ABSTRACT

SINGH, KUNWAR KRISHNA VEER. Frontiers in using LiDAR to Analyze Urban Landscape Heterogeneity. (Under the direction of committee chair Ross K. Meentemeyer.)

Light Detection and Ranging (LiDAR) technology has facilitated extraordinary advances in our ability to remotely sense precise details of both built and natural environments. The inherent complexity of urban landscapes and the massive data volumes produced by LiDAR require unique methodological considerations for big data remote sensing over large metropolitan regions. The heterogeneous landscapes of the rapidly urbanizing Charlotte Metropolitan Region of North Carolina provided an ideal testing ground for developing methods of analysis for urban ecosystems over large regional extents, including: (1) fusion of LiDAR digital surface models (DSMs) with Landsat TM imagery to balance spatial resolution, data volume, and mapping accuracy of urban land covers, (2) comparison of LiDAR-derived metrics to fine grain optical imagery – and their integration – for detecting forest understory plant invaders, and (3) data reduction techniques for computationally efficient estimation of aboveground woody biomass in urban forests.

In Chapter 1, I examined tradeoffs between potential gains in mapping accuracy and computational costs by integrating DSMs (structural and intensity) extracted from LiDAR with TM imagery and evaluating the degree to which TM, LiDAR, and LiDAR-TM fusion data discriminated land covers. I used Maximum Likelihood and Classification Tree algorithms to classify TM data, LiDAR data, and LiDAR-TM fusions. I assessed the relative contributions of LiDAR DSMs to map classification accuracy and identified an optimal spatial resolution of LiDAR DSMs for large area assessments of urban land cover. In Chapter 2, I analyzed combinations of datasets developed from categorized LiDAR-derived variables (*Overstory*, *Understory*, *Topography*, and *Overall Vegetation Characteristics*) and IKONOS imagery (*Optical*) to detect and map the understory plant invader, *Ligustrum sinense*, using Random Forest (RF) and logistic regression (LR) algorithms, and I assessed the relative contributions of sensors and forest landscape structures. I compared the top performing models developed using RF and LR and used the best overall model to map the distribution of *L. sinense* occurrence across the urbanizing forest landscapes of the region. In chapter 3, I examined the effects of LiDAR point density and landscape context on the estimation of

biomass (of general Urban Forest and of three specific Forest Types) using multiple linear regression. I compared biomass estimation accuracies of the Urban Forest and Forest Type models and between the top-performing models of these two Forest categories. For the effect of landscape context, I quantified the degree to which the presence of built development influenced biomass estimation, and I analyzed the effect of canopy stratification on the estimation of biomass.

A unifying theme of my dissertation is to advance LiDAR analytics for accurate and detailed estimation of urban landscape heterogeneity over large regional extents. The results of the three studies suggest that establishing optimal resolution and point density for LiDAR data is a highly effective method of pursuing large area studies of urban landscape heterogeneity, and the fusion of LiDAR-derived variables and multispectral data is beneficial in some applications such as improving class discrimination of spectrally similar land cover types. Finally, the direct measurement of forest understory and overstory structure through LiDAR has proven valuable for the study of complex and heterogeneous ecosystems like urban forests.

Frontiers in Using LiDAR to Analyze Urban Landscape Heterogeneity

by
Kunwar Krishna Veer Singh

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Forestry and Environmental Resources

Raleigh, North Carolina

2014

APPROVED BY:

_____	_____
Ross K. Meentemeyer Committee Chair	Helena Mitasova
_____	_____
Hugh Devine	Stacy A. C. Nelson
_____	
Marguerite Madden	

DEDICATION

My happiness is the pursuit of the *Rêves et Désirs*. I will chase them into being. This is who I am and who I will be! I dedicate this dream to my beloved father, *Udai Raj Singh*, and to my beloved mother, *Kusum Singh*, for their relentless support in making this dream a reality.

BIOGRAPHY

Kunwar '*Krishna'* Singh was born on December 19, 1978 to a Rathour family in a village, Zafarpur, about thirty miles northeast (27.128730° N and 81.321231° E) of Lucknow, Uttar Pradesh, India. Krishna spent the early years of his childhood in the village with his grandparents, and then moved to the city with his parents. The social and environmental setting in the village taught Krishna about the knowledge in art and environment. His keen interests in pottery, gardening and venturing into the wilderness shaped his personality as an independent thinker with a never-ending list of dreams and quests for an adventurous life. Krishna's father, an expert in soil and fertilizer with math background, taught him to value nature and art from a scientific perspective.

Krishna's childhood experiences living in a village instilled in him a keen awareness of all living things and of the importance of ecosystems and the services they provide. This appreciation led Krishna to the University of Lucknow where he received his Bachelor of Science in Biology in 1999, and a Master's degree in Environmental Science in 2001. Krishna continued his education by obtaining a Master of Technology degree in 2005 in Satellite Remote Sensing and Geographic Information Science with an emphasis on forest and ecological applications to improve his understanding of interconnected spatial, temporal, and human dimensions of landscape change. His passion for using geospatial technologies to study human-environment interactions brought him to St. Cloud State University, where he graduated with a Master's of Geography, and then to UNC Charlotte in 2008 to pursue a doctoral degree on the topic of applied remote sensing. Finally, Krishna transferred to NC State to accompany his academic advisor, mentor and a great friend, Dr. Ross Meentemeyer, and he completed his PhD in NC State's Department of Forestry and Environmental Resources and Center for Geospatial Analytics.

Krishna has been actively involved in extra-curricular activities throughout his time in Minnesota and North Carolina. He enjoys a wide variety of arts and cultural activities, including photography, collecting currencies, road trips, and traveling with friends. Krishna has been very fortunate to be surrounded by great friends who recognize his colorful, mysterious, charming, and infectious personality. His dedication to friends is unwavering.

Krishna is particularly proud of his strong work ethic and a cross-cultural perspective that inspire him to excel as he moves forward in his career. He views cultural, religious, and ethnic differences as springboards for collaboration rather than barriers. And it is imperative to him that he follow the path of a responsible citizen no matter where his life takes him. Krishna's next journey will involve interdisciplinary research on human and natural impacts to ecological systems. He is very excited to take that next, first step to work with Dr. Jill Johnstone in the Department of Biology at the University of Saskatchewan in Saskatoon, Canada.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1. Mapping Urban Landscape Heterogeneity

CHAPTER 2. Detecting Understory Plant Invasion

CHAPTER 3. Effects of LiDAR Point-density

LIST OF FIGURES

CHAPTER 1. Mapping Urban Landscape Heterogeneity

CHAPTER 2. Detecting Understory Plant Invasion

CHAPTER 3. Effects of LiDAR Point-density

INTRODUCTION

The three-dimensional (3D) structural characteristics of Light Detection and Ranging (LiDAR) remote sensing have facilitated extraordinary advances in our understanding of urban and environmental systems. Structural characteristics of LiDAR have been applied *'alone'* and through *'fusion'* with optical satellite remote sensing data in a range of applications focused on the built and natural environment. For example, LiDAR has been explored extensively in natural systems at broader scales for quantifying structural characteristics of forest stands, plant species distribution, forest fuel estimation, forest fire management, and classification of coastal vegetation and wetlands (Erdody and Moskal 2010; Gilmore et al. 2008; Hudak et al. 2002; Jones et al. 2010; Koetz et al. 2008; Popescu et al. 2002). In urban systems, LiDAR applications include developing digital surface models of urban landscapes and infrastructure, land-use classification, three-dimensional modeling of buildings, feature extraction, and delineating flood zones (Awrangjeb et al. 2010; Chen et al. 2009; Guo et al. 2011; Lloyd and Atkinson 2002; Meng et al. 2009). The speed in which urbanization is generating complexity in these dynamic systems requires methodological advancements in LiDAR applications to map, measure, and model spatial heterogeneity over large regions.

LiDAR remote sensing holds great promise to resolve three immediate challenges associated with mapping spatial heterogeneity of urban landscapes. First, the spatial heterogeneity inherent to urban environments represents a substantial challenge to discriminating spectrally similar land-use and land-cover (LULC) types found along urban-rural gradients. Integrating structural characteristics of LiDAR with optical remote sensing (ORS) may increase classification performance by improving discrimination among spectrally similar land use types such as forest, farmland, and managed clearings. However, LiDAR-fusion studies are typically limited to small spatial extents due to LiDAR's high cost, small footprint, and large data volume, leaving unanswered questions regarding the efficacy of using LiDAR-optical data fusion approaches to map LULC over large urbanizing regions. Second, the spatial and vertical heterogeneity of complex landscapes exposes the limitations of using conventional approaches to detect and map the growing problem of exotic plant

species invasions in the understory of urban forests (e.g. integrating optical remote sensing with species distribution models (SDMs)). While the predictive accuracy of SDMs varies significantly (Elith et al. 2010) and ORS has limited ability to detect the structure and composition of forested landscapes, LiDAR-derived metrics and their integration with high-resolution ORS data is unexplored to investigate the distribution of understory invasive plant, *Ligustrum sinense*, commonly known as Chinese privet. Third, we are faced with a growing need to better understand the spatial extent of forest resources in urbanizing regions and the ecosystem services they provide. This requires that we overcome the challenge of mapping aboveground biomass of remnant forests over large urbanizing regions with approaches intend to minimize LiDAR point-density while maximizing computational efficacy and accuracy estimates. LiDAR may provide a top-down solution for each of these challenges that also aids the development and implementation of urban forest management.

In the dissertation, I address three research objectives: 1) examine the fusion of LiDAR digital surface models with Landsat TM imagery to improve accuracy of mapping urban landscape heterogeneity over large regional extents, 2) evaluate the performance of LiDAR-derived metrics compare to IKONOS imagery and the combination of these data for detecting and mapping the spatial distribution of invasive understory plant, *Ligustrum sinense*, invasion in urbanizing forest landscapes, and 3) examine the effects of LiDAR point density and landscape context on the estimation of aboveground biomass of remnant forests in urbanizing landscapes. Each research objective represents both an individual chapter in the dissertation and a publishable manuscript. The first chapter is published in the ISPRS Journal of Photogrammetry and Remote Sensing in November 2012 entitled, 'LiDAR-Landsat data fusion for large-area assessment of urban land cover: balancing spatial resolution, data volume and mapping accuracy'.

REFERENCES

Awrangjeb, M., Ravanbakhsh, M., & Fraser, C.S. (2010). Automatic detection of residential buildings using LIDAR data and multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 65*, 457-467

Chen, Y.H., Su, W., Li, J., & Sun, Z.P. (2009). Hierarchical object oriented classification using very high resolution imagery and LIDAR data over urban areas. *Advances in Space Research, 43*, 1101-1110

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution, 1*, 330-342

Erdody, T.L., & Moskal, L.M. (2010). Fusion of LiDAR and imagery for estimating forest canopy fuels. *Remote Sensing of Environment, 114*, 725-737

Gilmore, M.S., Wilson, E.H., Barrett, N., Civco, D.L., Prisloe, S., Hurd, J.D., & Chadwick, C. (2008). Integrating multi-temporal spectral and structural information to map wetland vegetation in a lower Connecticut River tidal marsh. *Remote Sensing of Environment, 112*, 4048-4060

Guo, L., Chehata, N., Mallet, C., & Boukir, S. (2011). Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*, 56-66

Hudak, A.T., Lefsky, M.A., Cohen, W.B., & Berterretche, M. (2002). Integration of lidar and Landsat ETM plus data for estimating and mapping forest canopy height. *Remote Sensing of Environment, 82*, 397-416

Jones, T.G., Coops, N.C., & Sharma, T. (2010). Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific Northwest, Canada. *Remote Sensing of Environment, 114*, 2841-2852

Koetz, B., Morsdorf, F., van der Linden, S., Curt, T., & Allgower, B. (2008). Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data. *Forest Ecology and Management, 256*, 263-271

Lloyd, C.D., & Atkinson, P.M. (2002). Deriving DSMs from LiDAR data with kriging. *International Journal of Remote Sensing, 23*, 2519-2524

Meng, X.L., Wang, L., Silvan-Cardenas, J.L., & Currit, N. (2009). A multi-directional ground filtering algorithm for airborne LIDAR. *ISPRS Journal of Photogrammetry and Remote Sensing, 64,* 117-124

Popescu, S.C., Wynne, R.H., & Nelson, R.F. (2002). Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. *Computers and Electronics in Agriculture, 37,* 71-95

CHAPTER 1. Mapping Urban Landscape Heterogeneity

Manuscript as published in ISPRS Journal of Photogrammetry and Remote Sensing.

Singh, K.K., Vogler, J.B., Shoemaker, D.A., & Meentemeyer, R.K. (2012). LiDAR-Landsat data fusion for large-area assessment of urban land cover: Balancing spatial resolution, data volume and mapping accuracy. *ISPRS Journal of Photogrammetry and Remote Sensing, 74*, 110-121

LiDAR-Landsat data fusion for large-area assessment of urban land cover: balancing spatial resolution, data volume and mapping accuracy

Kunwar K. Singh*, John B. Vogler, Douglas A. Shoemaker, and Ross K. Meentemeyer

Center for Applied Geographic Information Science

Department of Geography and Earth Sciences

University of North Carolina

9201 University City Blvd

Charlotte, NC 28223

USA

*Corresponding author: ksingh9@uncc.edu; phone: 01-704-359-7139; fax: 01-704- 687-5966

john.vogler@uncc.edu

d.shoemaker@uncc.edu

rkmeente@uncc.edu

Abstract

The structural characteristics of Light Detection and Ranging (LiDAR) data are increasingly used to classify urban environments at fine scales, but have been underutilized for distinguishing heterogeneous land covers over large urban regions due to high cost, limited spectral information, and the computational difficulties posed by inherently large data volumes. Here we explore tradeoffs between potential gains in mapping accuracy with computational costs by integrating structural and intensity surface models extracted from LiDAR data with Landsat Thematic Mapper (TM) imagery and evaluating the degree to which TM, LiDAR, and LiDAR-TM fusion data discriminated land covers in the rapidly urbanizing region of Charlotte, North Carolina, USA. Using supervised maximum likelihood (ML) and classification tree (CT) methods, we classified TM data at 30m and LiDAR data and LiDAR-TM fusions at 1 m, 5 m, 10 m, 15 m and 30 m resolutions. We assessed the relative contributions of LiDAR structural and intensity surface models to classification map accuracy and identified optimal spatial resolution of LiDAR surface models for large-area assessments of urban land cover. ML classification of 1 m LiDAR-TM fusions using both structural and intensity surface models increased total accuracy by 32% compared to LiDAR alone and by 8% over TM at 30m. Fusion data using all LiDAR surface models improved class discrimination of spectrally similar forest, farmland, and managed clearings and produced the highest total accuracies at 1 m, 5 m, and 10 m resolutions (87.2%, 86.3% and 85.4%, respectively). At all resolutions of fusion data and using either ML or CT classifier, the relative contribution of the LiDAR structural surface models (canopy height and normalized digital surface model) to classification accuracy is greater than the intensity surface. Our evaluation of tradeoffs between data volume and thematic map accuracy for this study system suggests that a spatial resolution of 5 m for LiDAR surface models best balances classification performance and the computational challenges posed by large-area assessments of land cover.

Keywords: LiDAR, Landsat, fusion, land cover, large-area assessment, mapping accuracy, managed clearings

1. Introduction

Land-cover change in rapidly urbanizing regions is significantly changing the way societies experience their environment, with wide-ranging economic and ecological implications, including local climate change (Guneralp and Seto 2008), biodiversity loss (Gagne and Fahrig 2011), flood hazards (White and Greer 2006), fragmentation (Irwin and Bockstael 2007), and the degradation of ecosystem services (Turner 2010) and aesthetic value (Sander et al. 2010). The speed at which urbanization generates spatial heterogeneity and landscape fragmentation (Irwin and Bockstael 2007) makes it difficult to accurately track land-use and land-cover (LULC) changes at a desired scale and reasonable cost (Esch et al. 2009). Remotely-sensed data and imagery provide a comprehensive, scalable means for detecting and quantifying LULC change, and its use in mapping urban growth, estimating population density, and modeling sustainability and quality of life is becoming increasingly popular as the scale, cost, and spatial-temporal coverage improve (Rogan and Chen 2004). However, the spatial heterogeneity inherent to urban environments represents substantial challenges to discriminating LULC types using remotely sensed data. Spectral mixtures of vegetation and impervious surfaces common in transitory urbanizing landscapes challenge the ability of spectral-based, hard classification algorithms, such as maximum likelihood (ML), to detect unique signatures and accurately assign pixels to a probable dominant class (Lo and Choi 2004). Standard LULC classification schema for moderate-resolution data at regional and greater scales often lack the specificity (e.g., "mixed" class) and completeness (e.g., "other" class) necessary for accurate representation of complex urbanizing landscapes. For example, the Anderson Classification (Anderson 1976) mixes the concepts of LULC at all levels, which is insufficient to measure environmental impacts and facilitate urban growth prediction and planning (Ridd 1995). These systemic sources of error are manifested in a number of classification products, such as the widely used Landsat-based National Land Cover Database in which accuracy varies tremendously along both spatial and temporal scales (Wickham et al. 2010) with recognized difficulties in distinguishing low-density development (Irwin and Bockstael 2007).

Alternative methodologies exist that move beyond per-pixel, spectral-based approaches in an effort to improve LULC discrimination in urban systems. For example, spectral mixture models, fuzzy sets, and neural networks produce sub-pixel multivariate outputs amenable to the environmental continuum (Carpenter et al. 1999; Wu and Murray 2003; Zhang and Foody 2001). Object-based algorithms may also be used to exploit the spatial configuration of groups of neighboring pixels to improve accuracy (Chen et al. 2012). Other approaches integrate data from sensors with much finer spectral resolutions (eg., hyperspectral) or that capture the vertical structure of the built and natural environment (eg., radar, laser) (Guo et al. 2011; Koch 2010). Spatial and spectral resolutions are foremost considerations in the choice of different sensors, image classifiers, and classification schemes to accurately represent the heterogeneity of human-made and natural features found along urban-rural gradients. Moderate-resolution, multispectral imagery, such as Landsat series data, is one of the primary sources of concurrent, globally-available imagery used in mapping landscapes over time and at broad geographical scales. However, the limited spatial and spectral resolution provided by Landsat sensors makes discrimination of spectrally-similar vegetated covers prevalent in transitioning urban and suburbanizing landscapes difficult, leading to unsatisfactory classification results (Wickham et al. 2010).

The Light Detection and Ranging (LiDAR) airborne laser scanner has emerged as an increasingly popular tool for collecting very high-resolution structural data representing the vertical dimension of the Earth's surface by measuring the travel time of laser pulses between the sensor and earth objects (Jensen 2007). LiDAR data have been used in a range of environmental applications such as creating digital surface models of urban landscapes and infrastructure (Lloyd and Atkinson 2002; Meng et al. 2009), quantifying structural characteristics of forest stands (Dubayah and Drake 2000; Lim et al. 2003; Popescu et al. 2002), and delineating flood zones (Raber et al. 2007). However, the relatively narrow range of spectral information has limited the utility of LiDAR in LULC assessments (Chen 2007).

Fusion of LiDAR and optical remote sensing data has been explored in natural systems for mapping forest structural attributes, plant species distribution, forest fuel estimation, forest fire management, and classification of coastal areas and wetlands (Erdody

and Moskal 2010; Gilmore et al. 2008; Hudak et al. 2002; Jones et al. 2010; Koetz et al. 2008; Popescu et al. 2002). In urban systems, LiDAR and LiDAR-optical fusion data have been applied to LULC classification, 3D modeling, and feature extraction (Awrangjeb et al. 2010; Chen et al. 2009; Guo et al. 2011; Meng et al. 2009). Though promising, fusion studies are typically limited to small spatial extents due to LiDAR's high cost, small footprint, and large data volume, leaving questions unanswered regarding the efficacy of using LiDAR-optical data fusion approaches to map land cover over large, urbanizing regions (1000s of km^2). For example, how can we best leverage the high resolution of LiDAR data and its structural and intensity surface models to identify coarser resolutions that, when fused with moderate-resolution multispectral imagery, strikes a balance between computational efficiency and classification accuracy for mapping LULC over large, spatially and vertically heterogeneous regions?

This study examines the degree to which fusion of structural and intensity surface models of LiDAR with Landsat TM imagery improves mapping accuracy of spectrally similar LULC types found along urban-rural gradients, including forest, farmland, and managed clearings. Using Mecklenburg County in the Greater Charlotte Metropolitan area of North Carolina as a case study, we classified Landsat TM imagery at 30 m, and LiDAR and LiDAR-TM fusions at five resolutions (1 m, 5 m, 10 m, 15 m and 30 m) by applying two commonly used classification algorithms: maximum likelihood (ML) and classification tree (CT). We quantified the accuracy of resulting thematic maps using total accuracy and three class-level accuracies. This evaluation allowed us to 1) determine the best performing combination of data inputs and classifier, 2) assess individual and combined contributions of LiDAR structural and intensity surface models to classification accuracy using fusion data, and 3) identify an optimal spatial resolution for processing LiDAR surface models that maximizes computational efficiency while maintaining high classification accuracy. The outcome of this study is intended to facilitate further application of LiDAR surface models over large regions in conjunction with moderate resolution remote sensing data to improve LULC classification along urban-rural gradients.

2. Methods

2.1. Study system

Mecklenburg County (1415 km^2) is located within the Piedmont physiographic province in the center of the Charlotte Metropolitan Area of North Carolina, USA (Fig. 1a). Regional topography of this rapidly urbanizing region is characterized by a rolling landscape which increases in elevation from approximately 91 m in the eastern part of the county to about 457 m in the west. Historically, the region was dominated by oak-hickory-pine forests, many of which were logged and converted to agricultural lands. Since the mid-1970s, the Charlotte region has experienced rapid growth, transitioning from a primarily forest and agricultural landscape to a mix of high- to low-density urbanized land uses. Between 1985 and 2008, the county converted 33% of its tree cover and 3% of its open space to urbanized lands. Over the same period, the area of impervious surfaces, such as roads, buildings and parking lots, increased 60%, a trend that is expected to continue (Meentemeyer et al. 2013).

2.2. Landsat TM data

We acquired two Landsat TM images dated 21 May 2007 from the USGS EROS center. We then subset the three visible bands (blue, 0.45 - 0.52 µm; green, 0.52 - 0.60 µm; red, 0.63 - 0.69 µm) and three short-wave infrared bands (nir, 0.76 - 0.90 µm; mir1, 1.55 - 1.75 µm; mir2, 2.08 -2.35 µm) with spatial resolution of 30 m (Fig. 1b) and excluded the thermal band from further analysis. We reprojected the 6-band images into the State Plane Coordinate System (NC FIPS 3200, NAD 1983, meters). We calibrated the images to exoatmospheric (at-sensor) reflectance and atmospherically corrected them using the image-based COST method (Chavez 1996). We converted the raw digital numbers to at-sensor radiance based on published TM standard calibration coefficients followed by conversion to at-sensor reflectance (Mahiny and Turner 2007).

2.3. LiDAR data

We obtained 691 tiles of airborne LiDAR data, each with dimensions 1524 m x 1524 m, (Fig. 1b) from the Storm Water Services Division of Charlotte-Mecklenburg County government office. Original data acquisition by Kucera International, Inc. (Willoughby, Ohio, USA) occurred during 12-15 February 2007. LiDAR data were captured using a Leica ALS50 sensor coupled with a POSAV Applanix global positioning system (GPS)/inertial navigation system (INS) system mounted to a fixed wing aircraft. During the survey, GPS/INS data were acquired and processed in forward and reverse directions and concurrent ground-based GPS base station data was collected for georeferencing the LiDAR sensor's position, angle, and orientation relative to the horizontal and vertical coordinate datums. The sensor recorded first and last returns and intensity at a platform altitude of 1676.4 m above mean terrain (AMT) with an average point spacing of 1.4 m between any two neighboring points. Horizontal and vertical accuracies are 0.30 m RMSE and 0.18 m RMSE, respectively, and meet the Geospatial Positioning Accuracy Standards Part 3: National Standard for Spatial Data Accuracy (NSSDA) published by the Federal Geographic Data Committee (Jensen 2007).

2.3.1. LiDAR data processing

Each LiDAR tile is comprised of a dense collection of georeferenced points with the following attributes: x coordinate, y coordinate, return height (z), return type (first (top-of-canopy height) or last (ground level height)), and intensity, a spectral property of LiDAR data that measures the amount of energy backscattered from features on the Earth’s surface. We interpolated point data for each LiDAR tile using the LiDAR Analyst extension for ArcGIS 9 (LiDAR Analyst Edition 4.2, Visual Learning Systems, Inc., Missoula, MT) and produced 1 m, 5 m, 10 m, 15 m, and 30 m resolution raster datasets representing first return (Fig. 2a), last return (Fig. 2b), bare-earth (Fig. 2c) and intensity. Using these basic inputs, we developed three LiDAR surface models at key resolutions for each of the 691 tiles, including a canopy height model (Fig. 2d), a normalized digital surface model (Fig. 2e), and an

intensity surface (Fig. 2f). Processing surface models at multiple, coarser resolutions allowed us to investigate alternate resolutions for regional extent analyses and significantly reduced file size while increasing data processing efficiency. For example, the county-wide canopy height model was reduced from approximately 5 GB (gigabytes) at 1 m resolution to approximately 200 MB (megabytes) at 5 m resolution.

2.3.2. Canopy height model

The canopy height model (CHM) provided estimates of vegetation height in tenths of meters (Fig. 2d). We derived the CHM by subtracting the LiDAR last return raster (Fig. 2b) from the first return (Fig. 2a). Because vegetation typically provides a first return (top-of-canopy) and last return (ground level), this simple approach is effective for distinguishing vegetation canopy from human-made features in heterogeneous urbanized landscapes. Previous studies subtracted bare earth digital terrain models (DTM) from the LiDAR first return, which is more suitable in homogeneous forested landscapes (Zimble et al. 2003). CHM as derived in this study accurately represents the extent of canopy cover along urban-rural gradients.

2.3.3. Normalized digital surface model

The normalized digital surface model (nDSM) represents the absolute elevation of aboveground natural and human-made features, excluding the influence of topography (Fig. 2e) (Hofle et al. 2012). We developed the nDSM by subtracting the bare-earth DTM from the LiDAR last return. We derived the DTM by filtering the first- and last-return data to remove all non-ground points and then interpolating ground elevation. We then applied a low-pass filter to the resulting DTM to remove noise associated with any remaining non-ground features. Subtracting the bare-earth DTM from the LiDAR last return surface produced the nDSM, or absolute elevations of aboveground features. Discrimination of natural and human-made features using nDSM alone is not straightforward and Haala and Brenner (1999) suggest that nDSM cannot be used for differentiation and extraction of streets or LULC.

However, by combining nDSM with the CHM and intensity surface models, it is possible to effectively highlight built-up areas and enhance discrimination of LULC (Salah et al. 2011).

2.3.4. Intensity surface

We developed an intensity surface (IS) by applying the inverse distance weighting (IDW) interpolator to intensity values associated with all LiDAR return points. The resulting 8-bit panchromatic IS contained cells values ranging from 0 – 255 (Fig. 2f). While the IS is somewhat limited in its use for distinguishing between some impervious surfaces and tree canopy in urbanized landscapes, it contains useful information for effectively discriminating many other LULC types (Jensen 2007). Depending on acquisition season, LiDAR intensity values vary appreciably for different targets. For example, Yoon et al. (2008) found higher intensity values associated with grass when compared to other vegetation types. Other studies demonstrated the efficacy of LiDAR intensity data for tree species classification, discerning age classes, and distinguishing deciduous and coniferous forest classes (Holmgren et al. 2008; Singh et al. 2010).

2.4.Data fusion

We fused LiDAR surface models and TM data at multiple resolutions to assess differences in classification accuracy over a range of spatial resolutions and to ascertain individual and combined contributions of LiDAR structural (CHM + nDSM) and intensity (IS) surface models. We first resampled 30 m TM data to 1, 5, 10, and 15 m resolutions using nearest neighbor (NN) resampling to minimize loss of original pixel values at finer resolutions (Gardner et al. 2008; Khan et al. 1995; Raptis et al. 2003). We then used layerstacking to combine LiDAR surface models and TM data into three types of composite images at the five key resolutions: 1) CHM + nDSM + IS + TM , 2) CHM + nDSM (LiDAR structural) + TM, and 3) IS (LiDAR intensity) + TM (Huang et al. 2007). In addition, we produced a LiDAR only composite image by layerstacking CHM, nDSM, and IS models.

Including the original TM imagery at 30 m resolution, we prepared 21 datasets for classification and accuracy assessment.

2.5. Classification schema

We identified six classes that characterize the dominant environments of the Charlotte region's urban-rural gradient: impervious surfaces, managed clearings, farmland, forest, water and barren land. These classes follow the Anderson Level 1 schema (Anderson 1976) with one notable exception: to improve specificity and better represent the complexity of the urban landscape, we divided the urban or built-up class of the Anderson Level I schema into impervious surfaces and managed clearings categories (Table 1). Impervious surfaces consist of residential, commercial, and industrial infrastructure, including buildings, parking lots, airports, and roads. Managed clearings are the common vegetated land surfaces in metropolitan regions that are highly managed, transitional, and play an important role in the functioning of ecosystem services in urban landscapes. Examples include golf course fairways and greens, grassy medians and shoulders along roads, lands under electrical transmission lines, lawns and recreational sports complexes. Neither developed nor natural, these managed clearings in aggregate represent an important urban component of particular interest due to their ability to alter ecosystem function, potential for state change, and cultural importance (e.g. recreational green space) (Hunhammar and Bolund 1999).The farmland class represents multiple constituent land covers including various agricultural croplands and pasture. Our remote sensing data was acquired primarily during the "leaf-off", non-growing season when constituent farmland classes possess more similar spectral and height signatures.

2.6. Classification algorithms

We classified TM, LiDAR, and LiDAR-TM fusion datasets using two classifiers: supervised maximum likelihood (ML) (Esch et al. 2009) and classification trees (CT) (Friedl and Brodley 1997). Both ML and CT classifiers required training data representative of the

six selected classes. Table 2 lists the per-class numbers of classification training sites for the classifiers. Following Ediriwickrema and Khorram (1997) and to support the ML classifications, we identified 155 training areas (polygons) of homogeneous groups of pixels representative of each class. Ground reference used for verification of training area classes included July 2006 aerial imagery from the National Agricultural Imagery Program (NAIP) and March 2007 Google imagery. To support CT classification, which uses training points, we randomly sampled 10 locations within each of the 155 ML training areas. Analysis of variance (ANOVA) tests revealed no significant spectral differences between the ML training areas and the sampled CT points (F (1, 153) = 1.02, p = 0. 903).

The commonly used Gaussian-based, supervised ML algorithm is a simple, yet robust classifier for analyzing pixels across multiple bands of data and assigning class membership based on statistical probability (Bork and Su 2007). During ML estimation, a probability density function is calculated for each class based on mean and covariance statistics generated from training data. Here, training data consisted of collections of homogenous pixels inside polygons of the classes of interest (Table 2). By comparing image pixel values with the class-level statistics, the ML classifier is able to assign each pixel to its most probable class (Hagner and Reese 2007). ML performance is, however, generally limited by frequency distribution assumptions (Tso and Mather 2009), and classification time increases significantly with higher data dimensionality (Jia and Richards 1994). For comparison, and in an effort to circumvent the ML limitations, we also applied a CT classifier, which is a non-parametric approach to pattern recognition.

Decision trees such as CT provide an alternative LULC categorization method for remote sensing data using a hierarchical splitting mechanism (Tso and Mather 2009). CT has potential advantages over the ML estimator as it 1) processes data at different scales with no assumptions regarding the frequency distributions of the image data (Tooke et al. 2009) and 2) requires minimal computational time (Pal and Mather 2003). The CT classifier identifies and constructs a binary structured decision tree by recursively splitting image data at each node on the basis of a statistical test that increases the homogeneity of the training data (Friedl and Brodley 1997). Once a binary decision tree is constructed, it is applied to the

entire image dataset to produce a categorical map. We used training points sampled from homogenous ML training areas of known class types as the dependent variable. Landsat TM and LiDAR surface models values at these same locations served as independent variables in developing the binary decision tree through recursive splitting. We applied Akaike's information criterion and deviance pruning method with cross-validation to construct binary decision trees for each class of the 21 datasets.

2.7. Accuracy assessments

To assess classification performance, we generated 446 randomly distributed points across the study system for comparing each ML and CT classification map to ground reference data. The number of validation points (Table 2) is based on a multinomial distribution (Congalton and Green 1999), and we specified a minimum distance of 0.2 km between points. This design insured that all locations in the study system had an equal probability of selection and that all classes were adequately represented. Ground reference data consisted of high-resolution aerial NAIP imagery (July 2006) and March 2007 Google imagery. We used the same set of validation points for each thematic map to insure consistent and comparable accuracy assessments. We evaluated and compared the total map accuracy of 42 classification maps. Total accuracy reflects the percentage of validation points correctly classified when compared to ground reference.

We quantified and compared class-level accuracies using three metrics: producer's accuracy, user's accuracy, and an overall class accuracy. Overall class-level accuracy accounts for both errors of omission and commission in a single metric and is derived for each class as follows:

$$N_{correct} \; / \; (N_{correct} + N_{omission} + N_{commission})$$

where $N_{correct}$ is the number of correctly classified validation points for a class, $N_{omission}$ is number of omission errors, and $N_{commission}$ is the number of commission errors.

3. Results

3.1. Land cover mapping and classification performance

The LiDAR-TM fusion approach outperformed both the Landsat TM- and LiDAR-alone approaches in total classification accuracy. Analysis of 1 m fusion data using all LiDAR surface models and ML classifier produced the highest total accuracy (87.2%), improving on TM data alone and LiDAR data alone by 8% and 32%, respectively (Table 3). Total accuracy increased with finer resolution input data for both the fusion and LiDAR-only approaches regardless of classifier. The ML classifier generally performed better than the CT classifier with the exception of the LiDAR-only approach, where CT accuracies were on average 8-9% higher than ML classifications at all resolutions. ML classification accuracies using fusion data and all LiDAR surface models ranged from 81% to 87% for 30 m and 1 m fused data, respectively. CT classification accuracies increased from 78% to 84%. On average, classification accuracies based on fusion data and ML classifier were 4% higher than accuracies obtained using CT classifier. By comparison, ML classification of 30 m TM data resulted in a modest 0.7% improvement in total accuracy over the CT classifier. The 1 m, 5 m, and 10 m fusion inputs using the ML classifier produced total accuracies exceeding 85%, the accepted minimum standard for total accuracy in land cover mapping investigations (Anderson 1976; Rogan et al. 2003).

3.2.Contributions of LiDAR surface models

LiDAR's structural surface models (CHM + nDSM), when fused with TM data, contributed more to total classification accuracy than the intensity surface model regardless of classifier applied. On average, ML classification accuracies across all resolutions using LiDAR structural-TM fusion data were 2.6% higher than accuracies obtained from fusion of intensity surface model with TM data, and 3.8% higher than TM alone. CT classification accuracies derived from fused structural surface models with TM data were, on average, 1.6% higher than accuracies using fusion of the intensity surface model with TM data, and 1.2% higher than TM alone (Table 3). Applying the ML classifier to LiDAR structural-TM

fusion data at 1 m resolution produced a total accuracy of 85%, which meets the accepted minimum standard for total accuracy.

3.3.Discrimination of land cover classes

Evaluation of land cover discrimination at the class level revealed general uptrends in performance associated with finer resolution fusion data using structural and intensity surface models with either ML or CT classifier. The ML classifier provided consistently higher producer's and user's accuracies across multiple, higher resolutions of fusion data, particularly at 1 m, 5 m, and 10 m. Table 4 compares class-level accuracies from the analyses of TM data (30 m and ML classifier) and the top-performing fusion approaches (1 m, 5 m, and 10 m using ML classifier exceeded mapping accuracy standards). LiDAR-only results were sufficiently poor (Table 3) to warrant no further comparison. Fusion data using all LiDAR surface models resulted in net gains in producer's accuracies over TM data for all classes except impervious surfaces and managed clearings. Net increases in user's accuracies were realized for all classes and resolutions, barren land being one exception (decreased in all cases), and water, which exhibited no change at 5 m and 10 m fusion resolutions. Figure 3a illustrates the net change (gains or losses) in class-level accuracies based on TM data vs. 1 m fusion data using all LiDAR surface models. While producer's accuracy decreased 9% for impervious and only marginally for managed clearings, farmland discrimination improved nearly 40% using 1 m fusion data. Improvements in user's accuracy exceeded 10% for impervious surfaces and 15% for the spectrally-similar managed clearings, farmland, and forest classes, while water improved slightly and barren land decreased 8%. Accounting for both errors of omission and commission in the overall class-level metric showed that 1 m fusion data using all LiDAR surface models provided net gains in mapping accuracy for all classes (Fig. 3a) with similar gains obtained utilizing 5 m and 10 m fusion data (Table 4).

Figure 3b illustrates net change in class-level accuracies based on 1 m LiDAR structural-TM fusion data vs. LiDAR intensity-TM fusion. The addition of LiDAR structural surface models improved producer's, user's, and overall class-level accuracies for

impervious surfaces, farmland, and forest over LiDAR intensity-TM fusion data. In particular, producer's and overall farmland class accuracies increased 18% and 15%, respectively. User's and overall class accuracies for impervious surfaces saw 12% accuracy gains. In contrast, class-level accuracies for managed clearings and water decreased slightly compared to LiDAR intensity-TM fusion data. Barren land showed no change in producer's accuracy and modest gains in user's and overall class accuracy.

3.4.Optimal data resolution for large area land-use assessments

Classification accuracy generally decreased as we processed input fusion datasets at coarser resolutions (see Table 3 and Table 5). Figure 4 illustrates how processing inputs at coarser resolutions decreased file sizes disproportionately to reductions in classification accuracy as measured by the total accuracy based on ML classifier. Use of 5 m vs. 1 m fusion data reduced input file size by 16% with a < 1% reduction in total accuracy. Use of 10 m fusion data reduced input file size by 38% accompanied by a 1.8% reduction in the total accuracy which still exceeded 85%, the accepted minimum standard for total accuracy in land cover mapping (Anderson 1976; Rogan et al. 2003). The 5 m and 10 m fusion inputs also produced mapped land cover proportions consistent with the 1 m fusion data (Fig. 5). For large-area assessment of urban land-cover heterogeneity, we found that classified maps based on 5 m or 10 m LiDAR-TM fused data provided acceptable spatial resolution and mapping accuracy while increasing efficiency of data management and analysis due to smaller data volume.

4. Discussion

Rapidly urbanizing regions require frequent and accurate assessment of available land covers, but their regional extents, spatial heterogeneity, and landscape fragmentation present substantial challenges to accurate classification. In this Charlotte region case study, we found that the 85% mapping accuracy standard (Anderson 1976; Rogan et al. 2003) was only met with classifications produced from fusions of Landsat TM imagery and surface models

extracted from LiDAR data. Overall, the use of fusion data for classification resulted in accuracies that exceeded those produced from TM or LiDAR data alone (Table 3) and improved discrimination between structurally variable land covers (Fig. 3). Total and class-level accuracies increased as the resolution of fused input data approached the higher native resolution of LiDAR (Donoghue and Watt 2006; Hudak et al. 2002). These results corroborate the findings of Lu et al. (2011) and Wickham et al. (2010), which note the inadequacy of TM data to meet map accuracy standards in highly heterogeneous urban landscapes. Higher spatial resolution multispectral data such as IKONOS or WorldView2 may provide adequate classification mapping accuracy in complex urban environments; however, they lack the vertical information provided by LiDAR, and perhaps more importantly, when compared to Landsat data, these commercial satellite data have much smaller footprints, higher data volumes, and potentially prohibitive costs for large-area land cover assessments.

Analysis of ML classifications produced from LiDAR structural-TM fusion data revealed gains in both total accuracy at multiple resolutions (Table 3) and class-level accuracies of impervious surfaces, farmland, and forest (Fig. 3b). This demonstrates the contribution of the height-above-ground information, extracted in the CHM and nDSM surface models, in differentiating the vertical features of these classes from those classes with minimal vertical structure (e.g., managed clearings, water, and barren land). In contrast, fusion of LiDAR intensity-TM data provided the same or only slightly better (at higher resolutions) total accuracy above TM alone and contributed to modest accuracy improvements to managed clearing and water classes (Garcia et al. 2010). The individual contributions of LiDAR structural and intensity surfaces to different land cover classes (Fig. 3b) and their combined contribution to overall class-level accuracies (Fig. 3a and Table 4) supports fusion of all LiDAR surface models with TM data .

We identified two primary sources of error in the ML classification of TM data and fusion data: 1) errors of omission associated with misclassification of barren land as impervious surfaces, and 2) errors of omission and commission between farmland and managed clearings classes. While classification of fusion data using all LiDAR surface

models did not completely ameliorate the barren land-impervious surfaces confusion, discrimination of vegetated land covers improved, in particular, enhanced discrimination of farmland and managed clearings. This is a significant outcome given the prevalence of managed clearings in this urbanizing region (Fig. 5 and Fig. 6) and noting the utility of our modification of the Anderson Level I schema to include this important land cover that is often subsumed within the urban/built-up category. Other issues with LiDAR data quality and processing also impact classification accuracy. Gaps between LiDAR tiles are one potential source of error introduced into fusion data. Missing or incomplete LiDAR point data within LiDAR tiles may also contribute to misinterpretation in classification approaches. And LiDAR intensity was not normalized to a standard range, which could also negatively impact mapping results.

Overall LiDAR-only results were very poor (Table 3), however, high class-level accuracy of forest mapping across multiple resolutions supports the capability of LiDAR in characterizing forests attributes (Dubayah and Drake 2000; Lim et al. 2003; Popescu et al. 2002). These results were offset by poor performance of LiDAR-only data at distinguishing managed clearings from farmland where structural differences were not distinct due to leaf-off season and given low variability in intensity values. This warrants a comparative study between leaf-on and leaf-off LiDAR data for the study system and also highlights the importance of integration with optical remote sensing data. Further improvements in mapping of farmland and managed clearings classes may be realized if LiDAR data are acquired during the growing season as it would increase variation in nDSM and IS (Yaokui et al. 2010).

The selection of suitable parametric and non-parametric classifiers, depending upon the data type, can help in managing large data volume and data dimensionality. Multisource data provides greater dimensionality with increased data volume, and therefore, increases the chances of Hughes phenomenon (Hughes 1968), which can hinder efficient utilization of classification algorithms. Total mapping accuracy using LiDAR-TM fusions and ML or CT classifier improved with both increasing spatial resolution and increasing number of input bands, seemingly unaffected by data dimensionality. This observation fits with the findings

of Dalponte et al. (2009) and can be potentially explained by the limited number of bands (maximum nine) with a substantially high number of training sites used in this study.

Although we did not record computational time, we found the CT classifier to be more flexible and efficient in handling higher dimensional data with minimum required computational time. However, the performance of the ML classifier in terms of mapping accuracy was consistently better when applied to fusion data (Table 3). Higher CT accuracies using LiDAR-only data and CT's equivalent accuracy to ML using TM-only could be ascribed to the non-Gaussian distribution of the LiDAR surface models with respect to the Gaussian distribution of the TM data, which is consistent with research presented by Pal and Mather (2003). Frey (1995) found that the ML classifier performs equally well using Gaussian or non-Gaussian image data, which may explain its consistently better performance with fusion data. The CT classifier's relatively poor performance using fusion data may be due to the spatial distribution of training points; however, we found no significant spectral differences between the ML training areas and the sampled CT points used in the classifications. The mix of Gaussian and non-Gaussian image data in the fusion data could also explain CT classifier's poor performance. Likewise, the CT classifier constructs binary decision trees by recursively splitting each band of a composite image, it may be that multisource remote sensing data classifications require a combination of bands for proper representation of a class type (Li et al. 2011; Pal and Mather 2003).

This study illustrated that the use of coarser resolutions of fusion data and correspondingly smaller file sizes was accompanied by a reduction in the total classification accuracy (Fig. 4). However, the difference in total accuracies between the use of 1 m and 10 m fusion data was modest and disproportionate to a significant reduction in file size. These results prompted a practical methodological question as to the optimal resolution for analyzing fusion data that enhances data handling and computational efficiency while providing sufficient thematic map accuracy. In this research, our evaluation of tradeoffs between data volume and thematic map accuracy suggests that a spatial resolution of 5 m for fusion data best balances classification performance and the computational challenges posed by large-area assessments of land cover using LiDAR. Ultimately, the ideal resolution

depends heavily on specific research objectives, study system extent and spatial resolution requirements, and available computing resources.

5. Conclusions

Accurate and detailed mapping of urban landscape heterogeneity at regional scales is important for its significant implications on sustainable planning and measuring urban growth's impact on available natural resources. Our results suggest that fusion of higher-resolution structural and intensity surface models from LiDAR with TM data improved total and class-level classification accuracies when mapping urban landscape heterogeneity over regional extents vs. Landsat TM or LiDAR alone. LiDAR structural surface models (CHM + nDSM), when fused with TM data, contributed relatively more to total classification accuracy than LiDAR intensity, emphasizing the importance of the vertical dimension in urban LULC mapping. Classification of fusion data using structural and intensity surface models from LiDAR enhanced discrimination of spectrally similar land covers, in particular managed clearings, farmland and forest. Our modification of the Anderson Classification Level I schema to accommodate the managed clearings class is significant as it is the dominant cover type in Mecklenburg County, NC, and plays a unique role in the functioning and provision of ecosystem services in urban landscapes, yet is frequently miss-categorized as urban or built-up in the Anderson Classification Level I schema. This research further highlights the importance of establishing an optimal resolution for LiDAR surface models before integrating with moderate resolution multispectral remote sensing data for mapping urban landscape heterogeneity over large areas. Our evaluation of tradeoffs between data volume and mapping accuracy for this study system suggests that a spatial resolution of 5 m for LiDAR surface models best balances classification performance and the computational challenges posed by large-area land cover assessments.

Acknowledgements

We gratefully acknowledge financial and logistical support from the Renaissance Computing Institute (RENCI) and the National Science Foundation ULTRA-Ex program (BCS-0949170). We also express sincere thanks to Storm Water Services Division – Charlotte-Mecklenburg County Government Office and Engineering and Property Management, Land Development Services, City of Charlotte, for providing LiDAR data. Finally, we thank students and staff from the Center for Applied Geographic Information Science, UNC Charlotte, and Q. Meng for valuable comments and feedback on this manuscript.

REFERENCES

Anderson, J.R. (1976). *A Land use and land cover classification system for use with remote sensor data*. Washington: U.S. Govt. Print. Off.

Awrangjeb, M., Ravanbakhsh, M., & Fraser, C.S. (2010). Automatic detection of residential buildings using LIDAR data and multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 65*, 457-467

Bork, E.W., & Su, J.G. (2007). Integrating LIDAR data and multispectral imagery for enhanced classification of rangeland vegetation: A meta analysis. *Remote Sensing of Environment, 111*, 11-24

Carpenter, G.A., Gopal, S., Macomber, S., Martens, S., Woodcock, C.E., & Franklin, J. (1999). A neural network method for efficient vegetation mapping. *Remote Sensing of Environment, 70*, 326-338

Chavez, P.S. (1996). Image-based atmospheric corrections revisited and improved. *Photogrammetric Engineering & Remote Sensing, 62*, 1025-1036

Chen, G., Hay, G.J., Carvalho, L.M.T., & Wulder, M.A. (2012). Object-based change detection. *International Journal of Remote Sensing, 33*, 4434-4457

Chen, Q. (2007). Airborne lidar data processing and information extraction. *Photogrammetric Engineering & Remote Sensing, 73*, 109-112

Chen, Y.H., Su, W., Li, J., & Sun, Z.P. (2009). Hierarchical object oriented classification using very high resolution imagery and LIDAR data over urban areas. *Advances in Space Research, 43*, 1101-1110

Congalton, R.G., & Green, K. (1999). *Assessing the accuracy of remotely sensed data : principles and practices*. Boca Raton, Florida: Lewis Publications

Dalponte, M., Bruzzone, L., Vescovo, L., & Gianelle, D. (2009). The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. *Remote Sensing of Environment, 113*, 2345-2355

Donoghue, D.N.M., & Watt, P.J. (2006). Using LiDAR to compare forest height estimates from IKONOS and Landsat ETM+ data in Sitka spruce plantation forests. *International Journal of Remote Sensing, 27*, 2161-2175

Dubayah, R.O., & Drake, J.B. (2000). Lidar remote sensing for forestry. *Journal of Forestry, 98,* 44-46

Ediriwickrema, J., & Khorram, S. (1997). Hierarchical maximum-likelihood classification for improved accuracies. *IEEE Transactions on Geoscience and Remote Sensing, 35,* 810-816

Erdody, T.L., & Moskal, L.M. (2010). Fusion of LiDAR and imagery for estimating forest canopy fuels. *Remote Sensing of Environment, 114,* 725-737

Esch, T., Himmler, V., Schorcht, G., Thiel, M., Wehrmann, T., Bachofer, F., Conrad, C., Schmidt, M., & Dech, S. (2009). Large-area assessment of impervious surface based on integrated analysis of single-date Landsat-7 images and geospatial vector data. *Remote Sensing of Environment, 113,* 1678-1690

Frey, E. (1995). An examination of distributional assumptions in landsat tm imagery. In

Friedl, M.A., & Brodley, C.E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment, 61,* 399-409

Gagne, S.A., & Fahrig, L. (2011). Do birds and beetles show similar responses to urbanization? *Ecological Applications, 21,* 2297-2312

Garcia, M., Riano, D., Chuvieco, E., & Danson, F.M. (2010). Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. *Remote Sensing of Environment, 114,* 816-830

Gardner, R.H., Lookingbill, T.R., Townsend, P.A., & Ferrari, J. (2008). A new approach for rescaling land cover data. *Landscape Ecology, 23,* 513-526

Gilmore, M.S., Wilson, E.H., Barrett, N., Civco, D.L., Prisloe, S., Hurd, J.D., & Chadwick, C. (2008). Integrating multi-temporal spectral and structural information to map wetland vegetation in a lower Connecticut River tidal marsh. *Remote Sensing of Environment, 112,* 4048-4060

Guneralp, B., & Seto, K.C. (2008). Environmental impacts of urban growth from an integrated dynamic perspective: A case study of Shenzhen, South China. *Global Environmental Change-Human and Policy Dimensions, 18,* 720-735

Guo, L., Chehata, N., Mallet, C., & Boukir, S. (2011). Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*, 56-66

Haala, N., & Brenner, C. (1999). Extraction of buildings and trees in urban environments. *ISPRS Journal of Photogrammetry and Remote Sensing, 54*, 130-137

Hagner, O., & Reese, H. (2007). A method for calibrated maximum likelihood classification of forest types. *Remote Sensing of Environment, 110*, 438-444

Hofle, B., Hollaus, M., & Hagenauer, J. (2012). Urban vegetation detection using radiometrically calibrated small-footprint full-waveform airborne LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing, 67*, 134-147

Holmgren, J., Persson, A., & Soderman, U. (2008). Species identification of individual trees by combining high resolution LIDAR data with multi-spectral images. *International Journal of Remote Sensing, 29*, 1537-1552

Huang, H., Legarsky, J., & Othman, M. (2007). Land-cover classification using Radarsat and Landsat imagery for St. Louis, Missouri. *Photogrammetric Engineering & Remote Sensing, 73*, 37-43

Hudak, A.T., Lefsky, M.A., Cohen, W.B., & Berterretche, M. (2002). Integration of lidar and Landsat ETM plus data for estimating and mapping forest canopy height. *Remote Sensing of Environment, 82*, 397-416

Hughes, G.F. (1968). On Mean Accuracy of Statistical Pattern Recognizers. *IEEE Transactions on Information Theory, 14*, 55-+

Hunhammar, S., & Bolund, P. (1999). Ecosystem services in urban areas. *Ecological Economics, 29*, 293-301

Irwin, E.G., & Bockstael, N.E. (2007). The evolution of urban sprawl: Evidence of spatial heterogeneity and increasing land fragmentation. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 20672-20677

Jensen, J.R. (2007). *Remote sensing of the environment : an earth resource perspective*. (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall

Jia, X.P., & Richards, J.A. (1994). Efficient Maximum-Likelihood Classification for Imaging Spectrometer Data Sets. *IEEE Transactions on Geoscience and Remote Sensing, 32*, 274-281

Jones, T.G., Coops, N.C., & Sharma, T. (2010). Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific Northwest, Canada. *Remote Sensing of Environment, 114*, 2841-2852

Khan, B., Hayes, L.W.B., & Cracknell, A.P. (1995). The Effects of Higher-Order Resampling on Avhrr Data. *International Journal of Remote Sensing, 16*, 147-163

Koch, B. (2010). Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS Journal of Photogrammetry and Remote Sensing, 65*, 581-590

Koetz, B., Morsdorf, F., van der Linden, S., Curt, T., & Allgower, B. (2008). Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data. *Forest Ecology and Management, 256*, 263-271

Li, G.Y., Lu, D.S., Moran, E., & Hetrick, S. (2011). Land-cover classification in a moist tropical region of Brazil with Landsat Thematic Mapper imagery. *International Journal of Remote Sensing, 32*, 8207-8230

Lim, K., Treitz, P., Wulder, M., St-Onge, B., & Flood, M. (2003). LiDAR remote sensing of forest structure. *Progress in Physical Geography, 27*, 88-106

Lloyd, C.D., & Atkinson, P.M. (2002). Deriving DSMs from LiDAR data with kriging. *International Journal of Remote Sensing, 23*, 2519-2524

Lo, C.P., & 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*, 2687-2700

Lu, D.S., Li, G.Y., Moran, E., Batistella, M., & Freitas, C.C. (2011). Mapping impervious surfaces with the integrated use of Landsat Thematic Mapper and radar data: A case study in an urban-rural landscape in the Brazilian Amazon. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*, 798-808

Mahiny, A.S., & Turner, B.J. (2007). A comparison of four common atmospheric correction methods. *Photogrammetric Engineering & Remote Sensing, 73*, 361-368

Meentemeyer, R.K., Tang, W.W., Dorning, M.A., Vogler, J.B., Cunniffe, N.J., & Shoemaker, D.A. (2013). FUTURES: Multilevel Simulations of Emerging Urban-Rural Landscape Structure Using a Stochastic Patch-Growing Algorithm. *Annals of the Association of American Geographers, 103,* 785-807

Meng, X.L., Wang, L., Silvan-Cardenas, J.L., & Currit, N. (2009). A multi-directional ground filtering algorithm for airborne LIDAR. *ISPRS Journal of Photogrammetry and Remote Sensing, 64,* 117-124

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

Popescu, S.C., Wynne, R.H., & Nelson, R.F. (2002). Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. *Computers and Electronics in Agriculture, 37,* 71-95

Raber, G.T., Jensen, J.R., Hodgson, M.E., Tullis, J.A., Davis, B.A., & Berglund, J. (2007). Impact of lidar nominal post-spacing on DEM accuracy and flood zone delineation. *Photogrammetric Engineering & Remote Sensing, 73,* 793-804

Raptis, V.S., Vaughan, R.A., & Wright, G.G. (2003). The effect of scaling on land cover classification from satellite data. *Computers & Geosciences, 29,* 705-714

Ridd, M.K. (1995). Exploring a V-I-S (Vegetation-Impervious Surface-Soil) Model for Urban Ecosystem Analysis through Remote-Sensing - Comparative Anatomy for Cities. *International Journal of Remote Sensing, 16,* 2165-2185

Rogan, J., & Chen, D.M. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning, 61,* 301-325

Rogan, J., Miller, J., Stow, D., Franklin, J., Levien, L., & Fischer, C. (2003). Land-cover change monitoring with classification trees using Landsat TM and ancillary data. *Photogrammetric Engineering & Remote Sensing, 69,* 793-804

Salah, M., Trinder, J.C., & Shaker, A. (2011). Performance evaluation of classification trees for building detection from aerial images and LiDAR data: a comparison of classification trees models. *International Journal of Remote Sensing, 32,* 5757-5783

Sander, H., Polasky, S., & Haight, R.G. (2010). The value of urban tree cover: A hedonic property price model in Ramsey and Dakota Counties, Minnesota, USA. *Ecological Economics, 69*, 1646-1656

Singh, K.K., Vogler, J.B., Meng, Q., & Meentemeyer, R.K. (2010). Mapping Land Use Patterns in an Urbanizing Landscape Using LiDAR Intensity Data. In, *Optical Remote Sensing of the Environment* (p. OMC2). Tucson, AZ: Optical Society of America

Tooke, T.R., Coops, N.C., Goodwin, N.R., & Voogt, J.A. (2009). Extracting urban vegetation characteristics using spectral mixture analysis and decision tree classifications. *Remote Sensing of Environment, 113*, 398-407

Tso, B., & Mather, P.M. (2009). *Classification methods for remotely sensed data*. (2nd ed.). Boca Raton, Florida: CRC Press

Turner, M.G. (2010). Disturbance and landscape dynamics in a changing world. *Ecology, 91*, 2833-2849

White, M.D., & Greer, K.A. (2006). The effects of watershed urbanization on the stream hydrology and riparian vegetation of Los Penasquitos Creek, California. *Landscape and Urban Planning, 74*, 125-138

Wickham, J.D., Stehman, S.V., Fry, J.A., Smith, J.H., & Homer, C.G. (2010). Thematic accuracy of the NLCD 2001 land cover for the conterminous United States. *Remote Sensing of Environment, 114*, 1286-1296

Wu, C.S., & Murray, A.T. (2003). Estimating impervious surface distribution by spectral mixture analysis. *Remote Sensing of Environment, 84*, 493-505

Yaokui, C., Kaiguang, Z., Wenjie, F., & Xiru, X. (2010). Using airborne lidar to retrieve crop structural parameters. In, *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International* (pp. 2107-2110). Honolulu, Hawaii

Yoon, J.S., Shin, J.I., & Lee, K.S. (2008). Land Cover Characteristics of Airborne LiDAR Intensity Data: A Case Study. *IEEE Geoscience and Remote Sensing Letters, 5*, 801-805

Zhang, J., & Foody, G.M. (2001). Fully-fuzzy supervised classification of sub-urban land cover from remotely sensed imagery: statistical and artificial neural network approaches. *International Journal of Remote Sensing, 22*, 615-628

Zimble, D.A., Evans, D.L., Carlson, G.C., Parker, R.C., Grado, S.C., & Gerard, P.D. (2003). Characterizing vertical forest structure using small-footprint airborne LiDAR. *Remote Sensing of Environment, 87,* 171-182

Table 1. Modified Anderson Level I classification schema.

Anderson Level I	Modified Anderson Level I	Description
1. Urban or Built-up Land	1. Impervious surfaces	Intensive use areas characterized by impervious, human-made structures and materials
	2. Managed clearings	Intensively managed green areas within urban environments, such as lawns, golf courses and ranges, urban parks, and electrical utility corridors
2. Agricultural Land	3. Farmland	Croplands and pastures
3. Rangeland †		
4. Forest Land	4. Forest	Deciduous, evergreen, mixed, and managed forests
5. Water	5. Water	Streams, canals, lakes and reservoirs
6. Wetland †		
7. Barren Land	6. Barren Land	Areas of exposed soil, sand or rock
8. Tundra †		
9. Perennial Snow or Ice†		

† Not present in study system.

Table 2. Classification training and validation data.

Class	Training		Validation points
	ML[a] (polygons)	CT[b] (points)	
Impervious surfaces	19	190	71
Managed clearings	29	290	74
Farmland	45	450	83
Forest	21	210	98
Water	14	140	61
Barren land	27	270	59
	155	1550	446

[a] Maximum likelihood
[b] Classification tree

Table 3. Comparison of total accuracies from ML and CT classifications of 30 m TM data and multiple resolutions of LiDAR and Fusion data combinations.

Data	TM		LiDAR		LiDAR-TM Fusion					
			CHM + nDSM + IS		CHM + nDSM + IS+ TM		LiDAR Structural (CHM + nDSM) + TM		LiDAR Intensity IS + TM	
Resolution	Classifier and Total Accuracy (%)									
(meters)	ML	CT	ML	CT	ML	CT	ML	CT	ML	CT
1			54.8	63.7	87.2†	84.8	85.0‡	81.8	81.4	79.1
5			54.0	63.7	86.3†	82.3	84.1	80.9	80.5	78.5
10			53.4	62.8	85.4†	80.7	82.7	78.7	80.5	77.6
15			52.7	61.2	83.0	79.1	81.6	77.8	80.0	76.7
30	79.2	78.5	52.9	60.5	81.2	78.5	81.6	77.3	79.6	76.5

ML = maximum likelihood; CT = classification tree; CHM = canopy height model; nDSM = normalized digital surface model; IS = intensity surface; TM = Landsat Thematic Mapper

† Exceeds accepted minimum standard for total mapping accuracy (Anderson 1976; Rogan et al. 2003)

‡ Meets accepted minimum standards for total mapping accuracy.

Table 4. Comparison of class-level accuracies from ML classifications of TM data and top-performing† LiDAR-TM fusion data using all LiDAR surface models.

	Producer's (%)			User's (%)			Overall (%)		
	TM	Fusion (1 m)	Net Δ	TM	Fusion (1 m)	Net Δ	TM	Fusion (1 m)	Net Δ
Impervious surfaces	94	85	-9	80	92	12	68	77	9
Managed clearings	76	75	-1	57	74	17	48	57	9
Farmland	49	88	39	60	79	19	37	71	34
Forest	92	93	1	80	99	19	90	92	2
Water	93	100	7	99	100	1	93	100	7
Barren land	71	86	15	89	81	-8	66	72	6
	TM	Fusion (5 m)	Net Δ	TM	Fusion (5 m)	Net Δ	TM	Fusion (5 m)	Net Δ
Impervious surfaces	94	79	-15	80	92	12	68	72	4
Managed clearings	76	75	-1	57	69	12	48	54	6
Farmland	49	88	39	60	79	19	37	72	35
Forest	92	93	1	80	99	19	90	92	2
Water	93	100	7	100	100	0	93	100	7
Barren land	71	86	15	89	81	-8	66	72	6
	TM	Fusion (10 m)	Net Δ	TM	Fusion (10 m)	Net Δ	TM	Fusion (10 m)	Net Δ
Impervious surfaces	94	78	-16	80	92	12	68	72	4
Managed clearings	76	69	-7	57	70	13	48	53	5
Farmland	49	87	38	60	77	17	37	69	32
Forest	92	93	1	80	99	19	90	92	2
Water	93	100	7	100	100	0	93	100	7
Barren land	71	86	15	89	77	-12	66	69	3

ML = maximum likelihood; TM = Landsat Thematic Mapper

† Exceeds accepted minimum standard for total mapping accuracy (Anderson 1976; Rogan et al. 2003)

Table 5. Classification matrices and total accuracies for key comparative classifications.

			Actual class[†]							
			1	2	3	4	5	6	Sum	Total accuracy
LiDAR-TM Fusion CHM + nDSM + IS+ TM	1 m, ML	Predicted class[†] 1	60	2	1	0	0	2	65	
		2	2	53	6	7	0	4	72	
		3	0	18	73	0	0	2	93	
		4	0	0	1	91	0	0	92	87.2%
		5	0	0	0	0	61	0	61	
		6	9	1	2	0	0	51	63	
		Sum	71	74	83	98	61	59	446	
	5 m, ML	Predicted class[†] 1	56	3	0	0	0	2	61	
		2	6	53	7	7	0	4	77	
		3	0	17	73	0	0	2	92	
		4	0	0	1	91	0	0	92	86.3%
		5	0	0	0	0	61	0	61	
		6	9	1	2	0	0	51	63	
		Sum	71	74	98	98	61	59	446	
	10 m, ML	Predicted class[†] 1	55	5	0	0	0	0	60	
		2	3	51	8	6	0	5	73	
		3	1	17	72	1	0	3	94	
		4	0	0	1	91	0	0	92	85.4%
		5	0	0	0	0	61	0	61	
		6	12	1	2	0	0	51	66	
		Sum	71	74	83	98	61	59	446	
CHM + nDSM + TM	1 m, ML	Predicted class[†] 1	68	1	0	0	5	0	74	
		2	2	48	8	4	0	12	74	
		3	1	23	72	0	0	6	102	
		4	0	0	1	94	0	0	95	85.0%
		5	0	1	0	0	56	0	57	
		6	0	1	2	0	0	41	44	
		Sum	71	74	83	98	61	59	446	
TM IS + TM	1 m, ML	Predicted class[†] 1	67	1	0	0	4	12	84	
		2	1	51	22	1	0	0	75	
		3	2	20	57	7	0	5	91	
		4	0	0	1	90	0	1	92	81.4%
		5	0	1	0	0	57	0	58	
		6	1	1	3	0	0	41	46	
		Sum	71	74	83	98	61	59	446	
	30 m, ML	Predicted class[†] 1	67	2	0	0	4	11	84	
		2	1	56	38	3	0	0	98	
		3	2	15	41	5	0	5	68	
		4	0	0	1	90	0	1	92	79.2%
		5	0	0	0	0	57	0	57	
		6	1	1	3	0	0	42	47	
		Sum	71	74	83	98	61	59	446	

Table 5. (continued).

ML = maximum likelihood; CHM = canopy height model; nDSM = normalized digital surface model; IS = intensity surface; TM = Landsat Thematic Mapper
† 1 - Impervious surfaces; 2 - Managed clearings; 3 - Farmland; 4 - Forest; 5 - Water; 6 - Barren land

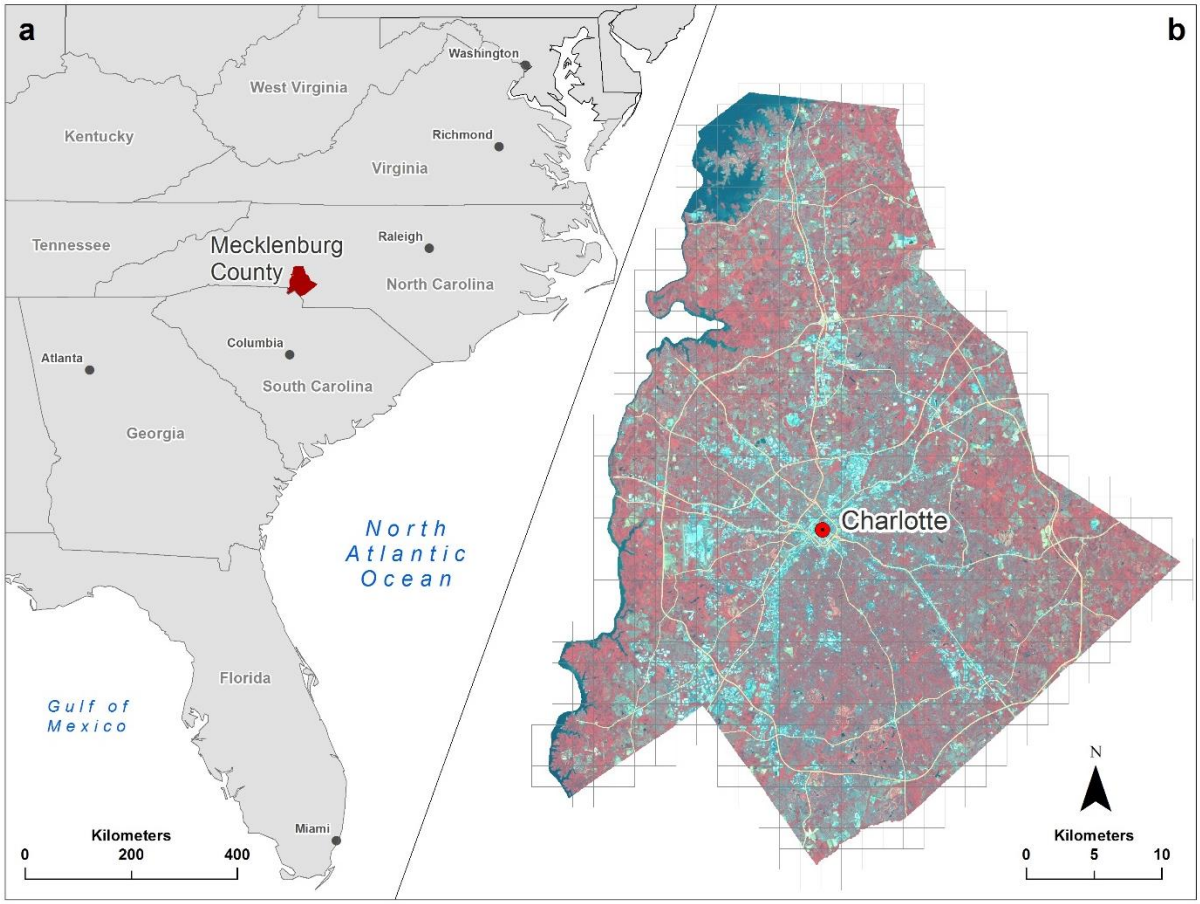


Fig. 1. Study system. a) Mecklenburg County and Charlotte Metropolitan Area, North Carolina, USA, and b) Landsat Thematic Mapper imagery (false color infrared display) acquired May 2007 with LiDAR tile index overlay. A total of 691 LiDAR tiles of dimension 1524 m x 1524 m used in developing surface models.

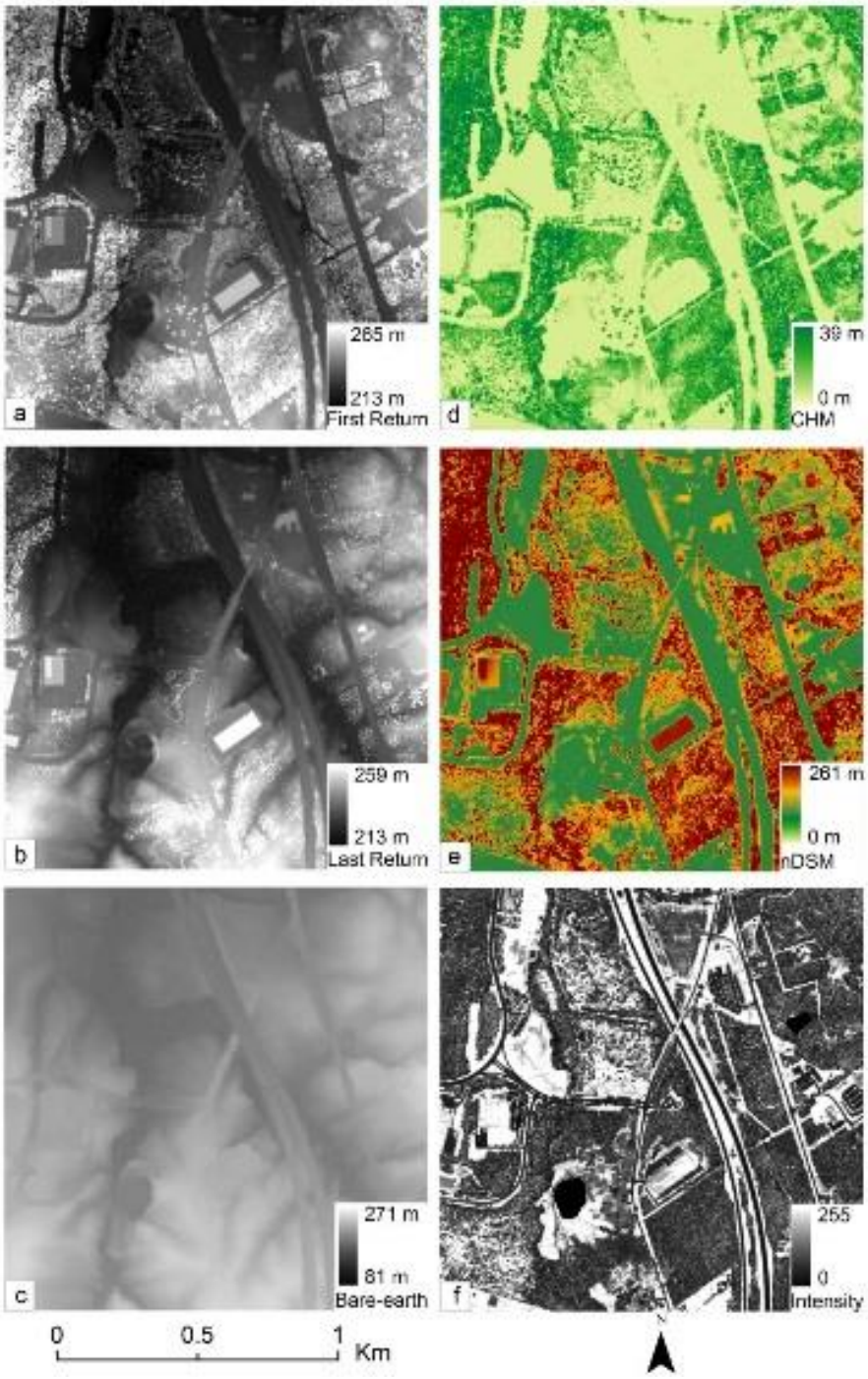


Fig. 2. Rasterized LiDAR data and 1 m resolution surface models. a) first return, b) last return, and c) bare-earth digital terrain model (DTM) used to develop surface models: d) canopy height model (CHM) and e) normalized digital surface model (nDSM). LiDAR point intensity converted to: f) intensity surface model.

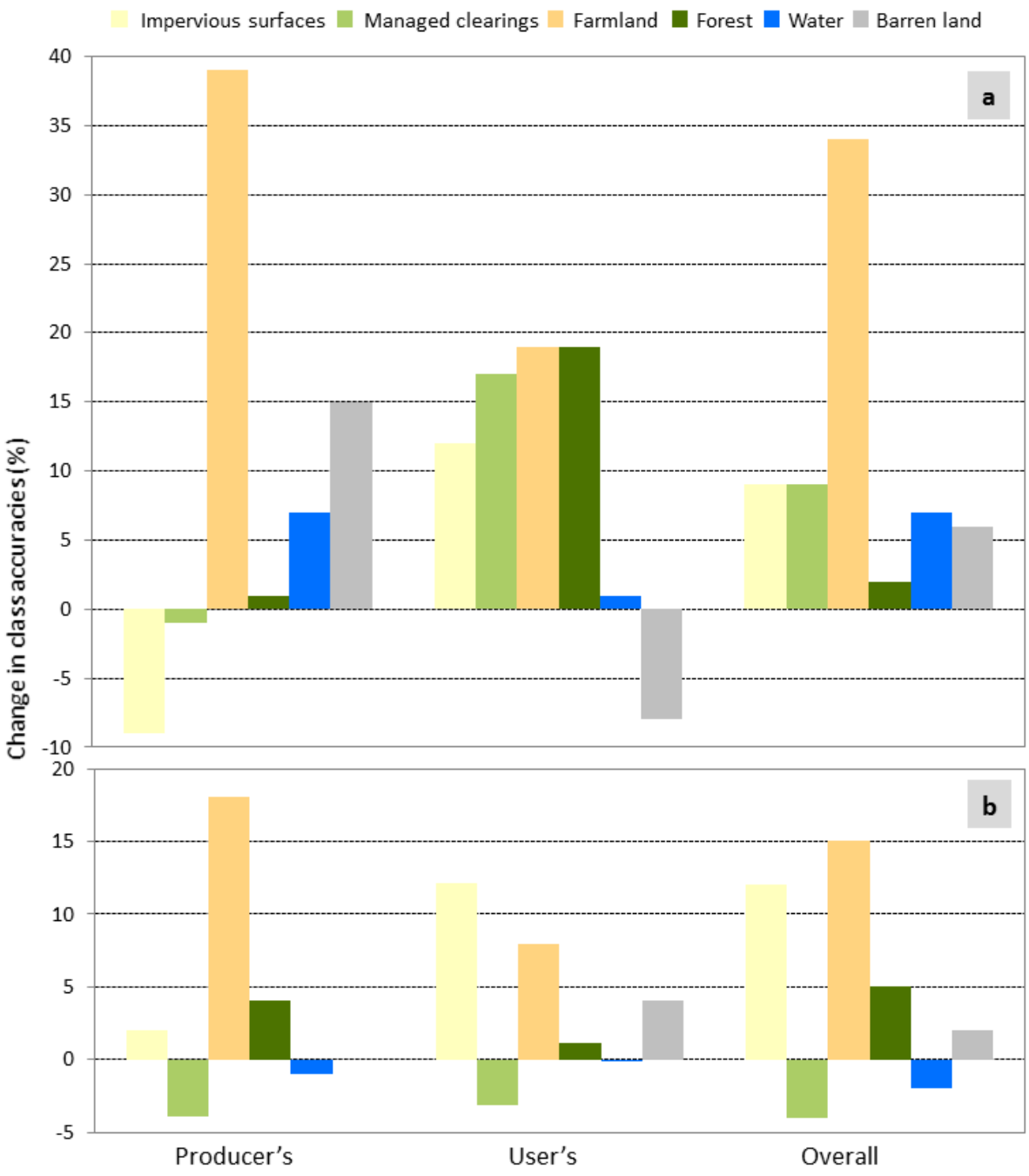


Fig. 3. Net change (gain or loss) in ML class-level accuracies realized using: a) 1 m LiDAR-TM fusion data with all LiDAR surface models vs. TM data alone, and b) 1 m LiDAR structural-TM fusion data vs. LiDAR intensity-TM fusion data. Missing bars indicate no change.

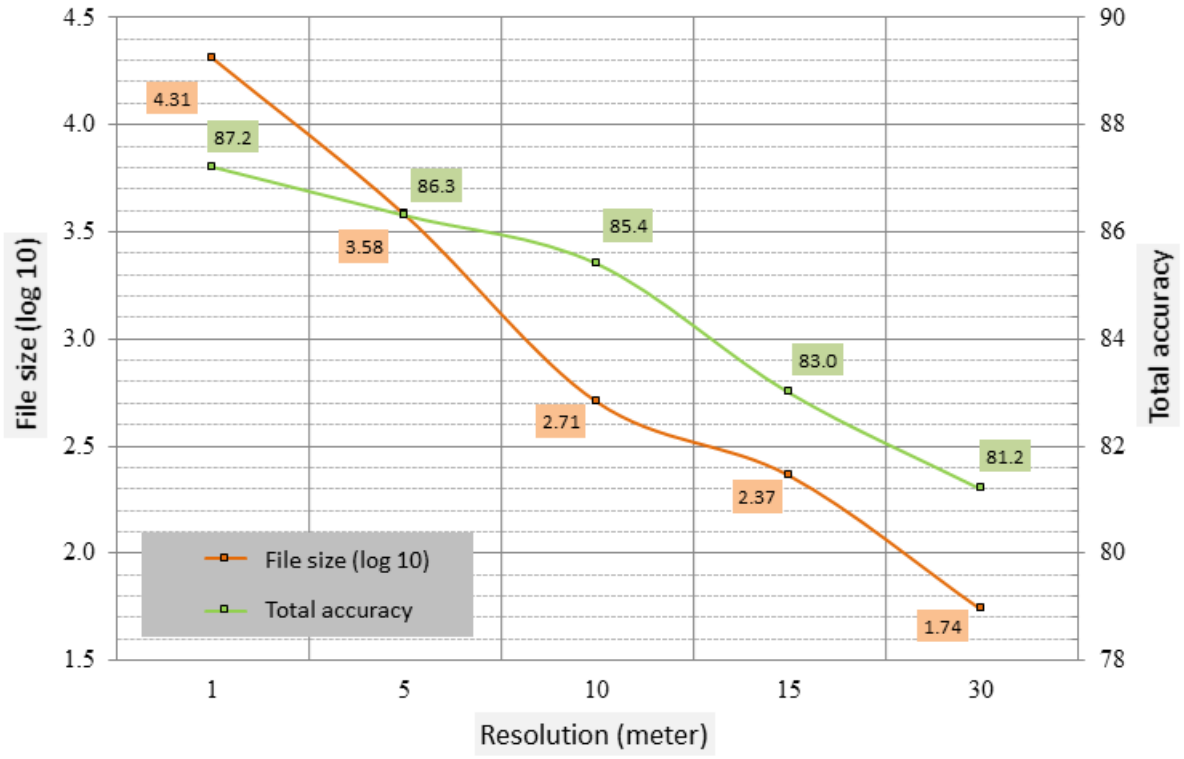


Fig. 4. Change in total accuracy and overall data volume (log base 10 of GB) across multiple resolutions of LiDAR-TM fusion data using all LiDAR surface models.

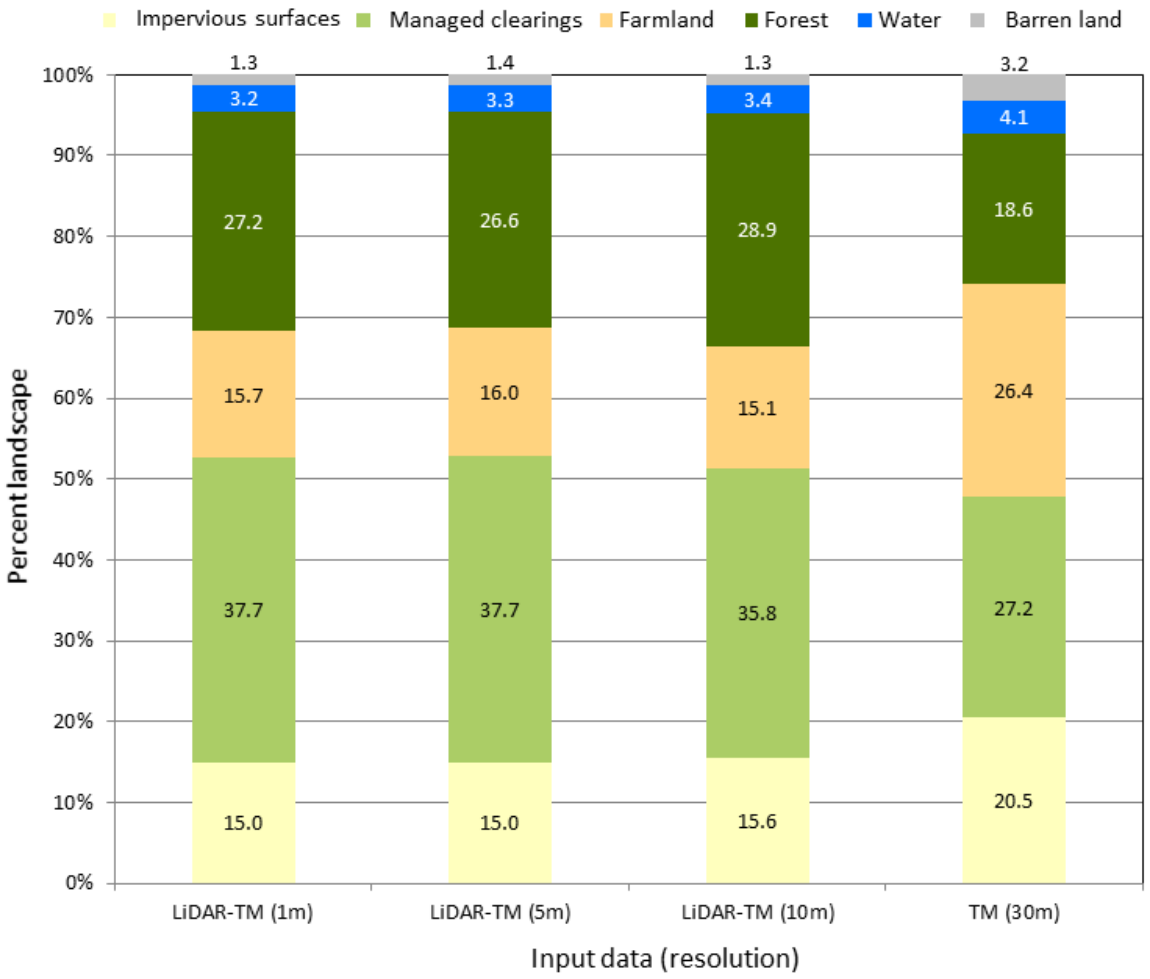


Fig. 5. Proportions of mapped land covers derived from top-performing LiDAR-TM fusion data (exceeding map accuracy standards) at 1 m, 5 m, and 10 m resolutions vs. proportions mapped from the 30 m TM data.

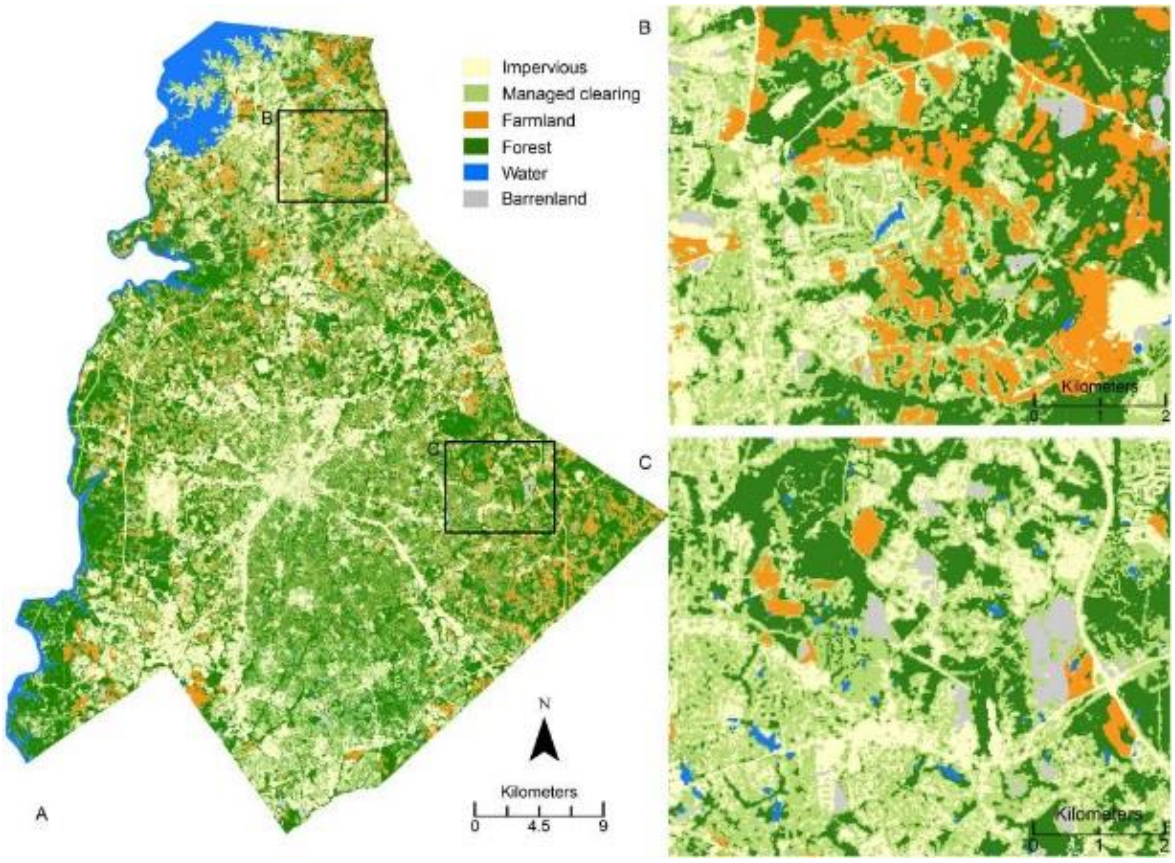


Fig. 6. A) Overall and B-C) detail views of land cover derived from ML classification of 1 m LiDAR-TM fusion data (using all LiDAR surface models), highlighting important distinction between managed clearings, farmland, and forest.

CHAPTER 2. Detecting Understory Plant Invasion

Detecting understory plant invasion in urban forests using LiDAR

Kunwar K. Singh[1,2]*, Amy J. Davis[3], and Ross K. Meentemeyer[1,2]

[1]Center for Geospatial Analytics
[2]Department of Forestry and Environmental Resources
North Carolina State University
3120 Jordan Hall
Raleigh, NC 27695
USA

[3]Department of Geography and Earth Sciences
University of North Carolina
9201 University City Blvd
Charlotte, NC 28223
USA

*Corresponding author: kksingh@ncsu.edu; phone: 01-704-359-7139; fax: 01-919- 515-3439
ajstephe@uncc.edu
rkmeente@ncsu.edu

Abstract

Light Detection and Ranging (LiDAR) data are increasingly used to measure structural characteristics of urban forests but rarely used to detect the growing problem of exotic understory plant invaders. We explored the merits of using LiDAR-derived metrics alone and through integration with spectral data to detect the spatial distribution of the exotic understory plant *Ligustrum sinense* (commonly known as Chinese privet), a rapidly spreading invader in the urbanizing region of Charlotte, North Carolina, USA. We uniquely analyzed regional-scale *L. sinense* field data collected over the course of three years with the combinations of datasets developed from categorized LiDAR-derived variables (*Overstory*, *Understory*, *Topography*, and *Overall Vegetation Characteristics*) and IKONOS spectral features (*Optical*) using Random Forest (RF) and logistic regression (LR) classifiers at the experimental extent to assess the relative contributions of sensors and forest landscape structures in detection of *L. sinense*. We compared top performing models developed for the experimental extent using RF and LR classifiers and used the best overall model to produce predictive map of the spatial distribution of *L. sinense* across the urbanizing forest landscapes of the county. RF classification of LiDAR-derived *Topography* metrics produced the highest mapping accuracy estimates, outperforming IKONOS data by 17.5% and the integration of LiDAR and IKONOS data by 5.3%. The top-performing model from the RF classifier produced the highest kappa of 64.8%, improving on the parsimonious LR model kappa by 31.1% with a moderate gain of 6.2% over the county extent model. Our results demonstrate the superiority of LiDAR-derived metrics over spectral data and fusion of LiDAR and spectral data for accurately mapping the spatial distribution of the forest understory invader *L. sinense*.

Keywords: Biological invasion, Chinese privet, data integration, forest health, invasive species, IKONOS, LiDAR, *Ligustrum sinense*, random forest, urban ecology

1. Introduction

Biological invasions are one of the leading contributors to changes in regional and global biodiversity (Asner and Vitousek 2005; Dantonio and Vitousek 1992; Simberloff et al. 2013). The rapid spread of invaders across a range of ecosystems makes it challenging to accurately assess the spatial distribution and abundance in a timely and cost-effective manner at a desired spatial-temporal scale (Huang and Asner 2009). Remote sensing has facilitated advances in mapping and monitoring biological invasions across a range of ecosystems (Ustin et al. 2004), including deciduous forests (Resasco et al. 2007), wetlands (Laba et al. 2008), rangelands (Everitt et al. 1995; Glenn et al. 2005; Noujdina and Ustin 2009), coastal environments (Hestir et al. 2008; Underwood et al. 2003), and tropical forests (Asner et al. 2008). These successes are primarily attributable to sensor characteristics, recent advancements in mapping algorithms, and synergistic use of datasets from multiple instruments. While some success has been achieved with high-to-moderate resolution imagery, hyperspectral imagery has demonstrated the promising results (Huang and Asner 2009). The growing number of invasions in forests pushes the limits of remote sensing performance, particularly in detecting exotic plant invasions in the understory where regeneration and recruitment are most impacted by invaders (Becker et al. 2013). Canopy cover closure, canopy gap shadowing, mortality of native plant species, and terrain variability in the understory of forest ecosystems limit accurate detection of understory invaders (Asner et al. 2008). Successful detection often depends upon open views of the species of interest (Resasco et al. 2007) and the timing of remote sensing imagery to take advantage of phenological patterns in the overstory, i.e. after leaf abscission (Becker et al. 2013; Tuanmu et al. 2010). To date, our knowledge of understory plant invasions is based on field surveys, but cost and data collection times necessarily limit the number of locations that can be assessed. Although spatially explicit models such as species distribution models (SDMs) have been used to predict the probability of invader presence in non-sampled locations, there has been little success to date in applying these models to new regions outside the study extent. Successful management of broad-scale exotic invaders in the

understory of forest ecosystems requires new techniques using minimal field data for mapping their spatial distribution and abundance.

Several studies have examined the potential of using field-measured SDMs – augmented with optical remote sensing - to detect understory plant invaders (Estes et al. 2010; Rushton et al. 2004). Andrew and Ustin (2009) suggest that the mapping of the distribution of invasives and sites prone to invasion using remote sensing alone can be improved upon by including covariates characterizing the environmental suitability of the site for the invader, such as slope, solar radiation index, and soil moisture index. Other studies have examined the efficacy of moderate-to-high resolution and multi-to-hyperspectral remote sensing data to detect understory invaders based on variability in seasonal and environmental responses (Wilfong et al. 2009). For example, spectral differences in the seasonal phenology of understory vegetation and deciduous tree canopy during leaf-off conditions can be used to detect understory vegetation, but success is limited to narrow windows of senescence time (Chastain and Townsend 2007). The spatial and vertical heterogeneities of forest vegetation limit the utility of spectral-based approaches for detecting the structure and composition of understory vegetation even where distinguishing physiological and phenotypic characteristics of an invader exist (Homolova et al. 2013). Studies have also attempted to improve the detection of understory vegetation by using sub-pixel multivariate algorithms, such as artificial neural networks, linear mixture models, and maximum entropy (Carpenter et al. 1999; Linderman et al. 2004; Tuanmu et al. 2010; Wang et al. 2009). While some studies have shown potential for using spectral remote sensing to detect understory invasives by coupling the phenological patterns and improved algorithms, questions remain regarding the efficient use of structural remote sensing.

Light Detection and Ranging (LiDAR) offers the possibility of accurately detecting understory invasive shrubs on the basis of their structural characteristics as distinguished from canopy species since it provides three dimensional data on the structure and arrangement of vegetation across all forest strata (Dubayah and Drake 2000; Falkowski et al. 2009; Goetz et al. 2010; Lefsky et al. 2002). LiDAR data have been widely used for quantifying structural characteristics (e.g., tree diameter, height, density, and biomass) of

forest stands (Hudak et al. 2008; Jaskierniak et al. 2011; Lefsky et al. 2002; Miura and Jones 2010), mapping snags and understory shrubs (Martinuzzi et al. 2009), characterizing successional stages of forest stands (Falkowski et al. 2009), estimating native understory vegetation cover (Estornell et al. 2011; Seidel et al. 2012; Wing et al. 2012), and identifying tree species (Kim et al. 2009). To complement LiDAR's narrow spectral range (Chen 2007), structural covariates of LiDAR have been combined with multi-to-hyperspectral remote sensing imagery to map forest structural attributes, plant species distribution, forest biomass, and forest canopy fuels (Bork and Su 2007; Erdody and Moskal 2010; Mutlu et al. 2008; Popescu et al. 2004; Vierling et al. 2002). To date, LiDAR technologies 'alone' or through 'fusion' with spectral data have not been leveraged to improve our ability to detect understory plant invaders in urban forest ecosystems at regional scales. In addition, detection information is typically limited to direct LiDAR measurement (e.g. shrub height) of the plant individuals themselves while ignoring other forest and landscape characteristics (e.g. topographic position) that may affect the distribution and growth of invaders.

In this study, we examine the utility of multi-return LiDAR data, IKONOS imagery, and the combination of data from both sensors for detecting and mapping understory invaders. Using *Ligustrum sinense* (commonly known as Chinese privet) as a case study, we focus on the urban to rural forests located in the Charlotte Metropolitan Region (CMR) of North Carolina to analyze the performance of LiDAR data, IKONOS, and the data combination. We incorporate additional forest and landscape characteristics to detect this plant invader by analyzing combinations of datasets developed from categorized LiDAR-derived variables (*Overstory*, *Understory*, *Topography*, and *Overall Vegetation Characteristics*) and IKONOS spectral features (*Optical*) using Random Forest (RF) and logistic regression (LR) classifiers. We then compare the top-performing models and apply the best model to map the probability of *L. sinense* occurrence across the heterogeneous study region.

2. Methods

2.1.Study system

The naturalized *L. sinense* has invaded over one million hectares of forests across a large portion of the southeastern United States, and the species is listed among the top invasives in many states (Miller et al. 2008). It grows in a wide variety of habitats and environmental conditions, including abandoned farmlands and disturbed areas, where it forms impenetrable thickets with a typical height range of 1 to 5 m, occasionally reaching heights of 10 m (Fig. 1) (Langeland and Burks 1998; Wilcox and Beck 2007). It is a prolific seed producer with a high germination rate (Morris et al. 2002). Its ability to tolerate both flooding and shade makes it well suited to the floodplains and riparian zones of forests (Brown and Pezeshki 2000). The abundance of *L. sinense* threatens forest ecosystems by altering the diversity, composition, and structure of forests (Greene and Blossey 2012; Hart and Holmes 2013). The growing threats of large-scale ecosystem modification are matters of extreme concern for land managers and conservationists in the United States (Hanula, Horn, and Taylor 2009; Merriam 2003).

A growing number of studies have examined *L. sinense* invasion and impacts on ecosystem services across much of the southeastern United States. The invasive has threatened to displace endangered Miccosukee gooseberry (*Ribes echinellum*) in Florida (Langeland and Burks 1998). Invasions by *L. sinense* have been linked to several negative outcomes including reduced herbaceous species in understory and suppressed tree regeneration in a mixed hardwood forest in western North Carolina (Merriam and Feil 2002); decreased beetle diversity in northeastern forests of Georgia (Ulyshen et al. 2010); and outcompeted the population of the Schweintz's sunflower (*Helianthus schweinitzii*) found in the central Piedmont region of North and South Carolina (USFWS 1991). Faulkner et al. (1989) found that dense thickets of *L. sinense* in forested areas of Tennessee produce large quantities of litter and act like 'umbrellas' preventing the infiltration of leaf litter from the native tree canopy. *L. sinense* ranked fourth among the top invasive plants in North Carolina with high prevalence in urban areas (EDDMapS 2014).

The highly fragmented and urbanizing forested landscape of Mecklenburg County (1,415 km^2), located in the CMR of North Carolina within the Piedmont province of the southeastern United States (Fig. 2a), provides an ideal system for this case study. The study system is characterized by rolling topography with a forest-farmland mosaic that comprises a mix of secondary growth oak-hickory-pine remnant forests dominated by loblolly pine on clear-cuts and abandoned farmland.

2.2.Field data collection

We established a network of 346 randomly distributed field plots within 75 forested sites in the study system (Fig. 2b). We collected data at 139 field plots during the leaf-on season of 2009 and 2010 from mid-May to the end of July and at 207 field plots during the leaf-off season of 2012 from mid-January to mid-March. Impassability to field plots due to thick understory vegetation in the leaf-on season forced us to collect data in the leaf-off season at the peak of senescence and abscission. Prior to the data collection, we identified a network of field plots in the study system divided into urban, suburban, and rural forested landscapes using upland and lowland characteristics. The use of an urban-rural gradient and geomorphological categories (lowland and upland) helped us account for the favorable habitat (moist habitats, especially abundant along stream banks) of *L. sinense*. This also ensured appropriate representation of forests across urban-rural gradients. At each site, we established a 10 x 10 meter (0.01 ha) plot by defining a center point and recording coordinates using a Garmin GPSMAP 62s GPS device. We used a modified Braun-Blanquet scale to collect presence/absence data following an assessment of abundance by estimating the percentage of the plot area occupied by aerial parts of *L. sinense*. We noted whether *L. sinense* was present or absent, and if present, collected data as percentage cover (up to 20%, 40%, 60%, 80%, and 100%) (Table 1).

2.3.LiDAR data acquisition

We obtained 1,896 tiles of discrete return airborne LiDAR data covering Mecklenburg County from the GIS Mapping and Project Services of Charlotte-Mecklenburg County government. Data was provided in the State Plane Coordinate System (NC FIPS 3200, NAD 1983, m) with each tile having dimensions 914.40 m x 914.40 m (Fig. 2b).. Pictometry International Corp. (Rochester, New York) acquired the data over six missions between April 11-14, 2012 using an Optech ALTM Gemini 3100 LiDAR system. The on-board GPS (Global Positioning System) data from the aircraft and concurrent ground-based station data were post-processed using Applanix POSPac v4.4 to generate a smoothed best estimate trajectory (SBET). Optech's DashMap software package was used to produce LiDAR point cloud data from the raw laser range data combined with SBET, and further processed in TerraMatch software to measure and apply small adjustments to the system's orientation angles in order to ensure proper alignment of data between flight lines. Additionally, TerraScan was used to classify returns as either ground or non-ground points and to ensure the quality of the classification. The sensor recorded four returns plus intensity with an average point spacing of 1 m between any two neighboring points over the study area.

2.3.1. LiDAR data processing

Topography and forest strata play a key role in the distribution and growth of understory plant invasions (Royo and Carson 2006). We developed LiDAR metrics representing overstory and understory forest structure, and topography. We processed each LiDAR tile using the BCAL LiDAR tools, based on IDL programing language (BCAL LiDAR Tools 2013), and produced 5 m resolution rasters for the entire county. Processing LiDAR data at 5 m spatial resolution allowed us to capture the variability within the plot while maintaining computational efficiency during the analysis. To account for forest vertical structures in the analyses, we divided forest into four vertical strata (≤1 m; > 1 m to ≤ 5 m; > 5 m to ≤ 10 m; and > 10 m) following the height range of *L. sinense*. LiDAR data collected in

the leaf-off season vary significantly both in point-density and intensity values for different forest types such that contrast among different forest strata and topography is emphasized. We converted LiDAR intensity into various vegetation and topography intensity metrics (Table 2). We also developed height percentile metrics to represent the proportion of vegetation below a given percentile. For example, 99th percentile defines a measure of the maximum height, whereas the remaining percentiles quantify variation in height across the vegetation profile (Jaskierniak et al. 2011).

We developed indices using a LiDAR-derived digital elevation model (DEM) to characterize moisture content and site conditions. Indices include an integrated moisture index (IMI), an elevation to relief ratio (ERR), a topographic roughness index (TRI), and a solar radiation index (SRI). An IMI with increasing values indicates increasing moisture accumulation and retention (Davies et al. 2010; Iverson et al. 1997). The ERR index measures the extent to which topography has been opened up by erosion (Martinuzzi et al. 2009). The TRI is a measure of variability in the landscape surface and a proxy for disturbances (Stambaugh and Guyette 2008). The SRI affects biological processes that influence species distributions across landscapes (Keating et al. 2007). We used a three-band composite dataset containing a canopy height model, a normalized digital surface model, and a differenced surface model of these two to classify forested landscapes into deciduous and evergreen types (Singh et al. 2010). We created a total of sixty-six variables from LiDAR, consisting of 43 canopy and 23 topographic metrics at the county extent (Table 2). We further post-processed these metrics using a mask of vegetation cover to exclude developed land-cover types from forested landscapes.

2.4.IKONOS data acquisition

We acquired a cloud-free multispectral IKONOS image from the leaf-off season dated February 21, 2010, and covering a portion of Mecklenburg County (Fig. 2c) (hereafter referred to as “experimental extent”). The image consists of three pan-sharpened visible bands (blue, 0.45-0.52 µm; green, 0.51-0.60 µm; red, 0.63-0.70 µm) and one near infrared

band (nir, 0.76-0.85 μm) at 1 m spatial resolution. We projected the four-band composite image into the State Plane Coordinate System (NC FIPS 3200, NAD 1983, m). The selection of leaf-off IKONOS images facilitated spectral discrimination between *L. sinense* and neighboring deciduous trees. We applied radiometric correction to ground reflectance by applying a simple dark pixel subtraction method. We co-registered calibrated IKONOS data to the LiDAR intensity surface model using landmark control points (Jones et al. 2010).

2.4.1. IKONOS data processing

We developed fourteen spectral features, including a pan-sharpened, four-band image, a three-band hue-intensity-saturation (HIS) image, a four-band texture image, and three vegetation indices: (1) enhanced vegetation index (EVI), (2) modified soil adjusted vegetation index (MSAVI), and (3) the atmospherically-resistant vegetation index (ARVI). Studies have demonstrated the efficacy of pan-sharpened imagery in the classification of forest tree species in urban areas, estimation of vegetation cover, detection of disturbance severity and canopy heterogeneity, and for understanding the spatiotemporal distributions of forest structural and dynamic variables in the tropical forest (Carleer and Wolff 2004; Malhi and Roman-Cuesta 2008; Pu 2011; Rich et al. 2010). We transformed the pan-sharpened, four-band image to HIS and a texture image as texture images have been used in retrieving forest structure variables and distinguishing subtle differences in vegetation types (Huete et al. 1997; Van Delm and Gulinck 2011). The EVI is suitable for detecting changes in areas of high biomass. The MSAVI distinguishes areas with a high degree of exposed soil surface. The ARVI, being less sensitive to atmospheric effects (aerosols), is useful for measuring and monitoring plant growth, vegetation cover, and biomass production. Finally, we resampled these features to 5 m resolution to match spatial grain of the LiDAR-derived variables and facilitate comparisons.

2.5.Datasets and modeling schema

To understand the contribution of forest structure in detection and mapping of *L. sinense*, we categorized the LiDAR-derived variables into *Overstory*, *Understory*, *Topography*, and *Overall Vegetation Characteristics* variable types and assigned IKONOS spectral features to the *Optical* types (Table 2). We used the description of variables and their established applications in forestry-related studies for the purpose of categorization (Asner et al. 2008; Donoghue et al. 2007; Falkowski et al. 2009; Hudak et al. 2008; Jaskierniak et al. 2011; Martinuzzi et al. 2009; Singh et al. 2012; Yu et al. 2011). Using the five 'variable types', we developed thirty-one unique combinations of models at the IKONOS experimental extent (Fig. 2c; Table 3), and then further applied the RF (Breiman 2001) classifier to analyze relative contributions of sensors and forest landscape structures in detection *L. sinense*. Finally, we compared the top performing model developed from selected predictor variables (PVs) using the RF classifier to the best parametric model developed from LR (Orka et al. 2012), and used the best overall model to map the probability of *L. sinense* occurrence across the heterogeneous county extent.

2.6.Statistical analysis

We implemented the RF and LR classifiers in the IDL (Waske et al. 2012) and R programming environment (R Core Team 2013), respectively. Prior to analysis and to account for the time lag between field data and remotely sensed data, we performed visual assessment of the field data collected prior 2012 using LiDAR canopy cover to eliminate those field plots affected by forest loss.

Traditional statistical methods do not address nonlinear complex interactions among high-dimensional remote sensing variables (Cutler et al. 2007). To overcome these limitations, we used a non-parametric RF classifier, a decision-tree-based ensemble classifier, which has proven effective in the classification of high-dimensional remote sensing and ecological data (Cutler et al. 2007; Goetz et al. 2010; Martinuzzi et al. 2009). We selected the RF classifier based on: (1) its robustness and accurate performance on complex

datasets with an array of input variables (higher dimensionality) (Guo et al. 2011; Stumpf and Kerle 2011), (2) its ability to determine significant variables in the classification, and (3) a waiver from cross-validation as it generates an internal unbiased estimate of test error (Cutler et al. 2007). The RF classifier provides estimates of variable importance in the classification to assist with reducing data dimensionality. The classifier operates by growing a large number of individual decision trees from randomized subsets of training samples to maximum size without pruning, and then selecting only the best split among a random subset at each node (Falkowski et al. 2009; Prasad et al. 2006). The optimal classification is then determined by selecting the most common classification results at each node within the group of multiple decision trees (Breiman 2001). Due to the randomness, the bias of RF usually increases, but its variance also decreases due to averaging, therefore constructing the most parsimonious classification model that retains the highest classification accuracy (Cutler et al. 2007; Goetz et al. 2010). For comparison, and due to the binary nature of the output, we also applied the LR classifier to plot level datasets. This approach offers a robust discriminant function without assuming a specific distribution in the input variables.

We applied the LR classifier to analyze the presence of *L. sinense*, where field data served as response variables and remote sensing covariates as predictor variables (PVs). Our selection of LR is based on its flexibility related to key statistical assumptions (linearity, normality, and homoscedasticity), and the straightforward interpretation of results based on odds ratios and statistical hypothesis testing. We modeled the logit of the presence of *L. sinense* as:

$$y = \log_e [P / (1 - P)] = logit(P) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

where y is a linear combination function of included predictor variables $x_1, \ldots x_k$, the parameters $\beta_0 + \beta_{1,} \ldots + \beta_k$ are the regression coefficients to be estimated, *P* is the probability of the occurrence of *L. sinense*, and *logit(P)* is the odds ratio.

We used the variance inflation factor (VIF) ≤ 10 for the identification and removal of collinear PVs, and we applied the backward elimination technique by deleting one predictor

variable for each iteration using the Akaike's Information Criterion (AIC) until no further improvement in the model was observed (Orka et al. 2012). We used a maximized sum of specificity and sensitivity criterion (MST – maximized sum threshold) for selecting the threshold of occurrence for *L. sinense* detection and the Hosmer and Lemeshow goodness-of-fit test to evaluate the final model (Donoghue et al. 2007). Once an optimal LR was established, we applied it to the entire dataset to produce a categorical map of presence and absence of *L. sinense*.

2.7.Evaluation of model performance

We evaluated model performance using five accuracy measures: F-1 measure, sensitivity, specificity, kappa, and the area under the receiver operating characteristic curve (ROC). For the robustness of the results, we computed resubstitution and 10-fold cross-validated (k-fold CV) estimates for all models from RF and LR classifiers. The area under the ROC curve ranges from 0 to 1 (none correctly classified to all correctly classified) and provides a quantitative measure of model performance. We obtained the ROC by plotting the true positive proportion of correctly predicted presence (sensitivity) on the Y-axis against the false positive proportion of correctly predicted absence (specificity) on the X-axis. Since the RF classifier provides a reliable error estimate using data that is randomly withheld from each iteration of tree development (the "out-of-bag" portion), an accuracy assessment using independent testing data is not required (Lawrence et al. 2006). However, for purpose of model comparison and analytical robustness, we computed a k-fold CV by dividing the field data into ten equal sets, where nine sets were input for training and one set for testing, and repeated this procedure ten times (Cutler et al. 2007).

3. Results

3.1.Understory detection and mapping performance

The LiDAR-derived PVs produced the highest accuracy estimates, outperforming both IKONOS alone and the integration of LiDAR and IKONOS data. The RF classifier,

with selected PVs (10 variables; Figure 4a and 4b) at experimental and county extents, performed substantially better with higher accuracy estimates compared to the LR classifier, while the LR classifier showed less variability between resubstitution and cross-validated accuracy estimates. The RF model at the experimental extent produced the highest resubstitution kappa of 64.8%, improving on the LR model by 31.1% with a moderate gain of 6.2% compared to the county extent (Table 3 and 4). The k-fold CV revealed the robustness of the RF classifier by producing a similar kappa (1% difference) estimate at both extents. In addition, the kappa estimate was 21% and 40% higher as compared to the model from the LR classifier and models created by the combination of variable type using the RF classifier, respectively (Table 4). The F-1 estimate between resubstitution and k-fold CV varies on average 7.84% using the RF classifier compared to 1.2% using LR. The RF classifier predicted nearly three times the area of *L. sinense* spread in the understory of forests at the experimental extent compared to the LR classifier (Figure 6a). The ratio of *L. sinense* to forest cover is nearly consistent across urban-rural gradients. However, the ratio of *L. sinense* to developed land cover in rural areas is 71% higher than in suburban areas, and 140% higher than in urban areas (Figure 6b). The estimated sensitivities across the models, created by the combination of variable type, are relatively higher ≥ 0.72 than specificities, which differ significantly except at the county extent where sensitivity is lower than specificity. However, with higher positive and negative productivity (ratio between true positive/negative and total positive/negative), the RF classifier performed substantially better at the county extent than RF at the experimental extent in detecting invasion.

3.2.Relative contribution of IKONOS imagery and LiDAR data

The IKONOS-derived PVs alone or when analyzed together with the LiDAR PVs did not improve classification accuracy estimates of detection and mapping *L. sinense*. The F-1 and kappa estimates obtained from RF classification of *Optical* PVs were only 1.25% and 1.9% higher, respectively, than the *Understory* model, and contributed only marginally to improvements in overall accuracy estimates when combined with *Overall Vegetation Characteristics* and *Understory* PVs (Table 3). Among the *Optical* PVs, texture contributed

slightly to improved detection when combined with *Overall Vegetation Characteristics* while no other *Optical* PVs were significant in the top performing RF and LR analysis.

3.3.Contributions of forest landscape structures

The *Topography* PVs alone produced the highest accuracy estimates and contributed more to detection of *L. sinense* than other variable types across all models created by the combination of variable type using the RF classifier at the experimental extent. The F-1 estimate was 11.4% higher than average estimates obtained from other variable types alone, followed by *Overstory* and *Overall Vegetation Characteristics* (Table 3). Gains in accuracies did not improve with increasing numbers of variable types except for a slight improvement of ≥ 0.2 kappa with the *Topography* PVs. However, the RF variable importance plot of the M_{ALL} model (Figure 4a) produced higher number of PVs from *Overall Vegetation Characteristics* contributing in the overall accuracy estimates followed by the *Topography* and *Overstory* variable type. The LR analysis also identified similar sets of significant variable types with the exception of *Understory* ($V_{STRATUM-1}$) (Table 5). Two PVs – TMI and VI_{SD} – are significant in both the top-performing models from RF classifier and model from the LR classifier. LiDAR intensity-derived forest landscape variables (VI_{MEAN} / VI_{MAX} and VI_{MIN}) contributed in both the top performing RF and LR models. However, the lower performance suggests that the LR classifier was unable to address adequately some nonlinear structure for matching or higher accuracy estimates. The LR classifier revealed the magnitude and direction of PVs' contribution in the detection of *L. sinense*. Examination of the RF variable importance plots showed that in addition to the TMI, three PVs (topographic roughness index, vegetation density and forest cover types) contributed considerably to improved detection with higher accuracy estimates at both experimental and county extents (Figure 4b).

3.4.Large-area assessment of *Ligustrum sinense*

The higher performance of LiDAR PVs using the RF classifier encouraged us to evaluate county level field data for improved detection of *L. sinense*. Our analysis produced a similar k-fold CV kappa estimate with only slight variation in resubstitution accuracy estimates. Use of county level field data increased positive, and negative predictive ratio by 33% as compared to the ratio at the experimental extent accompanied by a 13.4% difference between sensitivity and specificity, which suggests an improved prediction. The variable importance plot indicated TMI as the most significant PV for large-area assessment of *L. sinense*, followed by VC_{TYPE}, TRI, and vegetation density (Figure 4b). The analysis of LiDAR PVs and the RF classifier with the county level field data mapped the spatial distribution of *L. sinense* that occupies 5.95% of forest understory, predominantly in drainage ways and lowland areas of forested landscapes along urban-rural gradients in the county (Figure 7).

4. Discussion

Regional assessment of understory invasions requires an approach that overcomes complicacies of spatial and vertical heterogeneity of forested landscapes, and utilizes structural and spectral remote sensing data in a coherent manner. In this study, we found that structural and spectral data from LiDAR can be used to detect *L. sinense* over large areas with greater accuracy than using multispectral IKONOS data alone or the combination of predictors derived from LiDAR and IKONOS. We obtained high accuracy estimates using the RF classifier with selected LiDAR PVs at both experimental and the county extents (Table 4). Topographic attributes, including TMI and TRI, together with vegetation height and LiDAR's spectral metrics contributed the most in the detection of *L. sinense*, regardless of the classifier used.

Contrary to our expectations, high resolution IKONOS did not contribute to the detection of *L. sinense*, revealing the limited ability of spectral remote sensing to unravel vertical structure of heterotypic forest stands. The ineffectiveness of IKONOS in the

detection of *L. sinense* could be due to cumulative or individual constraints in spectral, spatial and temporal resolutions. The heterogeneity of forest strata is responsible for the mixed spectral response resulting in scale-dependent variance. This could be further complicated with an off-nadir effect of tall trees on understory vegetation as found in (Hsieh et al. 2001) and (Myeong et al. 2006). Decreasing the spatial resolution of IKONOS spectral features to 5 m for spatial consistency across the analysis and the uneven geo-registration between IKONOS and LiDAR may have also contributed to poor results (Nelson et al. 2006). A time lag in acquisition (a temporal gap) may also contribute to problematic integration of two remote sensing data sources, which is also the case in this study. Accurate detection of understory plant invasions is scale (scale of the study) and data dependent, considering the types of data analyzed, the existing data dimensionality and nonlinearity, and the data acquisition period. The choice of classification techniques also influenced computational efficiency and detection success. Moreover, detection success is most dependent upon the characteristics of forested landscapes, including physiology of invasive shrubs and the host environment (Chastain and Townsend 2007; Resasco et al. 2007).

Our approach of dividing PVs from LiDAR into *variable types* representing characteristics of forest structure and forested landscapes added a new element into the design of our analytical framework. The RF accuracy estimates of models based on LiDAR-derived variable type combinations (Table 3) illustrate the significance of forest structure and landscapes characteristics and their contribution in detection and mapping *L. sinense*. The *Topography* variables contributed the most with improved detection once combined with *Overall Vegetation Characteristics*, and *Overstory* using the RF classifier. Among our broad variable types, topographic moisture, roughness, vegetation height, and vegetation intensity profile were key variables in the RF and LR models at both extents (Figure 4a and 4b, Table 4). Topographic moisture and roughness play important roles in the growth and distribution of *L. sinense* in riparian forests of the southeastern United States (Hudson et al. 2013; Wang and Grant 2012), therefore the influence of these variables in the analysis is not surprising. The significance of vegetation height, density and intensity profiles variables supports our concept that physiological characteristics of *L. sinense* are the most suitable for LiDAR-

based detection. The prominence of vegetation density and intensity variables in our analysis also suggests that LiDAR with higher point-density, acquired in the peak of senescence period, can improve detection and mapping by providing distinct spectral contrast and vegetation densities between deciduous and compact evergreen understory, which corroborates the findings of Straatsma and Middelkoop (2007).

Our k-fold CV of model performance helped minimize the over-fitting tendency of the RF classifier and provided a direct comparison between RF and LR models. The RF classifier outperformed the LR classifier using the same set of selected PVs. Cutler et al. (2007) observed higher performance using the RF classifier when there were strong interactions among the PVs. RF's ability to model complex interactions among PVs perhaps addresses the complicated interactions of LiDAR-derived PVs in the top-performing model and explains the better performing RF classifier. The common PVs in both classifiers, for example, TMI and VI_{SD}, and the marginal difference in AUC value suggests that the LR-derived probabilities might be useful for making rapid adjustments to thresholds that alter false positive/false negative rates to improve prediction of *L. sinense*. One disadvantage of LR is that it requires non-collinear variables, whereas RF is not affected by multicollinearity because the algorithm 'spreads' the variable importance across all the variables and keeps those which are good predictors of the response even when correlated with other predictor variables (Cutler et al. 2007). The unbalanced sample size is also one potential reason for the performance disparity between the two classifiers. The RF makes no distributional assumptions about the predictor or response variables and can automatically reanalyze the presence and absence class to achieve equal size, providing another advantage over LR. In general, a higher number of presence data in the model poses a challenge of overestimation. On the contrary, due to a higher number of absence data, more than twice the presence data at both extents, we expect that neither classifier overestimated the spatial distribution of *L. sinense*.

We identified two primary sources of error in the overestimation of 'sensitivities' in models: 1) the presence of other evergreen understory shrubs, and 2) the use of field plots containing a fewer number of seedlings, juveniles, and matures as presence in the classifiers.

However, the effects of higher evergreen shrub coverage and plots with sparser *L. sinense* plants on detection and mapping are matters for further analysis. Other technical issues influencing accuracy estimates were GPS positional errors from the fieldwork, loss of information during data conversion (e.g., float to integer), and the quality of LiDAR data and derived variables at regional extents, including gaps between LiDAR tiles and non-calibrated LiDAR intensity. Computational challenges that arose due to large data volumes associated with the large-area LiDAR application suggest that we need to explore data reduction techniques and optimal resolutions for detection and mapping of understory plant invaders.

We observed little difference in accuracy estimates when implementing the RF classifier at the experimental and county extents, with only slightly reduced variability between resubstitution and 10-fold cross validation at the county extent. This reduced variability is attributable to the larger amounts of field data at the county extent. However, similar accuracy estimates at both extents suggests the suitability of the RF classifier when less field data is available, which is consistent with research presented by Martinuzzi et al. (2009). County-level analyses showed that topographic moisture index (TMI) and vegetation cover types – deciduous and evergreen - (V_{CVTYPE}) are the most important variables for predicting the distribution of *L. sinense* at that scale. Application of the RF classifier at the county extent prompted two methodological avenues for future research: 1) to analyze LiDAR-derived PVs with spatial resolution matching the field plot size (i.e. 10 m x 10 m) to enhance data handling and computational efficiency while providing sufficient accuracy estimates, and 2) to measure the variability in accuracy estimates when using other evergreen understory invasive species found in the field plots. Ultimately, using county-level field data with the RF classifier is suitable for large-area assessment of *L. sinense*, and this approach can be applied to other local and regional sites.

5. Conclusions

Successful management of native biodiversity and ecosystem services at regional scales requires methodological advances in the application of remote sensing technologies aiming to accurately quantify the spatial distribution of invasive species. Our results suggest

that LiDAR data provides 1) adequate information for accurately detecting and mapping the spatial distribution of the understory invasive *L. sinense* over large urbanizing regions and 2) improved results as compared to the analysis of IKONOS optical imagery. The *Topography* variables (TMI + TRI), when combined with *Overall Vegetation Characteristics* (density + spectral) and *Overstory* (height), produced the highest accuracy estimates. These findings highlight the significance of the structural and spectral characteristics of LiDAR for quantifying the landscape-level topographic, forest stand, and physical and structural plant attributes used for detecting and mapping the distribution of *L. sinense*. LiDAR intensity-derived variables were significant in the successful detection and mapping of *L. sinense*, emphasizes the importance of the spectral attributes of LiDAR data for assessing understory invasives. This research demonstrates how categorizing LiDAR-derived variables into variable types representing *Overstory* and *Understory* forest structure and *Topography* contributed to our understanding of the forest and landscape characteristics that influence the distribution, growth, and detection of *L. sinense*. Our study suggests an even distribution of *L. sinense* in forests along the urban-rural gradient of Mecklenburg County, NC, found predominantly in drainage ways, lowlands, and disturbed areas of forested landscapes as observed during fieldwork. This research further highlights the importance of LiDAR data acquisition during the leaf-off season for mapping *L. sinense*. Our methodological framework addresses data complexities, including multidimensionality and nonlinearity in multisource data, and provides a generalized approach for rapid assessments of *L. sinense* in forested landscapes over large urbanizing regions to support development and implementation of management plans for conserving native biodiversity.

Acknowledgements

This research was supported by the Garden Club of America (GCA) Zone VI fellowship in urban forestry, Casey Trees Endowment Fund, and the AAG (Association of American Geographers) Dissertation Research grants. We express sincere thanks to Storm Water Services Division - Charlotte-Mecklenburg County Government Office, and Engineering and Property Management, Land Development Services, City of Charlotte, for

providing LiDAR data. We thank S. Gagné (UNC Charlotte) for providing IKONOS data. We also thank fellow graduate students and doctoral dissertation committee members from the Center for Geospatial Analytics, North Carolina State University, for their comments and feedback on the manuscript.

REFERENCES

Andrew, M.E., & Ustin, S.L. (2009). Habitat suitability modelling of an invasive plant with advanced remote sensing data. *Diversity and Distributions, 15*, 627-640

Asner, G.P., Knapp, D.E., Kennedy-Bowdoin, T., Jones, M.O., Martin, R.E., Boardman, J., & Hughes, R.F. (2008). Invasive species detection in Hawaiian rainforests using airborne imaging spectroscopy and LiDAR. *Remote Sensing of Environment, 112*, 1942-1955

Asner, G.P., & Vitousek, P.M. (2005). Remote analysis of biological invasion and biogeochemical change. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 4383-4386

BCAL LiDAR Tools (2013). Idaho State University, Department of Geosciences In. Boise, Idaho: Boise Center Aerospace Laboratory (BCAL)

Becker, R., Zmijewski, K., & Crail, T. (2013). Seeing the forest for the invasives: mapping buckthorn in the Oak Openings. *Biological Invasions, 15*, 315-326

Bork, E.W., & Su, J.G. (2007). Integrating LIDAR data and multispectral imagery for enhanced classification of rangeland vegetation: A meta analysis. *Remote Sensing of Environment, 111*, 11-24

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5-32

Brown, C.E., & Pezeshki, S.R. (2000). A study on waterlogging as a potential tool to control Ligustrum sinense populations in western Tennessee. *Wetlands, 20*, 429-437

Carleer, A., & Wolff, E. (2004). Exploitation of very high resolution satellite data for tree species identification. *Photogrammetric Engineering & Remote Sensing, 70*, 135-140

Carpenter, G.A., Gopal, S., Macomber, S., Martens, S., & Woodcock, C.E. (1999). A neural network method for mixture estimation for vegetation mapping. *Remote Sensing of Environment, 70*, 138-152

Chastain, R.A., & Townsend, P.A. (2007). Use of landsat ETM and topographic data to characterize evergreen understory communities in appalachian deciduous forests. *Photogrammetric Engineering & Remote Sensing, 73*, 563-575

Chen, Q. (2007). Airborne lidar data processing and information extraction. *Photogrammetric Engineering & Remote Sensing, 73*, 109-112

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., & Hess, K.T. (2007). Random forests for classification in ecology. *Ecology, 88*, 2783-2792

Dantonio, C.M., & Vitousek, P.M. (1992). Biological Invasions by Exotic Grasses, the Grass Fire Cycle, and Global Change. *Annual Review of Ecology and Systematics, 23*, 63-87

Davies, K.W., Petersen, S.L., Johnson, D.D., Davis, D.B., Madsen, M.D., Zvirzdin, D.L., & Bates, J.D. (2010). Estimating Juniper Cover From National Agriculture Imagery Program (NAIP) Imagery and Evaluating Relationships Between Potential Cover and Environmental Variables. *Rangeland Ecology & Management, 63*, 630-637

Donoghue, D.N.M., Watt, P.J., Cox, N.J., & Wilson, J. (2007). Remote sensing of species mixtures in conifer plantations using LiDAR height and intensity data. *Remote Sensing of Environment, 110*, 509-522

Dubayah, R.O., & Drake, J.B. (2000). Lidar remote sensing for forestry. *Journal of Forestry, 98*, 44-46

EDDMapS (2014). Early Detection & Distribution Mapping System. In. The University of Georgia - Center for Invasive Species and Ecosystem Health: The University of Georgia

Erdody, T.L., & Moskal, L.M. (2010). Fusion of LiDAR and imagery for estimating forest canopy fuels. *Remote Sensing of Environment, 114*, 725-737

Estes, L.D., Reillo, P.R., Mwangi, A.G., Okin, G.S., & Shugart, H.H. (2010). Remote sensing of structural complexity indices for habitat and species distribution modeling. *Remote Sensing of Environment, 114*, 792-804

Estornell, J., Ruiz, L.A., Velazquez-Marti, B., & Fernandez-Sarria, A. (2011). Estimation of shrub biomass by airborne LiDAR data in small forest stands. *Forest Ecology and Management, 262*, 1697-1703

Everitt, J.H., Anderson, G.L., Escobar, D.E., Davis, M.R., Spencer, N.R., & Andrascik, R.J. (1995). Use of remote sensing for detecting and mapping leafy spurge (Euphorbia esula). *Weed Technology*, 599-609

Falkowski, M.J., Evans, J.S., Martinuzzi, S., Gessler, P.E., & Hudak, A.T. (2009). Characterizing forest succession with lidar data: An evaluation for the Inland Northwest, USA. *Remote Sensing of Environment, 113*, 946-956

Faulkner, J.L., Clebsch, E.E.C., & Sanders, W.L. (1989). Use of Prescribed Burning for Managing Natural and Historic Resources in Chickamauga and Chattanooga National Military Park, USA. *Environmental Management, 13*, 603-612

Glenn, N.F., Mundt, J.T., Weber, K.T., Prather, T.S., Lass, L.W., & Pettingill, J. (2005). Hyperspectral data processing for repeat detection of small infestations of leafy spurge. *Remote Sensing of Environment, 95*, 399-412

Goetz, S.J., Steinberg, D., Betts, M.G., Holmes, R.T., Doran, P.J., Dubayah, R., & Hofton, M. (2010). Lidar remote sensing variables predict breeding habitat of a Neotropical migrant bird. *Ecology, 91*, 1569-1576

Greene, B.T., & Blossey, B. (2012). Lost in the weeds: Ligustrum sinense reduces native plant growth and survival. *Biological Invasions, 14*, 139-150

Guo, L., Chehata, N., Mallet, C., & Boukir, S. (2011). Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*, 56-66

Hart, J.L., & Holmes, B.N. (2013). Relationships between Ligustrum sinense Invasion, Biodiversity, and Development in a Mixed Bottom land Forest. *Invasive Plant Science and Management, 6*, 175-186

Hestir, E.L., Khanna, S., Andrew, M.E., Santos, M.J., Viers, J.H., Greenberg, J.A., Rajapakse, S.S., & Ustin, S.L. (2008). Identification of invasive vegetation using hyperspectral remote sensing in the California Delta ecosystem. *Remote Sensing of Environment, 112*, 4034-4047

Homolova, L., Maenovsky, Z., Clevers, J.G.P.W., Garcia-Santos, G., & Schaeprnan, M.E. (2013). Review of optical-based remote sensing for plant trait mapping. *Ecological Complexity, 15*, 1-16

Hsieh, P.F., Lee, L.C., & Chen, N.Y. (2001). Effect of spatial resolution on classification errors of pure and mixed pixels in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing, 39*, 2657-2663

Huang, C.-y., & Asner, G.P. (2009). Applications of remote sensing to alien invasive plant studies. *Sensors, 9*, 4869-4889

Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E., & Falkowski, M.J. (2008). Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment, 112*, 2232-2245

Hudson, J.R., Hanula, J.L., & Horn, S. (2013). Removing Chinese privet from riparian forests still benefits pollinators five years later. *Biological Conservation, 167*, 355-362

Huete, A.R., Liu, H.Q., Batchily, K., & vanLeeuwen, W. (1997). A comparison of vegetation indices global set of TM images for EOS-MODIS. *Remote Sensing of Environment, 59*, 440-451

Iverson, L.R., Dale, M.E., Scott, C.T., & Prasad, A. (1997). A GIS-derived integrated moisture index to predict forest composition and productivity of Ohio forests (USA). *Landscape Ecology, 12*, 331-348

Jaskierniak, D., Lane, P.N.J., Robinson, A., & Lucieer, A. (2011). Extracting LiDAR indices to characterise multilayered forest structure using mixture distribution functions. *Remote Sensing of Environment, 115*, 573-585

Jones, T.G., Coops, N.C., & Sharma, T. (2010). Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific Northwest, Canada. *Remote Sensing of Environment, 114*, 2841-2852

Keating, K.A., Gogan, P.J.P., Vore, J.M., & Irby, L.R. (2007). A simple solar radiation index for wildlife habitat studies. *Journal of Wildlife Management, 71*, 1344-1348

Kim, S., McGaughey, R.J., Andersen, H.E., & Schreuder, G. (2009). Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sensing of Environment, 113*, 1575-1586

Laba, M., Downs, R., Smith, S., Welsh, S., Neider, C., White, S., Richmond, M., Philpot, W., & Baveye, P. (2008). Mapping invasive wetland plants in the Hudson River National Estuarine Research Reserve using quickbird satellite imagery. *Remote Sensing of Environment, 112*, 286-300

Langeland, K.A., & Burks, K.C. (1998). Identification and Biology of Non-Native Plants in Florida's Natural Areas. In K.A. Langeland, & K.C. Burks (Eds.) (p. 165). Gainesville, University of Florida

Lawrence, R.L., Wood, S.D., & Sheley, R.L. (2006). Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). *Remote Sensing of Environment, 100*, 356-362

Lefsky, M.A., Cohen, W.B., Parker, G.G., & Harding, D.J. (2002). Lidar remote sensing for ecosystem studies. *BioScience, 52,* 19-30

Linderman, M., Liu, J., Qi, J., An, L., Ouyang, Z., Yang, J., & Tan, Y. (2004). Using artificial neural networks to map the spatial distribution of understorey bamboo from remote sensing data. *International Journal of Remote Sensing, 25*, 1685-1700

Malhi, Y., & Roman-Cuesta, R.M. (2008). Analysis of lacunarity and scales of spatial homogeneity in IKONOS images of Amazonian tropical forest canopies. *Remote Sensing of Environment, 112,* 2074-2087

Martinuzzi, S., Vierling, L.A., Gould, W.A., Falkowski, M.J., Evans, J.S., Hudak, A.T., & Vierling, K.T. (2009). Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. *Remote Sensing of Environment, 113,* 2533-2546

Merriam, R.W., & Feil, E. (2002). The Potential Impact of An Introduced Shrub on Native Plant Diversity and Forest Regeneration. *Biological Invasions, 4*, 369-373

Miller, J.H., Chambliss, E.B., & Oswalt, C.M. (2008). Maps of occupation and estimates of acres covered by nonnative invasive plants in southern forests using SRS FIA data. In

Miura, N., & Jones, S.D. (2010). Characterizing forest ecological structure using pulse types and heights of airborne laser scanning. *Remote Sensing of Environment, 114*, 1069-1076

Morris, L.L., Walck, J.L., & Hidayati, S.N. (2002). Growth and reproduction of the invasive Ligustrum sinense and native Forestiera ligustrina (Oleaceae): Implications for the invasion and persistence of a nonnative shrub. *International Journal of Plant Sciences, 163*, 1001-1010

Mutlu, M., Popescu, S.C., Stripling, C., & Spencer, T. (2008). Mapping surface fuel models using lidar and multispectral data fusion for fire behavior. *Remote Sensing of Environment, 112*, 274-285

Myeong, S., Nowak, D.J., & Duggin, M.J. (2006). A temporal analysis of urban forest carbon storage using remote sensing. *Remote Sensing of Environment, 101*, 277-282

Nelson, S.A.C., Cheruvelil, K.S., & Soranno, P.A. (2006). Satellite remote sensing of freshwater macrophytes and the influence of water clarity. *Aquatic Botany, 85*, 289-298

Noujdina, N.V., & Ustin, S.L. (2009). Mapping downy brome (Bromus tectorum) using multidate AVIRIS data

Orka, H.O., Wulder, M.A., Gobakken, T., & Naesset, E. (2012). Subalpine zone delineation using LiDAR and Landsat imagery. *Remote Sensing of Environment, 119*, 11-20

Popescu, S.C., Wynne, R.H., & Scrivani, J.A. (2004). Fusion of small-footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, USA. *Forest Science, 50*, 551-565

Prasad, A.M., Iverson, L.R., & Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems, 9*, 181-199

Pu, R.L. (2011). Mapping urban forest tree species using IKONOS imagery: preliminary results. *Environmental Monitoring and Assessment, 172*, 199-214

R Core Team (2013). R: A Language and Environment for Statistical Computing. In. Vienna, Austria: R Foundation for Statistical Computing.

Resasco, J., Hale, A.N., Henry, M.C., & Gorchov, D.L. (2007). Detecting an invasive shrub in a deciduous forest understory using late-fall Landsat sensor imagery. *International Journal of Remote Sensing, 28*, 3739-3745

Rich, R.L., Frelich, L., Reich, P.B., & Bauer, M.E. (2010). Detecting wind disturbance severity and canopy heterogeneity in boreal forest by coupling high-spatial resolution satellite imagery and field data. *Remote Sensing of Environment, 114*, 299-308

Royo, A.A., & Carson, W.P. (2006). On the formation of dense understory layers in forests worldwide: consequences and implications for forest dynamics, biodiversity, and succession. *Canadian Journal of Forest Research, 36*, 1345-1362

Rushton, S.P., Ormerod, S.J., & Kerby, G. (2004). New paradigms for modelling species distributions? *Journal of Applied Ecology, 41*, 193-200

Seidel, D., Albert, K., Fehrmann, L., & Ammer, C. (2012). The potential of terrestrial laser scanning for the estimation of understory biomass in coppice-with-standard systems. *Biomass & Bioenergy, 47*, 20-25

Simberloff, D., Martin, J.L., Genovesi, P., Maris, V., Wardle, D.A., Aronson, J., Courchamp, F., Galil, B., Garcia-Berthou, E., Pascal, M., Pysek, P., Sousa, R., Tabacchi, E., & Vila, M. (2013). Impacts of biological invasions: what's what and the way forward. *Trends in Ecology & Evolution, 28*, 58-66

Singh, K.K., Vogler, J.B., Meng, Q., & Meentemeyer, R.K. (2010). Mapping Land Use Patterns in an Urbanizing Landscape Using LiDAR Intensity Data. In, *Optical Remote Sensing of the Environment* (p. OMC2). Tucson, AZ: Optical Society of America

Singh, K.K., Vogler, J.B., Shoemaker, D.A., & Meentemeyer, R.K. (2012). LiDAR-Landsat data fusion for large-area assessment of urban land cover: Balancing spatial resolution, data volume and mapping accuracy. *ISPRS Journal of Photogrammetry and Remote Sensing, 74*, 110-121

Stambaugh, M.C., & Guyette, R.P. (2008). Predicting spatio-temporal variability in fire return intervals using a topographic roughness index. *Forest Ecology and Management, 254*, 463-473

Straatsma, M., & Middelkoop, H. (2007). Extracting structural characteristics of herbaceous floodplain vegetation under leaf-off conditions using airborne laser scanner data. *International Journal of Remote Sensing, 28*, 2447-2467

Stumpf, A., & Kerle, N. (2011). Object-oriented mapping of landslides using Random Forests. *Remote Sensing of Environment, 115*, 2564-2577

Tuanmu, M.N., Vina, A., Bearer, S., Xu, W.H., Ouyang, Z.Y., Zhang, H.M., & Liu, J.G. (2010). Mapping understory vegetation using phenological characteristics derived from remotely sensed data. *Remote Sensing of Environment, 114*, 1833-1844

Ulyshen, M.D., Horn, S., & Hanula, J.L. (2010). Response of beetles (Coleoptera) at three heights to the experimental removal of an invasive shrub, Chinese privet (Ligustrum sinense), from floodplain forests. *Biological Invasions, 12*, 1573-1579

Underwood, E., Ustin, S., & DiPietro, D. (2003). Mapping nonnative plants using hyperspectral imagery. *Remote Sensing of Environment, 86*, 150-161

USFWS (1991). Endangered and threatened wildlife and plants: determination of *Helianthus schweinitzii* (Schweinitz's sunflower) to be an endangered species. *Federal Register, 56*, 21087-21091.

Ustin, S.L., Roberts, D.A., Gamon, J.A., Asner, G.P., & Green, R.O. (2004). Using Imaging Spectroscopy to Study Ecosystem Processes and Properties. *BioScience, 54*, 523-534

Van Delm, A., & Gulinck, H. (2011). Classification and quantification of green in the expanding urban and semi-urban complex: Application of detailed field data and IKONOS-imagery. *Ecological Indicators, 11*, 52-60

Vierling, L.A., Rowell, E., Chen, X.X., Dykstra, D., & Vierling, K. (2002). Relationships among airborne scanning LiDAR, high resolution multispectral imagery, and ground-based inventory data in a ponderosa pine forest. *Igarss 2002: Ieee International Geoscience and Remote Sensing Symposium and 24th Canadian Symposium on Remote Sensing, Vols I-Vi, Proceedings*, 2912-2914

Wang, H.H., & Grant, W.E. (2012). Determinants of Chinese and European Privet (Ligustrum sinense and Ligustrum vulgare) Invasion and Likelihood of Further Invasion in Southern U.S. Forestlands. *Invasive Plant Science and Management, 5*, 454-463

Wang, T.J., Skidmore, A.K., & Toxopeus, A.G. (2009). Improved understorey bamboo cover mapping using a novel hybrid neural network and expert system. *International Journal of Remote Sensing, 30*, 965-981

Waske, B., van der Linden, S., Oldenburg, C., Jakimow, B., Rabe, A., & Hostert, P. (2012). imageRF - A user-oriented implementation for remote sensing image analysis with Random Forests. *Environmental Modelling & Software, 35*, 192-193

Wilcox, J., & Beck, C.W. (2007). Effects of Ligustrum sinense Lour. (Chinese privet) on abundance and diversity of songbirds and native plants in a southeastern nature preserve. *Southeastern Naturalist, 6*, 535-550

Wilfong, B.N., Gorchov, D.L., & Henry, M.C. (2009). Detecting an Invasive Shrub in Deciduous Forest Understories Using Remote Sensing. *Weed Science, 57*, 512-520

Wing, B.M., Ritchie, M.W., Boston, K., Cohen, W.B., Gitelman, A., & Olsen, M.J. (2012). Prediction of understory vegetation cover with airborne lidar in an interior ponderosa pine forest. *Remote Sensing of Environment, 124*, 730-741

Yu, X.W., Hyyppa, J., Vastaranta, M., Holopainen, M., & Viitala, R. (2011). Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. *ISPRS Journal of Photogrammetry and Remote Sensing, 66,* 28-37

Table 1. Distribution of 100 m^2 field plots along urban-rural gradients.

Class	County extent		Total	Experimental extent		Total
	Presence	Absence		Presence	Absence	
Urban	68	88	156	16	11	27
Suburban	27	113	140	7	43	50
Rural	4	46	50	0	4	4
	99	247	346	23	58	81

[a]Field data used in the analyses at the experimental extent

Table 2. Predictor variables used in random forests and logistic regression classifiers for detection and mapping of *Ligustrum sinense*. LiDAR-derived variables assigned to *Overstory*, *Understory*, *Topography*, and *Other* types while IKONOS-derived variables are *Optical* type.

Variable type	Variable name	Description
Overstory	H_{MAX}	Height maximum
	$H_{PCT\ (75\text{-}95)}$	Height percentile 75th, 90th, and 95th
	V_{HT3}†	Vegetation height > 5 m and ≤ 10 m
	V_{HT4}†	Vegetation height > 10 m
	$V_{STRATUM\ (4\text{-}6)}$	Percentage of vegetation returns in the height > 10 m and ≤ 20 m, 20 m and ≤ 30 m, and > 30 m
Understory	H_{MIN}	Height minimum
	$H_{PCT\ (5\text{-}50)}$	Height percentile 5th, 10th, 25th, and 50th (median)
	V_{HT1}†	Vegetation height > 0 m and ≤ 1 m
	V_{HT2}†	Vegetation height > 1 m and ≤ 5 m (understory vegetation)
	$V_{STRATUM\ (0\text{-}3)}$	Percentage of vegetation returns in the height range > 0 m and ≤ 1 m, > 1 m and ≤ 2.5 m, and > 2.5 m and ≤ 10 m
Topography	ASP†	Aspect
	BE_{AR}	Bare earth absolute roughness
	BE_{GPD}	Ground Point-density
	BE_{IMIN}	Bare earth intensity minimum
	BE_{IMAX}	Bare earth intensity maximum
	BE_{IMEA}	Bare earth intensity mean
	BE_{ISD}	Bare earth intensity standard deviation
	BE_{Lmin}	Bare earth elevation minimum
	BE_{Lmax}	Bare earth elevation maximum
	BE_{Lmea}	Bare earth elevation mean
	BE_{LR}	Bare earth local roughness
	$BE_{SLPCOSASP}$	Bare earth slope cosine aspect
	$BE_{SLPSINASP}$	Bare earth slope sine aspect
	DEM†	Digital elevation model
	ERR†	Elevation relief ratio
	Crv†	Curvature surface
	H_{DE}†	Hill-shade
	N_{LGR}	Number of LiDAR ground returns
	PGR	Percentage ground return
	SLP	Slope surface (degree)
	SRI†	Solar radiation index
	TMI†	Topographic moisture index
	TRI†	Topographic roughness index

Table 2. (continued).

Variable type	Variable name	Description
Overall Vegetation Characteristics	H_{AAD}	Height average absolute deviation
	H_{CV}	Height coefficient of variation
	H_{IQR}	Height interquartile range
	H_{KUR}	Height kurtosis
	H_{MAD}	Height median absolute deviation
	H_{MEAN}	Height mean
	H_{RNG}	Height range
	H_{SD}	Height standard deviation
	H_{SKE}	Height skewness
	H_{TEX}	Height texture
	H_{VAR}	Height variance
	I_{ALL}	Intensity surface overall
	I_{MIN}	Intensity surface minimum
	I_{MAX}	Intensity surface maximum
	I_{MEAN}	Intensity surface mean
	I_{SD}	Intensity surface standard deviation
	N_{LR}	Number of LiDAR returns
	N_{LVR}	Number of LiDAR vegetation returns
	V_{TVD}	Total vegetation density
	VC_{TYPE} [±]	Vegetation cover type (deciduous, evergreen and mixed)
	VI_{MIN}	Vegetation intensity minimum
	VI_{MAX}	Vegetation intensity maximum
	VI_{MEAN}	Vegetation intensity mean
	VI_{SD}	Vegetation intensity standard deviation
Optical	Texture	Texture
	ARVI[±]	Atmospherically resistant vegetation index
	MSAVI[±]	Modified soil adjusted vegetation index
	IHS[±]	Hue, intensity and saturation
	EVI[±]	Enhanced vegetation index
	Pan	Pan-sharpened bands (three visible and one infrared)

Listed predictor variables were developed using BCAL LiDAR tools, based on IDL programing language (ENVI 5.0), [†]ArcGIS 10.1, and [±]ERDAS Imagine software. See Appendices for schematic details used for creating predictor variables.

Table 3. Model accuracy estimates using random forests classifier at the experimental extent.

Models	Accuracy estimate†				
	F-1 measure±	Kappa	Specificity	Sensitivity	AUC‡
M_{Os}	57.16	0.239	0.636	0.754	0.695
M_{Us}	41.34	0.000	0.000	0.722	0.361
M_{To}	60.04	**0.308**	0.636	0.790	**0.713**
M_{Ot}	47.46	0.086	0.417	0.734	0.576
M_{Op}	42.58	0.019	0.125	0.724	0.425
M_{OsUs}	43.48	0.038	0.167	0.724	0.445
M_{OsTo}	57.05	0.260	0.488	0.767	0.628
M_{UsTo}	55.70	0.213	0.448	0.742	0.595
M_{OsOt}	43.25	0.034	0.250	0.727	0.489
M_{UsOt}	42.69	0.000	0.125	0.725	0.425
M_{ToOt}	52.93	0.213	0.375	0.746	0.560
M_{OsOp}	48.61	0.149	0.625	0.730	0.677
M_{UsOp}	41.63	0.000	0.000	0.722	0.361
M_{ToOp}	55.39	0.270	0.604	0.760	0.682
M_{OtOp}	47.55	0.101	0.500	0.733	0.616
M_{OsUsOt}	43.79	0.038	0.375	0.739	0.557
M_{UsToOt}	55.67	0.255	0.500	0.763	0.631
M_{OsToOt}	55.93	0.244	0.563	0.758	0.660
M_{OsUsTo}	55.53	0.236	0.388	0.765	0.576
M_{OsUsOp}	41.63	0.014	0.000	0.735	0.368
M_{OsToOp}	56.09	0.270	0.625	0.762	0.694
M_{UsToOp}	54.86	0.250	0.475	0.758	0.616
M_{OsOtOp}	48.28	0.129	0.500	0.733	0.617
M_{UsOtOp}	46.68	0.100	0.250	0.728	0.489
M_{ToOtOp}	53.51	0.184	0.469	0.802	0.635
$M_{OsUsToOt}$	52.69	0.180	0.375	0.751	0.563
$M_{OsUsToOp}$	51.94	0.205	0.500	0.747	0.623
$M_{OsUsOpOt}$	42.97	0.030	0.125	0.739	0.432
$M_{OsToOpOt}$	55.68	0.238	0.531	0.749	0.640
$M_{UsToOpOt}$	54.32	0.196	0.542	0.759	0.650
M_{ALL}	54.79	0.229	0.500	0.757	0.628

†10-fold cross-validated accuracy estimate

±*F-1 measure* = the harmonic mean of precision and recall;

‡*AUC* = the area under the curve of the receiver operating characteristic. It is defined as the average of sensitivity and the specificity, also known as balanced accuracy.

Table 4. Accuracy metrics for top performing models with observed and predicted frequencies at the experimental and county extents.

	Experimental extent						County extent		
	RF			LR			RF		
	Predicted			Predicted			Predicted		
Observed	Presence	Absence	% Correct	Presence	Absence	% Correct	Presence	Absence	% Correct
Presence	11	9	55.00	15	20	42.90	92	31	74.80
Absence	0	61	100.00	5	41	89.10	35	187	84.20
Overall % correct			88.89			69.14			80.87
Accuracy estimates	F-1 measures:	71.00		F-1 measures:	54.50		F-1 measures:	73.60	
	Kappa:	0.648		Kappa:	0.337		Kappa:	0.586	
	Sensitivity:	1.000		Sensitivity:	0.750		Sensitivity:	0.724	
	Specificity:	0.871		Specificity:	0.672		Specificity:	0.858	
	AUC:	0.936		AUC:	0.787		AUC:	0.791	
10-fold cross-validation	F-1 measures:	61.55		F-1 measures:	53.30		F-1 measures:	67.37	
	Kappa:	0.537		Kappa:	0.331		Kappa:	0.547	

RF = random forests; LR = logistic regression; F-1 measure = the harmonic mean of precision and recall; AUC = the area under the curve of the receiver operating characteristic.

Table 5. Parameters and fit statistics for logistic regression model.

Variable type	Predictor	Description	β	*SE* β	*z*	*p*	e^{β} (odds ratio)
	Constant		-0.933	0.602	-1.551	0.1208	NA
Understory	$V_{STRATUM-1}$	Percentage of vegetation returns	0.035	0.014	2.470	0.0135	1.036
Topography	BE_{ISD}	Bare earth intensity standard deviation	-0.051	0.022	-2.308	0.0210	0.950
	TMI	Topographic moisture index	0.014	0.003	4.558	<0.0001	1.014
Overall	H_{KUR}	Height kurtosis	-0.612	0.279	-2.188	0.0287	0.542
Vegetation	I_{ALL}	Intensity surface overall	-0.013	0.004	-2.924	0.0035	0.987
Characteristics	VI_{MEAN}	Vegetation intensity mean	-0.459	0.170	-2.695	0.0070	0.631
	VI_{SD}	Vegetation intensity standard deviation	0.005	0.002	2.012	0.0442	1.005

	Test		X^2	*df*	*p*
	Overall model fit	Likelihood ratio test	68.356	7	< 0.0001
		Wald test	13.900	2	0.0009
	Goodness-of-fit test	Hosmer & Lemeshow	7.569	8	0.4766

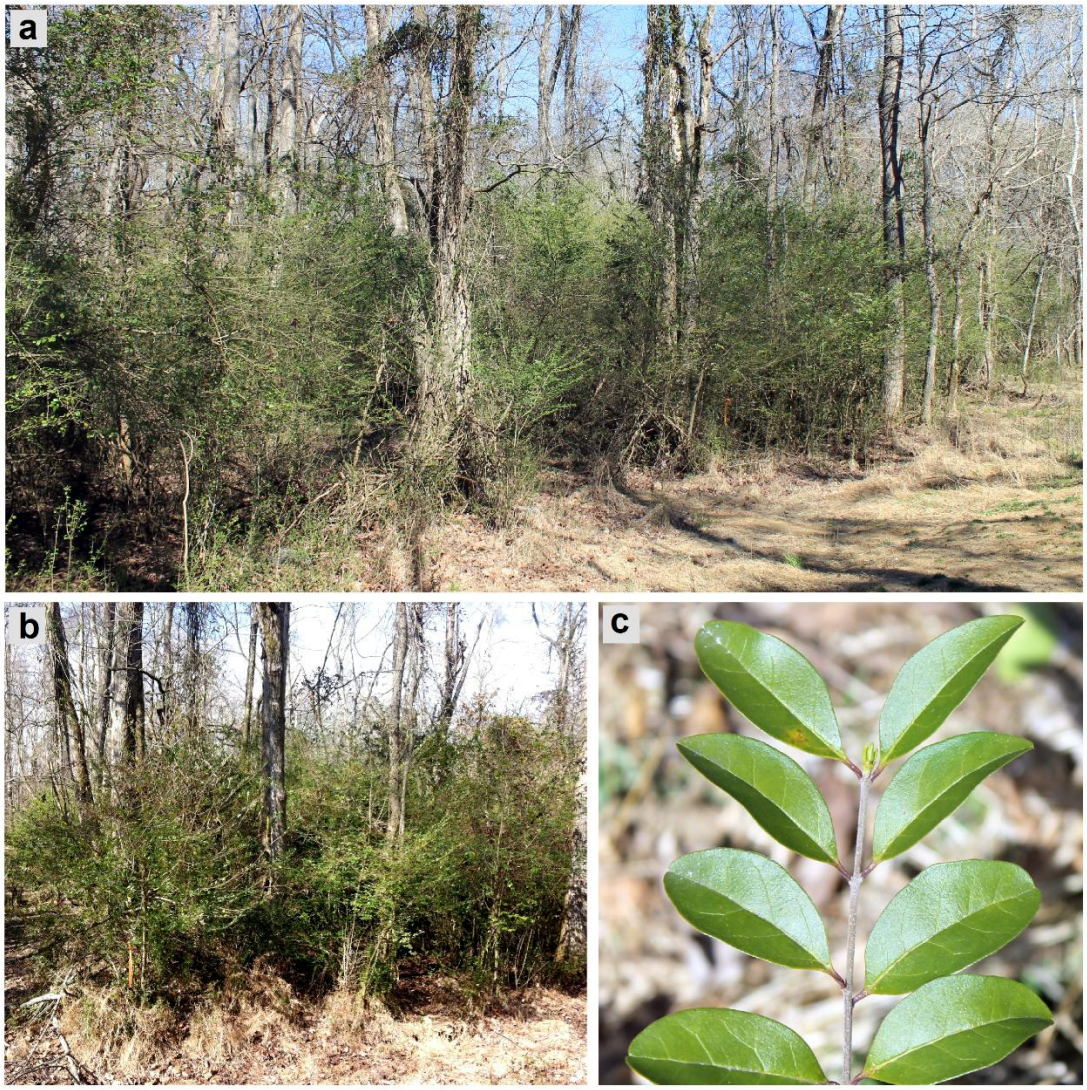


Fig. 1. *Ligustrum sinense* during the leaf-off season from backyards to backwoods. (a) a dense stand of *L. sinense* in the forest understory within the riparian zone along Toby Creek greenway in Charlotte, North Carolina, USA, (b) a dense monotypic stand of *L. sinense*, and (c) typical foliage exhibits thick leaves with a glossy upper surface and light green lower surface.

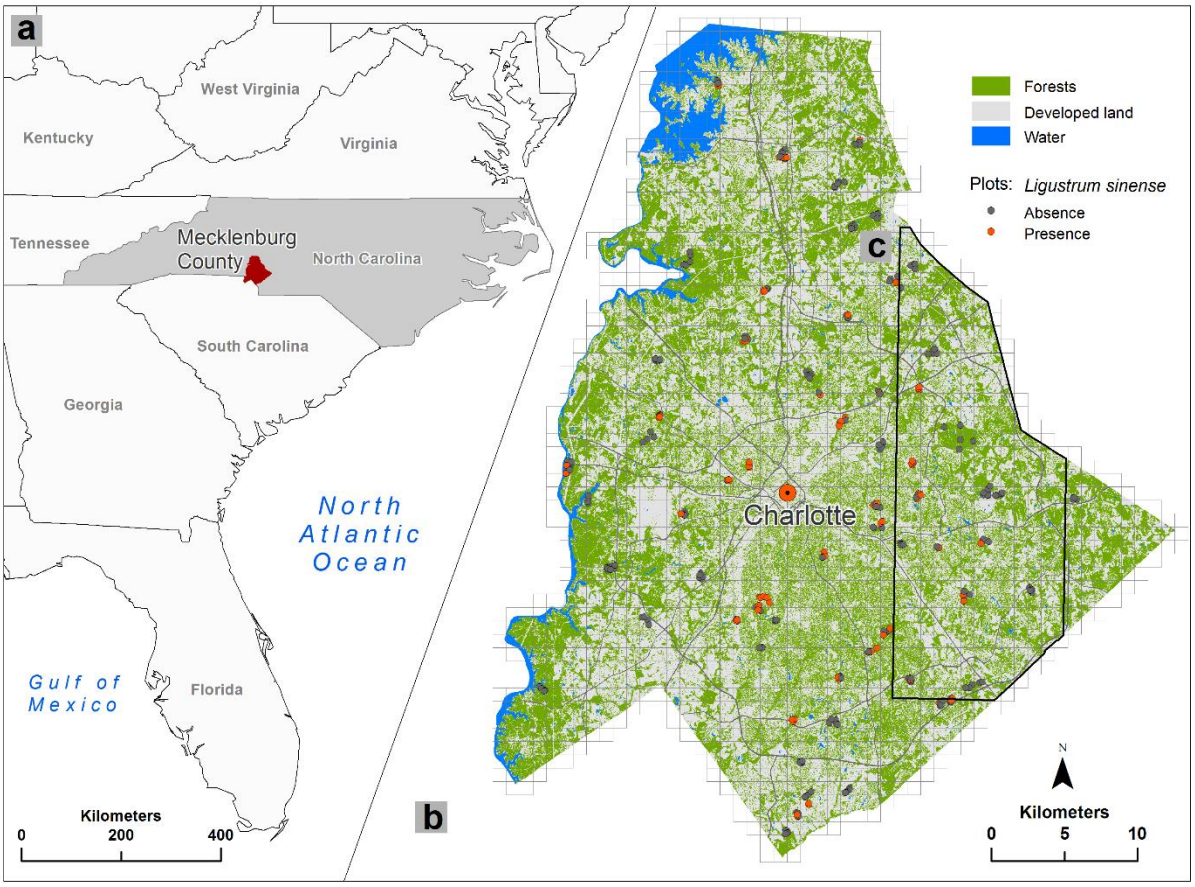


Fig. 2. Study system. (a) Mecklenburg County at the center of the Charlotte Metropolitan Area of North Carolina, USA, (b) the distribution of forest cover across the county with overlays of LiDAR tile index and locations of field plots. A total of 1896 LiDAR tiles of dimension 914.40 m x 914.40 m used for developing predictor variables, and (c) the extent of IKONOS coverage.

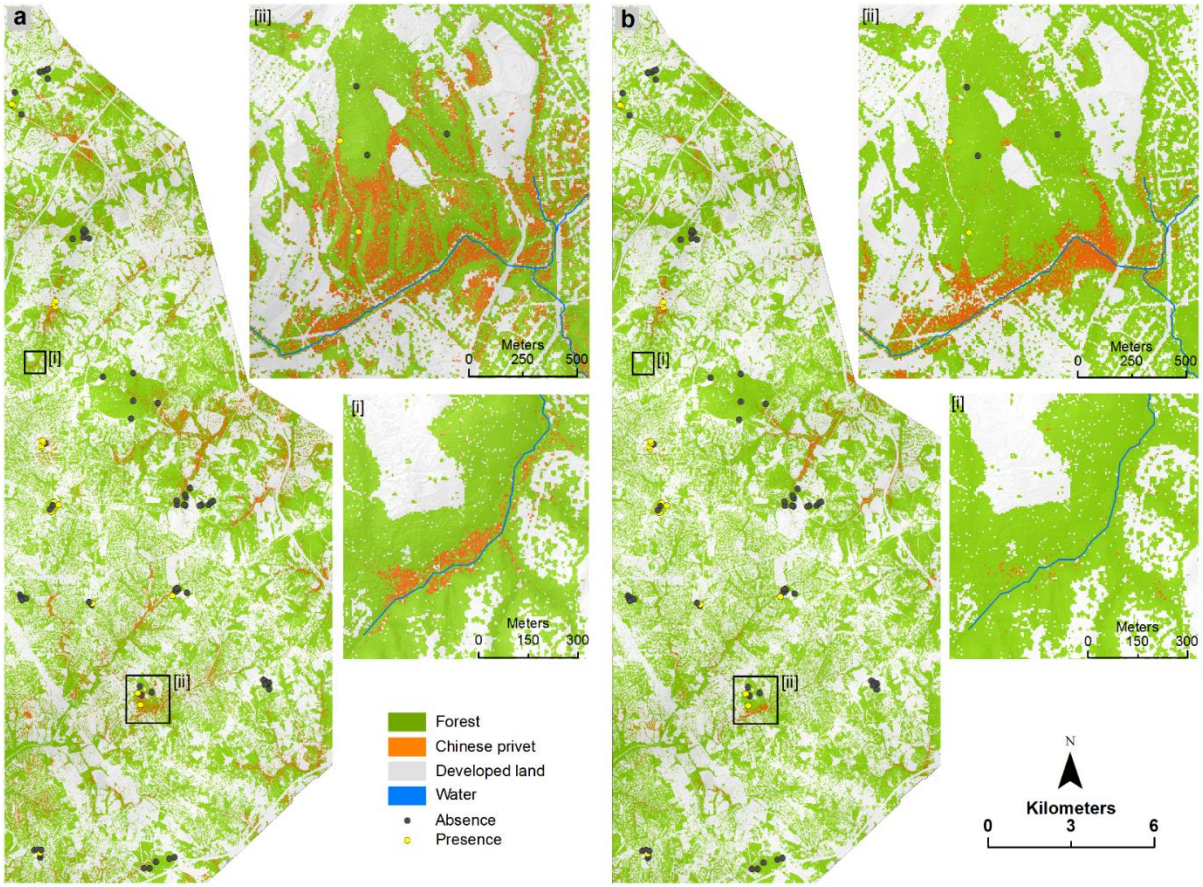
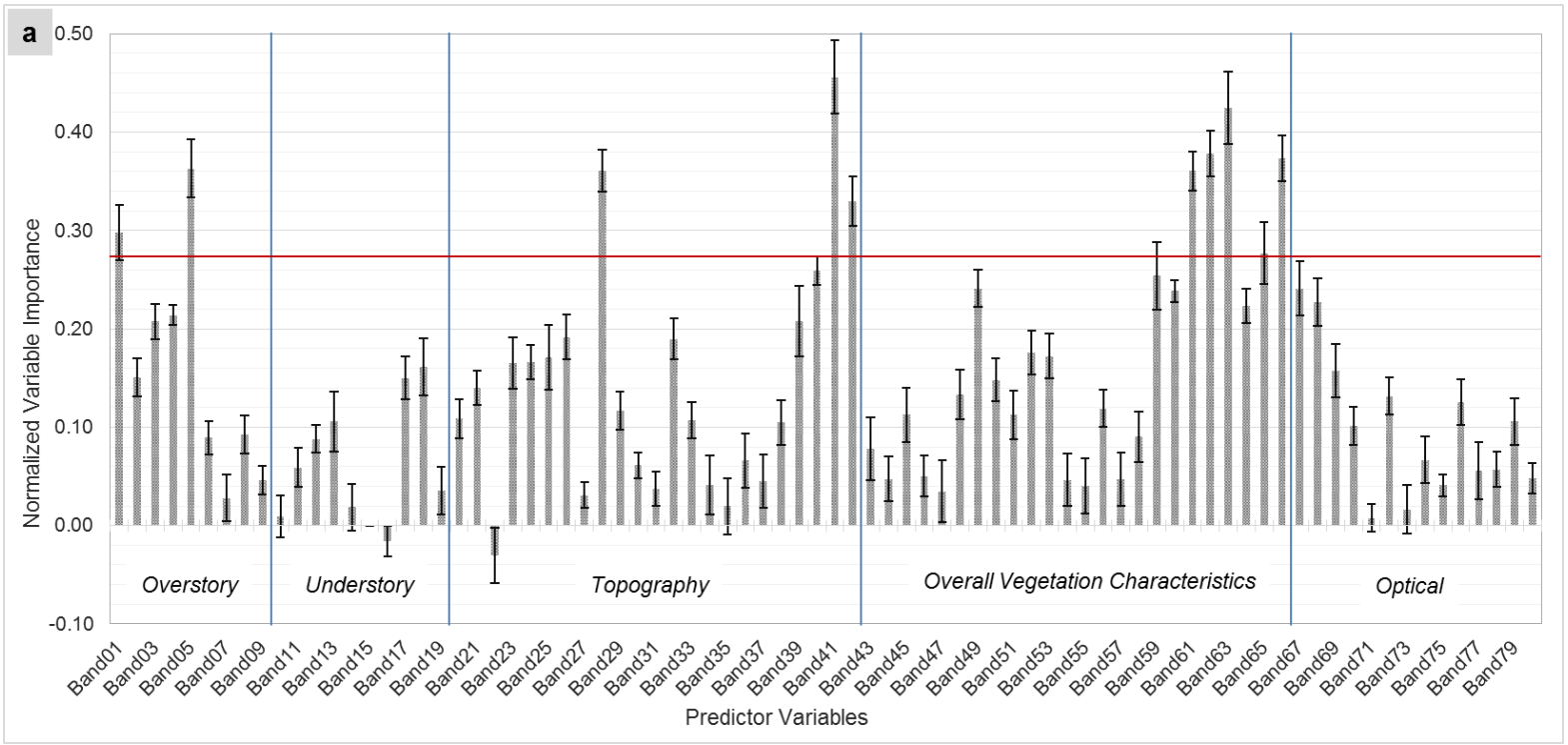
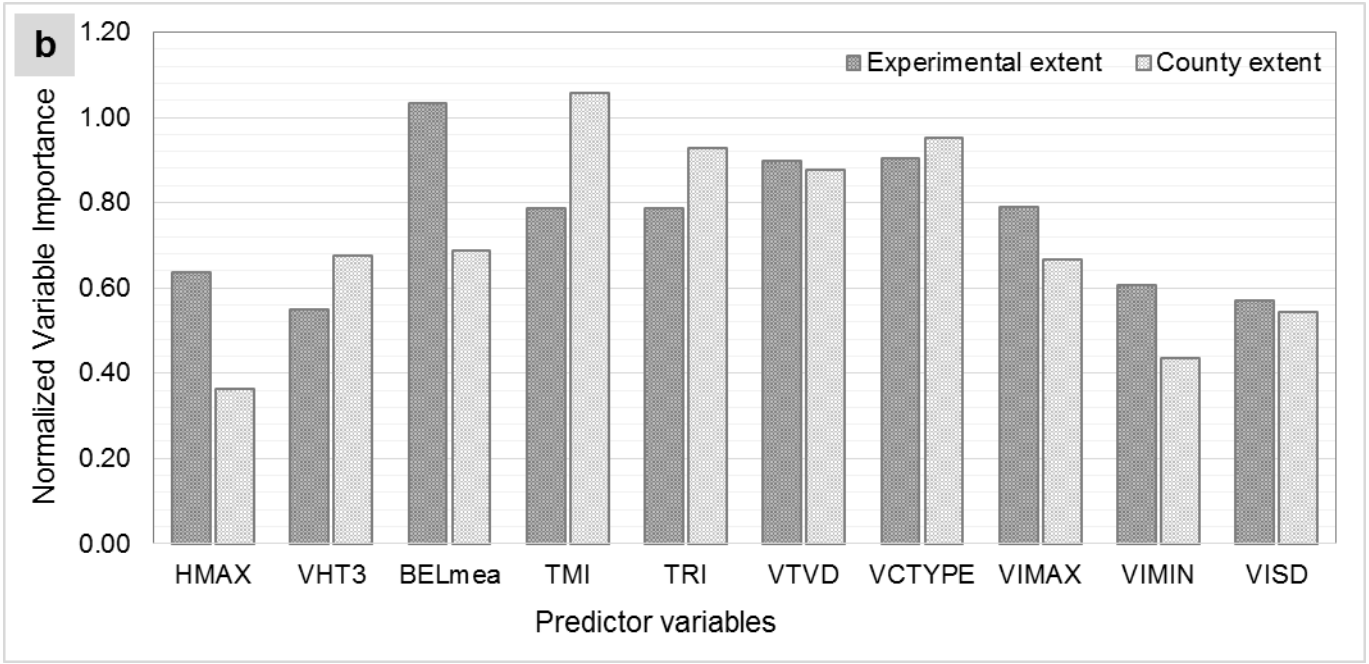


Fig. 3. The spatial distribution of *Ligustrum sinense* in the forest understory at experimental extent based on: (a) the top performing random forests model (F-1 accuracy = 69.90%) developed using selected predictor variables with high variable importance, and (b) a parsimonious logistic regression model (F-1 accuracy = 54.50%).

Fig. 4. Normalized variable importance (NVI) plot from the random forests classifier. NVI is a ratio between raw variable importance and the respective standard deviation. Plot represents the relative importance of each predictor variable in the model. Higher values indicate greater contribution to the model. (a) NVI plot for the model (M_{ALL}) with all predictor variables at experimental extent, and b) NVI plot for models developed from selected predictor variables at experimental and county extents.


a
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
Normalized Variable Importance
Overstory
Understory
Topography
Overall Vegetation Characteristics
Optical
Band01 Band03 Band05 Band07 Band09 Band11 Band13 Band15 Band17 Band19 Band21 Band23 Band25 Band27 Band29 Band31 Band33 Band35 Band37 Band39 Band41 Band43 Band45 Band47 Band49 Band51 Band53 Band55 Band57 Band59 Band61 Band63 Band65 Band67 Band69 Band71 Band73 Band75 Band77 Band79
Predictor Variables


b
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Normalized Variable Importance
Experimental extent
County extent
HMAX
VHT3
BELmea
TMI
TRI
VTVD
VCTYPE
VIMAX
VIMIN
VISD
Predictor variables

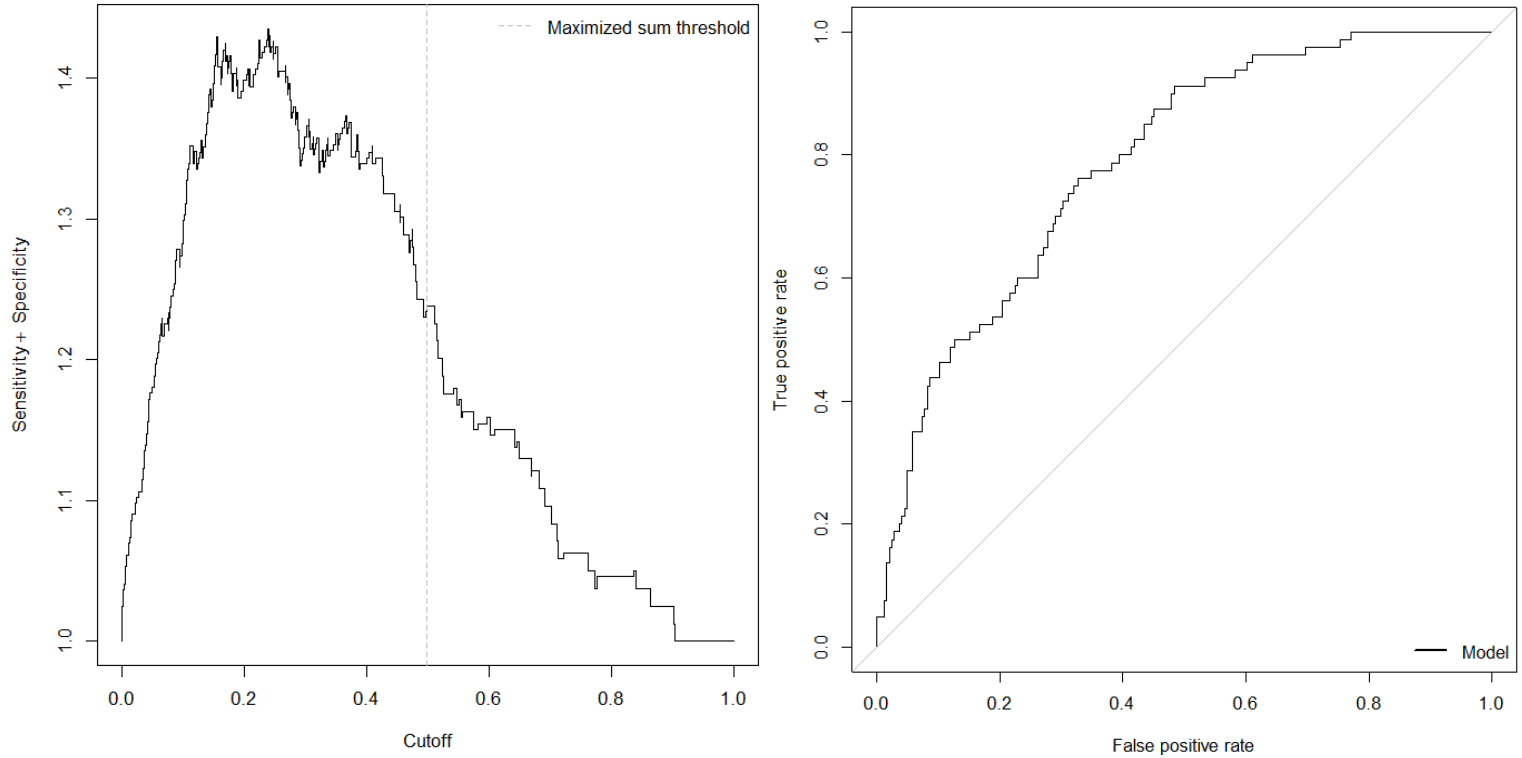


Fig. 5. Logistic regression results at the experimental extent. (a) A maximized sum of sensitivity and specificity criterion for selecting the threshold of occurrence, and (b) The receiver operating characteristic (ROC) curve for the logistic regression model. The stair-stepped line represents the ROC curve (area under the curve, 0.787 [95% CI, 0.74 to 0.83]).

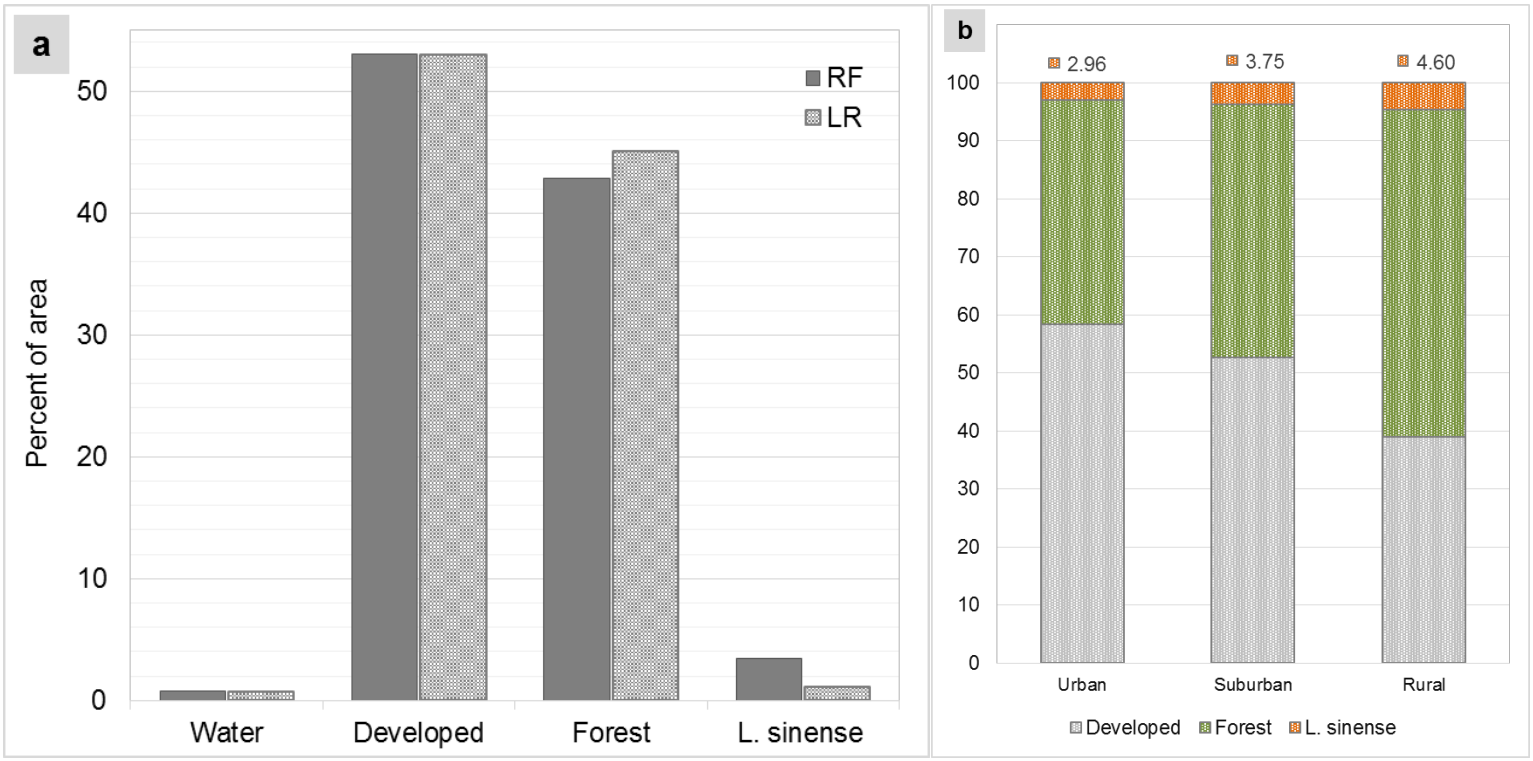


Fig. 6. Estimated *Ligustrum sinense* cover (% of area) in the forest understory at experimental extent. (a) The distribution of *L. sinense* and other land covers derived from top-performing random forests (RF) and logistic regression (LR) models, and b) the proportions of *L. sinense* cover along the urban-rural gradient.

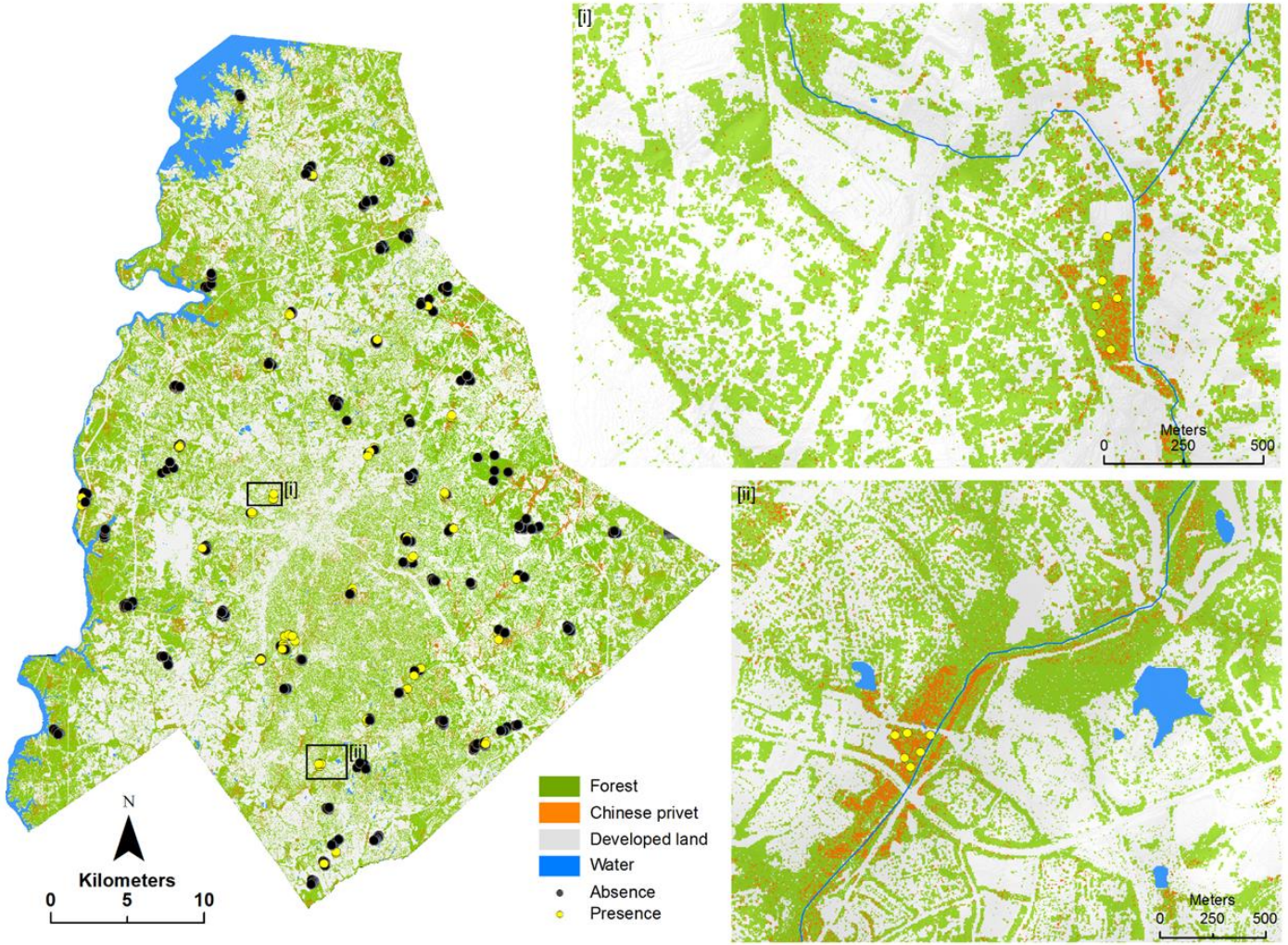


Fig. 7. The spatial distribution of *Ligustrum sinense* using the random forests classifier in the forest understory at the county extent.

CHAPTER 3. Effects of LiDAR Point-density

Effects of LiDAR point density and landscape context on estimates of urban forest biomass

Kunwar K. Singh[1,2]*, Gang Chen[3], James B. McCarter[1,2], and Ross K. Meentemeyer[1,2]

[1]Center for Geospatial Analytics

[2]Department of Forestry and Environmental Resources

North Carolina State University

Raleigh, North Carolina 27695

USA

[3]Department of Geography and Earth Sciences

University of North Carolina

9201 University City Blvd

Charlotte, NC 28223

USA

*Corresponding author: kun2001@gmail.com; phone: 01-704-359-7139; fax: 01-919-515-3439

gang.chen@uncc.edu

jbmccart@ncsu.edu

rkmeente@ncsu.edu

Abstract

Light Detection and Ranging (LiDAR) data is being increasingly used as an effective alternative to conventional optical remote sensing to accurately estimate aboveground forest biomass ranging from individual tree to stand levels. Recent advancements in LiDAR technology have resulted in higher point densities and improved data accuracies accompanied by challenges for procuring and processing voluminous LiDAR data for large-area assessments. Reducing point density lowers data acquisition costs and overcomes computational challenges for large-area forest assessments. However, how does lower point density affect the accuracy of biomass estimation in forests containing a great level of anthropogenic disturbance? We evaluate the effects of LiDAR point density on the biomass estimation of remnant forests in the rapidly urbanizing region of Charlotte, North Carolina, USA. We used multiple linear regression to establish the statistical relationship between field-measured biomass and predictor variables (PVs) derived from LiDAR point cloud data with varying densities. We compared the estimation accuracies between a general Urban Forest type and three Forest Type models (evergreen, deciduous, and mixed) and quantified the degree to which landscape context influenced biomass estimation. The explained biomass variance of the Urban Forest model, using *adjusted* R^2, was consistent across the reduced point densities, with the highest difference of 11.5% between the 100% and 1% point densities. The combined estimates of Forest Type biomass models outperformed the Urban Forest models at the representative point densities (100% and 40%). The Urban Forest biomass model with development density of 125 m radius produced the highest *adjusted* R^2 (0.83 and 0.82 at 100% and 40% LiDAR point densities, respectively) and the lowest RMSE values, highlighting a distance impact of development on biomass estimation. Our evaluation suggests that reducing LiDAR point density is a viable solution to regional-scale forest assessment without compromising the accuracy of biomass estimates, and these estimates can be further improved using development density.

Keywords: LiDAR, aboveground biomass, development density, canopy stratification, multiple linear regression

1. Introduction

Loss of forest biomass in urbanizing regions is a growing concern worldwide (Seto et al. 2012). Changes in biomass impact critical ecological and environmental processes necessary for the maintenance of biodiversity and ecosystem health, often for the long-term and with slowed or no opportunity for recovery. Long-term impacts include changes in regional nitrogen and carbon storage and flux (Groffman et al. 2006; Magnani et al. 2007; Nadelhoffer et al. 1999), increases in urban heat island effects (Imhoff et al. 2010; Zhang et al. 2013), and increased concentration of atmospheric carbon dioxide (Rowntree 1991 (Nowak and Greenfield 2012)), as well as changes in human perceptions of environmental quality and well-being (Grove et al. 2006). As urban forests decrease in size and number, we face an ever-increasing need to quantify the remaining resources. However, some of the most vulnerable forests around growing U.S. cities are the least accessible for ecological measurement due to the high proportion of privately owned land typical of metropolitan landscapes (Meentemeyer et al. 2013; Robertson et al. 2014). Since biomass directly relates to tree structure -- height and diameter at breast height -- cost-effective regional-scale remote sensing data that characterize the structure of forest stands are needed to quantify biomass across metropolitan landscapes. Light Detection and Ranging (LiDAR) provides structural data for forest analysis studies (Lefsky et al. 2002b).

Airborne and terrestrial LiDAR has emerged as a key remote sensing technology for the accurate estimation of forest biomass ranging from individual tree to stand levels (Allouis et al. 2013; Mascaro et al. 2011; Zolkos et al. 2013), and has been successfully applied for quantify and measure biomass of tropical forests (Drake et al. 2003; Dubayah et al. 2010a), shrubs (Estornell et al. 2011), understory vegetation (Seidel et al. 2012), unmanaged Mediterranean forests (Garcia et al. 2010), and urban forests (He et al. 2013). However, these applications have not characterized forest types found along urban-rural gradients, rarely address the impact of landscape context on biomass estimation across the urbanizing landscapes, and lack large-area assessment. Recent advancements in LiDAR technology have resulted in improved data accuracies using higher point densities (Csanyi and Toth 2007), which poses a challenge to the procurement and processing of voluminous LiDAR data for

large-area assessments. To overcome this challenge, large-area assessments are typically based on plot-level regression models either ‘alone’ (Drake et al. 2002) or through ‘fusion’ of sampled LiDAR transects with spectral data (Popescu et al. 2004). However, these approaches to large-area assessments are often not suitable for urban forest studies due to the presence of different forest types, and the high degrees of heterogeneity at fine scales caused by repeated anthropogenic disturbances, such as conversion of forests to development. Therefore, reducing point density may provide a solution for cutting procurement costs and overcoming computational challenges for large-area assessments. This knowledge may also help analyze features in point clouds derived from automated methods (e.g. Structure from Motion technique, (Brostow et al. 2008)) designed to generate a 3D point cloud from video sequences filmed from new mobile sensors.

Data procurement and processing costs limit the extent to which LiDAR is useful for large area studies (Wulder et al. 2008; Wulder et al. 2012). To achieve both cost-effective and accurate results, data acquisition parameters are optimized (Lovell et al. 2005; Naesset 2009) while tradeoffs are made between point-density and estimation accuracy (Jakubowski et al. 2013; Magnusson et al. 2007; Zhao et al. 2009). To date, studies considering the optimization of LiDAR point-densities, such as Gobakken and Naesset (2008), Jakubowski et al. (2013), Lim et al. (2008), and Treitz et al. (2012), were mainly focused on natural environments and rarely accounted for the influence of landscape context (e.g., surrounding urban development or forest stratification following disturbance) on forest structure and biomass estimation.

Forest structural heterogeneity and factors modifying it, such as urbanization (McHale et al. 2009), are important determinants in biomass estimation. Studies documenting the effects of urbanization on forest structure have found lower stem densities in young stands (Lawrence 1995), increased forest edge opening (Moran 1984), and increased biomass growth compared to rural stands (Gregg et al. 2003; O'Brien et al. 2012). These structural variations affect the total biomass, while also potentially affecting the capacity to measure biomass using LiDAR. Canopy stratification has been suggested to overcome potential estimation errors due to structural variation (Swatantran et al. 2011). Canopy stratification is

a useful organizational tool for the study of the vertical distribution of plants and animals (Baker and Wilson 2000), and is considered an index of vertical structure where the higher number of canopy strata represents increased complexity in the forest stands (Parker and Brown 2000). Fundamental differences in the heterogeneity of forests types in urbanizing landscapes suggest the need for a broader perspective and innovative approaches for the use of LiDAR point density in forest biomass estimation.

In this study, we evaluate the effects of LiDAR point density on the estimation of aboveground biomass (AGB) of remnant forests in the rapidly urbanizing region of Charlotte, North Carolina, USA. Using multiple linear regression (MLR), we established relationships between field-based biomass estimates and LiDAR-derived predictor variables (PVs) for a general urban forest type and three specific forest type biomass models, referred to as Urban Forest and Forest Types (coniferous, deciduous, and mixed). For the Urban Forest, we developed models using PVs of LiDAR data at eight point densities, while for the Forest Type models we used the original LiDAR data and the optimal point density. We then compared accuracies of the Urban Forest and Forest Type models in addition to comparison of top-performing biomass models between the Urban Forest and a combined estimate for the Forest Type models. We quantified the degree to which the presence of built development (e.g., buildings, roads, and parking lots) influenced biomass estimation. Finally, we analyzed the effect of canopy stratification on biomass estimation for both the Urban Forest and Forest Type models.

2. Materials and methods

2.1.Study system

This study focuses on the urban remnant forests of Mecklenburg County, North Carolina. The county is located within the Piedmont physiographic province in the center of the Charlotte Metropolitan Area and covers 1,415 km^2 (Fig. 1). The region's topography is characterized by rolling flat lands with elevation ranging from 252 m in the northern part of the county to about 159 m in the south. Forested landscapes in the area are primarily

comprised of oak-hickory-pine forests that have developed on former timber plantation sites as well as through natural regeneration on abandoned farmland. In recent years, urban sprawl, with low- to medium-density housing, has converted forest and farmland dominated landscapes into an array of developed land cover types with highly fragmented and complex urban forests. According to the American Forests report in 2008, Mecklenburg County has experienced a 33% decline in its tree canopy between 1985 and 2008, and will suffer an additional loss of ~3% by the year 2015 under recent trends.

2.2.Field data

We conducted field measurements during the years 2010-2012 as part of the Charlotte ULTRA-EX (Urban Long-Term Research Areas Exploratory) study designed to analyze socio-ecological interactions driving the persistence of private forest. Within each forest site, we established three to five 11.5 m fixed-radius random field plots (400 m^2) (Fig. 1b). We measured the diameter at breast height (*dbh*) of all native and invasive woody plants greater than 5 cm diameter, including vines. Other parameters measured in each plot include geographic coordinates, merchantable height, species name and type (deciduous vs. evergreen), and predominant land cover type. Prior to the analysis, we categorized all field plots into deciduous, coniferous, and mixed forest types using a threshold of 75% predominant land cover type. If a plot consisted of over 75% deciduous or coniferous trees, we labeled the plot as deciduous or coniferous, respectively. We labeled plots comprised of less than 75% deciduous or coniferous forest as mixed plots. To maintain the uniformity in modeling biomass across the forest types, we selected a similar number of field plots in each forest type resulting in a collection of 70 plots comprised of22 deciduous, 23 coniferous, and 25 mixed plots.

2.3.LiDAR data

We used leaf-off multiple return LiDAR data acquired April 11-14, 2012, and obtained from the Storm Water Services Division of Charlotte-Mecklenburg County

government office. Original data acquisition was carried out by Pictometry International (Rochester, USA) using Optech's ALTM Gemini 3100 LiDAR system with an average point spacing of 1 m between any two neighboring points over the study system. See section 2.3. of the chapter 2nd for more details.

2.4.LiDAR data processing and data reduction

We reduced the original LiDAR point-density to 80%, 60%, 40%, 20%, 10%, 5%, and 1% densities (Fig. 2; Table 1) using the 'percentage of the total points' reduction algorithm developed at Boise Center Aerospace Laboratory, Boise, Idaho (BCAL LiDAR Tools 2013). This algorithm reduces LAS files (LiDAR data format) by the percentage of the total points using random selection from the pool of equal height points in each return. We elected this data reduction approach over other approaches, such as point density or point spacing based on two primary considerations: 1) the inconsistency of point spacing and point-density throughout the LiDAR tiles and field-plots, and 2) thinning by percentage offers nearly consistent sampling across the tiles for large scale analysis (Anderson et al. 2006; Magnusson et al. 2007). Finally, we clipped the reduced and original LiDAR datasets at 12 m radius plot size to account for any possible misalignment due to GPS positional errors. This resulted in eight plot-level LiDAR datasets, including seven sets of reduced point densities.

2.5.Processing and extracting variables

2.5.1. Field-based aboveground biomass estimation

To identify an optimal LiDAR point density, we applied a generalized allometric equation (Eq. (1)) (Jenkins et al. 2003) to the tree species found in each field plot to estimate biomass for individual trees using field based *dbh* and parameters from the appropriate species group as documented in Jenkins et al. (2003) and listed in Table 2.

$$bm = Exp(\beta_0 + \beta_1 * \ln(dbh)) \quad (1)$$

where *bm* is total aboveground biomass (kg dry weight) for trees > 5 cm in *dbh*, *Exp* is the exponential function, *dbh* is the diameter at breast height in centimeters (cm), ln is the natural log base e (2.718282), and β_0 and β_1 are parameters for hardwood and softwood tree species groups (Table 2).

We aggregated tree species observed in the field into hardwood and softwood species groups. We used parameters of each respective species group to calculate biomass for individual trees and then averaged the biomass at plot level by applying a tons per hectare conversion unit. We used *dbh* to estimate biomass for three reasons: 1) *dbh* is the most stable predictor for regression models compared to predictors that incorporate LiDAR-derived tree height and crown diameter (Popescu 2007), 2) *dbh* is a more consistent estimation of biomass than direct estimates from LiDAR-derived metrics, and 3) species-independent biomass equations only require *dbh* compared to species-dependent.

2.5.2. LiDAR metrics extraction

We extracted plot-level metrics from LiDAR data using FUSION software developed at Pacific Northwest Research Station, Seattle, WA (McGaughey 2014). We normalized the clipped plot-level LiDAR datasets using 1 m resolution digital elevation model (DEM) derived from LiDAR ground returns. This process removed topography effects from LiDAR points across the datasets for yielding *height-above-ground* values. For extracting plot-level tree metrics, we used the height range 1.5 m to 35 m to exclude understory vegetation, and objects taller than the trees found within the region. We calculated descriptive statistics for normalized LiDAR point densities across the entire vertical profile (e.g., minimum, maximum, variance, interquartile distance, percentiles, etc.) and canopy related metrics (e.g., percentage of first returns above 3 m, percentage of first return above mean, and mode, etc.) (Frazer et al. 2011; Jakubowski et al. 2013). We derived and further identified a total of 37 metrics from previous studies at the plot-level (Table 3). Since the estimation of biomass is

directly related to structural metrics, we did not consider the intensity component of LiDAR in the analysis.

2.5.3. Development density estimation

To examine the effect of development and the distance effect on biomass estimation, we calculated development density (Eq. (2)) at multiple buffer distances of increasing radii around each plot (Table 3).

$$Ds = \frac{\sum_i^N d_i a_i^2}{N a_i^2} \quad (2)$$

where *Ds* is the development density, d_i is the number of developed cells in the circular plot, a_i is the spatial resolution of land cover, and *N* is the total number of cells in the circular plot. We reclassified land cover types, developed from Landsat TM and LiDAR structural data (Singh et al. 2012), into a 5 m raster representing developed (impervious surfaces and managed clearings) and undeveloped land cover (farmland, forest and barren land). We then calculated the total area of developed land cover within the multiple buffers to estimate the development density.

2.5.4. Canopy stratification

Reducing point density could dilute the effect of vertical complexity and impact the accuracy of biomass estimates. Therefore, we analyzed the effect of canopy stratification on the biomass estimations using the Landscape Management System (LMS) (McCarter 2001; McCarter et al. 1998) and the Forest Vegetation Simulator (FVS) model (Dixon 2002; Stage 1973; Wykoff et al. 1982). FVS uses internal model diameter/height relationships to fill in height and crown information that was not measured in the plots. In addition, FVS provides volume estimates for measured trees. We used LMS to calculate canopy layers using a canopy stratification algorithm based on work by Baker and Wilson (2000). The default canopy overlap parameter (-5) imposed a 5 ft. gap between the bottom of one canopy layer

and the top of the next lower canopy layer. For evergreen forests in our datasets, this produced few canopy layers because of a more continuous canopy structure. Thus, we adjusted the overlap parameter to produce results for a 5 ft. gap (-5), no overlap or gap (0), and a 5 ft. overlap (+5), allowing for the detection of multiple canopy layers in the more continuous canopy structure of the sampled stands (Fig. 7). We examined the results for each overlap parameter and selected the +5 ft. overlap parameter since it revealed very distinct multiple layers and complex structures in the data.

2.6.Statistical analysis and model development

We used MLR to evaluate the statistical relationship between field-based biomass and LiDAR-derived PVs (Table 3). First, we developed biomass models for Urban Forest using PVs derived from reduced LiDAR point densities to identify the optimal point density suitable for biomass estimation. Second, to account for biomass variation due to forest types across urbanizing landscapes and to test for differences among forest types, we developed separate regression models for evergreen, deciduous, and mixed forest stands using PVs from the identified optimal point density (i.e., LiDAR percentage that produced matching or improved biomass estimates compared to the original data) (Chen and Hay 2011; Popescu and Wynne 2004; Zhao et al. 2009). Third, we used Urban Forest biomass models developed using PVs derived from 100%, 40% and 10% point densities to compare and analyze the effect of development density on the estimation of biomass. We used Urban Forest models based on from three point densities to insure that the effect of development is not by chance and is point density independent. Finally, to analyze the effect of canopy stratification, we developed regression models where the residuals of the top-performing biomass models are treated as the new response variable against the PVs of interest and canopy stratification. This analysis determined the strength of the relationship between biomass and canopy structural complexity of the urban forests.

We implemented the MLR analyses in three steps: 1) identification of outliers, 2) variable selection followed by data transformation, if needed, and 3) model development and performance evaluation. We started by identifying a list of LiDAR metrics (Table 3) from

previous studies that are the best predictors of plot-level biomass in order to simplify the variable selection process (Dubayah et al. 2010b; Frazer et al. 2011; García et al. 2010; Hall et al. 2005). We applied the variance inflation factor (VIF) and the Pearson correlation coefficients to select optimal PVs for MLR that minimized data dimensionality and the presence of multicollinearity within the PVs, and to overcome issues of over fitting (Allouis et al. 2013). Due to many potential non-collinear PVs, we also applied an automated approach by employing the 'regsubsets' function from the LEAPS package (Lumley and Lumley 2013) in the R statistical software (R Core Team 2013) to view ranked models based on different scoring criteria (*adjusted* R^2, Mallows' Cp statistic, Bayesian information criterion (BIC) etc.). We used the lowest Mallow's Cp score and BIC for selecting the model while Akaike information criterion (AIC) was used in model performance evaluation. We also used logarithmic transformations during the development of the MLR models achieve linearity between field-based biomass and non-linear forest structural parameters (Frazer et al. 2011; Hudak et al. 2006; Lefsky et al. 2002a).

2.7.Evaluating model performance

We compared the performance of MLR models using an *adjusted* R^2, AIC, and root mean squared error (RMSE) based on a 10-fold cross validation (k-fold CV) analysis. The *adjusted* R^2 is the percentage of variation explained by the PVs that influence the response variable. The *adjusted* R^2 corrects upward bias, especially in small samples. We used the AIC to evaluate the performance of biomass models resulting from development density with increasing radii around each plot. We selected k-fold CV method due to the relatively low number of observations contained in each model. In this approach, the dataset is divided into *k* subsets, and at each iteration, one of the *k* subsets is used as testing data and the other *k-1* subset as training data. The variance of the resulting estimate is reduced as *k* is increased. We also used *F*-test to determine if the variance of observed vs. predicted biomass was equal to biomass models derived from PVs of reduced point-densities.

3. Results

3.1.LiDAR point density effects on Urban Forest biomass model performance

We observed similar biomass predictions for Urban Forest across all the reduced LiDAR point densities. The amount of biomass variance (*adjusted* R^2) explained by LiDAR-derived PVs was consistent across the reduced point densities (Fig. 3) with the highest difference of 11.5% observed between the 100% and 1% point density biomass models. The RMSE and k-fold CV values of Urban Forest models showed a similar trend with the maximum difference of 8.2 t/ha and 9.6 t/ha, respectively (Table 4). The biomass estimates from the 40% point density models were similar to the 100% point-density except for the significance of the *PFRame* variable (Table 3). *PFRame* was present in all models below 80% point density with the exception of the 1% model. The biomass variance across the reduced point densities was best explained by the tree height variance (Hvar (>60%)), followed by height percentile ($HP_{(1-99)}$), percentage of first returns above mean (*PFRame*) (if significant in the model), and the median of the absolute deviations from the overall mode (*HMADMo*). *F*-tests (equality of two variances) revealed no significant differences between the predicted and observed biomass estimates (Table 4) with virtually consistent distributions across the models except for slight differences at the 60% and 20% LiDAR point densities (Fig. 4).

3.2.Performance of Forest Type biomass model

While Urban Forest biomass models performed consistently well across the reduced point-densities, due to similar *adjusted* R^2 we selected 100% and 40% point densities for developing Forest Type biomass models. Forest Type biomass models outperformed (individually as well as cumulatively) the Urban Forest models at both the 100% and 40% point densities. Forest Type biomass models produced, on average, 4 t/ha to 6 t/ha lower RMSE values than Urban Forest models. The evergreen biomass model based on the 100% point density produced the highest *adjusted* R^2 and lowest RMSE followed by mixed and deciduous types (Fig. 5; Table 5). We found a moderate but significant difference in

evergreen and mixed types biomass variance (on average 4.56%) between the 100% and 40% point densities (Fig. 5) except for the deciduous type which showed similar *adjusted* R^2 (79%) and RMSE (29 t/ha) values. Biomass estimates of evergreen and mixed models performed similarly as we introduced the basal area ratio variable into the models. The average difference of RMSE values of Forest Type biomass models based on 100% and 40% point densities was 2 t/ha. However, we observed a difference of 6.2 t/ha and 5 t/ha at 100% and 40% densities, respectively, between Urban Forest and the combined estimates of Forest Type models (Table 5).

3.3.Development density effects on biomass model performance

Development density exhibited a negative relationship with biomass, and contributed on average 2.3% to the total explained biomass variance of models using 100%, 40% and 10% point densities. Equivalent contributions of development density suggest that development density is independent of point density and not due to chance. We achieved lower AIC and RMSE values with higher significance levels using development densities between 100 m and 175 m radius buffers, peaking at 125 m with $p<0.01$ significance level for all three reduced point densities (Fig. 6a, 6b and 6c). Table 6 illustrates how biomass variance improved with the addition of development density. The biomass models at 125 m radius produced the highest *adjusted* R^2 (0.83, 0.82, and 0.78 at 100%, 40%, and 10% point-densities, respectively) and the lowest k-fold CV-RMSE values (30.4 t/ha, 32.9 t/ha, and 38.2 t/ha at 100%, 40%, and 10%, respectively). Analysis of variance (ANOVA) tests revealed that biomass models using development densities at 100 m, 125 m, 150 m, and 175 m are significantly different from the models that exclude development density at all three reduced point densities (Table 6).

3.4.Contribution of canopy stratification on biomass estimation

Canopy stratification did not contribute to the Urban Forest biomass model and only marginally improved the unexplained variance of the Forest Type biomass models. On

average, the canopy stratification improved Forest Type biomass models by 1.4 t/ha and 1.2 t/ha at 100% and 40% point densities, respectively; the deciduous model showed the greatest improvement of 2.2 t/ha at both densities (Table 7).

4. Discussion

Accurate and repeated biomass estimation over large urbanizing areas has become a necessity to manage remnant forests and their potential impacts on regional carbon dynamics. This research explored the effects of LiDAR point density and landscape context on the accuracy of biomass estimates in urban remnant forests at a regional scale. We found that reducing point density provided a viable solution to regional scale assessment without compromising accuracy of biomass estimates. Differences in *adjusted* R^2 and RMSE values between 100% and 1% LiDAR point density cases were modest and disproportionate to the reduction in average point spacing and file size (Table 1). Jakubowski et al. (2013) found similar results and suggested that the accuracy of predicted forest structure metrics remains relatively high until low point densities are incorporated in the model. This raises two important methodological questions not addressed in this current study. First, what is the minimum number of points required to sufficiently represent the three-dimensional structure to provide suitable metrics for the estimation of biomass estimation? Second, how to reduce LiDAR data without compromising the quality of structural attributes? In this study, the data reduction algorithm and the structural predictors contributed to the performance of biomass estimation across the reduced point densities.

LiDAR data reduction by the 'percentage of the total points' algorithm was selected over the 'point density' and 'point spacing' algorithms since the latter two methods vary tile to tile and are determined by the complexities found in forest stands. For example, the overall point spacing for the county level LiDAR data was 1 m, while plot-level LiDAR data reduced to 20% density produced the equivalent point spacing. Table 1 shows that LiDAR data at lower point densities decreased the average point spacing disproportionately in comparison to reductions in file size. These findings are consistent with Anderson et al.

(2006) in that the reduction of LiDAR data to a certain extent does not affect the statistical properties of the elevation models. Therefore, LiDAR data can be reduced without significantly impacting the accuracy of biomass estimates. For example, models using 40% LiDAR point density with significant reduction in density, spacing and file size produced comparable biomass estimates models using the full LiDAR dataset (Table 1). Such data reduction can considerably minimize data processing time for large-area assessments without compromising the accuracy of biomass estimates. Point density reductions may lower data costs for large-area assessments, especially in cases where repeated coverage is needed.

The cumulative Forest Type biomass model outperformed the Urban Forest model at both 100% and 40% point densities. The observed variability in model performance across the three forest types at both point densities corroborates the findings of Zolkos et al. (2013). The structural complexities (e.g., canopy stratification, presence of understory vegetation, canopy density, etc.) of forest strata are one potential source of variability in model performance. Other issues such as leaf-on vs. leaf-off LiDAR data also impact model performance. Naesset (2005) reported that variability of the height distribution tends to increase from leaf-on to leaf-off conditions, which may cause variation in biomass estimates. This research utilized leaf-off data, and so this effect could not be evaluated. Likewise, less variation in the heights of conifers could lead to higher accuracies for biomass estimates in coniferous stands compared with hardwood stands (Nelson et al. 2004). This disparity may also depend upon the height to biomass relationship difference between hardwood (more biomass in branches) and conifer species (Nelson et al. 2007). However, differences in the performance of the Forest Type (evergreen and mixed) models using100% and 40% of LiDAR data were unexpected. Adding the 'basal area ratio' variable to these models resulted in similar estimates, suggesting that the reduced point density does not fully explain the complexity found in the lower strata of evergreen trees in urban remnant forests either in evergreen or mixed forest type field plots. This is a significant finding given the prevalence of conifers in the study system making 'basal area ratio' an important variable to effectively compensate for low point densities in the evergreen and mixed forest type biomass models.

Urbanization alters tree growth and understory recruitment along urban-rural gradients (Nitschke and Innes 2008; O'Brien et al. 2012). Evaluation of biomass models using development density at increasing radii revealed development density as an important contributor to biomass estimates and the quantification biomass of urban remnant forests. Our results illustrate improved performance from 100 m to 175 m buffer distance, with no change beyond 175 m (Fig. 6; Table 6). Insignificant contributions of development density to biomass estimation from the edge of the field plot to 100 m may be due to the deliberate distribution of field plots within forest sites with the least amount of developed land cover. The absence of a development density effect beyond175 m could be ascribed to the maximum developed land cover around each plot given the prevalence of developed land cover in this urbanizing region. To verify this significant outcome was not by chance and was independent of point density, we modeled biomass at three different LiDAR point densities (100%, 40%, and 10%) and found the same outcomes. O'Brien et al. (2012) suggested that anthropogenic impacts on urban forests could be multi-faceted; therefore, additional research is required to fully understand which land cover type has more impact on forest biomass estimates along urban- rural gradients.

Evaluation of multiple linear regression models in our analysis illustrated that descriptive statistics metrics (e.g., tree-height-variance (Hvar)) explained more variance than direct canopy height related metrics and were overall key predictors of biomass. The second largest contributor to overall model performance was *HMADmo* (median of the absolute deviations from the overall mode), which is a robust measure of variability within a data sample. This implies that field-measured biomass is primarily related to overall vertical forest structure. Zimble et al. (2003) identified tree-height-variance as an effective metric for classifing various vertical forest structural configurations. Additionally, leaf-off LiDAR has been recommended to improve the estimation accuracy of biophysical properties and prediction in a mixed forest, with further improvements by adding forest stratification (Naesset 2005). However, the contributions of these two variables in overall model performance could be a potential explanation for the relatively poor performance of canopy stratification in our Urban Forest biomass model.

5. Conclusions

Management of large scale urban remnant forests, including the maintenance of ecosystem functioning and services, can be enhanced by the resourceful use of LiDAR remote sensing for improved assessment of regional scale urban forests biomass. This study advances the application of LiDAR data for effective assessment of large-area biomass. Our results suggest that reducing LiDAR point density maintains biomass-modeling performance posed by large-area urban forests biomass assessments. Point density reduction may offer an effective solution for minimizing regional scale data procurement costs, and overcoming computational challenges posed by voluminous LiDAR data and point clouds derived by Structure from Motion technique. Equivalent biomass estimates across multiple point densities suggests the 'percentage of the total points' data reduction algorithm is suitable for large-area applications. Lower individual and cumulative RMSE values of the Forest Type biomass models emphasize the consideration of forest types for reasonable biomass estimates at regional scales. As expected, land cover type impacts the productivity and growth of forests in urbanizing landscapes, and when we added development density to the biomass models, our estimates improved. We also observed that canopy stratification did not improve our biomass estimation for urban forests but did lead to a slight improvement when applied to the deciduous biomass model. This suggests complexity in the vertical strata of deciduous forests affects biomass estimation. Our evaluation of all biomass models revealed tree-height-variance as the most useful biomass predictor, corroborating it's suitability for classifying various vertical forest structure configurations (Zimble et al. 2003). Overall, our findings suggest that reduced density LiDAR data may lower data procurement costs for repeated assessments without sacrificing accuracy, and biomass estimates can be improved by including landscape characteristics (e.g., land cover types, development density, etc.) in the regional analysis.

Acknowledgements

This research was supported by the Garden Club of America (GCA) Zone VI fellowship in urban forestry, the Casey Trees Endowment Fund, and the Association of American Geographers Dissertation Research grants. We are very grateful to the North Carolina Space Grant Consortium for additional financial support. We express sincere thanks to the Storm Water Services Division - Charlotte-Mecklenburg County Government Office, and Engineering and Property Management, Land Development Services, City of Charlotte for providing LiDAR data. We thank students, research staff and doctoral dissertation committee members from the Center for Geospatial Analytics, North Carolina State University, for their comments and feedback on the manuscript.

REFERENCES

Allouis, T., Durrieu, S., Vega, C., & Couteron, P. (2013). Stem Volume and Above-Ground Biomass Estimation of Individual Pine Trees From LiDAR Data: Contribution of Full-Waveform Signals. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 6*, 924-934

Anderson, E.S., Thompson, J.A., Crouse, D.A., & Austin, R.E. (2006). Horizontal resolution and data density effects on remotely sensed LIDAR-based DEM. *Geoderma, 132*, 406-415

Baker, P.J., & Wilson, J.S. (2000). A quantitative technique for the identification of canopy stratification in tropical and temperate forests. *Forest Ecology and Management, 127*, 77-86

BCAL LiDAR Tools (2013). Idaho State University, Department of Geosciences In. Boise, Idaho: Boise Center Aerospace Laboratory (BCAL)

Brostow, G.J., Shotton, J., Fauqueur, J., & Cipolla, R. (2008). Segmentation and Recognition Using Structure from Motion Point Clouds. *Computer Vision - Eccv 2008, Pt I, Proceedings, 5302*, 44-57

Chen, G., & Hay, G.J. (2011). An airborne lidar sampling strategy to model forest canopy height from Quickbird imagery and GEOBIA. *Remote Sensing of Environment, 115*, 1532-1542

Csanyi, N., & Toth, C.K. (2007). Improvement of lidar data accuracy using lidar-specific ground targets. *Photogrammetric Engineering and Remote Sensing, 73*, 385-396

Dixon, G. (2002). *Essential (FVS): A User's Guide to the Forest Vegetation Simulator*. Fort Collins, CO: Forest Management Service Center, US Forest Service

Drake, J.B., Dubayah, R.O., Knox, R.G., Clark, D.B., & Blair, J.B. (2002). Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment, 81*, 378-392

Drake, J.B., Knox, R.G., Dubayah, R.O., Clark, D.B., Condit, R., Blair, J.B., & Hofton, M. (2003). Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: factors affecting the generality of relationships. *Global Ecology and Biogeography, 12*, 147-159

Dubayah, R.O., Sheldon, S.L., Clark, D.B., Hofton, M.A., Blair, J.B., Hurtt, G.C., & Chazdon, R.L. (2010a). Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *Journal of Geophysical Research-Biogeosciences, 115*

Dubayah, R.O., Sheldon, S.L., Clark, D.B., Hofton, M.A., Blair, J.B., Hurtt, G.C., & Chazdon, R.L. (2010b). Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *Journal of Geophysical Research, 115*, G00E09

Estornell, J., Ruiz, L.A., Velazquez-Marti, B., & Fernandez-Sarria, A. (2011). Estimation of shrub biomass by airborne LiDAR data in small forest stands. *Forest Ecology and Management, 262*, 1697-1703

Frazer, G.W., Magnussen, S., Wulder, M.A., & Niemann, K.O. (2011). Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. *Remote Sensing of Environment, 115*, 636-649

Garcia, M., Riano, D., Chuvieco, E., & Danson, F.M. (2010). Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. *Remote Sensing of Environment, 114*, 816-830

García, M., Riaño, D., Chuvieco, E., & Danson, F.M. (2010). Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. *Remote Sensing of Environment, 114*, 816-830

Gobakken, T., & Næsset, E. (2008). Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research, 38*, 1095-1109

Gregg, J.W., Jones, C.G., & Dawson, T.E. (2003). Urbanization effects on tree growth in the vicinity of New York City. *Nature, 424*, 183-187

Groffman, P.M., Pouyat, R.V., Cadenasso, M.L., Zipperer, W.C., Szlavecz, K., Yesilonis, I.D., Band, L.E., & Brush, G.S. (2006). Land use context and natural soil controls on plant community composition and soil nitrogen and carbon dynamics in urban and rural forests. *Forest Ecology and Management, 236*, 177-192

Grove, J.M., Troy, A.R., O'Neil-Dunne, J.P.M., Burch, W.R., Cadenasso, M.L., & Pickett, S.T.A. (2006). Characterization of households and its implications for the vegetation of urban ecosystems. *Ecosystems, 9*, 578-597

Hall, S.A., Burke, I.C., Box, D.O., Kaufmann, M.R., & Stoker, J.M. (2005). Estimating stand structure using discrete-return lidar: an example from low density, fire prone ponderosa pine forests. *Forest Ecology and Management, 208*, 189-209

He, C., Convertino, M., Feng, Z.K., & Zhang, S.Y. (2013). Using LiDAR Data to Measure the 3D Green Biomass of Beijing Urban Forest in China. *Plos One, 8*

Hudak, A.T., Crookston, N.L., Evans, J.S., Falkowski, M.J., Smith, A.M.S., Gessler, P.E., & Morgan, P. (2006). Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral satellite data. *Canadian Journal of Remote Sensing, 32*, 126-138

Imhoff, M.L., Zhang, P., Wolfe, R.E., & Bounoua, L. (2010). Remote sensing of the urban heat island effect across biomes in the continental USA. *Remote Sensing of Environment, 114*, 504-513

Jakubowski, M.K., Guo, Q.H., & Kelly, M. (2013). Tradeoffs between lidar pulse density and forest measurement accuracy. *Remote Sensing of Environment, 130*, 245-253

Jenkins, J.C., Chojnacky, D.C., Heath, L.S., & Birdsey, R.A. (2003). National-scale biomass estimators for United States tree species. *Forest Science, 49*, 12-35

Lawrence, H.W. (1995). Changing forms and persistent values: historical perspectives on the urban forest. *See Bradley, 1995*, 17-40

Lefsky, M.A., Cohen, W.B., Harding, D.J., Parker, G.G., Acker, S.A., & Gower, S.T. (2002a). Lidar remote sensing of above-ground biomass in three biomes. *Global Ecology and Biogeography, 11*, 393-399

Lefsky, M.A., Cohen, W.B., Parker, G.G., & Harding, D.J. (2002b). Lidar remote sensing for ecosystem studies. *BioScience, 52*, 19-30

Lim, K., Hopkinson, C., & Treitz, P. (2008). Examining the effects of sampling point densities on laser canopy height and density metrics. *Forestry Chronicle, 84*, 876-885

Lovell, J.L., Jupp, D.L.B., Newnham, G.J., Coops, N.C., & Culvenor, D.S. (2005). Simulation study for finding optimal lidar acquisition parameters for forest height retrieval. *Forest Ecology and Management, 214*, 398-412

Lumley, T., & Lumley, M.T. (2013). Package 'leaps'. In

Magnani, F., Mencuccini, M., Borghetti, M., Berbigier, P., Berninger, F., Delzon, S., Grelle, A., Hari, P., Jarvis, P.G., Kolari, P., Kowalski, A.S., Lankreijer, H., Law, B.E., Lindroth, A., Loustau, D., Manca, G., Moncrieff, J.B., Rayment, M., Tedeschi, V., Valentini, R., & Grace, J. (2007). The human footprint in the carbon cycle of temperate and boreal forests. *Nature, 447*, 848-850

Magnusson, M., Fransson, J.E.S., & Holmgren, J. (2007). Effects on estimation accuracy of forest variables using different pulse density of laser data. *Forest Science, 53*, 619-626

Mascaro, J., Detto, M., Asner, G.P., & Muller-Landau, H.C. (2011). Evaluating uncertainty in mapping forest carbon with airborne LiDAR. *Remote Sensing of Environment, 115*, 3770-3774

McCarter, J.B. (2001). Landscape management system (lms): background, methods, and computer tools for integrating forest inventory, gis, growth and yield, visualization and analysis for sustaining multiple forest objectives. In (p. 101): University of Washington

McCarter, J.B., Wilson, J.S., Baker, P.J., Moffett, J.L., & Oliver, C.D. (1998). Landscape management through integration of existing tools and emerging technologies. *Journal of Forestry, 96*, 17-23

McGaughey, R.J. (2014). FUSION/LDV: Software for LIDAR data analysis and Visualization. In. Seattle, WA: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station

McHale, M., Burke, I., Lefsky, M., Peper, P., & McPherson, E. (2009). Urban forest biomass estimates: is it important to use allometric relationships developed specifically for urban trees? *Urban Ecosystems, 12*, 95-113

Meentemeyer, R.K., Tang, W.W., Dorning, M.A., Vogler, J.B., Cunniffe, N.J., & Shoemaker, D.A. (2013). FUTURES: Multilevel Simulations of Emerging Urban-Rural Landscape Structure Using a Stochastic Patch-Growing Algorithm. *Annals of the Association of American Geographers, 103*, 785-807

Moran, M.A. (1984). Influence of adjacent land use on understory vegetation of New York forests. *Urban Ecology, 8*, 329-340

Nadelhoffer, K.J., Emmett, B.A., Gundersen, P., Koopmans, C.J., Schleppi, P., Tietema, A., & Wright, R.F. (1999). Nitrogen deposition and carbon sequestration - Reply. *Nature, 400*, 630-630

Naesset, E. (2005). Assessing sensor effects and effects of leaf-off and leaf-on canopy conditions on biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment, 98*, 356-370

Naesset, E. (2009). Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment, 113*, 148-159

Nelson, R., Short, A., & Valenti, M. (2004). Measuring biomass and carbon in Delaware using an airborne profiling LIDAR. *Scandinavian Journal of Forest Research, 19*, 500-511

Nelson, R.F., Hyde, P., Johnson, P., Emessiene, B., Imhoff, M.L., Campbell, R., & Edwards, W. (2007). Investigating RaDAR-LiDAR synergy in a North Carolina pine forest. *Remote Sensing of Environment, 110*, 98-108

Nitschke, C.R., & Innes, J.L. (2008). A tree and climate assessment tool for modelling ecosystem response to climate change. *Ecological Modelling, 210*, 263-277

Nowak, D.J., & Greenfield, E.J. (2012). Tree and impervious cover change in US cities. *Urban Forestry & Urban Greening, 11*, 21-30

O'Brien, A.M., Ettinger, A.K., & HilleRisLambers, J. (2012). Conifer growth and reproduction in urban forest fragments: Predictors of future responses to global change? *Urban Ecosystems, 15*, 879-891

Parker, G.G., & Brown, M.J. (2000). Forest canopy stratification - Is it useful? *American Naturalist, 155*, 473-484

Popescu, S.C. (2007). Estimating biomass of individual pine trees using airborne lidar. *Biomass & Bioenergy, 31*, 646-655

Popescu, S.C., & Wynne, R.H. (2004). Seeing the trees in the forest: Using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogrammetric Engineering & Remote Sensing, 70*, 589-604

Popescu, S.C., Wynne, R.H., & Scrivani, J.A. (2004). Fusion of small-footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, USA. *Forest Science, 50*, 551-565

R Core Team (2013). R: A Language and Environment for Statistical Computing. In. Vienna, Austria: R Foundation for Statistical Computing.

Robertson, M., BenDor, T.K., Lave, R., Riggsbee, A., Ruhl, J.B., & Doyle, M. (2014). Stacking ecosystem services. *Frontiers in Ecology and the Environment, 12*, 186-193

Seidel, D., Albert, K., Fehrmann, L., & Ammer, C. (2012). The potential of terrestrial laser scanning for the estimation of understory biomass in coppice-with-standard systems. *Biomass & Bioenergy, 47*, 20-25

Seto, K.C., Guneralp, B., & Hutyra, L.R. (2012). Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences of the United States of America, 109*, 16083-16088

Singh, K.K., Vogler, J.B., Shoemaker, D.A., & Meentemeyer, R.K. (2012). LiDAR-Landsat data fusion for large-area assessment of urban land cover: Balancing spatial resolution, data volume and mapping accuracy. *ISPRS Journal of Photogrammetry and Remote Sensing, 74*, 110-121

Stage, A. (1973). Prognosis model for stand development. *Research paper/Intermountain forest and range experiment station. USDA (INT-137)*

Swatantran, A., Dubayah, R., Roberts, D., Hofton, M., & Blair, J.B. (2011). Mapping biomass and stress in the Sierra Nevada using lidar and hyperspectral data fusion. *Remote Sensing of Environment, 115*, 2917-2930

Treitz, P., Lim, K., Woods, M., Pitt, D., Nesbitt, D., & Etheridge, D. (2012). LiDAR Sampling Density for Forest Resource Inventories in Ontario, Canada. *Remote Sensing, 4*, 830-848

Wulder, M.A., Bater, C.W., Coops, N.C., Hilker, T., & White, J.C. (2008). The role of LiDAR in sustainable forest management. *The Forestry Chronicle, 84*, 807-826

Wulder, M.A., White, J.C., Nelson, R.F., Næsset, E., Ørka, H.O., Coops, N.C., Hilker, T., Bater, C.W., & Gobakken, T. (2012). Lidar sampling for large-area forest characterization: A review. *Remote Sensing of Environment, 121*, 196-209

Wykoff, W.R., Crookston, N.L., & Stage, A.R. (1982). User's guide to the stand prognosis model

Zhang, H., Qi, Z.F., Ye, X.Y., Cai, Y.B., Ma, W.C., & Chen, M.N. (2013). Analysis of land use/land cover change, population shift, and their effects on spatiotemporal patterns of urban heat islands in metropolitan Shanghai, China. *Applied Geography, 44*, 121-133

Zhao, K., Popescu, S., & Nelson, R. (2009). Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment, 113*, 182-196

Zimble, D.A., Evans, D.L., Carlson, G.C., Parker, R.C., Grado, S.C., & Gerard, P.D. (2003). Characterizing vertical forest structure using small-footprint airborne LiDAR. *Remote Sensing of Environment, 87*, 171-182

Zolkos, S.G., Goetz, S.J., & Dubayah, R. (2013). A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sensing of Environment, 128*, 289-298

Table 1. Distribution of point density and point spacing, and file size at plot level for different LiDAR data reductions.

LiDAR(%)	Point-density (point/m^2)			Point-spacing (m)			File size (MB)
	Min	Max	Average	Min	Max	Average	
100	2.64	13.67	5.77	0.21	0.65	0.43	10.1
80	2.12	11.12	4.57	0.23	0.71	0.48	8.11
60	1.66	8.29	3.46	0.27	0.82	0.56	6.10
40	1.06	5.37	2.30	0.35	1.04	0.71	4.07
20	0.51	2.78	1.16	0.51	1.52	1.03	2.05
10	0.26	1.32	0.58	0.77	2.30	1.52	1.06
5	0.13	0.68	0.29	1.10	3.29	2.25	0.55
1	0.03	0.14	0.06	2.27	8.90	5.48	0.14

% Percentage of original LiDAR data
MB = megabyte

Table 2. Parameters used for estimating aboveground biomass for all hardwood and softwood species found in the study system. Developed by Jenkins et al. (2003).

	Species group	Parameters		R^2
		β_0	β_1	
Hardwood	Soft maple/birch	–1.9123	2.3651	0.958
	Mixed hardwood	–2.4800	2.4835	0.980
	Hard maple/oak/hickory/beech	–2.0127	2.4342	0.988
Softwood	Cedar/larch	–2.0336	2.2592	0.981
	Pine	–2.5356	2.4349	0.987

Table 3. Predictor variables derived from each LiDAR point density, and used in multiple regression models for estimating plot level aboveground biomass.

Variable	Variable Description
TRC	Total return count
H_{min}	Height minimum
H_{max}	Height maximum
H_{mean}	Height mean
H_{mode}	Height mode
H_{SD}	Height standard deviation
H_{var}	Height variance
H_{cv}	Height coefficient of variation
H_{ske}	Height skewness
H_{kur}	Height kurtosis
H_{AAD}	Height average absolute deviation
H_{MAD}	Height median absolute deviation
HMADMe	Median of the absolute deviations from the overall median
HMADMo	Median of the absolute deviations from the overall mode
H_{IQR}	Height interquartile range
$HP_{(1-99)}$	Height percentile: 1^{th}, 5^{th}, 10^{th}, 20^{th}, 25^{th}, 30^{th}, 40^{th}, 50^{th} (median), 60^{th}, 70^{th}, 75^{th}, 80^{th}, 90^{th}, 95^{th}, and 99^{th}
CRR	Canopy relief ratio
PFRame	Percentage first returns above mean
PFRamo	Percentage first returns above mode
FRaMe	First returns above mean
FRaMo	First returns above mode
ARaMe	All returns above mean
ARaMo	All returns above mode
DevDen†	Development density radius: 50 m, 75 m, 100 m , 125 m , 150 m, 175 m, 200 m, and 500 m
BAratio‡	Ratio between coniferous and hardwood basal area in the plot
C_{STRATA}	Canopy stratification

† Variable derived from 2007 land cover estimated using LiDAR structural and Landsat TM data at 5 m spatial resolution (Singh et al., 2012).

‡ Variable derived from field-based *dbh*

Table 4. Multiple linear regression models for predicting aboveground biomass using predictor variables derived from original LiDAR data and point density reductions (by percentage).

LiDAR(%)	Model with significant variables	R^2	R^2 *(adj)*	RMSE (t/ha)	Xval† (t/ha)	*F*-test‡
100	$y = 44.39 + 0.31_{Hvar} - 3.72_{HMADmo} + 3.46_{HP05}$	0.8203	0.8114	31.99	34.02	($F_{1, 63} = 1.21$, $p = 0.43$, $CI = 0.74/1.99$)
80	$y = 49.26 + 0.31_{Hvar} - 3.65_{HMADmo} + 3.28_{HP05}$	0.8086	0.7992	33.02	35.34	($F_{1, 63} = 1.24$, $p = 0.39$, $CI = 0.75/2.03$)
60	$y = 5.40 + 0.28_{Hvar} - 2.58_{HMADmo}** + 3.12_{HP05} + 0.62_{PFRame}^{\times}$	0.8076	0.7948	33.10	36.93	($F_{1, 63} = 1.24$, $p = 0.39$, $CI = 0.75/2.03$)
40	$y = 11.13 + 0.28_{Hvar} - 3.36_{HMADmo} + 2.67_{HP05} + 0.75_{PFRame}*$	0.8210	0.8090	31.94	34.06	($F_{1, 63} = 1.22$, $p = 0.43$, $CI = 0.74/1.99$)
20	$y = 26.16 + 0.28_{Hvar} - 3.44_{HMADmo} + 2.70_{HP05} + 0.51_{PFRame}^{\times}$	0.7872	0.7730	34.82	37.58	($F_{1, 63} = 1.27$, $p = 0.34$, $CI = 0.77/2.08$)
10	$y= 19.57 + 0.27_{Hvar} - 3.64_{HMADmo}* + 2.84_{HP05} + 0.62_{PFRame}^{\times}$	0.7747	0.7597	35.82	38.87	($F_{1, 63} = 1.29$, $p = 0.30$, $CI = 0.79/2.12$)
5	$y= 34.12 + 0.28_{Hvar} - 3.47_{HMADmo}* + 2.30_{HP05} + 0.56_{PFRame}^{\times}$	0.7624	0.7466	36.79	39.35	($F_{1, 63} = 1.31$, $p = 0.28$, $CI = 0.80/2.15$)
1	$y= -119.15 + 1.27_{TRC}* + 2.56_{Hmin} + 0.17_{Hvar} + 259.90_{CRR}$	0.7164	0.6975	40.19	43.63	($F_{1, 63} = 1.40$, $p = 0.18$, $CI = 0.85/2.29$)

% Percentage of original LiDAR data

†10-fold cross validation

‡Between observed and predicted aboveground biomass

Level of significance: 0.001 '**' 0.01 '*' 0.05 '×' 0.1

Table 5. Multiple linear regression models for predicting aboveground biomass for different forest types using original LiDAR point density (100%) and 40% point density reduction.

LiDAR(%)		Coefficients	R^2	R^2 Adjusted	RMSE (t/ha)
100	Evergreen	$y = -65.62 + 0.03_{TRC} + 27.08_{Hmin} + 0.32_{Hvar} -4.58_{HMADmo}$	0.8681	0.8371	23.37
	Deciduous	$y = 178.85 + 0.30_{Hvar} -73.78In_{(HMADmo)} + 3.38_{HP05}$	0.8267	0.7942	29.13
	Mixed	$y = 121.89 -79.20In_{(TRC)} + 0.15_{Hvar} -15.30_{HP01} + 4.76_{HP05}$	0.8490	0.8135	24.79
40	Evergreen	$y = -239.07 + 0.03In_{(TRC)} + 0.25_{Hvar} -2.67_{HMADmo} + 8.21_{EP01}$	0.8362	0.7977	26.05
	Deciduous	$y = 210.21 + 0.31_{Hvar} -83.92In_{(HMADmo)} + 2.98_{HP05}$	0.8257	0.7930	29.22
	Mixed	$y = 73.12 -0.04_{TRC} + 0.25_{Hvar} -8.48_{HP01} + 6.54_{HP05}$	0.8056	0.7598	28.13
40†	Evergreen†	$y = -370.27 + 52.7In_{(TRC)} + 0.25_{Hvar} -2.69_{HMADmo} + 29.53_{Hmin} -0.79_{BAratio}$ ×	0.8599	0.8161	24.09
	Mixed†	$y = 129.12 -0.05_{TRC} + 0.23_{Hvar} -8.90_{HP01} + 5.94_{HP05} -89.23_{BAratio}$ *	0.8555	0.8104	24.25

% Percentage of original LiDAR data

† Model with BAratio variable

Level of significance: 0.01 '*' 0.05 '×' 0.1

Table 6. Contribution of development density, derived at eight different radii, in the regression models for predicting aboveground biomass using predictor variables derived at 100% and 40% LiDAR point densities.

LiDAR(%)	Radius (m)	Coefficients	R^2	R^2 Adjusted	AIC	RMSE (t/ha)	Xval† (t/ha)	*ANOVA*‡
	00	$y = 44.39 + 0.31_{Hvar} - 3.72_{HMADmo} + 3.46_{HP05}$	0.8203	0.8114	644.99	31.99	34.02	
	50	$y = 42.51 + 0.31_{Hvar} - 3.72_{HMADmo} + 3.49_{HP05} + 15.73_{DevDen}$ ±	0.8207	0.8087	646.85	31.96	34.69	
	75	$y = 48.71 + 0.31_{Hvar} - 3.69_{HMADmo} + 3.42_{HP05} - 33.79_{DevDen}$ ±	0.8218	0.8100	646.43	31.86	35.52	
	100	$y = 54.60 + 0.30_{Hvar} - 3.51_{HMADmo} + 3.49_{HP05} - 99.43_{DevDen}$ *	0.8336	0.8226	641.97	30.78	33.10	$p = 0.03$
100	125	$y = 55.97 + 0.30_{Hvar} - 3.39_{HMADmo} + 3.34_{HP05} - 112.66_{DevDen}$ **	0.8380	0.8272	640.26	30.38	30.39	$p = 0.01$
	150	$y = 55.91 + 0.30_{Hvar} - 3.35_{HMADmo} + 3.51_{HP05} - 106.57_{DevDen}$ *	0.8370	0.8261	640.65	30.47	33.76	$p = 0.01$
	175	$y = 53.67 + 0.30_{Hvar} - 3.45_{HMADmo} + 3.50_{HP05} - 84.11_{DevDen}$ *	0.8316	0.8204	642.77	30.97	33.67	$p = 0.05$
	200	$y = 54.23 + 0.30_{Hvar} - 3.49_{HMADmo} + 3.45_{HP05} - 72.48_{DevDen}$ ±	0.8295	0.8181	643.58	31.16	34.93	
	500	$y = 52.55 + 0.31_{Hvar} - 3.66_{HMADmo} + 3.48_{HP05} - 43.33_{DevDen}$ ±	0.8248	0.8132	645.32	31.59	37.76	
	00	$y = 55.34 + 0.29_{Hvar} - 3.71_{HMADmo} + 3.12_{HP05}$	0.8049	0.7953	650.33	33.34	34.53	
	50	$y = 56.32 + 0.29_{Hvar} - 3.71_{HMADmo} + 3.10_{HP05} - 8.70_{DevDen}$ ±	0.8050	0.7920	652.29	33.33	37.56	
	75	$y = 62.52 + 0.29_{Hvar} - 3.67_{HMADmo} + 3.06_{HP05} - 60.89_{DevDen}$ ±	0.8100	0.7974	650.59	32.89	35.56	
	100	$y = 65.96 + 0.28_{Hvar} - 3.42_{HMADmo} + 3.24_{HP05} - 123.91_{DevDen}$ *	0.8254	0.8137	645.12	31.53	33.83	$p = 0.01$
40	125	$y = 66.79 + 0.28_{Hvar} - 3.29_{HMADmo} + 3.30_{HP05} - 134.85_{DevDen}$ **	0.8297	0.8183	643.50	31.15	32.87	$p = 0.001$
	150	$y = 66.41 + 0.28_{Hvar} - 3.24_{HMADmo} + 3.27_{HP05} - 125.45_{DevDen}$ **	0.8275	0.8160	644.33	31.35	34.53	$p = 0.001$
	175	$y = 64.03 + 0.29_{Hvar} - 3.34_{HMADmo} + 3.22_{HP05} - 96.49_{DevDen}$ *	0.8194	0.8074	647.31	32.07	35.16	$p = 0.03$
	200	$y = 65.15 + 0.29_{Hvar} - 3.41_{HMADmo} + 3.17_{HP05} - 85.38_{DevDen}$ *	0.8175	0.8053	647.99	32.24	36.48	
	500	$y = 62.83 + 0.29_{Hvar} - 3.57_{HMADmo} + 3.19_{HP05} - 47.89_{DevDen}$ ±	0.8103	0.7977	650.50	32.87	35.84	

% Percentage of original LiDAR data
± Insignificant in the model
†10-fold cross validation
‡Between models with and without development density
Level of significance: 0.001 '**' 0.01 '*' 0.05 '×' 0.1

Table 6. (continued).

LiDAR(%)	Radius (m)	Coefficients	R^2	R^2 Adjusted	AIC	RMSE (t/ha)	Xval† (t/ha)	*ANOVA*‡
	00	$y = 43.03 + 0.25_{Hvar} - 1.54_{HMADmo} + 3.13_{HP05}$	0.7598	0.7480	663.84	36.99	41.67	
	50	$y = 42.47 + 0.25_{Hvar} - 1.54_{HMADmo} + 3.13_{HP05} + 4.84_{DevDen}^{\pm}$	0.7598	0.7438	665.83	36.98	40.84	
	75	$y = 50.03 + 0.24_{Hvar} - 1.49_{HMADmo} + 3.11_{HP05} - 61.92_{DevDen}^{\pm}$	0.7651	0.7494	664.40	36.58	41.47	
	100	$y = 56.47 + 0.24_{Hvar} - 1.38_{HMADmo} + 3.24_{HP05} - 138.43_{DevDen}$**	0.7857	0.7714	658.42	34.93	40.50	$p = 0.01$
10	125	$y = 58.09 + 0.24_{Hvar} - 1.29_{HMADmo} + 3.26_{HP05} - 148.82_{DevDen}$**	0.7909	0.7769	656.83	34.51	38.23	$p = 0.001$
	150	$y = 58.02 + 0.24_{Hvar} - 1.26_{HMADmo} + 3.23_{HP05} - 141.19_{DevDen}$**	0.7895	0.7754	657.27	34.63	39.30	$p = 0.001$
	175	$y = 55.19 + 0.24_{Hvar} - 1.33_{HMADmo} + 3.17_{HP05} - 109.58_{DevDen}$*	0.7792	0.7645	660.37	35.46	39.95	$p = 0.03$
	200	$y = 56.15 + 0.24_{Hvar} - 1.38_{HMADmo} + 3.11_{HP05} - 95.61_{DevDen}$*	0.7760	0.7611	661.30	35.72	42.89	
	500	$y = 50.89 + 0.25_{Hvar} - 1.52_{HMADmo} + 3.10_{HP05} - 38.70_{DevDen}^{\pm}$	0.7634	0.7477	664.85	36.71	42.13	

% Percentage of original LiDAR data
± Insignificant in the model
†10-fold cross validation
‡Between models with and without development density
Level of significance: 0.001 '**' 0.01 '*' 0.05 '^' 0.1

Table 7. Contribution of canopy stratification in predicting aboveground biomass for Urban Forest, and Forest Type using predictor variables derived at 100% and 40% LiDAR point densities.

LiDAR[(%)]	Urban Forest		Forest Type		
	Biomass model	Residuals ~ C_{STRATA}		Biomass model	Residuals ~ C_{STRATA}
	RMSE (t/ha)			RMSE (t/ha)	
100	31.99	31.91	Evergreen	23.37	22.32
			Deciduous	29.13	26.97
			Mixed	24.79	23.71
40	31.94	31.78	Evergreen†	26.05	25.07
			Deciduous	29.22	27.00
			Mixed†	28.13	27.82

% Percentage of original LiDAR data

† Model without BAratio (coniferous and hardwood basal area ratio) variable

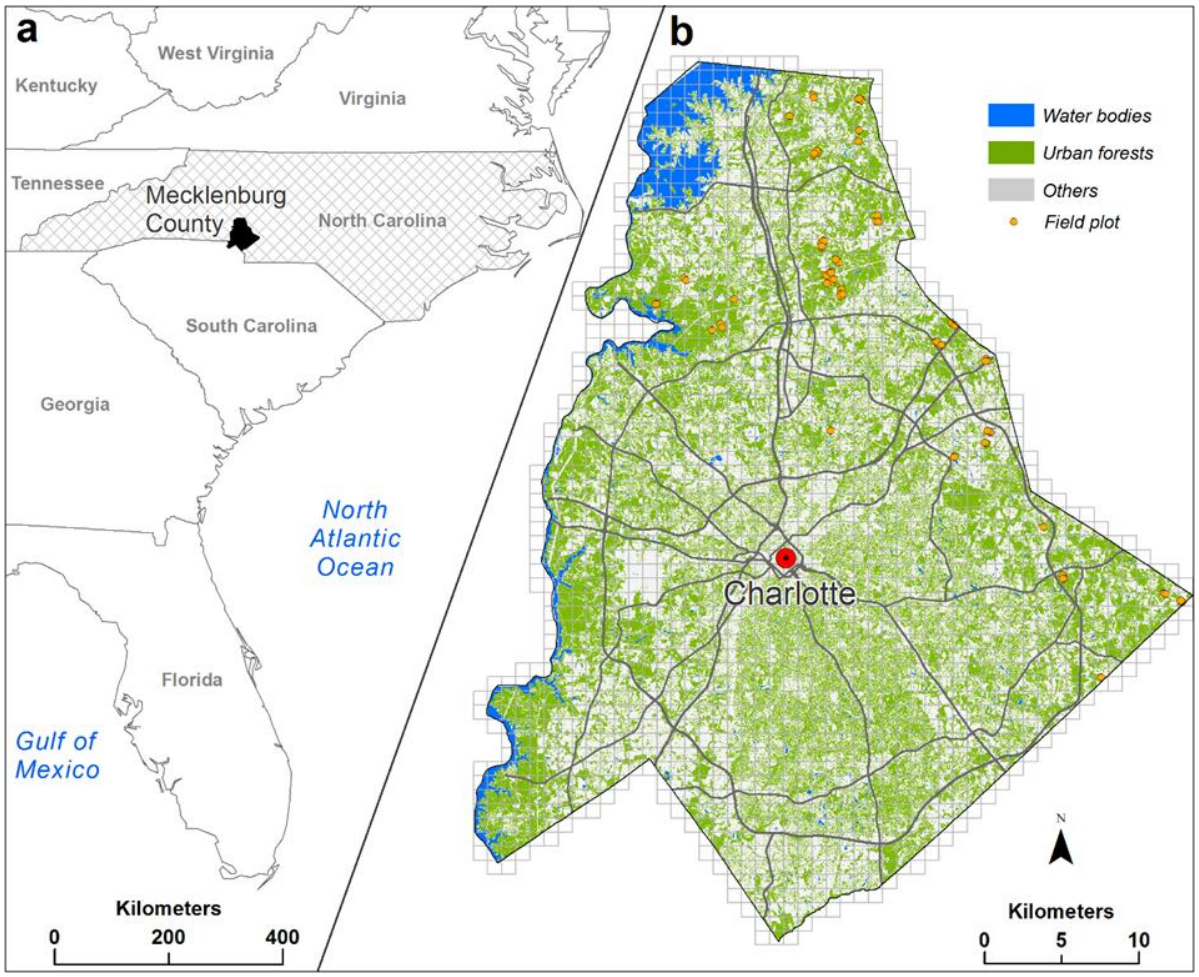


Fig. 1. Study system. (a) Mecklenburg County in the center of Charlotte Metropolitan Region of North Carolina, USA, and (b) the distribution of forest cover across the county with an overlay of LiDAR tiles and locations of field plots.

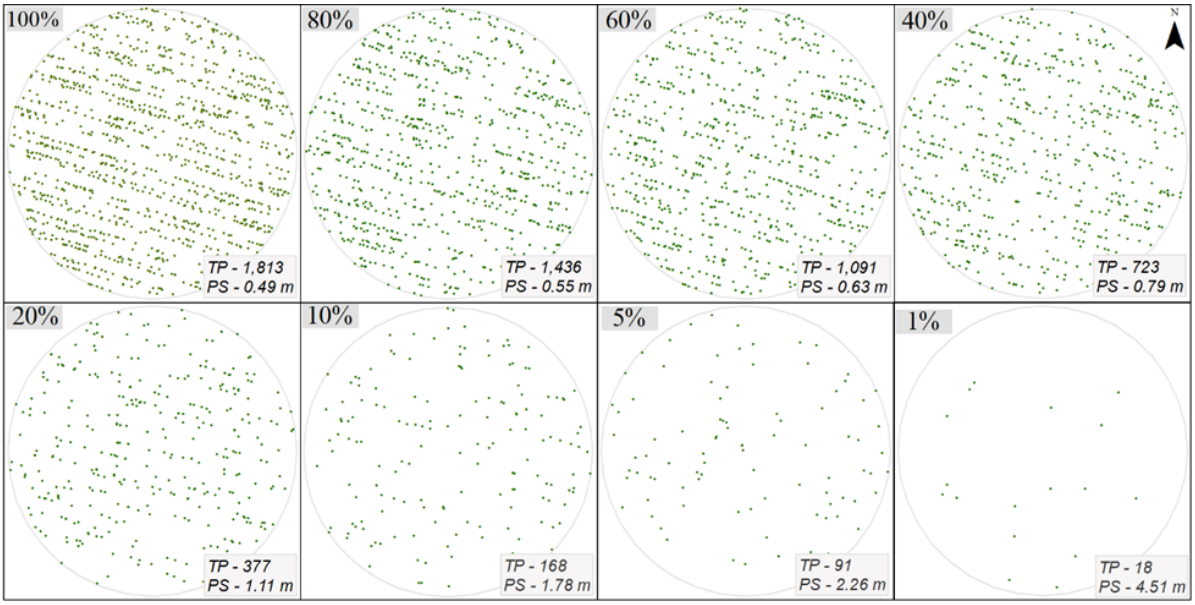


Fig. 2. An illustration of percentage based LiDAR data reduction. The total number of LiDAR points (TP) and average point spacing (PS) are shown for each point density reduction at plot-level (400 m^2).

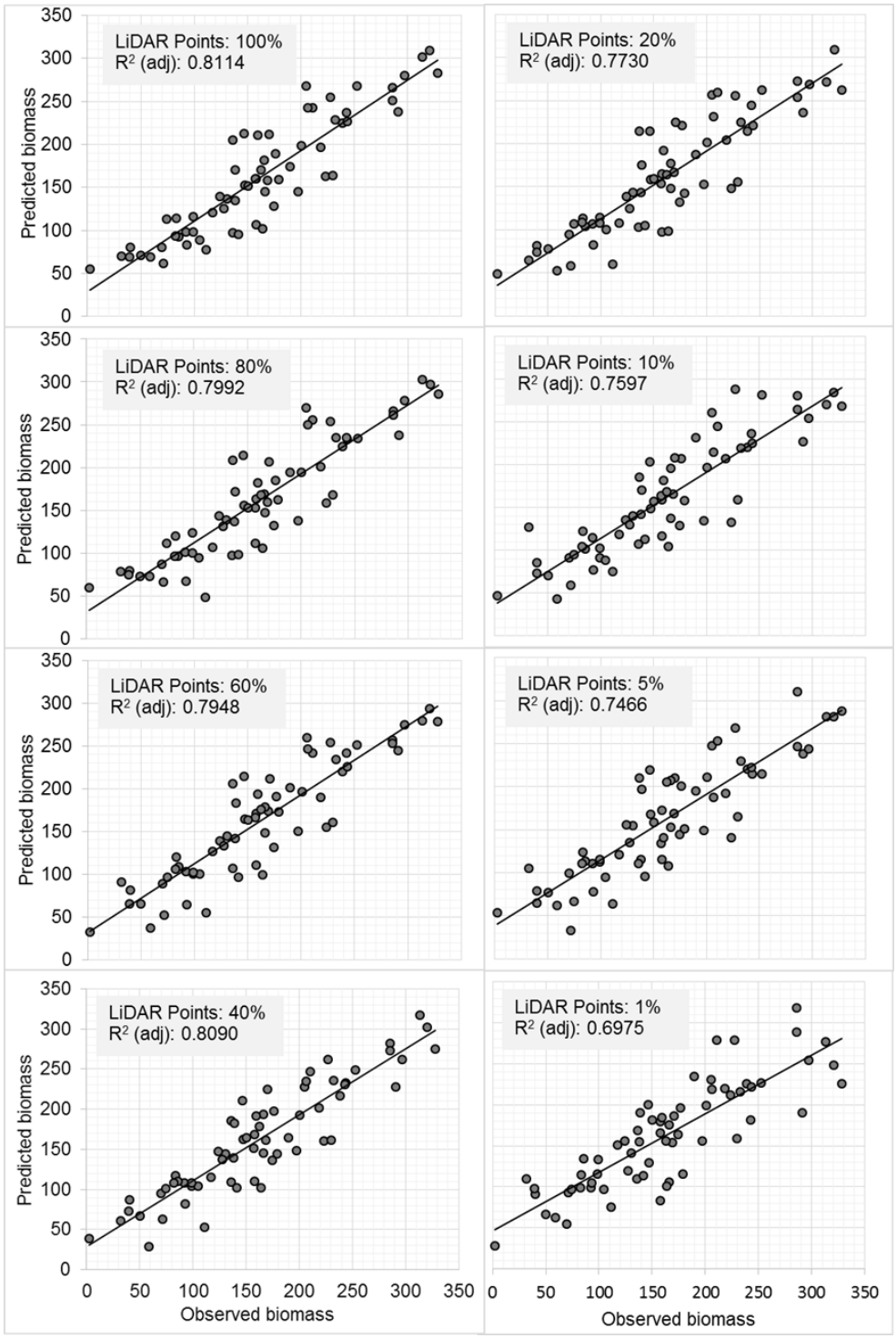


Fig. 3. Predicted versus observed aboveground biomass at each LiDAR point density reduction.

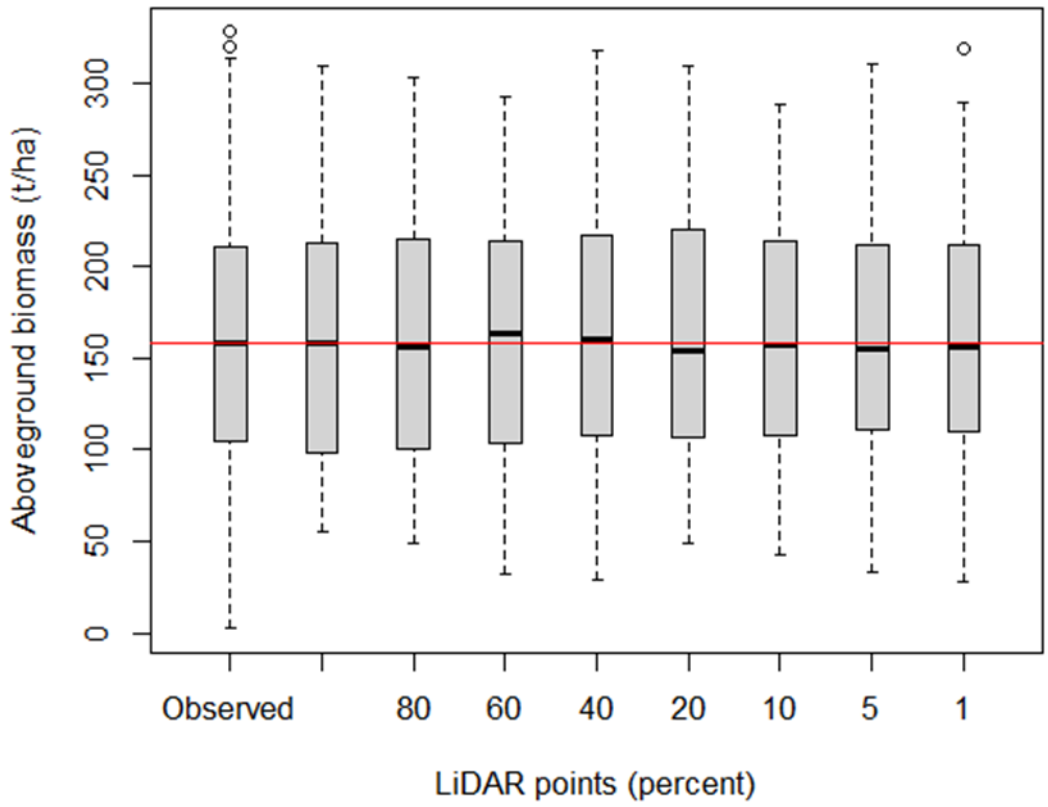


Fig. 4. Predicted aboveground biomass categorized by LiDAR point density reductions with red horizontal line indicating the median of observed biomass.

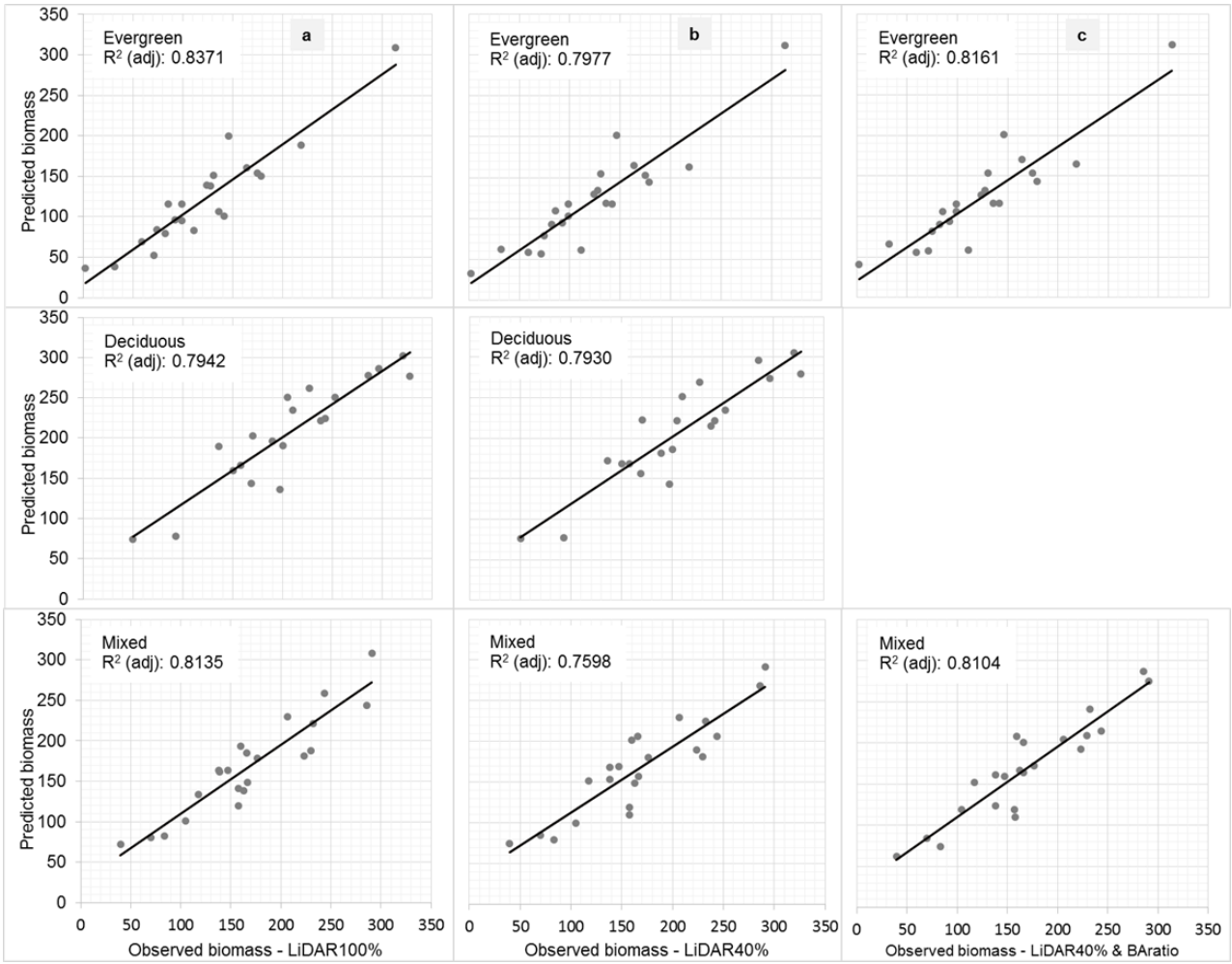


Fig. 5. Predicted versus observed aboveground biomass by Forest Type: (a) model based on 100% point density, (b) model based on 40% point density, and (c) effect of basal area ratio (BAratio) on the evergreen and mixed forest type biomass models.

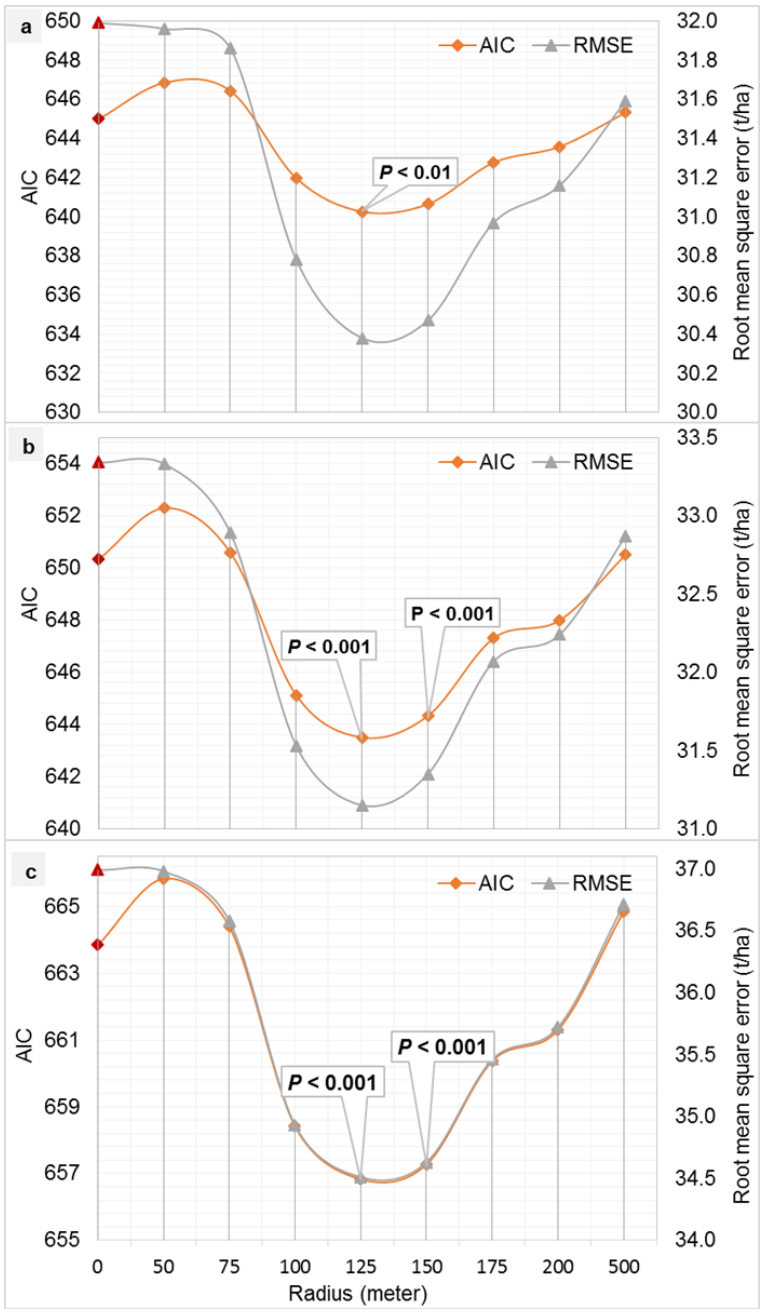


Fig. 6. Change in AIC (Akaike Information Criterion) and RMSE values with increasing radii for development density. Contribution of development density in regression models for predicting aboveground biomass using predictor variables derived at the point density reductions of (a) 100%, (b) 40%, and (c) 10%.

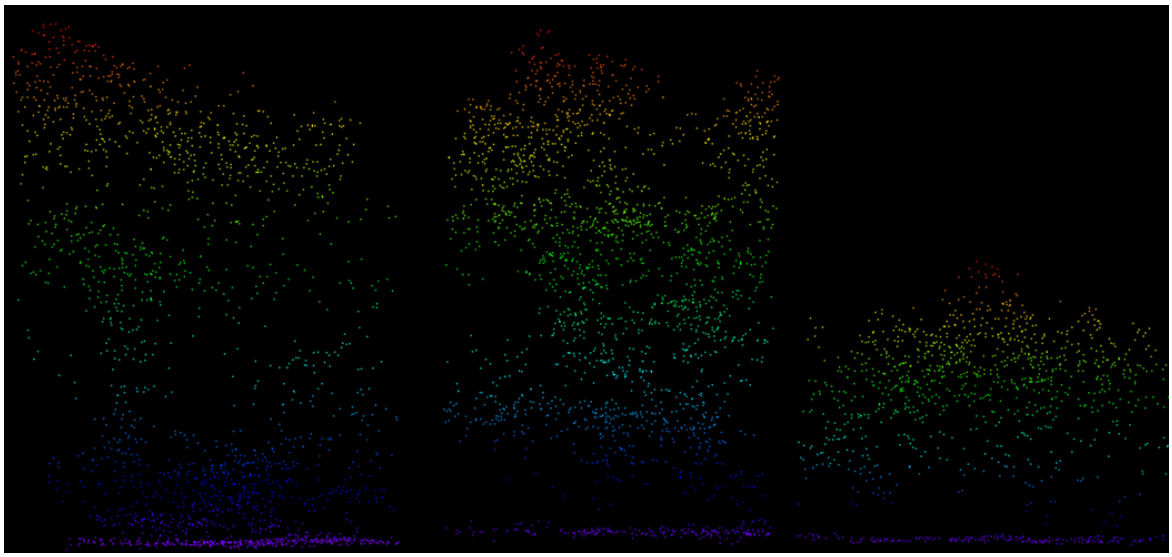


Fig. 7. Canopy stratification

CONCLUSIONS

LiDAR remote sensing is a promising source of structural data with great potential to map, measure, and model the spatial heterogeneity of urban landscapes at both fine and regional scales. LiDAR applications are generally limited to small area due to prohibitive cost and the computational difficulties posed by large data volumes. In this collection of work, I investigated the use of LiDAR data for mapping, measuring, and modeling spatial heterogeneity of urban landscapes at regional scales. In the first study, the fusion of LiDAR digital surface models with Landsat TM imagery was performed to improve accuracy of mapping urban landscape heterogeneity over large regional extents. The second study addressed the use of LiDAR for detecting and mapping the spatial distribution of understory invasive plants in urbanizing forest landscapes. In the final study, I evaluated the effects of LiDAR point density and landscape context on the estimation of aboveground biomass of remnant forests in urbanizing landscapes. The results suggest that leveraging high resolution LiDAR data 'alone' and through 'fusion' with spectral remote sensing improves assessments of emerging complexities – heterogeneity along urban-rural gradients, understory invasion, and forest biomass – in urban landscapes over large regions.

The digital surface models derived from an optimal resolution of LiDAR data, when fused with moderate-resolution multispectral imagery, not only strikes a balance between computational efficiency and classification accuracy for mapping land cover over large heterogeneous regions, but also provides improved discrimination of spectrally similar land cover types. There is also evidence that the need-based modification of land-use and land-cover classification schema may considerably improve overall mapping accuracy.

The leaf-off LiDAR data 'alone' is suitable for accurately detecting and mapping the spatial distribution of the understory invasive plant, *Ligustrum sinense*, over large urbanizing regions. Findings highlight the implication of the structural and spectral characteristics of LiDAR for quantifying landscape-level topography, forest stand characteristics, and physical plant attributes for successful detection and mapping of such understory invasive plants. The selection of appropriate classification algorithms also contributes to overcoming data

complexities, including multidimensionality and nonlinearity in multisource data, and provides a generalized approach for rapid assessments over large urbanizing regions.

Moderate to low point density LiDAR data does not affect derived structural configuration of tree stands. Therefore, reducing point density provides an effective solution for minimizing large-area data procurement costs and overcoming computational challenges while maintaining biomass estimate accuracy. Disaggregating urban forests into different forest types also allows improvement in accuracy estimates of biomass at regional scales. The addition of landscape context in the LiDAR-based biomass model results in improved models with higher biomass estimates compared to models using LiDAR alone.

These observations led to the conclusion that leveraging high resolution LiDAR data 'alone' and through 'fusion' may help improve our understanding of emerging complexities in urban landscapes. The goal of this work was to advance LiDAR analytics for accurate and detailed estimation of urban landscape heterogeneity over large regional extents. Collectively, these studies suggest that establishing optimal resolution/and point-density for LiDAR data is a highly effective method of pursuing large-area studies of urban landscape heterogeneity, and the fusion of LiDAR-derived variables with multispectral data is beneficial in some applications, such as improving class discrimination of spectrally similar land cover types. Finally, the direct measurement of forest understory and overstory structure using LiDAR has proven valuable for the study of complex and heterogeneous ecosystems like urban forests.

APPENDICES

Appendix

Schematics used to create predictor variables

1. Elevation relief ratio

The elevation relief ratio (ERR) index is a measure of the extent to which topography has been opened up by natural and anthropogenic activities. ERR provides hypsometric information about a watershed.

$ERR = (Z_{MEAN} - Z_{MIN}) / (Z_{MAX} - Z_{MIN})$ (Pike and Wilson 1971)

I used ArcGIS based 'Geomorphometry and Gradient Metrics' tools (Evans 2014) to develop ERR index using 'rectangle' analysis window and '3 x 3' cells neighborhood settings.

2. Curvature

The curvature (Crv) is useful for identifying areas of rapid change and the flow across a surface. I used the *Curvature tool* from the *Spatial Analyst tools* of ArcGIS to create a 'curvature raster'. Since the horizontal and vertical units were in the same units of measure (NC FIPS 3200, NAD 1983, m), I used the default z-factor which is '1'.

A positive curvature value indicates that the surface is upwardly convex at that particular cell, and a negative value indicates that the surface is upwardly concave while 0 value means the surface is flat.

3. Hillshade

A hillshade raster is used to show illumination ratio of surfaces. It is created by applying an illumination to the elevation raster source at a user-specified azimuth and altitude. I used *Hillshade tool* from the *Spatial Analyst tools* of ArcGIS to create a hillshade raster. For this, I used the default z-factor of 1 since the horizontal and vertical values are all

in the same unit of measure (meters) and considered local illumination angle by disabling the *model shadows* option in the tool. Integer value ranges from 0 to 255 in a hillshade raster.

4. Solar radiation index

Topography of a landscape play a key role in determining the amount of solar energy incident at a location on the Earth's surface. Variability in elevation, slope, slope orientation, and shadowing, can create strong local gradients in solar radiation that directly and indirectly affect ecological and environmental systems including energy and water balances and primary production. I calculated the direct solar irradiance for the study area from the solar elevation angle, solar azimuth angle, slope, and aspect of the slope using the formula described in Dubayah and Rich (1995).

$$Solar\ Radiation = [(\cos \Theta_O * \cos S) + (\sin \Theta_O * \sin S * \cos (\Phi_O - A))] * So \exp (-To/\cos \Theta_O)$$

Where,
Θ_O = solar elevation angle; S = slope; Φ_O = solar azimuth; A = aspect of slope; So = standard exoatmospheric flux (= 1353w/m^2), and To = optical depth.

Output values ranged from completely shadowed (0 and negative values) to completely lit (positive values). The units are w/m^2. I used the Model Builder of ArcGIS to create a SRI raster. To calculate the direct solar irradiance, I estimated the solar elevation angle and solar azimuth using the NOAA Solar Position Calculator. I used the Spring Equinox (March 21st, 2012) at 10:00AM because of LiDAR data were collected during the spring of 2012. The structure of equation and model (Fig. 1) looks like this:

*Filename (SRI) = ((Cos (elevation div deg) * Cos (slope_map div deg)) + (Sin (elevation div deg) * Sin (slope_map div deg) * Cos (azimuth div deg – aspect_map div deg)) * (1353 * exp(-optical depth div Cos (elevation div deg)))*

5. Topographic moisture index

The topographic moisture index (TMI) is a wetness index. I created TMI by taking the natural logarithm of the ratio of the upslope contributing area (a) and the local slope ($tan\beta$).

$TMI = ln(a/tan\ B)$

where, a = Upslope contributing area in m^2, and β = Slope raster.

I used the Model Builder in order to automate the entire process (Fig. 2). It involved creating a flow direction raster from a filled DEM, flow accumulation raster using flow direction raster, and a slope raster. Finally, I used the above equation in Raster Calculator to create TMI.

6. Topographic roughness index

The terrain ruggedness index (TRI) expresses the amount of elevation difference between adjacent cells of a DEM. It is calculated by estimating the difference between the elevation value of a cell and the mean of the eight surrounding cells. I used ArcGIS to create TRI in three steps. First, I used the Focal Statistics tool of ArcGIS to create minimum neighborhood raster (input = DEM, neighborhood = rectangle, size = 3×3, units = cells, *Statistics Type = minimum*). Second, I created maximum neighborhood raster following the same procedure (input = DEM, neighborhood = rectangle, size = 3×3, units = cells, *Statistics Type = maximum*). Finally, I derived TRI by taking the square root of the difference between squared value of these two rasters in Raster Calculator followed by reclassifying resulted values into seven categories as described in Riley et al. (1999) using the Reclassify tool.

REFERENCES

Dubayah, R., & Rich, P.M. (1995). Topographic Solar-Radiation Models for Gis. *International Journal of Geographical Information Systems, 9*, 405-419

Pike, R.J., & Wilson, S.E. (1971). Elevation-relief ratio, hypsometric integral, and geomorphic area-altitude analysis. *Geological Society of America Bulletin, 82*, 1079-1084

Riley, S.J., Degloria, S.D., & Elliot, R. (1999). A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences, 5*, 23-27

Evans JS, Oakleaf J, Cushman SA, Theobald D (in prep) A Toolbox for Surface Gradient Modeling. Jeffrey S. Evans spatial ecology website. Available: http://evansmurphy.wix.com/evansspatial. Accessed: 2013 May 3rd.

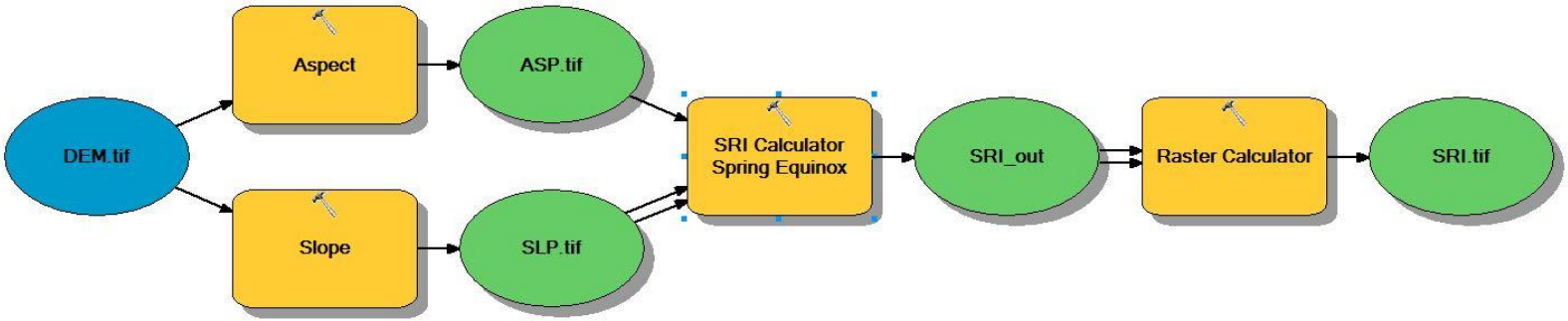


Fig. 1. Schematic for creating the solar radiation index.

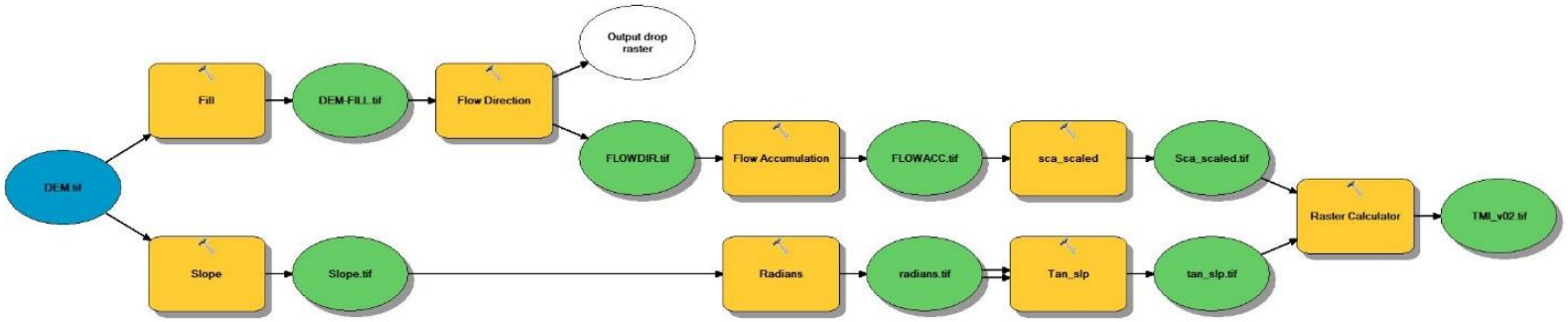


Fig. 2. Schematic for creating the topographic moisture index.