rule to calculate to and fro from the metric square meters to the English acres and vice versa. In particular, the spatial reference for the allocation maps is: NAD_1983_UTM_Zone_17N. The datum is: D_North_American_1983.

Final allocation preference map slicing, policy allocation, and cell statistics

Once the preference maps are masked, they need to be transformed from the continuous data to the discrete data so that cells with certain values can be extracted, calculated, and allocated. The continuous data are the data that contain floating points that are continuously stretched between 0 and 1. There are no unique values that can be derived from the value of a certain cell in the continuous data. As a result, the continuous data do not have an attribute table in ArcGIS 9.3. The discrete data, however, incorporate integers, which have unique values for cells. They have attribute tables for the values of all cells. This study adopts a technique called "slice" to convert continuous data to discrete data by slicing or dividing a large area into a number of smaller equal zones (Carr and Zwick, 2007) for allocations based on their cell values. As a result, the cells are extracted and grouped according to the magnitude of their values. The biggest preference values are assigned to the biggest sliced values (Carr and Zwick, 2007) when they are being allocated after slicing. As a result, the highest sliced values (i.e., the highest preference values) are the first to be allocated, and then follows the cells with smaller numbers so that the smaller numbers are allocated later in a secondary status, and the smallest are allocated in the last.

The common slicing method has two alternatives: one is to use the Slice tool in ArcGIS 9.3; the other is to use Single Output Map Algebra by making a preference map and multiplying a certain value, e.g., 100,000. Because the Slice tool in ArcGIS 9.3 sometimes does not create sufficiently small units for allocation, which the allocation process prefers, this study uses the second method to slice certain areas in a preference map because this second method can potentially create sufficiently small units of land by multiplying a sufficiently large constant, e.g., 100,000, in this study, with the preference maps. Overall, the slice technique can be used to extract discrete values from a continuous dataset, e.g., a preference map, which need conflict areas or infill lots as masks in order to seize those masked cells for allocation. In this case, the vector infill parcel map must be converted to a raster dataset first. This study applies the slice technique (the second method) to slice preference maps for certain areas covered by cells in conflict or infill lots.

When a cell is still large after slicing, this study adopts a policy allocation method to further extract necessary cells from it. For example, an area contains 12,070 cells after slicing and only 7,000 or so cells are needed based on the development forecasts. This study utilizes a policy allocation method to allocate these 7,000 cells, in which cells that are located within a jurisdiction with clear geographic boundaries are chosen as high priority compared to cells that are not located within a certain jurisdiction in order to mimic local governments' policy roles for urban development. In the above example, within 1,000 meters of urban buffer area is taken for development, which yields 7,132 cells, and these 7,132 cells are what should be allocated.

For the areas that are overlapped, e.g., industrial-warehouses infill development overlaps with the commercial-institutional-transportation conflict areas, this study adopts a Cell Statistics tool to extract overlapped cells so that the formal counts of industrial-warehouses infill cells as well as the commercial-institutional-transportation acreages can exclude the overlapping cells. The Single Output Map Algebra module can also be used in this regard. This study uses cell statistics to accomplish this goal, however.

Final Allocation Results

BAU Scenario

As mentioned previously, because the year 2010 has passed, the allocation for 2010 has the database support from the parcel data in the Alachua County Property Appraiser's Office. As a result, the allocation map for 2010 is a direct count from the parcel data for the four urban uses that reflect the parcel acreages and their changes throughout the years. This 2010 map, titled

“Current Plan,” is essentially the same for all five development scenarios for the same year. In reality, the development for 2020 and 2030 for all five scenarios starts from this map.

Figures 4-12 through 4-17 illustrate the allocation based on the parcel maps from the county appraiser’s office for the four urban uses for 2010 as well as the final allocation maps for the BAU scenario for 2020 and 2030, respectively. From the map, it is evident that the urban sprawl has occurred from 2010 because cells have been scattered throughout the county starting from 2010 although the level of the sprawl is not very severe in that year. The 2020 and 2030 BAU maps show a continued sprawled development pattern, which is quite serious because the leapfrog development continues outside urban buffer areas in the greenfield areas where the greenfield land is gradually being filled out to become urbanized areas in the outskirts of Alachua County, outside the urban buffer areas. In 2010, the leapfrog single-family development outside urban buffer areas is largely located in the areas of the College West Estates, Scattered Oaks Estates, Sunny Meadows, and Windy Acres. It is also located in the area of Windy Hills; Pinesville; the area of the University Country Estates; Wacahoota; the area of the Grassy Lake Estates; and the Monteocha vicinity. The leapfrog urban development in the greenfield starts to grow from these areas in 2020 and in 2030; as a result, the single-family development continues in the above areas for the above years. The growth becomes larger in area and more concentrated, which spreads to nearby parcels and fills the gaps between the urban buffer areas in the greenfield areas in 2020 and 2030. Even the areas in east Alachua County, where the single-family development is minor in 2010, have incredible growth in the areas such as the Wacahoota vicinity, the Micanopy vicinity, the Winsor vicinity, the Melrose vicinity, the Monteocha vicinity, the Wood Meadows Neighborhood, the Traxler vicinity, the Bellamy Road Estates, and the Bellamy Forest Neighborhood in 2020. Besides, additional single-family growth is scattered around in the outskirt of the county other than the development in the above areas in 2020. In 2030, the single-family growth pattern continues to fill out the gaps between the urban buffer areas, and the vacant land in the gap areas are largely filled out. However, there are still significant areas inside the urban buffer areas that are vacant, which are suitable for development but are not developed. Overall, it is obvious that for the BAU urban development, the growth pattern is a sprawled type, which is characterized as severe leapfrog development in the greenfield areas outside the urban buffer areas.

Another noticeable development is the commercial-institutional-transportation development, which is developed along State Route 441 on the north side of Gainesville as well as south of the Gainesville Regional Airport next to the Alachua County Agricultural Center and Fairgrounds. The area that is in the Jonesville vicinity south of NW 39th Avenue is also a big commercial-institutional-transportation development. Nevertheless, this development is all within the urban buffer areas and there is no sprawled pattern outside the urban buffer areas for the development of this land use. Tables 4-26 through 4-29 illustrate the developed acreage of land in 2010, 2020, and 2030, respectively, with comparisons between the allocated acreages and the demanded acreages in those years.

Infill Scenario

Figures 4-18 through 4-21 describe allocations for the infill development scenario. The infill scenario causes the development to occur in the infill parcels before the acreages are allocated to the conflict areas. The infill acreages can either be located within 1,000 meters of urban buffer areas or outside the areas. The infill scenario begins in 2020 because the infill scenario allocation in 2010 shares the same map with the BAU scenario. The single-family development in the infill scenario for 2020 is all located within the infill areas, which have 57,938 cells, equivalent to 12,885 acres (Table 4-31). The single-family development mostly occurs within the 1,000 meters of urban buffer areas in 2020. They are basically located in the nodal areas where major transportation corridors intersect. These nodal areas include the City of Alachua, the City of High Springs, the City of Newberry, the City of Archer, the City of Waldo, and the City of Hawthorne as Figure 4-2 shows. The areas in east Alachua County and west of Newnan's Lake also contain noticeable single-family development. Because most of the single-family development occurs within the urban buffer areas, not much leapfrog development is evident in the infill scenario in 2020. There are a few, minor acreages of the commercial-institutional-transportation use that are scattered outside the urban buffer areas, however. Multi-family and industrial-warehouses uses are all located within 1,000 meters of urban buffer areas.

Leapfrog development is quite evident for the single-family development in 2030. The leapfrog single-family development occurs fairly intensively in the greenfield areas outside urban buffer areas in 2030. Examples include the area of County Route 232 intersecting with County Route 235; the area of State Route 121 intersecting with NW 156 Avenue; the area of NE 156 Avenue intersecting with NE 39th Street; the area of County Route 1469 intersecting with NE US 301; the area of County Route 234 intersecting with County Route 1474; and the area of SW 137 Avenue intersecting with SW 91 Street. Figure 4-2 shows the county's major roads. The single-family infill scenario for 2030 allocates a total of 25,990 cells, equivalent to 5,780 acres (Table 4-31), all within the infill areas in the county's parcel data (Table 4-30).

The multi-family development allows some acreages to occur in the infill areas and some in the conflict areas in 2020, but when the infill acreages are used up in 2020, the development, therefore, goes to the conflict areas (Table 4-32). For the commercial-institutional-transportation use, there is some development (about 45 percent) in the infill areas, and when it is used up in 2020, the development goes to the conflict areas in 2030. The commercial-institutional-transportation use is all located within urban buffer areas and all within the conflict areas in 2030, however. All the industrial-warehouses development happens in the infill areas both for 2020 and 2030 (Table 4-30). The four urban uses with their development in the infill areas as well as in the conflict areas are illustrated in Table 4-30. The four urban uses with their acreage demand and allocation are illustrated in Tables 4-31 through 4-34.

Increased Density Development Scenario

The increased density development scenario follows the infill scenario with “shrunk” urban development acreages for the four urban development patterns. The single-family development is characterized as the infill development in 2020 and 2030, but with inward development patterns inside the 1,000 meters of urban buffer areas. Most of the multi-family development is located within conflict areas for 2020, and the multi-family development all occurs in the conflict areas in 2030 after the development consumes all the infill acreages in 2020. For the commercial-institutional-transportation development, the development pattern is infill-oriented with a few being located in the conflict areas in 2020. The commercial-institutional-transportation development in 2030 is mostly located in the conflict areas when the infill acreages are filled out. The industrial-warehouses development is all located in the infill areas in both 2020 and 2030. The proposed development acreages in the infill areas as well as in the conflict areas in 2020 and 2030 are shown in Table 4-35.

The allocated land versus demanded land for the four urban uses is shown in Tables 4-36 through 4-39. Increased single-family use in 2020 and 2030 all occurs within the urban buffer areas, which means no leapfrog sprawl takes place during the above period. The same is true for the multi-family use, the commercial-institutional-transportation use, and the industrial-warehouses use: all of the development happens within urban buffer areas. Apart from some leapfrog development in 2010, the urban development in the increased development scenario in

2020 and 2030 is all located within the urban buffer areas, with no leapfrog development in those years.

Figures 4-22 through 4-25 show that growth in 2020 happens in the central city—the City of Gainesville—as well as in the nodal areas where major road corridors intersect. These nodal areas include the City of Alachua, the City of High Springs, and the City of Newberry (Figure 4-2). The growth inside the Urban Cluster areas in 2020 is largely located in east Gainesville close to the Newnan's Lake (Figures 4-22 through 4-25). Growth in 2030 continues the pattern in 2020, but more intensively in the City of Gainesville as well as in the nodal areas. The growth fills in the gap areas in east Gainesville, outside 1,000 meters of urban buffer areas, in north Gainesville, in west Gainesville, in south Gainesville, as well as around the airport areas. Significant amounts of development are evident in the above areas; they are fairly intensive although they are all located within the urban buffer areas.

Redevelopment Scenario

The redevelopment scenario further shrinks the development acreages for the four urban uses similar to the increased density development scenario. The redevelopment allocation combines the redevelopment scenario, the infill scenario, the increased density development scenario, and the mixed-use together. Because there is 15 percent of urban development occurring within the central city—the City of Gainesville—as well as the Urban Cluster areas for redevelopment, the new urban development generally occurs within the central city, in the Urban Cluster areas, as well as in the nodal areas where major transportation corridors intersect such as the City of Alachua (Figure 4-2). Most of the redevelopment growth is located within the City of Gainesville city limits as well as in the county Urban Cluster areas in 2020 and 2030, respectively. In 2020, the single-family development is largely located within both the Urban Cluster areas in east Gainesville and west Gainesville. The multi-family development, in 2020, is located in the Urban Cluster areas as well in both east Gainesville and west Gainesville, although the development is not much. The commercial-institutional-transportation use, in 2020, is largely located within west Gainesville, in the nodal areas around the City of Alachua, and along State Route 441. The industrial-warehouses use is located at the intersection of State Route 441 and NW 34th Street in 2020. The above urban development is all located within the urban buffer areas, however, with no sprawl in 2020, although there is some sprawled development in 2010.

In 2030, the single-family development continues to grow in the Urban Cluster areas, most of which is still located within east Gainesville and west of the Newnan's Lake area. But some growth occurs within the nodal areas of the City of High Springs instead of the City of Alachua. In 2030, multi-family development grows within west Gainesville (on the west side of the Urban Cluster areas). The amount of development is still not very much compared to the amount of development in 2020. In 2030, the major commercial-institutional-transportation development is located in the area where Waldo Road intersects with NE 39th Avenue. Also, some development is located north of NE 39th Avenue and east of the North Main Street area as well as south of State Route 441 and west of NW 43rd Street. In 2030, the industrial-warehouses use is largely located at the intersection of NW County Route 235 and NW 173 Street as well as in the area north of the Gainesville Regional Airport (Figure 4-2).

In 2020, all single-family development occurs in infill parcels within 1,000 meters of the urban buffer areas. Most of the multi-family development occurs within the conflict areas with a small amount of multi-family development (54 acres) in the infill areas. Most of the commercial-institutional-transportation use is located within the infill areas while a small fraction of the commercial-institutional-transportation use (78 acres) is located in the conflict area. The industrial-warehouses use is all situated within the infill areas, however.

In 2030, single-family development still occupies the infill areas because the infill parcels for single-family haven't been filled up thus far. For multi-family development, when the infill multi-family cells are filled up in 2020, the development in 2030 receives land in the conflict areas. The commercial-institutional-transportation occupies both the infill areas and the conflict areas in 2030 because there is still some land available for infilling (583 acres). Same as 2020, the industrial-warehouses use takes all infill land based on the development sequence that infill development is prioritized over the development in the conflict areas.

The mixed-use development usually occurs inside the commercial-institutional-transportation use, the multi-family development, and along major transportation corridors. This study imitates land that is 200 meters away from the major transportation corridors to be seized for mixed-use development, in which they are also located inside a typical commercial-institutional-transportation use as mixed-use land. As a result, they are basically located within the central city—the City of Gainesville—in the Urban Cluster areas, or along State Route 441 for 2020 and 2030. In the City of Gainesville, the mixed-use land is located along these transportation corridors: NW 53rd Avenue, NE 39th Avenue, North Main Street, NE 31st Avenue, West Newberry Road, NW 23rd Avenue, SW 24th Avenue, SW 75th Street, and SW Archer Road (Figure 4-2). The mixed-use land occupies a total of 687 acres in 2020 and it develops to a total of 2,029 acres in 2030. Table 4-40 presents the acreages that are located within infill areas as well as in the conflict areas for the four urban uses in 2020 and 2030, respectively. Tables 4-41 through 4-44 illustrate the urban acreages demand verses urban acreages allocation for the four urban uses in 2020 and 2030, respectively. Figures 4-26 through 4-29 show the urban redevelopment in 2020 and 2030, respectively.

Conservation Scenario

Similar to the redevelopment scenario, acreages in the conservation scenario also "shrink" so that land to be allocated is smaller in acreage compared to the BAU scenario, the infill scenario, and the increased density development scenario so as to curb urban sprawl. This scenario adopts a 15 percent redevelopment benchmark, same as the value for the redevelopment scenario. However, different from the infill scenario, the increased density scenario, and the redevelopment scenario, in which infill land is taken priority over the land in the conflict areas, the conservation uses land only from the conflict areas. As a result, the total development acreages for the conservation scenario for the four urban uses can be located either within the 1,000 meters of urban buffer areas or outside. Because all land is located within conflict areas, the conservation scenario does not have overlaps because it is already separated in the conflict analysis where land is divided into categories based on different conflict scores each urban cell receives (Table 4-25).

Figures 4-30 through 4-33 illustrate the urban development in 2020 and 2030, respectively, based on the conservation scenario. Taking an in-depth look into the allocation map for the conservation scenario, most of the single-family development (80 percent) will be located inside the 1,000 meters of urban buffer areas in 2020 although there is some single-family development that is scattered outside the urban buffer areas. Furthermore, not much single-family development occurs within the City of Gainesville city limits, most of which, however, is located in the Urban Cluster areas as well as within the city limits of other incorporated cities in the county such as the City of Alachua, the City of High Springs, the City of Newberry, and the City of Archer (Figure 4-2). In 2020, about 83 percent of the multi-family development is situated within 1,000 meters of urban buffer areas although the multi-family development is not much, which has only a total of 247 acres in 2020. For the commercial-institutional-transportation use, however, 99 percent of the development occurs within 1,000 meters of the urban buffer areas, most of which is located within the City of Gainesville city limits. Most of the industrial-warehouses development takes place along State Route 441 and NW County Route 235 inside 1,000 meters of the urban buffer areas, which occupies about 63 percent of the total industrial-warehouses acreages. Overall, the urban sprawl with outward leapfrog development outside 1,000 meters of urban buffer areas is not severe in the conservation scenario compared to the BAU scenario, in that both scenarios take land from the conflict areas without infill cells, because most of the development in the conservation scenario allocates land inside 1,000 meters of urban buffer areas.

For 2030, about 57 percent of the single-family development happens within 1,000 meters of urban buffer areas. Scattered single-family development is evident outside 1,000 meters of the urban buffer areas where they are concentrated in the gap areas surrounded by the urban buffer areas; however, their locations are outside 1,000 meters of urban buffer areas. For the multi-family development, about 86 percent of the development is located within 1,000 meters of the urban buffer areas and they are mostly situated near SW 20th Avenue and NW 39th Avenue in 2030 (Figure 4-2). The commercial-institutional-transportation development is all located inside the 1,000 meters of urban buffer areas in 2030, where their sites are on the north side of the Gainesville Regional Airport, south of NE 39th Avenue, intersection of SW 24th Avenue and SW 75th Street, intersection of State Route 121 and SW 63rd Boulevard, north of SW 63rd Avenue, intersection of NW 43rd Street and State Route 441, and intersection of State Route 121 and County Route 231 (Figure 4-2). The industrial-warehouses development has 46 percent of land developed within 1,000 meters of the urban buffer areas in 2030. The major development outside 1,000 meters of the urban buffer areas for the industrial-warehouses use is in the Cadillac vicinity. From the above comparisons, it is apparent that the single-family development has the most prominent leapfrog development which constitutes urban sprawl because of the development's acreage and its locations. The other three urban uses are not as severe as the single-family use in terms of scattered and leapfrog developments, however.

Like the redevelopment scenario, the conservation scenario has also mixed-use development. The mixed-use in the conservation scenario is essentially the same as the redevelopment scenario, most of which are located inside a commercial-institutional-transportation use as well as along major transportation corridors. These major transportation corridors accountable for mixed-uses are NW 53rd Avenue, NE 39th Avenue, North Main Street, NE 31st Avenue, West Newberry Road, NW 23rd Avenue, SW 24th Avenue, SW 75th Street, and SW Archer Road (Figure 4-2). Same as the redevelopment scenario, the mixed-use land in the conservation scenario has a total of 687 acres in 2020 and a total of 2,029 acres in 2030. Table 4-45 illustrates the acreages of development that are located inside the 1,000 meters of urban buffer areas and outside based on 2020 and 2030, respectively. Tables 4-46 through 4-49 illustrate the acreages demanded and allocated for the four urban uses in 2010, 2020, and 2030, respectively, in terms of the conservation scenario.

Table 4-1. The baseline forecasting method for urban development acreages (single-family)

Baseline	Year	Single-Family Acres	Single-Family Population (People)	Gross Single-Family Density (People/Acre)	Changes (Acres)
Single-Family	2003	48,673	121,415	2.49	--
	2010	60,899	131,732	2.16	12,226
	2020	73,739	159,507	2.16	12,840
	2030	79,455	171,871	2.16	5,716

Table 4-2. The baseline forecasting method for urban development acreages (multi-family)

Baseline	Year	Multi-Family Acres	Multi-Family Population (People)	Gross Multi-Family Density (People/Acre)	Changes (Acres)
Multi-Family	2003	3,469	83,699	24.13	--
	2010	3,559	101,684	28.57	90
	2020	4,431	126,593	28.57	872
	2030	5,055	144,429	28.57	624

Table 4-3. The baseline forecasting method for urban development acreages (total residential)

Baseline	Year	Total Acres	Total Population (People)	Total Gross Residential Density (People/Acre)	Changes (Acres)
Total Residential	2003	52,142	205,114*	3.93	--
	2010	64,458	233,416*	3.62	12,317
	2020	78,170	286,100**	3.66	13,712
	2030	84,510	316,300**	3.74	6,340

(Source *: U.S. Census Bureau (1980, 1990, 2000, 2010);
**: Medium number (BEBR, March 2009))

Table 4-4. The baseline forecasting method for urban development acreages (commercial-institutional-transportation)

Baseline	Year	Total Acres	Commercial-Institutional-Transportation Employment (People)	Gross Commercial-Institutional-Transportation Density (People/Acre)	Changes (Acres)
Commercial-Institutional-Transportation	2003	22,010	68,042***	3.09	--
	2010	22,167	68,497	3.09	157
	2020	27,891	86,183	3.09	5,724
	2030	32,530	100,518	3.09	4,639

(Source: ***: BEBR (2003))

Table 4-5. The baseline forecasting method for urban development acreages (industrial-warehouses)

Baseline	Year	Total Acres	Industrial-Warehouses Employment (People)	Gross Industrial-Warehouses Density (People/Acre)	Changes (Acres)
Industrial-Warehouses	2003	2,819	9,130***	3.24	--
	2010	2,823	9,145	3.24	4
	2020	3,576	11,586	3.24	753
	2030	4,170	13,511	3.24	594

(Source: ***: BEBR (2003))

Table 4-6. The baseline forecasting method for urban development acreages (total urban)

Baseline	Year	Total Acres	Total Population (People)	Total Gross Urban Density (People/Acre)	Changes (Acres)
Total Urban	2003	76,971	205,114*	2.66	--
	2010	89,448	233,416*	2.61	12,477
	2020	109,637	286,100**	2.61	20,189
	2030	121,210	316,300**	2.61	11,573

(Source *: U.S. Census Bureau (1980, 1990, 2000, 2010);
**: Medium number (BEBR, March 2009))

Table 4-7. The baseline forecasting method for urban acreage changes

Year	Single-Family		Multi-Family		Commercial-Institutional-Transportation		Industrial-Warehouses	
	Acres	Cells	Acres	Cells	Acres	Cells	Acres	Cells
2010	12,226	54,976	90	404	157	706	4	18
2020	12,840	57,738	872	3,920	5,724	25,737	753	3,388
2030	5,716	25,702	624	2,807	4,639	20,860	594	2,671
Total	30,782	138,415	1,586	7,131	10,520	47,303	1,351	6,077

Table 4-8. Baseline urban acreage changes with percentages

	Urban Acreage Changes (Acres)			Percentage Of Urban Acreage Change (%)		
Year	2010	2020	2030	2010	2020	2030
Single-Family	12,226	12,840	5,716	97.99%	63.60%	49.39%
Multi-Family	90	872	624	0.72%	4.32%	5.39%
Residential	12,317	13,712	6,340	98.71%	67.92%	54.78%
Commercial-Institutional-Transportation	157	5,724	4,639	1.26%	28.35%	40.08%
Industrial-Warehouses	4	753	594	0.03%	3.73%	5.13%
Total	12,478	20,189	11,573	100.00%	100.00%	100.00%

Table 4-9. The increased density forecasting method for urban development acreages (single-family)

Increased Density	Year	Single-Family Acres	Single-Family Population (People)	Gross Single-Family Density (People/Acre)	Changes (Acres)
Single-Family	2003	48,673	121,415 *	2.49	--
	2010	60,899	131,732	2.16	12,226
	2020	64,663	159,507	2.47	3,764
	2030	69,635	171,871	2.47	4,972

(Source *: U.S. Census Bureau (1980, 1990, 2000, 2010))

Table 4-10. The increased density forecasting method for urban development acreages (multi-family)

Increased Density	Year	Multi-Family Acres	Multi-Family Population (People)	Gross Multi-Family Density (People/Acre)	Changes (Acres)
Multi-Family	2003	3,469	83,699 *	24.13	--
	2010	3,559	101,684	28.57	90
	2020	3,815	126,593	33.18	256
	2030	4,358	144,429	33.14	543

(Source *: U.S. Census Bureau (1980, 1990, 2000, 2010))

Table 4-11. The increased density forecasting method for urban development acreages (total residential)

Increased Density	Year	Total Acres	Total Population (People)	Total Gross Residential Density (People/Acre)	Changes (Acres)
Total Residential	2003	52,142	205,114 *	3.93	--
	2010	64,458	233,416 *	3.62	12,317
	2020	68,478	286,100 **	4.18	4,020
	2030	73,993	316,300 **	4.27	5,515

(Source *: U.S. Census Bureau (1980, 1990, 2000, 2010);
**: Medium number (BEBR, March 2009))

Table 4-12. The increased density forecasting method for urban development acreages (commercial-institutional-transportation)

Increased Density	Year	Total Acres	Commercial-Institutional-Transportation Employment (People)	Gross Commercial-Institutional-Transportation Density (People/Acre)	Changes (Acres)
Commercial-Institutional-Transportation	2003	22,010	68,042 ***	3.09	--
	2010	22,167	68,497	3.09	157
	2020	23,845	86,183	3.61	1,678
	2030	27,881	100,518	3.61	4,035

(Source: ***: BEBR (2003))

Table 4-13. The increased density forecasting method for urban development acreages (industrial-warehouses)

Increased Density	Year	Total Acres	Industrial-Warehouses Employment (People)	Gross Industrial-Warehouses Density (People/Acre)	Changes (Acres)
Industrial-Warehouses	2003	2,819	9,130 ***	3.24	--
	2010	2,823	9,145	3.24	4
	2020	3,043	11,586	3.81	221
	2030	3,560	13,511	3.80	517

(Source: ***: BEBR (2003))

Table 4-14. The increased density forecasting method for urban development acreages (total urban)

Increased Density	Year	Total Acres	Total Population (People)	Total Gross Urban Density (People/Acre)	Changes (Acres)
Total Urban	2003	76,971	205,114 *	2.66	--
	2010	89,448	233,416 *	2.61	12,477
	2020	95,367	286,100 **	3.00	5,919
	2030	105,433	316,300 **	3.00	10,067

(Source *: U.S. Census Bureau (1980, 1990, 2000, 2010);
**: Medium number (BEBR, March 2009))

Table 4-15. The increased density forecasting method for urban acreage changes

Year	Single-Family		Multi-Family		Commercial-Institutional-Transportation		Industrial-Warehouses	
	Acres	Cells	Acres	Cells	Acres	Cells	Acres	Cells
2010	12,226	54,976	90	404	157	706	4	18
2020	3,764	16,924	256	1,151	1,678	7,545	221	993
2030	4,972	22,356	543	2,442	4,035	18,145	517	2,323
Total	20,962	94,257	889	3,997	5,870	26,396	742	3,335

Table 4-16. The redevelopment forecasting method for urban development acreages (single-family)

Redevelopment	Year	Single-Family Acres	Single-Family Population (People)	Gross Single-Family Density (People/Acre)	Changes (Acres)	After 15% Redevelopment
Single-Family	2003	48,673	121,415 *	2.49	7,448	--
	2010	60,899	131,732 *	2.16	12,226	12,226
	2020	64,098	159,507	2.49	3,764	3,199
	2030	68,324	171,871	2.52	4,972	4,226

(Source *: U.S. Census Bureau (1980, 1990, 2000, 2010))

Table 4-17. The redevelopment forecasting method for urban development acreages (multi-family)

Redevelopment	Year	Multi-Family Acres	Multi-Family Population	Gross Multi-Family Density (People/Acre)	Changes (Acres)	After 15% Redevelopment
Multi-Family	2003	3,469	83,699 *	24.13	--	--
	2010	3,559	101,684 *	28.57	90	90
	2020	3,776	126,593	33.52	256	218
	2030	4,238	144,429	34.08	543	462

(Source *: U.S. Census Bureau (1980, 1990, 2000, 2010))

Table 4-18. The redevelopment forecasting method for urban development acreages (total residential)

Redevelopment	Year	Total Acres	Total Population (People)	Total Gross Residential Density (People/Acre)	Changes (Acres)	After 15% Redevelopment
Total Residential	2003	52,142	205,114 *	3.93	--	--
	2010	64,458	233,416 *	3.62	12,317	12,317
	2020	67,875	286,100 **	4.22	4,020	3,417
	2030	72,562	316,300 **	4.36	5,515	4,688

(Source *: U.S. Census Bureau (1980, 1990, 2000, 2010);
**: Medium number (BEBR, March 2009))

Table 4-19. The redevelopment forecasting method for urban development acreages (commercial-institutional-transportation)

Redevelopment	Year	Total Acres	Commercial-Institutional-Transportation Employment (People)	Gross Commercial-Institutional-Transportation Density (People/Acre)	Changes (Acres)	After 15% Redevelopment
Commercial-Institutional-Transportation	2003	22,010	68,042 ***	3.09	--	--
	2010	22,167	68,497	3.09	157	157
	2020	23,594	86,183	3.65	1,678	1,426
	2030	27,024	100,518	3.72	4,035	3,430

(Source ***: BEBR (2003))

Table 4-20. The redevelopment forecasting method for urban development acreages (industrial-warehouses)

Redevelopment	Year	Total Acres	Industrial-Warehouses Employment (People)	Gross Industrial-Warehouses Density (People/Acre)	Changes (Acres)	After 15% Redevelopment
Industrial-Warehouses	2003	2,819	9,130 ***	3.24	--	--
	2010	2,823	9,145	3.24	4	4
	2020	3,010	11,586	3.85	221	188
	2030	3,449	13,511	3.92	517	439

(Source ***: BEBR (2003))

Table 4-21. The redevelopment forecasting method for urban development acreages (total urban)

Redevelopment	Year	Total Acres	Total Population (People)	Total Gross Urban Density (People/Acre)	Changes (Acres)	After 15% Redevelopment
Total Urban	2003	76,971	205,114*	2.66	8,650	8,650
	2010	89,448	233,416 *	2.61	12,477	12,477
	2020	94,479	286,100 **	3.03	5,919	5,031
	2030	103,036	316,300 **	3.07	10,067	8,557

(Source *: U.S. Census Bureau (1980, 1990, 2000, 2010);
**: Medium number (BEBR, March 2009))

Table 4-22. The redevelopment forecasting method for urban acreage changes

Year	Single-Family		Multi-Family		Commercial-Institutional-Transportation		Industrial-Warehouses	
	Acres	Cells	Acres	Cells	Acres	Cells	Acres	Cells
2010	12,226	54,976	90	404	157	706	4	18
2020	3,199	14,386	218	978	1,426	6,413	188	844
2030	4,226	19,003	462	2,075	3,430	15,423	439	1,975
Total	19,651	88,365	769	3,458	5,013	22,542	631	2,837

Table 4-23. Conflict scores and equivalent descriptions

Conflict Score	Conflict Score Descriptions
4444	all high preference all in conflict
4443	high preference single-family conflicts with high preference multi-family and high preference commercial-institutional-transportation
4442	high preference single-family conflicts with high preference multi-family and high preference commercial-institutional-transportation
4441	high preference single-family conflicts with high preference multi-family and high preference commercial-institutional-transportation
4434	high preference single-family conflicts with high preference multi-family and high preference industrial-warehouses
4424	high preference single-family conflicts with high preference multi-family and high preference industrial-warehouses
4414	high preference single-family conflicts with high preference multi-family and high preference industrial-warehouses
4144	high preference single-family conflicts with high preference commercial-institutional-transportation and high preference industrial-warehouses
4244	high preference single-family conflicts with high preference commercial-institutional-transportation and high preference industrial-warehouses
4344	high preference single-family conflicts with high preference commercial-institutional-transportation and high preference industrial-warehouses
4433	high preference single-family conflicts with high preference multi-family
4432	high preference single-family conflicts with high preference multi-family
4431	high preference single-family conflicts with high preference multi-family
4423	high preference single-family conflicts with high preference multi-family
4422	high preference single-family conflicts with high preference multi-family
4421	high preference single-family conflicts with high preference multi-family
4413	high preference single-family conflicts with high preference multi-family
4412	high preference single-family conflicts with high preference multi-family
4411	high preference single-family conflicts with high preference multi-family
4341	high preference single-family conflicts with high preference commercial-institutional-transportation
4241	high preference single-family conflicts with high preference commercial-institutional-transportation
4141	high preference single-family conflicts with high preference commercial-institutional-transportation
4341	high preference single-family conflicts with high preference commercial-institutional-transportation
4342	high preference single-family conflicts with high preference commercial-institutional-transportation
4343	high preference single-family conflicts with high preference commercial-institutional-transportation
4334	high preference single-family conflicts with high preference commercial-industrial-warehouses

Table 4-23. Continued

Conflict Score	Conflict Score Descriptions
4324	high preference single-family conflicts with high preference commercial-industrial-warehouses
4314	high preference single-family conflicts with high preference commercial-industrial-warehouses
4234	high preference single-family conflicts with high preference commercial-industrial-warehouses
4224	high preference single-family conflicts with high preference commercial-industrial-warehouses
4214	high preference single-family conflicts with high preference commercial-industrial-warehouses
4134	high preference single-family conflicts with high preference commercial-industrial-warehouses
4124	high preference single-family conflicts with high preference commercial-industrial-warehouses
4114	high preference single-family conflicts with high preference commercial-industrial-warehouses
4333	single-family preference dominates
4332	single-family preference dominates
4331	single-family preference dominates
4323	single-family preference dominates
4322	single-family preference dominates
4321	single-family preference dominates
4313	single-family preference dominates
4312	single-family preference dominates
4311	single-family preference dominates
4211	single-family preference dominates
4212	single-family preference dominates
4213	single-family preference dominates
4221	single-family preference dominates
4222	single-family preference dominates
4223	single-family preference dominates
4231	single-family preference dominates
4232	single-family preference dominates
4233	single-family preference dominates
4111	single-family preference dominates
4112	single-family preference dominates
4113	single-family preference dominates
4121	single-family preference dominates
4122	single-family preference dominates
4123	single-family preference dominates
4131	single-family preference dominates

Table 4-23. Continued

Conflict Score	Conflict Score Descriptions
4132	single-family preference dominates
4133	single-family preference dominates
3333	all moderate preference all in conflict
3444	high preference multi-family conflicts with high preference commercial-institutional-transportation and high preference industrial-warehouses
3441	high preference multi-family conflicts with high preference commercial-institutional-transportation
3442	high preference multi-family conflicts with high preference commercial-institutional-transportation
3443	high preference multi-family conflicts with high preference commercial-institutional-transportation
3414	high preference multi-family conflicts with high preference industrial-warehouses
3424	high preference multi-family conflicts with high preference industrial-warehouses
3434	high preference multi-family conflicts with high preference industrial-warehouses
3144	high preference commercial-institutional-transportation conflicts with high preference industrial-warehouses
3244	high preference commercial-institutional-transportation conflicts with high preference industrial-warehouses
3344	high preference commercial-institutional-transportation conflicts with high preference industrial-warehouses
3411	multi-family preference dominates
3412	multi-family preference dominates
3413	multi-family preference dominates
3421	multi-family preference dominates
3422	multi-family preference dominates
3423	multi-family preference dominates
3431	multi-family preference dominates
3432	multi-family preference dominates
3433	multi-family preference dominates
3141	commercial-institutional-transportation preference dominates
3241	commercial-institutional-transportation preference dominates
3341	commercial-institutional-transportation preference dominates
3142	commercial-institutional-transportation preference dominates
3242	commercial-institutional-transportation preference dominates
3342	commercial-institutional-transportation preference dominates
3143	commercial-institutional-transportation preference dominates
3243	commercial-institutional-transportation preference dominates
3343	commercial-institutional-transportation preference dominates
3114	industrial-warehouses preference dominates
3214	industrial-warehouses preference dominates
3314	industrial-warehouses preference dominates

Table 4-23. Continued

Conflict Score	Conflict Score Descriptions
3124	industrial-warehouses preference dominates
3134	industrial-warehouses preference dominates
3224	industrial-warehouses preference dominates
3234	industrial-warehouses preference dominates
3324	industrial-warehouses preference dominates
3334	industrial-warehouses preference dominates
3332	moderate preference single-family conflicts with moderate preference multi-family and moderate preference commercial-institutional-transportation
3331	moderate preference single-family conflicts with moderate preference multi-family and moderate preference commercial-institutional-transportation
3313	moderate preference single-family conflicts with moderate preference multi-family and moderate preference industrial-warehouses
3323	moderate preference single-family conflicts with moderate preference multi-family and moderate preference industrial-warehouses
3343	commercial-institutional-transportation preference dominates
3133	moderate preference single-family conflicts with moderate preference commercial-institutional-transportation and moderate preference industrial-warehouses
3233	moderate preference single-family conflicts with moderate preference commercial-institutional-transportation and moderate preference industrial-warehouses
3433	multi-family preference dominates
2222	all moderate preference all in conflict
2444	high preference multi-family conflicts with high preference commercial-institutional-transportation and high preference industrial-warehouses
2441	high preference multi-family conflicts with high preference commercial-institutional-transportation
2442	high preference multi-family conflicts with high preference commercial-institutional-transportation
2443	high preference multi-family conflicts with high preference commercial-institutional-transportation
2414	high preference multi-family conflicts with high preference industrial-warehouses
2424	high preference multi-family conflicts with high preference industrial-warehouses
2434	high preference multi-family conflicts with high preference industrial-warehouses
2144	high preference commercial-institutional-transportation conflicts with high preference industrial-warehouses
2244	high preference commercial-institutional-transportation conflicts with high preference industrial-warehouses
2344	high preference commercial-institutional-transportation conflicts with high preference industrial-warehouses
2411	multi-family preference dominates
2412	multi-family preference dominates
2413	multi-family preference dominates
2421	multi-family preference dominates

Table 4-23. Continued

Conflict Score	Conflict Score Descriptions
2422	multi-family preference dominates
2423	multi-family preference dominates
2431	multi-family preference dominates
2432	multi-family preference dominates
2433	multi-family preference dominates
2141	commercial-institutional-transportation preference dominates
2241	commercial-institutional-transportation preference dominates
2341	commercial-institutional-transportation preference dominates
2142	commercial-institutional-transportation preference dominates
2242	commercial-institutional-transportation preference dominates
2342	commercial-institutional-transportation preference dominates
2143	commercial-institutional-transportation preference dominates
2243	commercial-institutional-transportation preference dominates
2343	commercial-institutional-transportation preference dominates
2114	industrial-warehouses preference dominates
2214	industrial-warehouses preference dominates
2314	industrial-warehouses preference dominates
2124	industrial-warehouses preference dominates
2134	industrial-warehouses preference dominates
2224	industrial-warehouses preference dominates
2234	industrial-warehouses preference dominates
2324	industrial-warehouses preference dominates
2334	industrial-warehouses preference dominates
2223	industrial-warehouses preference dominates
2221	moderate preference single-family conflicts with moderate preference multi-family and moderate preference commercial-institutional-transportation
2212	moderate preference single-family conflicts with moderate preference multi-family and moderate preference industrial-warehouses
2232	commercial-institutional-transportation preference dominates
2242	commercial-institutional-transportation preference dominates
2122	moderate preference single-family conflicts with moderate preference commercial-institutional-transportation and moderate preference industrial-warehouses
2322	multi-family preference dominates
2422	multi-family preference dominates
1111	all low preference all in conflict
1444	high preference multi-family conflicts with high preference commercial-institutional-transportation and high preference industrial-warehouses
1441	high preference multi-family conflicts with high preference commercial-institutional-transportation
1442	high preference multi-family conflicts with high preference commercial-institutional-transportation

Table 4-23. Continued

Conflict Score	Conflict Score Descriptions
1443	high preference multi-family conflicts with high preference commercial-institutional-transportation
1414	high preference multi-family conflicts with high preference industrial-warehouses
1424	high preference multi-family conflicts with high preference industrial-warehouses
1434	high preference multi-family conflicts with high preference industrial-warehouses
1144	high preference commercial-institutional-transportation conflicts with high preference industrial-warehouses
1244	high preference commercial-institutional-transportation conflicts with high preference industrial-warehouses
1344	high preference commercial-institutional-transportation conflicts with high preference industrial-warehouses
1411	multi-family preference dominates
1412	multi-family preference dominates
1413	multi-family preference dominates
1421	multi-family preference dominates
1422	multi-family preference dominates
1423	multi-family preference dominates
1431	multi-family preference dominates
1432	multi-family preference dominates
1433	multi-family preference dominates
1141	commercial-institutional-transportation preference dominates
1241	commercial-institutional-transportation preference dominates
1341	commercial-institutional-transportation preference dominates
1142	commercial-institutional-transportation preference dominates
1242	commercial-institutional-transportation preference dominates
1342	commercial-institutional-transportation preference dominates
1143	commercial-institutional-transportation preference dominates
1243	commercial-institutional-transportation preference dominates
1343	commercial-institutional-transportation preference dominates
1114	industrial-warehouses preference dominates
1214	industrial-warehouses preference dominates
1314	industrial-warehouses preference dominates
1124	industrial-warehouses preference dominates
1134	industrial-warehouses preference dominates
1224	industrial-warehouses preference dominates
1234	industrial-warehouses preference dominates
1324	industrial-warehouses preference dominates
1334	industrial-warehouses preference dominates
1112	industrial-warehouses preference dominates
1113	industrial-warehouses preference dominates

Table 4-23. Continued

Conflict Score	Conflict Score Description
1121	commercial-institutional-transportation preference dominates
1131	commercial-institutional-transportation preference dominates
1141	commercial-institutional-transportation preference dominates
1211	multi-family preference dominates
1311	multi-family preference dominates
1411	multi-family preference dominates

Table 4-24. Conflict scores for the BAU scenario

	Conflict Scores	Cells	Acres
Single-Family Dominates	2111	104,976	23,346
	3111	39,501	8,785
	3112	115	26
	3121	2	0
	3211	272	60
	4111	12,398	2,757
	4112	67	15
	4211	45	10
	Subtotal	157,376	34,999
Multi-Family Dominates	2311	12	3
	1411	96	21
	1311	495	110
	1211	5,986	1,331
	Subtotal	6,594	1,466
Commercial-Institutional-Transportation Dominates	1231	3	1
	1141	11	2
	1132	9	2
	1131	474	105
	1121	220	49
	Subtotal	717	159
Industrial-Warehouses Dominates	2114	1	0
	1214	2	0
	1213	2	0
	1114	410	91
	1113	662	147
	1112	17,660	3,927
	Subtotal	18,737	4,167

Table 4-24. Continued

	Conflict Scores	Cells	Acres
Minor Conflict	2212	1	0
	2211	1,762	392
	2121	4	1
	2112	804	179
	1221	11	2
	1212	76	17
	1122	3	1
	Subtotal	2,661	592
Major Conflict	1111	1,563,159	347,631
	Subtotal	1,563,159	347,631
	Total	1,749,244	389,014

Table 4-25. Conflict scores for the infill scenario

	Conflict Scores	Cells	Acres
Single-Family Dominates	4311	10	2
	4211	22	5
	4112	52	12
	4111	9,956	2,214
	3211	154	34
	3112	96	21
	3111	34,088	7,581
	2111	75,787	16,854
	Subtotal	120,165	26,723
Multi-Family Dominates	2411	7	2
	2311	205	46
	1412	5	1
	1411	305	68
	1312	36	8
	1311	1,892	421
	1211	1,679	373
	Subtotal	4,129	918
Commercial-Institutional-Transportation	1121	325	72
Dominates	Subtotal	325	72
Industrial-Warehouses Dominates	2114	1	0
	1213	2	0
	1114	359	80
	1113	502	112
	1112	13,997	3,113
	Subtotal	14,861	3,305

Table 4-25. Continued

	Conflict Scores	Cells	Acres
Minor Conflict	3311	45	10
	2211	785	175
	2112	574	128
	1314	1	0
	1221	2	0
	1212	10	2
	1122	25	6
	Subtotal	1,442	321
Major Conflict	1111	844,868	187,890
	Subtotal	844,868	187,890
	Total	985,790	219,230

Table 4-26. Single-family acreage demand and allocation in 2010, 2020, and 2030 for the BAU scenario

BAU Single-Family	2010		2020		2030	
	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	12,226	54,976	12,840	57,738	5,716	25,702
Allocated	12,226	54,976	12,853	57,796	5,740	25,812

Table 4-27. Multi-family acreage demand and allocation in 2010, 2020, and 2030 for the BAU scenario

BAU Multi-Family	2010		2020		2030	
	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	90	404	872	3,920	624	2,807
Allocated	90	404	991	4,455	666	2,996

Table 4-28. Commercial-institution-transportation acreage demand and allocation in 2010, 2020, and 2030 for the BAU scenario

BAU Commercial-	2010		2020		2030	
Institutional-Transportation	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	157	706	5,724	25,737	4,639	20,860
Allocated	157	706	5,710	25,675	4,841	21,767

Table 4-29. Industrial-warehouses acreage demand and allocation in 2010, 2020, and 2030 for the BAU scenario

BAU Industrial-Warehouses	2010		2020		2030	
	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	4	18	753	3,388	594	2,671
Allocated	4	18	825	3,710	797	3,584

Table 4-30. Infill scenario infill development acreages and conflict development acreages in 2020 and 2030

Infill Scenario	2020						2030					
	Within Infill Areas*		Within Conflict Areas **		Total		Within Infill Areas*		Within Conflict Areas **		Total	
	Acres	Cells	Acres	Cells	Acres	Cells	Acres	Cells	Acres	Cells	Acres	Cells
Single-Family Allocated	12,885	57,938	0	0	12,885	57,938	5,780	25,990	0	0	5,780	25,990
Multi-Family Allocated	55	247	828	3,723	883	3,970	0	0	835	3,755	835	3,755
Commercial-Institutional-Transportation Allocated	2,663	11,974	3,214	14,453	5,877	26,427	0	0	4,664	20,974	4,664	20,974
Industrial-Warehouses Allocated	759	3,414	0	0	759	3,414	707	3,181	0	0	707	3,181
Total	16,362	73,573	4,042	18,176	20,404	91,749	6,487	29,171	5,499	24,729	11,987	53,900

*: located either within 1,000 meters of urban buffer areas or outside
**: without overlapping with the infill areas

Table 4-31. Single-family acreage demand and allocation in 2010, 2020, and 2030 for the infill scenario

Infill Single-Family	2010		2020		2030	
	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	12,226	54,976	12,840	57,738	5,716	25,702
Allocated	12,226	54,976	12,885	57,938	5,780	25,990

Table 4-32. Multi-family acreage demand and allocation in 2010, 2020, and 2030 for the infill scenario

Infill Multi-Family	2010 Acres	2010 Cells	2020 Acres	2020 Cells	2030 Acres	2030 Cells
Demanded	90	404	872	3,920	624	2,807
Allocated	90	404	883	3,970	835	3,755

Table 4-33. Commercial-institution-transportation acreage demand and allocation in 2010, 2020, and 2030 for the infill scenario

Infill Commercial-Institutional-Transportation	2010 Acres	2010 Cells	2020 Acres	2020 Cells	2030 Acres	2030 Cells
Demanded	157	706	5,724	25,737	4,639	20,860
Allocated	157	706	5,877	26,427	4,664	20,974

Table 4-34. Industrial-warehouses acreage demand and allocation in 2010, 2020, and 2030 for the infill scenario

Infill Industrial-Warehouses	2010 Acres	2010 Cells	2020 Acres	2020 Cells	2030 Acres	2030 Cells
Demanded	4	18	753	3,388	594	2,671
Allocated	4	18	759	3,414	707	3,181

Table 4-35. Increased density development scenario development acreages and conflict development acreages in 2020 and 2030

Increased Density Development Scenario	2020						2030					
	Within infill areas*		Within conflict areas **		Total		Within infill areas*		Within conflict areas **		Total	
	Acres	Cells	Acres	Cells	Acres	Cells	Acres	Cells	Acres	Cells	Acres	Cells
Single-Family Allocated	3,859	17,353	0	0	3,859	17,353	5,322	23,929	0	0	5,322	23,929
Multi-Family Allocated	54	244	245	1,102	299	1,346	0	0	594	2,670	594	2,670
Commercial-Institutional-Transportation Allocated	1,678	7,546	78	352	1,756	7,898	318	1,428	3,787	17,029	4,105	18,457
Industrial-Warehouses Allocated	331	1,489	0	0	331	1,489	552	2,480	0	0	552	2,480
Total	5,922	26,632	323	1,454	6,245	28,086	6,192	27,837	4,381	19,699	10,573	47,536

*: located within 1,000 meters of urban buffer areas
**: without overlapping with the infill areas

Table 4-36. Single-family acreage demand and allocation in 2010, 2020, and 2030 for the increased density development scenario

Increased Density Single-Family	2010		2020		2030	
	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	12,226	54,976	3,764	16,924	4,972	22,356
Allocated	12,226	54,976	3,859	17,353	5,322	23,929

.

Table 4-37. Multi-family acreage demand and allocation in 2010, 2020, and 2030 for the increased density development scenario

Increased Density Multi-Family	2010		2020		2030	
	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	90	404	256	1,151	543	2,442
Allocated	90	404	299	1,346	594	2,670

Table 4-38. Commercial-institution-transportation acreage demand and allocation in 2010, 2020, and 2030 for the increased density development scenario

Increased Density Commercial-Institutional-Transportation	2010		2020		2030	
	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	157	706	1,678	7,545	4,035	18,145
Allocated	157	706	1,756	7,898	4,105	18,457

Table 4-39. Industrial-warehouses acreage demand and allocation in 2010, 2020, and 2030 for the increased density development scenario

Increased Density Industrial-Warehouses	2010		2020		2030	
	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	4	18	221	993	517	2,323
Allocated	4	18	331	1,489	552	2,480

Table 4-40. Redevelopment scenario development acreages and conflict development acreages in 2020 and 2030

Redevelopment Scenario	2020						2030					
	Within Infill Areas*		Within Conflict Areas **		Total		Within Infill Areas*		Within Conflict Areas **		Total	
	Acres	Cells	Acres	Cells	Acres	Cells	Acres	Cells	Acres	Cells	Acres	Cells
Single-Family Allocated	3,389	15,239	0	0	3,389	15,239	4,234	19,037	0	0	4,234	19,037
Multi-Family Allocated	54	244	245	1,101	299	1,345	0	0	455	2,047	455	2,047
Commercial-Institutional-Transportation Allocated	1,403	6,307	78	352	1,481	6,659	583	2,620	2,998	13,479	3,580	16,099
Industrial-Warehouses Allocated	232	1,042	0	0	232	1,042	435	1,955	0	0	435	1,955
Total	5,078	22,832	323	1,453	5,401	24,285	5,252	23,612	3,453	15,526	8,704	39,138

*: located within 1,000 meters of urban buffer areas

**: without overlapping with the infill areas

Table 4-41. Single-family acreage demand and allocation in 2010, 2020, and 2030 for the redevelopment scenario

Redevelopment	2010		2020		2030	
Single-Family	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	12,226	54,976	3,199	14,386	4,226	19,003
Allocated	12,226	54,976	3,389	15,239	4,234	19,037

Table 4-42. Multi-family acreage demand and allocation in 2010, 2020, and 2030 for the redevelopment scenario

Redevelopment Multi-	2010		2020		2030	
Family	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	90	404	218	978	462	2,075
Allocated	90	404	245	1,101	455	2,047

Table 4-43. Commercial-institution-transportation acreage demand and allocation in 2010, 2020, and 2030 for the redevelopment scenario

Redevelopment Commercial-Institutional-Transportation	2010		2020		2030	
	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	157	706	1,426	6,413	3,430	15,423
Allocated	157	706	1,481	6,659	3,580	16,099

Table 4-44. Industrial-warehouses acreage demand and allocation in 2010, 2020, and 2030 for the redevelopment scenario

Redevelopment Industrial-Warehouses	2010		2020		2030	
	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	4	18	188	844	439	1,975
Allocated	4	18	232	1,042	435	1,955

Table 4-45. Conservation scenario development acreages and conflict development acreages in 2020 and 2030

Conservation Scenario	2020						2030					
	Within Urban Buffer Areas		Outside Urban Buffer Areas		Total		Within Urban Buffer Areas		Outside Urban Buffer Areas		Total	
	Acres	Cells	Acres	Cells	Acres	Cells	Acres	Cells	Acres	Cells	Acres	Cells
Single-Family Allocated	2,670	12,006	685	3,078	3,355	15,084	2,423	10,897	1,847	8,305	4,270	19,202
Multi-Family Allocated	205	920	42	189	247	1,109	451	2,029	72	326	524	2,355
Commercial-Institutional-Transportation Allocated	1,540	6,926	8	37	1,549	6,963	3,544	15,936	0	0	3,544	15,936
Industrial-Warehouses Allocated	120	541	72	322	192	863	198	891	237	1,064	435	1,955
Total	4,535	20,393	807	3,626	5,343	24,019	6,616	29,753	2,156	9,695	8,773	39,448

Table 4-46. Single-family acreage demand and allocation in 2010, 2020, and 2030 for the conservation scenario

Conservation	2010		2020		2030	
Single-Family	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	12,226	54,976	3,199	14,386	4,226	19,003
Allocated	12,226	54,976	3,355	15,084	4,270	19,202

Table 4-47. Multi-family acreage demand and allocation in 2010, 2020, and 2030 for the conservation scenario

Conservation	2010		2020		2030	
Multi-Family	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	90	404	218	978	462	2,075
Allocated	90	404	247	1,109	524	2,355

Table 4-48. Commercial-institution-transportation acreage demand and allocation in 2010, 2020, and 2030 for the conservation scenario

Conservation	2010		2020		2030	
Commercial-Institutional-Transportation	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	157	706	1,426	6,413	3,430	15,423
Allocated	157	706	1,549	6,963	3,544	15,936

Table 4-49. Industrial-warehouses acreage demand and allocation in 2010, 2020, and 2030 for the conservation scenario

Conservation	2010		2020		2030	
Industrial-Warehouses	Acres	Cells	Acres	Cells	Acres	Cells
Demanded	4	18	188	844	439	1,975
Allocated	4	18	192	863	435	1,955

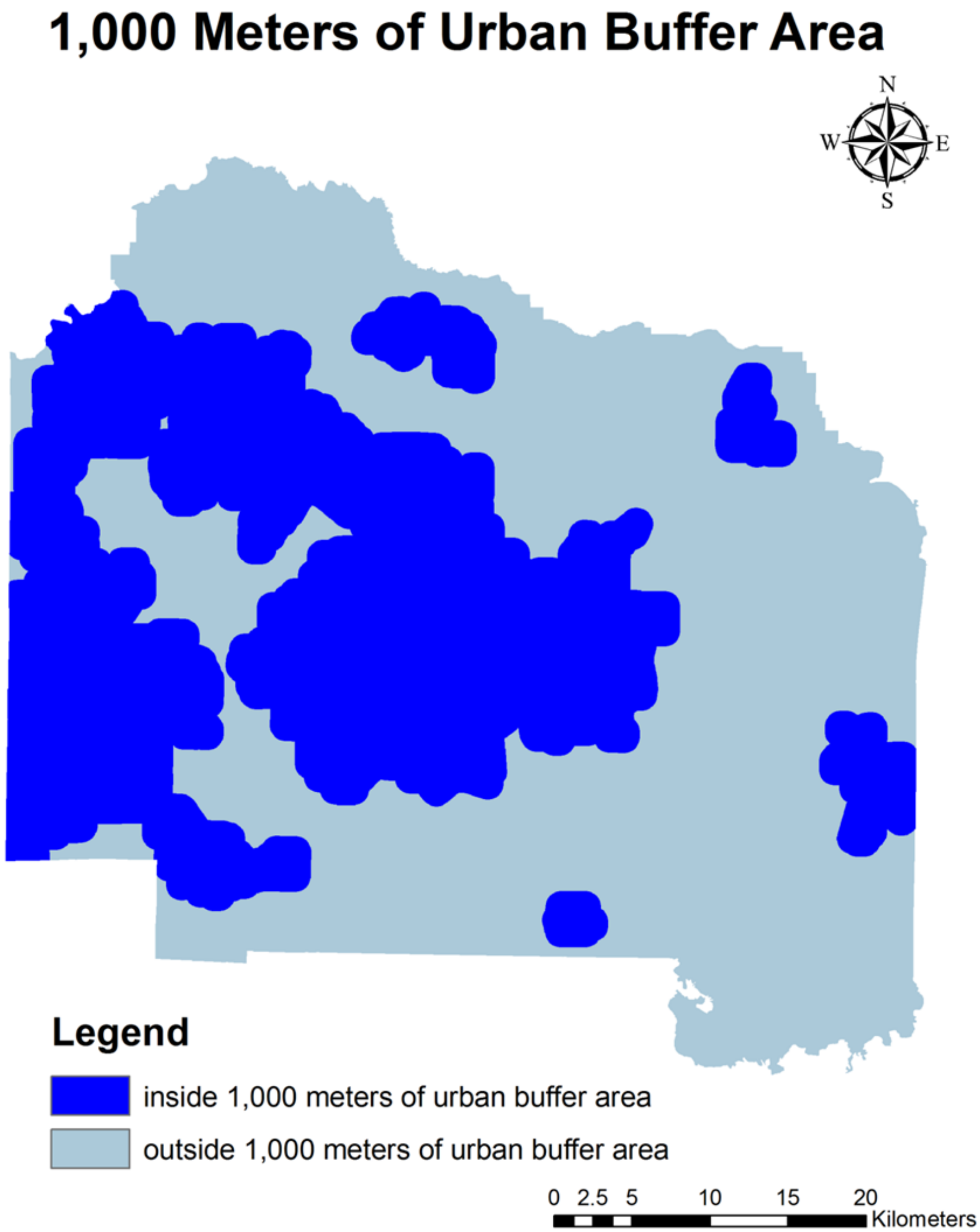

Figure 4-1. Areas within urban buffer areas and outside urban buffer areas (Raw data source: Alachua County Growth Management Department)

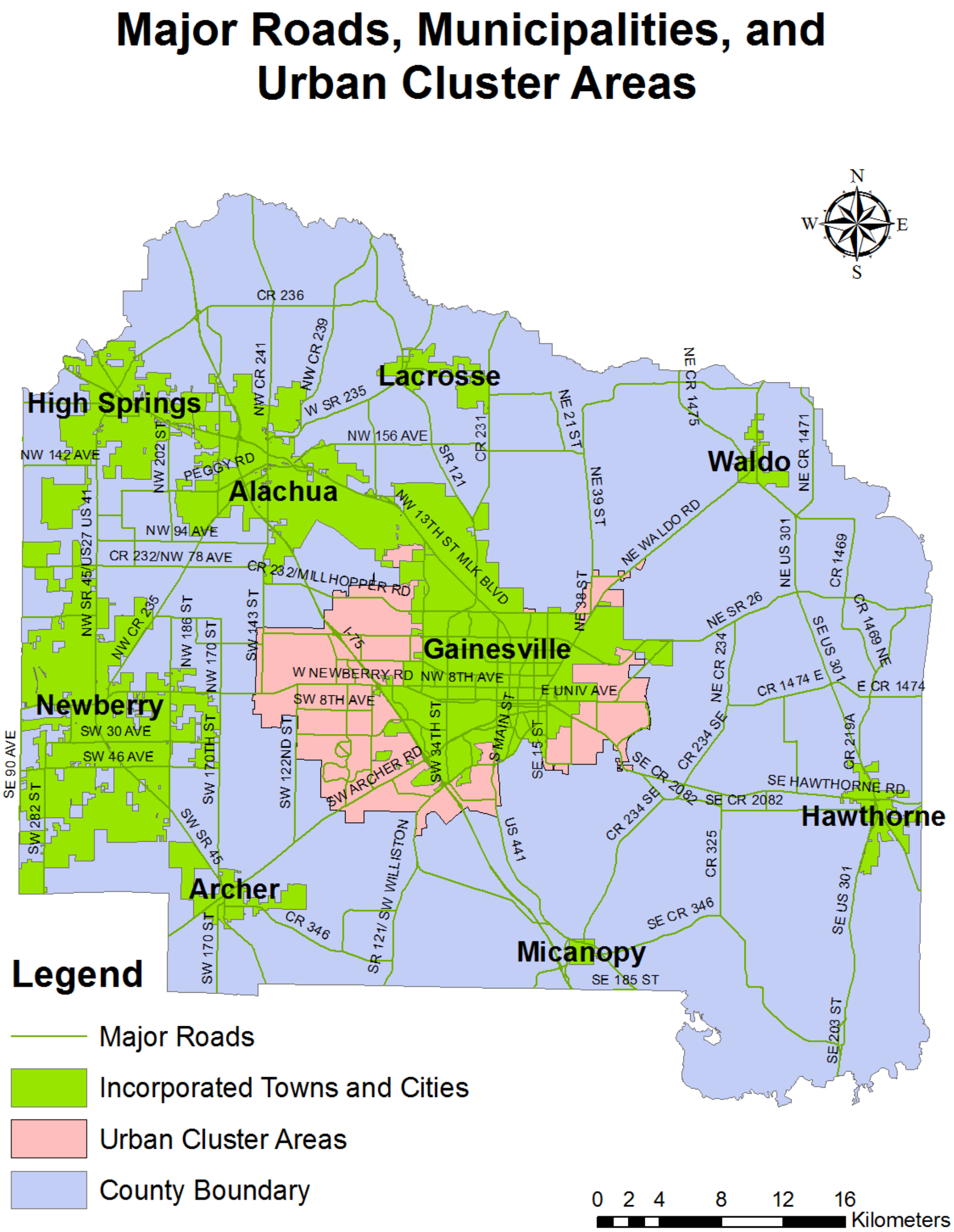

Figure 4-2. Major roads in Alachua County, incorporated towns and cities, and Urban Cluster areas (Raw data source: Alachua County Growth Management Department)

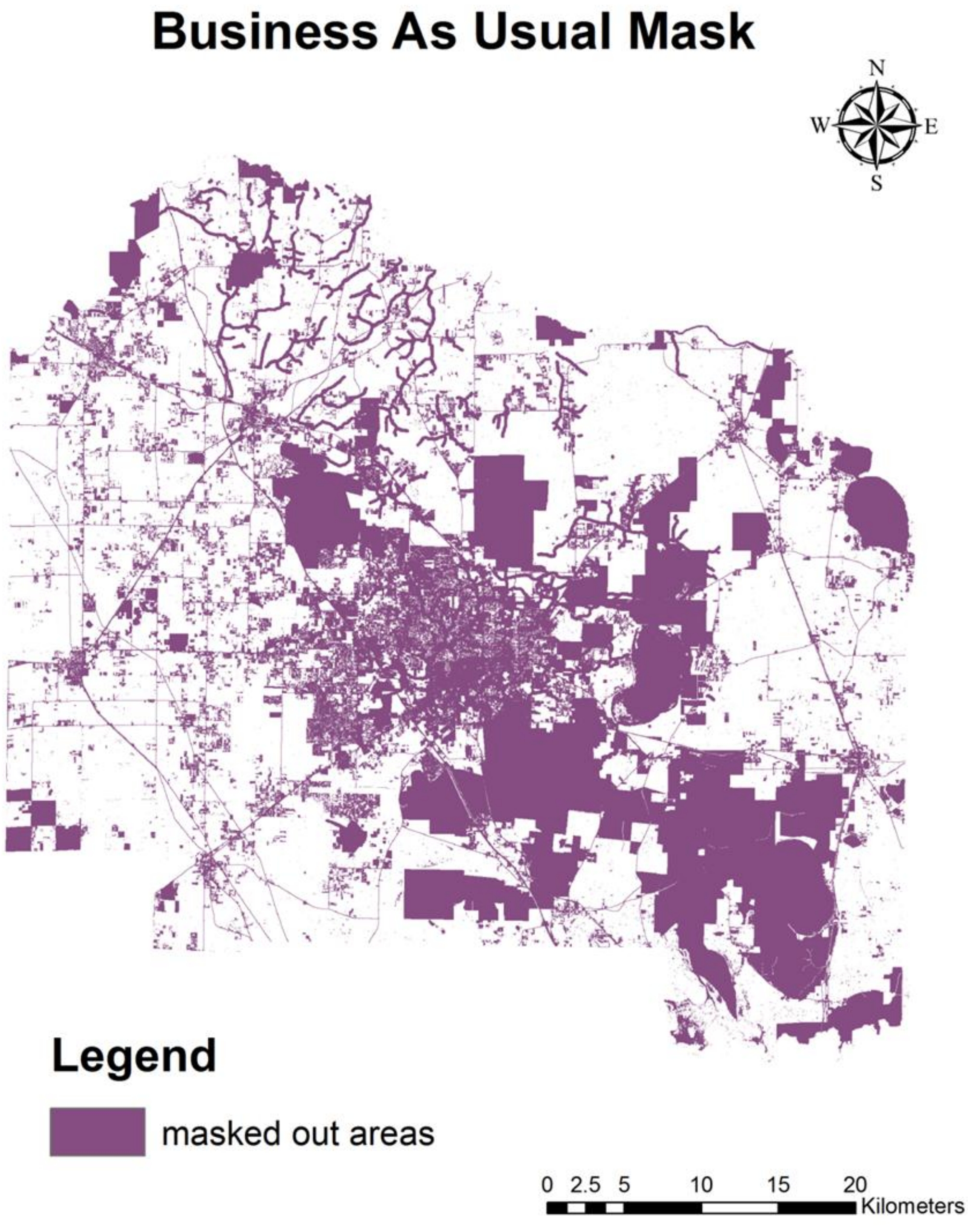

Figure 4-3. The BAU mask (Raw data source: Alachua County Growth Management Department)

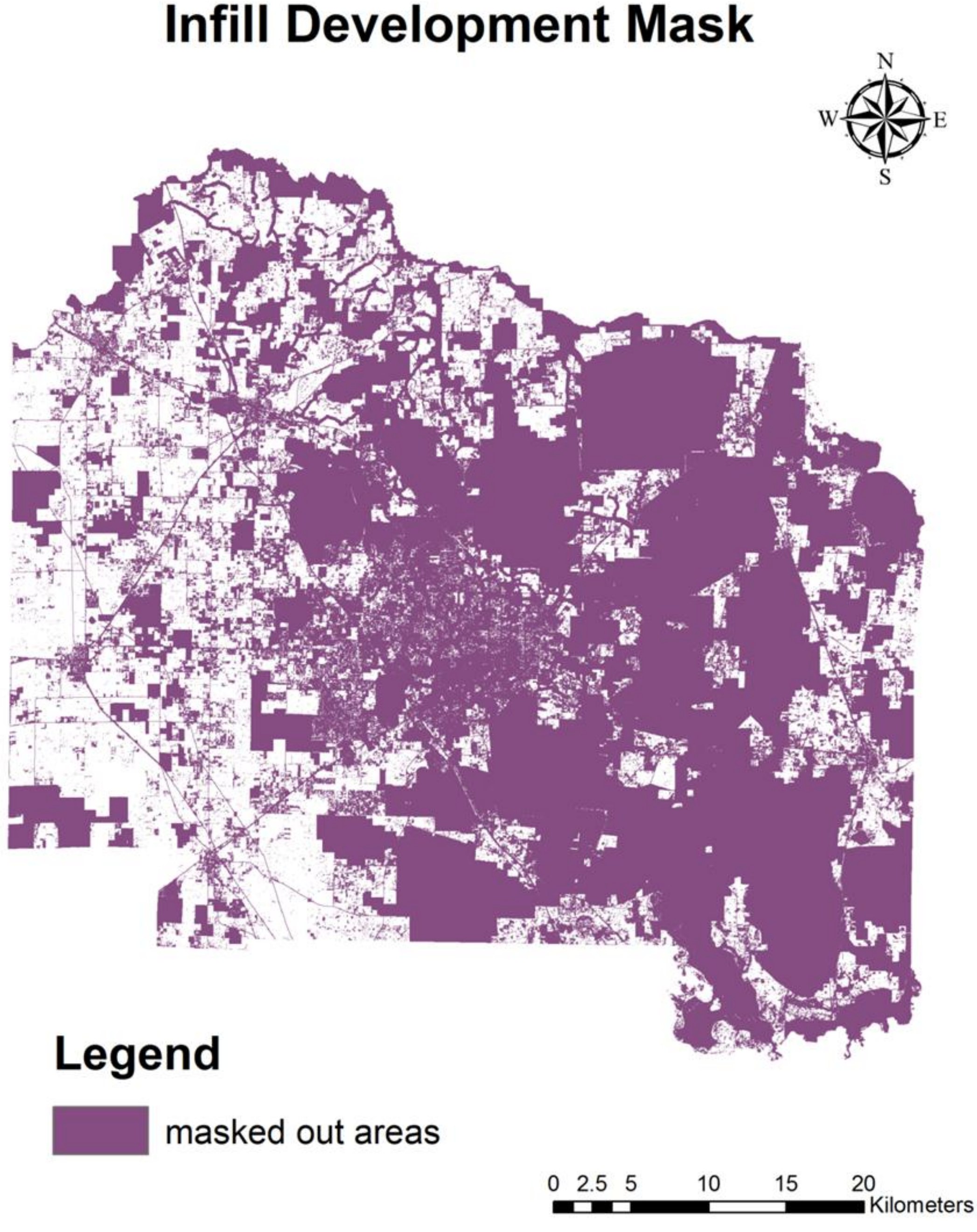

Figure 4-4. The infill development mask (Raw data source: Alachua County Growth Management Department)

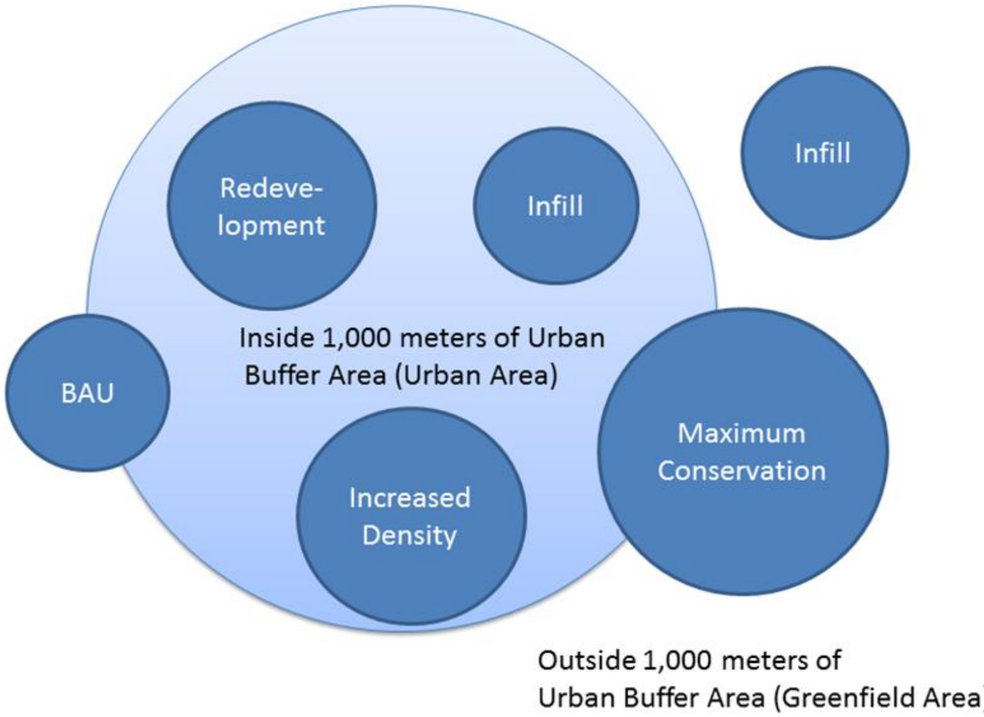

Figure 4-5. Five scenarios with their relationship with the urban area and the greenfield areas, where the urban area is within 1,000 meters of the urban buffer areas and the greenfield area is outside 1,000 meters of the urban buffer areas.

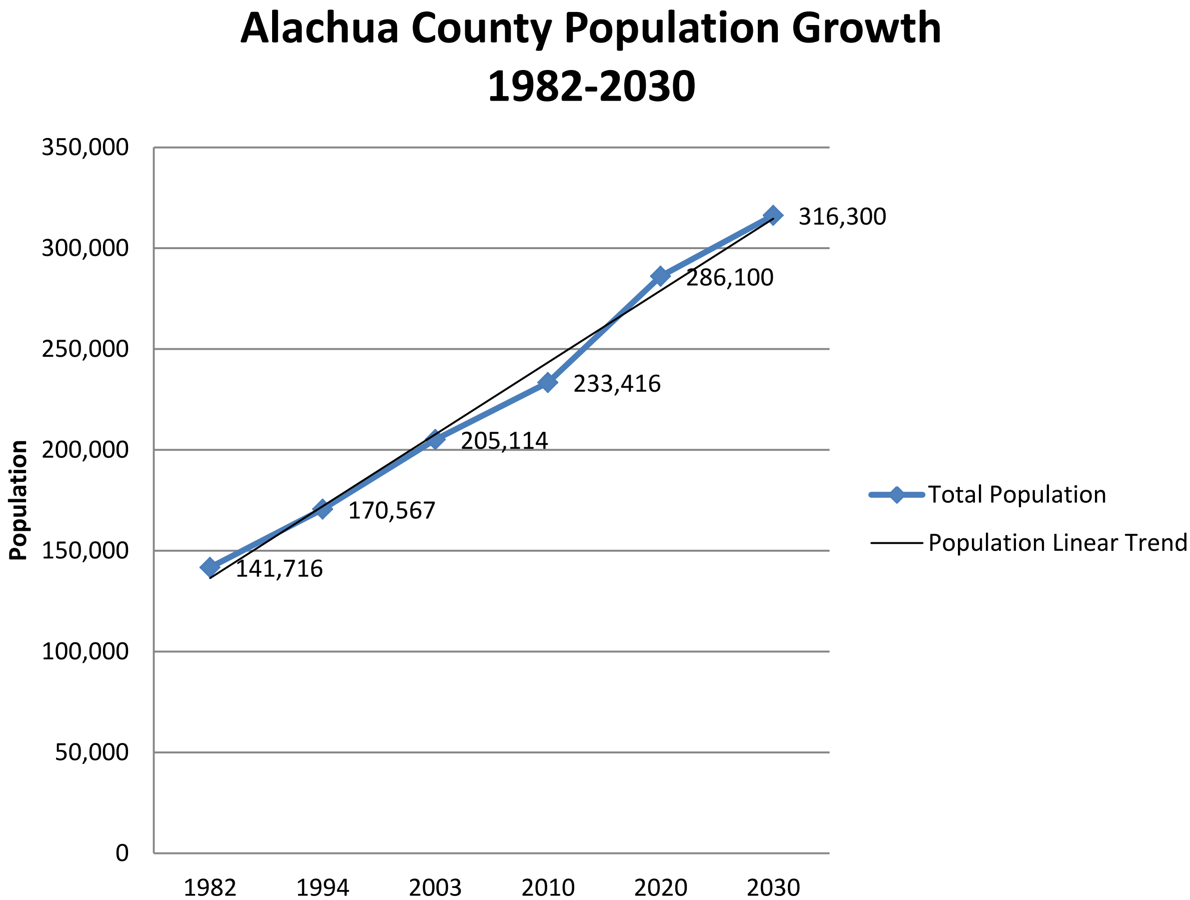

Figure 4-6. Population growth in Alachua County from 1982-2030. Considering the natural growth of county's population, student population increase is not included. (Source: U.S. Census Bureau (1980, 1990, 2000, 2010) and BEBR (2009, March))

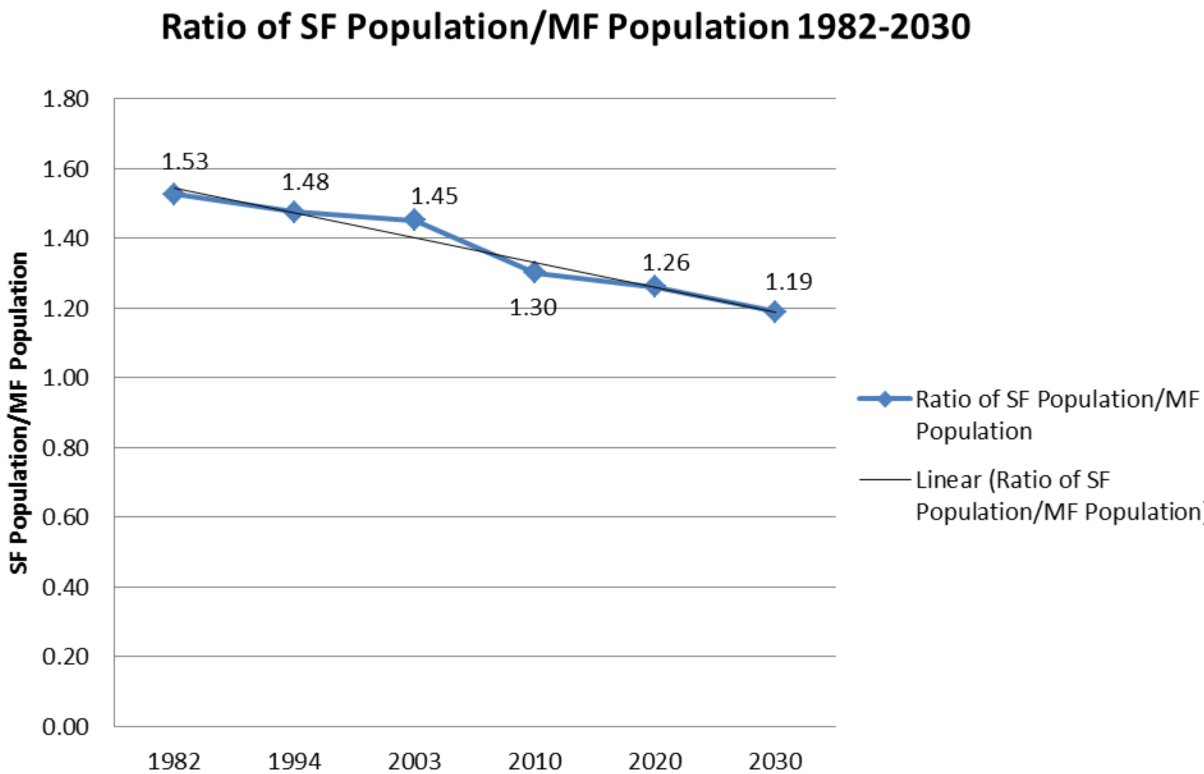

Figure 4-7. Ratio of single-family population/multi-family population from 1982-2030.

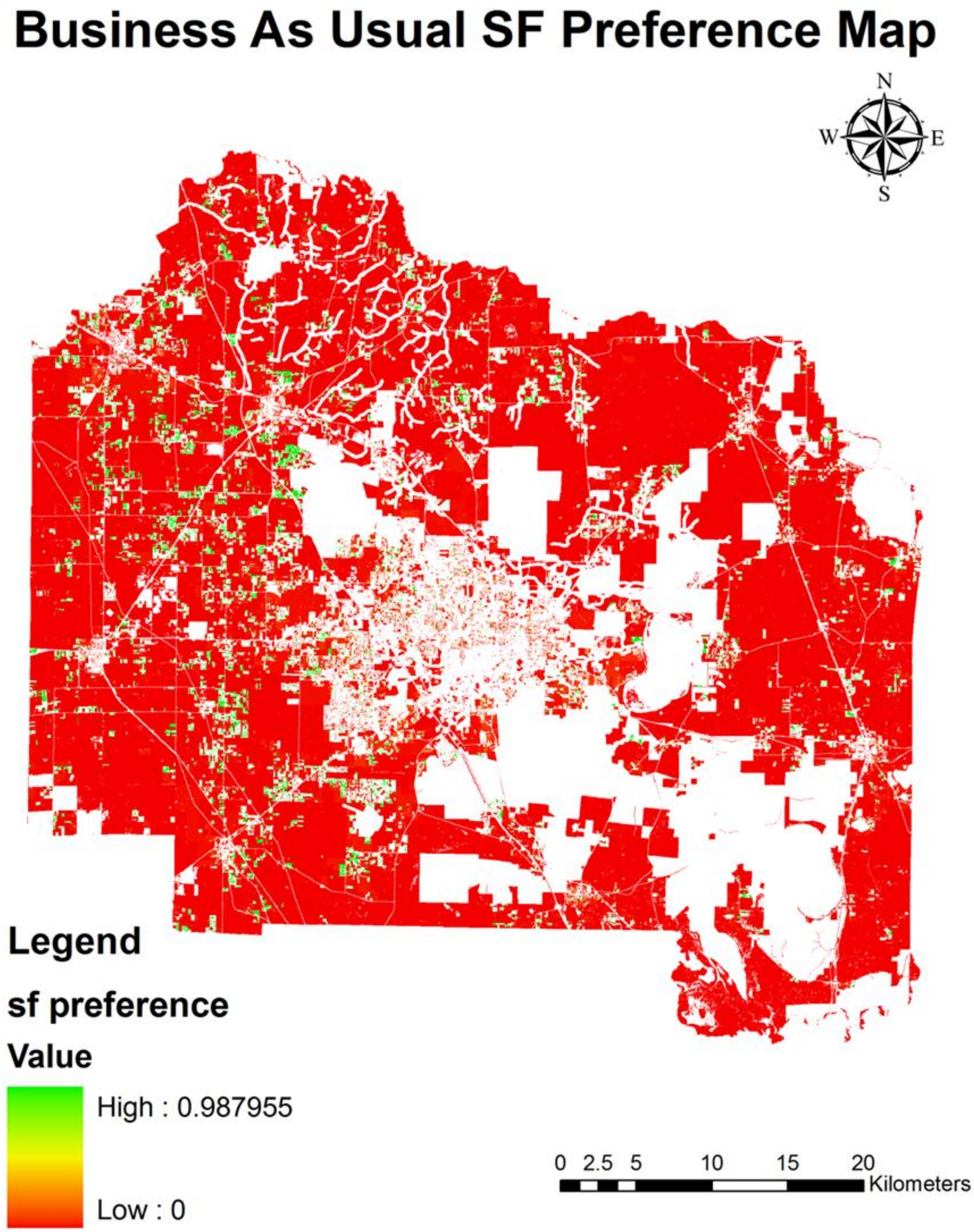

Figure 4-8. BAU preference map for single-family development

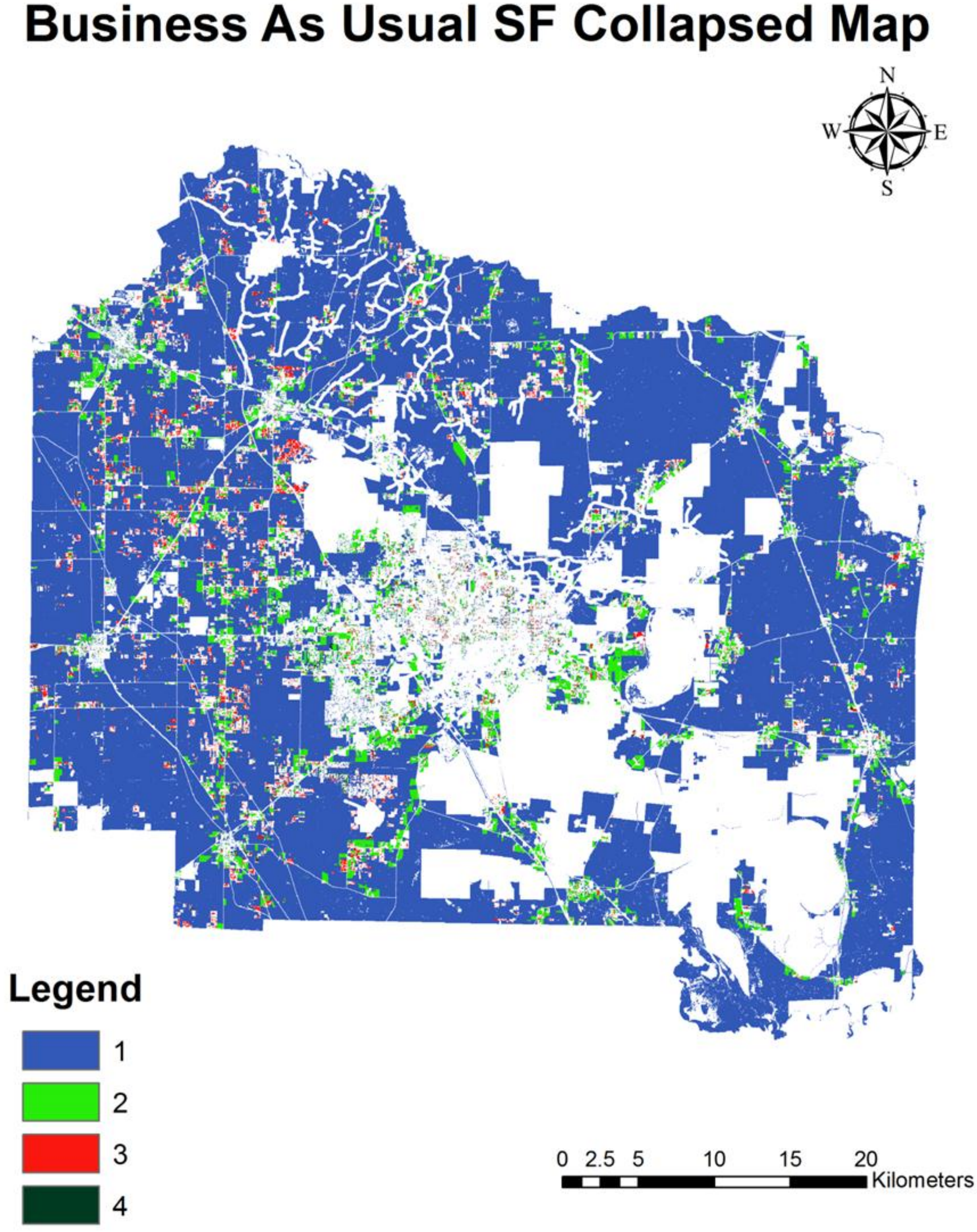

Figure 4-9. Collapsed map for single-family use for the BAU scenario

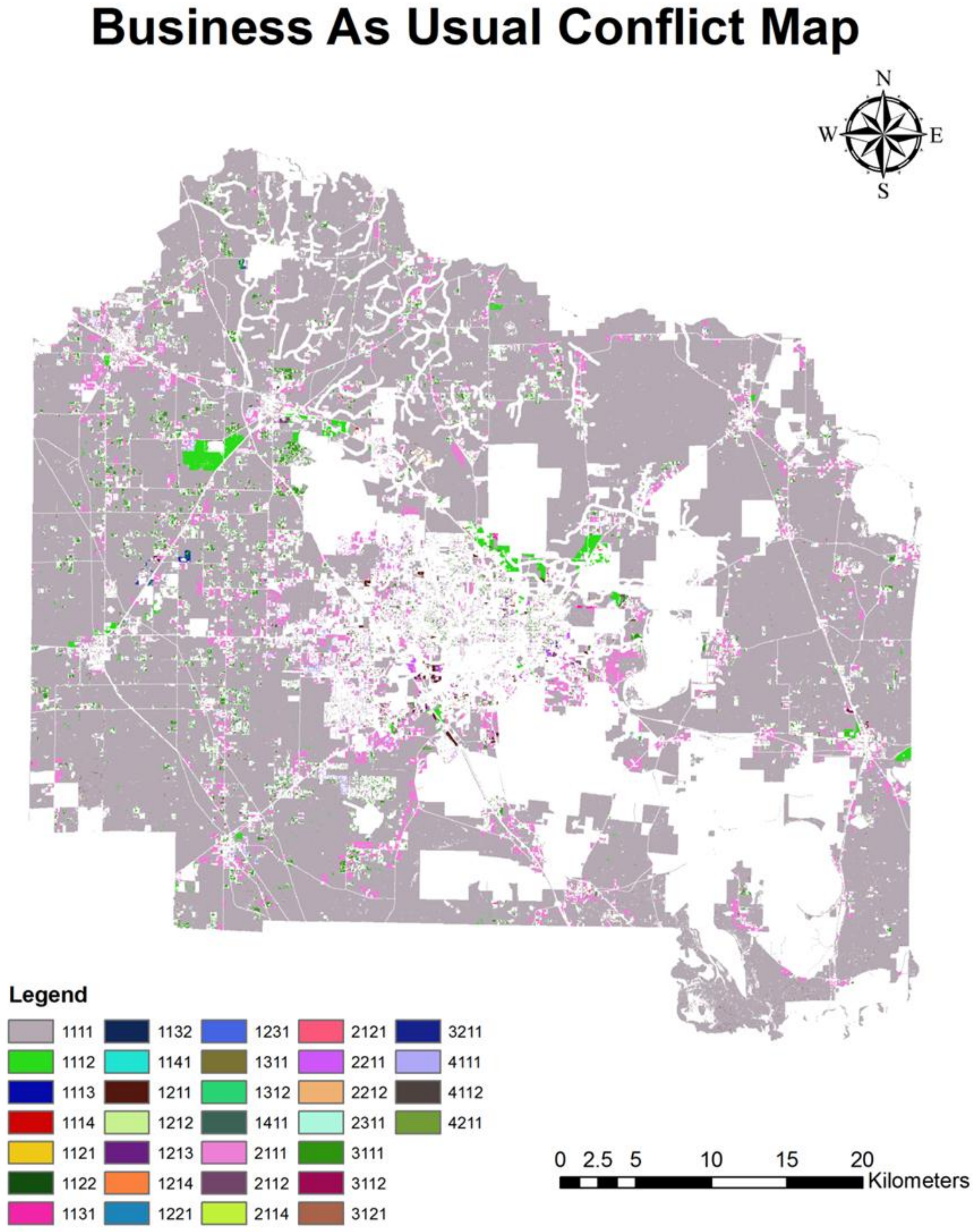

Figure 4-10. Conflict map for the BAU scenario

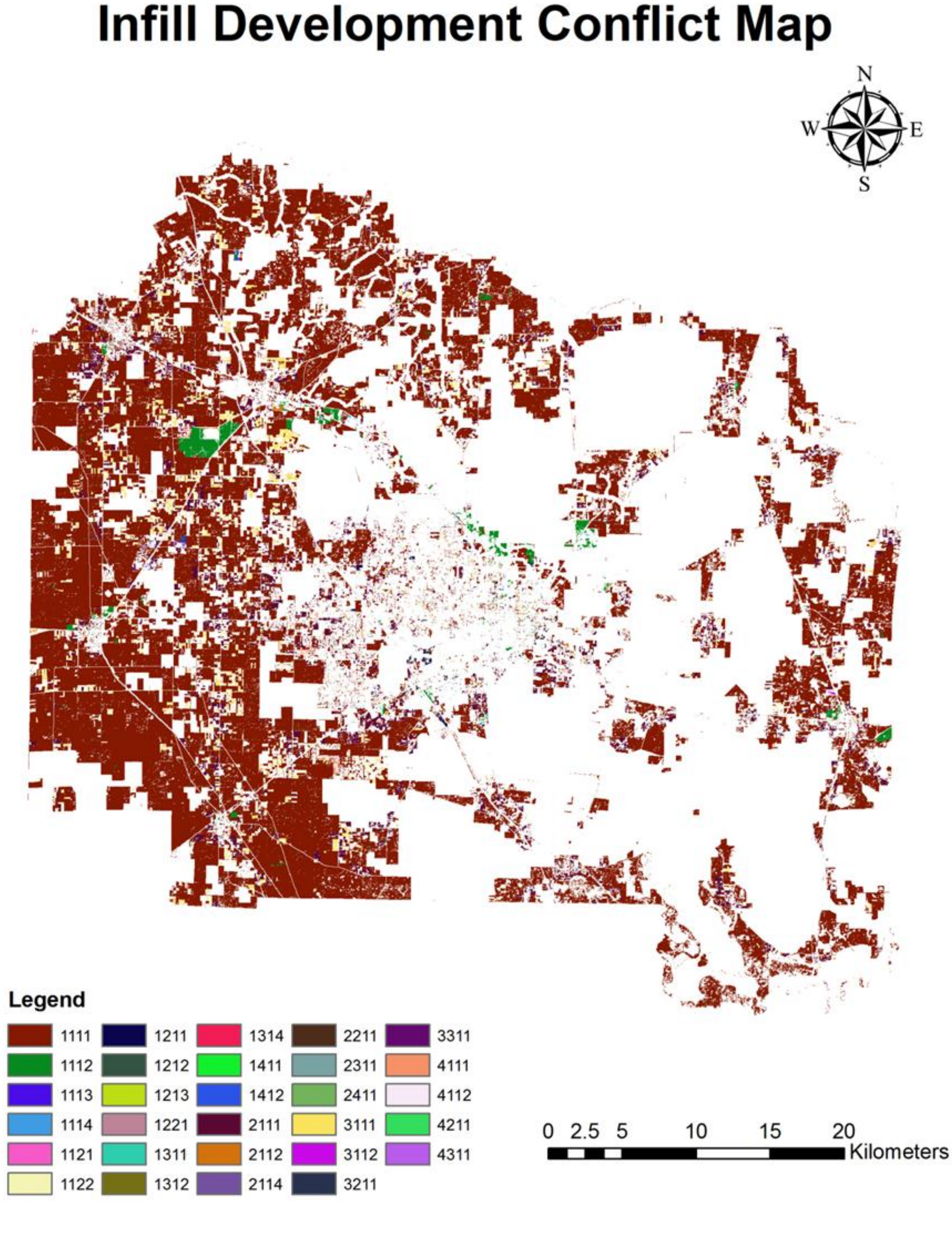

Figure 4-11. Conflict map for the infill scenario

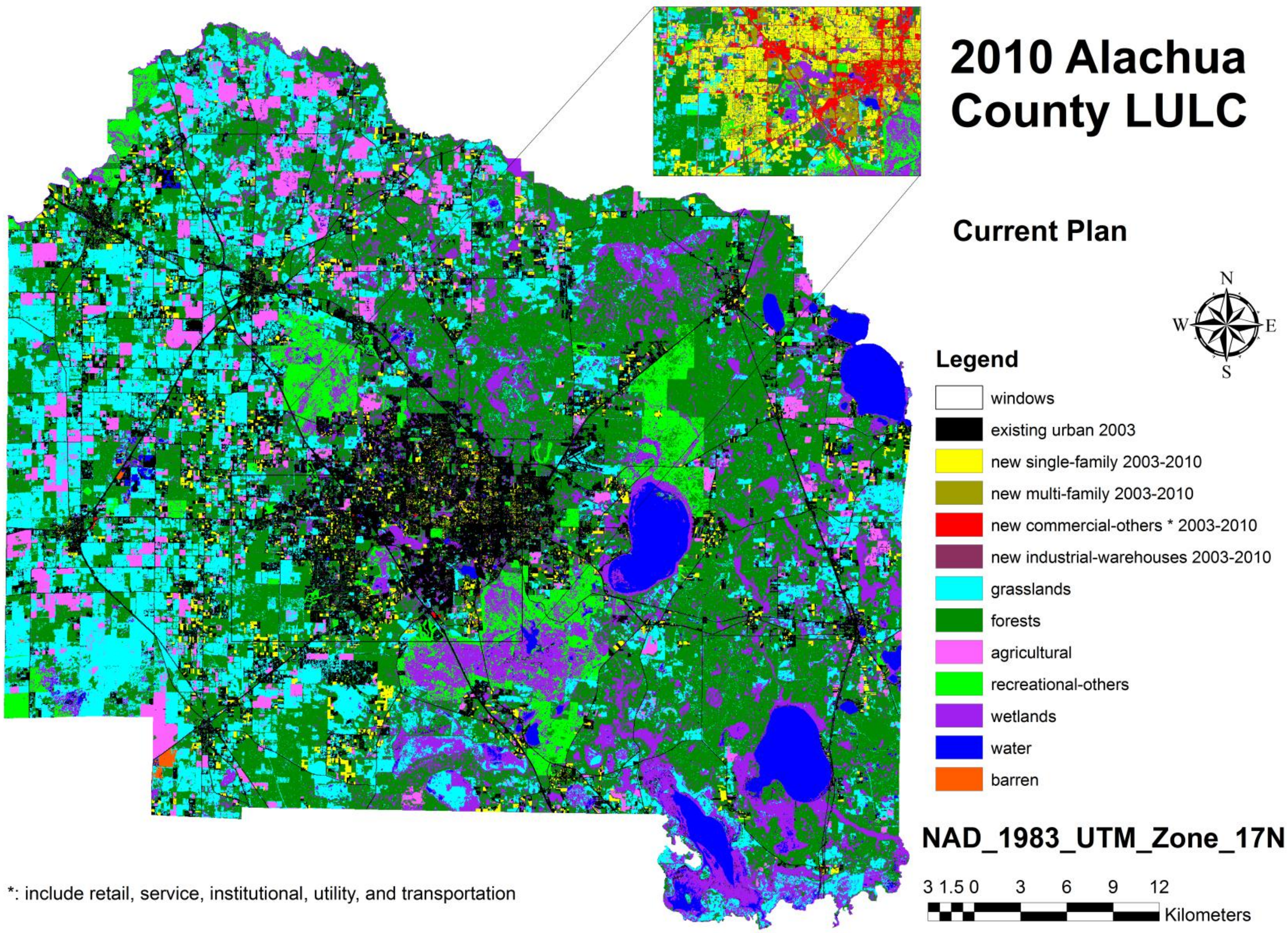

Figure 4-12. 2010 Alachua County LULC Current Plan 1

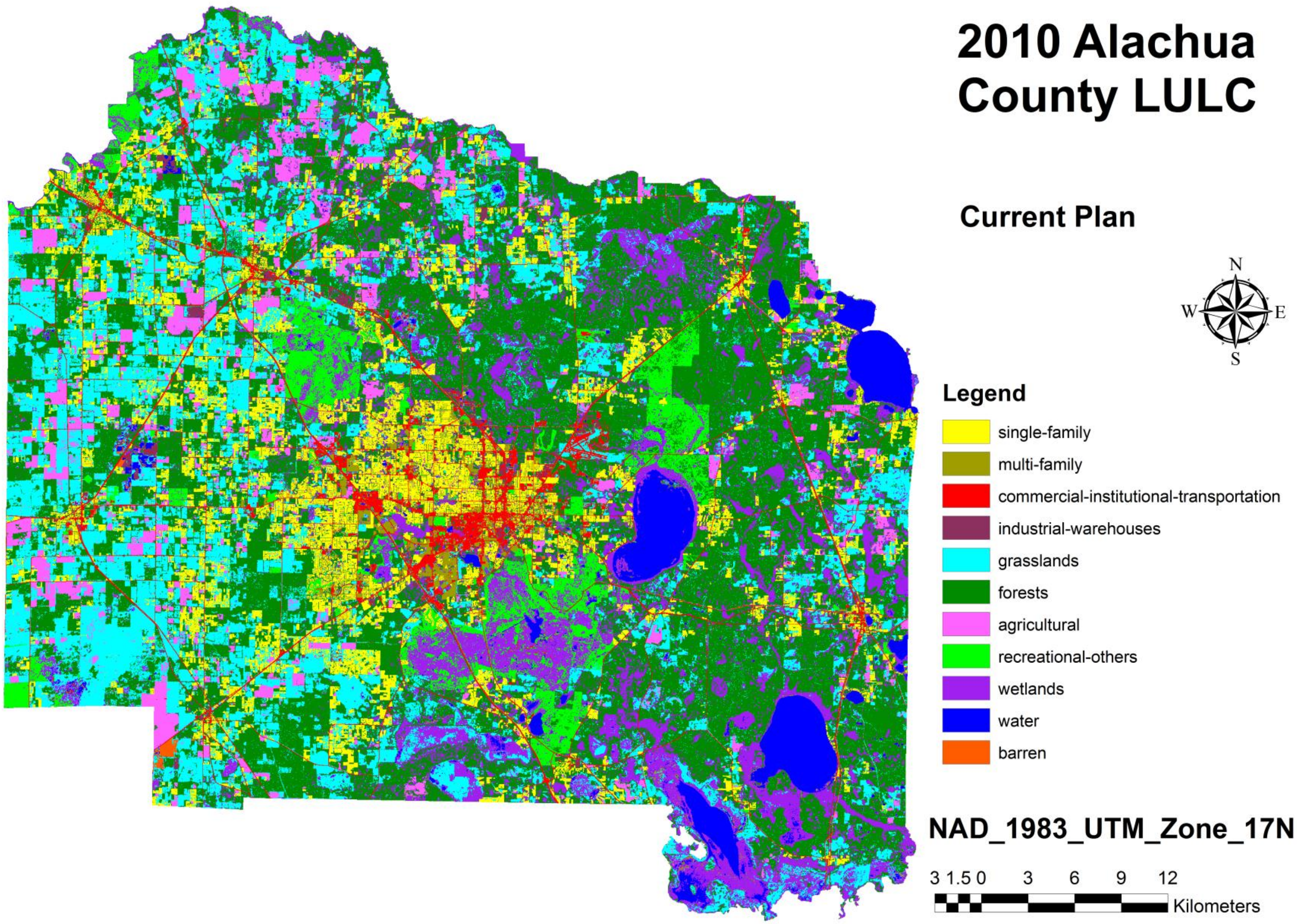

Figure 4-13. 2010 Alachua County LULC Current Plan 2

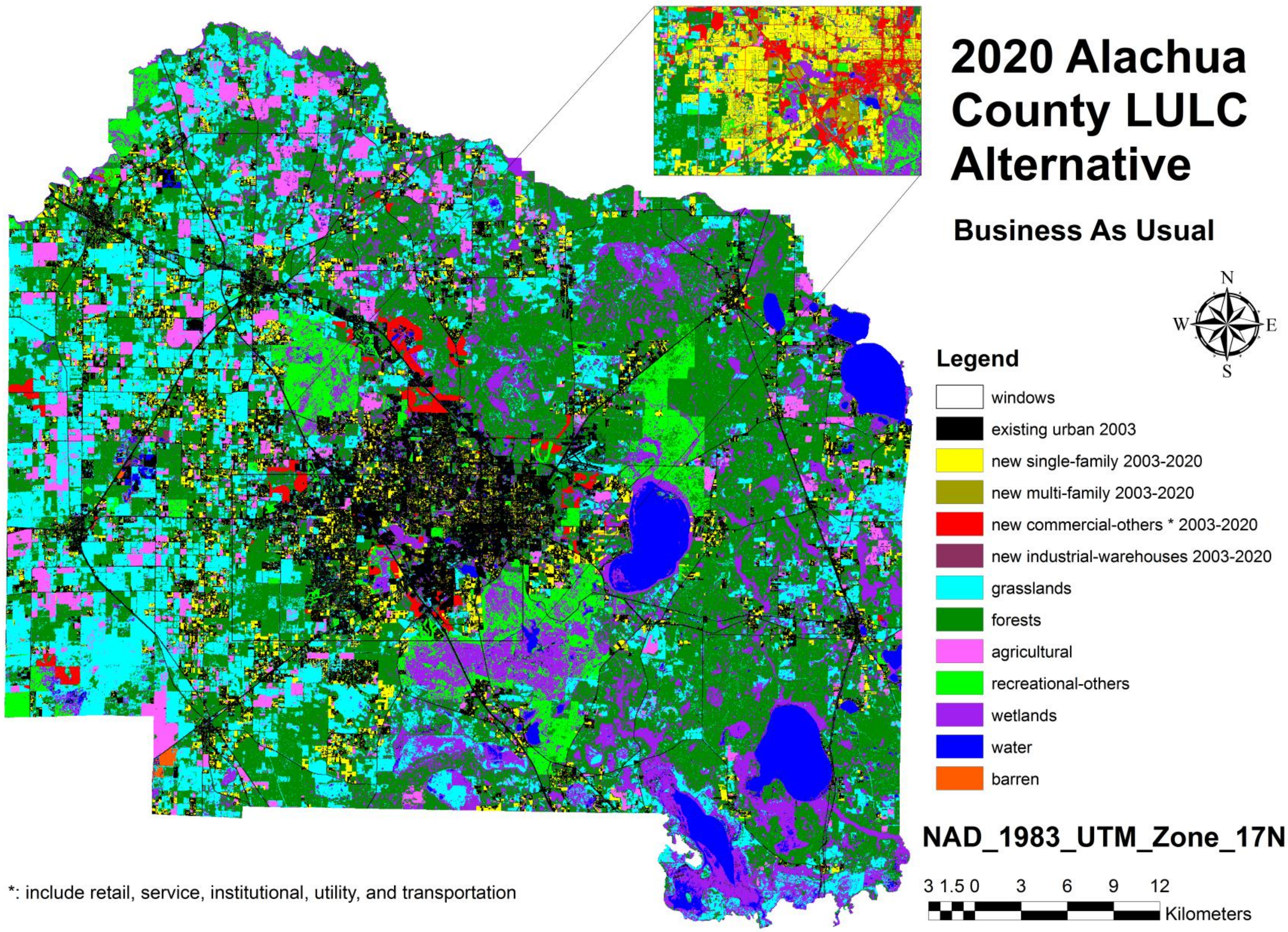

Figure 4-14. 2020 Alachua County LULC Alternative Business As Usual Scenario 1

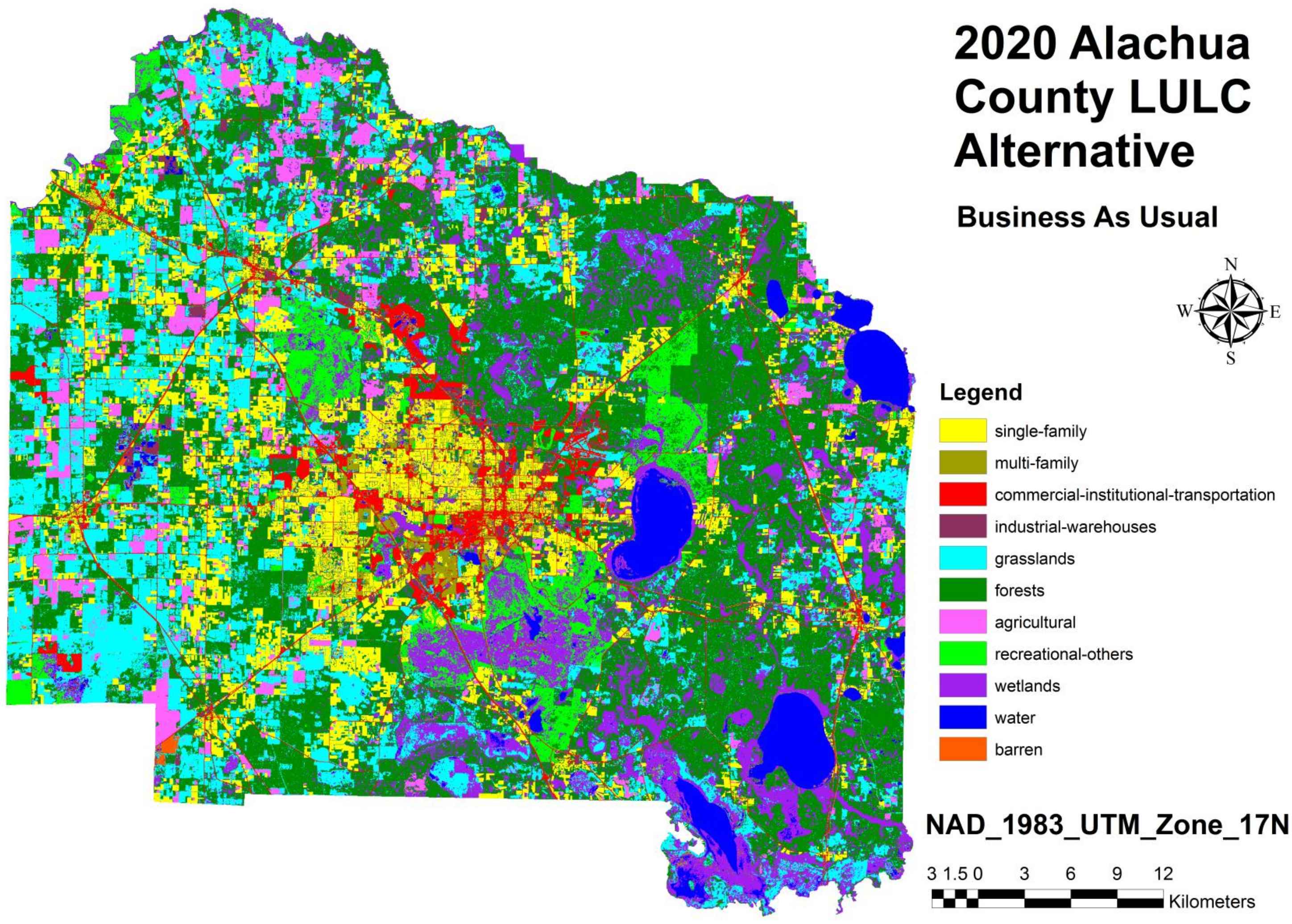

Figure 4-15. 2020 Alachua County LULC Alternative Business As Usual Scenario 2

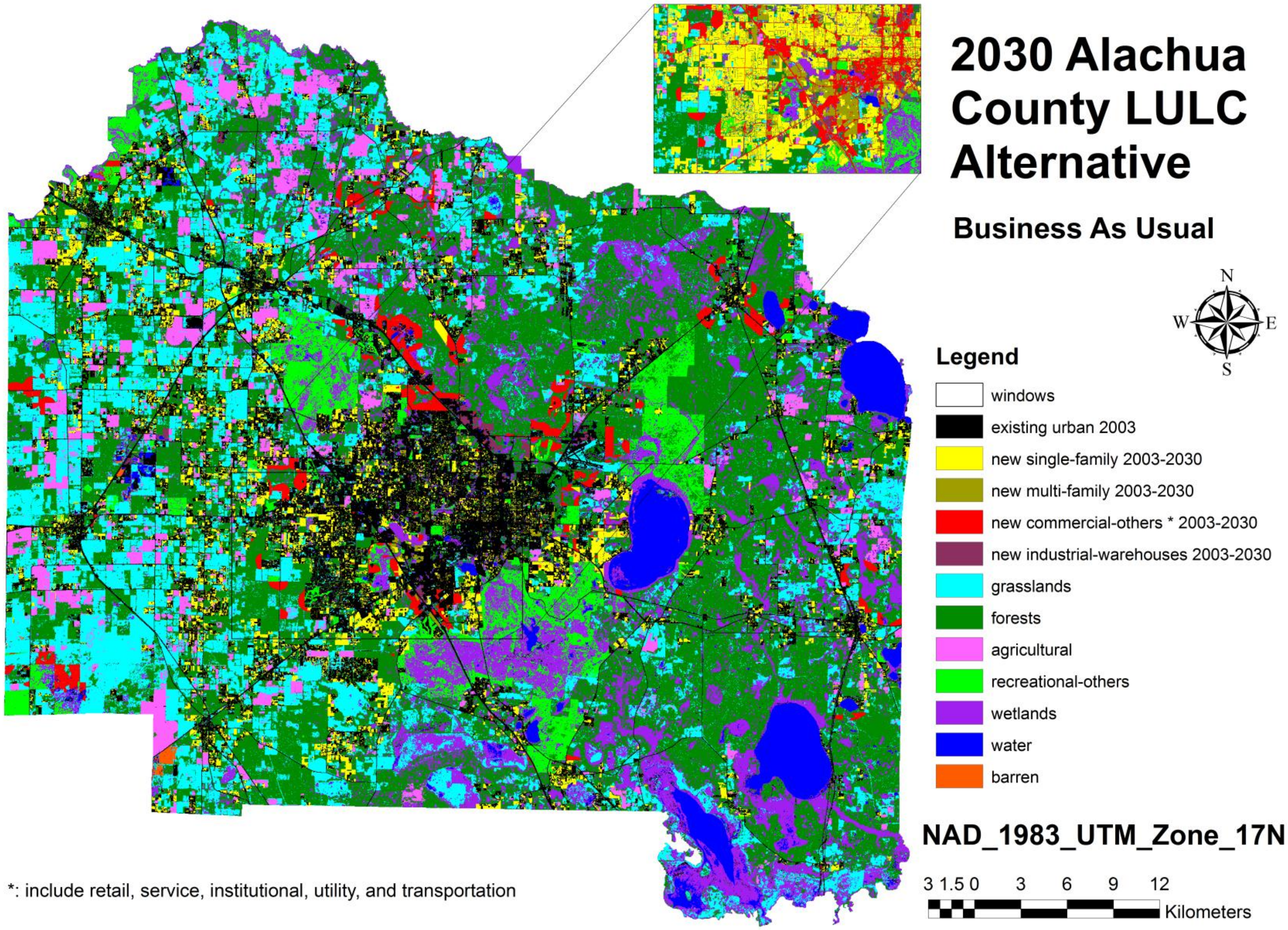

Figure 4-16. 2030 Alachua County LULC Alternative Business As Usual Scenario 1

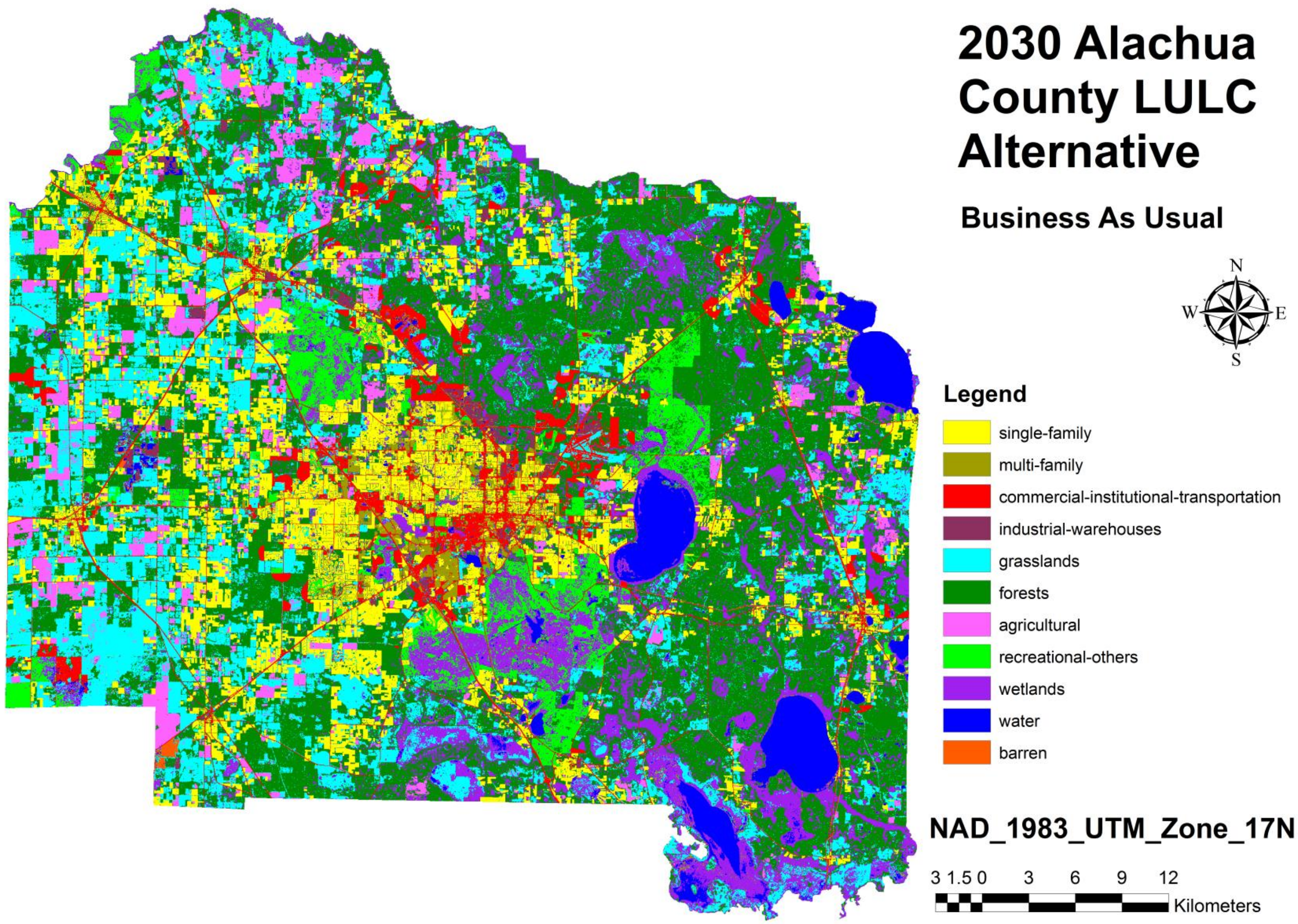

Figure 4-17. 2030 Alachua County LULC Alternative Business As Usual Scenario 2

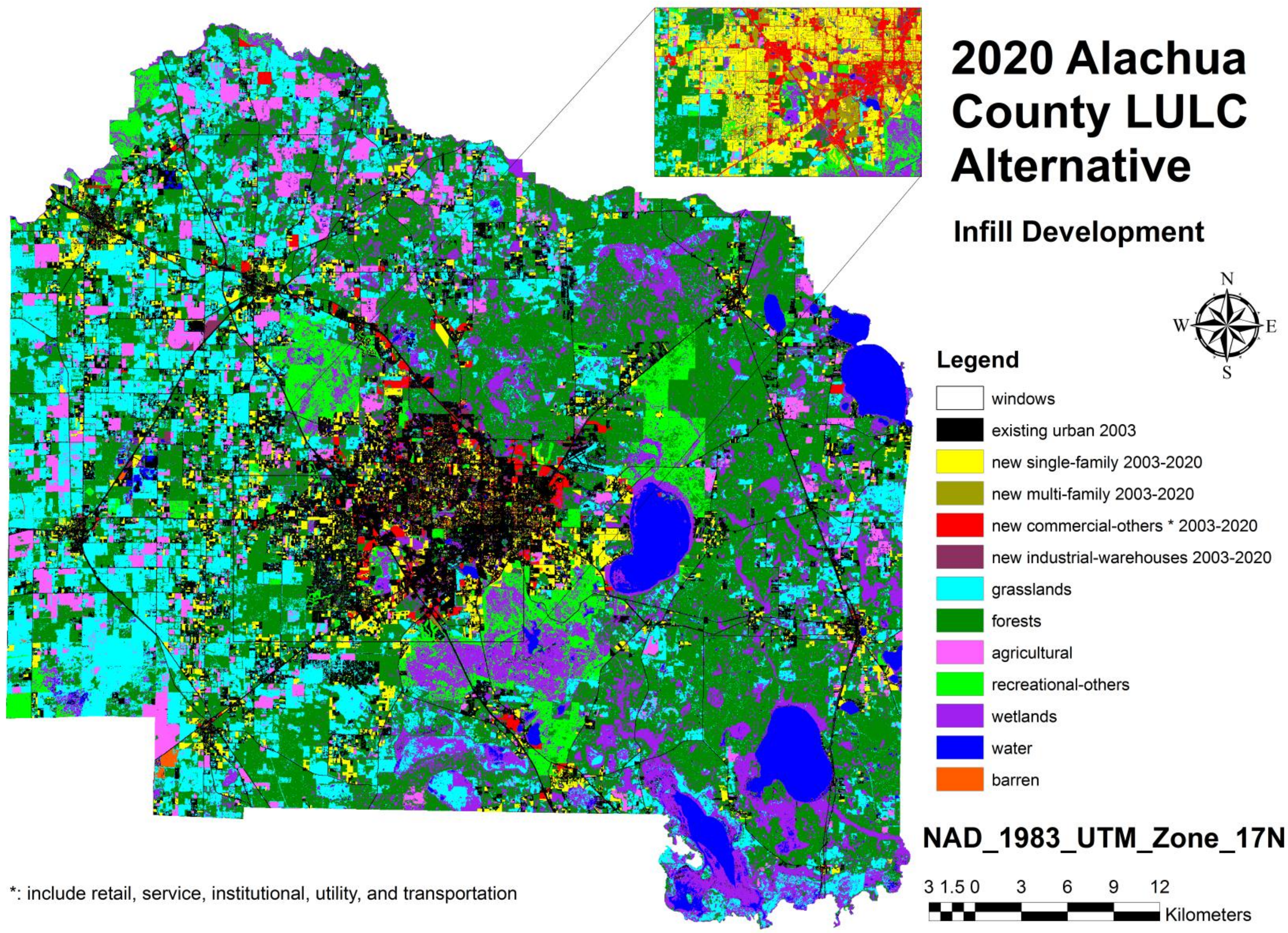

Figure 4-18. 2020 Alachua County LULC Alternative Infill Development Scenario 1

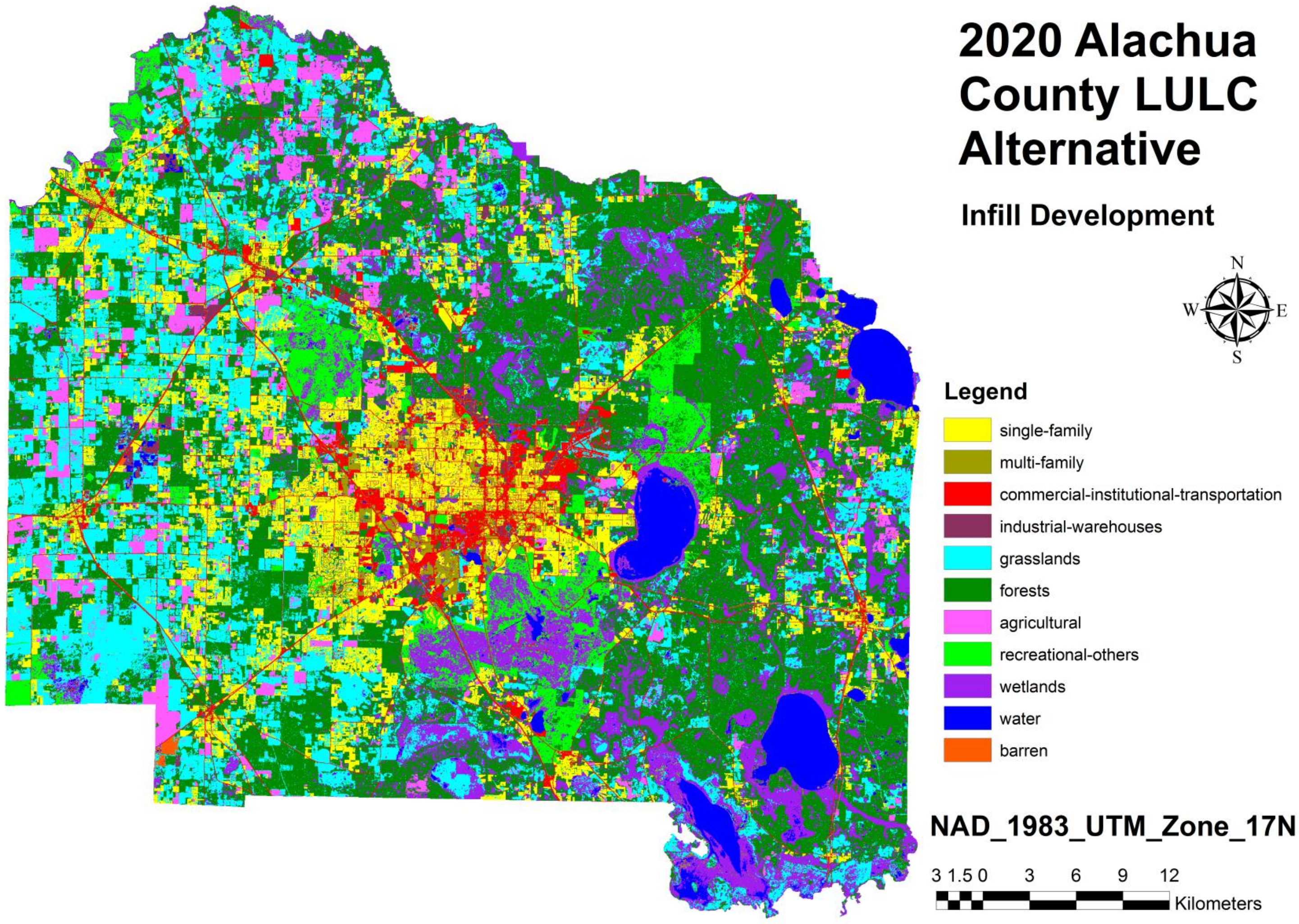

Figure 4-19. 2020 Alachua County LULC Alternative Infill Development Scenario 2

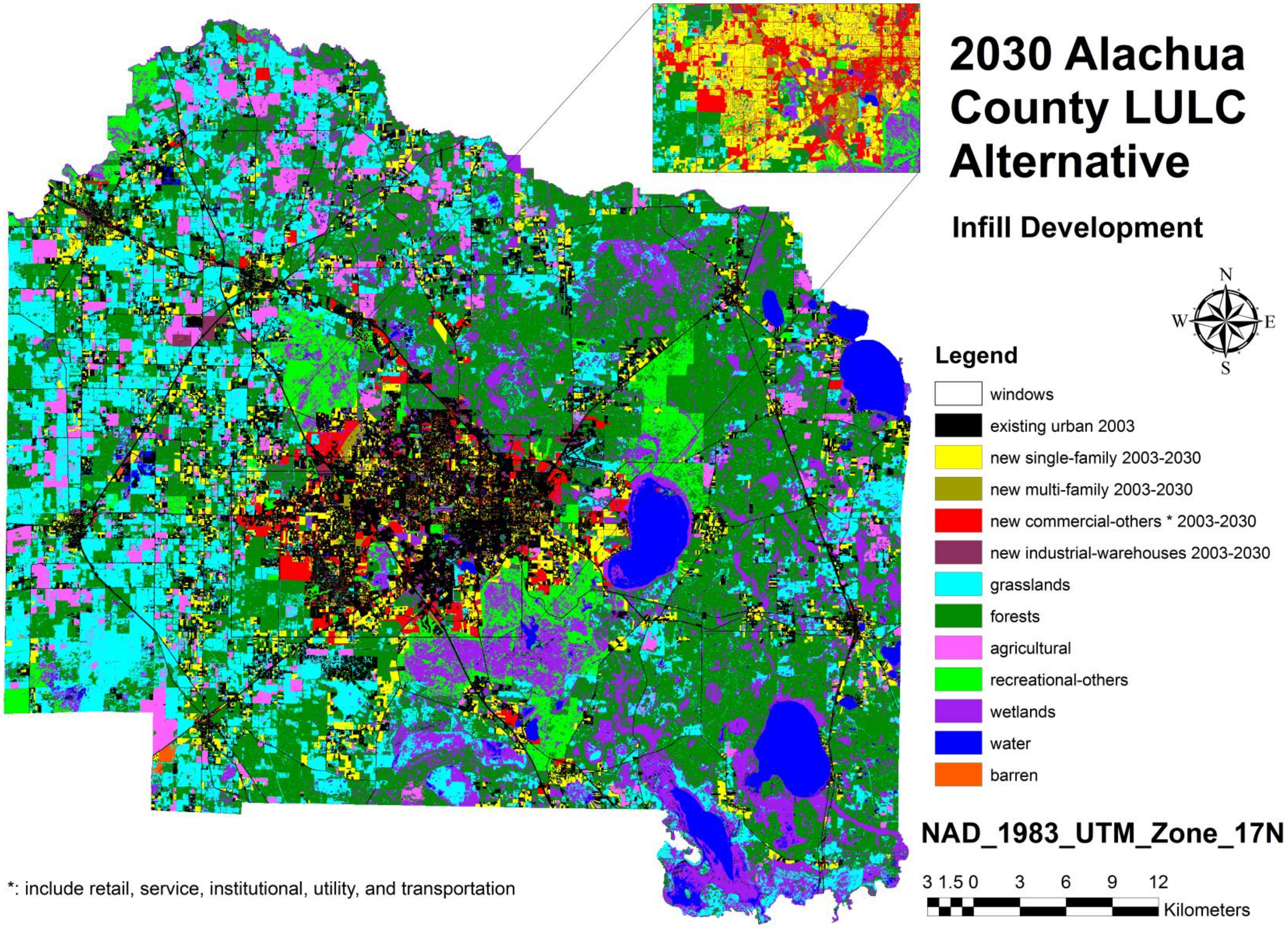

Figure 4-20. 2030 Alachua County LULC Alternative Infill Development Scenario 1

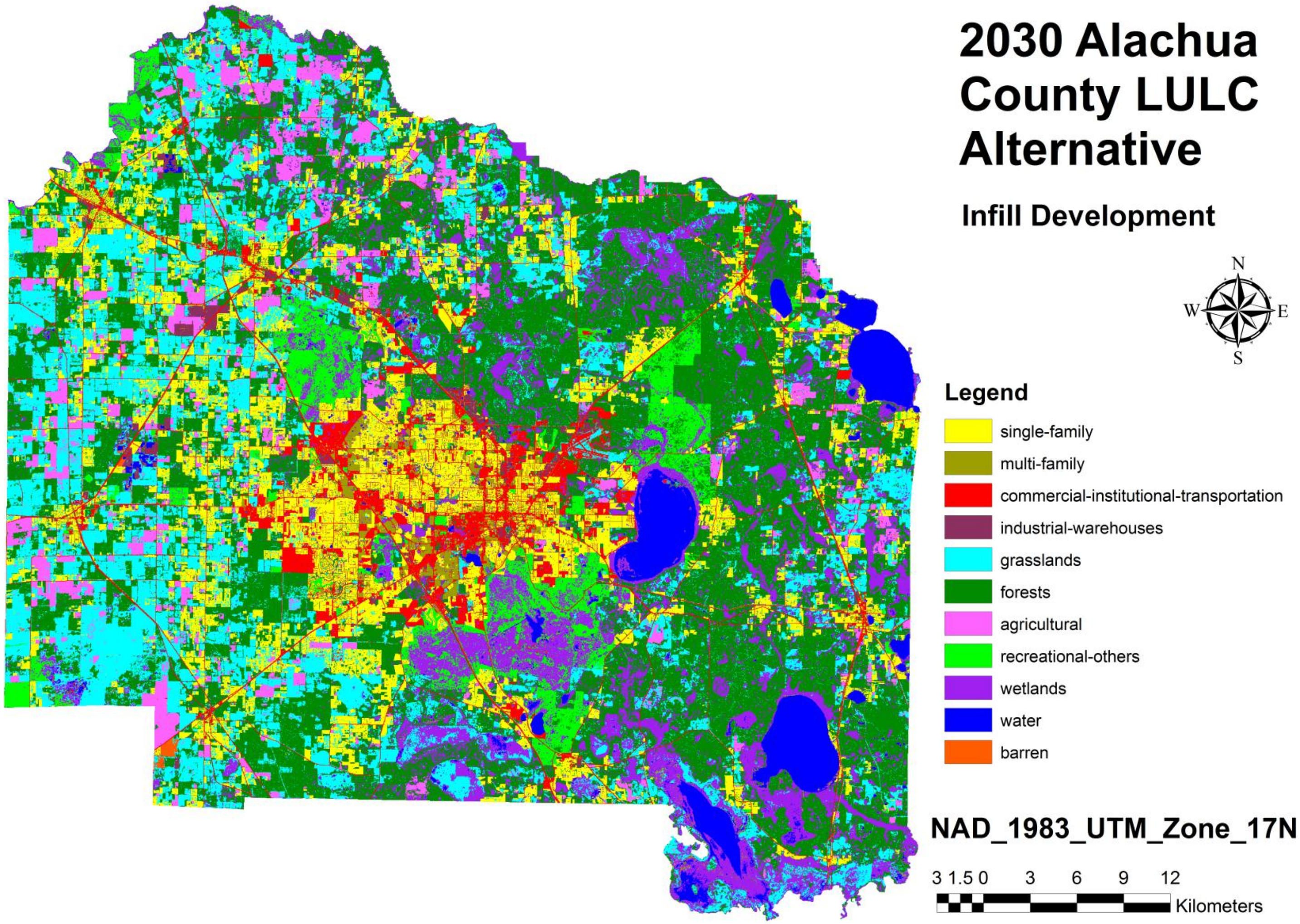

Figure 4-21. 2030 Alachua County LULC Alternative Infill Development Scenario 2

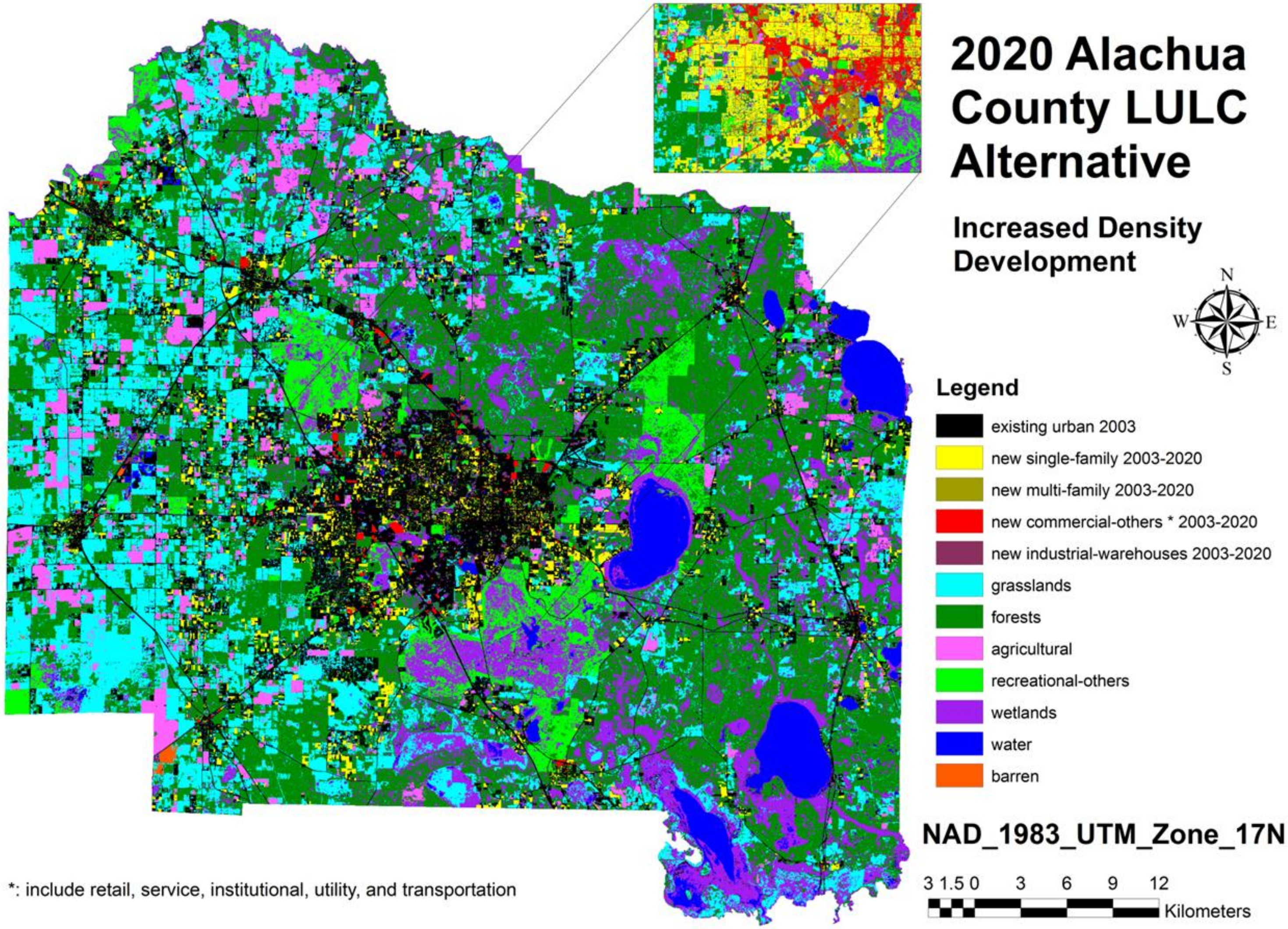

Figure 4-22. 2020 Alachua County LULC Alternative Increased Density Development Scenario 1

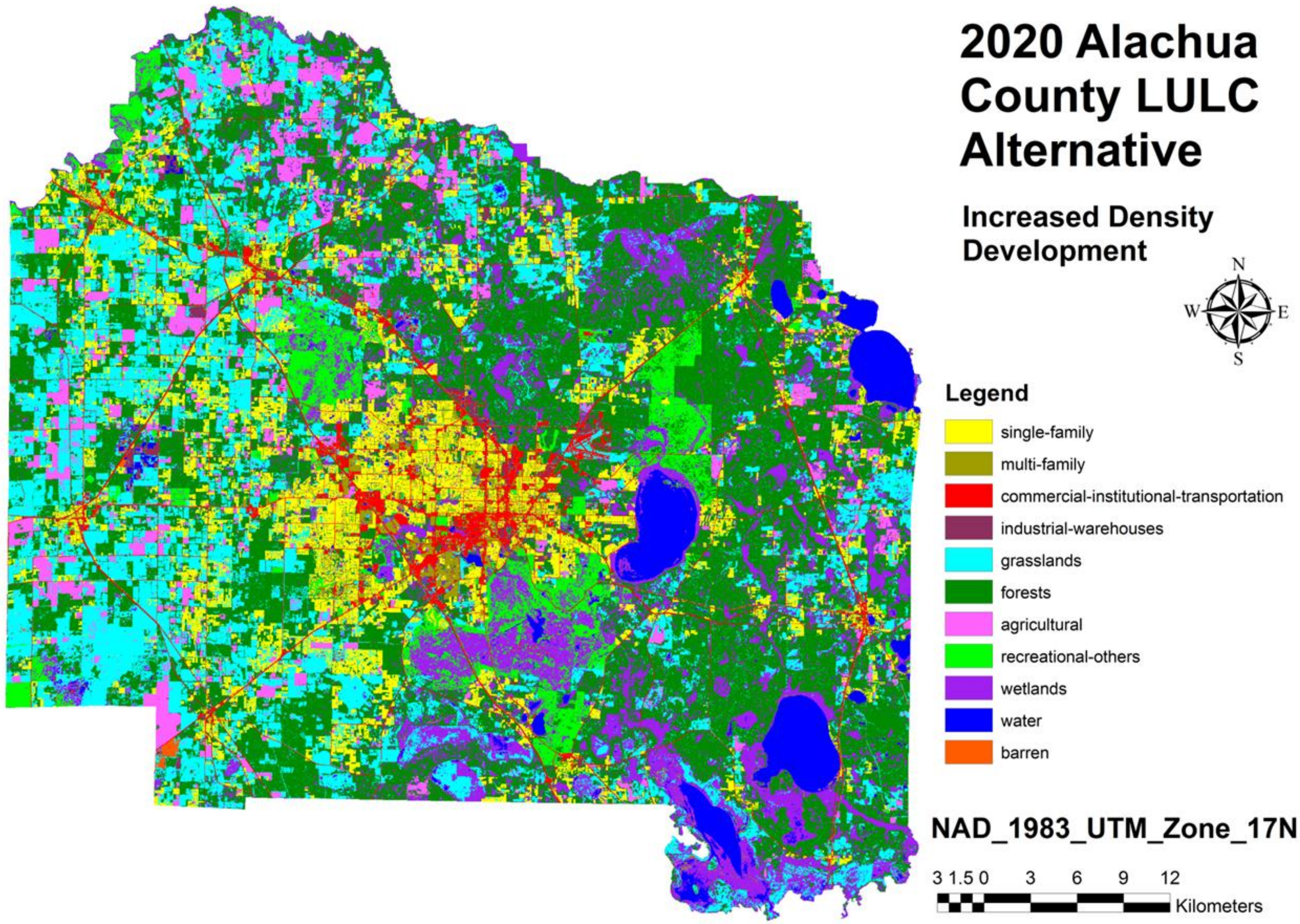

Figure 4-23. 2020 Alachua County LULC Alternative Increased Density Development Scenario 2

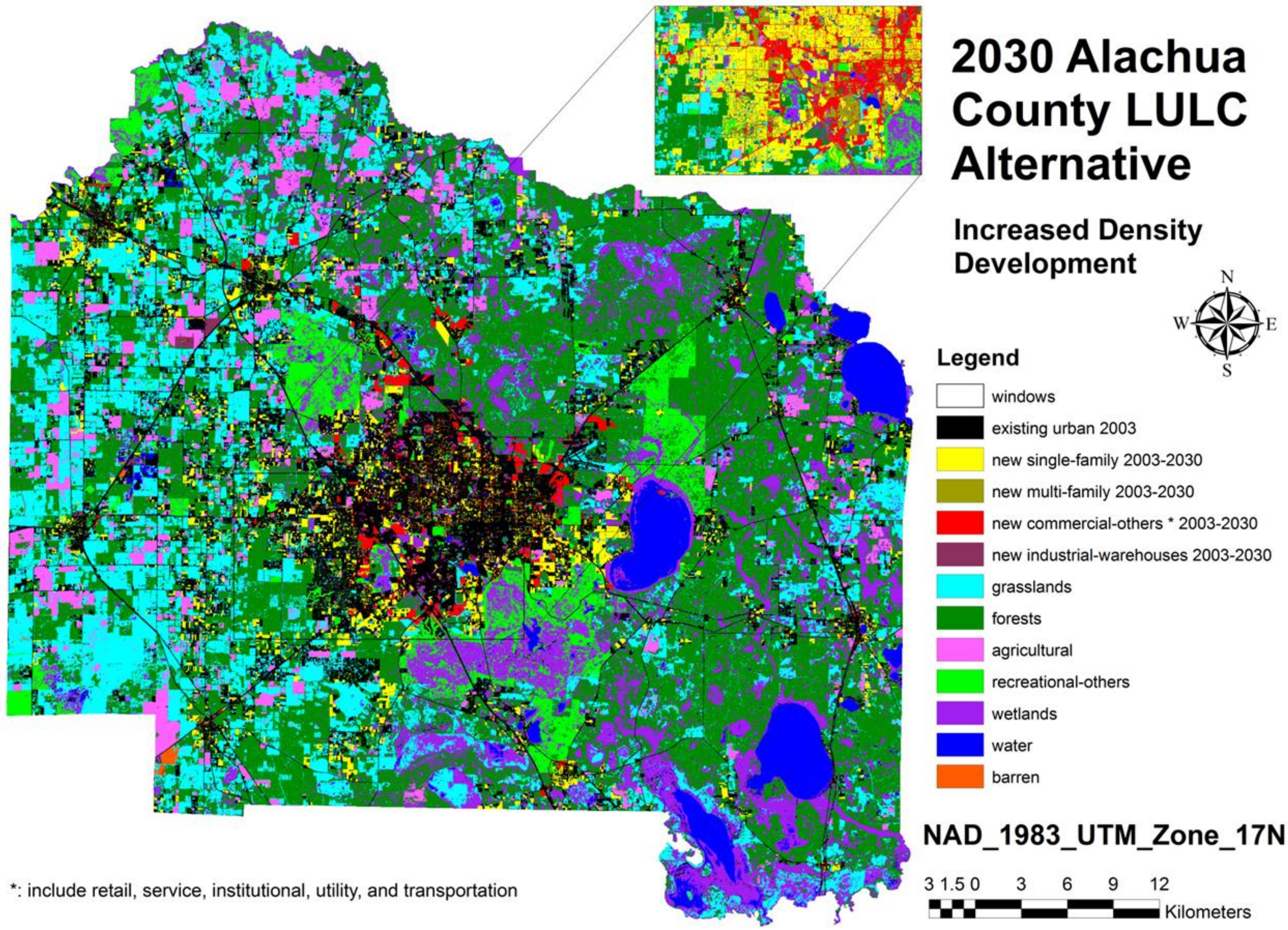

Figure 4-24. 2030 Alachua County LULC Alternative Increased Density Development Scenario 1

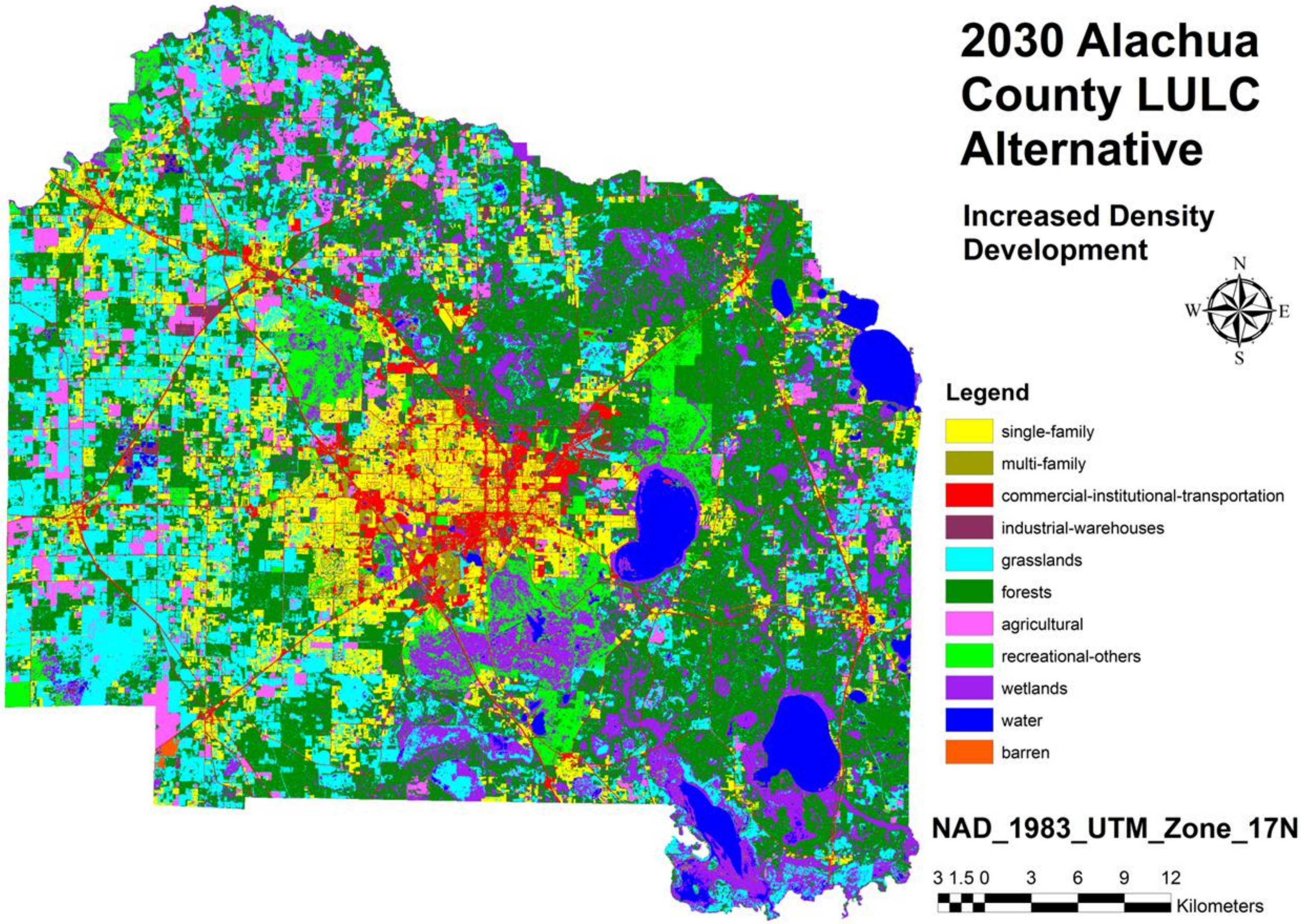

Figure 4-25. 2030 Alachua County LULC Alternative Increased Density Development Scenario 2

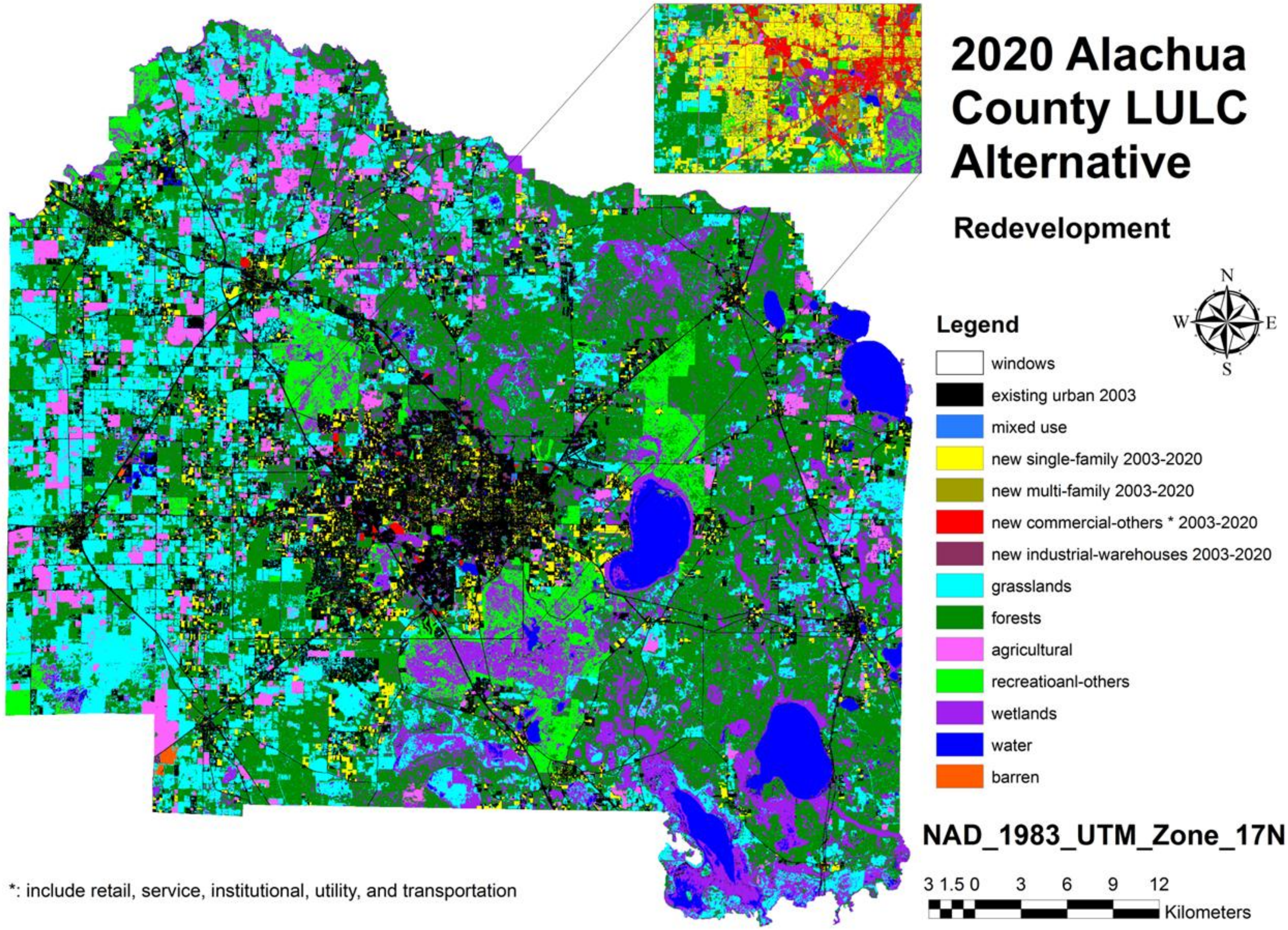

Figure 4-26. 2020 Alachua County LULC Alternative Redevelopment Scenario 1

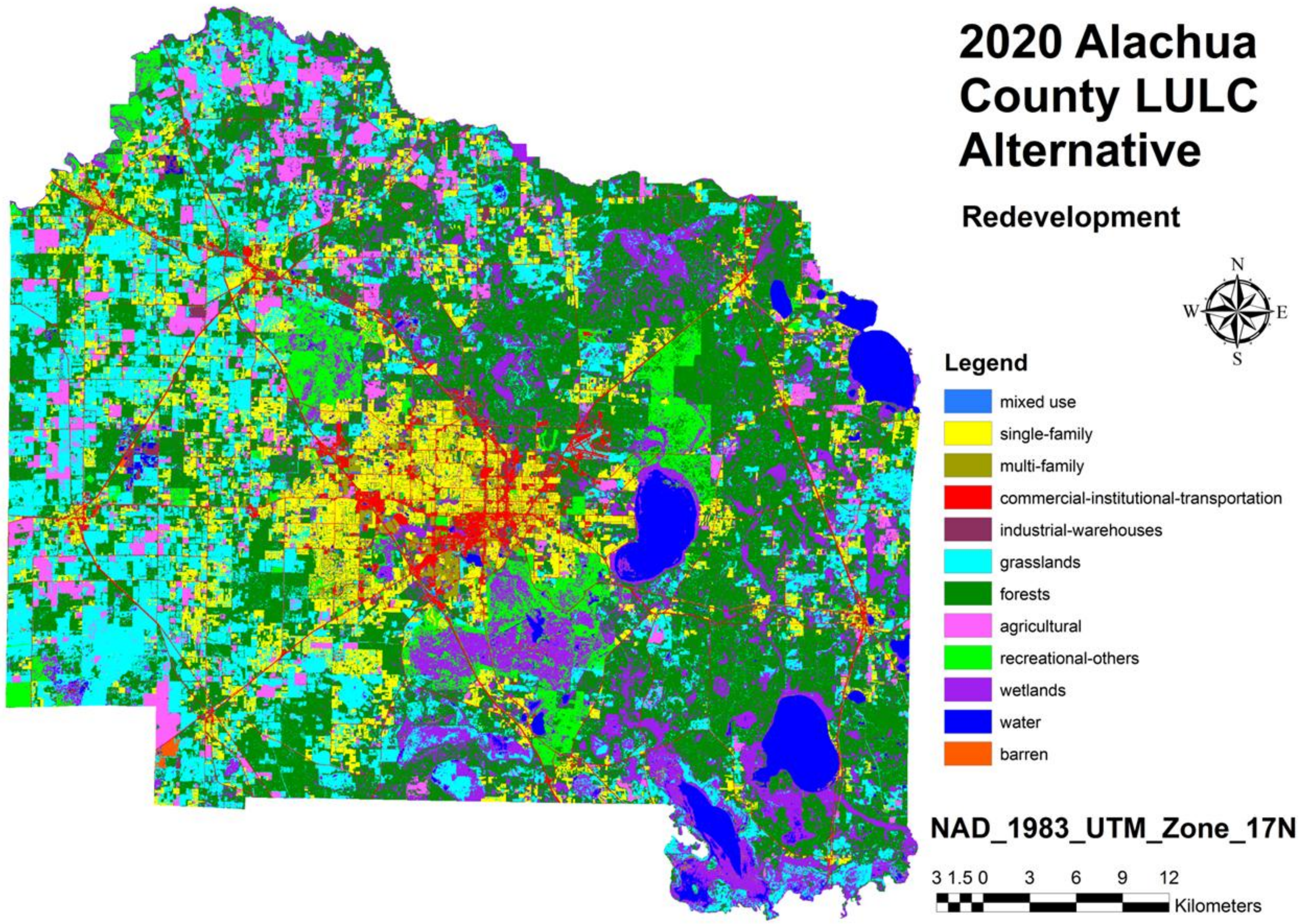

Figure 4-27. 2020 Alachua County LULC Alternative Redevelopment Scenario 2

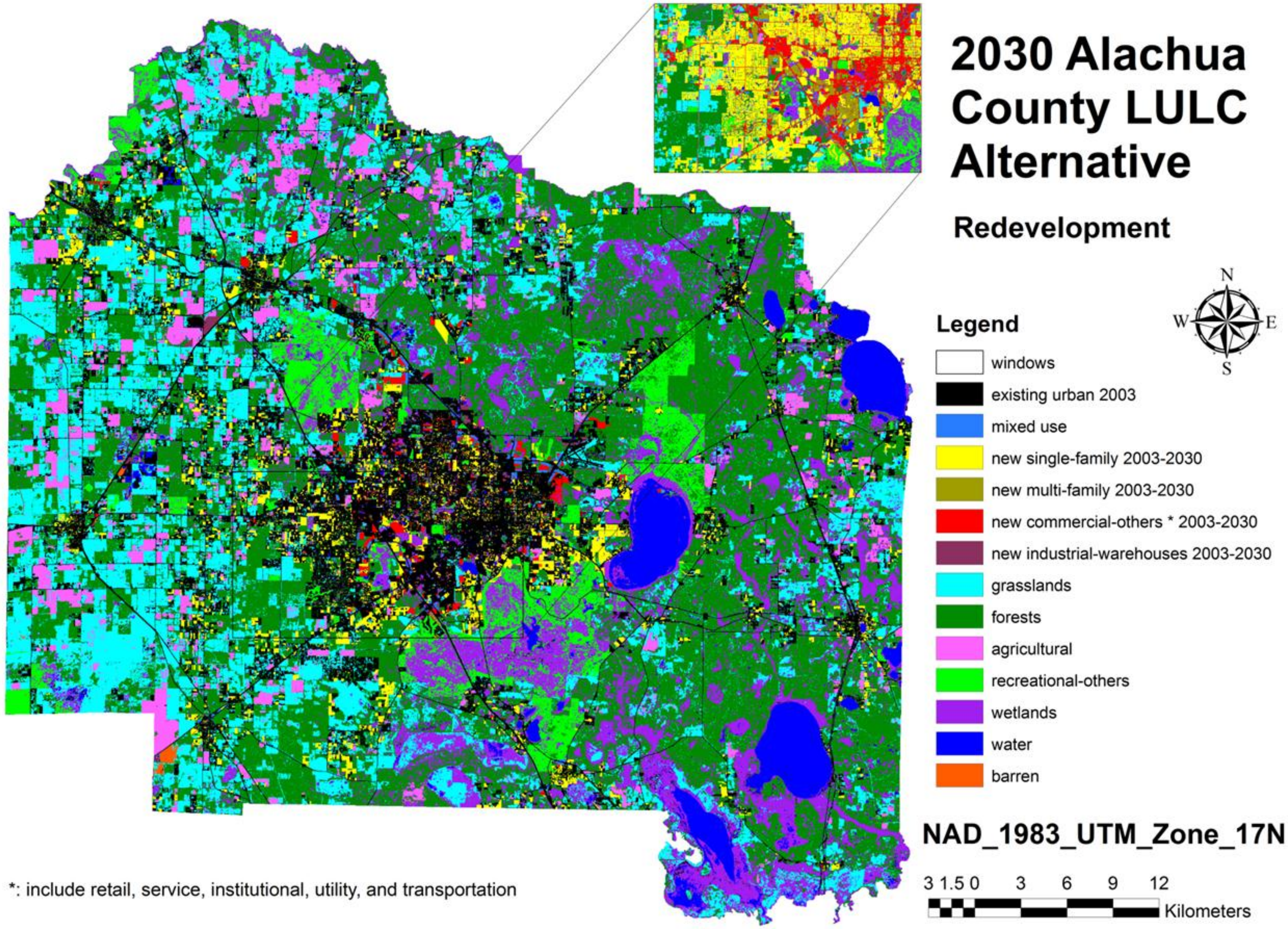

Figure 4-28. 2030 Alachua County LULC Alternative Redevelopment Scenario 1

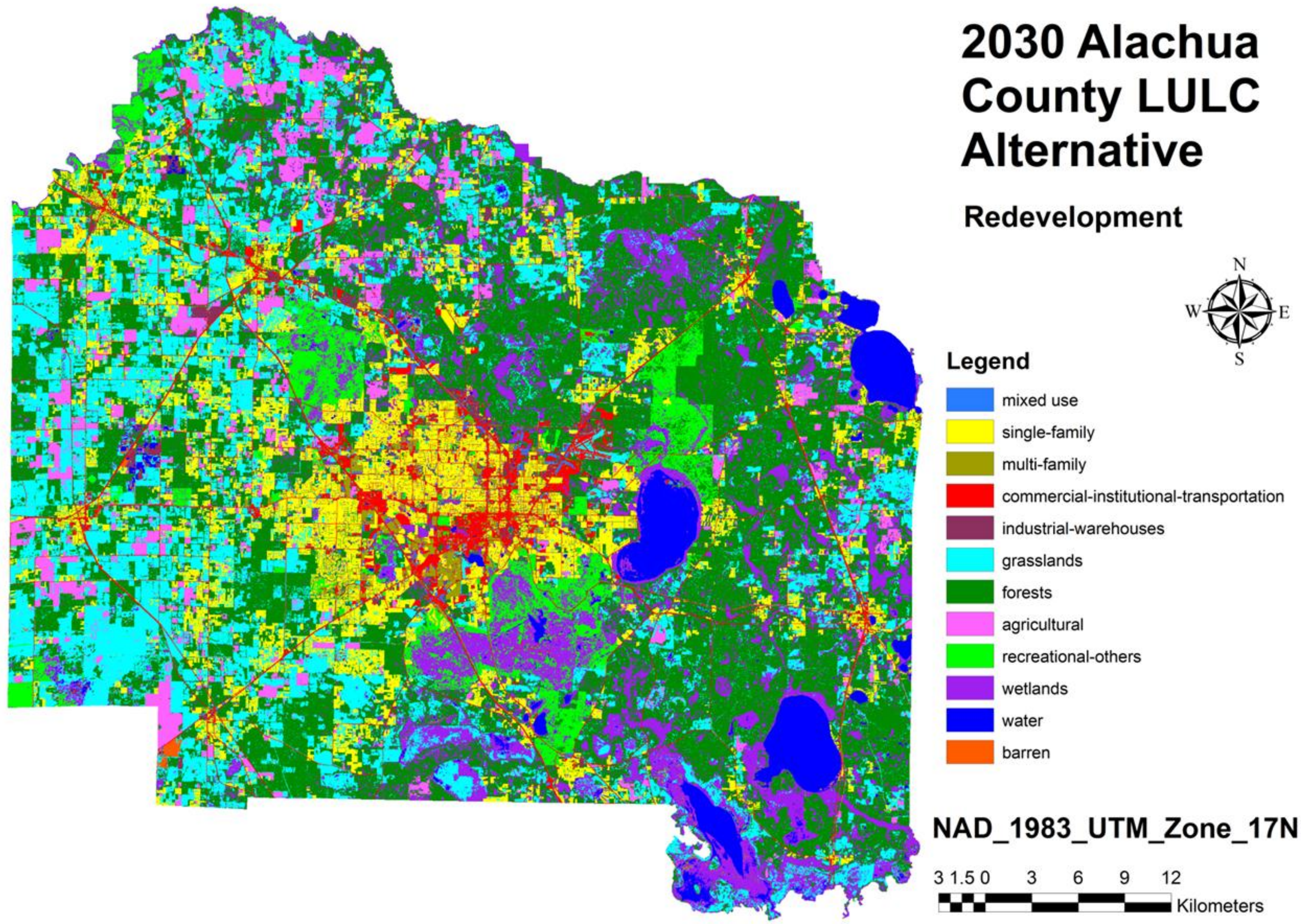

Figure 4-29. 2030 Alachua County LULC Alternative Redevelopment Scenario 2

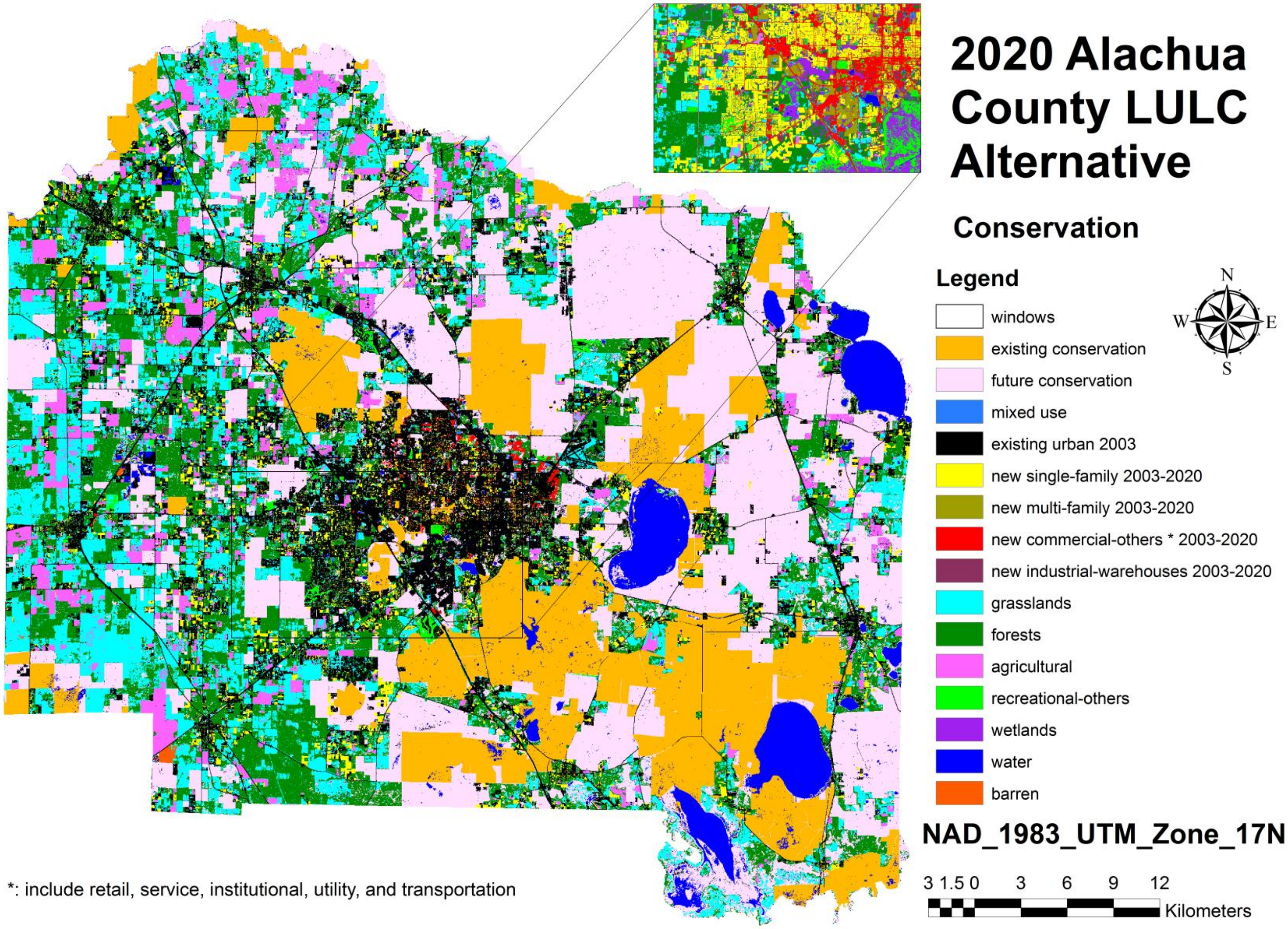

Figure 4-30. 2020 Alachua County LULC Alternative Conservation Scenario 1

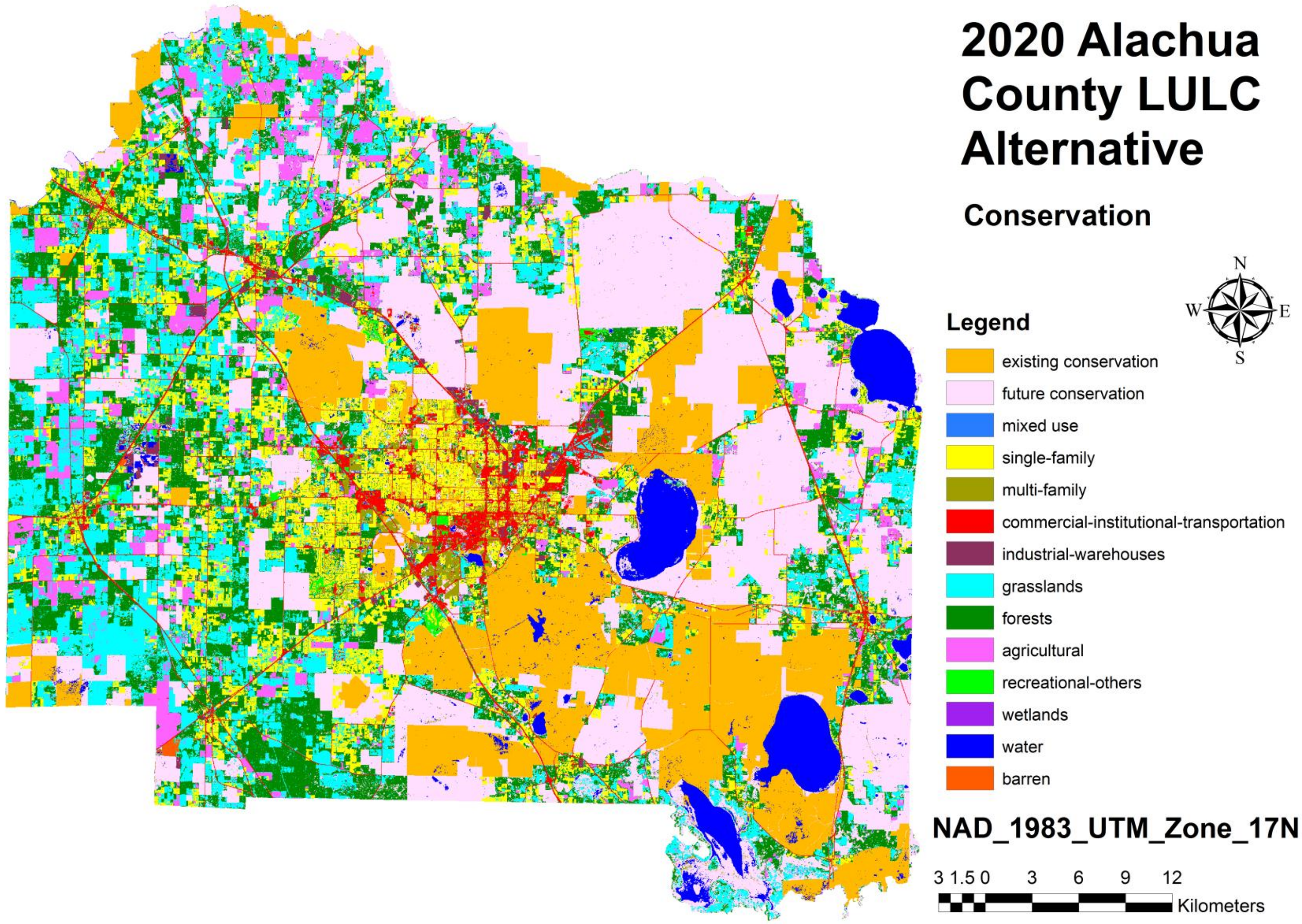

Figure 4-31. 2020 Alachua County LULC Alternative Conservation Scenario 2

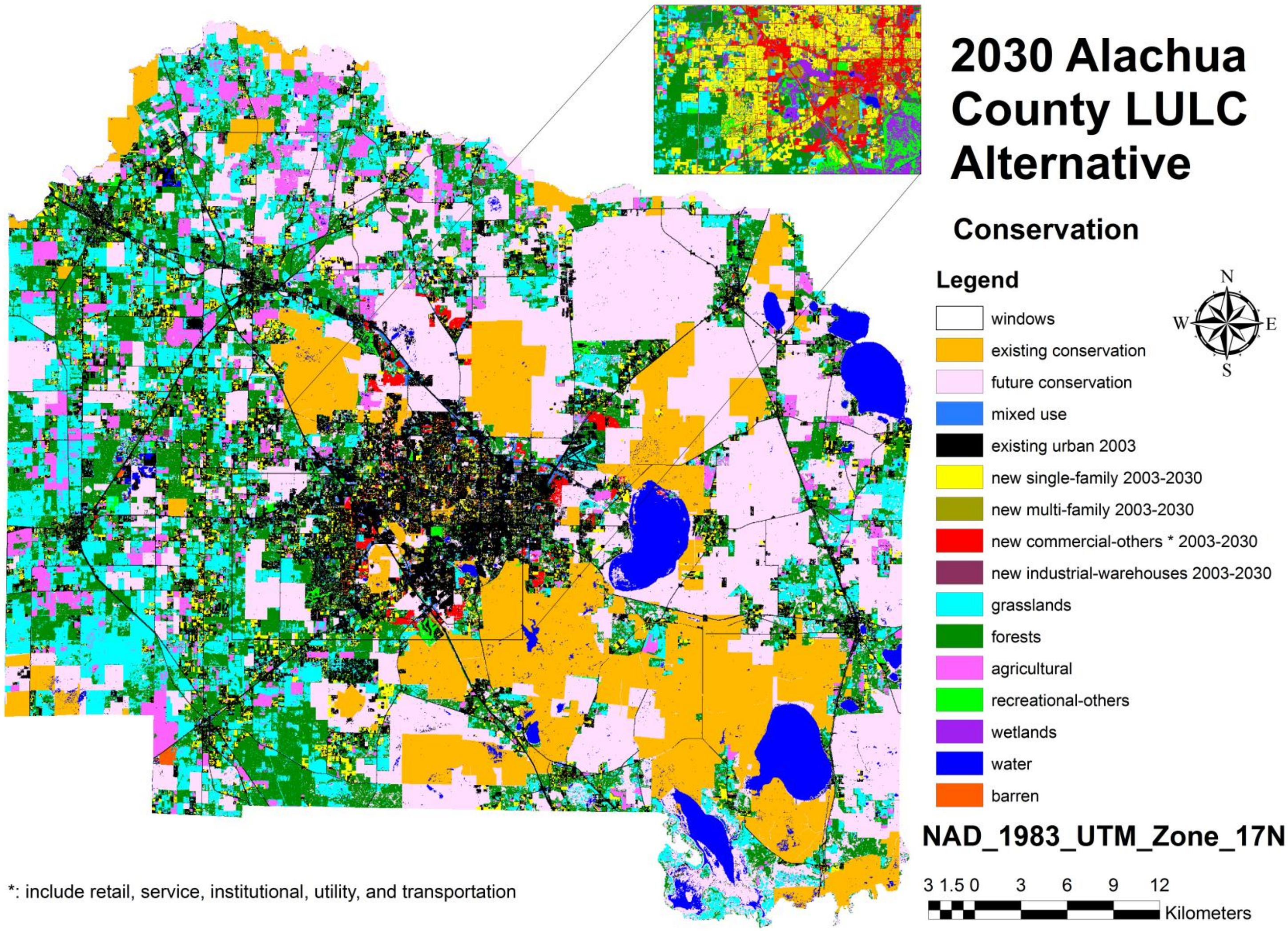

Figure 4-32. 2030 Alachua County LULC Alternative Conservation Scenario 1

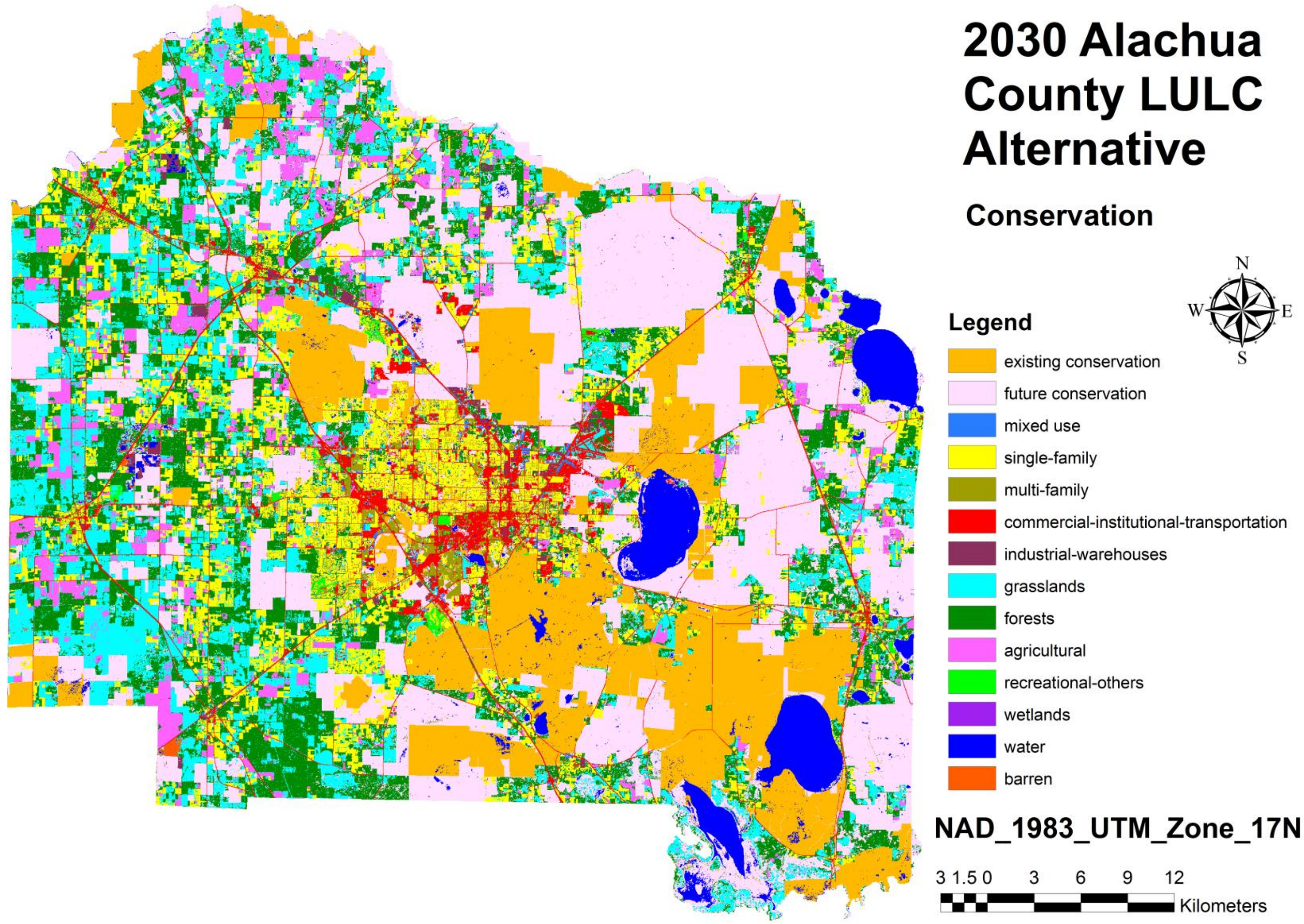

Figure 4-33. 2030 Alachua County LULC Alternative Conservation Scenario 2

CHAPTER 5
CONCLUSIONS AND DISCUSSION

This research has successfully classified satellite imagery for eleven urban and natural LULC classes for 1982, 1994, and 2003, and also successfully projected four urban LULC classes for both 2020 and 2030, in Alachua County, Florida, based on five sets of articulated assumptions. From the study, past, present, and future spatial patterns of urban development are well presented, in order to compare changes over time resulting from the variation in adopted assumptions.

Using remote sensing data alone or parcel data alone did not result in a satisfactory method of LULC classification, so remote sensing data and GIS parcel data are used concomitantly. The comparison of classification strategies, included in the study, proves that this combined approach results in the highest degree of accuracy.

The high accuracy of classification maps in Alachua County in 1982, 1994, and 2003 cannot have been obtained without applying ancillary data such as parcel information. This is the major advantage of the CART method as it includes ancillary data into considerations. In addition, the CART method best represents land uses rather than land covers, especially for small parcels. However, the land covers in large parcels can still be distinguished by the CART method. As a result, sub-parcel land cover details such as wetlands can be identified well in this research. Generally speaking, the inclusion of parcel data as well as the information derived from the parcel data is the most important step for successful county-wide LULC classifications.

Because the computing power that processes the ROIs in ENVI 4.4 RuleGen 1.02 (Chapter 2) to classify urban LULCs is overburdened, reduced ROI signatures are sought based on the random sampling tool. If it is known in advance that the ROIs will overburden the computer, reduced ROI signatures will be sought in the beginning. This effort can save significant CPU computing time: at least half of the pixel amounts can be cut at the outset.

The logistic regression model built in Chapter 3 was used to simulate urban growth in Alachua County using four urban LULC classes. This class level is equivalent to USGS Level III. Because the urban LULC is accurately classified in Chapter 2, the logistic regression model is able to yield high accuracy for each of the four urban LULC classes for urban growth simulation. This study simulated urban growth in the county for 2003 LULC classes based on the urban LULC classifications in 1982, 1994, and 2003, which are based on model construction, model calibration, model interpretation, model refinement, and sensitivity analysis. When compared with actual 2003 urban land use, the model accurately predicted the spatial distribution of the four LULC classes.

This logistic regression model has advantages compared to other AI models such as the CA model, the ABM model, the Neural Network (NN) model, and the Genetic Algorithm (GA) model. For example, the logistic regression model includes consideration of historical factors with longitudinal data being imported into the model, and it is also capable of including spatial variables into the model. As a result, the research goals and objectives of this dissertation are effectively achieved; the overall accuracy level for the 2003 simulated LULC map reaches 97.30%. Because of the empirical trials of various spatial independent variables into the model that are tested incrementally on a one-by-one basis, the accuracy level for each of the four urban dependent variables can reach more than 97%.

In Chapter 4, this study allocates urban LULC for Alachua County for 2020 and 2030. The allocation of urban LULC is based on the articulation of five development scenarios, which cover almost all the possibilities for the county’s future urban development. As a result, the possible future development patterns for each scenario can be assessed and mapped.

Recapping the entire dissertation, it is found that although this dissertation has accomplished the author’s research goals, the study hasn’t had a chance to try additional intuitive classification methods other than the CART method, the V-I-S method, and supervised classification methods (Chapter 2). The new and intuitive classification methods—including the NN method, the thermal method, and the fuzzy supervised method—are worth trying. Although adding these classification methods to the county-based LULC classifications may produce the same results, the county-based LULC classification research will be more interesting and exciting. Similarly, the satellite SPOT data can also be tried for LULC classifications to further elevate accuracy; QuickBird and IKONOS imageries may also be used in this regard.

In addition, the author hasn’t conducted the simulation for redevelopment using the MLR model. According to Landis and Zhang (1997), the MLR model is capable of simulating redevelopment. More research shall be conducted in this regard to simulate redevelopment in Alachua County; for example, a hybrid model that combines the MLR model and the CA model can be considered in this manner. Correspondingly, because of time constraints, the author has conducted research for a rural county—Alachua County—only. The author hopes that similar urban LULC simulation can be conducted for an urban county such as Duval County (the City of Jacksonville), Orange County (the City of Orlando), or Miami-Dade County (the City of Miami) so as to compare the rural county with an urban county. Furthermore, because the MLR model has the disadvantage that it is less temporally dynamic (Chapter 3), this model needs a complicated allocation process to overcome. The author hopes to try the CA model and the Markov-Chain Model (MCM) to simulate urban growth for Florida counties. As a result, midterm simulation results can be obtained which may be interesting and exciting. Three modules can be applied: the CELLATOM module, the CA_MARKOV module, and the MARKOV module. They are available in the IDRISI software. Suitability analyses will be needed in this regard for model building rather than probabilities in the MLR model.

Because this study didn't consider LULC changes for natural land, the future study shall take this factor into consideration. To do this work, a Conversion of Land Use and Its Effects (CLUE) model can be applied. The CLUE model has a LULC conversion matrix, in which codes can be used to represent the changes and non-changes between different pairs of LULC classes (Verburg, no date); for example, 0 represents that LULC classes can be changed from LULC A to LULC B, and 1 represents that LULC classes cannot be changed. As a result, more research should be undertaken to see if the CLUE model is capable of simulating redevelopment, because the LULC conversion matrix in the CLUE model may potentially be used to simulate redevelopment.

This study adopted a twenty-year timeframe (to 2030) for urban growth simulation. The examination of land use change would be more meaningful if the time frame could be extended beyond 20 years. A research time frame that extended 50 years or more is optimal.

In sum, this study classifies eleven LULC classes up to the level of USGS Level III and simultaneously results in high accuracy. Because of the highly accurate LULC that have been classified, the MLR model is built with high accuracy to predict urban LULC classes. In addition, because the MLR model is built with high accuracy, urban LULC allocation is conducted with high accuracy as well, which can allow researchers to interpret the county's future urban growth. These are the noteworthy elements of this study. In fact, the satellite imagery classifications, the MLR model building, and urban LULC allocation in this study have potential to be widely applied in ways that benefit both the public and the private sectors.

LIST OF REFERENCES

Ainsworth. (no date). *Logistic Regression: Continued. Psy 524. Course Presentation*. Retrieved December 30, 2009, from http://www.csun.edu/~ata20315/psy524/docs/Psy524%20lecture%2019%20logistic_cont.ppt.

Almeida, C.M., Monteiro, A.M.V., and Câmara, G. (2003). Modeling the Urban Evolution of Land Use Transitions Using Cellular Automata and Logistic Regression. In *Geoscience and Remote Sensing Symposium, IGARSS '03, Proceedings, 2003 IEEEInternational. 21-25 July 2003* (pp. 1564- 1566).

Anderson, J.R., Hardy, E.E., Roach, J.T., and Witmer, R.E. (1976). *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*. Geological Survey Professional Paper 964. United States Government Printing Office, Washington.Retrieved October 26, 2008, from http://landcover.usgs.gov/pdf/anderson.pdf.

Banman, C. (no date). *Supervised and Unsupervised Land Use Classification*. Retrieved October 31, 2009, from http://www.emporia.edu/earthsci/student/banman5/perry3.html.

Batty, M. and Longley, P. (1994), *Fractal Cities: A Geometry of Form and Function.* London: Academic Press.

BEBR (2003). *Florida Statistical Abstract 2003.* Bureau of Economic and Business Research: Warrington College of Business Administration. Gainesville: The University of Florida Press

BEBR (2009, March). *Florida Population Studies: Projections of Florida Population by County, 2008-2035*. Bureau of Economic and Business Research: Warrington College of Business Administration. 42(153): 1-8. Gainesville: The University of Florida Press

Benfield F.K., Raimi M.D., and Chen D.D.T. (1999). *Once There Were Greenfields: How Urban Sprawl Is Undermining America's Environment, Economy, and Social Fabric*. Natural Resources Defense Council.

The Boston Indicators Project (no date). Car ownership and Vehicle Miles Traveled, Boston and Metro Boston. Retrieved October 24, 2008, from http://www.bostonindicators.org/indicatorsproject/transportation/indicator.aspx?id=2056

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. Belmont, Calif: Wadsworth International Group.

Broos, M.J. and Day, R. (no date). *Measuring Urban Sprawl and Predicting Land Use Change using Geospatial Technologies.* Retrieved October 25, 2008, from http://lal.cas.psu.edu/Research/sprawl.asp

Bureau of Transportation Statistics. (2002). *National Transportation Statistics 2001 (July, 2002).* U.S. Department of Transportation. Retrieved December 21, 2008, from http://answers.google.com/answers/threadview?id=146894

Carr, M.H. and Zwick, P.D. (2007). *Smart Land-Use Analysis: The LUCIS Model*. Redlands, CA: ESRI Press.

Census Bureau (1980, 1990, 2000, 2010). *Profile of General Population and Housing Characteristics*. Retrieved September 18, 2011 from http://www.census.gov.

Cieslewicz, D.J. (2002). The Environmental Impacts of Sprawl. In Squires, G.D. (Ed.) Urban Sprawl: Causes, Consequences & Policy Responses (p.23 – 38). Washington, D.C.: The Urban Institute Press.

Cowen, D.J., Jensen, J.R., Bresnahan, G., Ehler, D., Traves, D., Huang, X., Weisner, C., and Mackey, H.E. (1995). The Design and Implementation of an Integrated GIS for Environmental Application. *Photogrammetric Engineering & Remote Sensing*, 61(1): 1393-1404

Cramér, H. (1999). *Mathematical Methods of Statistics.* Princeton: Princeton Press.

Dewey, J.F. and Denslow, D.A., (2001). *Growth and Infrastructure in Alachua County: Does Conventional Development Pay Its Share of Public Costs*? Bureau of Economic and Business Research, University of Florida.

Downs, A. (1994). *New Visions for Metropolitan America*. Washington DC: The Brookings Institution.

Duany, A., Plater-Zyberk, E., and Speck, J. (2000). *Suburban Nation*. New York: North Point Press.

Field, A. (2000). *Discovering Statistics using SPSS for Windows.* London, Thousand Oaks, Dew Delhi: Sage Publications.

Garson, G. D. (2009). *Nominal Association: Phi, Contingency Coefficient, Tschuprow's T, Cramer's V, Lambda, Uncertainty Coefficient*. From Statnotes, North Carolina State University, Public Administration Program, Retrieved December 30, 2009, from http://faculty.chass.ncsu.edu/garson/PA765/assocnominal.htm.

Herold, N.D., Koeln, G. and Cunnigham, D. (2003). Mapping Impervious Surfaces and Forest Canopy Using Classification and Regression Tree (CART) Analysis. *ASPRS 2003 Annual Conference Proceedings.* May 2003. Anchorage, Alaska. Retrieved April 19, 2008, from http://www.nemo.uconn.edu/tools/impervious_surfaces/pdfs/Herold_etal_2003.pdf

Hosmer, D.W., Jr. and Lemeshow, S. (1989). *Applied Logistic Regression.* Wiley Series in Probability and Mathematical Statistics. New York, Chichester, Brisbane, Toronto, Singapore: A Wiley-Interscience Publication.

Hu, Z. and Lo, C.P. (2007); Modeling Urban Growth in Atlanta Using Logistic Regression. *Computers, Environment and Urban Systems*, 31(6): 667-688.

Huang, X. and Jensen, J.R. (1997). A Machine-Learning Approach to Automated Knowledge-Base Building for Remote Sensing Image Analysis with GIS Data. *Photogrammetric Engineering & Remote Sensing*, 63(10): 1185-1194.

Hung, M.C. (2003). Remote Sensing and GIS for Urban Environmental Modeling, Monitoring and Visualization. (Doctoral Dissertation. University of Utah. 2003).

Hung M-C and Ridd, M.K. (2002). A Subpixel Classifier for Urban Land-Cover Mapping Based on a Maximum-Likelihood Approach and Expert System Rules. *Photogrammetric Engineering & Remote Sensing*. 68(11): 1173-1180.

IDRISI Help. (no date). *IDRISI Andes*. Clark Labs.

Jackson, K.T. (1985). *Crabgrass Frontier*: The Suburbanization of the United States. New York, Oxford: Oxford University Press.

Jargowsky, P.A. (2002). Sprawl, Concentration of Poverty, and Urban Inequality. In Squires, G.D. (Ed.) *Urban Sprawl: Causes, Consequences and Policy Responses*. (p.39 –117). Washington, DC: Urban Institute.

Jensen, J.R. and Toll, D. (1982). Detecting Residential Land-Use Development at the Urban Fringe. *Photogrammetric Engineering & Remote Sensing*, 48(4): 629-643.

Jensen, J.R. and Cowen, D.C. (1999). Remote Sensing of Urban/Suburban Infrastructure and Social Economic Attributes. *Photogrammetric Engineering & Remote Sensing*, 65(5): 611-622.

Jensen, J.R. (2005). *Introductory Digital Imaging Processing: A Remote Sensing Perspective* (Third Edition). Upper Saddle River, New Jersey: Pearson Prentice Hall.

Kahn, M.E. (2006). *The Benefits of Sprawl*. Environmental and Urban Economics. Thoughts on Environmental and Urban Issues from an Economics Perspective. Retrieved December 6, 2009, from http://greeneconomics.blogspot.com/2006/03/benefits-of-sprawl.html.

Kim, H.J. (2007). *Spatiotemporal Analysis of Urban Growth in Local Communities*. (Doctoral Dissertation. Kent State University. 2007).

Kramber, W.J. and Morse, A. (1994). *Integrating Image Interpretation and Unsupervised Classification.* Retrieved October 31, 2009, from http://libraries.maine.edu/Spatial/gisweb/spatdb/acsm/ac94037.html.

Landis, J.D. (1994). The California Urban Futures Model: A New Generation of Metropolitan Simulation Models. Environment and Planning B: Planning and Design. 21(4): 399-420.

Landis, J.D. (1995). Imagining Land Use Futures: Applying the California Urban Futures Model. Journal of the American Planning Association. 61(4): 438-457.

Landis, J.D. and Zhang, M. (1997). Modeling Urban Land Use Change: the Next Generation of the California Urban Futures Model. Paper Presented at the National Center for Geographic Information and Analysis (NCGIA) the 1997 Land Use Modeling Workshop. Retrieved December 30, 2009, from http://www.ncgia.ucsb.edu/conf/landuse97/papers/landis_john/paper.html.

Lawrence, R. and Wright, A. (2001). Rule-Based Classification Systems Using Classification and Regression Tree (CART) Analysis. *Photogrammetric Engineering & Remote Sensing*, 67(10): 1137-1142.

Levy, J. (2003). Contemporary Urban Planning (Six Edition). Upper Saddle River, New Jersey: Prentice Hall.

Lewis, R.J. (2000). *An Introduction to Classification and Regression Tree (CART) Analysis.* Presented at the 2000 Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California. Retrieved April 19, 2008, from http://www.saem.org/download/lewis1.pdf.

LCI (Land Cover Institute). (2007). *NLCD Land Cover Class Definitions.* Retrieved October 26, 2008, from http://landcover.usgs.gov/classes.php#herb

Liou, C-Y and Yang, K-D.O. (no date). Unsupervised Classification of Remote Sensing *Imagery with Non-negative Matrix Factorization.* Retrieved October 31, 2009, from http://red.csie.ntu.edu.tw/publications/ICONIP05_Unsupervised%20Classification%20of%20Remote%20Sensing%20Imagery.pdf.

Lo, C.P. and Choi, J. (2004). A Hybrid Approach to Urban Land Use/Cover Mapping Using Landsat 7 Enhanced Thematic Mapper Plus (ETM+) Images. *International Journal of Remote Sensing.* 25:14, 2687-2700.

Lu, D. and Weng, Q. (2004). Spectral Mixture Analysis of the Urban Landscape in Indianapolis with Landsat ETM+ Imagery. *Photommetric Engineering & Remote Sensing*, 70(9): 1053-1062.

Lucy, W.H. and Phillips, D.L. (2006). *Tomorrow's Cities, Tomorrow's Suburbs*. Chicago, Illinois; Washington, DC: American Planning Association.

Menard, S. (1995). *Applied Logistic Regression Analysis*. Thousand Oaks, London, New Delhi: Sage Publications.

McFadden, D. S. (1973). *Conditional Logit Analysis of Qualitative Choice Behavior*. In Zarembka, P. (Ed.), Frontiers in Econometrics (p.105-142). New York: Academic Press.

Mowrer, H.T. and Congalton, R.G. (2000). *Introduction: the Past, Present, and Future of Spatial Uncertainty Analysis*. In Mowrer, H.T. and Congalton, R.G.(Ed.) *Quantifying Spatial Uncertainty in Natural Resources.*(p.xv-xxiv). Chelsea, Michigan: Ann Arbor Press.

Mumford, L. (1961). *The City in History: Its Origins, Its Transformations, and Its Prospects*. San Diego, New York, London: Harcourt Brace & Company.

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.

Myint, S.W., Wentz, E.A., and Purkis, S.J. (2007). Employing Spatial Metrics in Urban Land-Use/Land-Cover Mapping: Comparing the Getis and Geary Indices. *Photogrammetric Engineering & Remote Sensing*, 73(12): 1403-1415.

NCGIA. (no date). *Meeting Summary: the State of the Art of Land use Modeling.* Retrieved October 28, 2009, from http://www.ncgia.ucsb.edu/conf/landuse97/summary.html

Paul, O.V.D. (2007). *Remote Sensing: New Applications for Urban Areas*. Proceedings of the IEEE. Volume 95, Issue 12, Dec. 2007, 2267 – 2268.

Phinn, S., Stanford, M., Scarth, P., Murray, A.T., and Shyy, P.T. (2002). Monitoring the Composition of Urban Environments based on the Vegetation-Impervious Surface-Soil(VIS) Model by Subpixel Analysis Techniques. *International Journal of Remote Sensing*. 23(20): 4131-4153.

Poelmans, L. and Rompaey, A.V. (2010). Complexity and Performance of Urban Expansion Models. *Computers, Environment and Urban Systems.* 34(1): 17-27.

Pontius, R.G., Jr. and Schneider, L.C. (2001). Land-Cover Change Model Validation by an ROC Method for the Ipswich Watershed, Massachusetts, USA. *Agriculture, Ecosystems & Environment.* 85(1-3): 239-248.

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 Remote Sensing.* 16(2): 2165-2185.

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

Seaman, M. (2001). *Categorical Data*. Retrieved December 30, 2009, from http://edpsych.ed.sc.edu/seaman/edrm711/questions/categorical.htm.

Sirkin, R.M. (1999). *Statistics for the Social Sciences (2 Edition)*. Thousand Oaks, London, New Delhi: Sage Publications.

Short, N.M., Sr. (no date). *Supervised Classification.* End to End Remote Sensing Tutorial Page 1-17. RST. Retrieved October 31, 2009, from http://rst.gsfc.nasa.gov/Sect1/Sect1_17.html.

Squires, G.D. (2002). *Urban Sprawl and the Uneven Development of Metropolitan American*. In Squires, G.D. (Ed.) Urban Sprawl: Causes, Consequences & Policy Responses (p.1 – 22). Washington, D.C.: The Urban Institute Press.

Toll, D.L. (1984). An Evaluation of Simulated Thematic Mapper Data and Landsat MSS Data for Discriminating Suburban and Regional Land Use and Land Cover. *Photogrammetric Engineering & Remote Sensing*, 50(12): 1713-1724.

Verburg, P. (no date). *The CLUE-S model: Tutorial CLUE-s (version 2.4) and DYNACLUE (version 2).* Wageningen University, Netherlands. Retrieved August 21, 2009, from http://www.cluemodel.nl/.

Verburg, P.H., Eickhout, B., and Meijl, H.V. (2007). A Multi-Scale, Multi-Model Approach for Analyzing the Future Dynamics of European Land Use. *The Annals of Regional Science*. 42(1): 57-77.

Ward, D., Phinn, S.R. and Murray. A.T. (2000). Monitoring Growth in Rapidly Urbanizing Areas Using Remotely Sensed Data. *Professional Geographer*. 52(3): 371-386.

Wassmer, R.W. (2001). *The Connection between Local Government Finance and the Generation of Urban Sprawl in California*. Retrieved December 6, 2009, from http://www.csus.edu/calst/government_affairs/reports/ffp44.pdf.

Yang, X. (2000). *Integrating Image Analysis and Dynamic Spatial Modeling with GIS in a Rapidly Suburbanizing Environment*. (Doctoral Dissertation. University of Georgia. 2000).

Zambon, M., Lawrence, R., Bunn, A., and Powell, S. (2006). Effect of Alternative Splitting Rules on Image Processing Using Classification Tree Analysis. *Photogrammetric Engineering & Remote Sensing*, 72(1): 25-30.

BIOGRAPHICAL SKETCH

Yong Hong Guo was born in Shanghai, China, in 1968. He graduated from Tongji University in Shanghai, China, in 1991, holding a bachelor's degree in urban planning. In 1999, he graduated from the Virginia Polytechnic Institute and State University (Virginia Tech) having a master's degree in urban planning. In 2006, he enrolled in the Urban and Regional Planning Department at the University of Florida to pursue the degree of Doctor of Philosophy in urban planning. He received his Ph.D. from the University of Florida in the spring of 2012. He has been a member of the American Institute and Certified Planners (AICP) since 2007.