Forest Change Dynamics Across Levels of Urbanization in the Eastern US

Yi-Jei Wu

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science

In

Forest Resources and Environmental Conservation

Valerie A. Thomas (Co-Chair)

Robert D. Oliver (Co-Chair)

John A. McGee

Michael G. Sorice

August 14, 2014

Blacksburg, VA

Keywords: nation land cover database (NLCD), Landsat, multitemporal, change detection

Forest Change Dynamics Across Levels of Urbanization in the Eastern US

Yi-Jei Wu

Abstract

The forests of the eastern United States reflect complex and highly dynamic patterns of change. This thesis seeks to explore the highly variable nature of these changes and to develop techniques that will enable researchers to examine their temporal and spatial patterns. The objectives of this research are to: 1) determine whether the forest change dynamics in the eastern US differ across levels of the urban hierarchy; 2) identify and explore key micropolitan areas that deviate from anticipated trends in forest change; and 3) develop and apply techniques for 'Big Data' exploration of Landsat satellite images for forest cover analysis over large regions.

Results demonstrate that forest change at the micropolitan level of urbanization differs from rural and metropolitan forest dynamics. The work highlights the dynamic nature of forest change within the Piedmont Atlantic megaregion, largely attributed to the forestry industry. This is by far the most dominant change phenomenon in the region – but is not necessarily indicative of permanent forest change. A longer temporal analysis may be required to separate the contribution of the forest industry from permanent forest conversion in the region.

Techniques utilized in this work suggest that emerging tools that provide supercomputing/parallel processing capabilities for the analysis of 'big' satellite data open the door for researchers to better address different landscape 'signals' and to investigate large regions at a high temporal and spatial resolution. The opportunity now exists to conduct initial assessments regarding spatio-temporal land cover trends in the southeast in a manner previously not possible.

Dedication

I want to dedicate this thesis to my family. Thank you for your love, encouragement and support.

Acknowledgements

First, special thank you to my advisors: Valerie A. Thomas and Robert D. Oliver. I appreciate them both for giving me this opportunity and for their guidance and patience throughout this process.

I would also like to thank my committee as a whole, which includes Valerie, Robert, John A. McGee, and Michael G. Sorice. I thank each of these professors for the patience, enthusiasm and expertise that they brought to this project. The constructive feedback and insight that they provided were invaluable to this research.

A special thanks to the Google for providing me a valuable opportunity to utilize the Google Earth Engine platform as a trusted tester for my research.

Finally, I would like to thank the entire Department of Forest Resources and Environmental Conservation at Virginia Tech for providing me lots of assists. It has been a privilege to study and work among the faculty, staff and students of this department.

Attribution

My advisors, Drs. Thomas and Oliver, helped and guided behind the research presented in this thesis.

Valerie A. Thomas is currently a professor in forest resources and environmental conservation department at Virginia Tech. Robert D. Oliver is currently a professor in geography department at Virginia Tech. Both of them were the co-authors on this thesis, aided to frame the overall questions and improve the communicative effort.

Table of Contents

List of Figures

List of Tables

Chapter 1 – Introduction

1. Introduction

In 2012, Rutkow (2012, 345) sought to make the "nation's treescape more legible" in his seminal book *American Canopy*. Covering more than 400 years of history, *American Canopy* skillfully outlines the relationship between America's growth and prosperity and the country's tree species. The book highlights that "[t]he language we use to describe forests has shifted immensely" reflecting both attitudes and ideologies (Rutkow 2012, 346). As fiber and fuel, trees have sheltered, transported and clothed Americans. Whether thought of in singular or plural, tree(s) have provided places of recreation and reflection, have been the subject of commodification and conservation, and served as crucial symbols of our character and culture. Central to *American Canopy* and other research efforts is the story of the nation's continental forest loss, from a staggering one billion acres (estimated) to a low of 600 million acres by the early twentieth century (Houghton and Hackler 2000; Loveland and Acevedo 2006; Napton et al 2010; Rutkow 2012). In particular, the USGS Land Cover Trends Project conducted research concerning the land cover change rates, trends, causes, and consequences for the United States for the post 1973 period (Loveland and Acevedo 2006; Napton et al 2010).Although recent figures illustrate that a forest 'revival' has occurred with roughly 750 million acres now covering the continental US, Rutkow (2012, 346) reminds us that "forest quantity is not the same as forest quality." While Rutkow nicely illustrates how forests helped make suburbs, cities and America possible, his aim was to broadly explain the historical utility and necessity of our forest stands, not to offer a specific judgment on a region, pattern or process.

This thesis has a more narrow focus but it derives motivation from one of the concluding sentences of *American Canopy*: "Trees appear frozen in time, and the invisibility of gradual change can make problems difficult to spot." For some time, researchers have been monitoring the extent (and health) of America's forests and have sought to document change(s) using a variety of methods. A key challenge in remote sensing, forestry and urban research is to develop a mean to spot an underlying signal of forest change in urban counties over different time periods across various spatial scales (from megaregions to micropolitan areas). By employing remote sensing, statistical analysis, and change detection methods, this thesis explores the highly variable nature of forest change in the eastern United States and offers an approach that will enable researchers to more fully examine the spatial and temporal patterns of forest change.

1.1 Thesis Objectives

The objectives of this research are to:

(i) determine whether the forest change dynamics in the eastern US differ across levels of the urban hierarchy (i.e., from rural → micropolitan → metropolitan → megaregion); (Chapter 2)

(ii) identify and explore key micropolitan areas that deviate from anticipated trends in forest change patterns based on prior land cover change research for this region; (Chapter 2)

(iii) develop and apply techniques for 'Big Data' exploration of Landsat satellite images for forest cover analysis over large regions (Chapters 2 and 3)

1.2 Thesis Structure and Attribution

This thesis is divided into five chapters. The present chapter provides context and addresses the issue of research objectives and motivation. The second chapter discusses three broad conceptual issues that have relevancy to the remote sensing community. Distinguishing between land cover and land use and illustrating the definitional complexity of the word forest, are the first tasks in Chapter Two. The chapter ends with a discussion of the computational challenge of big data for the analysis of land cover change. Chapter Three is a manuscript prepared for a special issue on Micropolitan America for the *Southeastern Geographer* (expected publication date is Winter 2014). I conducted the analysis (in Chapter 3) using Landsat NLCD and Google Earth Engine Imagery. Chapter Four, and the appendix, describe in more detail the Google Earth Engine analysis that was used to support spatial analytical techniques. This chapter also presents additional results that did not fit into the scope of the special issue. Chapter Five is a short summary chapter highlighting the overall conclusions and the importance of the work.

Literature Cited

Houghton, R.A. and Hackler, J.L. 2000. Changes in terrestrial carbon storage in the United States. 1: The roles of agriculture and forestry. *Global Ecology and Biogeography* 9: 125 – 144.

Loveland, T.R. and Acevedo W. 2006. Land-cover change in the eastern United States. United States Geological Survey. http://landcovertrends.usgs.gov/eastResults.html

Napton, D.E., Auch, R.F., Headley, R. and Taylor, J.L. 2010. Land changes and their driving forces in the Southeastern United States. Regional Environmental Change 10(1): 37 – 53.

Rutkow, E. 2012. American Canopy: Trees, Forests, and the Making of a Nation.

Chapter 2 – Conceptual Background

1. Land Cover versus Land Use

When conducting and interpreting land cover analysis, there is the potential for confusion between the concept of land cover and land use, which have been convolved in classification schemes for many years (Anderson et al. 1976). This convolvement has increased in recent years and has led to some misinterpretation of results from analyses and monitoring programs that draw inference from land classifications (Fisher and Unwin 2005). Di Gregorio and Jansen (2000) mention that there are few land classification schemes that employ solely either land cover or land use in their definition of classes.

Land cover and land use are fundamentally different. While land cover denotes the physical state of the land and is determined by the direct observation of the earth's surface, land use denotes the human use of the land and can be defined as socio-economic references for human activities on the land surface. Land cover classes are usually defined by straight interpretations of remotely-sensed imagery for the primary purpose of natural science, for example: ecology, environmental science, etc. Land use classifications are based on the surveys from planning authorities and can be used for the sake of planning human or the natural environment. Another way of thinking about this is that there is a many-to-many relationship that exists between the two: the same land use type could be in disparate land cover composites and various land use classes could be composed of the same land cover; on top of that, not every same land use type certainly corresponds to the same land cover type (Fisher and Unwin 2005).

Why is this challenging for forest change analysis? There are over 800 accepted definitions of forests and wooded areas in the literature (Lund 2007). Consider the case of forest change in the southeastern United States, where forestry and logging have big impacts on the landscape. In a land cover classification scheme, after a forest has been clearcut, it is no longer considered to be forest, even if it has been replanted (as is the case with the National Land Cover Database classification scheme for forests). More likely, it will be classified as either grassland/herbaceous or barren depending on the condition of the site. Depending on the timing of the harvesting cycles, this could mean that some areas show significant decreases in the amount forest for that time. In contrast, when the replanted clearcut grows back into a forest, a significant gain in forested land might be seen.

In a land-use classification scheme, if the clearcut is intended to be regrown into a forest it is still considered a forest. One could consider a clearcut to be the first successional stage. From this perspective, the cyclical loss and gain of forest lands due to harvesting and regrowth is not considered to be “change”. A land use classification of forests would show very different results for change over time.

The importance of a careful consideration of definitions is strikingly obvious when considering the Global Forest Change Product by Hansen et al (2013), which shows a patchwork style of loss and gain in southeast forests. This product is a continuous canopy change product – using a land cover definition. It shows the southeastern US to be one of the most dynamic areas in the world for land cover change, largely attributed to forest harvesting by the forestry industry. This is by far the most dominant change signal in the region – but is not necessarily indicative of permanent forest loss or gain.

2. Definition of Forest Class

Given the confusion between land use and land cover, and the hundreds of definitions of forests, there are no specific guidelines on the most appropriate technique to classify remote sensing imagery or geographic information system (GIS) polygons as forest land. As a result, various forest definitions have been adopted across various databases, programs, organizations, administrative unit (i.e., state, nation), and ecological or miscellaneous properties (Lund 2007).

The Food and Agriculture Organization (FAO) of the United Nations (UN) began to publish the assessments for the world's forest resources every 5 to 10 years since the 1940s in order to provide consistent data that describes the world's forest and its change. Primary data sources are the country reports by National Correspondents and remotely-sensed images. However, there are two problems: the first one is that country boundaries shift over time (particularly since 1940) and secondly, the definition of forest land changes through time and tree characteristics – for instances, tree cover area, tree height, and tree type. According to the FAO, forests are land surface where more than 0.5 hectares with 10% or more tree canopy cover and trees higher than 5m (or having the potential of reaching 5m). However, the National Land Cover Database (NLCD) that we used in this research has a totally different forest definition. According to the NLCD, forests are areas characterized by tree cover (natural or semi-natural woody vegetation, generally greater than 5 meters tall); tree canopy accounts for more than 20% of the vegetation cover (NLCD, 2001). The type of vegetation that qualifies as forest is different from country to country. For example, the 2010 Country Report of Vietnam for UN Food and Agriculture Program states that areas that are dominated by bamboo are considered as forest in Vietnam, but bamboo is considered a grass in many other countries (UNFAO 2010).

Because the myriad of factors that area taken into account when defining forests, the choice and application of forest definitions by researchers can result in divergent outcomes. Romijn (2013) proves that the selection of forest definition can have a great impact on the estimates of forest cover, deforestation rate and forest degradation areas; moreover, it can also affect the assessments of potential drivers of deforestation as well as carbon emission, and the development of reference emission level (GOFC-GOLD 2012; Magdon and Kleinn 2012; Romijn 2013; Azuma 2014).

The appropriate definition of a forest is clearly context-specific. There is neither a globally accepted definition, nor full agreement among scientists regarding how to define forests. However, the IPCC recommends using the definition given by the FAO which is becoming more internationally accepted for global research (IPCC 2006). It should be noted that in the United States, the forest assessment report was conducted by the United States Department of Agriculture (USDA) Forest Services, Forest Inventory and Analysis program (FIA) using a forest definition that is different from the FAO. No matter how one defines forest land, it is vital that the implications of excluding or including different patches of trees within the forest class are carefully considered.

3. Computational Challenges for Land Cover Change Analysis

Performing a regional analysis of change detection at high resolutions requires an enormous amount of data and preprocessing, making tailored analysis very challenging for researchers. The Landsat 7 satellite archive is divided into “scenes” which fall along the orbit path, and cover an area of roughly 185 km across-track by 180 km along-track. Although the actual size of an image (in terms of computer storage) will vary depending on the specific format and data type, a single scene is typically about 450 MB. This is, of course, raw data without any preprocessing, classification, or index calculation which could double or triple the size of the data in some cases. Also, this is for a single date and by definition change detection involves at least two dates, but often more. For context, even a subset of the region of interest under investigation for this thesis, such as Piedmont Atlantic Megaregion discussed in Chapter Two, requires at least 16 Landsat scenes to cover the entire region and a minimum of three dates (2001, 2006, and 2011). This equates to roughly 2 TB of raw data, without even conducting any analysis. Such computational requirements have restricted researchers into using either: 1) products already generated, such as: the National Land Cover Database product (produced for 1992, 2001, 2006, and 2011) – which may have an unsuitable classification scheme for the research question; 2) coarse resolution products (such as those derived from the MODIS satellite at about 1km spatial resolution) – which may obscure localized change; 3) very narrowly constrained research questions; or 4) some high performance solutions but complicated parallel processing and supercomputing.

As part of my thesis research, I requested and obtained “Google Trusted Tester” authentication which enabled access to the classification and image analysis capabilities from the Graphical User Interface (GUI) of Google Earth Engine “Beta” (discussed in Chapters Two and Three of this thesis). Google maintains the largest Landsat satellite archive outside of the United States Geological Survey (USGS), including full coverage of North America amongst other global regions, as well as different types of ‘preprocessed’ data. These data can be traced back nearly 40 years and are available online via the Google Earth Engine platform which includes extremely valuable tools for researchers in this infancy stage of ‘Big Data’ science. The Google Earth Engine platform enables independent researchers, scientist groups, and larger groups to tailor regional analysis in a more efficient manner, yet also ensures that such investigations remain computationally feasible from a desktop computer. A knowledgeable and enthusiastic user group, including Google employees, remote sensing scientists, other computing specialists and students, are also part of a community email listserv which discusses potential issues and ideas, which I made full use of throughout my research.

Literature Cited

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. *U.S. Geological Survey, Professional Paper* 964, p.28, Reston, VA.

Azuma, D.L. and Gray, A. 2014. Effects of changing forest land definitions on forest inventory on the West Coast, USA. *Environ Monit Assesss* 186: 1001 – 1007.

De Gregorio, A. and Jansen, L.J.M. 2000. Land Cover Classification System (LCCS): Classification concepts and user manual. *Environment and Natural Resources Service (SDRN)*, FAO, Rome.

Fisher, P. and Unwin, D. 2005. Land use and Land cover: Contradiction or Complement in *Re-Presenting GIS* pp.85 – 98. Chichester: Wiley.

GOFC-GOLD 2012. A Sourcebook of Methods and Procedures for Monitoring and Reporting Anthropogenic Greenhouse Gas Emissions and Removals Associated with Deforestation, Gains and Losses of Carbon Stocks in Forests Remaining Forests, and Forestation. *GOFC-GOLD Report* Version COP 18- 1. GOFC-GOLD Land Cover Project Office, Wageningen University, The Netherlands.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O. and Townshend, J.R.G. 2013. High-resolution global maps of 21st-century forest cover change. Science 342: 850 – 853.
http://earthenginepartners.appspot.com/science-2013-global-forest.

IPCC 2006. *2006 IPCC Guidelines for National Greenhouse Gas Invertories*. IGES, Japan, Prepared by the National Greenhouse Gas Inventories Programme.

Lund, H. G. 2007. Definitions of Forest, Deforestation, Afforestation, and Reforestation. *http://home.comcast.net/~gyde/DEFpaper.htm*

Magdon, P. and Kleinn, C 2012. Uncertainties of forest area estimates caused by the minimum crown cover criterion – a scale issue relevant to forest cover monitoring. *Environmental Monitoring and Assessment* 1 – 16.

National Land Cover Database (NLCD) 2001. U.S. Geological Survey, U.S. Department of the Interior. *http://www.mrlc.gov/nlcd11_leg.php*

Romijn, E., Ainembabazi, J.H., Wijaya, A., Herold, M., Angelsen, A, Verchot, L and Murdiyarso, D. 2013. Exploring different forest definitions and their impact on developing REDD+ reference emission levels: A case study for Indonesia. *Envrionmental Science and Policy* 33: 246 – 259.

United Nations Food and Agriculture Program (2010). Global forest resources assessment, FRA 2010. Country Report, Vietnam. *http://www.fao.org/docrep/013/al664E/al664e.pdf*

Chapter 3 – Forest Change Dynamics across Levels of Urbanization in the Eastern US

Yi-Jei Wu, V.A. Thomas, and R.D. Oliver

Submitted to the *Southeastern Geographer*

Abstract

Extensive urbanization of the Eastern Seaboard has meant that the forests of the eastern United States now reflect complex and highly dynamic patterns of change depending on the spatial and temporal scale of analysis. Although land cover and forest change in the US has been studied by numerous authors, forest change within political or administrative units, and along urbanization gradients, requires additional consideration. The objectives of this research are to: (i) determine whether the forest change dynamics in the eastern US differ across levels of the urban hierarchy (i.e., from rural → micropolitan → metropolitan → megaregion); and (ii) identify and explore key micropolitan areas that deviate from anticipated trends in forest change patterns based on prior land cover change research for this region. Results highlight the dynamic nature of forest change within the Piedmont Atlantic megaregion and illustrate the highly variable nature of forest change within micropolitan regions.

1. Introduction

The world's temperate deciduous forests are host to some of the most altered landscapes on the planet with less than 1% of the temperate deciduous forests in Europe and North America remaining undisturbed (Reich and Frelich 2002). In North America, the extensive urbanization of the Eastern Seaboard (from northern New England to central Florida and west to the Mississippi) has produced a legacy of large scale deforestation and a general increase in landscape fragmentation (Steyaert and Knox 2008). The competition for land use has meant that the forests of the eastern United States now reflect complex and highly dynamic patterns of change. Land cover change patterns in the southeastern United States have been evident since at least 1950 (Napton et al. 2010). More recently, the high resolution of 2000 to 2012 Global Forest Cover Change map produced by Hansen et al. (2013) describes the southeastern United States as a region of high variability with both significant forest loss and forest gain (often the result of planting/harvesting cycles). Although the 2000-2012 Global Forest Change product provides valuable insights into the dynamics of this region at a fine spatial unit (i.e., 30m), the change patterns over coarser spatial units, including political or administrative units, require additional consideration.

Despite the fact that land use trends in the United States have received considerable attention, most of the literature focuses on exploring the trends in the scale of nation, ecoregion (Drummond 2010; Napton et al. 2010), state, and county (Waisanen and Bliss 2002). Another way of thinking about forest change is to explore trends based on levels of urbanization. This could be done based on the census designations of Core Based Statistical Areas (CBSA) from rural to micropolitan to metropolitan areas. Large portions of the nation can also be grouped into megaregions, which have shared natural resources and ecosystems, and linked economic and transportation systems (America2050, 2011). By shifting the spatial context we argue that there is the opportunity to develop additional insights on forest change dynamics in the eastern United States. The objectives of this research are to: (i) determine whether the forest change dynamics in the eastern US differ across levels of the urban hierarchy (i.e., from rural → micropolitan → metropolitan → megaregion); and (ii) identify and explore key micropolitan areas that deviate from anticipated trends in forest change patterns based on prior land cover change research for this region.

2. Background

Millions of hectares of forested land are projected to be converted to developed areas and other land uses over the coming decades. It is not surprising to find that researchers are interested in identifying critical locations of conversion as well as developing models to quantify and project the scale and intensity of change (Griffith et al. 2003). As Drummond and Loveland (2010, p 286) highlight, "land-use changes are recognized as fundamentally important to understanding a range of ecological, biophysical, social, and climate consequences." By combining remotely-sensed data, statistical sampling, and change detection methods, researchers have sought to expose the spatial and temporal patterns of land conversion across the United States. It is now widely understood that naturally forested lands in the eastern United States were significantly altered to accommodate the expansion of the human settlement system. Despite the heavy conversion of forest land for crop production, pasture, residential and industrial uses, as well as the cutting of forest for fuel and other uses, forested land remains the dominant land cover type in the eastern US (Napton et al. 2010; Loveland and Acevedo 2006). As such, many current investigations seek to better understand the changes in land cover occurring at the metropolitan fringe (Daniels 1999), rural-urban fringe (Myers and Beegle 1947), exurban zone (Berube et al. 2006) or other similar terms employed to characterize the transition zone between dense urban settlements and rural environments, because these are the sites where forest resources are most likely to be influenced by continued urbanization. As Drummond and Loveland (2010) point out, land-use intensification and urban expansion continue to overshadow the amount of forest gain from initiatives such as the Conservation Reserve Program (CRP). We aim to shift the landscape analysis from ecoregions to concentrate on the grouping of settlements that are transitioning between rural and metro (i.e., the micropolitan regions). Further, we stratify micropolitan regions into megaregions as defined by America 2050. America 2050 has initially identified eleven megaregions in the conterminous United States for pursuit of the new High-Speed Rail mobility system, with four of the megaregions presently located in the eastern United States (Figure 1). America 2050 defines megaregions as regions with a shared ecosystem, natural resources, as well as common infrastructure system (i.e., transportation), and the interlocked economic systems that link these population centers together (American 2050 2011). It is important to note that the megaregion designation is completely separate from the Census Bureau's designations, with each megaregion containing multiple metropolitans, micropolitans, and rural areas. The utility of this stratification is that it helps to limit the confounding effects of ecology and cultural differences on our understanding of the influence of the level of urbanization on forest change.

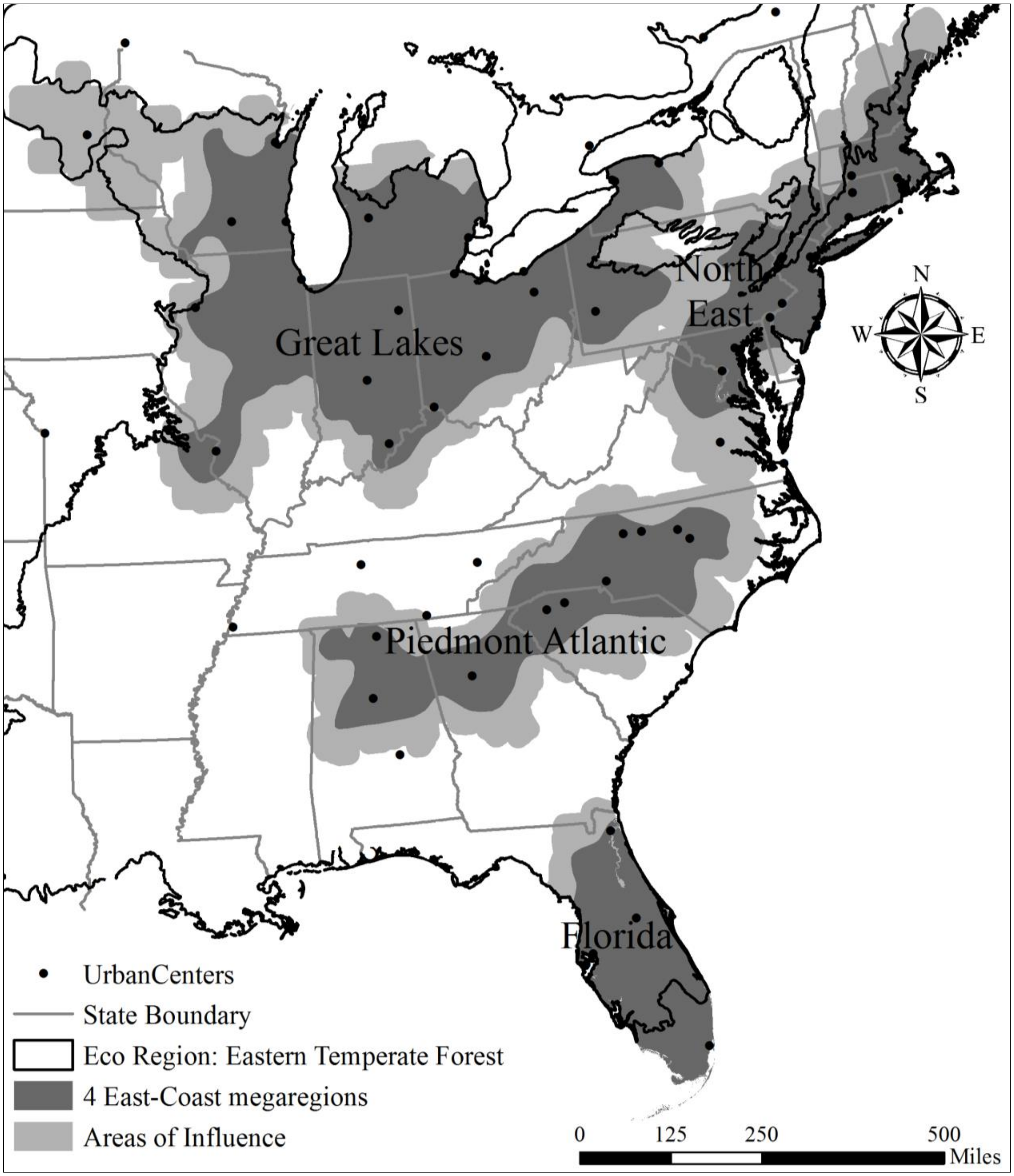

Figure 1. The four East-Coast megaregions and its urban centers as well as influenced area (based on information from America 2050, 2011).

3. Methods

3.1 Census Data and Study Area

Since CBSA designations vary as counties transition from rural to micropolitan to metropolitan (usually in the direction of more urbanization), it is important to use the archived Census Bureau designations appropriate for the time period under study. As such, we used the Census Bureau's 2006 CBSA designations. For the contiguous United States, this amounted to 2264 CBSA designations: including 358 metropolitans, 571 micropolitans, and 1335 non-designated counties.

We selected the four East-Coast megaregions as the study area because they all fall within the same ecological region (i.e., the Eastern Temperate Forest Level I Ecoregion) and as such have broadly similar ecosystems and type, quality and quantity of environmental resources (Omernik 1987, 1995), with the exception of a minor southern portion of the Florida megaregion. Although the total area of these four East-Coast megaregions only accounts for approximately 11% of the total US territory, their combined population represents 40% of the US total population. These four megaregions contain 774 counties and 427 CBSA areas (almost 20% of the national total, Figure 2, Table 1). In general, these four key megaregions can be characterized by increasing fragmentation and a general shift away from land designated for primary resource extraction.

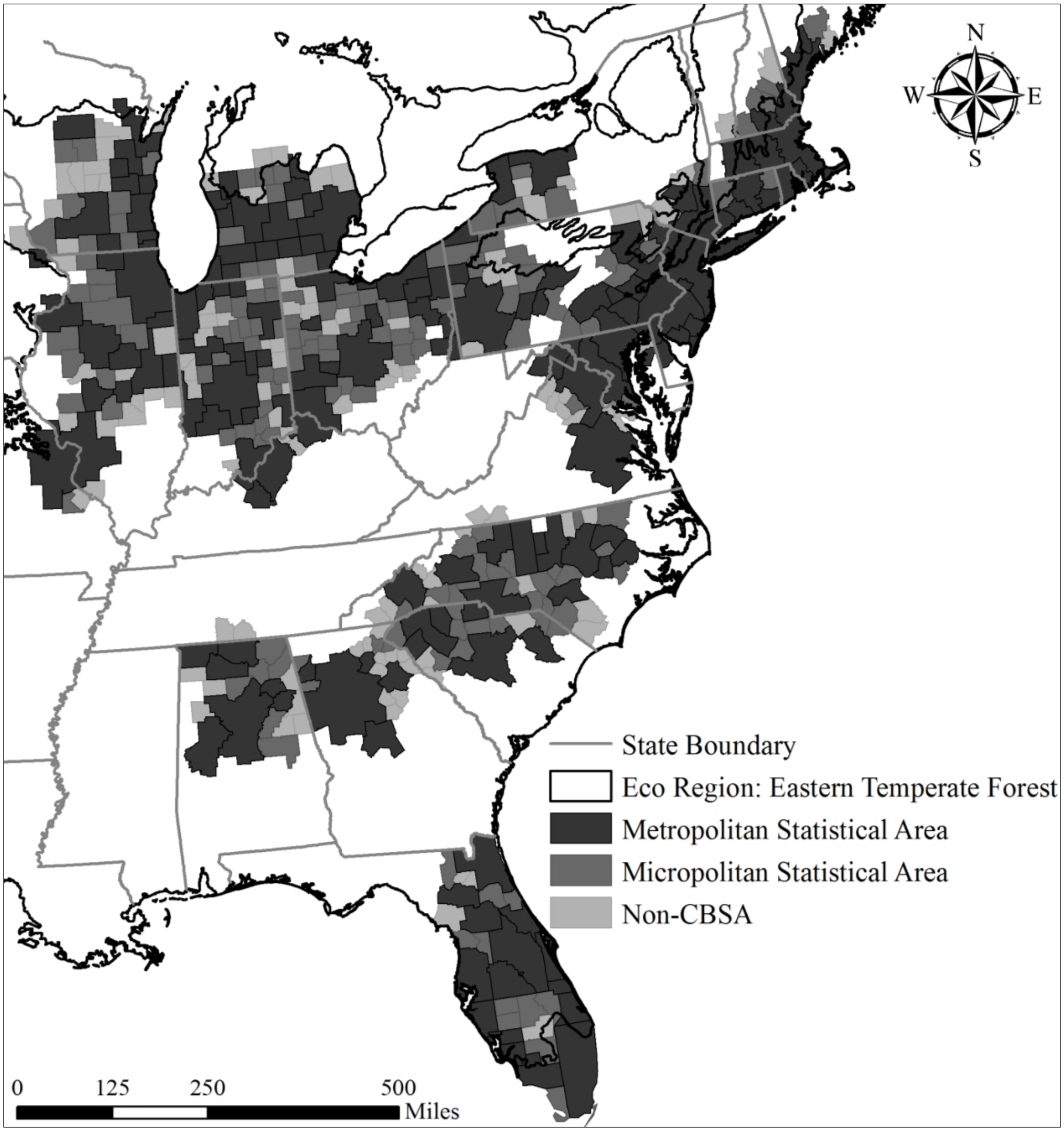

Figure 2. The distribution of the three CBSA categories in the four East-Coast megaregions.

Table 1. Summary table of four East-Coast megaregions.

Megaregion	CBSA Category	CBSA No.	No. County	Area (1000 km^2)
North East	Metropolitan	33	140	148.4
	Micropolitan	14	15	24.1
	Non-designated	18	18	21.8
Great Lakes	Metropolitan	71	212	281.8
	Micropolitan	91	95	147.6
	Non-designated	66	66	89.6
Piedmont Atlantic	Metropolitan	29	106	136.5
	Micropolitan	39	41	57.2
	Non-designated	36	36	44.5
Florida	Metropolitan	15	30	74.9
	Micropolitan	11	11	22.3
	Non-designated	4	4	7.0
		Total	774	1055.7

3.2 Assessment of Forest Change Dynamics

An analysis of satellite imagery is the most effective approach to assess forest change across such broad spatial extents. A number of satellite products could be used, depending on the desired spatial and temporal characteristics of the analysis. We chose Landsat and its derived products for several reasons: 1) the entire Landsat archive was made free to the public in 2008, providing imagery for the entire study region every two weeks for more than 30 years; 2) the imagery has a high spatial resolution (30 m) which allows for an easier interpretation of change results; 3) there exists well vetted national land cover and land cover change products, derived from Landsat imagery, that allow for an assessment of forest dynamics for large regions (i.e., the National Land Cover Database (NLCD) and the Land Cover Change product) for 1992, 2001, 2006 and 2011; and 4) much of the Landsat archive is now hosted on Google Earth Engine, which enables regional analysis of many scenes for research purposes previously not possible. Specifically, our initial analysis utilized the land cover change product from 2001-2006 to assess patterns across the varying levels of the urban hierarchy and to identify key areas for additional investigation. We then increased the temporal resolution of the analysis to assess change over shorter time periods.

3.2.1 Forest Change from 2001 - 2006

We used the National Land Cover Database (NLCD) land-cover change product for 2001-2006 (Fry et al 2011; Xian et al 2009), which was created by comparing the spectral characteristics of Landsat images between 2001 and 2006 to identify and label the changes in accordance with the trajectory from the 2001 NLCD. NLCD is a land cover product produced by the U.S. Geological Survey (USGS), and generated from the Multi-Resolution Land Characteristics (MRLC) consortium that maps consistent land cover information at the national scale for various applications, and is a 16-class land cover classification image based on the classification of Landsat images (ETM+) for the conterminous United States with a 30-meter spatial resolution (Vogelmann et al 2001). There are three forest classes in the classification scheme, which include locations where trees are generally taller than 5 meters and there is more than 20% total vegetation cover. These include: (1) deciduous forest – where over 75% of the tree species shed leaves seasonally; (2) evergreen forest – where over 75% of the tree species hold leaves all year round; and (3) mixed forest – where that neither deciduous nor evergreen forest dominate more than 75%. For our analysis, we combined these 3 classes to create a single forest class called forest.

The NLCD change product contains enough information to consider both total change, as well as the type of change for every class (i.e., prior and new class). To simplify the analysis we looked only at total change in the amount of forest cover (i.e. forest loss, forest gain, no change in forest cover). There are two ways that forest change could be examined for the CBSAs. The first is percent forest change, which reports change relative to the prior (2001) amount of forest cover. While this is a common way to report forest change, this approach fails to reflect the total area that has been impacted. In cases where the prior forest cover is very small, a small change in forest cover might be represented as a large percent change in actual forest. When examining county-scale land cover trends, we believe that it is more meaningful to report the fraction of the county affected by the change of interest. As such, we examined net change from 2001-2006 computed to the level of the rural, micropolitan, and metropolitan CBSA, and normalized to the area of the CBSA, as follows:

$$net\ forest\ change\ (\%\ of\ CBSA) = \left(\frac{2006\ forest\ area - 2001\ forest\ area}{total\ area\ of\ a\ statistical\ area}\right) * 100 \qquad (1)$$

Normalization by area will reduce the apparent forest change, such that only changes that are large relative to the area of the CBSA will be emphasized. Most CBSAs should have a relatively low net change (i.e., relative to the size of the CBSA).

We used the non-parametric Kruskal-Wallis ANOVA by Ranks Test to determine whether a significant difference in forest change exists across the different levels of urbanization (McDonald 2009). Box and whisker plots were created for a visual comparison of forest change, and to identify outliers where forest change patterns deviated from the norm. Basic descriptive statistics characteristics (i.e., mean and standard deviation) for each group were also calculated.

3.2.2 Examination of Outliers at Higher Temporal Resolution

Although the NLCD change product provides significant insight into change patterns at the national level, the time step it considers is coarse (i.e., 5 or 10 years). Unfortunately, analysis at a finer timestep, while maintaining the same spatial resolution, is computationally challenging for large regions due to the massive quantity of satellite imagery required. One tool that has recently become available to researchers is the graphical user interface of Google Earth Engine. This tool allows researchers to run algorithms on Google's satellite image archive, which includes almost all Landsat scenes of the United States. The tool uses Google's parallel processing platform, which puts the capabilities of high performance computing into the hands of researchers. We used this tool to classify the landscape into four land cover categories: forests, developed areas, water, and non-tree vegetation/ agriculture for 2001, 2003, and 2006 to enable an examination of forest change over a finer timestep. An accuracy assessment (see Appendix A) was conducted using high-resolution aerial photos (i.e., National Agriculture Imagery Program) to ensure that the result was comparable to the NLCD change product. The results from the Google Earth analysis were used to gain a deeper understanding of spatio-temporal dynamics of forest cover over the time period, particularly for areas the showed unanticipated patterns of forest change.

4. Results and Discussions

4.1 Forest Change, 2001 - 2006

Forest change, expressed as the percentage of the statistical area affected, is shown in Figure 3. As expected, much of the study region shows little change relative to the statistical area. Most of the statistical areas in the Great Lakes and North East megaregions experienced a net forest loss (Figure 3).   The Piedmont Atlantic Region is the most dynamic, with areas of significant loss and gain. This is consistent with the findings of Hansen et al. (2013), which noted that the US Southeast has some of the most dynamic forest change patterns in the world – largely attributed to forest harvesting activities. Of particular interest is the clear separation of forest dynamics along the state border of North Carolina and South Carolina. Areas near Charlotte, North Carolina have net forest gain. Several micropolitan areas in South Carolina, near the Charlotte metropolitan region experienced the highest net gains in the entire study area (up to 3.7% of the micropolitan statistical area). Significant forest lost can also be seen for the Atlanta metropolitan region, a finding that is commensurate with the pattern of low-density urban sprawl that has characterized Atlanta's urban expansion.

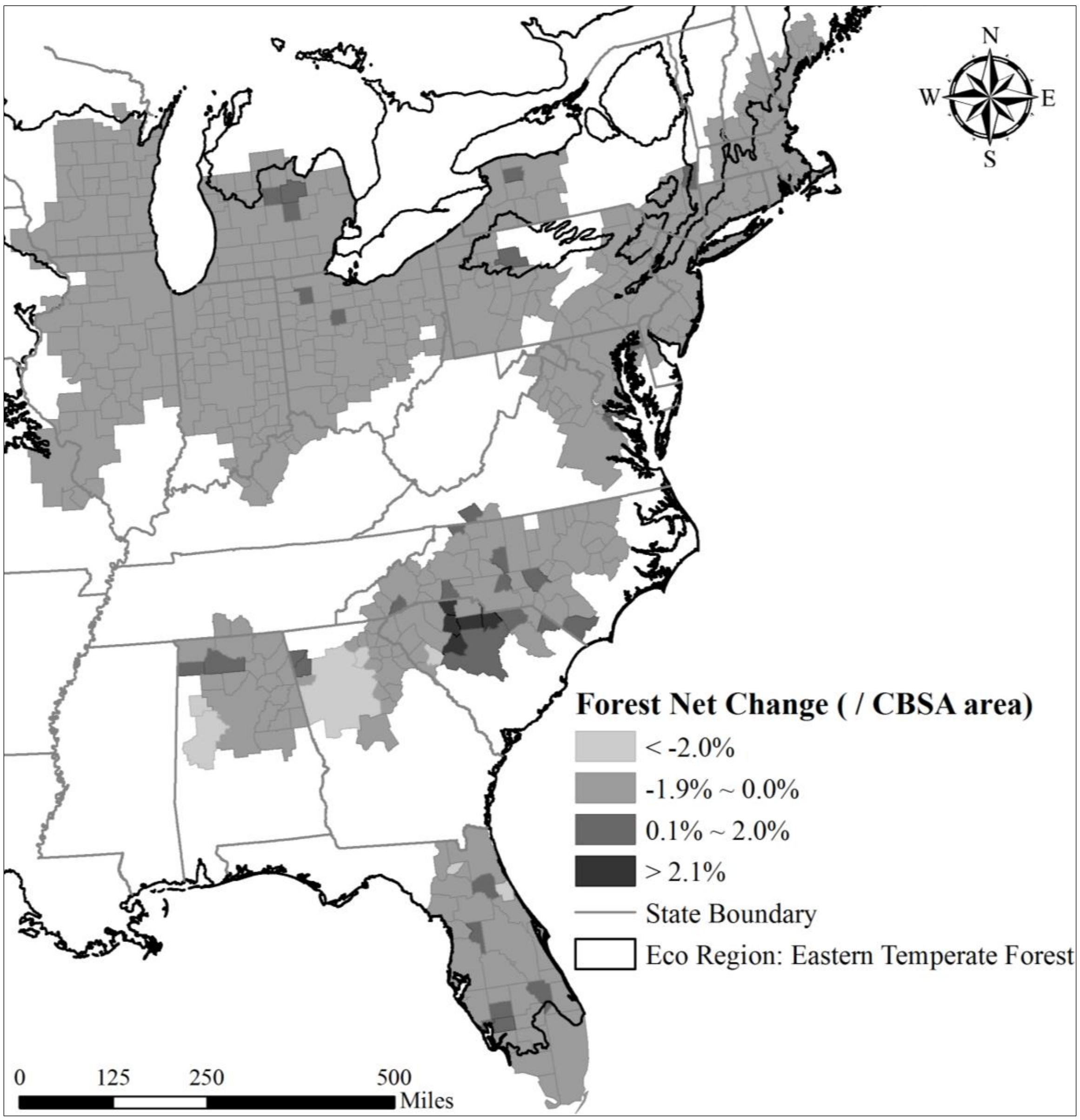

Figure 3. Forest net change between 2001 and 2006 at the scale of designation of the four East-Coast megaregions, calculated from NLCD.

Statistical analysis highlights that not only is the Piedmont Atlantic megaregion the most dynamic in terms of forest change patterns, it is also the only megaregion where there is no significant difference in forest change across the 3 levels of urbanization (Table 2). In the Great Lakes and North East megaregions, forest loss follows a predictable pattern across the three levels of urbanization, where rural areas experience the least amount of loss, and metropolitan areas have significantly greater loss (Table 2, Figure 4). This is consistent with what others have shown for the national patterns of development across the levels of urbanization (Oliver and Thomas 2014), supporting the common assumption of a link between development and forest loss. Florida shows an inverse trend, which can also be explained by development. There are only four rural counties in the Florida megaregion (Table 1) and those counties lie adjacent to metropolitan areas that are witnessing significant urban expansion and development pressure.

Table 2. Descriptive statistics of forest net change between 2001 and 2006 for each statistical area. The p value indicates the significance of the Krusall-Wallis test, showing whether or not differences exist across the levels of urbanization are significant for each megaregion.

Megaregion	CBSA category	Forest net change (%)		K-W, H value	p value (K-W test)
		Mean	SD		
North East	Metropolitan	-0.37	0.29	12.5	0.0019*
	Micropolitan	-0.25	0.26		
	Non-designated	-0.02	0.43		
Great Lakes	Metropolitan	-0.15	0.16	29.1	<0.0001*
	Micropolitan	-0.08	0.13		
	Non-designated	-0.08	0.14		
Piedmont Atlantic	Metropolitan	-0.71	0.76	4.0	0.1349
	Micropolitan	-0.08	1.18		
	Non-designated	-0.53	0.86		
Florida	Metropolitan	-0.11	0.18	6.8	0.0336*
	Micropolitan	-0.23	0.67		
	Non-designated	-0.80	1.02		

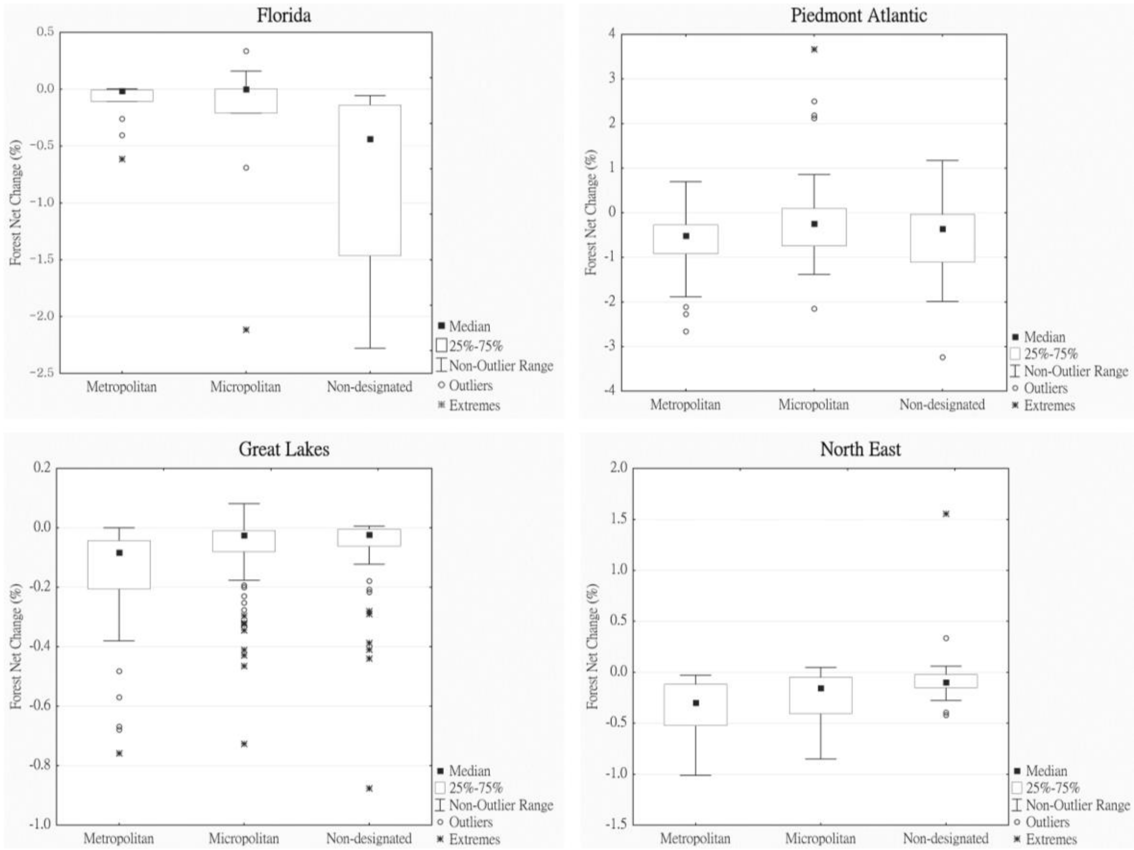

Figure 4. Boxplots of forest net change (%) for the three CBSA categories in the four East-Coast megaregions.

4.2 Anomalous Land Cover Change Pattern Implications

A closer look at the forest change within micropolitan areas showed significant differences across the 4 megaregions (Kruskal-Wallis by Ranks H=15.2, p=0.0017). The box plots (Figure 4) illustrate a number of outliers. Of particular interest are those micropolitan areas that underwent significant forest increase in the 5-year study period. As mentioned, a distinct spatial trend is evident along the border of North and South Carolina near the city of Charlotte, NC. Although state-level historic differences in settlement patterns and policies could explain a difference in total forest cover along the border, that is not likely to account for the difference in forest change over a five year period, particularly because the increase in forest area is not explained by a lack of development in these area. For example, Lancaster County in South Carolina (Figure 5) has been shown in other work (Oliver and Thomas 2014, this issue) with a significant development rate (7%) in the 5 year period – one of the highest rates of micropolitan development in the region. Forested lands represented the biggest contribution to that permanent conversion to development. Census records also show that the county had a 25% 10-year population growth rate. It has since been reclassified and become part of the Charlotte-Concord-Gastonia NC-SC metropolitan statistical area in 2013. Given these drivers of forest loss, what explains the significant gains of forest cover seen in the satellite imagery?

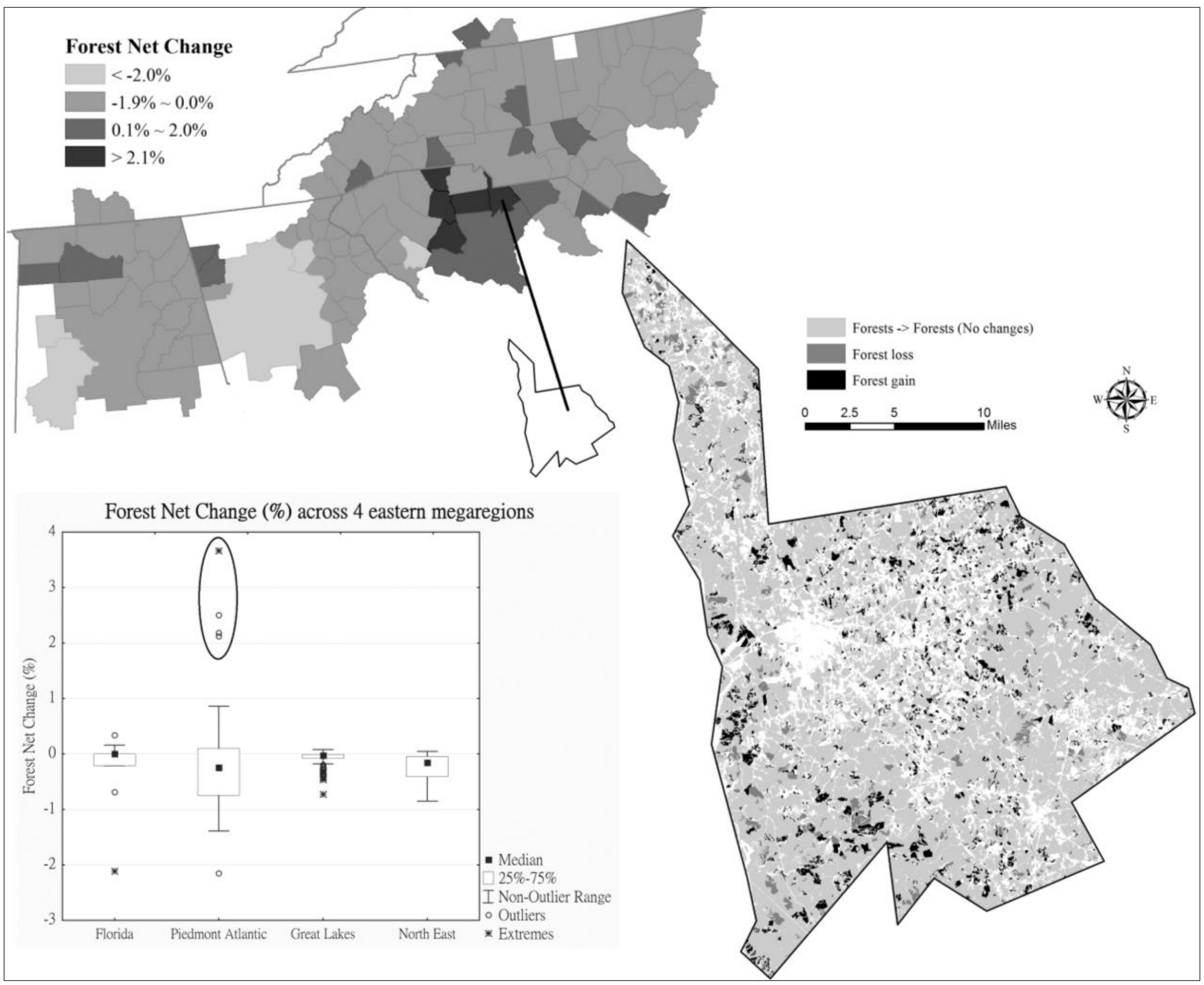

Figure 5. The outliers of the Piedmont Atlantic megaregion's micropolitans shown in the boxplot are the darkest polygons in the map.

A closer look at Lancaster county using the NLCD change product at the pixel level (Figure 5) as well as the Google Earth Engine analysis at the finer timestamp (Figure 6) reveals a few interesting findings. Most of the permanent conversion to development is in the northern tip of the county (i.e., influenced by Charlotte and surrounding areas). 60% of the areas that lost forest are converting to grassland/ herbaceous and 22% to the developed areas, and these areas are possibly clearcutting for further construction developments. Forest gains, evident in the middle and south of the county, were more prominent in 2001-2003. There appears to be two big drivers of forest growth in the time period. The first is regrowth after logging activities, as can be seen throughout the south and east of the county. The second is regrowth after clearing for some 'development' activity. An example of this can be seen at the eastern end of the airport runway at the Lancaster County Airport (west of Lancaster on the western border of the county). These finding suggest that the analysis of the shorter time period, made feasible by Google Earth Engine, provides insight into localized drivers and timing of forest dynamics, but that longer time scales (10-20 years) and consideration of the prior classes may be necessary to truly understand urban growth effects. In other words, the apparent gains in forest cover during the study period are likely part of a long-term cycle of forest loss and gain due to harvesting, with the 2001-2006 time period showing more regrowth. This is supported by the 'prior land cover class' information, which shows that 57% of the forest gain is occurring in areas that were grassland/herbaceous and 36% were pasture/hay (i.e., what a clearcut would look like on a satellite image). This is coupled with a more subtle signal of permanent forest loss due to development that is somewhat overshadowed in this 5 year period.

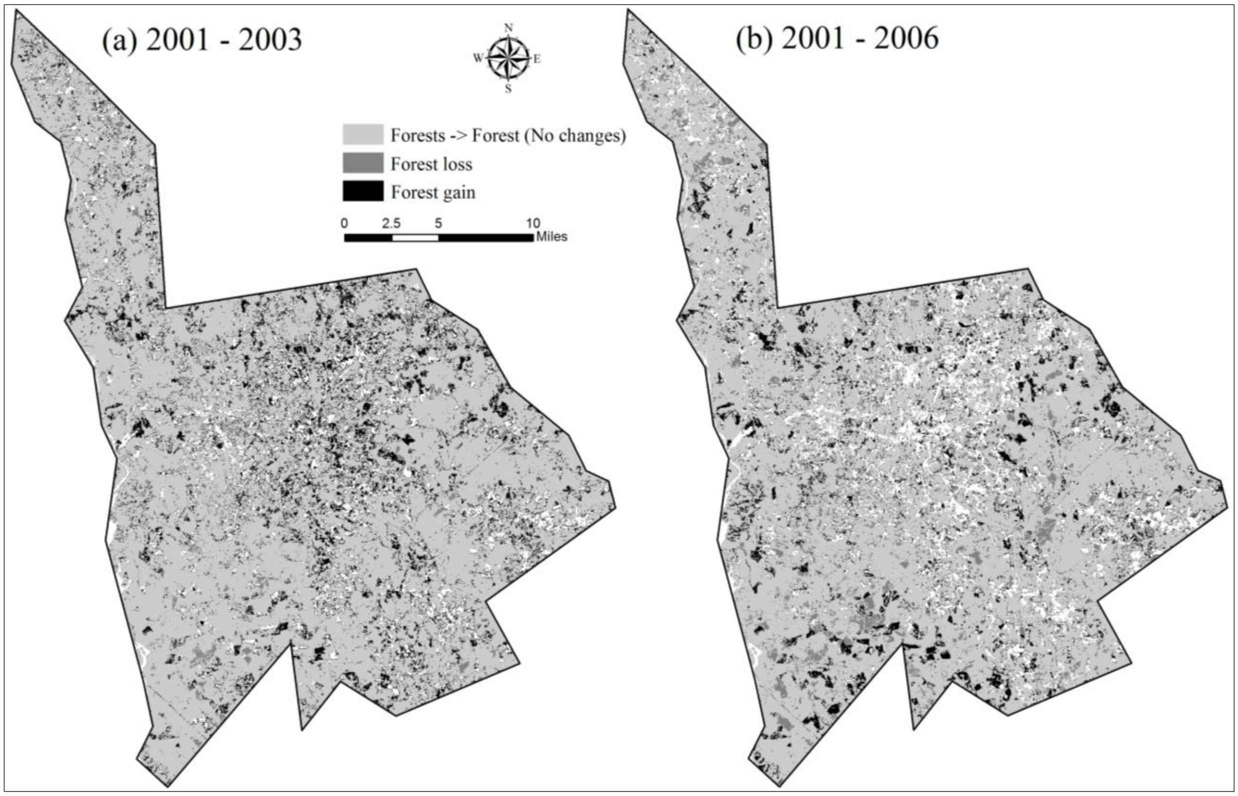

Figure 6. Forest change in Lancaster County, SC for (a) 2001-2003 and (b) 2001-2006 calculated from yearly Landsat image analysis using Google Earth Engine.

Combining the analysis of the high resolution NLCD product with supercomputing power of Google Earth Engine enabled a broad regional comparison and a detailed investigation of local drivers of land cover change. Our results, and the recent results of Hansen et al. (2013), suggest that some additional research is needed to fully uncover the temporal signal to land cover change patterns in the southeast. Short-term land cover dynamics (i.e., 5-10 years) may obscure the underlying signal of development that is impacting the landscape. Given that the rotation cycle for forest harvesting is generally longer than a decade (i.e., often around 20-25 years for loblolly pine in the southeast), it should now be possible to separate the cyclical dynamics of forestry in the region from the more permanent conversions to development in the satellite signal. It is likely that urbanization and development are impacting micropolitan regions more heavily than what is initially apparent from broad regional interpretations of satellite change detection analysis.

5. Conclusions

We have shown that forest land change in micropolitan areas differs from metropolitan and rural forest dynamics. The Piedmont Atlantic and Florida megaregions are the most variable, the Greater Atlanta area and state of Alabama have the greatest forest loss across the east coast. Other areas in the southeast are affected by logging activities and show a pattern of loss and gain that likely repeat over longer time periods. Our findings indicate that a long temporal analysis is required to account for this affect to achieve a better understanding of permanent forest conversion in the region and illustrate the highly variable nature of forest change within micropolitan regions. Techniques utilized in this paper suggest that emerging tools that provide supercomputing/parallel processing capabilities for the analysis of 'big' satellite data open the door for not only researchers to better address different landscape 'signals', investigate large regions at a high temporal and spatial resolution but also less-technical users to deal with change detection analysis. The occasion now exists to ask questions regarding spatio-temporal land cover trends in the southeast in a manner previously not possible.

Literature Cited

Berube, A., Singer, A., Wilson, J. and Frey, W. 2006. *Finding Exurbia: America's Fast-Growing Communities at the Metropolitan Fringe.* The Brookings Institution, Washington, DC.

Daniels, T.L. 1999. *When City and Country Collide: Managing Growth in the Metropolitan Fringe.* Washington, D.C.: Island Press.

Drummond, M.A. and Loveland, T.R. 2010. Land-Use pressure and a transition to forest-cover loss in the eastern United States. *Bioscience* 60(4):286 – 298.

Fry, J.A., Xian, G., Jin, S., Dewitz, J.A., Homer, C.G. and Yang, L. 2011. Completion of the 2006 national land cover database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing* 77(9): 858 – 864.

Griffith, J.A., Stehman, S.V. and Loveland, T.R. 2003. Landscape Trends in Mid-Atlantic and Southeastern United States Ecoregions. *Environmental Management*. 32(5): 572-588.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O. and Townshend, J.R.G. 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342: 850 – 853.
http://earthenginepartners.appspot.com/science-2013-global-forest.

Loveland, T.R. and Acevedo W. 2006. Land-cover change in the eastern United States. United States Geological Survey. *http://landcovertrends.usgs.gov/eastResults.html*

McDonald, J.H. 2009. Handbook of biological statistics (2nd ed.) Baltimore, Maryland: Sparky House Publishing.

Myer, R.R. and Beegle, J.A. 1947. Delineation and Analysis of the Rural-Urban Fringe. *Applied Anthropology* 6: 14 – 22.

Napton, D.E., Auch, R.F., Headley, R. and Taylor, J.L. 2010. Land changes and their driving forces in the Southeastern United States. *Regional Environmental Change* 10(1): 37 – 53.

Oliver, R.D. and Thomas, V.A. 2014. Micropolitan areas: Exploring the linkages between demography and land-cover change in the United States cities. *Cities* 38: 84 – 94.

Omernik, J.M. 1987. Ecoregions of the conterminous United States. Map (scale 1:7,500,500). *Annals of the Association of American Geographers* 77(1): 118 – 125.

Omernik, J.M. 1995. Ecoregions: A spatial framework for environmental management. Tools for Water Resource Planning and Decision Making. In *Biological Assessment and Criteria*, eds. W.S. Davis and T.P. Simon, 49 – 62. Lewis Publishers, Boca Raton, FL.

Reich, P.B. and Frelich, L. 2002. Temperate Deciduous Forests. Volume 2, The Earth system: biological and ecological dimensions of global environmental change. In *Encyclopedia of Global Environmental Change*, eds. H.A. Mooney and J.G. Canadell, 565 – 569. John Wiley & Sons, Ltd, Chichester.

Steyaert, L.T. and Knox, R.G. 2008. Reconstructed historical land cover and biophysical parameters for studies of land-atmosphere interactions within the eastern United States. *Journal of Geophysical Research* 113: 1984 – 2012.

Vogelmann, J.E., Howard, S.M., Yang, L., Larson C.R., Wylie, B.K. and van Driel, N. 2001. Completion of the 1990s National Land Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing* 67: 650 – 662.

Waisanen, P.J. and Bliss, N.B. 2002. Changes in population and agricultural land in conterminous United State counties, 1790 – 1997. *Global Biogeochecmical Cycles* 16(4): 84-1 – 84-19.

Xian, G., Homer, C.G. and Fry, J.A. 2009. Updating the 2001 national land cover database land cover classification to 2006 using Landsat imagery change detection methods. *Remote Sensing of Environment* 113(6): 1113 – 1147.

Chapter 4 – Big Data Analysis Using Google Earth Engine

As mentioned in Chapter One, it takes 16 Landsat scenes to cover the whole area of the Piedmont Atlantic Megaregion (and about 4 times this amount for all 4 megaregions examined in this thesis). If a researcher wants to examine the change at higher temporal scale than what is possible through the NLCD change products, alternative computing platforms are necessary. There are a number of ‘pre-processing’ steps required before a change detection output can be generated.

Google Earth Engine stores a number of Landsat-derived image products in its archives that are available to ‘Trusted Testers’ for analysis. To reduce pre-processing requirements, I used the ‘percentile composite data’ that was calculated from Landsat 7 reflectance as the source imagery for the land cover classification. The Landsat percentile composite data is a composite of top of atmosphere reflectance (i.e. not atmospheric corrected) as opposed to surface reflectance (i.e. atmospherically corrected data). Note that the Landsat surface reflectance datasets were still under revision by Google at the time of this research. While datasets with surface reflectance would typically be a more appropriate selection for doing change detection analysis (given the effect of atmosphere across scenes), some initial exploration of the data and results, as well as a comparison with other derived products, suggested that the percentile composite would provide acceptable and best-possible classification accuracy (Figure 10 and 11 in Appendix A). Nevertheless, future work should make use of the surface reflectance data as they become available in the archives.

Four land cover categories: forests, developed areas, water, and non-tree vegetation/ agriculture, were created for collecting spectral signatures by dropping sample points or drawing polygons (Figure 12, 13, and 14 in Appendix A). Sample polygons were created for each land cover class (Figure 15, 16, and 17 in Appendix A). After some exploration of possible classification algorithms, I adopted the Fast Naïve Bayes classifier, as it produced the highest accuracy across my classes. The Fast Naïve Bayes classifier is a probabilistic classifier based on Bayes' theorem with the assumption that there is strong and naïve independence between features. This classifier model is suitable for multiclass classification because it is fast to build, only requiring small amounts of training data to estimate the parameters, and easy to modify with new training data without having to rebuild the model. As such, I was able to use this classifier across different years. All spectral bands (except thermal) were used in the classification process.

I experimented with the number of sample polygons for each land cover class. After about 100 total samples, classification accuracy either did not increase, or actually decreased, based on image interpretation. I attribute this to misclassification and natural variance in pixel reflectance in agricultural land and young forests. Once I was satisfied with the performance of the classifier on the Google Earth Engine platform, I ran the classifier for the study region, downloaded the classified output and imported the classification result into a GIS for further analysis (i.e., accuracy assessments, change detection, and statistical tests) (Figure 18, 19, and 20 in Appendix A).

1. Accuracy Assessment and Comparison with NLCD

To have a better understanding of how the NLCD product compared to the Google Earth Engine classification, I first assessed change from 2001-2006 for both products. Obviously, the Google Earth Engine platform offers opportunities to look at the land cover change in a smaller time period than NLCD. As mentioned a smaller temporal scale was examined with Google Earth Engine, but an understanding of the error in both datasets could only be obtained by comparing the same time period against human interpretation. Both classification products will contain error.

The most appropriate method of performing an accuracy assessment is to compare the classified data layers with higher resolution aerial photos from the same time period (i.e., Virginia Base Mapping Program (VBMP) or National Agriculture Imagery Program (NAIP)). The classified output data layer can be used as a reliable land cover result if the percentage of the accuracy reaches the desired threshold (i.e., 75% for overall accuracy in my thesis).

In this thesis, Union county, South Carolina was used for the accuracy assessment (Figure 11). Not only is this county one of the micropolitan outliers in the Piedmont Atlantic megaregion, but it also has a few small ponds or lakes, several large patches of forest in the southern part, numerous agricultural lands in the north, and urban areas in the center (in other words, all classes are evident in the county). With the exception of large patches of forest, other land cover types are fragmented by roads and other human activties so that if the classified maps lack precision, these complicated land covers will not be captured correctly.

The 2006 high-resolution (i.e., 1 m) aerial photographs from the National Agriculture Imagery Program (NAIP), available free for download, were used as referenced images. I generated 100 sample points, dispresed across the county (Figure 7) and recorded the land cover class based on the NLCD, Google Earth Engine, and my interpretation of the high-resolution NAIP aerial images. Given that the main focus of the research was forest and forest change, it is essential to have a high classificaiton accuracy for forest land. My results show that the user's accuracy for forest class in both the NLCD and the classified Landsat images are over 85%; the producer's accuracy for NLCD is 80%, which is lower than that of the classified images with over 84% (Table 3, 4, and 5).

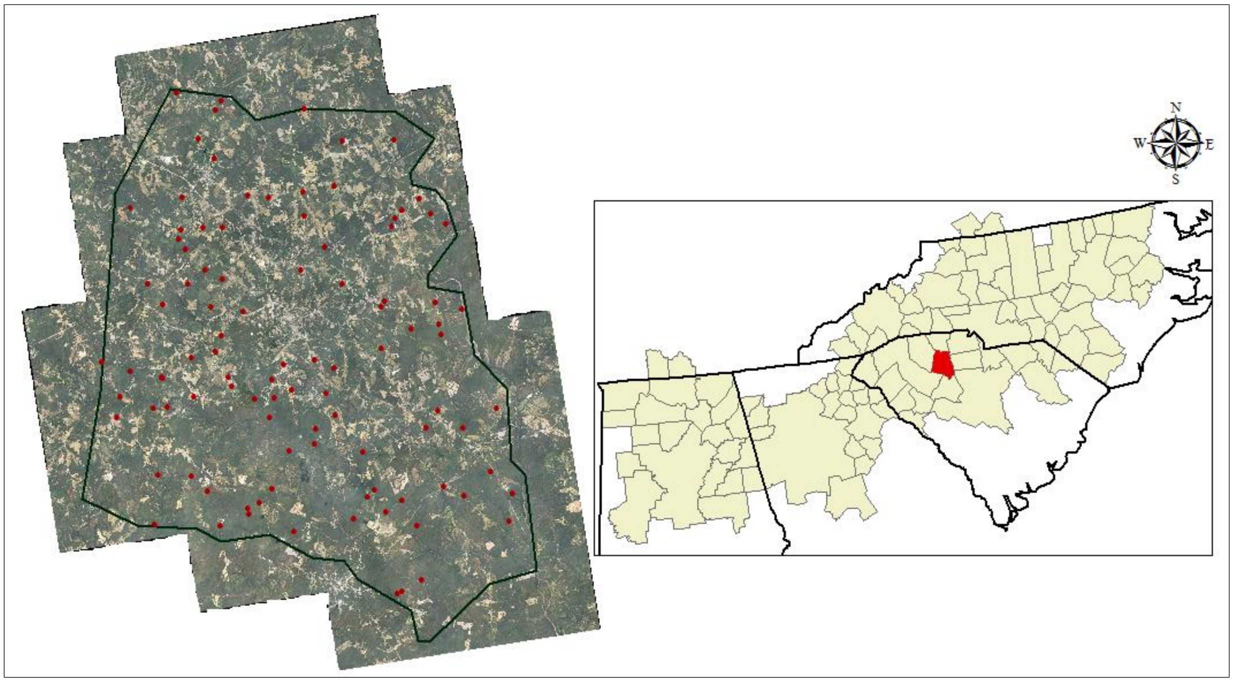

Figure 7. Right: Red county is the location of Union county, South Carolina. Left: Red dots are the 100 sample points generated in the Union county.

Table 3. Error matrix between NLCD and NAIP for year 2006.

	Referenced Data				
	Forests	Non-tree Veg/ Ag	Developed Areas	Water	Row Total
Forests	57	7	1	0	65
Non-tree Veg/ Ag	14	13	1	0	28
Developed Areas	0	2	1	0	3
Water	0	2	0	2	4
Column Total	71	24	3	2	100

Table 4. Error matrix between classified Landsat scenes and NAIP for year 2006.

	Referenced Data				
	Forests	Non-tree Veg/ Ag	Developed Areas	Water	Row Total
Forests	60	9	1	0	70
Non-tree Veg/ Ag	11	15	2	0	28
Developed Areas	0	0	0	0	0
Water	0	0	0	2	2
Column Total	71	24	3	2	100

Table 5. Error matrix between classified Landsat scenes and NAIP for year 2009.

	Referenced Data				
	Forests	Non-tree Veg/ Ag	Developed Areas	Water	Row Total
Forests	61	9	0	0	70
Non-tree Veg/ Ag	9	14	2	0	25
Developed Areas	0	1	1	0	2
Water	0	1	0	2	3
Column Total	70	25	3	2	100

2. Differences between Google Earth Engine Classification and NLCD

Although the overall agreement between the two classifications (NLCD and Google Earth Engine) was good, some differences were observed. Detailed investigation suggests that the definition of forests and developed land has a critical impact on our understanding of forest dynamics at the census designation scale. When comparing NLCD data closely with Landsat images at the pixel level (Figure 8), there are many locations that have significant canopy cover but remain excluded from being classified as forest in the NLCD. These areas include parks, golf courses, and urban street trees. When using annual Landsat imagery to assess forest land in developed regions, the NLCD change product is not directly comparable to a canopy cover change product (i.e., such as the continuous Global Forest Change map by Hansen et al 2013 or my classification of 'forest' which could also be described as 'trees'). In this case, the definition of the NLCD 'developed area' classes include areas with a mixture of constructed material and vegetation. This means that trees that are incorporated into a 'developed' class are excluded from the NLCD definition of forests. NLCD uses both land cover and land use in its description of classes. Forest is a land cover class in NLCD. However, the 'Developed' classes could be considered to be more closely aligned with a 'land use' definition. Recall from Chapter One that land cover and land use are fundamentally different from each other. To reiterate, while land cover denotes the physical state of the land and is determined by the direct observation of the earth surface, land use denotes the human use of the land and can be defined as socio-economic references for human activities on the land surface. While land cover class is usually defined by straight interpretations of remotely sensed imagery, land use type is based on the surveys from planning authorities (Fisher and Unwin 2005).

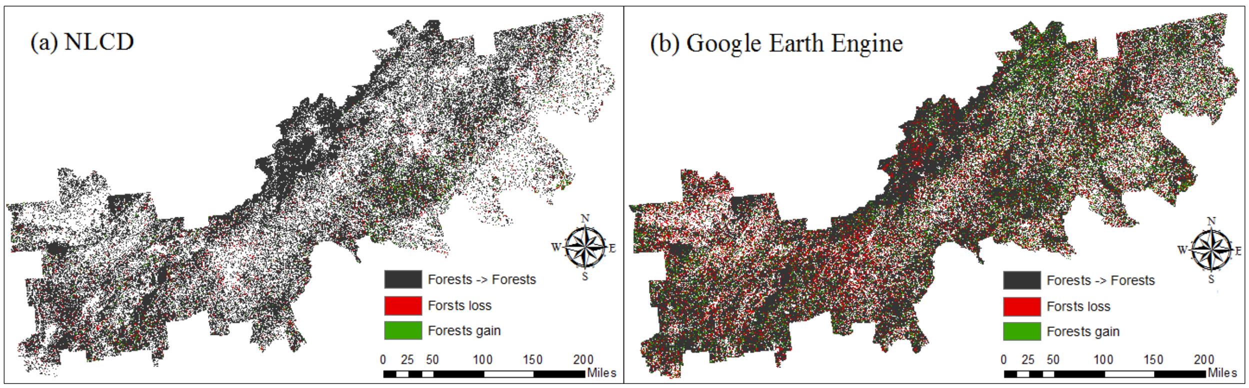

Figure 8. Comparison between NLCD classification and Google Earth Engine classification (in pixel level).

Results in pixel level analysis (Figure 8) reveals that there are both forest gains and forest losses in every statistical area; but when aggregated into census designations, each statistical area is limited to a single percentage of forest cover change (Figure 9). While forest loss and gain generally balance out, a few notable exceptions are evident. For example, the common forest conversion pattern for both NLCD (Figure 3) and Landsat images (Figure 9) indicates that the greater Atlanta metropolitan has experienced considerable forest net loss during 2001 to 2006.

In the Google Earth Engine analysis, I split the time period: (i) from 2001 to 2003 and (ii) from 2001 to 2006 to determine if the pattern of change was consistent over the entire time period (Figure 9 a, b). The results demonstrate the expected progression of forest loss in the Atlanta region. Outside Atlanta, some differences in the timing of forest loss are evident. Areas in the northern part of Figure 9 show a significant loss in the 2001-2003 period. Southern areas in Figure 9 show initial forest gain (2001-2003) followed by a period of forest loss from 2004-2006. This finding serves to illustrate that the Google Earth Engine analysis is performing effectively at both time steps and is capable of providing real change results at the finer time step. Put bluntly, this is not just noise or random error being reflected for the greater Atlanta area.

Confirmation of the expected results for the Atlanta region provides more confidence when examining forest change patterns at finer time steps in other regions. There are numerous examples where the overall change pattern from 2001-2006 is consistent for both the NLCD and Google Earth analysis; however, an investigation of a finer timestep provides significant insight into the timing of this change. For example, in Alleghany and Carroll Counties, both the NLCD and the Google Earth Engine analyses show an increase in forest cover after a 5-year period. Shifting to a finer temporal step made possible with the Google Earth Engine analysis showed an initial forest loss from 2001-2003, but a significant forest gain afterwards (mainly driven by agriculture conversion). I confirmed this decrease in agricultural land with the census data from the National Agriculture Statistics Service (NASS). NASS shows that the total area of all agricultural, crop, and pasture land have decreased from 2002 to 2007 in this area. As with the Atlanta example above, the utility of Google Earth Engine is that it can help elucidate 'moments' of change.

Unfortunately, my results also highlight some areas where the NLCD and Google Earth Engine analysis depicts different pattern of forest change over the 2001-2006 time period. For example, the NLCD map shows a stark pattern of forest conversion along the state border separating North and South Carolina, with significant gains on the South Carolina side. This could potentially suggest a difference in deforestation practices (or development activities) in the two states. This pattern is not as obvious with the Google Earth Engine Analysis (Figure 9). Note that border region is heavily impacted by the Charlotte metropolitan area. A close examination of the developed areas suggests that the differences in the two results is, again, partly due to the differences in how NLCD classifies development and forest, excluding many treed areas from the forest class.

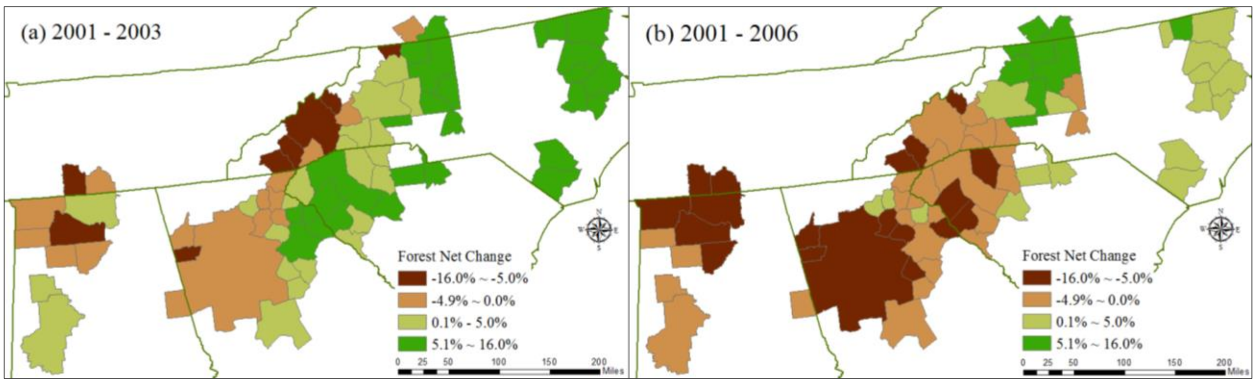

Figure 9. Forest net change between (a) 2001 – 2003 (b) 2001 – 2006 at the scale of the census designation, calculated from yearly Landsat images in Google Earth Engine. There are some statistical areas excluded due to cloud-cover in 2003 summer imagery.

The three examples discussed above help to legitimate the utility of Google Earth Engine's platform in the construction of accurate forest change detection maps. By accommodating finer time step analyses, there is also the potential to more meaningfully isolate the timing of the land cover conversion (forest cover or otherwise). These findings, along with those discussed in Chapter Two, suggest that future work with a much longer time series could help to separate the signal of forest harvesting practice from urban growth (or forest loss due to development).

Chapter 5 – Summary and Conclusions

This thesis has two primary contributions. First, forest change across a gradient of urbanization was investigated. Micropolitan areas were shown to have a different pattern of forest change than metropolitan and rural areas. In the east coast of the US, the Piedmont Atlantic and Florida megaregions are the most variable in terms of the pattern of loss and gain. The Greater Atlanta area and state of Alabama have experienced the greatest forest loss in the Eastern US for the 2001-2006 period. Other areas in the southeast are affected by logging activities and show a pattern of loss and gain that likely repeat over longer time periods. The cyclical nature of logging makes the production of land cover change maps cumbersome and consequently requires a careful consideration of timing and class definition. The emergence of Google Earth Engine provides researchers with an additional means to address the complexity of forest dynamics in the context of urban change. Techniques utilized in this thesis illustrate that the Google Earth Engine platform and its GUI tools—which provide supercomputing/parallel processing capabilities for the analysis of 'big' satellite data—open the door for researchers to better address different landscape spectral signatures and to increase the scale of their investigations. The occasion now exists to ask questions regarding spatio-temporal land cover trends in the southeast in a manner previously not possible.

Appendix A – Google Earth Engine Analysis

The addendum provides the step-by-step procedure of Landsat Images Analysis on the Google Earth Engine and the accuracy assessment report of the classified images.

Step 1: <u>Insert and customize the image layers</u>

In workspace, insert 3 Landsat 7 imagery layers – Percentile Composite. While this dataset is not a surface reflectance images which is not applicable so far in the Google Earth Engine, it is a composite of Top Of Atmosphere reflectance (i.e. not atmospheric corrected) as opposed to surface reflectance (i.e. atmospherically corrected data). For each layers, change the date from April to Auguest for three different years: 2001, 2003, and 2006, and lower the percentile to 45 in order to have greener vegetation and aviod cloud cover.

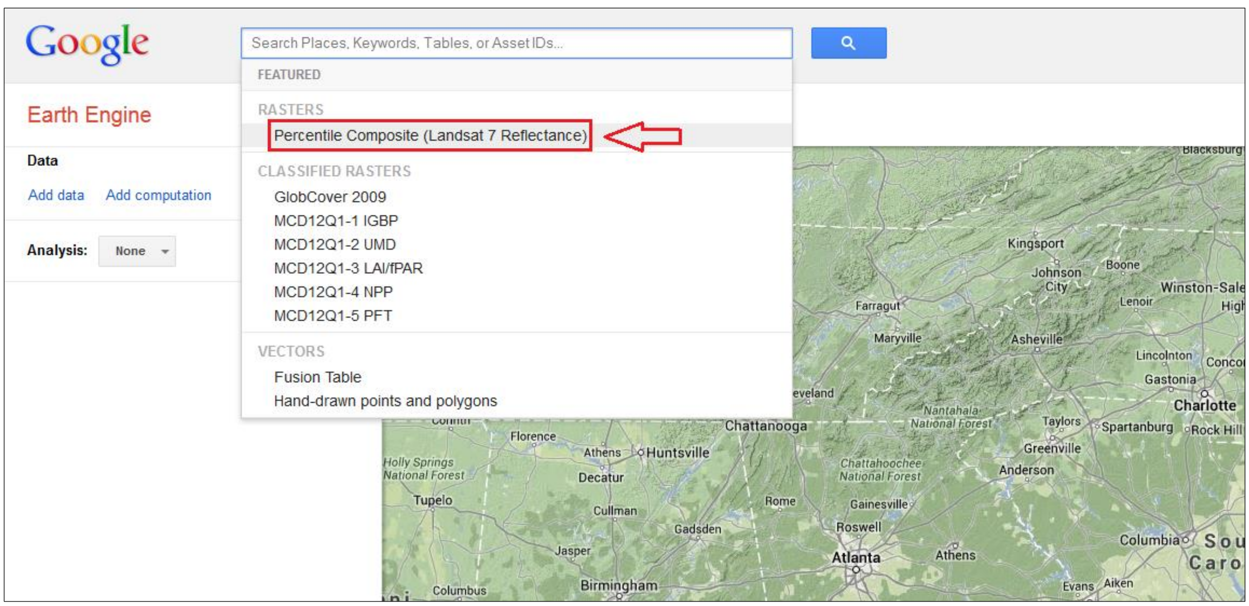

Figure 10. Insert Lansat scenes from dropdown menu.

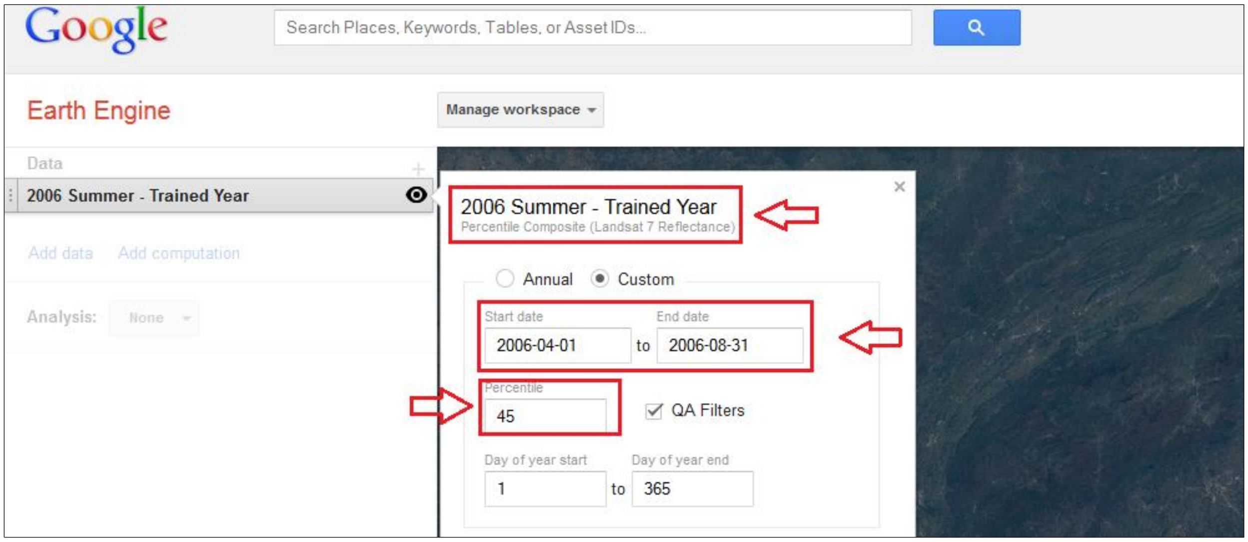

Figure 11. Rename the layer and select appropriate date and percentile.

Step 2: <u>Draw Area of Interest (AOI) and set classes</u>

Insert a blank layer by choosing Hand-drawn points and polygons in the drop-down menu for drawing AOI either in points or in polygons. Then, add classes (i.e., in this research analysis, I added four classes: forests, water, developed areas, and non-tree vegetation/ agriculture).

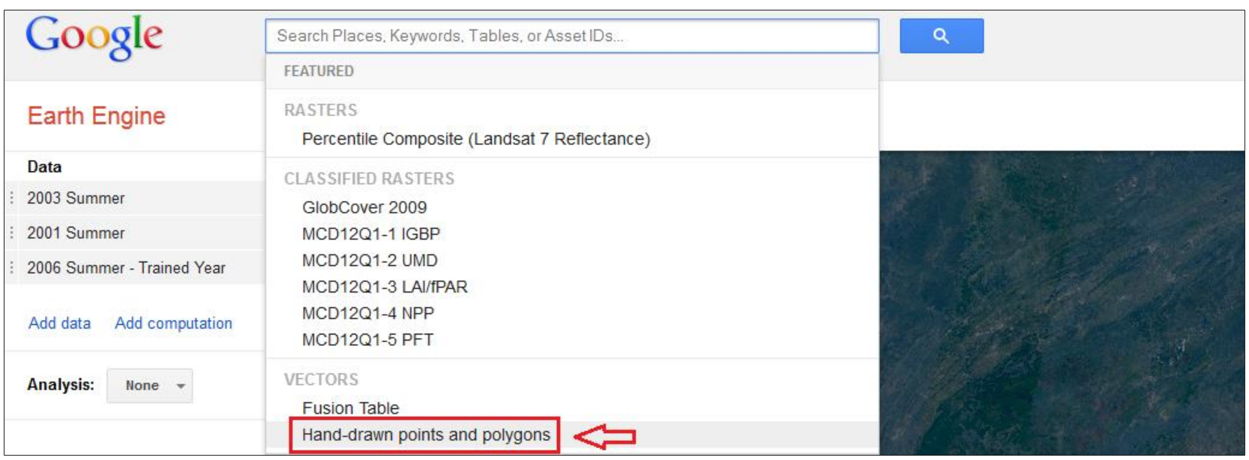

Figure 12. Insert hand-drawn points or polygons for storing the captured spectral signatures for each land cover class.

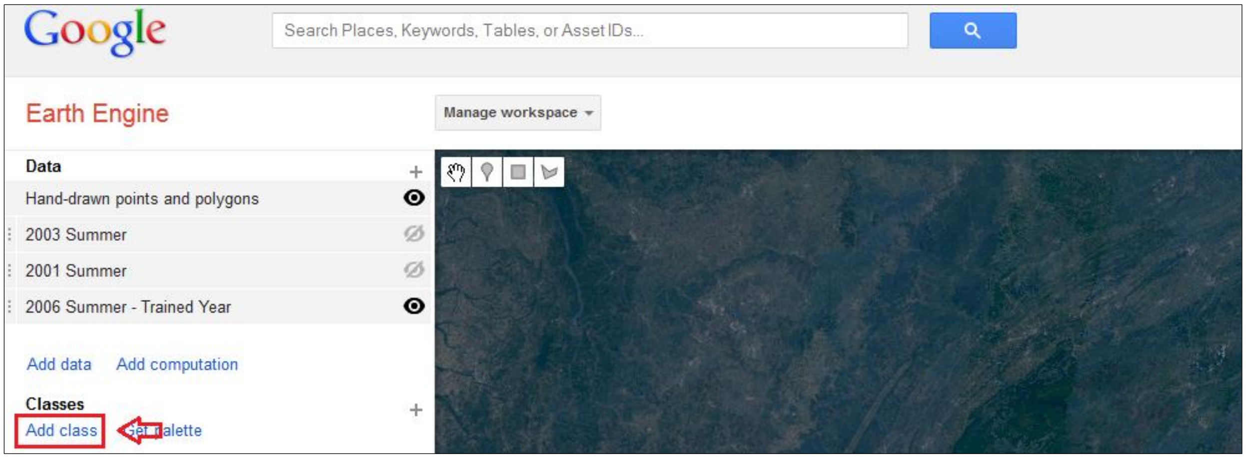

Figure 13. Add classes for different land cover types.

Step 3: <u>Capture spectral information for each class</u>

Drawing AOIs for each class and save those AOIs in the inserted blank layer. I applied polygons instead of points when drawing AOIs for the sake of capturing enough diverse spectral information for every class.

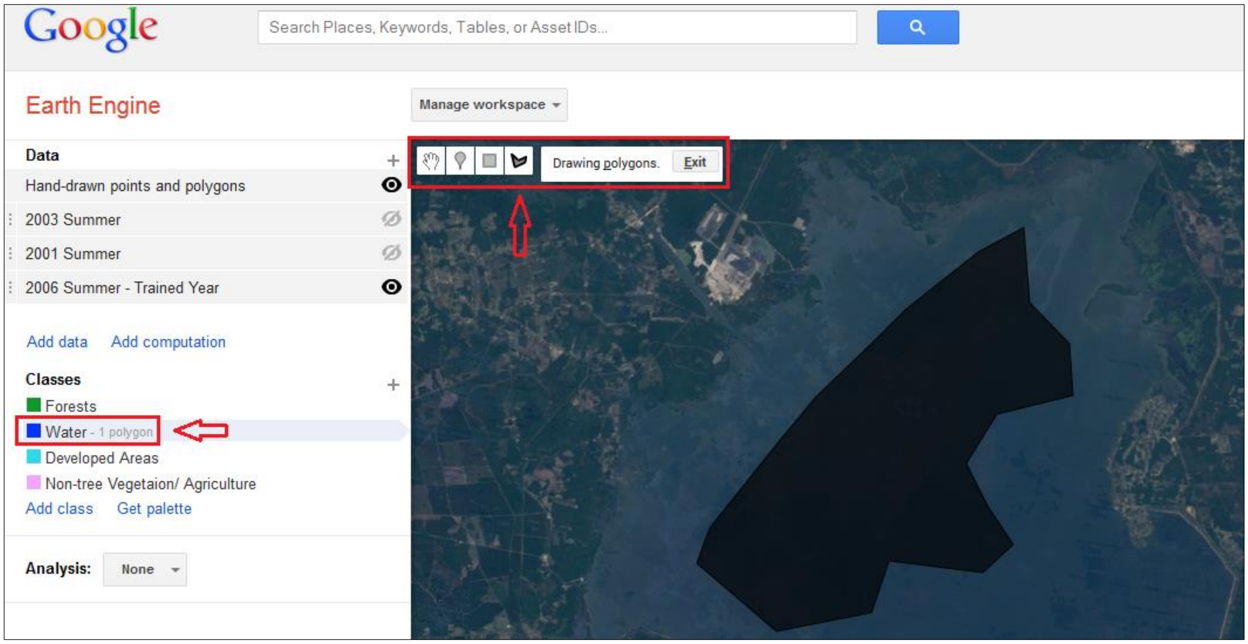

Figure 14. Select a land cover class and drop points or draw polygons for collecting enough spectral signatures.

Step 4: Apply classification method

Choose the appropriate classifier and set the resolution to be 30m same as NLCD in order to have further comparisons. Make sure to select the correct layer as input for classification training data, and exclude other layers. Repeat same steps for 3 different yearly Landsat images to get 3 classified images for 2001, 2003, and 2006. After the classifier is applied, another classified image based on the chosen for each class is added on the top of the layer list.

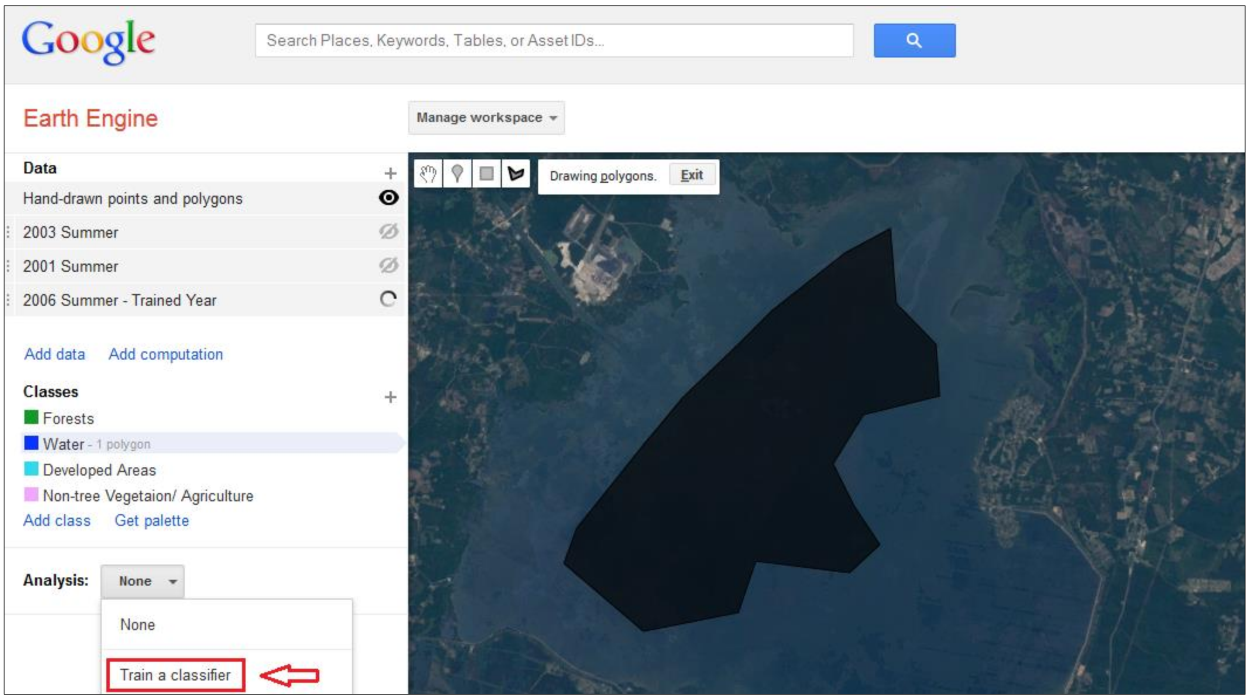

Figure 15. Apply classification based on collected spectral signatures for every land cover type.

Figure 16. Select appropriate classifier and resolution (in meter).

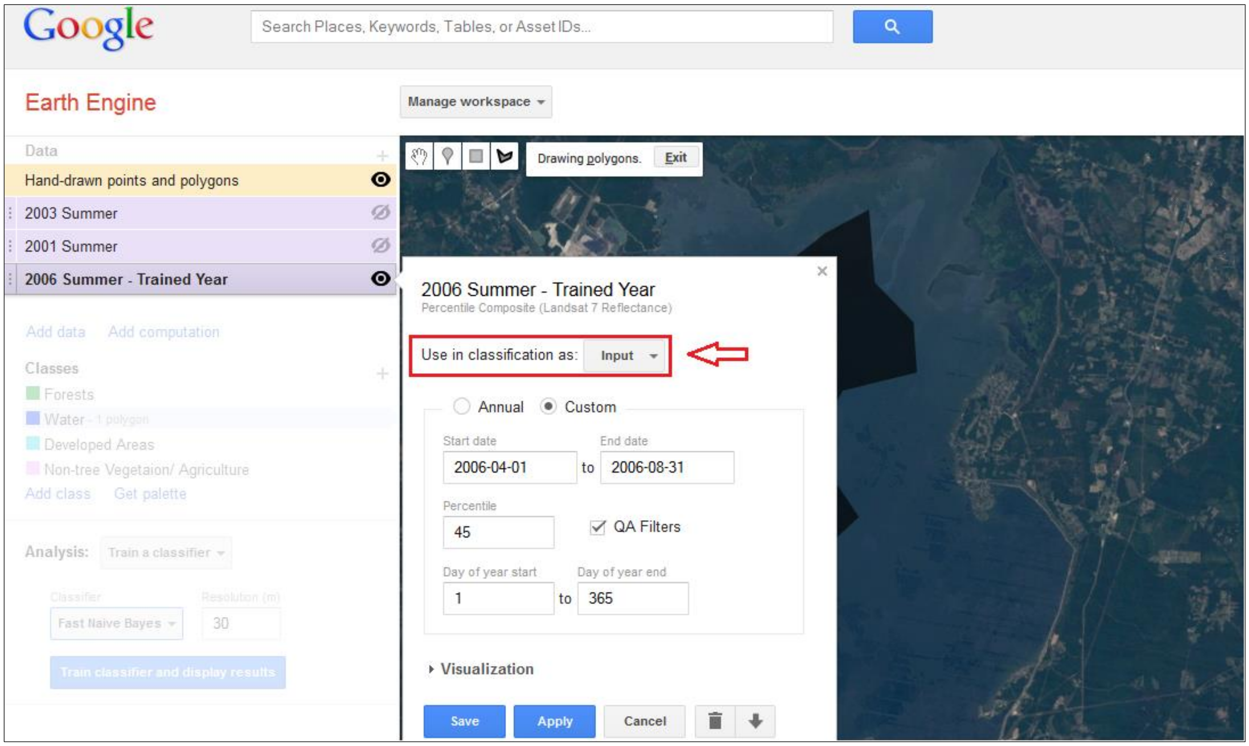

Figure 17. Decide which Landsat scenes to be the input layer for classification analysis.

Step 5: Download the output classified images

For each classified image, at the bottom of the error matrix, click the download button and set the proper download parameters. In this analysis, I set the resolution as 30m as well to match with the NLCD resolution and harmonize the spatial reference for all files. Repeat the same moves for 3 result maps.

The error matrix is obtained by comparing the classified image based on the AOIs drawn and the classifier chosen with the original satellite image.

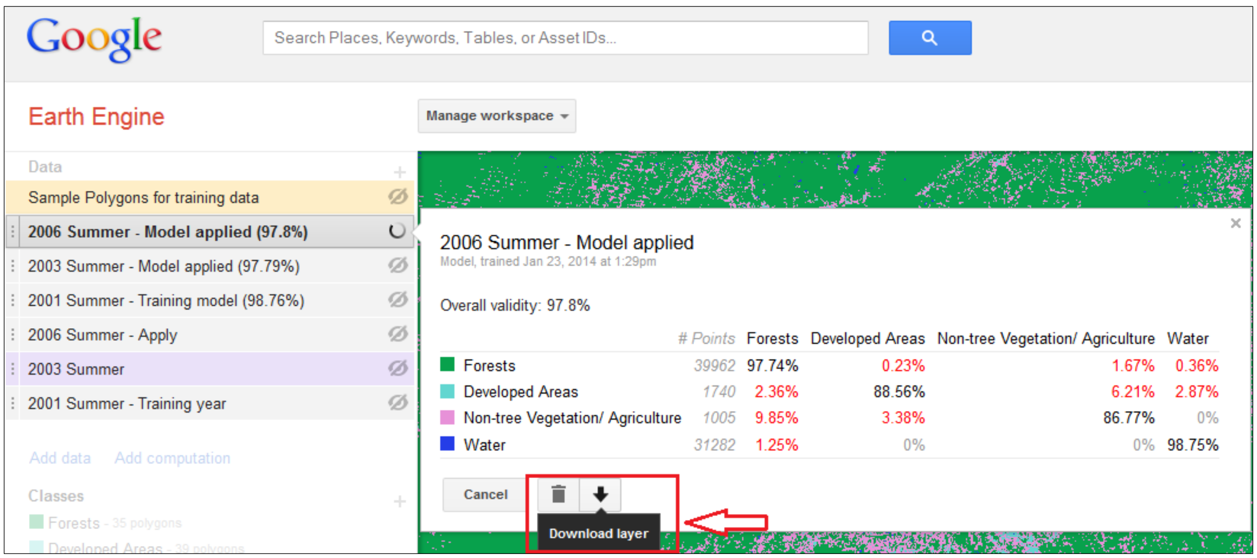

Figure 18. Download the output classified layer.

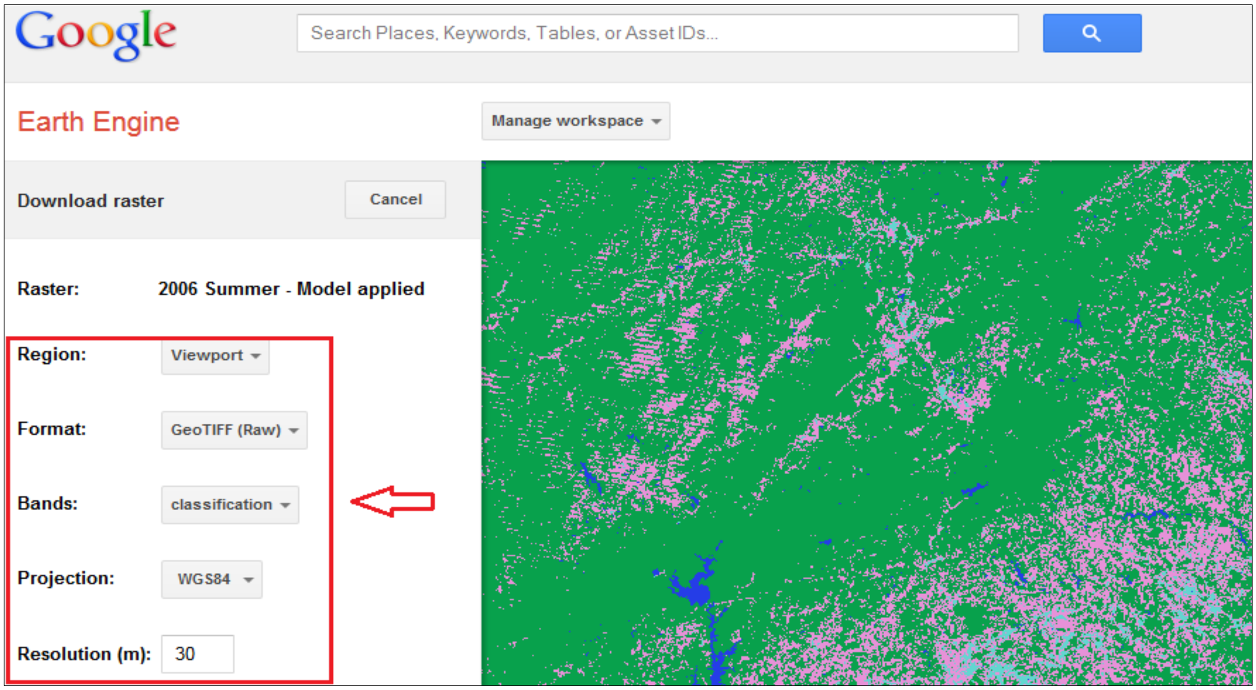

Figure 19. Select conditions for downloading the classified layer.

Step 6: <u>Import into ArcGIS for further analysis</u>

Add the output files into ArcMap for accuracy assessments. If the accuracy meets the threshold, the classified layer can be applied on further analysis.

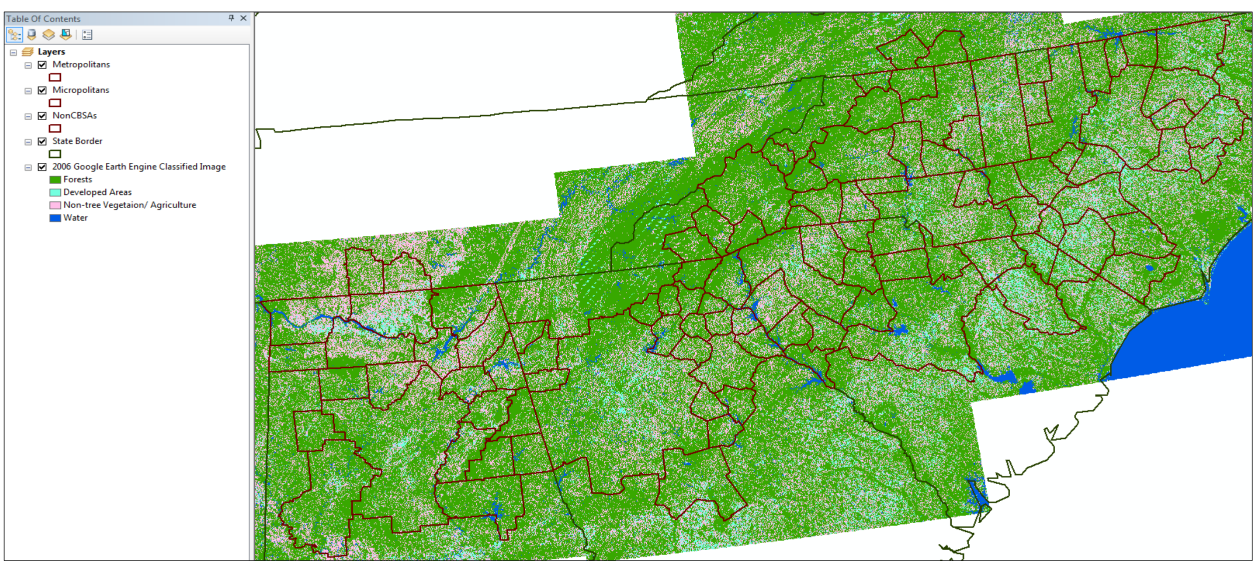

Figure 20. Further analysis on the classified layer through ArcMap and other software.