UNIVERSITY OF CALIFORNIA

Santa Barbara

Urban forest ecosystem analysis using fused airborne hyperspectral and lidar data

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy

in Geography

by

Michael Gerard Alonzo

Dissertation Committee:

Professor Dar A. Roberts, Chair
Professor Joseph P. McFadden
Professor Bodo Bookhagen, University of Potsdam, Institute of Earth and Environmental Science
Dr. David J. Nowak, USDA Forest Service, Northern Research Station

September 2015

The dissertation of Michael Gerard Alonzo is approved

_____

Dar A. Roberts

_____

Joseph P. McFadden

_____

Bodo Bookhagen

_____

David J. Nowak

September 2015

Urban forest ecosystem analysis using fused airborne hyperspectral and lidar data

by

Michael Gerard Alonzo

Acknowledgments

My work on this dissertation relied on the kind and consistent support of a large contingent of colleagues, friends, and family. First, I'd like to thank my PhD advisor, Dar Roberts for knowing that I wanted to be (and was capable of being) a remote sensor, even when I didn't know myself. For my 5 years in the VIPER Lab, Dar has been a model for me in terms of how to conduct science and how to communicate with other scientists. Further, the effort he puts towards creating opportunities (funding and otherwise) for his students is unsurpassed. Any success I've found thus far has been through the doors that Dar cracked open.

I was also very lucky to have Professors Joe McFadden and Bodo Bookhagen rounding out my UCSB dissertation committee. Somehow, each time I encountered an obstacle in my project, one of them or the other was ideally suited, and willing, to help chart a path forward. Specifically, I'd like to thank Joe for acting as what amounted to a co-advisor with respect to the urban ecology aspects of this research. Bodo offered me invaluable support especially with regards to working with lidar data and through our discussions of how to communicate impactful science. I look forward to many future collaborations with Dar, Bodo, and Joe.

Special thanks to my external committee member, Dave Nowak of the USDA Forest Service Northern Research Station. First, he has been a great help in terms of answering i-Tree Eco questions and opening up the algorithms so I could poke and prod at them. Second, his collaborations with my previous employer, Casey Trees, in Washington, DC, planted the seeds of my interest in urban forestry research and the early ideas for this dissertation.

Thanks to all members (past and present) of the VIPER Lab (including Dylan Parenti!) for providing technical/research support but, more importantly, for a being cool people who made showing up to work fun. Specifically, I owe a shout-out to Keely Roth and Seth Peterson. Seth is the lab's bedrock who, lucky for us, frequently seems to enjoy helping with our projects as much as working on his own. Keely has always been the person I could turn to when I need to hash out ideas and problem-solving approaches. She is an excellent teacher who never made me feel like my dumb questions were dumb. Thanks also to my cohort members Grant and Sarah for being good friends and also for keeping a Santa Barbara roof over my head even after I moved to L.A.

My parents, Kate and Jerry, deserve infinite thanks for stoking my geographical interests and for their unwavering support throughout the (many) years of Geography schooling. In my confidence that they'd always "be there", I felt more comfortable following the twists, turns, and risks leading me to (and through) UCSB. Finally, thanks to my wife, Claire. None of this would have been possible without her. She gamely uprooted from Washington, DC to move to the otherworldly Santa Barbara. She has supported me intellectually, emotionally and, let's be honest here, financially through the grad school process. I am extremely lucky to have someone who makes me smile and laugh while challenging me and making me think.

CV: Michael G. Alonzo

http://geog.ucsb.edu/~malonzo

mike.alonzo@geog.ucsb.edu 1309 Montecito Dr.
805-883-8821 Los Angeles, CA 90031

EDUCATION

Ph.D. University of California, Santa Barbara, Department of Geography (August 2015)
Certificate in College and University Teaching (January 2015)

M.A. University of Denver, Department of Geography (2007)
Thesis: Rural vs. Rural: An examination of the disparities in access to primary care among Colorado's non-metropolitan communities

B.A. Middlebury College, Department of Geography (2001)
Study abroad: Universidad de La Serena, Chile (July 1999 – February 2000)

PUBLICATIONS

Peer reviewed (undergraduate mentee co-author underlined)

Roberts, D.A., **Alonzo, M**., Weatherly, E., Dennison, P., & Dudley, K. (in press) Multiscale analysis of urban areas using mixing models. Invited book chapter for *Integrating Scale in Remote Sensing and GIS.* Quattrochi, D., Wentz, L., Lam, N., & Emerson, J. (Eds.)

Roth, K.L., Roberts, D.A., Dennison, P.E., **Alonzo, M.,** Peterson, S.H., Beland, M. (in press). Evaluating strategies for discriminating plant species across diverse ecosystems with imaging spectroscopy. *Remote Sensing of Environment*.

Alonzo, M., Bookhagen, B., McFadden, J.P., Sun, A., & Roberts, D.A. (2015) Mapping urban forest leaf area index with airborne lidar using penetration metrics and allometry. *Remote Sensing of Environment*, 162, 141-153.

Alonzo, M., Bookhagen, B., & Roberts, D. A. (2014). Urban tree species mapping using hyperspectral and lidar data fusion. *Remote Sensing of Environment*, 148, 70–83.

Alonzo, M., Roth, K., Roberts, D. (2013). Identifying Santa Barbara's urban tree species from AVIRIS imagery using canonical discriminant analysis. *Remote Sensing Letters*, 4(5), pp. 513-521.

Manuscripts in preparation

Alonzo, M., Van Den Hoek, J., & Ahmed, N. Quantifying forest disturbance from mine waste disposal using remote sensing under inaccessible and very cloudy conditions. Target journal: *PNAS*.

Alonzo, M., McFadden, J.P., Nowak, D.J., & Roberts, D.A. Spatially explicit modeling of urban ecosystem structure and function. Target journal: *Urban Forest & Urban Greening*.

Jenerette, G.D., Weller Clarke, L., Avolio, M.L., Pataki, D.E., Gillespie, T.W., Pincetl, S., Nowak, D.J., L.R., McHale, M., McFadden, J.P., & **Alonzo, M.** Climate filters and trait choices shape urban tree biodiversity: Beyond biome matching and homogenization. Target journal: *Ecology Letters*

Other publications

Howard, H. & **Alonzo, M**. (2010) Improving watershed health and air quality in Washington, DC. *Mapping Forestry.* P. Eredics (Ed.). Redlands, CA: Esri Press.

Alonzo, M. (2009) Casey Trees Study: Understanding the Residential Energy Benefits of Trees in the District of Columbia. *City Trees*. May/June 2009

FELLOWSHIPS AND AWARDS

Research fellowships and grants

NASA Postdoctoral Program Fellowship (2015 – 2017)
EPA-STAR Fellowship (2014) -- $84,000.
UCSB Graduate Division Dissertation Fellowship (2014) -- $12,647
Casey Trees Fellow, Garden Club of America Zone VI Fellowship in Urban Forestry (2014) -- $4,000
Robert N. Colwell Memorial Fellowship (2014) - $6,000 (American Society for Photogrammetry and Remote Sensing Foundation)
Imagery Grant, GeoEye Foundation (2011)

Awards

The Jack Estes Memorial Award ($1,000 for excellence in remote sensing research)
Outstanding Student Paper Award (Biogeosciences) – AGU 2014
Special Achievement in GIS, ESRI Users Conference (2010)

Travel scholarships

Graduate Student Association Travel Grant -- $200 (AGU 2014 in San Francisco, CA)
UCSB Graduate Senate Doctoral Student Travel Grant -- $1030 (ForestSat 2014 in Riva del Garda, Italy)
Dangermond Travel Scholarship - $1200 (ICUC9 in Toulouse, France)
Dangermond Travel Scholarship - $500 (AAG 2015 in Chicago, IL)
Dangermond Travel Scholarship - $600 (AAG 2014 in Tampa, FL)
Dangermond Travel Scholarship - $500 (AGU 2013 in San Francisco, CA)
Dangermond Travel Scholarship - $1200 (Riegl Lidar 2013 in Vienna, Austria)
Dangermond Travel Scholarship - $500 (AAG 2013 in Los Angeles, CA)
Dangermond Travel Scholarship - $580 (AAG 2012 in New York City, NY)

TEACHING EXPERIENCE

Instructor at UC Santa Barbara

Introduction to Geographic Information Systems (2014)
Introduction to Geographic Information Systems (2013)

Teaching Assistant at UC Santa Barbara

Oceans and Atmosphere (2014)
Measuring our Environment (2011)
Introduction to Remote Sensing (2011)
Maps and Spatial Reasoning (2010)

Teaching Assistant at University of Denver

The Human Population (2004)
Cities and Civilizations (2003)
Introduction to GIS (2003)
People, Places, and Landscapes (2002, 2003, 2004)

UNDERGRADUATE MENTORSHIP

Mike Zuanich and Alex Sun: Calculation of plot-level LAI using high-dynamic range NIR hemispherical photography. Novel approach for foliage binarization using segmentation in eCognition (2013).

Seth Gorelik: Trained as lead field data collector for an Urban Forest Effects (i-Tree Eco) modelplot-based urban forest inventory. Identified trees, measured crown structure, took NIR hemispherical photos. Led other field data collection interns (2012).

Keri Opalk and Alex Sun: Field crew for i-Tree Eco data collection (2012)

Bryan Wong: Atmospheric correction of AVIRIS imagery using MODTRAN5 and ATCOR-4 (2011)

SERVICE

AAG Remote Sensing Specialty Group Awards ad-hoc committee member (2015)
UCSB-ASPRS Workshop co-organizer / instructor: How to answer science questions with airborne lidar (2015)
UCSB-ASPRS Workshop co-organizer: Python for Scientific Computing and Image Analysis (2013)
UCSB-ASPRS Workshop co-organizer: Terrestrial lidar data acquisition, processing, and analysis (2013)
President UCSB Student chapter of the American Society for Photogrammetry and Remote Sensing (2012-13)

Conduct peer-review for:
Remote Sensing of Environment
Remote Sensing
Journal of Selected Topics in Applied Earth Observations and Remote Sensing
Urban Forestry and Urban Greening
Sensors
ISPRS Journal of Photogrammetry and Remote Sensing

PRESENTATIONS AND INVITED TALKS

*Presenting author unless noted with **

2015 Complete maps of urban forest structure and function using hyperspectral imagery and lidar data (AAG, Chicago, IL)
2014 Mapping Urban Forest Leaf Area Index Using Lidar: A Comparison of Gap Fraction Inversion and Allometric Methods (AGU, San Francisco, CA)
2014 Multi-sensor remote sensing of the extent and persistence of the 2005 Amazon drought (AGU, San San Francisco, CA)*
2014 Mapping Urban Forest Leaf Area Index with high scan angle lidar in discontinuous canopy (ForestSat2014,Riva del Garda, Italy)
2014 Mapping urban Leaf Area Index (LAI) using high point density lidar (AAG, Tampa, FL)
2013 Urban forest species mapping with imaging spectroscopy and lidar (AGU, San Francisco, CA)
2013 Urban forest inventory with imaging spectroscopy and lidar (Riegl UC, Vienna, Austria)
2013 Urban forest inventory with imaging spectroscopy and lidar (AAG, Los Angeles, CA)
2012 Crown scale fusion of imaging spectroscopy and lidar for urban tree species identification **(Invited)** (ForestSat2012, Corvallis, OR)
2012 Crown scale fusion of imaging spectroscopy and lidar for urban tree species identification (IGARSS, Munich, Germany)
2012 Discriminating Plant Species Across Diverse Ecosystems With Imaging Spectroscopy (IGARSS, Munich, Germany)*
2012 Mapping and quantifying the urban forest with imaging spectroscopy and lidar (ASPRS, Sacramento, CA)
2012 Mapping and quantifying the urban forest with imaging spectroscopy and lidar (AAG, New York, NY)
2010 Integrating ArcGIS Server and Flex: Highlighting Trees in Washington, DC (ESRI UC, San Diego, CA)
2010 The Benefits of Urban Trees (Center for Neighborhood Technology, Chicago, IL) **(Invited)**
2007 Incorporating GIS into Community Tree Inventories (Partners in Community Forestry, Baltimore, MD)
2007 Implementing the Urban Forest Effects Model in Washington, DC (Rutgers, NJ) **(Invited)**
2004 Physician location and retention in rural Colorado (Assoc. of American Geographers, Philadelphia, PA)

PROFESSIONAL EXPERIENCE

2013 Targeted Victory, Alexandria, VA – **GIS Consultant**
2007 – 2010 Casey Trees, Washington, DC – **GIS Specialist**
2004 – 2007 Epic Systems Corporation, Madison, WI – Application Manager
2003 – 2004 GVR Metropolitan District, Denver, CO – **GIS Consultant**

Summer 2002 FV Bligh Reef and Carmen Rose, Cordova and Bristol Bay, AK – Fisherman
2001 – 2002 Pictometry International, Rochester, NY – Aerial Data Acquisition
Summer 2001 Middlebury College Geography Department, Middlebury, VT – **GIS Intern**

PROFESSIONAL AFFILIATIONS

Association of American Geographers (2011 – present)
American Geophysical Union (2013 – present)
American Society for Photogrammetry and Remote Sensing (2011 – present)

Abstract

Urban forest ecosystem analysis using fused airborne hyperspectral and lidar data

by Michael G Alonzo

Urban trees are strategically important in a city's effort to mitigate their carbon footprint, heat island effects, air pollution, and stormwater runoff. Currently, the most common method for quantifying urban forest structure and ecosystem function is through field plot sampling. However, taking intensive structural measurements on private properties throughout a city is difficult, and the outputs from sample inventories are not spatially explicit. The overarching goal of this dissertation is to develop methods for mapping urban forest structure and function using fused hyperspectral imagery and waveform lidar data at the individual tree crown scale.

Urban forest ecosystem services estimated using the USDA Forest Service's i-Tree Eco (formerly UFORE) model are based largely on tree species and leaf area index (LAI). Accordingly, tree species were mapped in my Santa Barbara, California study area for 29 species comprising >80% of canopy. Crown-scale discriminant analysis methods were introduced for fusing Airborne Visible Infrared Imaging Spectrometry (AVIRIS) data with a suite of lidar structural metrics (e.g., tree height, crown porosity) to maximize classification accuracy in a complex environment. AVIRIS imagery was critical to achieving an overall species-level accuracy of 83.4% while lidar data was most useful for improving the discrimination of small and morphologically unique species. LAI was estimated at both the field-plot scale using laser penetration metrics and at the crown scale using allometry. Agreement of the former with photographic

estimates of gap fraction and the latter with allometric estimates based on field measurements was examined. Results indicate that lidar may be used reasonably to measure LAI in an urban environment lacking in continuous canopy and characterized by high species diversity. Finally, urban ecosystem services such as carbon storage and building energy-use modification were analyzed through combination of aforementioned methods and the i-Tree Eco modeling framework. The remote sensing methods developed in this dissertation will allow researchers to more precisely constrain urban ecosystem spatial analyses and equip cities to better manage their urban forest resource.

Table of Contents

Chapter 1

Introduction

1.1 Motivation and summary of chapters

More than 50% of all humans live in cities (UN-DESA, 2014). Urban areas, while representing less than 3% of the earth's land area, have a disproportionate impact on global climate change (Grimm et al., 2008). Conversely, cities are susceptible to the impacts of climate change such as storm surges associated with sea level rise and more intense heat waves, which can also exacerbate already poor air quality (Grimm et al., 2008; Nicholls & Cazenave, 2010). Unmitigated impacts will, given the densely settled nature of cities, be magnified and potentially lead to negative human health outcomes, property damage, and generally decreased livability (Georgi & Dimitriou, 2010; Kalkstein & Greene, 1997). The drivers of undesirable environmental conditions in cities can be linked to current urban morphology which is dominated by dark, manmade materials that reduce albedo, urban canyons that trap longwave radiation at night, and impervious surfaces that intensify stormwater runoff (Grimm et al., 2008; Oke, 1982). In response, many cities are working to augment and optimally arrange *green* infrastructure to lower temperatures, reduce air pollution, dampen and delay runoff from precipitation events, and even sequester carbon (McPherson et al., 2011; Myint et al., 2010).

Urban trees in particular are strategically important in a city's effort to both reduce their carbon footprint and improve mitigation of heat island effects, air pollution, and stormwater runoff (Myint et al., 2010; Nowak et al., 2014). Trees sequester carbon and, through shading and evapotranspiration, reduce power plant emissions associated with mechanized cooling (Akbari, 2002; McPherson & Simpson, 1999b). Lowered urban temperatures also improve air quality with further gains brought about by contaminant deposition on and absorption by leaves (Nowak et al., 2006a). Finally, peak flow of stormwater runoff is delayed due to tree canopy interception of precipitation (Xiao *et al.,* 2000). This increases the likelihood of properly functioning gray infrastructure and reduces physical damage from flooding and contamination of downstream water bodies.

Currently, the most common method for quantifying urban forest structure and ecosystem function is through implementation of the USDA Forest Service's i-Tree Eco (formerly Urban Forest Effects or UFORE) model (www.itreetools.org; Nowak et al., 2008a). The i-Tree Eco model requires intensive tree and ground measurements from 200+ sample plots as inputs. Key model outputs leveraged by cities in their planning processes include estimates of stem counts by species, air pollution removal, carbon sequestration, and building energy-use modification. However, field sampling methods come with potential drawbacks: (1) they are labor and time intensive, generally taking a trained, 2-person field crew ~14 weeks to sample about 200 one-tenth acre plots (Nowak et al., 2008b); (2) they can be expensive, depending upon local costs; (3) permission to access plots may be limited in some areas; (4) the magnitudes of fundamental outputs such as species and tree size distributions are subject to sampling

error that is dependent upon sample size and forest variability; (5) the results are averages or totals within classes and not spatially explicit across the city.

The goal of the research presented in this dissertation was to create a citywide, spatially explicit urban forest inventory through fusion of hyperspectral imagery with high point-density lidar data. To achieve this goal, I completed three projects leading to actionable maps of urban forest structure and function in my study area of Santa Barbara, California. First, I classified each tree in Santa Barbara's urban forest to either the species or leaf type level. Second, I developed methods for mapping urban leaf area index (LAI) using allometry and laser penetration metrics. Third, I produced spatially explicit models of the urban forest's potential for air pollution removal and carbon storage. The uncertainties associated with remotely estimated urban forest structure and function were also evaluated and compared to current, field-based methods.

In **Chapter 2,** I assess the extent to which species can be identified using fused 3.7 m resolution Airborne Visible/Infrared Imaging Spectrometer (Green et al., 1998) imagery and 22 pulses m^{-2} waveform lidar data. Here, I focus on discriminating 29 species representing approximately 80% of Santa Barbara's canopy cover at the individual tree crown scale (hereafter "crown scale"). From crowns that were isolated using watershed segmentation, I developed methods to extract 28 lidar structural metrics (e.g., tree height, crown porosity) directly from the 3-D point cloud. These measurements were concatenated to spectral information from all available crown pixels to form the input dataset to a canonical discriminant analysis (CDA) classification algorithm. Beyond developing a method for accurately mapping a high number of

species, this portion of the research yielded an improved understanding of the biochemical and biophysical determinants of species separability.

Leaf area index (LAI) plays a key role in mediating urban forest ecosystem functions such as air pollution removal and urban cooling. In **Chapter 3,** I explore the utility of lidar data at both the crown and field-plot scales for estimating LAI. For plot-scale estimates, validated against hemispherical photography, I developed lidar penetration metrics (LPM) essentially using the passage of lidar returns through canopy to ground as a proxy for gap fraction. Crown scale estimates were based on structural measurements (tree height, crown base height, crown diameter) available from the segmented lidar dataset and allometric scaling. These estimates were validated against LAI outputs using i-Tree Eco allometric equations and field-measured inputs. Additional contributions of this chapter include an improved understanding of the impact of leaf angle distribution on LPMs (particularly in diverse forests) and methods for maintaining accuracy under conditions of discontinuous canopy cover and high scan-angle lidar data.

In **Chapter 4,** I apply the species identification and LAI estimation methods developed in Chapters 2 and 3 to create wall-to-wall maps at the crown scale of each quantity. The species map was further validated against a City of Santa Barbara street tree dataset and broad spatial patterns were examined. Classification uncertainties in the species map were compared against sampling uncertainties that are a byproduct of field plot methods. The LAI map was used to demonstrate potential for better constraining estimates of building energy-use reduction. Finally, dry-weight carbon

stock was estimated from lidar structural measurements using three scaling equations, one for each leaf type present in the study area: broadleaf, needleleaf, and palm.

1.2 The urban forest

Defined broadly, the urban forest is said to include all ecosystem elements within a city or an urbanized area. According the UDSA Forest Service[1]:

> Urban forests broadly include urban parks, street trees, landscaped boulevards, public gardens, river and coastal promenades, greenways, river corridors, wetlands, nature preserves, natural areas, shelter belts of trees and working trees at industrial brownfield sites.

Effectively, an urban forest is not delineated in terms of a natural ecotone but with respect to management jurisdiction. While perhaps odd from an ecological perspective, matching the urban forest edge to the extents of a political boundary allow for good alignment of forest variables with existing socioeconomic and demographic data collection units such as U.S. Census tracts (Sanders, 1984). In this dissertation, the urban forest is more narrowly considered and includes only trees but all trees regardless of specific planting location. That is, a tree is part of an the urban forest whether it is part of a "natural" stand, introduced to a landscaped lawn, or amidst a sea of pavement in a supermarket parking lot **(Fig. 1.1)**.

[1] USDA Forest Service Urban and Community Forestry: http://www.fs.fed.us/ucf/program.shtml

Figure 1.1: Examples of urban forest scenes in the Santa Barbara area. All trees within a management unit generally comprise the urban forest.

Urban forest structure refers to plant characteristics such as species, species mix, age/size classes, leaf area, biomass, and health status (Sanders, 1984). This structure, with respect to tree species, crown dimensions, leaf area, and biomass, is highly variable across the United States and results from the interaction of natural (climate, soils) and anthropogenic (development, management practices) forces (Jenerette *et al., in prep*). For instance, Washington, DC is a temperate city located within a natural forest matrix. The set of species comprising more than 50% of its canopy are all native to the region (e.g., *Fagus grandifolia*, *Liriodendron tulipifera*; www.itreetools.org/resources/reports). By contrast, Santa Barbara is situated in a Mediterranean climate characterized by lower natural canopy cover. The canopy dominant in Santa Barbara is the native, *Quercus agrifolia* but many of the other common urban trees are introduced and frequently irrigated. Given the intensity of planning and development in urban areas, most urban forest attributes are strongly influenced by humans. The following are examples of common controls on urban forest structure:

- Historical planting patterns and management priorities (Dwyer et al., 2000)
- Utility management including planting small stature species under wires, utility pruning, and hazardous tree removal (Goodfellow, 1995).

- Land use intensity and land value (i.e., fewer trees downtown, Hutyra et al., 2011)
- Matching tree species with appropriate planting sites considering the planting space size/shape, water retention, traffic flow, shade, acceptability of litterfall, infrastructure conflicts, etc. (Manning, 2008; Urban, 1992)
- Ordinances prohibiting the removal of large/heritage trees (Schroeder et al., 2003)
- Community and personal preference (Fraser & Kenney, 2000)

1.3 Urban forest ecosystem services and disservices

Urban trees provide a broad array of ecosystem services that are governed by tree species, canopy structure, and locational context (Escobedo & Nowak, 2009; Manning, 2008; McCarthy & Pataki, 2010; McPherson et al., 2011; Simpson, 2002; Urban, 1992). Commonly quantified services include air pollution removal, carbon storage and sequestration, cooling, windbreak, and stormwater runoff reduction. Additional relationships have been established between the presence of tree canopy and increase in home value (Sander et al., 2010), improved health outcomes (Tzoulas et al., 2007; Ulrich, 1984), reduced crime (Kuo & Sullivan, 2001), and increased retail activity (Wolf, 2005). While the second set of more socially-oriented functions are important, there is, when compared to the first set, less direct physical basis for relating canopy amounts to the magnitude of the service. In this dissertation, given my choice to focus on the i-Tree Eco model, I will only examine the ecophysiological functions relating to CO_2, heat, water, and momentum fluxes.

Tree canopy can at least temporarily remove the pollutants O_3,NO_2,SO_2, CO, and PM-10[2] from the air via dry depositional uptake and intercellular suspension (Baldocchi et al., 1987; Hirabayashi et al., 2011). The magnitude of the removal is governed by leaf area, pollution concentrations, and meteorological conditions (Hirabayashi et al., 2011; Nowak et al., 2006a). Increased canopy leaf area, especially over paved surfaces, delays stormwater peak flow through interception of precipitation (Xiao & McPherson, 2002). Trees fix atmospheric CO_2 as woody biomass through the process of photosynthesis. Net carbon sequestration by the urban forest is a function current carbon storage, the carbon sequestration rate (i.e., the average growth rate), and the carbon loss from mortality and decay (Nowak et al., 2013a). Trees can lower urban surface and air temperatures through direct shading and evapotranspiration (Simpson, 2002). This effect has been shown on a broad spatial scale as a function of increased fractional cover and LAI and has further been shown to reduce summertime building cooling costs (Akbari, 2002; Georgi & Zafiriadis, 2006; Lu & Weng, 2006; Oke, 1989).

There is a relatively limited amount of literature associated with ecosystem disservices because, as Lyytimäki and Sipilä (2009) mentions, it's easier to spark enthusiasm among "bystanders" (the public) with talk of ecosystem "goods" rather than "bads". Lyytimäki and Sipilä (2009) notes some commonplace disservices from a Scandinavian perspective. They include: invasive species, nuisance species, pollen, and other health risks. Lyytimäki et al. (2008) note that many disservices are person specific and may result from differing viewpoints on what constitutes the ideal urban landscape. For example, families might view the ideal park as one that is big and open to facilitate ball playing. Yet, the authors note that compared to a recovering brownfield (perhaps on the neglected outskirts of town) this park is "barren green" while the brownfield may be home to significant biodiversity. Additionally, tree cover has been

[2] Refers to all particulate matter 10 microns or less.

linked to disservices ranging from pollen allergies and blocked views, to sidewalk damage and the production of litterfall (Roy et al., 2012).

1.4 Santa Barbara's urban forest

My study area is the land area of downtown Santa Barbara centered at 34° 25' 4"N and 119° 41' 48"W and stretching 6.8 km in the east-west direction and 5 km in the north-south direction **(Fig. 1.2)**. Santa Barbara is situated on a coastal plain, located between the Pacific Ocean to the south and Santa Ynez Range to the north. It benefits

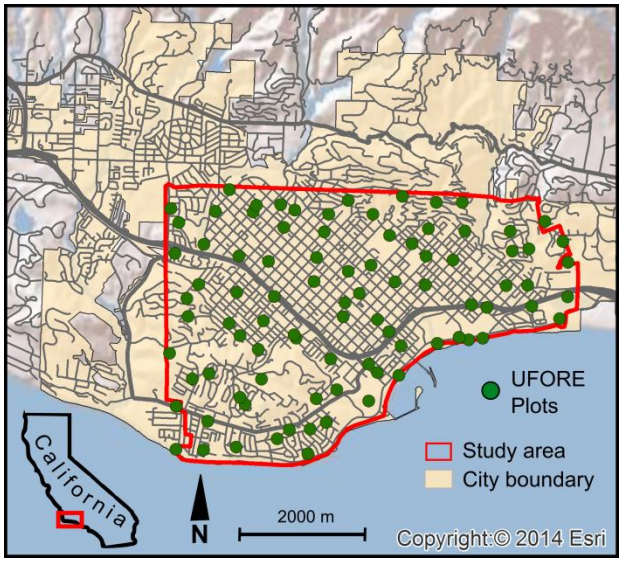

Figure 1.2: Study area with field plots shown inside border of the City of Santa Barbara.

from year-round mild temperatures associated with its Mediterranean climate.

The growing season begins following rains in November and dry-down occurs by the mid-summer. Summer mean minimum and maximum monthly temperatures are 14° C and 24° C respectively. Winter highs are near 18° C with lows of 7° C (US Climate Data, 2015). Precipitation is highest in the winter months with a monthly average of 10

cm of rain and lowest in the summer with less than 2 cm. However, over the last 4 rainy seasons (2011-2012 to 2014-2015), the average monthly precipitation has been ~6 cm (US Climate Data, 2015). Summer fog (June to September) also plays a role in mediating temperature and moisture availability along the coast and into some of the valleys.

The study area is 27.3 km^2, 97% of which lies within the city boundaries of Santa Barbara. Fifty four percent of the city of Santa Barbara's land area is contained by the study area. The downtown commercial districts of Santa Barbara are situated on the coastal plain and exhibit minimal topographic variation. The study-area range of elevation values is from 0 m at the beach to approximately 240 m in the foothills at study area's northernmost extent. Near the southwestern corner of the study area is The Mesa neighborhood, which lies atop one the coastal plain's anticlinal hills (Gurrola et al., 2014). The Mesa has a high point of about 120 m above sea level. The areas of Santa Barbara with greater topographic relief generally coincide with large single-family residential homes.

Table 1.1: Study area land uses (LU) in square km and by percent. Right column is the percent field plot area in each land use. Nat-Ag-Rec stands for Natural-Agricultural-Recreation.

LU class	Study area (km^2)	Study area (% total)	Plot area (%)
Civic	1.50	6.6	9.7
Commercial	1.64	7.3	7.5
Industrial	0.86	3.8	6.4
Multifamily	3.10	13.7	9.4
Nat-Ag-Rec	2.40	10.6	12.1
Single Family	8.02	35.5	29.2
Transportation	5.05	22.4	25.7

Santa Barbara is considered an “urbanized area” based on U.S. Census definitions since, with 89,000 residents, it is a place with greater than 50,000 people and has a population density of greater than 1,000 persons per square mile. Based on census block group data from 2010, the study area houses approximately 71,000 of Santa Barbara's residents. **Table 1.1** shows the distribution of land uses throughout the study area. It is clear that the landscape is dominated by single family residential development but that multi-family housing and commercial areas also play an important role in the spatial composition of the area.

Fractional canopy cover (*fCov*) was estimated in 2012 at 25.4% for the entire municipality of Santa Barbara using high-resolution digital imagery (City of Santa Barbara Urban Forest Management Plan, 2014, www.santabarbaraca. gov). In my more densely developed downtown study area, we estimated tree cover at about 23.7% using lidar data. The Santa Barbara area supports a wide variety of native, introduced, and invasive tree species. The number of species within the study area is unknown, however, 118 unique species were detected during our i-Tree Eco field campaign (105, 0.04 hectare plots). Also, the City of Santa Barbara maintains a geospatial dataset tracking approximately 36,000 publically managed trees. This dataset references more than 450 unique species. While this indicates very high diversity, many of these species are very rare and may even be specimens with only one local example. According to the i-Tree Eco output, the ten most common species account for just over half of the total trees in the study area and the top 25 species account for >80% of canopy cover **(Fig. 1.3)**.

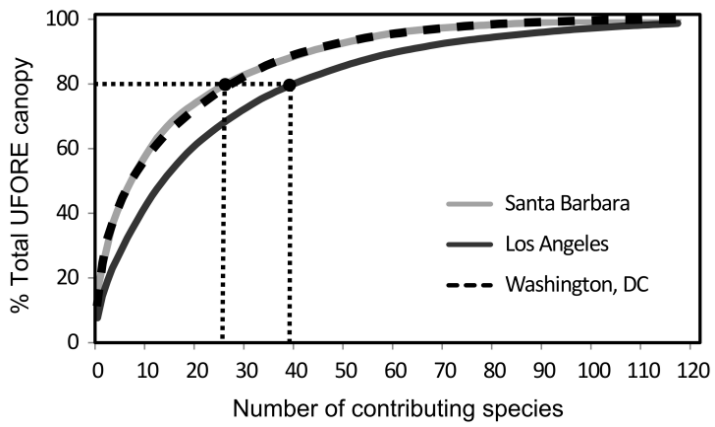

Figure 1.3: The most common species by canopy cover contribution in Santa Barbara compared to Washington, DC (Casey Trees, 2010) and Los Angeles (Nowak et al., 2010). The dotted lines indicate how many species are needed to reach 80% canopy cover. In L.A. there is no native canopy dominant species.

1.5 Monitoring and measuring the urban forest

In the United States, since the year 2000, most large scale efforts to analyze the urban forest ecosystem have made use of the USDA Forest Service's i-Tree Tools software suite. Many cities have specifically employed the i-Tree Eco application to broadly estimate citywide structure and function based on a set measurements taken at several hundred field plots (see i-Tree Eco reports at: www.itreetools.org/resources/reports).

The sampling and measurement methods employed in an Eco study are adapted from Forest Inventory and Analysis (FIA) protocols. Cities generally establish at least 200 field plots stratified by land use classes. Plots are ideally permanent, meaning that annual change in growth and yield can be quantified. This sampling design has generally led to a standard error of the estimate of citywide stem count of approximately 12% (Nowak et al., 2008a). However, the standard error of particular

species stocking levels can vary widely depending on the city's overall biodiversity as well as between-plot heterogeneity. In order to eliminate bias, plots, within each land use, are randomly distributed and all effort is made to gain permission to take measurements on private property. Each plot has a radius of 11.4 m and an area of 0.04 ha.

Within the boundaries of each plot the core tree measurements include species identification, diameter at breast height (dbh), crown dimensions, and tree health. Additional contextual measurements include ground cover, land use, and impervious surface fractions as well as proximity of trees to climate-controlled buildings (Nowak et al., 2008a). LAI is modeled from the crown dimension and species data using a log-linear formula specified for open-grown deciduous trees (Nowak, 1996). Details on the estimation of the ecosystem functions discussed in **Section 1.3** can be found in the following publications: 1) Carbon storage and sequestration (Nowak et al., 2013a); 2) Air pollution reduction (Hirabayashi et al., 2011; Nowak et al., 2006a); 3) Stormwater runoff reduction (Wang et al., 2008); and 4) Building energy modification (McPherson & Simpson, 1999a).

1.6 Remote sensing of the urban forest

In contrast to extensive field sampling, remote sensing can provide spatially extensive data, potentially with higher temporal resolution and lower cost. However, remote sensing has seen limited use in the development of urban forest inventories due to the spatial and spectral complexity of urban landscapes (Herold et al., 2004). For example, Landsat, a space-borne, multispectral sensor is not well suited to this problem

because of its 30 m ground instantaneous field of view, which is much larger than the characteristic scale of spatial variation found in urban areas (~5-10 m; Jensen & Cowen, 1999; Welch, 1982; Woodcock & Strahler, 1987). High spatial resolution (< 3 m), multispectral platforms such as IKONOS and GeoEye have allowed for precise mapping of urban forest canopy extents (MacFaden et al., 2012; Richardson & Moskal, 2014b) but limited spectral information largely precludes species identification or structural quantification.

1.6.1 Species identification using spectral information

Researchers using imaging spectrometry have successfully classified forest and urban tree species using both *in-situ* and airborne spectral data. At the leaf scale the primary controls on reflectance are leaf biochemical properties and leaf anatomy (Asner, 1998; Gates et al., 1965; Roberts et al., 2004). At the crown scale, volumetric structure and the spatial arrangement of foliage drive multiple scattering of photons in the near-infrared (NIR) and reflectance attenuation in the visible (VIS) wavelengths of light (Asner, 1998). Inclusion of non-photosynthetic vegetation (NPV) in the reflected signal will vary based on LAI, leaf angle distribution, and phenology (Roberts et al., 2006; Toomey et al., 2009). Asner and Martin (2008) found that if one prescreens for high LAI conditions – using the Normalized Difference Vegetation Index (NDVI: Rouse *et al.*, 1973) as a proxy – and has a consistent illumination geometry, airborne imaging spectrometry data can be used to assess leaf-level chemical properties at the crown scale.

Pu (2009) examined spectral features of 11 urban tree species in Tampa, Florida using *in-situ* spectra. He used pair-wise analysis of variance to track which vegetation

indices and narrow-band spectral features were most frequently useful in discriminating pairs of species. He found that indices sensitive to water content (e.g. normalized difference water index), along with positional and derivative spectral features related to pigment status, were most often distinct among species. Cochrane (2000) examined 11 tropical tree species with *in-situ* VIS and NIR spectra to determine how well they could be separated from mahogany by capitalizing on interspecies variation in spectral brightness and shape. He concluded that while discrimination is possible, significant error driven by within-class variability at both the leaf and crown level remains.

Clark *et al*. (2005) found that canopy structure led to relatively high separability of tropical rain forest species at the individual tree crown (ITC) level based on variable NIR and shortwave infrared region 2 (SWIR2) reflectance. They were able to discriminate among seven emergent tree species with 92% accuracy using Hyperspectral Digital Imagery Collection Experiment (HYDICE) imagery and manually delineated crowns. Following a previous field study, in which they classified 6 pine and hardwood species using spectrometer data, van Aardt and Wynne (2007) evaluated the classification of 3 of those species using airborne hyperspectral data at both the pixel and 3 x 3-smoothed pixel level. They achieved accuracies of 65% at the pixel level and up to 85% after median filtering. Xiao *et al*. (2004) used spectral mixture analysis applied to AVIRIS data to map 22 urban tree species in Modesto, CA with 70% accuracy at the species level and 94% division-level accuracy. Zhang and Qiu (2012) created crown objects from a lidar point cloud and identified tree-top pixels for 40 species in

Dallas, Texas. They used a neural network approach and achieved 69% classification accuracy.

Recent advances in chemometric techniques at the leaf and crown scale (e.g., Asner & Martin, 2008) have allowed for spectroscopic quantification of a greater number of leaf chemicals than was previously thought possible (Price, 1994). Precise accounting for leaf-level chemicals, in turn, improves our ability to classify at the species level given more physical variables. Still, in biodiverse environments where there can be hundreds of tree species, additional variables that exhibit low correlation with spectral data are needed. Structural variability can be accounted for with lidar range and intensity metrics (e.g. crown surface mean intensity, relative 90th height percentile). These have been used in urban settings to discriminate between coniferous and broadleaf trees and for classification at the genus level (Kim et al., 2011, 2009).

1.6.2 Species identification using lidar

Holmgren and Persson (2004) sought to discriminate between pine and spruce species using discrete return lidar with a footprint diameter of 0.26m. They delineated the ITCs automatically based on fitting a parabolic surface to the rasterized digital canopy model. Using this technique, they detected 91% of the stem volume. This level of tree delineation accuracy is less likely in a mixed or deciduous forest due to overlapping crowns with indistinct tree tops. Within each crown segment the authors computed range-based statistics on the lidar points belonging to that ITC. The features can be summarized as: height central tendency metrics; height percentile metrics; height standard deviation metrics; and proportions of first, single, and vegetated returns. Given a high degree of correlation among many of these variables, eight

groupings that retained some degree of orthogonality were created. When appropriate, the variables were normalized by maximum tree height to decrease sensitivity to changing laser scanners or seasonality. Linear discriminant analysis (LDA) was employed to classify the species and determine the most contributory variables. They achieved an overall accuracy of 95% with a set of six variables. The most valuable single variable (OA = 88.3%) was *pveg* (proportion of vegetation returns).

Kim *et al.* (2009) used lidar intensity metrics to discriminate among seven coniferous and eight broadleaf species grown in a mostly open (i.e. minimal crown overlap) park setting in Seattle, Washington using both leaf-on and leaf-off data. After isolating individual crowns the authors compiled nine intensity metrics for the points within each crown segment. The most effective individual variables as judged by classification accuracy using a linear discriminant function were *entire_all* (mean intensity for all crown returns) for the leaf-on dataset and *upper_all* (mean intensity values for the top 1m of crown returns) for the leaf-off dataset. Ultimately after a principal components rotation was applied an overall accuracy of 91% was achieved using both datasets in tandem. An interesting point to note was that in overlapping crowns, the intensities in the sections of overlap were almost double the values in the non-overlapping portion of the crown. Overlapping sections were excluded from analysis. Kim *et al.* (2011) also used three dimensional point-cloud homogeneity criteria to separate genera at the same park in Seattle. They used an unsupervised cluster analysis method called *k-medoid clustering*. This method is similar to k-means clustering in that it seeks to group similar features but since it's based on median values it is less sensitive to outliers. Clusters of leaf-off data were formed such that evergreen

and deciduous trees were separated. Accuracy declined when leaf-on data were classified.

1.6.3 Estimating LAI using lidar

There has been significant effort in the last decade to quantify forest LAI using airborne lidar. The utility of lidar in this context is manifold: First, appropriately high point density facilitates characterization of tree crowns in three dimensions. Second, modeling the attenuation of lidar beams through canopies is conceptually similar to modeling the extinction of sunlight. The latter, using Beer-Lambert's law is the basis for gap-fraction inversion methods of calculating LAI. Thus, model alignment between ground data in the form of hemispherical photos is not overly complex theoretically. Third, as an active sensor, lidar estimates are not adversely impacted by uneven illumination conditions.

Solberg et al. (2009a) successfully applied airborne laser scanning (ALS) for LAI estimation in a homogeneous forest. To create the relationship with LAI-2000 data and hemiphotos, they developed regression models based on the log-transformed inverse of the ratio of last to first echoes. The form of their equation relating LAI to ALS penetration rate is typical of many studies in that it is based on inversion of the Beer-Lambert law. They reached an r^2 of greater than 0.9. They note that in the case of relatively homogenous leaf angle distribution, gap fraction can be estimated from a vertical angle. Thus, lidar penetration rate may be used to estimate gap fraction assuming the beam divergence is generally smaller than the gaps.

Zhao and Popescu (2009) estimated the LAI of a pine-dominated forest in East Texas. In their study, they highlighted the many places one may encounter issues with respect to hemiphoto analysis, algorithm choice, lidar metric choice, and spatial alignment of photos with lidar data. For instance, gap fraction results were dependent on both the analyst and the algorithm. LAI estimates from one example photo differed by 0.28 when different algorithms were compared. Further, they list five impediments to developing regression models for LAI using lidar metrics:

1. One doesn't know the appropriate height threshold for separating ground and canopy
2. One doesn't know the field of view of the hemiphotos
3. Many effective lidar metrics remain site specific
4. Still lacking exact functional form of relationship between various lidar metrics and LAI
5. Differing forest types will influence relationships between LAI and lidar metrics

Perhaps the most consistently problematic of the above issues is the difficulty relating the field of view of hemiphotos to a certain lidar data trap size. Studies that take place under homogeneous canopy are less impacted by this issue as it is assumed that LAI will be consistent regardless of field of view. In heterogeneous stands this issue is most often addressed through correlation analysis in which the zenith angle of the hemiphoto and the radius of the lidar data trap are simultaneously varied until optimal correspondence is achieved (Morsdorf et al., 2006; Richardson et al., 2009a; Zhao & Popescu, 2009).

Of the previous work estimating LAI from discrete return lidar, the studies by Morsdorf et al. (2006) and Richardson et al. (2009a) come the closest to addressing the issues one will face taking these measurements in an urban setting. Morsdorf et al. (2006) introduced a method whereby LAI of a scene (i.e. the field of view of a hemiphoto) is estimated with lidar for canopy only and then rescaled based on the scene's fractional vegetation cover. This is applicable to urban areas where, within any given plot, there may be extreme heterogeneity in vegetation cover, ranging from 0 to 100%. This rescaling allows the researcher to exclude all areas of zero vegetation cover from the analysis. Richardson et al. (2009a) built on this work in the semi-urban setting of Washington Park Arboretum in Seattle. In the Washington Park Arboretum there is a wide variety of tree species and high variability in fractional cover. The authors tested the capacity of four different models to estimate LAI as measured by the LAI-2000 and hemiphotos. They found that the model formulation most closely matching Beer-Lambert's law of light attenuation yielded the highest r^2 values (0.78). They applied this model to map LAI throughout their study area. While they sought a spatial resolution of 3m for their map, they found that, many cases they did not have any ground returns within this small rectangle and could not, thus apply the logarithmic transformation in their selected model. The final map product was created with a 14 m spatial resolution which is not optimal for an urban setting. Finer resolution mapping will be possible with higher point density lidar datasets.

Chapter 2

Urban tree species mapping using hyperspectral and lidar data fusion

Authors:

Michael Alonzo[a,*], Bodo Bookhagen[b], and Dar A. Roberts[a]

[a] Geography Department, Ellison Hall 1832, UC Santa Barbara, CA 93106-4060, USA

[b] University of Potsdam, Institute of Earth and Environmental Science, Karl-Liebknecht-Str. 24-25, 14467 Potsdam-Golm, Germany

Citation for this chapter as published in *Remote Sensing of Environment:*

Alonzo, M., Bookhagen, B., & Roberts, D. A. (2014). Urban tree species mapping using hyperspectral and lidar data fusion. *Remote Sensing of Environment*, 148, 70–83.

2.1 Abstract

In this study we fused high-spatial resolution (3.7 m) hyperspectral imagery with 22 pulse/m^2 lidar data at the individual crown object scale to map 29 common tree species in Santa Barbara, California, USA. We first adapted and parallelized a watershed segmentation algorithm to delineate individual crowns from a gridded canopy maxima model. From each segment, we extracted all spectra exceeding a Normalized Difference Vegetation Index (NDVI) threshold and a suite of crown structural metrics computed directly from the three-dimensional lidar point cloud. The variables were fused and crowns were classified using canonical discriminant analysis. The full complement of spectral bands along with 7 lidar-derived structural metrics were reduced to 28 canonical variates and classified. Species-level and leaf-type level maps were produced with respective overall accuracies of 83.4% (kappa = 82.6) and 93.5%. The addition of lidar data resulted in an increase in classification accuracy of 4.2 percentage points over spectral data alone. The value of the lidar structural metrics for urban species discrimination became particularly evident when mapping crowns that were either small or morphologically unique. For instance, the accuracy with which we mapped the tall palm species *Washingtonia robusta* increased from 29% using spectral bands to 71% with the fused dataset. Additionally, we evaluated the role that automated segmentation plays in classification error and the prospects for mapping urban forest species not included in a training sample. The ability to accurately map urban forest species is an important step towards spatially explicit urban forest ecosystem assessment.

2.2 Introduction

As of 2011, more than 50% of all humans live in cities (UN-Habitat, 2011). Cities play an outsized role in driving global climate change (Schneider et al., 2010) and are uniquely susceptible to climate change impacts. Urban areas suffer from higher temperatures, poorer air quality, and increased peak flow of stormwater runoff, when compared to their rural neighbors (Voogt, 2002; Escobedo & Nowak, 2009; Lee & Bang, 2000). Optimally arranged green infrastructure in cities can reduce impacts by facilitating reduced urban temperatures, improving air quality, and dampening peak flow (Bolund & Hunhammar, 1999; Myint et al., 2010). Urban trees in particular provide a range of ecosystem services, along with some disservices (e.g., Lyytimäki et al., 2008), but the magnitude of service depends on tree species, structure, and locational context (Escobedo & Nowak, 2009; Manning, 2008; McCarthy & Pataki, 2010; Simpson, 2002; Urban, 1992; Xiao & McPherson, 2011). Presently, the Urban Forest Effects model (i-Tree Eco Nowak & Crane, 2000) is commonly implemented in urban areas worldwide to produce citywide estimates of urban forest structure, species diversity, and ecosystem function. However, urban forest inventory, particularly on private properties, is labor intensive and the results are not spatially explicit.

Mapping the extents of urban tree canopy using aerial or satellite imagery is currently operational (MacFaden et al., 2012; McGee et al., 2012). However, these maps rarely provide information on tree species, age class, or leaf area index (LAI) , which are common prerequisites to estimates of ecosystem function. Mapping tree species is challenging in urban environments due to the fine characteristic scale of spatial variation (Welch, 1982) and potentially very high species diversity. While some space-

borne, broadband sensors (e.g., IKONOS, GeoEye) are capable of achieving < 3 m multispectral spatial resolution, they lack the spectral range and resolution required to resolve the subtle chemical and structural signatures upon which species discrimination relies (Clark et al., 2005). Hyperspectral imagery has proven useful in mapping tree species at the pixel level based on variability in spectral reflectance at leaf to crown scales (Martin et al., 1998; Dennison and Roberts, 2003; Clark et al., 2005; van Aardt and Wynne, 2007; Boschetti et al., 2007; Franke et al., 2009; Yang et al., 2009; Youngentob et al., 2011). In an urban setting, Xiao et al. (2004) mapped 22 common species in Modesto, California with 70% accuracy at the species level and 94% accuracy at the leaf-type (i.e. broadleaf, conifer, palm) level.

Classification accuracies for pixel-based algorithms in highly mixed urban landscapes are limited by extreme spectral variation over small spatial extents. In response there has been increased use of object-based image analysis (OBIA), which relies on image segmentation routines to group spectrally similar and spatially proximate pixels into objects to reduce undesirable noise common in pixel-level results (Benz et al., 2004; Blaschke, 2010; Myint et al., 2011). This technique has been applied with some success to tree species identification using hyperspectral imagery either through crown-level spectral averaging or pixel-majority classification (Clark et al., 2005; van Aardt and Wynne, 2007; Zhang and Qiu, 2012; Alonzo et al., 2013). In a suburban setting north of Dallas, Texas, Zhang and Qiu (2012) achieved a classification accuracy of 69% for 40 tree species using a “treetop-based” approach. They selected the single highest pixel per crown object in order to ensure sunlit spectra whenever possible. Alonzo et al. (2013) showed that for manually delineated urban tree crowns in

Santa Barbara, the pixel majority approach using all crown pixels exceeding a Normalized Difference Vegetation Index (NDVI) threshold was effective, especially with limited training data. Their classification of 15 urban species with Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data resulted in an overall accuracy of 86%. Nevertheless, Castro-Esau et al. (2006), while producing strong species classification results using leaf-level spectra, show a linear decline in classification accuracies with increasing numbers of species. This suggests that 1) it may not be currently possible to map all species simultaneously in biodiverse forests and 2) that significantly expanding the classification feature space with non-spectral data may be required for significant advances.

Lidar data allow for the generation of a set of crown structural variables based on either the ranges and intensities of individual pulse returns or characterization of the full waveform. Lidar data have been employed frequently to measure forest parameters such as tree height (e.g., Lim et al., 2003; Anderson et al., 2006; Edson and Wing, 2011), biomass (e.g., Popescu et al, 2003; Næsset and Gobakken, 2008; Asner et al, 2011; Mascaro et al, 2011; Shrestha and Wynne, 2012), and LAI (e.g., Morsdorf et al., 2006; Solberg et al., 2009; Zhao and Popescu, 2009; Tang et al., 2012). Classification of trees using pulse range and intensity metrics has been undertaken at the leaf type (e.g., Kim et al., 2009; Ørka et al., 2009; Yao et al., 2012), genus (e.g., Kim et al., 2011), and species levels (e.g., Holmgren and Persson, 2004; Brandtberg, 2006). Other work has shown that retaining the full lidar waveform can provide a set of discriminatory variables derived from, for example, echo width and amplitude (Heinzel and Koch, 2011; Vaughn et al., 2012). Suites of canopy structural variables (e.g. tree height, crown

base height, vertical intensity profiles) extracted from the lidar point cloud at the individual tree level offer complementary information to the biochemical and biophysical data garnered from optical data. However, it has thus far not been demonstrated that lidar-variables alone are sufficient for discriminating among large numbers of species in biodiverse environments.

"Fusion" is a ubiquitous term in the remote sensing literature that generally refers to the combination of multisensor spatial data, at either the pixel, feature, or decision level (Pohl and Van Genderen, 1998). Increasingly, lidar and either multispectral (e.g., Holmgren et al., 2008; Ørka et al., 2012) or hyperspectral (e.g., Asner et al., 2008; Voss and Sugumaran, 2008; Dalponte et al., 2008; Jones et al., 2010; Liu et al., 2011; Dalponte et al., 2012; Dalponte et al., 2014) data are fused together at the pixel or feature level for tree species classification and quantification of forest inventory parameters (e.g., Lucas et al., 2008; Swatantran et al., 2011; Clark et al., 2011; Anderson, et al., 2008; Latifi et al., 2012). In some cases the value of fusion has come from the addition of structural variables (e.g., height, standard deviation of all height points within a pixel) that are minimally correlated with spectral bands (Dalponte, et al., 2008; Voss and Sugumaran, 2008; Jones et al., 2010; Dalponte et al., 2012). In others, fusion has added value indirectly through improved image segmentation and crown-object creation (Voss and Sugumaran, 2008; Zhang and Qiu, 2012; Alonzo et al., 2013; Dalponte et al., 2014). However, to the authors' knowledge, there has been minimal research focused on improving tree species classification using crown-object level fusion of hyperspectral imagery and structural metrics extracted directly from the 3-D lidar point cloud. Moreover, the prospects for mapping an *entire*, biodiverse urban

forest to the leaf-type level with hyperspectral-lidar data fusion, have not been evaluated. Finally, there is limited knowledge of how automated image segmentation impacts the accuracy of classification results in a highly complex urban environment.

The goal of this study is to improve the accuracy of tree species mapping in the biodiverse city of Santa Barbara, California, through crown-object level fusion of AVIRIS (Green et al., 1998) imagery and high point-density lidar data. This paper builds significantly on the work by Alonzo et al. (2013) which focused on classifying manually-delineated tree crowns using hyperspectral imagery. In particular, we now include lidar-derived structural metrics in classification algorithms and delineate crowns using watershed segmentation. The specific aims of this paper are:

1) For our urban study area, within crown objects delineated using watershed segmentation, classify 29 common tree species using crown-level fusion of hyperspectral imagery and lidar data.
2) Test the extent to which *all* of the urban forest's canopy can be classified to the leaf type level using classification functions developed for the 29 common species. Leaf-type level classification is frequently sufficient for parameterizing estimates of urban ecosystem function that are largely mediated by crown structure measurements and total leaf area.
3) Evaluate the impact of segmentation error on classification accuracy through comparison of results from automatically delineated and manually delineated crowns.
4) Isolate particular spectral regions and lidar-derived structural variables that hold promise for improving discrimination among urban tree species and leaf types.

Our study helps cities move closer to a spatially explicit accounting of the common species in their urban forest. Further, it facilitates better understanding of the spectral and structural contributions to species discrimination as well as the benefits and errors associated with object-oriented approaches.

2.3 Data and Methods

2.3.1 Study area and sample

This study was conducted in downtown Santa Barbara, California (34.42° N, 119.69° W) (**Fig. 2.1**). Santa Barbara is a city of about 90,000 residents located on a coastal plain between the Pacific Ocean to the south and the Santa Ynez mountains to the north. It has a Mediterranean climate and supports a diverse mix of native,

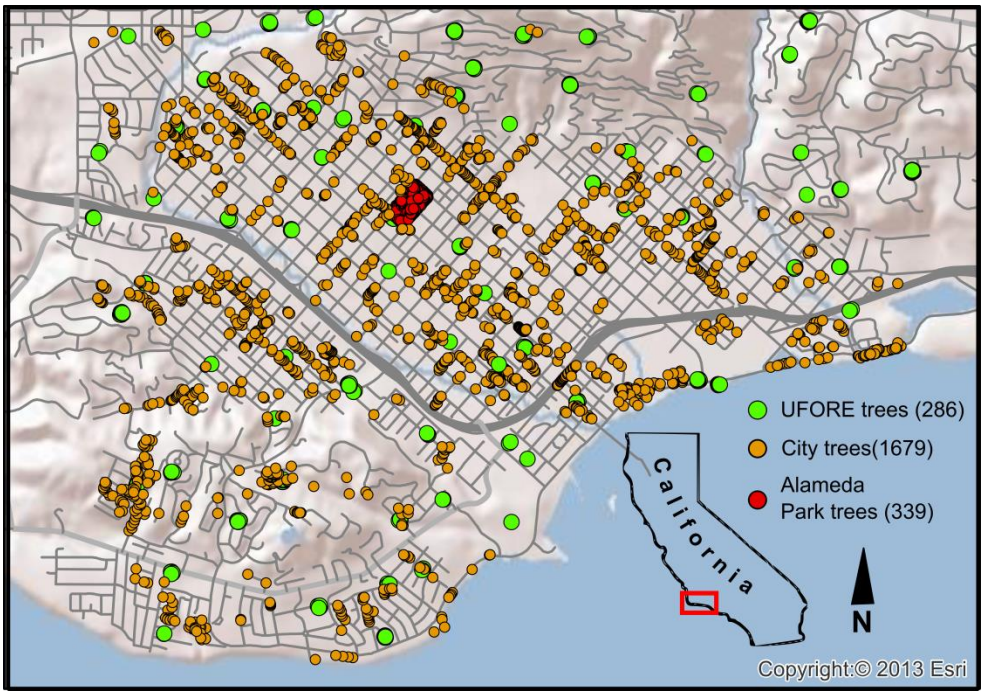

Figure 2.1: Study area of downtown Santa Barbara with the locations of all trees used for classification model calibration and validation.

introduced, and invasive urban forest species. A spatial database maintained by the City of Santa Barbara contains one or more specimens from >450 species. In a Fall 2012 inventory following i-Tree Eco protocols, 105 plots were sampled and 108 species recorded. Despite this diversity, far fewer species provide the bulk of the city's canopy cover: In Santa Barbara, based on i-Tree Eco and municipal data, we estimate that approximately 70% of the study area's trees represent over 80% of the city's canopy area yet comprise fewer than 30 species.

This study's first objective was to map approximately 80% of Santa Barbara's canopy to the species level by training a classifier on 29 common species. The 80% canopy cover threshold was chosen based on analysis of i-Tree Eco-derived cumulative canopy cover distributions in the cities of Santa Barbara, Washington, DC (Casey Trees, 2010) and Los Angeles, California (**Fig. 1.3**; Clarke et al., 2013). Twenty nine species (**Table 2.1**) were ultimately chosen for their large contributions to canopy cover and our ability to isolate training crowns (further details in **Appendix A.1**). The other 20% of the canopy (hereafter "less common" species) were modeled as one of the trained species (hereafter "common" species) and thus classified only to the leaf-type level. The tree crowns included in this study's training set (**Fig. 2.1**) were selected from: 1) The city's geospatial database (1,679 stems); 2) i-Tree Eco plots (286 stems);and 3) one additional park (Alameda Park) spanning two city blocks (339 stems). The species of each crown selected from the city's database was confirmed using Google Street View. The 2nd set was collected by the authors and included species identification and all i-Tree Eco-prescribed stem, crown, and positional measurements (Field methods in detail:

Table 2.1: The 29 species included in model training. Tree type: B = Broadleaf, C = Coniferous, P = Palm. Canopy area refers to total canopy area by species.

Species Code	Scientific Name	Tree Type	Stem Count	Canopy Area (m^2)
ARCU	*Archontophoenix cunninghamiana*	P	62	756
CICA	*Cinnamomum camphora*	B	57	5290
CUMA	*Cupressus macrocarpa*	C	55	4857
EUFI	*Eucalyptus ficifolia*	B	50	4596
EUGL	*Eucalyptus globulus*	B	58	9401
FIMI	*Ficus microcarpa*	B	56	9006
GEPA	*Geijera parviflora*	B	58	2777
JAMI	*Jacaranda mimosifolia*	B	76	6609
LIST	*Liquidambar styraciflua*	B	65	5081
LOCO	*Lophestemon confertus*	B	66	3465
MAGR	*Magnolia grandiflora*	B	63	7425
MEEX	*Metrosideros excelsa*	B	62	1581
OLEU	*Olea europaea*	B	81	6042
PHCA	*Phoenix canariensis*	P	99	5294
PICA	*Pinus canariensis*	C	73	4675
PIPI2	*Pinus pinea*	C	76	11387
PIUN	*Pittosporum undulatum*	B	96	6166
PLRA	*Platanus racemosa*	B	71	6933
POGR	*Podocarpus gracilior*	B	62	6214
PYKA	*Pyrus kawakamii*	B	55	3404
QUAG	*Quercus agrifolia*	B	108	8895
SCMO	*Schinus molle*	B	53	1971
SCTE	*Schinus terebinthifolius*	B	71	5863
STSI	*Stenocarpus sinuatus*	B	51	1112
SYAU	*Syzygium australe*	B	67	3982
SYRO	*Syagarus romanzoffiana*	P	130	2705
TISP	*Tipuana tipu*	B	58	7874
ULPA	*Ulmus parvifolia*	B	50	6370
WARO	*Washingtonia robusta*	P	87	1220

https://sites.google.com/site/ucsbviperlab/i-Tree Ecomethods). The third set was also collected by the authors and each tree was identified to the leaf-type level with stem position precisely recorded using differential GPS and a total station. The primary

utility of this final set was to offer a complex set of overlapping crowns on which to validate watershed segmentation algorithms. The total number of stems included in this analysis is 2,304, comprising approximately 100 species. This number is approximate because not all species in Alameda Park were identified beyond their leaf type. Common species make up 2,016 of the 2,304 total crowns in the sample and 91% of the total crown area (165,887 m^2).

2.3.2 Data

2.3.2.1 Lidar data and processing

Waveform lidar data were collected in August of 2010 with a helicopter-mounted Riegl Q560 laser scanner. The lidar data were georeferenced with two local differential GPS stations and stored in the UTM coordinate system (Zone 11N, NAD83). The waveform was discretized using standard Riegl processing procedures to an average last-return point density of 22 points/m^2 across the study area with additional returns available in high vegetation. Height values on flat surfaces were evaluated to be precise to within 2 cm (**Appendix A.2**). The point cloud was classified to ground, building, and vegetation using LAStools (LAStools v111216, http://lastools.org) with minimal adjustments to default settings. Buildings were discriminated from trees with 98% accuracy and there was no confusion between vegetation taller than 2 m and ground. A bare earth digital terrain model (DTM), and two canopy height models (CHM), one for buildings and the other for vegetation, were generated at 0.25 m pixel resolution. In this research we use the term "canopy height model" to refer to height above bare ground.

2.3.2.2 AVIRIS imagery and processing

Two AVIRIS flight lines spanning the study area were acquired from a Twin Otter aircraft flying at approximately 4000 m above sea level on November 1, 2010. The scene acquisition times were centered on 11:50 and 14:20 Pacific Standard Time with solar zenith angles of 50.5° and 54.1°, respectively. The 224-channel AVIRIS instrument samples upwelling radiance between 365 and 2500 nm with a field of view of 34° and instantaneous field of view of 1 mrad (Green et al., 1998). The resultant ground instantaneous field of view was 3.7 and 3.4 m for the two flight lines, respectively.

AVIRIS products are delivered after correction for aircraft motion and orthorectification using digital terrain. Surface reflectance was retrieved on each flight-line using ATCOR-4 (Richter and Schlaepfer, 2002). Bands within the following spectral regions were discarded due to water vapor contamination or low signal-to-noise ratios: 365—385 nm, 1323—1432 nm, 1811—2007 nm, and 2446 2496 nm. After confirming negligible reflectance bias between the two images, a mosaic was created (3.7 m) and registered to the gridded lidar data (0.25 m). The AVIRIS data were warped using Delaunay triangulation based on 137 ground control points and resampled using nearest neighbor resampling. The average root mean square error (RMSE) in the alignment cannot be calculated automatically when Delaunay triangulation is employed. Visual assessment suggests that the error was less than one AVIRIS pixel.

2.3.3 Crown segmentation

A general overview of the segmentation process along with the full methods workflow employed for this study is shown in **figure 2.2**. In short, tree canopy was isolated from abiotic scene components and low vegetation primarily based on the

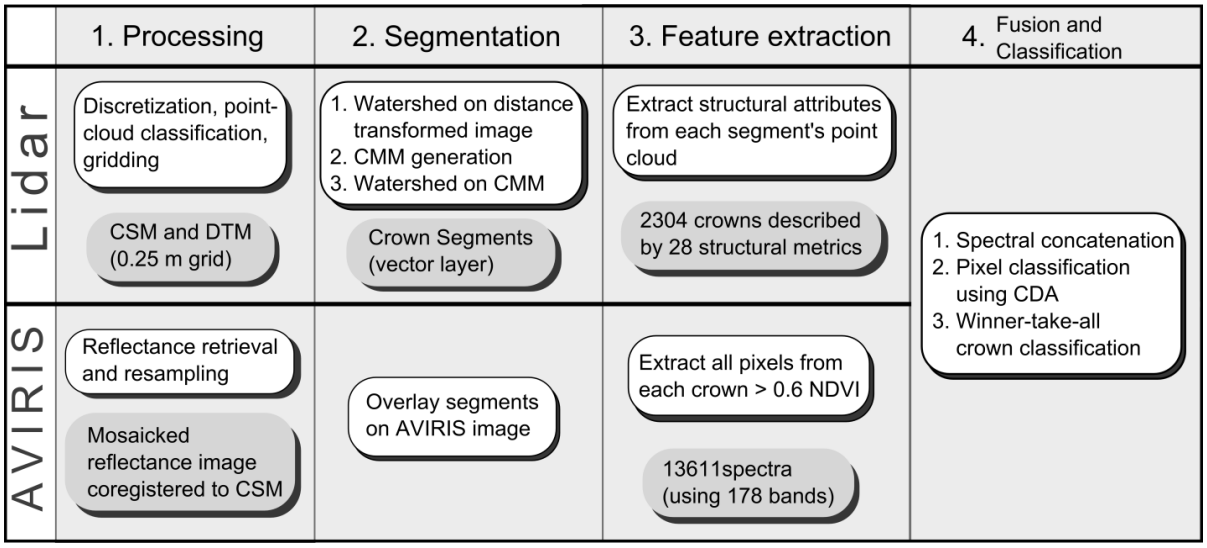

Figure 2.2: Workflow diagram for data preprocessing, crown delineation via watershed segmentation, data fusion, and classification. White boxes are processes and gray boxes are data products. AVIRIS = Airborne Visible Infrared Imaging Spectrometer, CSM = Canopy Surface Model , DTM = Digital Terrain Model, CMM = Canopy Maxima Model, NDVI = Normalized Difference Vegetation Index, CDA = Canonical Discriminant Analysis.

point cloud classification completed in LAStools. Additional refinement was conducted using morphological opening, closing, thickening, and majority filtering on the gridded DTM and CHMs. This allowed for removal of most power lines, isolated vegetation canopies < 1 m^2, and other image noise and resulted in a 0.25 m binary canopy image.

Building on the methods of Chen et al. (2006), marker-controlled watershed segmentation (Digabel and Lantuéjoul, 1978) was chosen to isolate individual tree crowns on the CHM. The watershed algorithm was originally proposed as a mean to segment grayscale images (Digabel and Lantuéjoul , 1978). In the creation of watershed segments, one considers the brightness values of an image essentially as topography that can be divided into catchment basins surrounding local minima. The lines that divide one catchment area from its neighbors define the boundaries of the segments. This segmentation algorithm is not robust to image noise and spurious local minima frequently lead to oversegmentation. Suppressing all but those local minima relating to

the image objects in question yields a better result (Chen et al., 2006). Marking only appropriate minima can be challenging in itself: many images, including CSMs, contain overlapping distributions of appropriate and extraneous local minima values making it difficult to simultaneously minimize omission and commission error.

For this project, we combined two watershed segmentation routines. The first was executed on an inverse distance transformed, binary canopy image, where local minima were marked at crown locations furthest from canopy edges. The second was executed on an inverted canopy maxima model (CMM) where markers were imposed in locations corresponding with maximum tree height. The second routine was enacted on each segment produced by the first routine thus further subdividing the initial segments. To create a CMM we used i-Tree Eco data to establish a local, empirical relationship between tree height and canopy width (see **Fig. A.2** in **Appendix A** for detail). This linear model was used to establish a variable search window size for local crown maxima based on the modeled tree width. As an input to watershed segmentation, this method is considered an improvement over use of a CHM because it is more likely that only legitimate tree tops will be marked (Chen et al., 2006). However, the relationship between tree height and width in Santa Barbara's urban forest, irrespective of species was relatively weak ($r^2 = 0.38$).

Segmentation accuracy was estimated by calculating the ratio of field-measured stems contained in exactly one segment to the sum of the total number of stems and the number of segments containing zero stems. It was beyond the scope of this study to evaluate the areal accuracy of the segmentation. For further processing details and

associated Matlab code relating to the watershed segmentation and CMM generation we refer the reader to **Appendix A.3**.

2.3.4 Spectral and structural feature extraction

The following sections describe how the spectral and structural variables were generated from crown-level segments overlaid on the AVIRIS imagery and the lidar point cloud (**Fig. 2.3**). These variables were then fused and used to generate the classification models. Multiple AVIRIS spectra were extracted for each crown and a single set of 28 structural metrics were computed from each watershed-crown clipped point cloud. **Section 2.3.5** on data fusion will discuss how these variables were combined for input into classification algorithms.

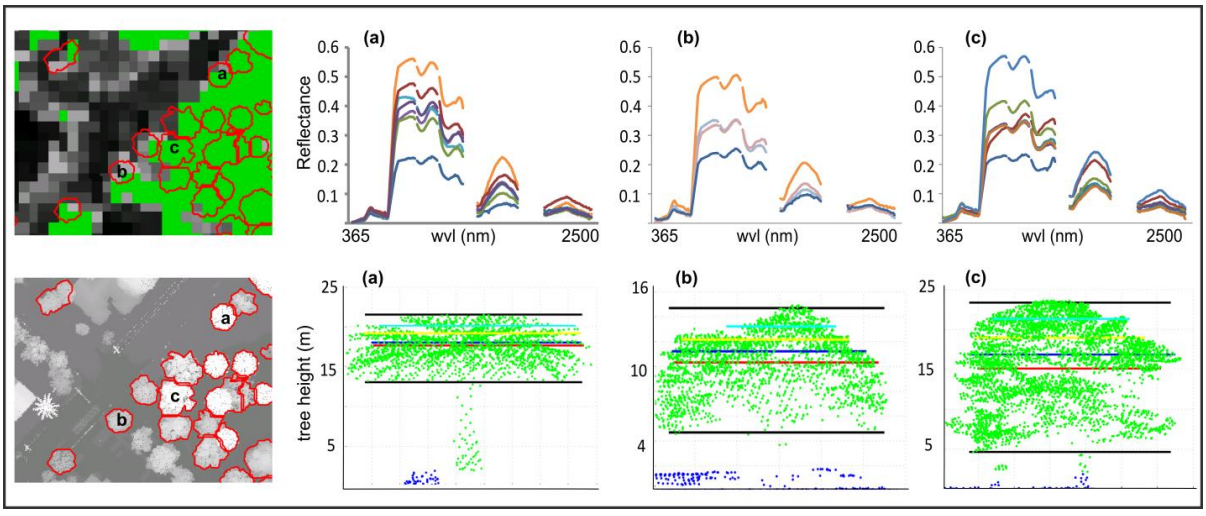

Figure 2.3: Top row: Multiple AVIRIS spectra extracted from each of three typical watershed crowns. The green mask highlights pixels with NDVI > 0.6. The three example crowns are: a) PHCA (Phoenix canariensis), a palm; b) LOCO (Lophostemon confertus) a broadleaf evergreen; and c) PICA (Pinus canariensis), a conifer. Bottom row: Watershed crown point-cloud extraction with selected structural metrics. Black lines indicate crown base and max heights. Red line is mean crown height. Dark blue is median height of returns in crown. Orange and violet show 75th and 90th percentile heights as well as widths at those heights

2.3.4.1 Extraction of spectra from AVIRIS imagery

Prior to extraction of spectra, the AVIRIS image was resampled from 3.7 m to 1 m resolution using nearest neighbor resampling. While this created redundant spectra, it was necessary to ensure that spectral information from valid 3.7 m canopy pixels located on segment edges was included in the analysis. Spectra were extracted from each crown segment (**Fig. 2.3**) using a variable NDVI (Rouse et al., 1973) threshold to reduce contamination by impervious surface, soil, or non-photosynthetic vegetation spectral information. The average NDVI for all crowns was 0.61. Therefore, the initial extraction threshold was set to 0.6. If no pixels in a given crown met that criterion, all pixels above an NDVI of 0.5 were extracted. If no pixels met this second threshold, the single pixel with the maximum NDVI value was selected (Alonzo et al., 2013). All redundant spectra in a given crown segment were eliminated prior to classification. For our sample of 2,304 crowns, 13,611 spectra were extracted with a median of 4 unique spectra per crown. Many of the species included in this study have small crowns that fully contain only one or two 3.7 m pixels. As such, 23% of crowns were represented in the classification stage by only one unique spectrum.

2.3.4.2 Extraction of structural metrics from lidar data

The lidar point-cloud subset associated with each tree crown was extracted from the scene tiles so that each crown could be processed individually (**Fig. 2.3**). This strategy allows for arbitrarily large tree crown datasets to be processed either in serial or with simple parallelization. Building upon previous work (e.g., Holmgren and Persson, 2004; Kim et al., 2009), we created 28 structural variables (**Table 2.2**). These metrics can be roughly categorized as relating to crown height (h), crown widths at

Table 2.2: Lidar-derived structural variables. Bold entries were selected for inclusion in classification models for watershed crowns.

Variable	Description
h_1	Max crown height
h_2	**Median height of returns in crown**
h_3	Crown surface height: 0.25 m spatial resolution
h_4	Crown surface height: 1 m spatial resolution
h_5	Crown base height
w_1	**Crown width at median height of returns in crown**
w_2	Crown width at 50th percentile height
w_3	Crown width at 75th percentile height
w_4	Crown width at 90th percentile height
hw_rat_1	Ratio of crown length to tree height
hw_rat_2	**Ratio of crown height to width: median height**
hw_rat_3	Ratio of crown height to width: 90th percentile height
hw_rat_4	Ratio of crown height to width: 75th percentile height
hw_rat_5	Ratio of width at 90th percentile height to mean height
hw_rat_6	Ratio of N-S width to E-W width
int_1	Average intensity above median height
int_2	**Average intensity below median height**
int_3	**Crown surface intensity: 0.25 m spatial resolution**
int_4	Crown surface intensity: 1 m spatial resolution
int_dist_1	Crown surface intensity /overall average crown intensity
int_dist_2	Skewness of intensity distribution through crown
int_dist_3	Surface intensity (0.25 m) / surface intensity (1 m)
int_dist_4	Return intensity above median crown height / below
cp_1	**Surface heights (0.25 m) / surface heights (1 m)**
cp_2	(Mean crown height - median height of returns) / crown height
cp_3	**Count of returns in 0.5 m vertical slice at 90th percentile height divided by width at that height**
cp_4	Count of returns in 0.5 m vertical slice at mean crown height divided by width at that height
cp_5	Count of returns in 0.5 m vertical slice at median height of crown returns divided by width at that height

selected heights (*w*), ratios of crown heights to widths at selected heights (*hw_rat*), direct measures of return intensity through the crown (*int*), distributions of intensity through the crown (*int_dist*) and crown porosity measured by return penetration into

the crown (*cp*). Details and Matlab code relating to the computation of these variables can be found in **Appendix A.4**. While overall correlation among variables was limited, there were several groupings exhibiting Pearson's Product Moment Coefficients (r) greater than 0.80 (see **Table A.1** in **Appendix A**). In order to exclude correlated variables and to choose the most effective variables for species separability, forward feature selection (FFS) was employed.

2.3.4.3 Forward feature selection (FFS)

FFS is a method used to reduce a high-dimensional dataset that may contain redundant discriminating variables (Hoffbeck and Landgrebe, 1996). The one variable that best discriminates among classes is first added to the classification model based on its ability to minimize the model's misclassification rate (MCR). Each remaining variable is sequentially tested to assess which will, when combined with those already included, offer the greatest marginal decrease in MCR. Alonzo et al. (2013) previously showed that a reduced set of spectral variables did not yield classification accuracies higher than the full complement of 178 bands and performed no better computationally than a reduced set of canonical variates. Thus, for purposes of improving model classification accuracy, only lidar variables were winnowed for further use using FFS. However, it is also an objective of this research (aim #4) to explore the contributions of different spectral regions to species separability in an urban forest. As such, each spectral band's unique contribution to separability was evaluated based in-part on frequency of that band's selection. A more general measure of separability, assessed as the ratio of among class sums of squares to within class sums of squares (F-ratio; Clark et al., 2005), was

also employed to isolate useful spectral regions and categories of structural metrics (e.g. all variables related to height).

To explore the sensitivity of the structural variables to crown segmentation error, FFS was run on both manually delineated crowns (hereafter manual crowns) and watershed segments (hereafter watershed crowns). FFS was run on each set 100 times with crowns randomly partitioned each run into training and validation sets to mitigate the impact of crown variability in the sample. The seven structural metrics that were chosen using manual crowns in more than 30% of the runs and that demonstrated minimal intercorrelation were retained for further use. Accordingly, the 7 most-frequently selected metrics exhibiting low correlation were chosen for the watershed crowns (**Table 2.2**).

2.3.5 Data fusion and classification

2.3.5.1 Fusing spectral and structural data at the crown level

In this study, the majority of tree crowns contained multiple, unique spectra meeting the 0.6 NDVI threshold. However, there was only *one* set of structural metrics extracted per crown (**Fig. 2.3**). Alonzo et al. (2013) demonstrated that retaining multiple pixels per crown and assigning a class to the crown object using a pixel majority ("winner-take-all") approach was more accurate than classification using a single, crown-mean spectrum. As such, for each crown, we chose to replicate the set structural metrics to correspond with the number of extracted spectra. The resulting data matrix for manual crowns contained 12,773 rows and 206 columns, where each row represents a unique spectrum and each column contains either one of 178 spectral

bands or one of 28 structural metrics. The same structure was created for the watershed crowns but with 13,317 rows.

2.3.5.2 Canonical discriminant analysis (CDA)

All classifications in this study were conducted using canonical variates in a linear discriminant analysis (LDA) classifier. LDA has proven useful previously in remote sensing research for separating highly overlapping classes in (e.g., Yu et al., 1999; Clark et al., 2005; Pu, 2009). In LDA, classification equations are formulated based on the pooled within-class covariance matrix of the set of independent variables. An observation is assigned to the class with the highest classification function score (Duda and Hart, 1973). In canonical discriminant analysis one replaces p original variables with up to $g-1$ derived canonical variates, where g is the number of classes (i.e. 29 tree species; Klecka, 1980). Whereas principal components analysis (PCA) and minimum noise fraction (MNF) summarize the total variability among the set of independent variables, the canonical rotation summarizes the between class variance among g classes. The derived canonical discriminant functions are linear combinations of the original variables where the coefficients maximize the between-group separation. Data reduction with this technique has been successfully applied to remote sensing classification problems including urban tree species discrimination (Zhao and Maclean, 2000; Pu and Liu, 2011; Alonzo et al., 2013). Compared to LDA on p original variables, CDA dramatically improves computational performance and, in the case of limited training data, can avoid the ill-posed problem where the number of variables is greater than the number of observations.

2.3.5.3 Classification candidate sets

The primary goal of this research was to assess the accuracy with which we could map tree species in a heterogeneous urban forest using fused hyperspectral and lidar data. We attempt to map 29 common species that comprise much of Santa Barbara's canopy area and provide the majority of urban-forest derived ecosystem services. We acknowledge that it is currently impossible to train a classification algorithm on all species present in an urban area. Thus, we trained our CDA classifier to label *all* crowns as one of the 29 common species. At the leaf-type level, the classification was deemed successful when a crown was labeled as a common species with a matching leaf type. For example, if a *Quercus suber* (less common species) was classified as *Quercus agrifolia* (common species) then the leaf-type classification was correct.

In order to assess classification accuracy separately for the 29 common species and the ~70 less common species, we subdivided the 2,304 total crowns into the four overlapping sets listed below (each corresponding research aim from the introduction is also noted):

1) Accuracy for mapping 29 common species (2,016 crowns) to the species level (aims #1 & #3)
2) Accuracy for mapping same 29 common species to the leaf-type level (aims #2 & #3)
3) Accuracy for mapping ~70 less common species (288 crowns) to the leaf-type level (aims #2 & #3)
4) Accuracy for mapping ~100 total species (2,304 crowns) to the leaf-type level (aims #2 & #3)

2.3.5.4 Variable combinations

Each candidate set listed in **section 2.3.5.3** was classified using four different variable subsets. The purpose of adding and holding out variables links with research aim #1: we seek to assess the respective values of hyperspectral data, lidar data, and object-level fusion of both in classification accuracy at the species and leaf-type levels. Prior to classification, for the sake of computational efficiency and methodological consistency, each variable combination was reduced to the maximum number of canonical variates with significant discriminating power (α=0.05). The rotated variable sets used to generate classification equations were thus:

1) All hyperspectral bands (178) and all lidar-based structure variables (28) reduced to 28 canonical variates (hereafter: *CDA-full*).
2) All spectral bands and the subset of 7 lidar variables selected using FFS, reduced to 28 canonical variates (*CDA-7fuse*).
3) All hyperspectral bands (178) reduced to 28 canonical variates (*CDA-spec*).
4) Seven FFS-selected lidar bands reduced to 5 significant canonical variates (*CDA-lid*).

2.3.5.5 Classification approach

Of the 2,304 crowns included in this study, 25 manually delineated crowns from each of the 29 common species, 725 in total, were randomly selected and permanently set aside for model training, leaving 1,579 for testing at the leaf-type level and 1,291 for testing at the species level. It is necessary to train an object-level classification model using manually delineated crowns in order to assure that the full segment area is composed of one and only one known species (Dalponte et al., 2014). The set of

watershed crowns that spatially aligned with the 725 manual crowns were also excluded from the testing to ensure disjoint training and test sets. Ultimately, the 1,579 manual crowns and the spatially coincident set of watershed crowns were each classified. The manual crowns were classified in order to evaluate the potential classification errors associated with automatic segmentation (aim #3).

To accommodate the dataset's high within-species structural and spectral variation and so to minimize the impact of outlier crowns on discriminant function generation, the 725 training crowns were subsampled with replacement for each of 50 model runs (mr). That is, for each model run, the discriminant functions were generated using (29 species) x (20 crowns/species) x (an average of 4 spectra per crown) = 2,320 fused spectra. Bootstrapping over more model runs (mr = 100) was also investigated but model stability was deemed adequate with mr = 50.

In each model run, discriminant functions were generated based on the current subset of training pixels. This set of 28 (*g-1)* functions was, in turn, multiplied through the run's training and testing datasets to produce the canonical variates. Pixel level LDA classification was carried out on the test set of canonical variates. Upon completion of each model run, a species label was assigned to each crown based on the pixel majority classification. After completion of all 50 runs, the mode crown-level result was calculated and retained for final map creation and accuracy assessment. Pixel-level classifications were also retained for comparison with results from object-oriented approaches.

The final classifications for the 1,579 manual crowns and the spatially coincident watershed segments were mapped in a GIS. A manually delineated ground-reference

map with species information for the 29 common species and leaf-type information for the less common species was used for spatial validation. Watershed crown accuracy was assessed only on a canopy-area basis by spatially intersecting the validation map with the classified segments. Percent correctly classified canopy area has been chosen in lieu of the number of correctly classified stems as the primary method for reporting results for two reasons: First, from an urban forest and ecosystem services management perspective it is more important to gather detailed information on species dominant in the local canopy. Second, it is not feasible to conduct stem-count accuracy assessment when the unit of analysis is the potentially-misaligned crown segment. Still, to better understand the utility of lidar for classifying smaller crowns stem count accuracy was assessed for manual crowns.

2.4 Results

2.4.1 Crown segmentation accuracy

Assessed against field observations, 83% of the watershed segments contained a single tree stem indicating overall good agreement (**Table 2.3**). However, the segmentation accuracy when evaluating only trees from i-Tree Eco plots or from Alameda Park decreased to 55%. This is because Alameda Park, in particular, is a highly complex urban forest setting, with significant crown overlap among trees of all sizes and species (**Fig. 4**). In this type of environment, as evident in figure 4, there is clear omission error. This is likely because the window size of the CMM was determined based on a weak relationship between tree height and width. It is particularly noticeable in this figure that the widths of tall but slender palm trees were not modeled

well leading to inclusion of neighboring stems in their segments. Still, evaluated against segmentation using the CHM, there was a small overall improvement in segmentation accuracy (1%). The improvement may be more pronounced in densely forested areas but this was not evaluated.

Table 2.3: Segmentation accuracy. In bold: 1960 stems were appropriately placed in one watershed crown. Eighty-five segments were without stems.

Stems in segment	0	**1**	2	3	4	5	6	7	8
Segment count	85	**1960**	75	27	10	5	1	1	1

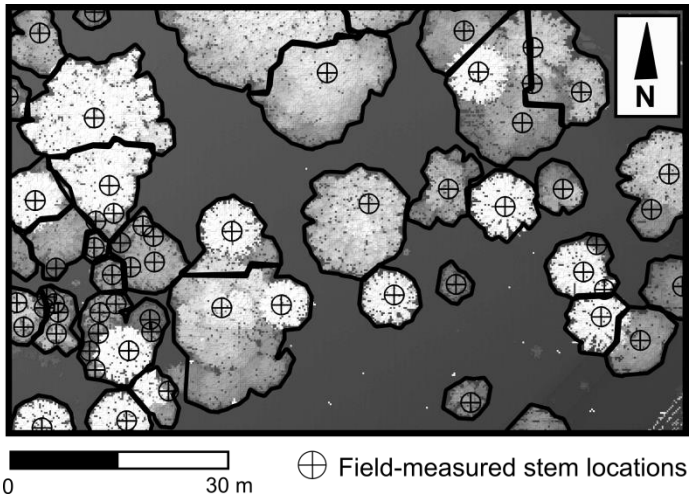

Figure 2.4: Subset of Alameda Park lidar canopy height model shaded by height. Watershed segments and field-measured stems. Note the undersegmentation in the densely vegetated southwest corner

2.4.2 Forward feature selection of structural variables and spectral bands

Using the cross-validated misclassification rate, 7 variables were selected for classifying manual crowns. The same number was selected for watershed crowns. Six of those variables appear in both selection sets, perhaps indicating that watershed crowns and manual crowns can be classified using the same set of structural metrics (**Fig. 2.5**).

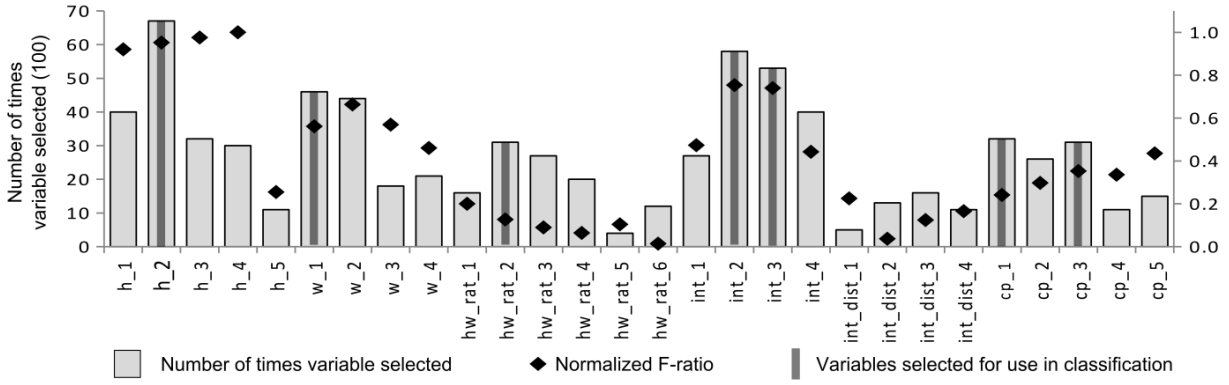

Figure 2.5: Column data show the number of times (out of 100 model runs) that each structural variable was selected using forward feature selection (FFS). The black diamonds represent the normalized F-ratio for each variable. The dark gray vertical stripes indicate that a variable was selected for use in classification of watershed crowns.

The only variables differing between the two sets were *h_1* (max crown height) and *h_2* (median height of returns in crown). For simplicity, and given strong intercorrelation, all further analysis was conducted using *h_2* and the six structural variables selected for both manual and watershed crowns. Overall, variables related to tree height and return intensity stand out with respect to their high between-class to within-class variance as quantified by the normalized F-ratio. When taking variable intercorrelation into account, one height metric (*h_2*), one width metric *(w_1*), one height-to-width ratio metric (*hw_rat_2*), two intensity metrics (*int_2* and *int_3*) and two crown porosity metrics (*cp_1* and *cp_3*) were selected in more than 30% of the watershed crown model runs (**Fig. 2.6**)**.** That the set of variables selected for each set of objects is nearly identical may highlight the success of segmenting an image largely comprising street trees. It may also indicate that the selected variables are robust to minor aberrations in morphology.

Spectral bands were chosen most consistently from the visible region of the spectrum (VIS, 394—734 nm; **Fig. 2.6**). This corresponds with an F-ratio that is

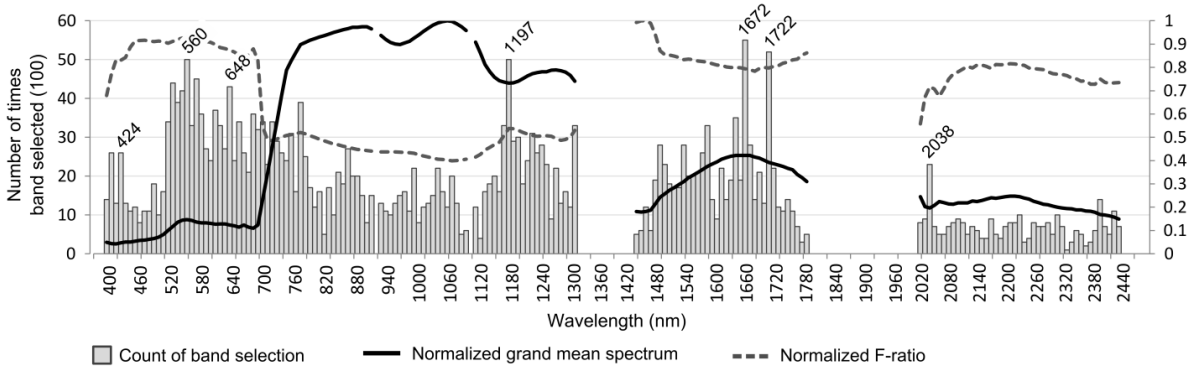

Figure 2.6: Column data show the number of times (out of 100 model runs) that each spectral band was selected using forward feature selection (FFS). The black line is the normalized grand mean spectrum for all pixels from trained (29) species. The dashed line is the normalized F-ratio for each band.

relatively high from 400 nm until the red edge at approximately 700 nm. In particular, bands were selected surrounding the green peak between 520 and 590 nm. Selection frequency in that region was driven by discriminatory power but also by relatively low correlation with neighboring bands in the VIS as well as with bands in the shortwave infrared 2 (SWIR2, 2018—2425 nm; **Appendix A.5**). The near infrared region (NIR, 744—1313 nm) displayed low F-ratios and yielded one band selected with particular frequency in the liquid water absorption feature centered on 1197 nm. The shortwave infrared 1 region (SWIR1, 1443—1802 nm) and SWIR2 regions yielded high F-ratios but only SWIR1 held bands selected in more than 30% of model runs. The lack of band selection from SWIR2 may be a result of high overall correlation with the VIS (r = 0.84) and the SWIR1 (r = 0.83).

2.4.3 Classification results

2.4.3.1 Classification of the 29 common species

The *CDA-7fuse* variable combination yielded the highest overall species-level classification accuracy (83.4% of canopy area, kappa = 82.6) for watershed crowns containing common species (**Fig. 2.7a**). Species-level classification accuracy with only hyperspectral data (*CDA-spec*) was 79.2%. Lidar data only (*CDA-lid*) yielded an accuracy of 32.9%. The best fused result using the manual crowns was 85.4%, suggesting minimal impact on classification accuracy by segmentation error (**Fig. 2.8**). These object-level results compare favorably to 68% pixel-level accuracy in which a single crown could contain several differently classified pixels. The species map accuracy provided by the best fused model (*CDA-7fuse*) was only 4.2 percentage points (pp) better than the spectral-only model (*CDA-spec)*, but there was significant variation when considering the accuracies of individual species (**Table 2.4**, **Fig. 2.9**).

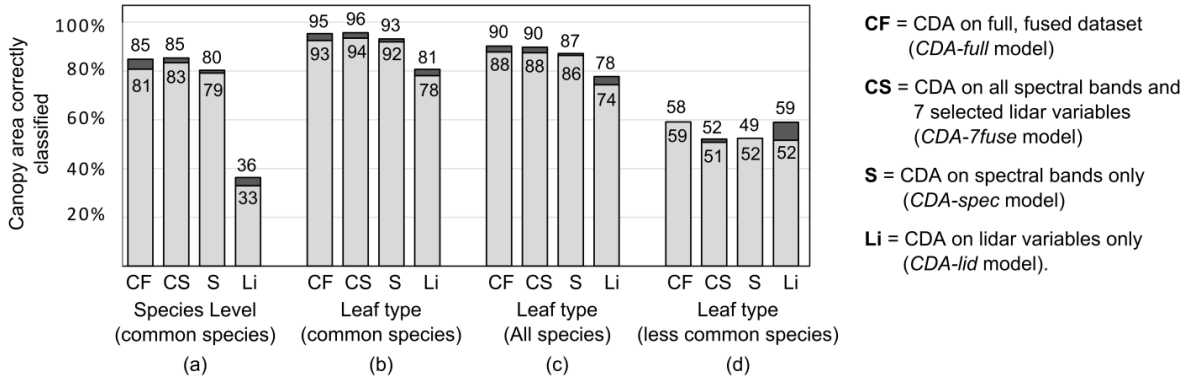

Figure 2.7: Canopy area mapping accuracy. Light gray bars show classification accuracy for watershed crowns while darker bars show the improvement when using manual crowns. (a) Species-level accuracy by model for classifying the 29 common species. (b) Leaf-type level accuracy by model for the 29 common species. (c) Leaf-type level accuracy for mapping all (~100) species. (d) Leaf-type level accuracy for mapping the ~70 less common species

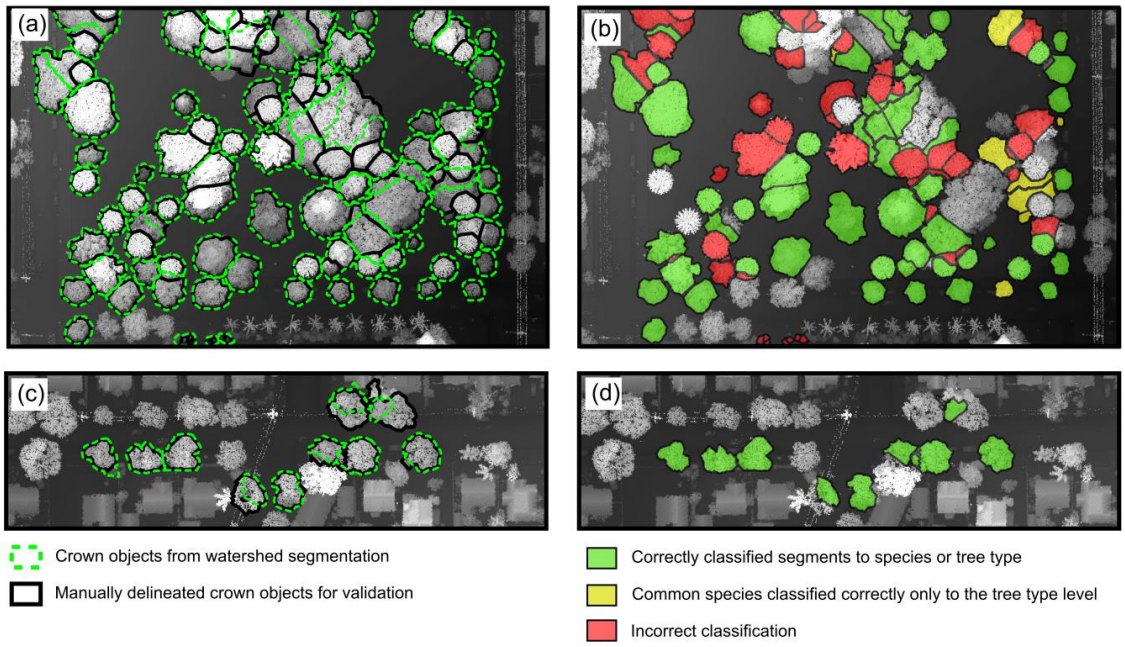

Figure 2.8: Subsets of lidar canopy height model showing: (a) Sample of segmentation results in Alameda Park with significant crown overlap. (b) Classification results in Alameda Park. Trees with no color were in the training set and thus not mapped. (c) Sample of street tree segmentation results. (d) Classification of street trees. Trees with no color were either in the training set or not part of the study.

Large, dense crowned species such as *Ficus microcarpa* (FIMI) were classified well by spectral data alone with an averaged user's and producer's accuracy of 97% (**Table 2.4**). Small broadleaf crowns like *Stenocarpus sinuatus* and *Metrosideros excelsa* were poorly classified by spectral data with average accuracies of 24% and 29% respectively. Small crowned species overall were poorly classified by *CDA-spec* and better classified by *CDA-7fuse* (**Fig. 2.9**). The classification accuracy of species with the 7 smallest crown sizes increased by an average of 17 pp with incorporation of lidar data. Palm species (ARCU, PHCA, SYRO, WARO) were classified using *CDA-spec* with 43% accuracy and by *CDA-7fuse* to 63%. Conifer species (CUMA, PICA, PIPI2) were classified with 84% accuracy using *CDA-spec* and 85% accuracy with *CDA-7fuse*. The

Table 2.4: Producer and User accuracies for CDA on the fused dataset compared to CDA on spectral data only. Refer to **Fig. 2.9** for a graphical depiction of the differences in accuracy.

Species Code	Scientific Name	Tree Type	Stem Count	Canopy Area (m^2)	Fused (*CDA-7fuse*) Prod. (%)	Fused (*CDA-7fuse*) User (%)	Spectral only (*CDA-spec*) Prod. (%)	Spectral only (*CDA-spec*) User (%)
ARCU	*Archon. cunninghamiana*	P	62	756	65	35	16	34
CICA	*Cinnamomum camphora*	B	57	5290	84	83	82	74
CUMA	*Cupressus macrocarpa*	C	55	4857	90	91	94	81
EUFI	*Eucalyptus ficifolia*	B	50	4596	89	61	58	94
EUGL	*Eucalyptus globulus*	B	58	9401	93	98	96	91
FIMI	*Ficus microcarpa*	B	56	9006	92	100	98	96
GEPA	*Geijera parviflora*	B	58	2777	82	69	39	55
JAMI	*Jacaranda mimosifolia*	B	76	6609	95	95	95	93
LIST	*Liquidambar styraciflua*	B	65	5081	87	93	90	77
LOCO	*Lophestemon confertus*	B	66	3465	61	75	82	66
MAGR	*Magnolia grandiflora*	B	63	7425	92	86	89	89
MEEX	*Metrosideros excelsa*	B	62	1581	42	46	26	32
OLEU	*Olea europaea*	B	81	6042	83	94	94	88
PHCA	*Phoenix canariensis*	P	99	5294	64	84	72	80
PICA	*Pinus canariensis*	C	73	4675	85	58	84	69
PIPI2	*Pinus pinea*	C	76	11387	97	89	84	93
PIUN	*Pittosporum undulatum*	B	96	6166	72	95	91	73
PLRA	*Platanus racemosa*	B	71	6933	83	97	86	86
POGR	*Podocarpus gracilior*	B	62	6214	92	86	62	82
PYKA	*Pyrus kawakamii*	B	55	3404	76	58	34	50
QUAG	*Quercus agrifolia*	B	108	8895	89	77	80	87
SCMO	*Schinus molle*	B	53	1971	32	43	40	41
SCTE	*Schinus terebinthifolius*	B	71	5863	93	94	93	85
STSI	*Stenocarpus sinuatus*	B	51	1112	19	18	17	30
SYAU	*Syzygium australe*	B	67	3982	87	80	87	69
SYRO	*Syagarus romanzoffiana*	P	130	2705	36	80	64	17
TISP	*Tipuana tipu*	B	58	7874	99	87	84	88
ULPA	*Ulmus parvifolia*	B	50	6370	69	84	78	51
WARO	*Washingtonia robusta*	P	87	1220	66	76	31	27

remaining 22 broadleaf species were classified with 73% accuracy with *CDA-spec* and improved to 78% accuracy with *CDA-7fuse*. Several species were classified worse using

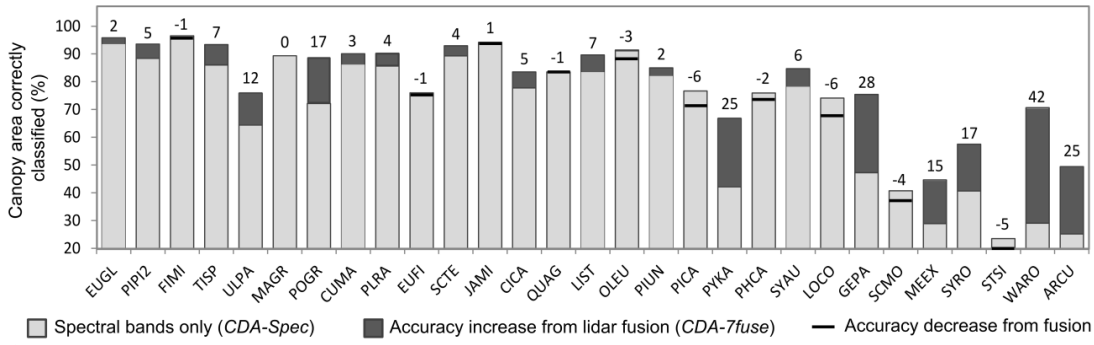

Figure 2.9: Classification accuracy by species for spectral bands only (CDA-spec) and lidar-hyperspectral fusion (CDA-7fuse). Horizontal bars illustrate cases where fusion with lidar reduced accuracy. For species botanical names refer to table 1. Species are sorted by average crown size with the largest species at left.

the *CDA-7fuse* model. Species whose accuracies declined by more than 1 pp included: OLEU, PICA, PHCA, LOCO, SCMO, and STSI.

To further determine the value of adding structural metrics to species classification, we evaluated the success of each variable combination in terms of stem count accuracy. This implies a result that is equally weighted across all species regardless of crown size and could relate to a goal of better understanding the diversity and spatial arrangement of species throughout a city. Stem count accuracy could only be assessed using manual crowns. The stem accuracy with the *CDA-spec* model was 63.0% and increased to a maximum of 71.5% with the *CDA-7fuse* model.

2.4.3.2 Classification to the leaf-type level

Given the 29 common species, the *CDA-7fuse* model reached 93.5% leaf-type accuracy on watershed crowns and 95.7% accuracy on manual crowns (**Fig. 2.7b**). For the same crowns, lidar alone was much more effective at the leaf-type level than at the species level reaching 78.1% accuracy on watershed crowns. Across all (~100) species, *CDA-full* achieved a mapping accuracy of 87.9% to the leaf-type level (**Fig. 2.7c**). When

mapping just the set of ~70 less common species, the accuracies decreased substantially (**Fig. 2.7d**). *CDA-full* mapped watershed segments containing less common species with 59.1% accuracy.

2.5 Discussion

2.5.1 Object oriented approach

Within a single tree crown, leaf-level spectral reflectance may vary considerably as a function of biochemistry and water content (Cochrane, 2000; Ustin et al., 2009). In a given AVIRIS pixel there is further within-class spectral variability driven by canopy architecture, exposure of woody biomass, and exposure of underlying substrate (Roberts et al., 2004; Clark et al., 2005; Asner et al., 2008). This variability potentially manifests in a classification result as a single tree crown containing pixels labeled as multiple species. The problem is exacerbated at fine-spatial resolutions since a given pixel's spectral response may deviate significantly from the crown mean spectrum. Our study reaffirms the utility of object-level analysis for relatively fine resolution (3.7 m) hyperspectral imagery. The overall accuracy for mapping the 29 common species using pixels was 68% and increased to 79% using only spectral information but with pixel majority aggregation at the crown-object level. This increase in accuracy at the object level is in line with previous research: van Aardt and Wynne (2007), classifying 3 pine species from AVIRIS imagery, improved their single pixel results by 20 pp (from 65% to 85%) using a 3x3 window to compute average spectra prior to classification. Clark et al. (2005) found that, on average, 10% of pixels in each *correctly* labeled crown were misclassified.

2.5.2 Contribution to classification: spectral regions and features

In correspondence with previous research, the hyperspectral imagery in our fusion study was the primary driver of classification accuracy (**Fig. 2.7**; Voss and Sugumaran, 2008; Jones et al., 2010; Dalponte et al., 2012). Alonzo et al. (2013) showed that AVIRIS bands spanning the entire solar reflected region (394 to 2425 nm) are required for accurate classification of urban tree species but that the VIS is particularly important. With 29 urban species, this study reconfirms previous research highlighting the discriminatory power and relatively unique spectral information contributed by green peak bands surrounding 560 nm (Fig. 6; Castro-Esau et al., 2006; Pu, 2009, Alonzo et al., 2013). Bands in the green peak region are commonly related to the expression of xanthophyll cycle pigments (Ustin et al., 2009). Chlorophyll absorption regions near 430 and 642 nm were also repeatedly included. We note a continuation of bands selected in >30% of runs (though a diminution of the F-ratio's value) along the red edge where spectral slope and relative spectral reflectance have been previously employed to discriminate tropical rain forest (Cochrane, 2000) and urban (Pu, 2009; Pu and Landry, 2012) tree species. In contrast to the results of Clark et al. (2005), we find limited discriminatory value in the near infrared (NIR, 744 to 1313 nm) range. It is possible that in their tropical rainforest study area they encountered greater between-class diversity with respect to phenology, LAI, and water status. Our results do correspond to those of Dalponte et al. (2012) who found that the NIR region was poorly suited for discrimination due to very high within-class variance. In our study, the one frequently selected band in the NIR was the prominent liquid water absorption band at 1197 nm. The two most frequently selected bands overall (1672 and 1722 nm) were in

the shortwave infrared (SWIR1) region spanning 1443 to 1802 nm. This likely corresponds to species separability driven by variable lignin and cellulose content found in foliar and non-photosynthetic plant matter (Kokaly et al., 2009). The shortwave infrared region (SWIR2) from 2018 to 2425 nm offered little in terms of marginal separability here. A high F-ratio indicates *potential* utility in discrimination but no bands were selected using FFS more than 30% of the time. This may be a product of consistently high sample LAI and leaf water content dampening the viability of lignin-cellulose absorption features (Kokaly et al, 2009). It may also be a product of high correlation among SWIR2, SWIR1, and VIS bands.

2.5.3 Contribution to classification: structural metrics

The selection of particular structural metrics improved classification accuracy compared to retention of all lidar variables by 2.6 pp, though this difference was not tested for statistical significance (**Fig. 2.7a**). Tree height is the single most common lidar variable used in tandem with spectral information as a means to improve classification results (Jones et al., 2010; Dalponte et al., 2008; Koetz et al., 2008). This is due in part to its clear utility in facilitating differentiation among spectrally similar tree species but also because it is simple to measure at either the pixel or crown-object scale and perhaps because it is robust to imperfect image segmentation. Our study also found height metrics to be the most important structural variables based on normalized F-ratio (**Fig. 2.5**). By the same measure, the second most important variable category contained return intensity metrics. In particular *int_2* (average intensity below median height of returns in crown) likely describes the arrangement of leaves and branches in a crown's interior while *int_3* (crown surface intensity: 0.25 m spatial resolution) may

characterize leaf reflectance values along the crown surface (Kim et al., 2009). A previous study found standard deviation of crown intensity values to be the 2nd most valuable variable (Holmgren et al., 2004). They postulated that this metric differentiates densely foliated crowns from those with larger internal gaps. A third grouping of unique variables were those related to crown widths at various heights. Variable *w_1* (crown width at median height of returns in crown) may describe crown morphology in a manner useful for separating upright growth forms from spreading forms. Holmgren et al. (2004) also found that growth form gave rise to important discriminatory variables. They included *segp*, a summary statistic relating to the shape of a parabola fit to the surface of their study crowns, in their final, successful, classification of two conifer species. Finally, we speculate that *cp_1* (surface heights in 0.25 m grid divided by surface heights in 1 m grid) and *cp_3* (count of returns in 0.5 m vertical slice at 90th percentile height divided by width at that height) are two ways to measure crown porosity.

It has been shown previously that structural metrics that are ratios of absolute metrics (e.g. *hw_rat_3*: ratio of crown height to width: 90th percentile height) are useful for species discrimination because they are more invariant to life stage and can capture between-species variability in crown morphology (Holmgren et al, 2004; Kim et al., 2011). Our urban forest study did not corroborate these findings. The derived structural variables in the set *hw_rat* were purposed to crown form description. However, they offered very limited value to species separability compared to their absolute crown-width analogs. Similarly, the *int_dist* family of variables was created out of the supposed need to normalize for uncalibrated intensity values but also offered

minimal discriminatory power. We suggest that ratio metrics were less valuable to this study because the sample set of tree crowns was mostly mature and within-class variance for *absolute* metrics such as heights and widths may have been kept sufficiently low through proactive urban forest management (e.g. pruning, training).

2.5.4 The impact of segmentation on classification accuracy

The impact of imperfect segmentation on classification accuracy was minimal. The decrease in accuracy of the *CDA-7fuse* classifier when moving from manual crowns to watershed crowns was only 2.0 pp. This is likely the case for several reasons: 1) In this urban study area, the segmentation algorithm successfully isolated 83% of the trees. This is, in part, due to a sample set dominated by street trees, which are easier to delineate than heavily-overlapping private property or park trees. Dalponte et al., (2014) showed a 13 pp reduction in accuracy when classifying three tree species in a more complex boreal forest using automatically-delineated lidar segments compared to manual crowns. 2) The basis on which accuracy was evaluated in this study was canopy *area* (i.e. *not* stem count) correctly classified. In this scenario, large crowns could, for example, be oversegmented yet classified correctly due to the classifier's heavy reliance on pixel-level spectral information. 3) Both the resultant classification map and the initial manual delineation of crowns exist only as 2-dimensional overlays on a gridded CHM. As such, even manual crowns are not perfect representations of 3-D crown morphology and exhibit, in a sense, "segmentation errors" in their own right. Three dimensional segmentation of the lidar point cloud itself, where crown assignment takes place via point clustering at the individual return level, is currently possible (e.g.,

Reitberger et al., 2009; Ferraz et al., 2012) and may be useful to implement in future classification projects.

2.5.5 The utility of lidar data

The overall increase in classification accuracy of 29 common species from the inclusion of lidar structural metrics was 4.2 pp. Previous work in complex forested settings has shown improvements from the addition of lidar to hyperspectral data of 1.1 pp for 23 classes (Dalponte et al., 2008), 1.2 pp for 11 species (Jones et al., 2010), and 6 pp for 6 tree species (Dalponte et al., 2012). There are likely several reasons why the addition of lidar data does not dramatically increase overall accuracy. First and foremost, the structural metrics do not drive species separability as much as spectral bands. In this study, only the height metrics surpassed the VIS spectral bands with respect to their relative F-statistics. Additionally, classification accuracy in this study was assessed in terms of canopy area. Species with large canopies can already be well characterized with only hyperspectral information and an object-oriented approach (e.g. Alonzo et al., 2013). Finally, extraction of accurate structural information is likely most difficult for large crowned species due to frequent intermingling with neighboring crowns, irregular crown shapes, and segmentation error (Chen et al., 2006; Kim et al, 2009). Despite minor increases in *overall* classification accuracy, each of the projects referenced above did demonstrate significant improvements in classification accuracy for certain species.

The value of lidar data is evident for small crowned species (**Fig. 2.9**). Of the 8 species whose classification accuracy improved by > 10 pp, 6 (ARCU, GEPA, MEEX, PYKA, SYRO, and WARO) were in the bottom half of the sample set in terms of average

crown area. The aggregated average crown size for those seven species was 30.4 m^2 compared to an overall average of 78.0 m^2. A particularly notable jump in accuracy (+42 pp) was made when the fused data were used to classify *Washingtonia robusta* (**Fig. 2.10**). WARO had the 2nd smallest average crown size and the 2nd highest total tree height. These attributes, on one hand, made WARO difficult to classify using coregistered hyperspectral data both due to its small crown area and because of horizontal crown displacement caused by differing view geometries. On the other hand,

Figure 2.10: *Washingtonia robusta* (Mexican fan palm)

these same attributes made WARO structurally unique, and relatively easy to classify upon inclusion of lidar data. This example highlights three key ways in which lidar likely improves classification accuracies. First, with a lidar point density of 22 pulses/m^2 we have the ability to map much smaller discrete objects than with the hyperspectral data alone. Secondly, since the crown objects were generated using the

gridded lidar dataset, there is no image registration error. Third, unique crown structural characteristics (e.g., height, crown length, crown shape) are not measureable with optical data alone. With increased availability of fine-spatial resolution hyperspectral data (< 1 m) such as AVIRIS Next Generation (Hamlin et al., 2011) or the Carnegie Airborne Observatory (Asner et al., 2007), it will be interesting to see how much classification accuracies can improve even without lidar data.

Some larger crowns were also classified more accurately with *CDA-7fuse* than with spectral data alone. There is evidence that higher crown porosity (possibly relating to lower LAI) may lead to a reduced capacity for accurate classification using spectral data alone. It has been shown that higher LAI strengthens spectral signals in the NIR and portions of the VIS (Asner, 1998). In this study one of the metrics relating to crown porosity was *cp_2*. This metric compares the position of the mean crown height (as a function of tree height and crown base height) to the median height of returns in the crown. Higher numbers suggest a dense upper crown that skews the vertical distribution of lidar returns upwards. Large crowns with the *least* crown porosity by this measure were FIMI and EUGL. FIMI and EUGL were both classified well with spectral data alone (97% and 93% average accuracies respectively) but they were also ranked 3rd and 1st respectively in terms of average crown size. Large and medium crowns with the *highest* porosity were PLRA, PYKA, PICA and ULPA. From the addition of lidar data they gained 4, 25, -6, and 12 pp, respectively. We assume that PICA was ranked highly by this metric more from a combination of upright crown geometry and off-nadir lidar pulses than actual high porosity.

2.5.6 Classification of less common species

The original choice to map 30 species was made because 80% of Santa Barbara's canopy cover comprises roughly 30 species and there appeared to be diminishing increases in canopy for each additional species added after this point (**Fig. 1.3**). In Santa Barbara, 23% of the total canopy cover sampled in the i-Tree Eco field collection was from two species: the native *Quercus agrifolia* (Coast live oak) and the introduced *Syagrus romanzoffiana* (Queen palm). This relationship between species mix and canopy cover may hold in other parts of the country as well. For instance, based on a 2009 i-Tree Eco study in Washington, DC (Casey Trees, 2010), roughly the same relationship was established with 25% of canopy comprised of two native species: *Fagus grandifolia* (American beech) and *Liriodendron tulipifera* (Tulip tree). In Los Angeles, with a very arid climate and a lack of native canopy dominants, the relationship shifts somewhat but 30 species would still equate to roughly 70% of canopy cover. Given an increase in availability of lidar and hyperspectral datasets, these species-canopy relationships indicate the transferability of the methods established in this paper to conduct similar assessments for the canopy dominants in other, larger cities.

In large cities, with established urban forest management programs, it is feasible to collect training data for and map ~30 species to the species level. However, for those, potentially, hundreds of species with low stem counts representing the remaining 20 or 30% of canopy area, it will be pragmatic to classify only to the leaf-type level. In this study, mapping to the leaf-type level meant modeling the less common species as one of the common species and checking for leaf-type agreement. Over the entire dataset of

2,304 crowns (~100 species), leaf-type mapping reached 87.9% accuracy using *CDA-full*. However, when only classifying the ~70 less common species the accuracy declined to 59.1%. We surmise that the low accuracy with which these species were classified is a product of our choice to use a CDA classifier. The classification functions generated were specifically tailored to maximize separability among the input training classes, which did not include the less common species. This leads to a well-tuned classifier for the common species but one that may not be able to capture the variation in the dataset comprising the less common species. Other classification methods may ultimately prove superior for hierarchical classification schemes wherein all trees are classified first to the leaf-type level and then common species are further classified to the species level. For example, Multiple Endmember Spectral Mixture Analysis (MESMA: Roberts et al., 1998) allows for constrained classification based on a target spectrum's similarity to reference spectra such that species not represented in a spectral library would rightly remain unclassified.

2.6 Conclusions

This research sought to improve species and leaf-type level mapping in the urban forest. We first selected 29 common species that dominate the canopy in Santa Barbara, California and classified them using CDA on combined hyperspectral and high point-density lidar data. We achieved a species-level accuracy among trained species of 83.4%. We mapped the *entire* set of sample crowns, including ~70 less common species, to the leaf-type level with 87.9% accuracy. We believe this study demonstrates the potential for separating highly overlapping species classes using data fusion at the

crown-object level. In an immediate, operational sense, the techniques described in this paper are likely applicable with high accuracy (and perhaps with lower point density lidar data) for discriminating among urban vegetation growth forms (e.g. herbs, shrubs, trees) where simple structural metrics could vastly improve separability when combined with either multi- or hyperspectral data. The data to accomplish this sort of classification are available in many cities today and the results even at this generalized level could yield actionable results relating to the spatial distribution of urban ecosystem structure and function.

In line with previous research, classification accuracies in this study were bolstered by lidar variables pertaining to tree height, crown morphology, and perhaps the internal arrangement of leaves and branches. In particular, we showed that small crowns and crowns with unique morphological characteristics were more apt to be correctly labeled with the inclusion of structural data. Further, we showed that classification following automated crown segmentation was more accurate than a pixel-level result and the diminution in accuracy introduced from segmentation error was quite small. As many cities have gained access to high-accuracy canopy coverage maps it is a reasonable next step to implement simple crown segmentation algorithms to generate serviceable crown objects for further analysis. Ultimately, the ability to both map dominant canopy species and inventory common but smaller species is important if we're to operationalize remotely sensed urban forest inventory.

2.7 Acknowledgments

The authors would like to thank the Naval Postgraduate School for funding this research through grant N00244-11-1-0028, "Quantifying the Structure and Function of an Urban Ecosystem Using Imaging Spectrometry, Thermal Imagery, and Small Footprint LiDAR". Special thanks also to: Seth Gorelik, Keri Opalk, Alex Sun, Randy Baldwin, and Matt Ritter for their assistance with field data collection; Joe McFadden for supporting field work and helping to refine the manuscript; the City of Santa Barbara for providing ground reference data; Keely Roth and Seth Peterson for help creating and fine-tuning the technical methods; and the four anonymous reviewers for their valuable comments and suggestions.

Chapter 3

Mapping urban forest leaf area index with airborne lidar using penetration metrics and allometry

Authors:

Michael Alonzo[a], Bodo Bookhagen[b], Joseph P. McFadden[a], Alex Sun[a], and Dar A. Roberts[a]

[a] Geography Department, Ellison Hall 1832, UC Santa Barbara, CA 93106-4060, USA

[b] University of Potsdam, Institute of Earth and Environmental Science, Karl-Liebknecht-Str. 24-25, 14467 Potsdam-Golm, Germany

Citation for this chapter as published in *Remote Sensing of Environment*:

Alonzo, M., Bookhagen, B., McFadden, J. P., Sun, A., & Roberts, D. A. (2015). Mapping urban forest leaf area index with airborne lidar using penetration metrics and allometry. *Remote Sensing of Environment*, 162, 141–153.

3.1 Abstract

In urban areas, leaf area index (LAI) is a key ecosystem structural attribute with implications for energy and water balance, gas exchange, and anthropogenic energy use. In this study, we estimated LAI spatially using airborne lidar in downtown Santa Barbara, California, USA. We implemented two different modeling approaches. First, we directly estimated effective LAI (LAI_e) using scan angle- and clump-corrected lidar laser penetration metrics (LPM). Second, we adapted existing allometric equations to estimate crown structural metrics including tree height and crown base height using lidar. The latter approach allowed for LAI estimates at the individual tree-crown scale. The LPM method, at both high and decimated point densities, resulted in good linear agreement with estimates from ground-based hemispherical photography (r^2 = 0.82, y = 0.99x) using a model that assumed a spherical leaf angle distribution. Within individual tree crown segments, the lidar estimates of crown structure closely paralleled field measurements (e.g., r^2=0.87 for crown length). LAI estimates based on the lidar crown measurements corresponded well with estimates from field measurements (r^2 = 0.84, y = 0.97x+0.10). Consistency of the LPM and allometric lidar methods was also strong at 71 validation plots (r^2 = 0.88) and at 450 additional sample locations across the entire study area (r^2 = 0.72). This level of correspondence exceeded that of the canopy hemispherical photography and allometric, ground-based estimates (r^2 = 0.53). The first-order alignment of these two disparate methods may indicate that the error bounds for mapping LAI in cities are small enough to pursue large scale, spatially explicit estimation.

3.2 Introduction

Urban trees provide a broad array of ecosystem services that are governed by tree species, canopy structure, and locational context (Escobedo & Nowak, 2009; Manning, 2008; McCarthy & Pataki, 2010; McPherson et al., 2011; Simpson, 2002; Urban, 1992). Leaf Area Index (LAI), commonly defined as one half of the total green leaf area per unit ground area (Chen & Black, 1992), is a critical structural attribute that has implications for urban energy balance, gas exchange, hydrological throughput, and anthropogenic energy use. It is an ecophysiological measure of leaf surface available for photosynthesis and transpiration (Chen et al., 1997). In addition, dry depositional uptake and intercellular suspension of air pollutants such as O_3, NO_2, SO_2, CO, and PM_x is partly mediated by effective leaf surface area (Baldocchi et al., 1987; Hirabayashi et al., 2011). In urban areas, this process has been related to spatial variation in air pollution reduction (e.g.,Escobedo & Nowak, 2009). Increased canopy leaf area, especially over paved surfaces, delays stormwater peak flow through interception of precipitation (Xiao & McPherson, 2002). Higher urban vegetation fractional cover (Lu & Weng, 2006; Myint et al., 2010) and higher LAI (Georgi & Zafiriadis, 2006; Hardin & Jensen, 2007; Oke, 1989; Peters & McFadden, 2010) have been correlated with lowered urban temperatures and reduced summertime building cooling costs. At the same time, tree cover has also been linked to ecosystem disservices ranging from pollen allergies to sidewalk damage and the production of litterfall (Roy et al., 2012).

Many cities have estimated urban LAI using the USDA Forest Service's i-Tree Eco model (Nowak et al., 2008a). The i-Tree Eco model produces estimates of urban forest structure, including LAI, and ecosystem function using field measurements of

tree species and crown dimensions acquired on ≥200 stratified random inventory plots across a city (Nowak et al., 2008a). The resulting estimates of ecosystem function are used by cities for urban forest management and planning (e.g., Million Trees LA: McPherson et al., 2011). However, the data collection process is labor intensive, and the results are only available at very coarse spatial resolution. Further, the LAI estimates become increasingly uncertain in regions where the model's allometric equations have not been parameterized by locally-evaluated, species-specific coefficients (Gower et al., 1999; Peper & McPherson, 2003). By contrast, the estimation of *effective* LAI (LAI_e) in an urban area from hemispherical photography (hereafter "hemiphotos") may be more robust to the varying mixtures of tree species than allometric methods. LAI_e differs from true LAI in that it does not account for the non-random distribution of foliage throughout the canopy and does not differentiate between foliar and woody plant components (Chen & Black, 1991). However, measurement challenges such as discontinuous canopy cover, variability in canopy height, occlusion of foliage by buildings and other structures, and difficulty of accessing private property at times when sky conditions are appropriate for the method have limited the use of this technique in cities (Jensen et al., 2009, 2012; Osmond, 2009; Peper & McPherson, 2003; Richardson et al., 2009b). Importantly, both allometry and hemispherical photography are field-sampling techniques that generate only point estimates of LAI that cannot easily be extended to a citywide map. Remote sensing data can be used to estimate and map urban LAI and LAI_e over large areas at fine spatial scales, possibly at significant cost savings compared to field campaigns (Nowak et al., 2008b).

Maps of LAI_e in natural forest settings are frequently produced using laser penetration metrics (LPM) calculated from airborne lidar (e.g.,Hopkinson et al., 2013; Korhonen et al., 2011; Solberg et al., 2009; Zhao & Popescu, 2009). LPMs, which report the penetration ratios of laser pulses through canopy, are favored in part due to the theoretical reliance on Beer-Lambert's law of light attenuation that they share with gap fraction calculated from hemiphotos. However, issues related to multi-scale clumping of foliage, the variable relationship between sensor scan angle and canopy path length, and the wide range of possible leaf angle distributions due to species diversity have largely precluded lidar mapping of LAI_e in heterogeneous areas (Holmgren et al., 2003; Morsdorf et al., 2008; van Gardingen et al., 1999). Despite these limitations, Richardson et al. (2009) showed that mapping LAI_e was possible in a biodiverse urban park and that the assumption of a spherical leaf angle distribution may be acceptable.

In this study we sought to improve LAI mapping capabilities in heterogeneous urban environments. We used two theoretically distinct modeling approaches and multiple types of validation evidence. It is important to acknowledge that indirect, ground-based measurements of LAI or LAI_e are problematic, exhibiting variability and bias with respect to true LAI and each other (Bréda, 2003; Peper & McPherson, 2003). We first examined the relationship between lidar estimates of LAI_e using a Beer-Lambert style approach and estimates from hemiphotos acquired at 71 field plots. Second, we adapted the allometric equations used in the i-Tree Eco model for use with crown dimension measurements (e.g., height, diameter) taken at the individual tree crown scale (hereafter "crown scale") using lidar. The specific objectives of this study were:

1. Map LAI_e in a heterogeneous, urban landscape at the field-plot scale through correlation of LPM derived LAI_e and hemiphoto gap fraction inversion.
2. Introduce methods for mitigating the effects of off-nadir lidar pulse angles and non-random foliage distribution on estimates of LAI_e in discontinuous canopy.
3. Map LAI of individual trees using automatically delineated crown objects, lidar-measured crown dimensions, and an allometric approach.
4. Compare plot-aggregated allometric LAI outputs with the plot-level outputs from the LPM method to characterize the covariation.

We anticipated that the plot-level metrics based on Beer-Lambert's law would offer a site-transferable means to estimate LAI_e with minimal model calibration from field data. This output could be useful for broad assessment and modeling of urban surface energy balance in terms of heat, moisture, and momentum fluxes (Grimmond et al., 2010) However, the resultant map resolution will not allow for estimates of urban tree ecosystem service provision in the manner desired by many cities (i.e., services that depend on crown location relative to buildings and impervious surfaces). Crown scale estimates of LAI validated against i-Tree Eco allometry offer a more direct path towards a spatially explicit urban forest inventory albeit one that internalizes the uncertainties of the i-Tree Eco model.

3.3 Materials and Methods

3.3.1 Study area and field plots

This study was conducted in downtown Santa Barbara, California (34.42° N, 119.69° W) **(Fig. 3.1)**. Santa Barbara is a city of about 90,000 residents, encompassing 51 km^2, located on a coastal plain between the Pacific Ocean to the south and the Santa Ynez mountains to the north. It has a Mediterranean climate and supports a diverse mix of native, introduced, and invasive urban forest species. Fractional canopy cover (*fCov*) was estimated in 2012 at 25.4% for the entire municipality of Santa Barbara using high-resolution digital imagery (City of Santa Barbara Urban Forest Management Plan, 2014, www.santabarbaraca.gov). Our study area was situated in the most densely built portion of the city and, according to i-Tree Eco estimates in 2012, *fCov* was approximately 20%.

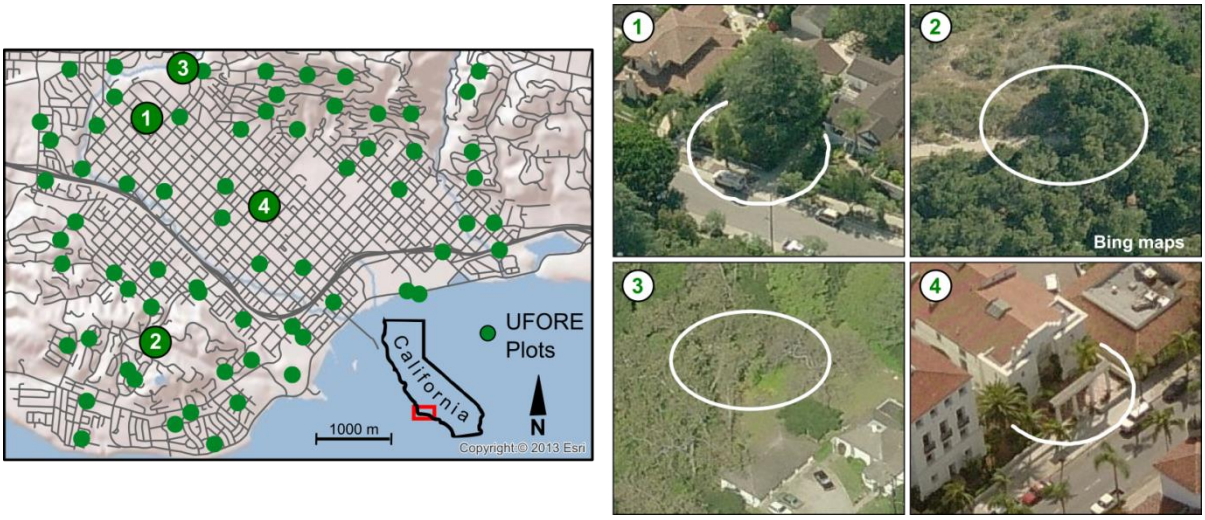

Figure 3.1: Study area located in downtown Santa Barbara, California. Green dots show grid-randomized distribution of 71 field plots. The shaded topographic relief background shows that most plots are on flat ground but some, in the northeast and southwest are on steep slopes. At right, oblique angle aerial imagery courtesy of Bing Maps showing 4 representative 11. 4 m radius plots.

In the fall of 2012, we inventoried vegetation within 105 plots, recording 108 unique species. The most commonly sampled species were the broadleaf persistent native *Quercus agrifolia* (Coast live oak) and the introduced *Syagrus romanzoffiana* (Queen palm). Each plot **(Fig. 3.1)** had a radius of 11.4 m in accordance with i-Tree Eco data collection protocols (i-Tree Eco User's Manual v. 4.1.0, www.itreetools.org). Species composition and structure in the plots was extremely heterogeneous: Thirty-eight plots had LAI values of < 1 and 10 plots had values >3 (mean = 1.39). Average canopy height was also highly variable, ranging between 2 and 23 meters with significant internal variation as well. The number of trees per plot ranged between 1 and 57 with a median stem count of 4 trees. Plot centers were geolocated using differentially corrected GPS and were positionally accurate to 30 cm with respect to the lidar data. The distance and direction of each stem was measured from plot center with an Opti-Logic laser range finder and a compass. Of the 105 sampled plots, 71, with colocated field measurements, hemiphotos, and lidar data were retained for analysis.

3.3.2 LAI_e estimates from hemispherical photography

To characterize the extreme heterogeneity in urban forest gap fraction, one hemiphoto was acquired at plot center along with four additional photos 5.5 m from the center in each cardinal direction. Plots where only one photo site was accessible represented 16% of the total. While best practice dictates that hemiphotos are acquired under diffuse light conditions, this was not always possible. Southern California autumn days are frequently cloudless and the high likelihood of a field plot falling on private property limited our flexibility in acquisition time. The photos were taken at 1 m above ground using a Nikon Coolpix 5400 digital camera retrofitted by removing the

manufacturer's infrared-blocking filter and replacing it with a filter that blocked wavelengths <590 nm such that it could record red and infrared light. The modified camera was used because hemiphotos acquired with near infrared (NIR) wavelengths can lead to more efficient and accurate image binarization of foliage (Chapman, 2007). This advantage is important in urban settings where hemiphotos frequently contain structures interspersed with foliage (Osmond, 2009).

At each hemiphoto location, we acquired images at three exposure settings: 1-stop underexposed, automatic exposure, and 1-stop overexposed. We combined these multiple exposures into a single high-dynamic range (HDR) image to enhance contrast between foliage and sky (Jonckheere et al., 2005; Zhang et al., 2005) and mitigate pixel saturation caused by direct beam radiation (Korhonen et al., 2011). HDR processing was completed with minimal changes to default settings in Dynamic-Photo HDR 5 (v 5.2.0). Foliage, plant stems, and branches were distinguished from all other scene components through an image segmentation rule-set applied to the hemiphotos using Trimble's eCognition software (v. 6.4, Munich, Germany). Multi-resolution segmentation and locally-thresholded classification allowed us to address variability in scene illumination and the complex mix of biotic and abiotic scene elements (Jonckheere et al., 2004). Nevertheless, as these photos were used for validation of remotely sensed estimates, some manual editing was required.

Gap fraction was calculated from the binary images at zenith bin midpoints (7°, 23°, 38°, 53°, and 68°) that, for compatibility with previous research (e.g., Korhonen et al., 2011; Solberg et al., 2009) , correspond with the concentric detector rings of the LAI-2000 Plant Canopy Analyzer (Li-Cor, Lincoln, Nebraska, USA). LAI_e can be

calculated from gap fraction using a discrete approximation of Miller's integral (Miller, 1967; Korhonen et al., 2011)**:**

$$LAI_e = 2\sum_{i=1}^{n} -\ln(\overline{P}_i)\cos(\theta_i)\, w_i \qquad (3.1)$$

where $\overline{P}_i$ are ring-wise gap fractions as functions of zenith angle (θ) averaged across all photo sites at a given plot (Ryu et al., 2010a) and w_i are weights corresponding to $\sin(\Theta)d\Theta$ of the midpoint angle of each zenith bin (Solberg et al., 2006).

3.3.3 LAI estimates from i-Tree Eco allometry

At the full set of 105 plots we identified and measured 612 trees following standard i-Tree Eco protocols. The crown measurements used by the i-Tree Eco model for estimation of leaf area include height of live top, crown base height, and average crown diameter. The log-linear allometric equation, initially created for full-crowned, deciduous, open-grown trees (Nowak, 1996) is:

$$\ln(LA) = -4.3309 + 0.2942L + 0.7312D + 5.7217S - 0.0148C + \epsilon \qquad (3.2)$$

where LA is leaf area, L (crown length) is equal to the height of crown live top minus crown base height, D is the average crown diameter, S is a species-specific average shading factor, C is the crown's outer surface area represented by: $\pi D(L+D)/2$ and ϵ is an error term. Following back-transformation and correction for logarithmic bias, further adjustments may be made in cases of crowns with dimensions beyond the limits for which the equations were developed and for crowns that exhibit leaf loss due to factors such as dieback, defoliation and pruning (D.J. Nowak, pers. comm., 2014). Plot level LAI estimates were computed by summing leaf area for all trees measured on the plot and then dividing by the total plot area of 408 m^2.

3.3.4 Lidar data acquisition and processing

Waveform lidar data were collected in August of 2010 aboard a helicopter with a front-mounted Riegl Q560 laser scanner (Riegl USA, Orlando, Florida). The lidar data were georeferenced with two local differential GPS stations and stored in the UTM coordinate system (Zone 11N, NAD83). The waveform was discretized using standard Riegl processing procedures to an average last-return point density of 22 points m^{-2} across the study area with additional returns (maximum of 4) available in tall vegetation. Height values on flat surfaces were evaluated to be precise to within 2 cm. Nominal scan angles ranged between 0 and 30° but the front-mounted sensor configuration resulted in a minimum *pulse angle* of 10° and a maximum of 30°. In this research the term *pulse angle* refers to the three dimensional, angular deviation from a theoretical pulse traveling perpendicularly to the ground. This is measured by constructing the 3-D line between a last return and a first return connected by their shared GPS time (Zhao & Popescu, 2009).

Each of the 71 plots was entirely sampled by at least two lidar flight lines and thus, in most cases, by multiple pulse angle distributions. Pulse angle was assigned to single echo pulses from multi-return neighbors in the same flight line or, in the absence of multiple returns, directly from the nominal scan angle. The point cloud was classified to ground, building, and vegetation using LAStools (LAStools v111216, http://lastools.org) with minimal adjustments to default settings and an overall classification accuracy, validated via manual image interpretation, exceeding 97%.

3.3.5 Plot-level estimates of LAI_e from lidar

We extracted 2.5, 5, 10, 11.4, 15, 20, and 25 m radius cylindrical subsets of the lidar point cloud centered at each hemiphoto site (n = 243) acquired across our set of 71 plots (methods workflow**: Fig. 3.2**). Hemiphotos and lidar pulses do not sample the same canopy (Morsdorf et al., 2006; Richardson et al., 2009b). The former is upward

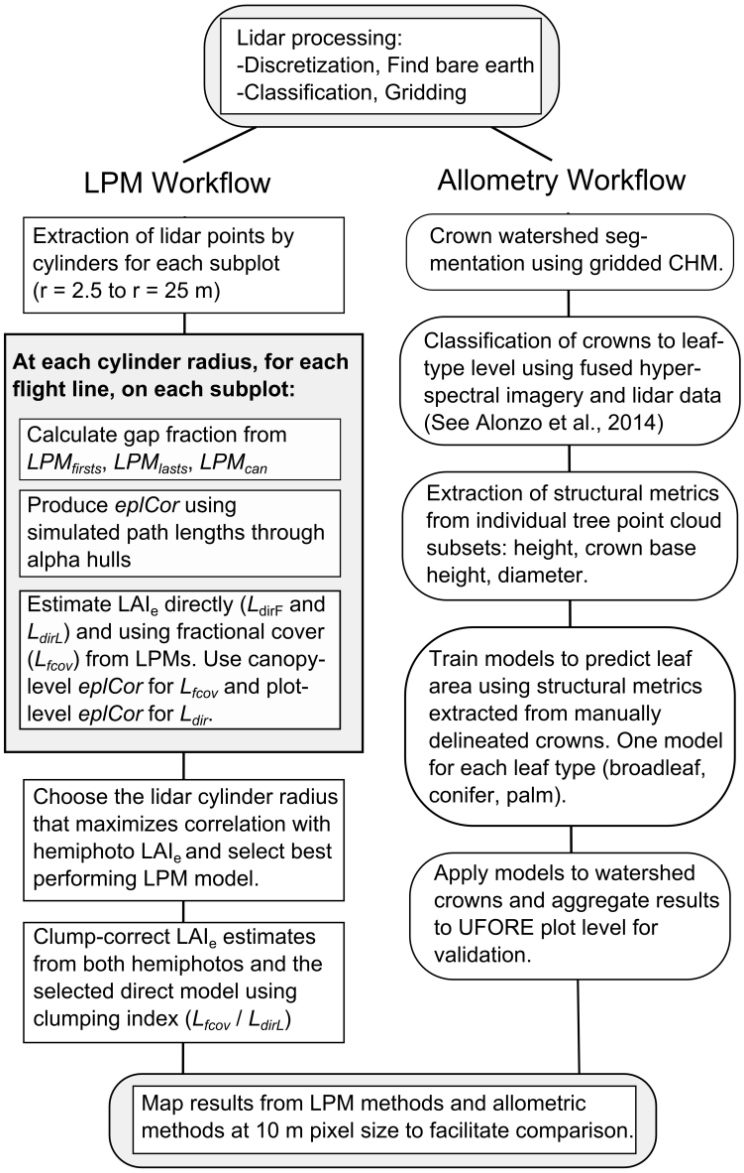

Figure 3.2: Workflow diagram for both mapping approaches described in detail in the text. Please refer to **Table 3.1** for full expansion of the acronyms in this figure.

looking with a conical field-of-view while the latter is downward looking and results in a cylindrical extrusion of the field plot boundary. In an urban setting with highly variable canopy height, the ideal angular restriction of hemiphoto view zenith and optimal sizing of the lidar cylinder radius is not known *a priori*. Thus, we extracted data at multiple radii in order to determine, on average, which cylinder size maximized the correlation between estimated LAI_e and the hemiphoto LAI_e estimates that were produced from each of the five aforementioned zenith angle ranges.

Table 3.1: Key notation used throughout study.

Notation	Explanation
LPM	Laser Penetration Metric
LAI	Refers to "true" Leaf Area Index or Leaf Area Index from allometry
LAI_e	Effective Leaf Area Index
LPM_{firsts}	LPM gap fraction formulated only using first returns (all pulses on plot)
LPM_{lasts}	LPM gap fraction formulated with inclusion of last returns (all pulses on plot)
LPM_{can}	LPM gap fraction for canopy only (only includes pulses intersecting with canopy)
fCov	Plot canopy fractional cover
L_{dirF}	Plot-level, direct LAI_e using LPM_{firsts}
L_{dirL}	Plot-level, direct LAI_e using LPM_{lasts}
L_{fCov}	Plot-level LAI_e using LPM_{can} multiplied by *fCov* (used for clumping correction)
eplCor	Expected path length correction

3.3.5.1 Gap fraction from Laser Penetration Metrics

Accurate gap fraction estimation from LPMs is dependent on canopy gap size and arrangement, acquisition parameters such as beam footprint and scan angle, and the specific formulation of the LPM (Hopkinson et al., 2013; Morsdorf et al., 2008; Zhao & Popescu, 2009). We implemented three frequency-based LPMs as proxy measures of gap fraction **(Table 3.1)**. The first two were calculated using all pulses extracted from a given plot (hereafter: “direct” method) and are adaptations of an LPM applied in a

natural forest setting by Solberg et al. (2006) and a managed park setting by Richardson et al. (2009). Prior to correction for clumping (discussed in the following section), the direct method will yield significant overestimates of gap fraction whenever foliage is clumped among spatially disaggregated trees. However, this is likely similar to the overestimation produced from hemiphoto analysis. We formulated the first LPM, simply as the inverse of fractional cover, only counting first returns, as follows:

$$LPM_{firsts} = \ 1 - fCov \tag{3.3}$$

Where:

$$fCov = \frac{F_c}{F_g + F_c} \tag{3.4}$$

In this model description, F denotes first and only returns, the subscript g is for ground, and subscript c is for canopy. When referring to lidar measurement from near overhead, there is no universally accepted differentiation between $1 - fCov$ and gap fraction. They may be considered equivalent (Hopkinson & Chasmer, 2009) or they may be distinguished in terms of the size of the gaps in question (Carlson & Ripley, 1997). To increase sensitivity to smaller gaps, gap fraction retaining last ground returns was calculated as the sum of first and last returns at ground level ($L_g + F_g$) divided by first ground, last ground, and first canopy returns:

$$LPM_{lasts} = \frac{L_g + F_g}{L_g + F_g + F_c} \tag{3.5}$$

A third LPM was calculated *only in and under canopies* as the ratio of ground last returns to the sum of ground last returns and first canopy returns:

$$LPM_{can} = \frac{L_g}{L_g + F_c} \tag{3.6}$$

This formulation does not include single returns that have penetrated through the canopy via larger gaps as these were accounted for in **equation 3.4** for fractional cover. Logarithmic inversion of the direct measures, LPM_{firsts} and LPM_{lasts} will result in direct estimates of plot-level LAI_e. The same inversion of LPM_{can} will result in an estimate *only of canopy* LAI_e. This result must be rescaled by *fCov* in order to yield a comparable, plot-level result. In this study LAI_e from LPM_{can} was only used to correct the two direct estimates, as well as hemiphoto-estimated LAI_e, for plot-scale clumping. To evaluate whether these metrics may be implemented with more commonly available lidar datasets, each LPM was additionally calculated for decimated pulse densities of 5 pts m^{-2} and 2 pts m^{-2}.

3.3.5.2 Pulse interception simulation for path length and clumping correction

Gap fraction estimates from LPMs will vary as a function of path length through the canopy. This component of variability is commonly mitigated through retention only of pulses with near-nadir scan angles (e.g., Morsdorf et al., 2008) but has also been accounted for using a cosine correction, which effectively normalizes path length by pulse angle (e.g., Zhao & Popescu, 2009). Relying on narrow swaths of data in cities is problematic if we hope to operationalize spatially extensive LAI_e measurements at reasonable cost. However, the application of a simple cosine correction (e.g., $1/\cos(\Theta)$ where Θ = pulse angle) may be inappropriate in areas of discontinuous canopy and variable crown morphology (Holmgren et al., 2003). Pulse angles in this study's lidar dataset were never less than 10 degrees off-nadir. Thus, in furtherance of objective #2, we simulated canopy pulse interception through a range of pulse angles and crown geometries to develop a more precise method for correction called *expected path length*

correction (*eplCor*). Simulator detail and code are available in **Appendix B** Portions of this MATLAB-based (vR2013b, The MathWorks, Inc.) simulator code were also used for the "on-the-fly" simulation using the lidar data referenced in the following section.

eplCor is the ratio of the expected pulse path length through the crown at a given pulse angle and crown geometry compared to the hypothetical, simulated nadir path length. *eplCor* is assessed at the canopy level by comparing the path lengths only of pulses intersecting the crown's alpha hull. Alpha hulls differ from convex hulls in that the requirement of convexity is relaxed when the user-defined alpha scale parameter is finite. Use of alpha hulls allows for more precise wrapping of a set of spatially disaggregated crowns. At the plot level, *all* plot pulses are included to account for the fact that, with upright crown geometry, as pulse angle increases, more pulses will pass through the crown. We examined the relationship between pulse angle and LAI_e for upright, intermediate, and spreading geometries based, respectively, on 75th, 50th, and 25th percentile height-to-diameter ratios of our 612 field-measured trees. Results are reported as "simulated LAI_e ($sLAI_e$)" and produced from equations following the form of **Eq. 3.7** for plot-level estimates and **Eq. 3.9** for canopy-level estimates rescaled by *fCov* (see **Section 3.3.5.3**).

The extent to which estimates of LAI_e will underestimate true LAI due to clumping is also partially determined by urban forest stand configuration. In an urban setting characterized by isolated trees, clumping of foliage occurs at the shoot, branch, *and* plot scales. While the error from clumping is presumed to be the same for both lidar and hemiphoto estimates, we believe that tree spacing in our urban study area will lead to underestimates that may be too large to ignore. Here, we again used pulse

interception simulation to examine how the distribution of foliage throughout a plot, from highly clumped (e.g., aggregated in a single, dense palm) to fully random, drives underestimates of direct LAI_e measures. To accomplish this, we set true plot LAI to values ranging from 0.11 to 1.67. At each value we varied the radius of a simulated tree, and thus the foliage density, from 3 m to 15 m. To produce a clumping ratio we divided simulated LAI_e calculated using **Eq. 3.9** by simulated LAI_e calculated using **Eq. 3.7 (See Section 3.3.5.3)** . This ratio presumes that a plot-level estimate of LAI_e made by first measuring only canopy LAI_e and then rescaling by *fCov* will not be impacted by the spacing / arrangement of the plot's trees (Ryu et al., 2010a).

3.3.5.3 Estimating LAI_e from lidar data

For each cylinder radius, at each hemiphoto site, for each flight line, LAI_e was calculated from lidar data directly at the plot level using only first returns (L_{dirF}, where the subscript F indicates first returns) and including last returns (L_{dirL} where the subscript L indicates last returns) as well as indirectly from canopy LAI_e multiplied by *fCov* (L_{fCov}). Formulations:

$$L_{dirF} = \frac{-\ln(LPM_{firsts})}{k \cdot eplCor} \quad (3.7)$$

$$L_{dirL} = \frac{-\ln(LPM_{lasts})}{k \cdot eplCor} \quad (3.8)$$

$$L_{fCov} = \frac{-\ln(LPM_{can})}{k \cdot eplCor} \cdot fCov \quad (3.9)$$

where LPM_x is the relevant gap fraction generated using the LPM **equations 3.3, 3.5 and 3.6.** We set k = 0.5 to correspond with the commonly reported spherical leaf angle distribution for leaves of any size (Chen et al., 1997). Given the high species diversity in our study area there was little basis on which to choose any LAD other than spherical.

We also report k as estimated using weighted linear regression and a no-intercept model to evaluate the deviation from this expected value (e.g., Solberg et al., 2006). Linear regression weighted by lidar-estimated LAI_e was selected in order to account for increasing error variance at higher values of LAI_e. Coefficients of determination were calculated based on sums of squared error and total sums of squares for the untransformed data relative to the weighted least squares fit. For the airborne lidar data, *eplCor* was estimated using on-the-fly simulation. That is, for each of the 243 subplots, an alpha hull was generated for canopy returns and expected path lengths were calculated based on median pulse angle, median pulse azimuth, and crown geometry. The flight line results, weighted by number of pulses, were averaged back to the full plot (n = 71) level. Clumping ratios were calculated and multiplied through L_{dirF}, L_{dirL} and the hemiphoto LAI_e results to account for plot-scale clumping.

3.3.6 Crown scale estimates of LAI using lidar-measured crown dimensions

Objective #3 of this study was to map LAI at the crown scale using allometric methods. Note: This study does *not* use LAI_e and LAI interchangeably. The latter metric represents an estimate of "true" LAI that is neither impacted by foliar clumping nor commingling of woody and leafy plant materials. The choice to forgo LPM methods in favor of allometry for mapping LAI at the crown scale is warranted for several reasons: 1) Currently, allometric estimation using the i-Tree Eco model is the most common means for estimating LAI in US cities; 2) LPMs can only be used when it is possible to tally the full set of ground-reflected pulses that have passed through a given tree crown. This is difficult with highly off-nadir pulse angles where the (X,Y) positions of ground returns are displaced significantly with respect to the (X,Y) positions of canopy returns;

3) Even with high-point density and low flight altitude, individual tree crowns with high LAI may not allow passage of *any* pulses, resulting in numerical overflow of the gap fraction inversion computation; 4) Objective #4 of this study was to compare the results of mapping methods with distinct theoretical underpinnings.

The following sections describe in detail how we used lidar to estimate the same structural inputs that are used in the i-Tree Eco allometric equation **(Eq. 3.2)** and then how we applied the results to all crowns in our study area **(Fig. 3.2)**.

3.3.6.1 Crown segmentation and leaf type classification

The allometric equation **(3.2)** used in i-Tree Eco to estimate leaf area from crown dimensions also incorporates a species-specific shading factor. This factor is included to account for the large species-driven variation in leaf size, shape, and arrangement that may be encountered within any measured crown volume. In this study it was not feasible to develop coefficients for each species due to limited training data. Instead we classified each tree to the leaf-type level (i.e., broadleaf, needle leaf, palm). Segmentation of canopy into individual crowns and leaf-type classification was undertaken prior to beginning this study. The process is detailed in Alonzo et al. (2014). A brief synopsis follows:

Canopy segmentation made use of the marker-controlled watershed algorithm on a gridded lidar canopy height model in the manner originally proposed by (Chen et al., 2006). Spectral data from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS, Green et al., 1998) and structural metrics (e.g., height, height-to-width ratios, porosity) were extracted from each crown and fused. Classification of 29 common species was carried out on the fused dataset using canonical discriminant analysis to

83.4% overall accuracy. Classification of all crown segments to the leaf-type level was carried out with 93.5% accuracy. This classification information was used in **Section 3.3.6.3** to facilitate the formulation of separate leaf area models, one for each leaf type.

3.3.6.2 Measurement of crown dimensions from the lidar point cloud

In this study, tree height and crown base height were estimated directly from the 3-D lidar point cloud by first finding the vertical midpoint between the highest and lowest return in a given segment that was classified as canopy. From the midpoint, a 0.25 m window was moved vertically up and down the tree, stopping when the slice contained fewer than a predetermined minimum number of returns. This method proved superior to a quantile-based method because of the significant presence of cultivated shrubbery below the crown **(Fig. 3.3a)** and power lines or branches of neighboring trees above **(Fig. 3.3b)**. Average crown diameter was calculated from watershed crown area after abstracting each segment to its circle of equivalent area. These measurements served as the raw inputs to **Eq. 3.2**.

3.3.6.3 Model formulation using crown measurements from manually delineated crowns

As in the i-Tree Eco model**,** total leaf area was estimated as a precursor to LAI. We trained the lidar model to predict leaf area using 109 manually delineated crowns with known leaf type and leaf area. We used a log-log adaption of **Eq. 2** due to the lognormal frequency distributions of each independent variable, the lognormal distribution of leaf area, and clear non-linear, bivariate relationships between each predictor and response:

$$\ln(LA) = \ln(b_0 + b_1 \cdot L + b_2 \cdot D + b_3 \cdot C) \quad (3.10)$$

where LA is leaf area, L is crown length, D is average crown diameter, C is crown surface area, and b_x are empirically determined coefficients. Coefficient estimation was carried out on both the pooled set of 109 crowns and the sets separated by leaf type using weighted least squares regression.

3.3.6.4 Model application on watershed crowns

The models formulated using manual crowns were applied to all watershed segments in the study area after estimating the crown dimensions for each segment. Following segment-level estimates of leaf area, LAI was calculated at the field-plot level

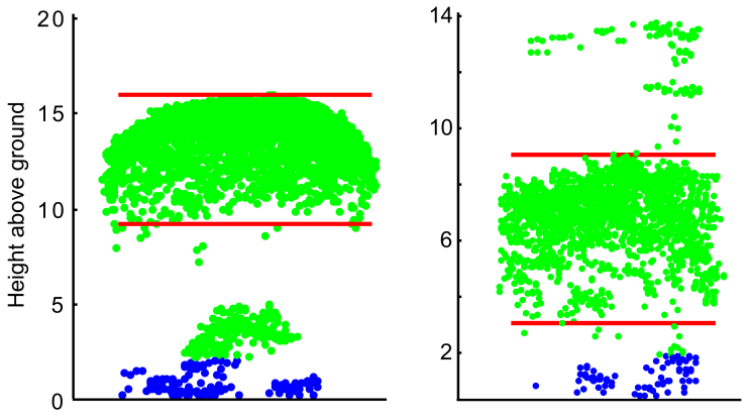

Figure 3.3: Two example crowns based on airborne lidar data collected during August 2010 with automatically estimated height and crown base height (red lines). Green dots indicate returns classified as canopy. (a) Presence of understory shrubbery. (b) Crown is overhung by power lines.

to allow for accuracy assessment against aggregated i-Tree Eco values. To clarify: it is not possible to validate individual crown results for watershed segments because it is not known whether a segment contains only a part of one, exactly one, or more than one tree. Thus, the leaf area estimates for the 1584 watershed crowns that intersected

71 i-Tree Eco plots were summed by plot and divided by the plot area (408 m^2) to yield plot-level LAI.

3.3.6.5 Map making and model intercomparison

A study-area wide map of LAI_e was generated at 10 m pixel resolution using the best performing of the L_{dirF} and L_{dirL} models. Finer resolution was not considered because it has been previously shown in an urban setting that, even with high pulse density, pixel sizes of 3 m led to data gaps due to lack of ground returns (Richardson et al., 2009). Moreover, given our study's mean pulse angle (20°) and mean canopy height (10 m) the horizontal displacement of ground returns relative to canopy returns was expected to exceed 50% of all pulses if 5 m pixels were generated. No such limitation existed for the allometric outputs which were mapped at the crown scale. The crown scale map was subsequently gridded to 10 m pixels through intersection and reapportionment operations in a GIS to allow for spatially explicit comparison with the selected LAI_e result**.** We compared the model results for the 71 study plots and also throughout a spatial subset of the study area using 450 randomly distributed sample points. Additionally, we compared all possible pairwise relationships (n = 12 when each ground and each lidar method is used as an independent and a dependent variable) to establish rough error bounds on LAI mapping in cities. For purposes of intercomparison, we report root-mean squared error (RMSE) and slope coefficient values from weighted least squares regression with inclusion of an intercept term in all cases.

3.4 Results

3.4.1 Comparison of ground-based estimates

Hemiphoto LAI_e and allometrically determined LAI from i-Tree Eco field measurements have different theoretical underpinnings. Nevertheless, there was a significant linear relationship between Hemiphoto LAI_e and i-Tree Eco allometric LAI (r^2 = 0.53), although the slope of 0.34 indicated that hemiphoto LAI_e was generally lower than allometric LAI **(Fig. 3.4)**. This relationship and all other linear relationships reported in this research were significant at the p = 0.001 level.

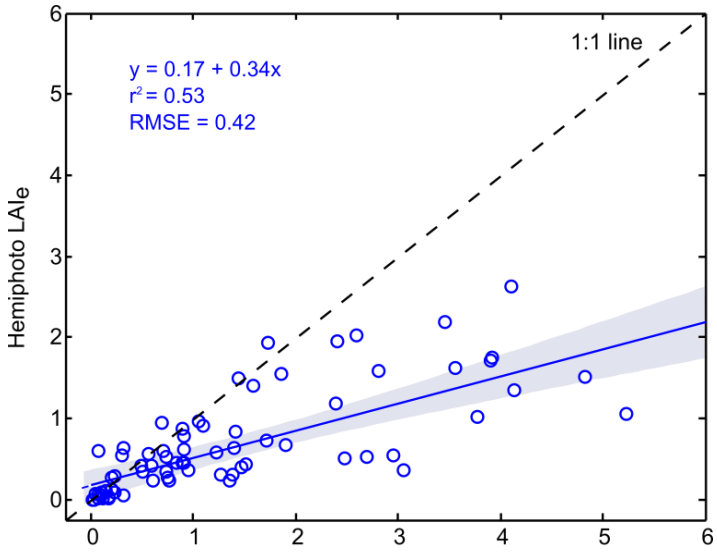

Figure 3.4: The relationship between UFORE allometric LAI from field measurements and hemiphoto derived LAI_e. Grey shading in this and all similar plots indicates the 95% confidence interval of the regression equation.

3.4.2 Plot fractional cover

Fractional cover was calculated for each plot using **Eq. 3.4** and it was compared to field estimates made using i-Tree Eco protocols. There was a significant linear relationship between the two estimates of fractional cover (r^2 of 0.85, y = 0.86x-0.92, data not shown), with lidar estimates generally resulting in 5-10% more fractional

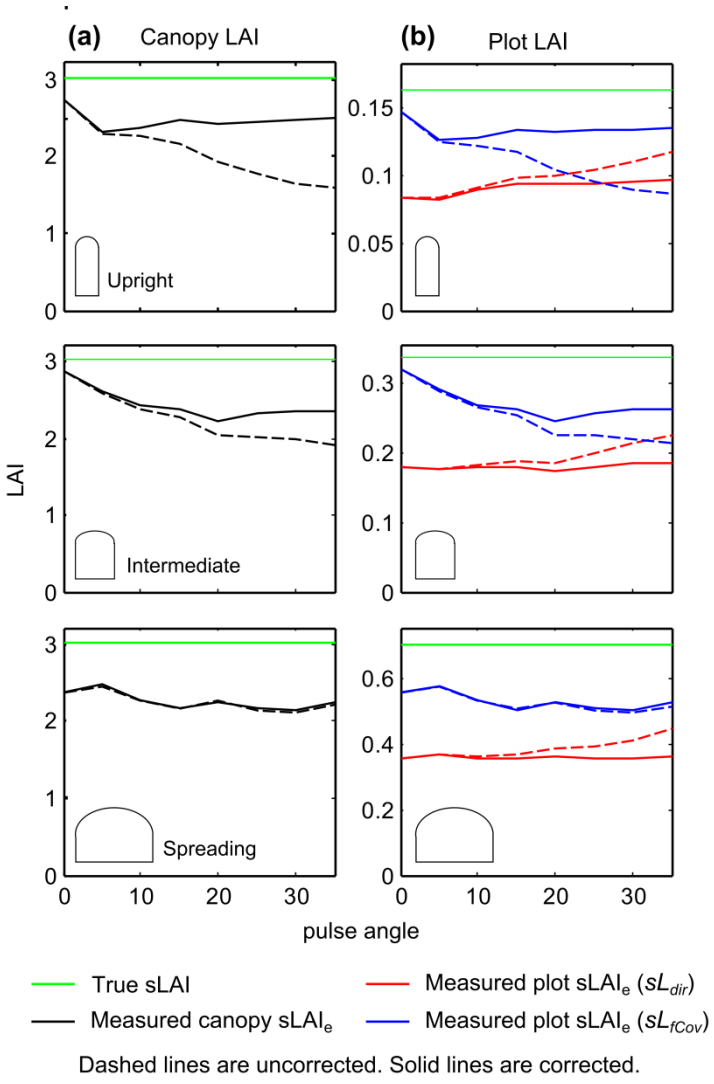

Figure 3.5: Adjustments to simulated LAI_e($sLAI_e$) using expected path length: (a) Canopy-level adjustment including only pulses intersecting the alpha hulls. (b) Plot level adjustments including all plot pulses in *eplCor* calculation. sL_{dir} uses the "direct" method (**Eq. 3.8**), sL_{fCov} uses the "canopy method" (**Eq. 3.9**). Upright, intermediate, and spreading crowns defined as 75th, 50th, and 25th percentile height-to-diameter ratios of field-measured trees

cover. The lidar results may be more accurate than the i-Tree Eco measurements because the latter produce relatively coarse estimates by dividing the sky into 4 quadrants and estimating the canopy cover in each visually.

3.4.3 Correcting for path length and plot-scale clumping using simulation

The expected path length correction method was evaluated in our simulation environment to determine its robustness with respect to crown geometry and distribution of foliage **(Fig. 3.5)**. LAI_e was underestimated for all crown geometries at both the canopy and plot levels with underestimation exacerbated by increased pulse angle, particularly for the upright and intermediate crowns. At the canopy level, for upright crowns, *eplCor* strongly compensated for the underestimation at pulse angles >5°. At 30° off nadir, the LAI estimate was corrected from 1.65 to 2.48. For the intermediate crown, the correction at 30° raised the LAI estimate from 1.98 to 2.38. Due to minimal change in pulse-angle dependent path length in the spreading crown, *eplCor* had negligible impact. Simulated plot level results based on rescaled canopy LAI_e mimicked the patterns produced in the canopy level results as they only differ by the scalar, *fCov*. In the case of the direct, plot-level results however, the net effect of *eplCor* (with non-canopy pulses now included) was to lower the estimated LAI_e by implicitly compensating for the higher number of pulses making contact with canopy at higher pulse angles.

Simulation of plot-scale clumping showed that underestimates of true plot LAI varied as a function of foliage density (the ratio of crown projection area to plot area) and LAI (**Fig. 3.6**). Underestimates were most severe when true plot LAI was highest (1.67) and the foliage density was highest (canopy radius = 3 m, plot radius = 15 m). In this case the measured LAI value was only 5% of true LAI. As foliage became more evenly distributed, estimates improved, following a logarithmic trajectory whose coefficients varied with true plot LAI. For instance, for a true plot LAI of 0.66, with all

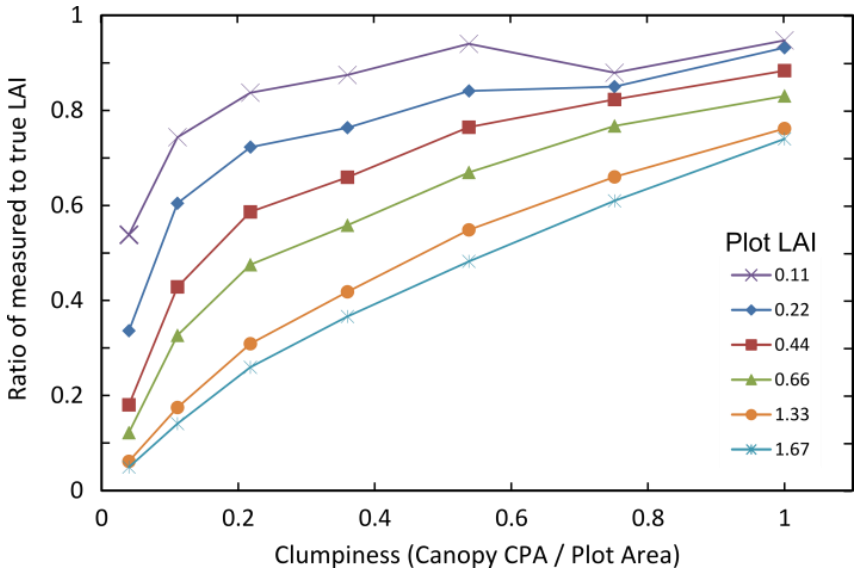

Figure 3.6: The impact of canopy clumping to deviation from true LAI on simulated 15 m plots. A ratio value of "1" indicates that measured LAI = true LAI. Each series represents one "tree" whose radius increases until it equals the plot radius. This plot highlights the difficulty of measuring plot-level LAI when foliage is densely clumped into trees with small canopy projection areas (CPA).

canopy clumped into a tree with a radius of 7 m, measurement yielded an LAI value that was 45% of true LAI. At a true plot LAI of 0.22 with foliage fully distributed throughout the 15 m radius plot, the measured LAI value was 90% of true LAI.

3.4.4 Plot LAI_e estimates from lidar data

On the assumption that view geometry alignment does not depend on LPM formulation, LAI_e was directly estimated using only L_{dirL} at all plot radii and all hemiphoto zenith bins with **Eq. 3.8**. The coefficient of determination was maximized for L_{dirL} using the range of hemiphoto zenith angles from 0 to 45° and a cylinder radius of 10 m **(Fig. 3.7)**. This correlation theoretically indicates a mean canopy height between 10 and 20 m, which indeed bounds the mean height of field-measured canopy: 12 m. This optimal zenith bin and lidar data cylinder radius was used in all further analyses.

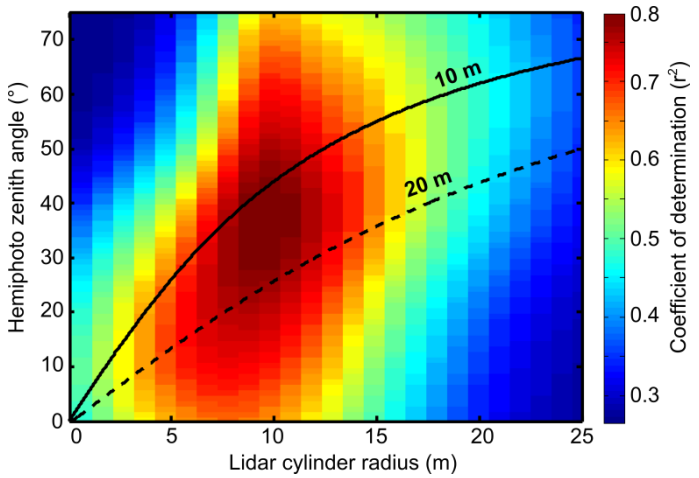

Figure 3.7: Coefficient of determination (r^2) matrix for the relationship between LAI_e from hemiphotos and LAI_e using lidar and the L_{dirL} model. Theoretical lines of maximum correlation for 10 m and 20 m canopy height are also displayed.

Model performance of L_{dirF} (first returns only) and L_{dirL} (inclusion of ground last returns) was assessed by simultaneous minimization of bias and maximization of the coefficient of determination with respect to hemiphoto LAI_e. After multiplication by the clumping ratio, the best model performance was exhibited by L_{dirL} with a no-intercept linear relationship of y = 0.99x, r^2 = 0.82, and RMSE = 0.41 **(Fig. 3.8)**. Examination of L_{dirF} (y = 0.73x, r^2 = 0.73, RMSE = 0.40) showed a linear relationship with hemiphoto LAI_e at lower values. At higher hemiphoto LAI_e values, significant overestimates were present. For simplicity, only L_{dirL} was considered for subsequent analyses. Changes to slope, r^2, and RMSE values after decimation to 5 pts m^{-2} were each within 1% of the full point-cloud values reported above. After decimation to 2 pts m^{-2}, the L_{dirL} slope decreased to 0.98, RMSE increased from 0.41 to 0.42 and the r^2 value decreased from 0.82 to 0.80. L_{dirF} slope decreased from 0.73 to 0.72, RMSE increased from 0.40 to 0.42 and r^2 decreased from 0.73 to 0.69.

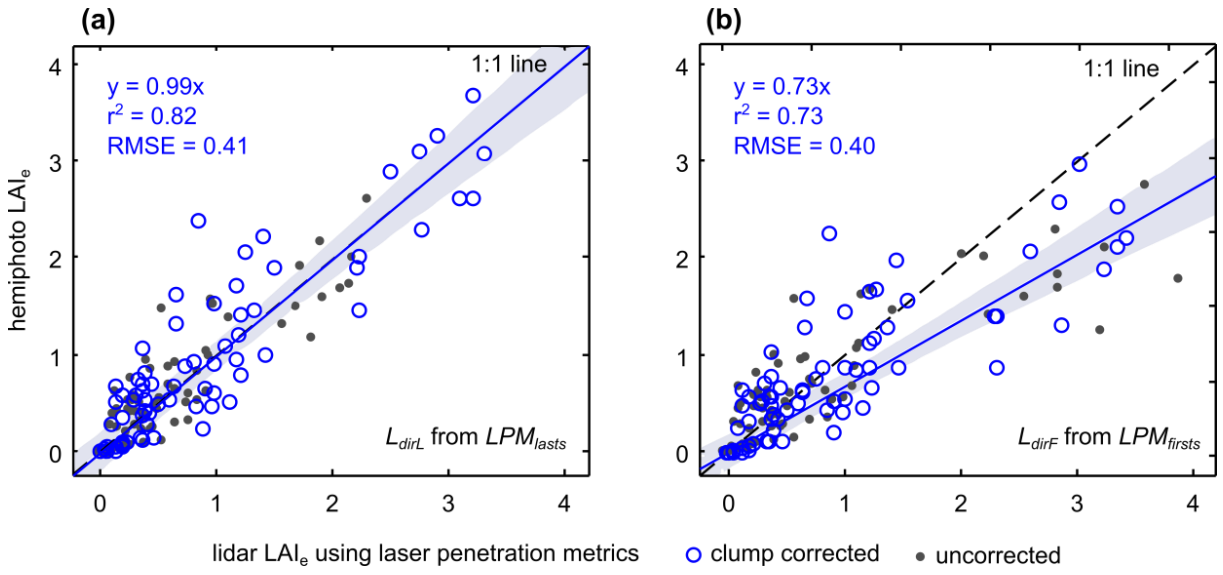

Figure 3.8: Weighted, no-intercept, linear relationships between hemiphoto LAI_e and LAI_e predicted using model L_{dirL} (a) and L_{dirF} (b). Clump-corrected results shown in blue and uncorrected result shown in gray. Results only displayed for full point cloud without decimation.

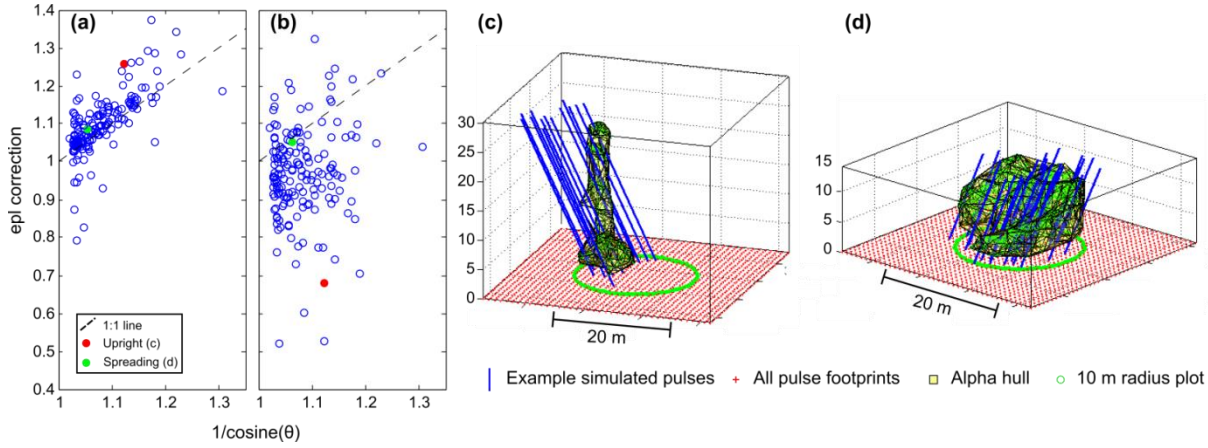

Figure 3.9: Demonstration of expected path length correction (*eplCor*) at plot and canopy levels. (a) Plot level correction compared to cosine correction. (b) Canopy-level correction compared to cosine correction. (c) Example of alpha hull wrapping a clump of upright palm trees. (d) An alpha hull wrapping a closed canopy of spreading oak trees.

Canopy-level *eplCor* was applied to correct the L_{fCov} model, while the plot-level analog was applied to correct the estimates from L_{dirL} **(Fig. 3.9)**.These plot-level corrections resulted in an 8% reduction in bias and a change in slope from 0.92 to 0.99. The canopy level correction, after conversion to a comparable plot-level result, resulted

in a 1% *increase* in bias with a decrease in slope from 0.74 to 0.73. This latter result is merely presented to show the effect of *eplCor*. We do not believe that the hemiphoto dataset offers comparability with L_{fCov} results due to different sensitivities to clumping. The mean value of the clumping ratio (L_{fCov}/L_{dirL}) was 1.4 (standard deviation = 0.26), which indicates similarity to the simulated values of 1.43, 1.42, and 1.49 for the upright, intermediate, and spreading crowns, respectively.

3.4.5 LAI estimates using lidar-measured crown dimensions and allometry

There was strong agreement between crown diameter, crown length, and crown-surface area calculated from lidar and their field-measured analogs, with r^2 values of 0.95, 0.87, and 0.94 respectively **(Fig. 3.10)**. Log-log fits by leaf type were established between the lidar structural measurements and the i-Tree Eco measurements:

$$\ln(LA_b) = \ln(1.76 + 0.60L_b + 2.32D_b - 0.44C_b) \quad (3.11)$$

$$\ln(LA_c) = \ln(-5.05 - 2.06L_c - 5.38D_c + 4.90C_c) \quad (3.12)$$

$$\ln(LA_p) = \ln(7.02 + 2.11L_p + 11.09D_p - 5.33C_p) \quad (3.13)$$

where the subscripts *b, c, and p* denote broadleaf, needle leaf, and palm leaf types respectively. Formulation of separate log models for each leaf type resulted in an r^2 value of 0.87 (RMSE = 0.39) compared to r^2 = 0.78 (RMSE = 0.48) for a combined model.

Application of the above models to the set of watershed segments intersecting our i-Tree Eco field plots (n_{segs} = 1584) resulted in an overall r^2 of 0.84 and a slight overall underestimate (y = 0.97x+0.10, RMSE = 0.53) compared to LAI estimated from field data using the native i-Tree Eco model equations **(Fig. 3.11)**. At the watershed-crown level, combined model r^2 = 0.81 (RMSE = 0.57, data not shown).

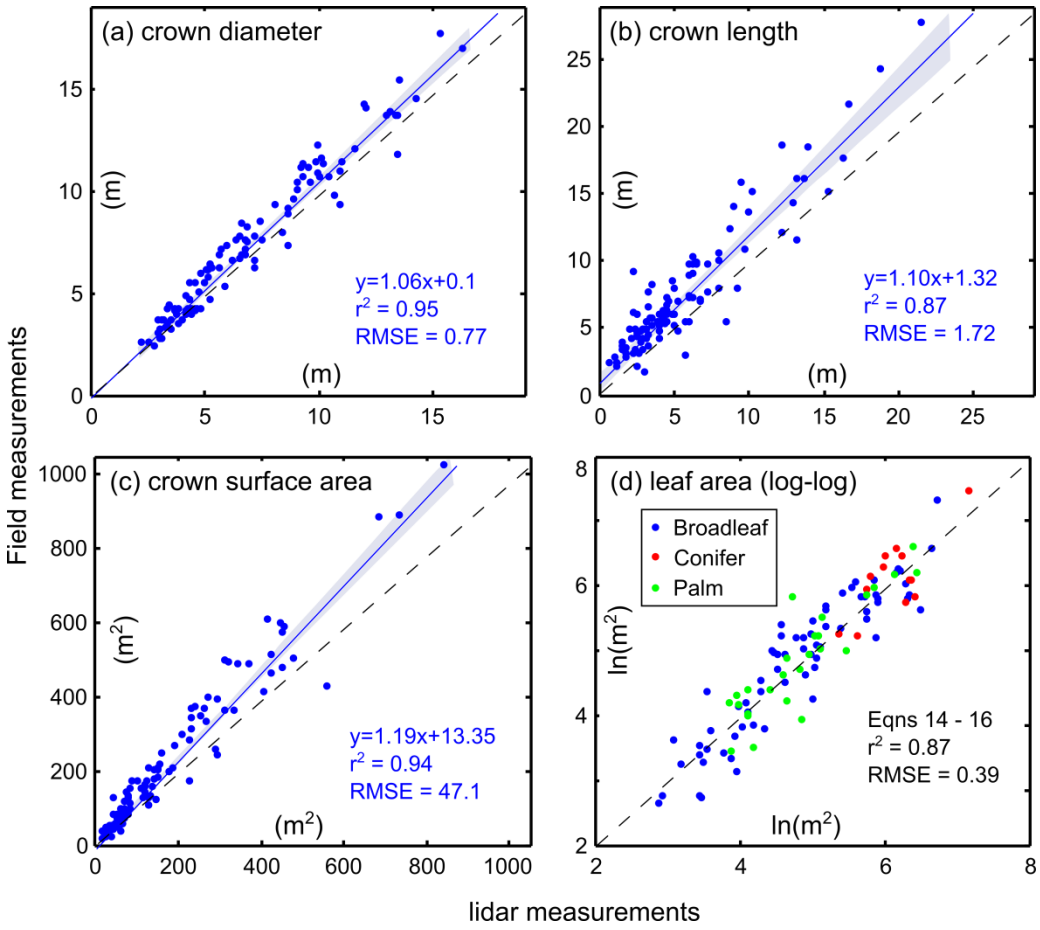

Figure 3.10: (a)-(c) Relationship between field-measured crown dimensions and the same measurements made using airborne lidar on 109 manually-delineated crowns. (d) Log-log fit of lidar-predicted leaf area and leaf area estimated from crown measurements in the field.

3.4.5.1 Comparing the lidar models and mapping the estimates

There was strong agreement (r^2 = 0.86, y=0.05+1.22x, RMSE = 0.59) between the LAI_e results from the L_{dirL} model and LAI measured allometrically using lidar-extracted crown measurements. Allometric LAI was underestimated slightly by L_{dirL} possibly, in part, because the latter does not account for shoot/branch scale clumping and because penetration metrics are prone to saturation at high LAI. All ground-based and lidar models were related with root-mean squared errors ranging between 0.39 and 0.93 and slope values between 0.45 and 1.21 **(Table 3.2).**

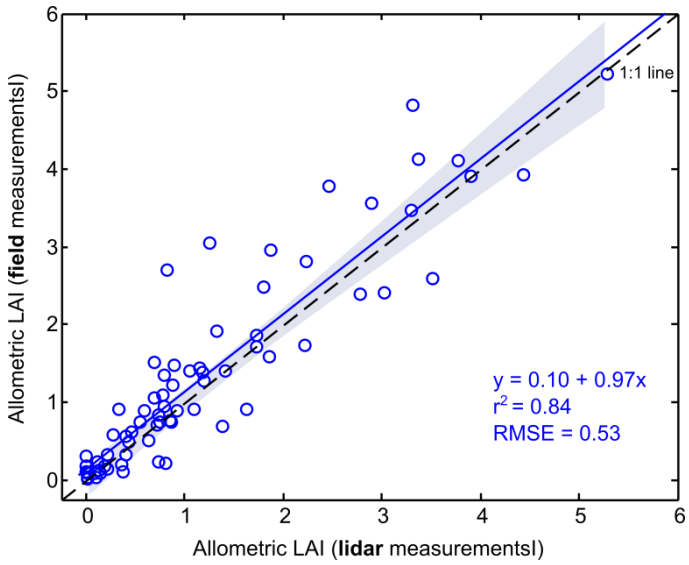

Figure 3.11: Allometric LAI from field measurements predicted by allometric LAI from lidar structural measurements using separate models for each leaf type (Eqs. 11-13). Aggregation of individual crown estimates to the field plot scale

Table 3.2: Weighted Least Squares root mean squared error (bold) and slope coefficients (in parentheses) for all combinations of field and lidar-estimated LAI or LAI_e. For comparability, all table entries are based on WLS regression with inclusion of an intercept term.

		Dependent variable			
		UFORE	hemiphoto	lidar LPM	lidar allom.
Independent variable	UFORE	--	**0.42** (0.45)	**0.57** (0.53)	**0.48** (0.87)
	hemiphoto	**0.93** (0.95)	--	**0.39** (0.89)	**0.74** (0.94)
	lidar LPM	**0.86** (0.89)	**0.4** (1.10)	--	**0.59** (1.21)
	lidar allm.	**0.53** (0.97)	**0.56** (0.69)	**0.44** (0.71)	--

Maps were generated using each of the models and are displayed here, for the purposes of visual clarity, as spatially corresponding subsets of the full study area **(Fig. 3.12)**. While produced initially at different spatial resolutions (10 m pixels for LAI_e versus crown scale for lidar allometry), the broad scale similarities are visually apparent. After resampling the crown-level map to 10 m resolution, we assessed the patterns of disagreement between the two methods **(Fig. 3.12c)**. As anticipated based

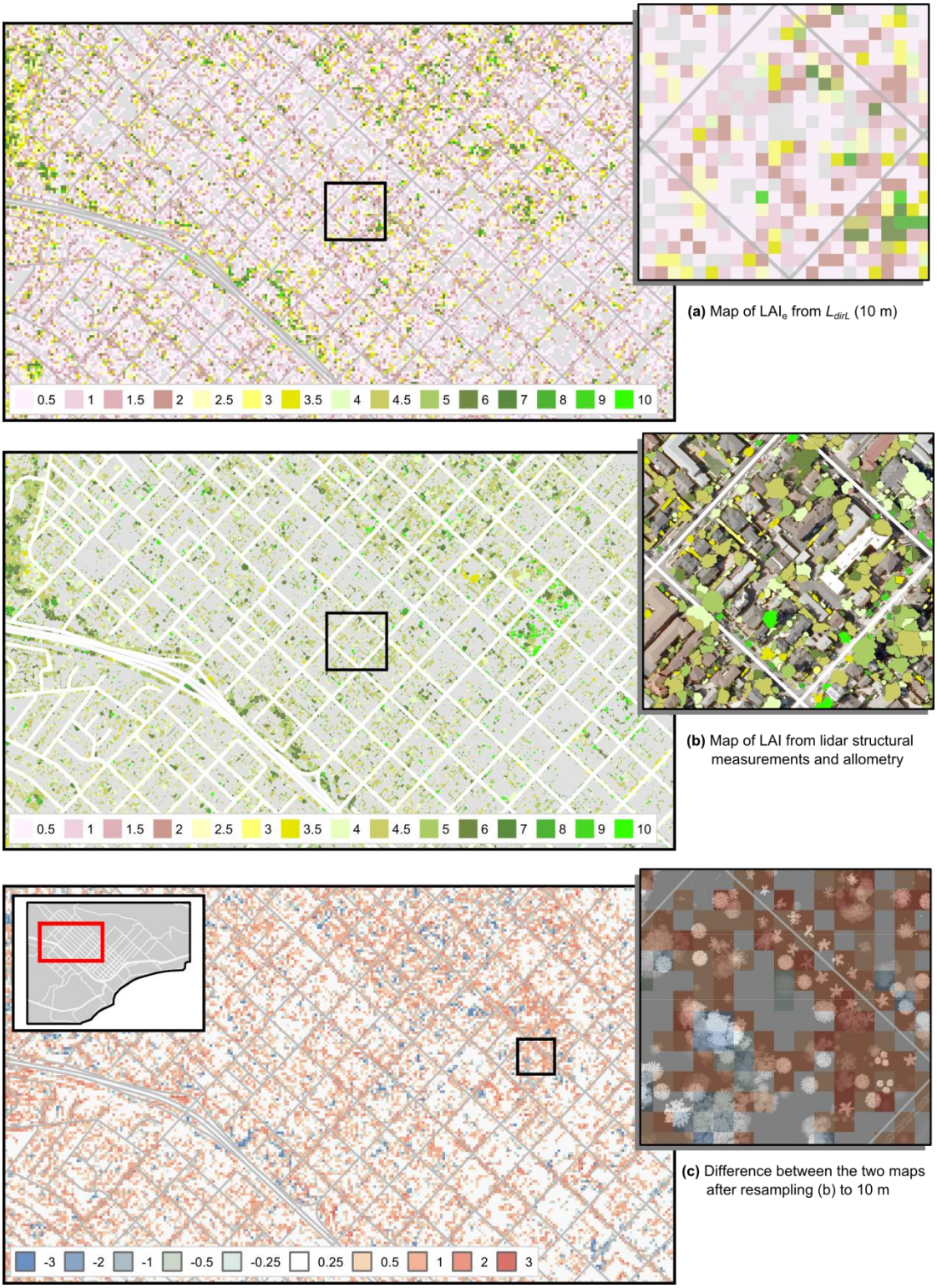

Figure 3.12: (a) Map of LAI$_e$ using the L_{dirL} model at 10 m pixel size for downtown Santa Barbara (see map inset for location). (b) Map of LAI from lidar extracted structural metrics and allometry at individual crown scale. (c) The L_{dirL} map subtracted from the allometric map after the latter was resampled to 10 m resolution. All legends are numbered with the class' highest value.

on plot-level results, the allometric model generally produced higher estimates. The L_{dirL} model produced higher LAI_e values mostly under large broadleaf canopies as illustrated in the inset map. To compare the maps more rigorously, we extracted pixel values from each 10 m map at 450 points. Map agreement between the two models yielded a coefficient of determination of 0.72. Interestingly, the bias in the map was the opposite of the bias in the plot-level comparison (0.78 instead of 1.22). In the map, at high LAI, L_{dirL} produced higher values. This may be due to the higher likelihood of 10 m pixels being under closed, broadleaf canopy compared to the 20 m diameter sample plots.

3.5 Discussion

3.5.1 Mapping LAI_e with laser penetration metrics (objective #1)

We found strong agreement between LAI_e estimated from hemiphoto gap-fraction inversion and LAI_e estimated using LPMs, as has been previously reported for natural forests (Hopkinson & Chasmer, 2009; Richardson et al., 2009b; Solberg et al., 2009b, 2006; Zhao & Popescu, 2009). Our best result (r^2 = 0.82, y = 0.99x, RMSE = 0.41) comparing LPM methods to hemiphotos was produced with the L_{dirL} model. We found that the inclusion of ground last returns (LPM_{lasts}) in our LPM formulation was necessary to retain model sensitivity at LAI_e >2 **(Fig. 3.8a)**. This was due to the fact that, at LAI >2, LPM_{firsts} was insensitive to change in gap fraction because the beam footprint was larger than the typical gap size (Lovell et al., 2003; Solberg et al., 2009b). For low canopy cover plots either metric is likely sufficient because LAI_e is largely governed by fractional cover. In regions of partial vegetation cover specifying LAI may

be redundant if fractional cover is already known (Carlson & Ripley, 1997). Still, it is unclear whether L_{dirF} leads to under- or overestimation of LAI_e as it is unknown whether we are more likely to digitize canopy returns where the pulse did not fully intersect with canopy or ground returns where part of the pulse *did* intersect with canopy (Hopkinson & Chasmer, 2009). The addition of last returns, beyond allowing for greater sensitivity to smaller gap sizes, increases the effective pulse density, which raises the likelihood that under- and overestimates at gap margins will cancel out.

In L_{fCov} only last returns on the ground were accounted for in the numerator. Thus, in dense canopy, the penetration rate may be zero, resulting in an LAI_e of infinity. This did not occur on our 71 plots because their 20 m diameter and relatively sparse canopy always allowed for some ground returns. However, on 73 out of 41,134 (0.2%) of our 10 m grid squares this numerical overflow forced exclusion of the L_{fcov} values. While this failure rate was low, it would increase at smaller pixel sizes and when applying a penetration metric to dense, individual crowns. Canopy density is only one cause of this error: For either individual crowns or small pixels, ground returns from non-nadir pulses will exhibit horizontal displacement with respect to associated canopy hits. The number of ground returns that should be associated with the pixel in question, but are actually counted in an adjacent pixel, will vary as a function of pixel size, canopy height, and pulse angle.

While it was necessary to include last returns in order to maintain model sensitivity at higher LAI_e, sensitivity was not impacted significantly through reduction in pulse density. The full point-cloud relationships depicted in **Figure 3.8a**, differed negligibly in terms of r^2 and bias from those generated at 2 pts m^{-2}. The only

noteworthy deviation occurred with 2 pts m^{-2} sampling density in plots with very low fractional cover. In several of these cases higher pulse density was required in order to detect the presence of small patches of vegetation. In high cover plots, the low pulse density was still sufficient, at least at 10 m radius plot scale, to produce similar ratios of ground returns to all returns. The fact that this method may be successful at lower pulse densities allows cities to consider similar analyses with existing lidar datasets.

3.5.1.1 Correcting for off-nadir pulse angles and clumping (objective #2)

Assessment of LAI_e using Beer-Lambert's law assumes that attenuation of light through the canopy is partially a product of its passage through *N* statistically independent horizontal layers of canopy (Jonckheere et al., 2004). When the *N* layers are sufficiently thin and foliage is distributed randomly, the probability of light interception in a given layer follows a Poisson distribution. For measurement of LAI_e from overhead lidar, we assume that this passage follows a near vertical path from sensor to ground. If, however, the pulse angle deviates significantly from 0°, we must consider how this change affects the path length of the pulse through the canopy. All else being equal, an increase in path length due to sensor positioning relative to a crown will lead to a spurious reduction in gap fraction. Given a continuous, uniform extent of canopy, it is reasonable to assume that path length correction can be applied in the form of $1/\cos(\Theta)$ where Θ is the angular deviation from nadir.

Our urban study area, however, was characterized by heterogeneous canopy such that path length was additionally dependent on crown geometry and pulse azimuth. We found that, at the canopy level, a simple cosine correction had little relationship with *eplCor* and, under most circumstances, prescribed an adjustment that erroneously

lowered the LAI_e estimate **(Fig. 3.9b)**. This effect can be seen clearly with the cluster of highly upright palms shown in **Figure 3.9c**. At the plot level, however, the utility of the cosine correction remained largely intact **(Fig. 3.9a)**. On plots dominated by spreading crowns (**Fig. 3.9d**) the distinction between plot and canopy-level correction is less pronounced. At the plot level there is a coherent linear relationship between *eplCor* and simple cosine correction **(Fig. 3.9a)**. Thus, it is not surprising that this use of *eplCor* only resulted in a 2% improvement over $1/\cos(\Theta)$. The true value of *eplCor* will more likely be realized should measurement of LAI_e become feasible at the individual crown scale.

Estimation of true LAI from hemiphotos requires a horizontally continuous, random distribution of foliage across all view zenith and azimuth angles as well as the ability to separate woody plant matter from photosynthetic material (Jonckheere et al., 2004). The same requirements apply to plot-level estimation using lidar. While difficult to discern from the literature, there appear to be a number of studies that utilize hemiphotos under discontinuous canopy conditions (e.g., Richardson et al., 2009; Solberg et al., 2006; Zhao & Popescu, 2009). Under these conditions, estimating LAI_e is still possible (and technically valid) but the magnitude of the departure from true LAI may become extreme due to significant spatial separation among clumps of foliage. Given the extremely discontinuous canopy in a semi-arid, urban environment, we attempted to account for clumping attributable to tree spacing.

We measured clumping as the ratio of LAI_e estimated using L_{fCov} to L_{dirL} and produced an average clump ratio value of 1.4 for our study area. We multiplied both the hemiphoto LAI_e estimates and the L_{dirL} estimates by the plot-specific clump ratio. This

result cannot be validated directly but we did find increased agreement with allometric results, which could indicate that the correction for plot-scale clumping shifts LAI_e values somewhat closer to true LAI**.** It may behoove future researchers to experiment with ground-based methods that allow for indirect LAI_e estimation of individual trees or coherent canopy clumps. Methods of this type that have been used in sparse canopy conditions include non-hemispherical digital canopy photography (Pekin & Macfarlane, 2009; Peper & McPherson, 2003; Ryu et al., 2010b) and the LAI-2000 with restrictions placed on the view zenith and azimuth (Peters & McFadden, 2010).

3.5.1.2 Leaf angle distribution

We used a fixed value of k = 0.5 for all LAI_e models because it represents the extinction coefficient for the spherical leaf angle distribution (LAD). In the least squares fit between L_{dirL} and hemiphoto LAI_e, our estimated slope coefficient was 0.99. This equates to a modeled k value of 0.495, suggesting that in a highly mixed urban forest, an initial k value of 0.5 is reasonable. This is consistent with Richardson et al. (2009), who estimated k = 0.485 in a semi-urban park setting with a mix of broadleaf and needle leaf species. Falster & Westoby (2003) digitized the LADs in three dimensions of 38 perennial species in sclerophyll woodland qualitatively similar to the oak / eucalypt assemblages common in southern California. They found that each species had a unimodal LAD with mean leaf angles ranging from 27° (planophile) to 74° (erectophile). Across their entire sample, the cross-species mean leaf angle was 52° which is approximately equidistant from the expected means for uniform random and spherical LADs. While 0.5 may be an appropriate theoretical value for k in mixed forests, it must be noted that the value additionally depends on lidar acquisition parameters such as

LPM formulation. Solberg et al. (2009) showed, using first returns only, that their regression's slope coefficient (equivalent to $1/k$ in this context) was stable across Norway spruce development classes (1.83 – 1.98). However, when using both first and last returns the slope coefficient mean across development classes increased to 2.47 with a range from 1.83 to 2.69.

3.5.2 Estimating LAI with lidar-derived structural metrics and allometry (objective #3)

Our urban study adds to the larger body of work in natural forests suggesting that lidar is effective for measuring basic crown structural attributes **(Fig. 3.10)** such as height (e.g., Andersen et al., 2006; Edson & Wing, 2011) and crown base height (e.g.,Popescu & Zhao, 2008; Reitberger et al., 2009). This is consistent with research suggesting laser return height distributions are less susceptible to increased scan angle than penetration metrics (Holmgren et al., 2003; Morsdorf et al., 2008). It is likely that metrics such as crown base height are *best* assessed with off-nadir data and in an urban forest characterized by open-grown trees. Allometric model formulation was improved by the incorporation of leaf type information (from $r^2 = 0.78$ and RMSE = 0.48 to $r^2 = 0.87$ and RMSE = 0.39). However, the model, when applied to the watershed segments improved by a smaller margin with this increased specificity ($r^2 = 0.81$ and RMSE = 0.57 to $r^2 = 0.84$ and RMSE = 0.53). This is possibly attributable to inaccuracies in the leaf-type classification of the watershed segments. Improvement was also likely limited due to aggregation of the results to the i-Tree Eco-plot level for validation purposes.

Implementation of the crown-scale allometric method produces fine-spatial scale results that are useful for calculation of urban forest ecosystem services such as

building energy-use reduction and stormwater runoff mitigation. However, a key limitation of this method, and a large source of uncertainty, is that shading coefficients have been directly estimated only for a small number of species and regions. While this coefficient is not included in **Eq. 3.11-3.13**, it is introduced implicitly when training the model against i-Tree Eco results. Of the 108 species sampled in the field, species-specific coefficients were only available for 17, and those coefficients were not likely estimated in Southern California. The remaining species were assigned coefficients corresponding to other trees of the same genus or leaf type. We expect the empirical scaling coefficients would need to be determined for other regions and forest types in order to obtain a similar degree of accuracy. Further error may have been introduced to the models generated in this study because they do not account for variation in tree condition or crown transparency. Future gains could likely be made by combining structural measurement of crown dimensions with a lidar penetration metric correlated with crown porosity.

3.5.3 Comparison of models and maps (objective #4)

In this study we made field measurements of LAI using crown structural measurements plus allometry and LAI_e using gap fraction inversion of hemispherical photos. We developed lidar models that roughly mimicked each method and then validated each model against its field analog. The final objective of this study was to compare the results from each method **(Table 2)**. We did this to estimate the error bounds on urban LAI estimates. This is not a statistically-based error bound and caution must be taken in comparing LAI and LAI_e. Nevertheless, we believe that the first-order similarity of these results indicates potential for wider application of lidar remote

sensing for mapping urban LAI. Notably, it was the comparison of the two ground-based methods, which resulted in the weakest relationship in terms of RMSE and bias.

The maps **(Fig. 3.12)** illustrate the broad similarities between the methods but also highlight the key difference in spatial scale of the output. In theory, the LPM map can be generated at smaller pixel sizes. However, as pixel size decreases, the number of pixels with infinite LAI_e due to numerical overflow and the percentage of ground points actually associated with neighboring pixels will increase. Richardson et al. (2009) found that, with their lidar point density and LPM formulation, they could produce an LAI map at 14 m but not at 3 m, due to the lack of ground returns in many canopy-filled pixels. These issues make the prospect of mapping individual tree LAI_e using penetration metrics difficult at the present time.

The patterns of map disagreement **(Fig. 3.12c)** clearly reflect the differences in model formulation and may shed light on which model is most appropriate for a particular application. Where canopy cover is low and clumped in isolated trees (see **Fig. 3.12c** inset) the allometric method may be better suited to capture the fine scale variability in LAI. In this scenario a pixel-level inversion of an LPM would be unduly influenced by extremely high gap fraction due to small crown size and clumped foliage. On the other hand, in areas with higher overall and relatively homogeneously distributed canopy cover, pixel-level LPMs are quite sensitive to small changes in leaf surface due to the nature of the logarithmic inversion. Further, in those areas, segmentation algorithms preceding the allometric approach are less likely to properly delineate overlapping crowns so the utility of leaf type classification is diminished.

3.6 Conclusions

This study demonstrated the potential for mapping Leaf Area Index (LAI) in a heterogeneous urban environment using two theoretically distinct methods. We first showed strong agreement between effective LAI (LAI_e) estimated from a laser penetration metric (LPM) and LAI_e measured in the field using hemispherical photography ($r^2 = 0.82$). In order to quantify the relationship between the two methods, we developed novel methods to correct for off-nadir pulse angles and plot-level clumping in a structurally diverse and discontinuous canopy. While we initially made use of a lidar dataset with very high point density (22 pts m^{-2}), we found that the results could be reproduced at decimated point densities down to 2 pts m^{-2}. This indicates that cities may have success implementing LPM methodology for calculating LAI_e using existing data.

Secondly, we showed that lidar-derived structural metrics such as height, crown base height, and crown segment area can be used as inputs to existing allometric equations for prediction of LAI. This result was compared against LAI allometrically estimated from field measurements of individual trees and yielded an r^2 of 0.84 when formulating separate models for each leaf type. A key difference in the mapped outputs between the methods was the spatial resolution. We found that a map produced using LPMs must have a pixel size large enough to allow for lidar beam penetration to ground, even under dense canopy. An allometric map has no theoretical lower bound on the size of the output crown object. However, maps generated using allometric equations are subject to an unknown amount of error associated with use of coefficients not developed for the trees or site in question. Thus, while this work demonstrates that LAI

can be mapped at citywide scales, it is still important that the practitioner be aware of the trade-offs inherent in each of the methods. The ability to map LAI across large urban areas offers new potential to constrain estimates from hydrological and atmospheric models and better understand the spatial distribution of urban ecosystem services at increasingly fine scale.

3.7 Acknowledgements

This research was funded by the Naval Postgraduate School (N00244-11-1-0028), the National Science Foundation (BCS-0948914), and the Garden Club of America Fellowship in Urban Forestry. Special thanks to Seth Gorelik and Keri Opalk for their assistance with field data collection, David Nowak for suggestions to improve the technical methods, and the good people of Santa Barbara for allowing us on their lawns.

Chapter 4

Mapping urban forest structure and function using hyperspectral imagery and lidar data

Authors:

Michael Alonzo[a], Joseph P. McFadden[a], David J. Nowak[b], and Dar A. Roberts[a]

[a] Geography Department, Ellison Hall 1832, UC Santa Barbara, CA 93106-4060, USA

[b] USDA Forest Service, Northern Research Station, 5 Moon Library, SUNY-ESF, Syracuse, NY 13210

This chapter will be submitted to *Urban Forestry and Urban Greening*.

4.1 Abstract

Cities measure the structure and function of their urban forest resource to optimize forest management and the provision of ecosystem services. Measurements made using plot sampling methods yield useful results including citywide or land-use level estimates of species counts, leaf area, biomass, and air pollution reduction. However, these quantities are statistical estimates made over large areas and thus are not spatially explicit. Maps of forest structure and function at the individual tree crown scale can enhance management decision-making and improve understanding of the spatial distribution of ecosystem services relative to humans and infrastructure. In this research we used hyperspectral imagery and waveform lidar data to directly map urban forest species, leaf area index (LAI), and carbon storage in downtown Santa Barbara, California. We compared these results with a field-plot study using the i-Tree Eco model. Remote sensing methods generally reduced uncertainty in species-level canopy cover estimates compared to field-plot methods. This was due to high classification accuracy for large canopy species (e.g., *Platanus racemosa* with ~90% average accuracy, *Pinus pinea* at ~93%, *Quercus agrifolia* at ~83%) and high standard error of the plot-based estimates due to the uneven distribution of canopy throughout the city. Average LAI in canopy based on lidar measurements was 4.47 while field measurements and allometry resulted in an LAI of 5.57. When accounting for fractional cover, citywide lidar-based LAI was 1.15 and field-based LAI was 1.29. Citywide carbon based on lidar measurements and allometry was estimated at 50,991 metric tons (t) and 55,900 t from plot-sampling. As others have noted, carbon density varied substantially by development intensity based largely on differences in fractional cover but less so when

only evaluating canopy. Using separate biomass equations for each leaf type (broadleaf, needleleaf, palm) resulted in a more accurate carbon map but a less accurate citywide estimate.

4.2 Introduction

Measurement of urban forest structure is a prerequisite to estimating urban forest ecological functions such as urban heat island mitigation, air pollution removal, carbon storage, building energy-use modification, and stormwater runoff reduction. Measurements of structure are generally based on a limited set of random field plots given the impracticality of a complete urban tree inventory (Nowak et al., 2008a). Estimates of urban forest function are commonly made using the USDA Forest Service's i-Tree Eco (formerly Urban Forest Effects or UFORE) model (www.itreetools.org; Nowak et al., 2008a). Many cities use the model outputs to improve urban forest management and inform environmental policy (e.g., McPherson et al., 2011). However, field sampling methods come with potential drawbacks: (1) they are labor and time intensive, generally taking a trained, 2-person field crew ~14 weeks to sample about 200 0.04-ha plots (Nowak et al., 2008b); (2) they can be expensive, depending upon local transportation and labor costs (3) permission to access plots may be limited in some areas; (4) the magnitudes of fundamental outputs such as species and tree size distributions are subject to sampling error that are dependent upon sample size and forest variability; (5) the results are averages or totals within classes and not spatially explicit across the city.

Generation of spatially explicit outputs is important for several reasons: First, it is critical for urban forest managers to know the specific locations of common species so they can prepare for threats from pests, fire, and blight (Laćan & McBride, 2008; Santamour, 1990; White & Zipperer, 2010). Second, the magnitudes of ecosystem functions vary spatially throughout the city in relation to historical tree planting patterns, natural regeneration, tree site suitability, water availability, infrastructure conflicts, maintenance, street traffic, etc. (Escobedo & Nowak, 2009; Manning, 2008; McCarthy & Pataki, 2010; McPherson et al., 2011; Nowak, 2012; Simpson, 2002; Urban, 1992). Third, ecosystem services may not be distributed fairly with respect to class and race: Communities that would benefit the most from improved air quality, cooler temperature and even reduced crime (Kuo & Sullivan, 2001) are frequently those with the least access to canopy (Heynen et al., 2006; Schwarz et al., 2015).

In contrast to extensive field sampling, remote sensing can provide spatially extensive data, potentially with higher temporal resolution and lower cost. However, the use of remote sensing in the development of urban forest inventories has been limited due to the spatial and spectral complexity of urban landscapes (Herold et al., 2004). For example, Landsat, a space-borne, multispectral sensor is not well suited to this problem because of its 30 m ground instantaneous field of view, which is much larger than the characteristic scale of spatial variation found in urban areas (~5-10 m; Jensen & Cowen, 1999; Welch, 1982; Woodcock & Strahler, 1987). High spatial resolution (< 3 m), multispectral platforms such as IKONOS and GeoEye have allowed for precise mapping of urban forest canopy extents (MacFaden et al., 2012; Richardson & Moskal, 2014a) but limited spectral information largely precludes species

identification or structural quantification. However, successful discrimination of urban forest tree species has been shown possible using either only hyperspectral imagery (Alonzo et al., 2013; Xiao et al., 2004) or hyperspectral imagery fused with light detection and ranging (lidar) data (Alonzo et al., 2014; Zhang & Qiu, 2012). Leaf area index (LAI) has also been accurately estimated at fine spatial scale using both lidar penetration ratios as a proxy for gap fraction and allometric estimates based on individual crown structural measurements (Alonzo et al., 2015).

While the ability to map urban forest structure at high spatial resolution has been demonstrated, citywide maps of structure and attendant functions have not yet been produced. The primary goal of this paper is to examine the capability of hyperspectral imagery and lidar data to produce citywide maps of key structural attributes: tree species, LAI, and carbon storage. Our secondary goal is to evaluate the statistical uncertainty of the maps and to assess how they may be used with other types of spatial data to address urban ecological questions. The specific objectives of this paper are as follows:

1. Scale the species identification methods developed by Alonzo et al. (2014) and the LAI estimation methods from Alonzo et al. (2015) into citywide maps.
2. Create a citywide map, at the individual tree crown scale, of dry-weight carbon storage using lidar crown measurements and allometric scaling.
3. Analyze the spatial patterns of each map to demonstrate the utility of spatially explicit representation and to characterize the uncertainties of the remote sensing methods.

4. Compare the remote sensing outputs with i-Tree Eco outputs generated using traditional plot-sampling methods

4.3 Methods

4.3.1 Study site description

This study was conducted in downtown Santa Barbara, California (34.42° N, 119.69° W) as discussed in **Chapter 1** and illustrated in **Fig 1.1**. Santa Barbara is a city of about 90,000 residents located on a coastal plain between the Pacific Ocean to the south and the Santa Ynez mountains to the north. It has a Mediterranean climate and supports a diverse mix of native, introduced, and invasive urban forest species. The study area represented 45% of the total area of the city. We chose the specific boundaries in order to focus on the most characteristically urban parts of the city and based on availability of contemporaneous lidar and hyperspectral data. The city of Santa Barbara maintains a spatial database that contains one or more specimens from >450 tree species. While collecting i-Tree inputs at 105 plots, 108 unique species were recorded. Despite this diversity, based on i-Tree and municipal data, we estimate that only 25 species represent approximately 80% of the city's canopy cover. The most common and canopy dominant native species is *Quercus agrifolia* (Coast live oak). The most common introduced species in terms of stem count is the *Syagrus romanzoffiana* (Queen palm).

4.3.2 Field data collection and the i-Tree Eco model

In the Fall of 2012, we took measurements at 105, 11.4 m radius field plots that were randomly distributed in 105 grid cells subdividing the study area. In this study, we screened out 16 of the northernmost plots because they fell within the boundaries of a flight line that was unusable due to errors in reflectance retrieval. Thus, 89 plots were retained for further analysis. At the time of data collection we were unable to gain permission to take measurements at 9 private property plots. For each access denial, a new plot location was randomly generated in the same grid square. Plots were not stratified by land use in this study but nonetheless were distributed roughly in proportion to the major land use classes as noted in **Chapter 1** and **Table 1.1**. The land use classes chosen for this study are common urban classes that are similar to those defined in other i-Tree Eco studies such as Los Angeles (Nowak et al., 2010) and Washington, DC (Nowak et al., 2006b). The aggregated class, Natural-Agricultural-Recreation includes open, undeveloped lands and some parks which house most of the area's native stands of *Quercus* and *Platanus*.

At each plot, we identified and measured all trees with diameter at breast height (dbh) > 2.5 cm (1 in) and a height of at least 1.83 m (6 ft). Tree measurements included dbh of all stems, tree height, height-to-live-top, crown diameter (North – South), crown diameter (East – West), crown base height, crown missing, and dieback. We also measured the distance and direction of each tree from each building that was less than 18 m (60 ft) away. At the plot level, we estimated percent canopy cover, ground cover class percentages, and land-use class percentages (i-Tree Eco User's Manual v. 4.1.0, www.itreetools.org). At each plot in which one or more trees were present, we took up

to 5 near infrared, hemispherical photographs to measure plot-level gap fraction and LAI (Alonzo et al., 2015).

We used the i-Tree Eco software application to generate a statistically based estimate of urban forest structure and function. In i-Tree, primary estimates of structure include stem counts by species, size class distributions, and LAI. While species and size class distributions are directly estimated from field sampling, LAI is allometrically estimated using an equation specifically developed for open-grown trees (Nowak, 1996). Ecosystem functions such as air pollution reduction, carbon storage, building energy modification, and stormwater runoff reduction are estimated from tree measurements and ancillary, contextual variables. Reduction of airborne CO, O_3, NO_2, SO_2, and $PM_{2.5}$ is modeled as a function of tree LAI, hourly meteorological data, and local hourly pollution concentrations. The economic value of the associated human health benefit additionally requires population density information (Nowak et al., 2014). Fresh weight carbon is estimated using measurements of dbh and height input into species, genus, or leaf-type specific equations from the literature (McHale et al., 2009; Nowak & Crane, 2002). Fresh weight is converted to total dry-weight carbon using literature values (Nowak & Crane, 2002). Trees modify building energy use through direct shading, evapotranspiration, and windbreak effects. These building energy benefits and avoided carbon emissions from associated power production are estimated from tree size, distance and direction to buildings, climate region, leaf type, canopy cover, regional energy production information, and building construction information

(McPherson & Simpson, 1999b). Stormwater runoff reduction can be estimated from canopy cover, LAI, land cover, and local weather data (Wang et al., 2008)1.

The i-Tree Eco model is subject to the following uncertainties: 1) measurement error in the field data collection, which is assumed to be zero; 2) sampling error that can be characterized by the standard error of the estimate; and 3) model error (e.g., error in applying allometric equations, conversion factors, pollution removal models) that is generally of unknown magnitude (Nowak et al., 2008a), but have been shown to provide reasonable estimates compared with field measurements (Morani et al., 2014; Peper & McPherson, 2003) or other standardized field/model estimates (Nowak et al., 2013a).

4.3.3 Remote sensing data and processing

Full waveform lidar data were collected in August of 2010 with a helicopter-mounted Riegl Q560 scanner. The waveform was discretized to an average, last-return pulse density of 22 pts m^{-2} with additional returns available in tree canopy. The lidar returns were classified as ground, building, or tree canopy using LASTools (LAStools v111216, http://lastools.org) with minimal adjustments to default settings. Finally, a gridded (0.25 m) canopy height model was exported for use in crown segmentation algorithms.

Hyperspectral imagery was collected using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS; Green et al., 1998). The two radiance images used in this study were acquired on November 1, 2010 at approximately 11:50 and 14:20 Pacific Standard Time (PST). AVIRIS is a 224 channel system that samples radiance at 10 nm intervals

[1] Stormwater runoff was not estimated in the i-Tree eco implementation discussed in this article.

across the solar-reflected spectrum from 365 nm to 2500 nm. This spectral range and resolution allows for identification of leaf pigment absorption features such as chlorophyll in the visible wavelengths of light (Ustin et al., 2009), characterization of leaf abundance and structure using the near infrared wavelengths (Roberts et al., 2004) and assessment of water status and lignin-cellulose features in the shortwave infrared (Kokaly et al., 2009). The radiance images were converted to surface reflectance using ATCOR-4 (Richter & Schlaepfer, 2002) and 178 bands with adequate signal-to-noise ratios and minimal water vapor contamination were retained for analysis. The two AVIRIS images were mosaicked together at a final pixel resolution of 3.7 m. The hyperspectral mosaic was co-registered to the lidar data based on 137 ground control points and Delaunay triangulation.

For analysis at the individual tree crown scale (hereafter "crown scale"), the remote sensing imagery must be segmented into crowns. Crown segmentation is difficult and frequently inaccurate particularly in forests characterized by overlapping, broadleaf crowns (Ferraz et al., 2012). However, urban areas contain many isolated trees that make automated segmentation more straightforward. The marker-controlled watershed segmentation algorithm used in this study is described in detail in Alonzo et al. (2014). In brief: Tree tops are isolated ("marked") on the gridded canopy height model and crowns are "grown" from each tree top until a local minimum is reached. This minimum ideally corresponds with the end of one crown and the beginning of another. This method was 83% accurate in the Santa Barbara study area with success defined as a watershed segment containing one-and-only-one tree stem. Areas of

inaccuracy primarily corresponded to natural forest stands which were relatively homogeneous with respect to species.

4.3.4 Species mapping

It is currently not possible to distinguish all species in a species diverse urban forest using either field sampling or remote sensing methods. Thus, in this study we used cumulative canopy cover distributions calculated from i-Tree data and the city's public tree database to choose 29 common species that accounted for the greatest proportion of the canopy cover (**Table 4.1 col A**). We chose to quantify canopy cover instead of stem counts because there is a more direct relationship between canopy area and the magnitude of ecosystem function and it is a more straightforward quantity to measure from an airborne platform. Alonzo et al. (2014) developed a method for classifying the common species in the Santa Barbara study area using fusion of lidar measurements and hyperspectral data at the crown scale. In brief, this method fused the information from 178 AVIRIS bands with 7 lidar-derived structural metrics (e.g., tree height, width-to-height ratios, crown porosity) for each segmented tree crown and classified that crown using canonical discriminant analysis (**Fig 4.1**). Crowns that were not one of the 29 common species were classified as one of the common species and their classification was deemed correct when leaf type (e.g., broadleaf, needleleaf, palm)

Table 4.1 (on following page): Comparison of i-Tree and remotely sensed citywide estimates of species composition including indicators of uncertainty. Column D is Relative Standard Error. Column K is weighted by the i-Tree-estimated canopy area. Remote sensing metrics (Quantity disagreement, Producer's/User accuracy) are based on a validation sample of 2,304 crowns. Quantity disagreement is the difference between the amount of canopy classified as a certain species and the actual amount of that species' canopy; location agreement is not considered. Producer's accuracy quantifies classification accuracy (quantity and location) in terms of error of omission. User accuracy is the same as Producer's but for errors of commission.

A	B	C	D	E	F	G	H	I	J
Common Species Botanical name	**Leaf Type**	**Eco Canopy Area (sq m)**	**Eco rel. Std. Err. (89 plots)**	**Rem. Sens. Canopy Area (sq m)**	**Rem.Sens. Quantity Disagreement**	**Rem. Sens. Producer's Accuracy**	**Rem. Sens. User Accuracy**	**Eco Rem. Sens. Difference**	**Eco / RS Area Weig. Difference**
Archontoph. cunningham.	P	56,476	80%	140,148	85%	65%	35%	148%	1%
Cupressus macrocarpa	C	240,367	83%	201,714	-1%	90%	91%	-16%	-1%
Eucalyptus globulus	B	507,152	66%	224,757	-5%	93%	98%	-56%	-5%
Ficus microcarpa	B	28,262	99%	57,771	-8%	92%	100%	104%	1%
Jacaranda mimosifolia	B	207,450	70%	215,859	-1%	95%	95%	4%	0%
Liquidambar styraciflua	B	96,553	72%	160,027	-7%	87%	93%	66%	1%
Lophestemon confertus	B	62,006	88%	70,072	-19%	61%	75%	13%	0%
Magnolia grandiflora	B	38,706	73%	148,203	7%	92%	86%	283%	2%
Metrosideros excelsa	B	88,613	80%	230,892	-9%	42%	46%	161%	3%
Olea europaea	B	205,110	51%	105,296	-12%	83%	94%	-49%	-2%
Phoenix canariensis	P	42,534	99%	68,888	-24%	64%	84%	62%	0%
Pinus canariensis	C	233,419	68%	262,396	47%	85%	58%	12%	1%
Pinus pinea	C	251,804	89%	183,910	9%	97%	89%	-27%	-1%
Pittosporum undulatum	B	539,596	40%	628,007	-24%	72%	95%	16%	2%
Platanus racemosa	B	290,274	99%	239,742	-14%	83%	97%	-17%	-1%
Podocarpus gracilior	B	57,421	95%	123,380	7%	92%	86%	115%	1%
Pyrus kawakamii	B	155,534	54%	173,642	31%	76%	58%	12%	0%
Quercus agrifolia	B	1,394,467	35%	1,353,001	15%	89%	77%	-3%	-1%
Schinus molle	B	17,203	99%	44,545	-25%	32%	43%	159%	0%
Schinus terebinthifolius	B	232,994	70%	176,253	-1%	93%	94%	-24%	-1%
Syzygium australe	B	126,894	39%	76,132	9%	87%	80%	-40%	-1%
Syagarus romanzoffiana	P	458,426	39%	49,241	-55%	36%	80%	-89%	-7%
Tipuana tipu	B	118,399	89%	156,739	14%	99%	87%	32%	1%
Ulmus parvifolia	B	20,795	99%	150,966	-18%	69%	84%	626%	2%
Washingtonia robusta	P	114,134	46%	46,724	-14%	66%	76%	-59%	-1%

agreement was achieved. For example, if a less common, persistent broadleaf, *Quercus virginiana*, was modeled as *Quercus agrifolia* or as *Magnolia grandiflora*, leaf type agreement was successful. In Alonzo et al. (2014) this method was tested on 2,304 crowns which represents approximately 1% of the total stems in the study area. In the current paper, we scale the method to map 290,000 stems (i-Tree estimate) or 179,043 segments generated from the canopy height model. This map allows for analysis of spatial patterns throughout the city and direct comparison to i-Tree estimates.

With the final map we offer additional analyses including: (1) error assessment based on comparison with the city of Santa Barbara's point dataset containing >450 species; (2) comparison of citywide canopy by species with i-Tree Eco estimates; (3) quantification of uncertainties in the remote sensing result due to segmentation and

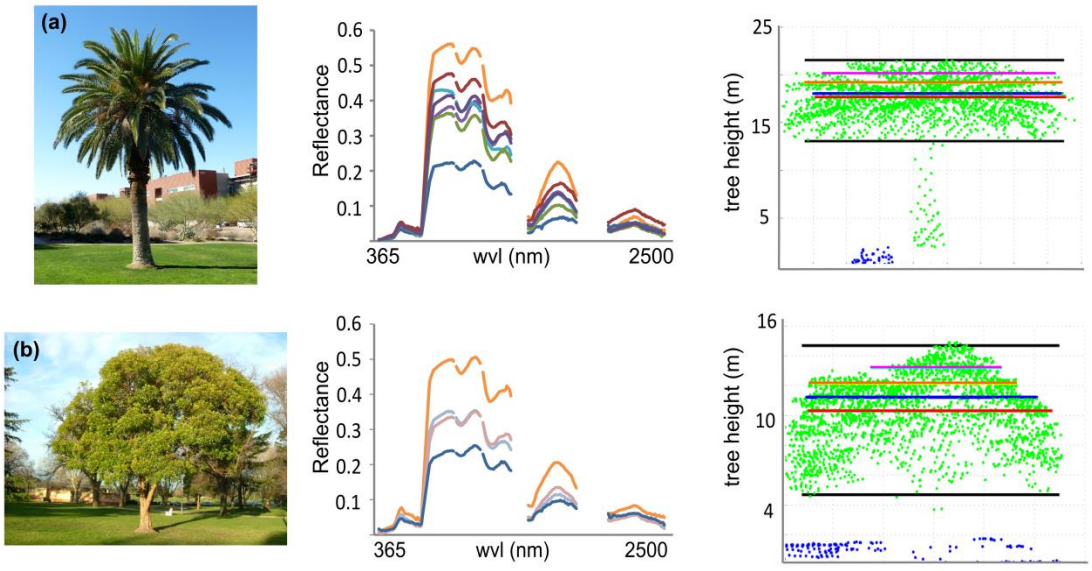

Figure 4.1 (a) *Phoenix canariensis photo*, AVIRIS spectra extracted from one crown, and set of crown-level structural measurements. Black lines indicate crown base and max heights. Red line is mean crown height. Dark blue is median height of returns in crown. Orange and violet show 75th and 90th percentile heights as well as widths at those heights. (b) Same set of information but for the broadleaf evergreen, *Lophostemon confertus*.

classification error and the uncertainties in the i-Tree Eco result stemming from sampling error of species and crown size distributions; and (4) analysis of how the classification algorithm mapped the uncommon species (i.e., whether leaf type agreement was achieved)

For canopy area, the i-Tree standard error of the estimate (σ_e) can be calculated based on the mean and standard deviation of canopy area (per species) across the 89 plots. The more evenly spread a species' canopy area is among 89 plots, the lower the σe. The error is driven by both the geographic distribution of a given species (presence/absence), and the sampling distribution of the species canopy area. It should be noted that in a standard i-Tree project, data are typically collected on 200+ plots which has yielded an average relative standard error in the stem count estimate of 12% (Nowak et al., 2008b). In this study, across 89 plots, for total stem count, σ_e = 19.3%. Thus, we expect that errors reported from the field-plot analysis are higher than normal. The most directly comparable estimate of remote sensing uncertainty is quantity disagreement. This metric is formulated for each species as (mapped canopy – actual canopy)/actual canopy. Like the i-Tree estimates, it does not take into account location disagreement, which is instead conveyed by the Producer's (error of omission) and User (error of commission) accuracies (**Table 4.1, cols F-H**, Congalton, 1991).

4.3.5 Leaf area index mapping

Leaf area index can be defined as the one-sided total leaf area per unit ground area (Chen & Black, 1992). It is an important ecophysiological quantity that, in urban areas, strongly mediates air pollution removal (e.g., Hirabayashi et al., 2011), dampening and delay of stormwater runoff (e.g., Xiao & McPherson, 2002), and cooling

through shading and evapotranspiration (e.g., Peters & McFadden, 2010). In i-Tree, LAI is measured allometrically using a formula initially created for deciduous, open-grown trees (Nowak, 1996):

$$\ln(\mathrm{LA}) = -4.3309 + 0.2942\mathrm{L} + 0.7312\mathrm{D} + 5.7217\mathrm{S} - 0.0148\mathrm{C} + \epsilon \qquad (3.1)$$

where *LA* is leaf area, *L* (crown length) is equal to the height of crown live top minus crown base height, *D* is the average crown diameter, *S* is a species-specific average shading factor, *C* is the crown's outer surface area represented by: $\pi D(L+D)/2$ and ϵ is an error term. LAI is computed for each sample tree by dividing *LA* by the crown projection area.

This study followed the method developed in Alonzo et al. (2015) for estimating LAI at the crown scale allometrically. In brief: From each segment, tree height and crown base height were estimated from the lidar point cloud. Average crown diameter was calculated from watershed crown area after abstracting each segment to its circle of equivalent area. The derived input variables were calculated with high accuracy: Crown diameter root mean squared error (RMSE) = 0.77 m (r^2 = 0.95), crown length RMSE = 1.72 m (r^2 = 0.87), and crown surface area RMSE = 47.1 m^2 (r^2 = 0.94). In place of a species-specific shading factor, one scaling equation was developed for each leaf type using weighted linear regression on the log-transformed predictor and response variables. The final equations were as follows:

$$\ln(\mathrm{LA_b}) = \ln(1.76 + 0.60\mathrm{L_b} + 2.32\mathrm{D_b} - 0.44\mathrm{C_b}) \qquad (3.2)$$

$$\ln(\mathrm{LA_n}) = \ln(-5.05 - 2.06\mathrm{L_n} - 5.38\mathrm{D_n} + 4.90\mathrm{C_n}) \qquad (3.3)$$

$$\ln(\mathrm{LA_p}) = \ln(7.02 + 2.11\mathrm{L_p} + 11.09\mathrm{D_p} - 5.33\mathrm{C_p}) \qquad (3.4)$$

where the subscripts b, n, and p denote broadleaf, needleleaf, and palm leaf types respectively. When applied to the segments intersecting with i-Tree field plots and validated against plot-aggregated values, the plot-level equation was y = 0.10 + 0.97x (RMSE = 0.53, r^2 = 0.84, where x is the lidar estimate and y is the field estimate) indicating good linear agreement.

In the present study, we (1) produce a citywide, crown-scale map of LAI; (2) sum total leaf area by distance and direction from single family homes as a first step to producing a citywide energy benefit map; (3) discuss errors and uncertainties of the remotely sensed map product compared to the i-Tree output.

4.3.6 Carbon mapping

Generally, inputs to allometric equations to predict carbon storage are either based on dbh alone, or dbh and tree height, and equations are specific for a given species when possible or matched by genus or leaf type otherwise (Pillsbury et al., 1998). It is difficult to estimate dbh accurately from airborne remote sensing platforms (Popescu, 2007). Instead, we relied on the same set of 28 structural metrics extracted from each crown segment by Alonzo et al. (2014). These metrics can be generally grouped as relating to tree height, tree width at various heights, crown porosity, vertical distribution of woody and foliar elements, and several derived ratio metrics. Using subsets of these structural measurements, we developed three equations, one for each leaf type, to directly predict dry-weight carbon storage. The general form of our scaling equation was:

$$C = b_0 (Var1)^{b_1} (Var2)^{b_2} \quad (3.5)$$

where C is dry-weight carbon (used synonymously with biomass) from field estimates, *Var1* and *Var2* are the best performing structural metrics and $b_{0,1,2}$ are empirically-determined coefficients. We also developed a single, pooled equation for the purpose of comparison. We hypothesized that the separate equations would predict biomass more accurately because they can more closely align with the species- or genus-specific equations found in i-Tree. Moreover, relationships between structure and biomass clearly vary by leaf type. For example, height has been shown to predict biomass well in a setting dominated by broadleaf trees (Raciti et al., 2014) but using only height to estimate palm tree biomass will likely lead to large overestimates.

Stepwise regression was used with a goal of minimizing the median of absolute deviation (MAD) of carbon estimates from remote sensing compared to 109 field measured crowns (70 broadleaf, 13 needleleaf, 26 palms). Two variables for each leaf type were selected. MAD was used instead of RMSE because it deemphasizes outlier error and should lead to higher map accuracy albeit at the cost of reducing the accuracy of the citywide estimate. For each leaf type and for the pooled model, the selected variables were used to determine coefficients for equation 5 and, following correction for logarithmic bias (Sprugel, 1983) applied to the entire study area where the leaf type information was available from the species classification result. The mapped result was validated by aggregating the crown level carbon estimates from lidar to the field-plot level and comparing with aggregated estimates from field measurements.

4.4 Results and discussion

4.4.1 Tree species mapping

A species map for the full study area was generated at crown scale and also, for visual clarity resampled to a 10 m grid (**Fig 4.2**). As previously reported in Alonzo et al. (2014) the sample overall accuracy of the map for the canopy area of the 29 common species is 83.4%. The individual species accuracies ranged between 37% for the small, broadleaf *Metrosideros excelsa*, and 96% for the large broadleaved *Eucalyptus globulus* and *Ficus microcarpa*. This number is reported as the average of the Producer's and User accuracies which are broken out in **Table 4.1 columns G and H**.

We conducted further error analysis on the citywide map (assessed against City of Santa Barbara street tree data) to confirm the stability of the sample results for the ten species with the most canopy area. It is not as straightforward to assess error for the full map because of the significant number of less common species that were mapped as one of the 29 common species. However, it is also interesting to note which uncommon species were most likely classified as one of the ten most common species (**Table 4.2**). This information can begin to address questions about plant functional similarities that manifest both spectrally and structurally. The Producer's accuracies for the street tree sample were consistent with the sample accuracies from Alonzo et al. (2014). In lieu of User accuracies, we show here the percent of canopy mapped as a common species that is actually an uncommon species. For example, the uncommon species mapped more than any other as *Eucalyptus globulus* was *Eucalyptus cornuta*, and overall, 42% of street trees mapped as *Eucalyptus globulus* were actually uncommon species. Likewise, 20% of canopy mapped as *Quercus agrifolia* was actually an uncommon species, most

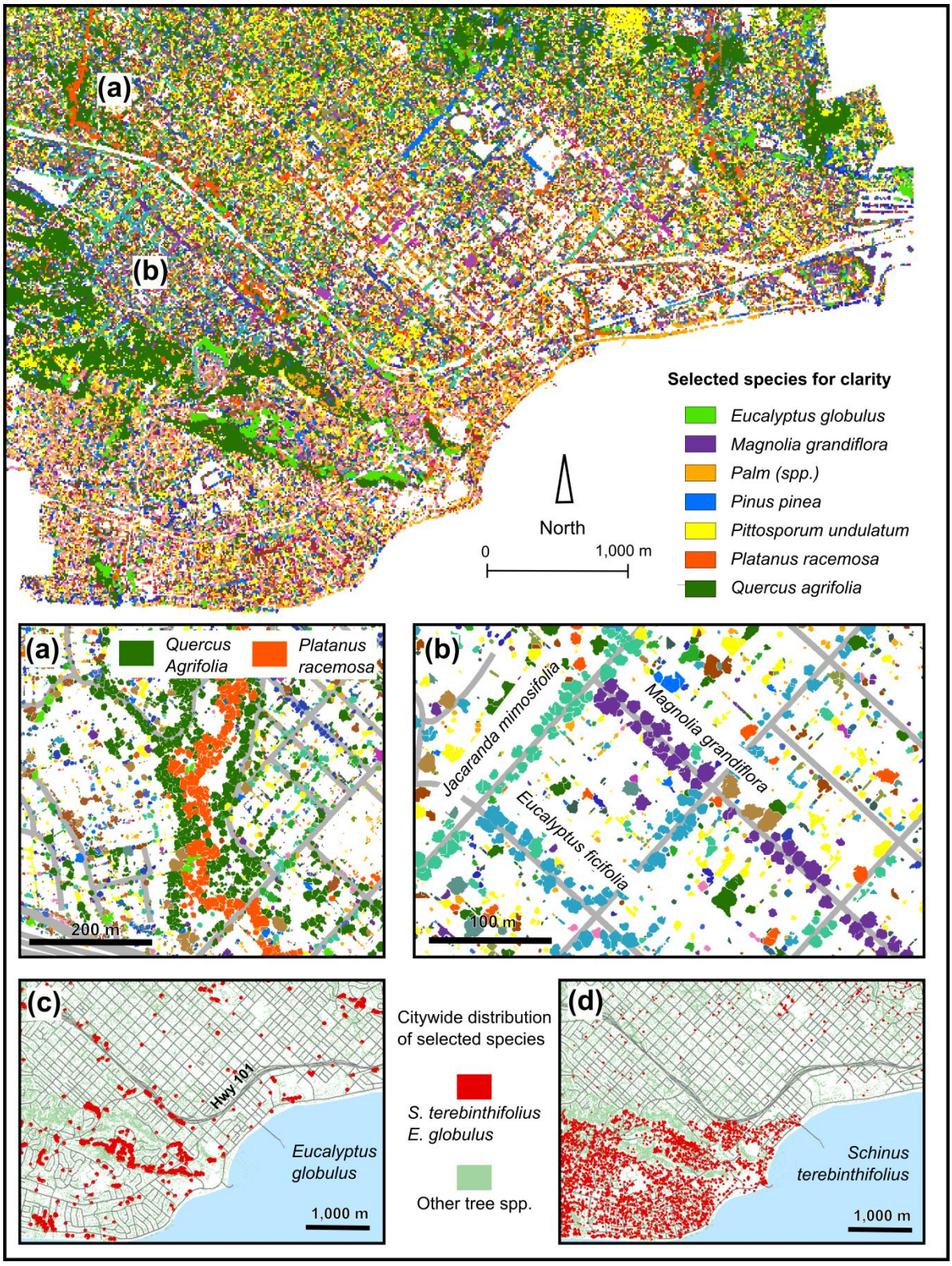

Figure 4.2: Species maps. The main map shows the species with the most canopy in each 10 m grid square. (A) Riparian corridor; (B) Linear street tree planting; (C) Distribution of *E. Globulus*; (D) Distribution of *S. terebinthifolius* that is partly an artifact of illumination and sensor geometry. See table 2 for species code key.

frequently *Quercus ilex*. In all cases but one (*Cupressus macrocarpa* mapping *Melaleuca quinquinervia*) leaf type agreement was achieved.

In certain cases, the degree to which a common species tended to be over-mapped relative to either other common species or uncommon species was flight-line dependent. *Schinus terebinthifolius* was much more frequently selected to map uncommon species, particularly *Cupaniopsis anacardioides*, in the southernmost flight line (**Fig 4.2d**). City tree planting data does indicate that more *Schinus terebinthifolius* were planted in this topographically distinct part of the city but the additional spatial correspondence with the flight line is indisputable. This outco11me can be attributed to poor lighting geometry: the solar zenith angle for this flight line was 54° which likely resulted in signal attenuation, especially in a forward scattering sensor view

Table 4.2: Secondary accuracy assessment for ten most important species in terms of canopy area using City of Santa Barbara street trees as validation dataset. Instead of User accuracy we show the extent to which uncommon species were mapped as one of these common species. For example, 40% of trees that were labeled on the map as *C. macrocarpa* were actually an uncommon species. The uncommon species that was most frequently labeled as *C. macrocarpa* was *M. quinquinervia*. The former is needleleaf while the latter is broadleaf so leaf type agreement, in this case, was not achieved.

Common species botanical name	Producers accuracy	Mapped uncommon	Most frequently mapped uncommon species	Leaf type agreement?
Cupressus macrocarpa	98%	40%	*Melaleuca quinquinervia*	No
Eucalyptus globulus	98%	42%	*Eucalyptus cornuta*	Yes
Jacaranda mimosifolia	87%	10%	*Maytenus boaria*	Yes
Pinus canariensis	57%	33%	*Pinus torreyana*	Yes
Pinus pinea	97%	19%	*Pinus halepensis*	Yes
Pittosporum undulatum	47%	34%	*Brachychiton discolor*	Yes
Platanus racemosa	82%	32%	*Koelreuteria bipinnata*	Yes
Quercus agrifolia	76%	20%	*Quercus ilex*	Yes
Schinus terebinthifolius	83%	33%	*Cupaniopsis anacardioides*	Yes
Syagrus romanzoffianum	26%	13%	*Trachycarpus fortunei*	Yes

configuration. That *Cupaniopsis anacardioides*, in particular, was most frequently mapped as *Schinus terebinthifolius*, is not surprising: They are quite similar in terms of crown structure, leaf size, and leaf color.

More often, species distributions were scattered throughout the city with locations more dependent on land use and ecological niches than any remote sensing artifact. In a natural setting, our map showed the interspersion of the common riparian species, *Platanus racemosa* among the *Quercus agrifolia* stands following a stream channel (**Fig 4.2a**). Our map also characterized the linear planting patterns along city streets well: Note the preferential planting of *Magnolia grandiflora*, *Jacaranda mimosifolia*, and *Eucalyptus ficifolia* in a single-family residential neighborhood (**Fig 4.2b**). Finally, on a citywide scale, one can see the locally-clumped yet globally distributed pattern of *Eucalyptus globulus* (**Fig 4.2c**). The clumps are likely naturally regenerating *Eucalyptus* stands frequently colocated with oak woodlands, while the linear patterns (most notably along Highway 101) likely represent intentional planting for aesthetic purposes and noise abatement.

4.4.2 Comparing remote sensing and i-Tree species results

Total canopy area for the 25 common species[1] that were sampled by both remote sensing and i-Tree was estimated as 5,584,589 m^2 based on field sampling and 5,288,304 m^2 based on remote sensing classification (5% difference). Before comparing species estimates and associated error metrics, the limitations of this analysis must be stated: (1) i-Tree canopy is multi-layered, therefore the canopy totaled across all

[1] Four of the 29 common species (*Cinnamomum camphora, Eucalyptus ficifolia, Geijera parviflora,* and *Stenocarpus sinuatus*) were considered "common" based on City of Santa Barbara data but were not recorded in any UFORE plots.

species includes understory trees. The remote sensing estimate is single-layered, meaning that the sum of all canopy cover equals the citywide total canopy cover. That said, the median crown light exposure for the 612 i-Tree trees was 4 (meaning 4 out of 5 "sides" of the tree are exposed to significant light), so understory trees are not prevalent. (2) The remote sensing estimates include uncommon species that were mapped as one of the common species. It is unknown whether the final remote sensing result over- or under- estimates the actual canopy of the common species.

The remote sensing estimates were checked for deviation from the i-Tree estimate (**Table 4.1, col I**) and also for area-weighted deviation based on the i-Tree canopy estimates (**Table 4.1, col J**). To allow for this exercise, the i-Tree estimates were considered to be "ground truth". *Quercus agrifolia* i-Tree estimates and remote sensing outputs were within 3% of one another, and within 1% when weighted by area. Other common, large canopy species, such as *Jacaranda mimosifolia* and *Pinus canariensis* were quantified similarly with both methods. The largest area-weighted discrepancies were found for the palm, *Syagarus romanzoffiana* and *Eucalyptus globulus*.

The methods we used do not permit definitive assertions about which result is more accurate, due in part to inflated i-Tree Eco species σ_e because of the low field plot count (n=89). Nevertheless, some inferences can be made based on measures of uncertainty. For all but two feather palm species (*Archontophoenix cunninghamia* and *Syagarus romanzoffiana*) the remote sensing quantity disagreement (**Table 4.1, col F**) was less than the i-Tree σ_e (**Table 4.1, col D**), frequently by a substantial amount. The largest discrepancies in uncertainties are found in species that were infrequently sampled in our field plots (e.g., *Platanus racemosa*, *Jacaranda mimosifolia*, *Magnolia*

grandiflora) but have generally large canopies which both increases the i-Tree σ_e and facilitates accurate classification using hyperspectral imagery. For example, there was only one *Platanus racemosa* sampled in the field but its crown projection area at 450 m^2 was more than twice as large as the next-largest species' average crown size. For a species like *Magnolia grandiflora*, remote sensing methods are likely more accurate because of the non-spatially random manner in which it is planted. Many of the city's *Magnolia grandiflora's* are planted as street trees on dedicated Magnolia street segments. This can lead to uneven presence-absence data from field sampling but adequate characterization from remote sensing due to large crowns and perhaps due to unique leaf properties. On the other hand, the small but common feather palms, *Archontophoenix cunninghamia* and *Syagarus romanzoffiana* are better estimated using field methods. This is especially true for *Syagarus romanzoffiana* because it was very common, relatively evenly distributed across land use classes, and is difficult to distinguish from *Archontophoenix cunninghamia* using remote sensing.

4.4.3 Leaf area index mapping

Using the lidar-based allometric method, the average LAI of the entire study area was estimated to be 1.15 with LAI within canopy of 4.47 (**Fig 4.3**). This was 11% less than the field-based estimate of 1.29 (canopy LAI = 5.57). The model uncertainties associated with lidar-based and field-based estimates of LAI are unknown. However, Alonzo et al. (2015) showed that the crown-scale, lidar-derived LAI values closely matched effective LAI estimates made using laser penetration metrics, which are lidar proxies for gap fraction. That two methods relying on different theoretical

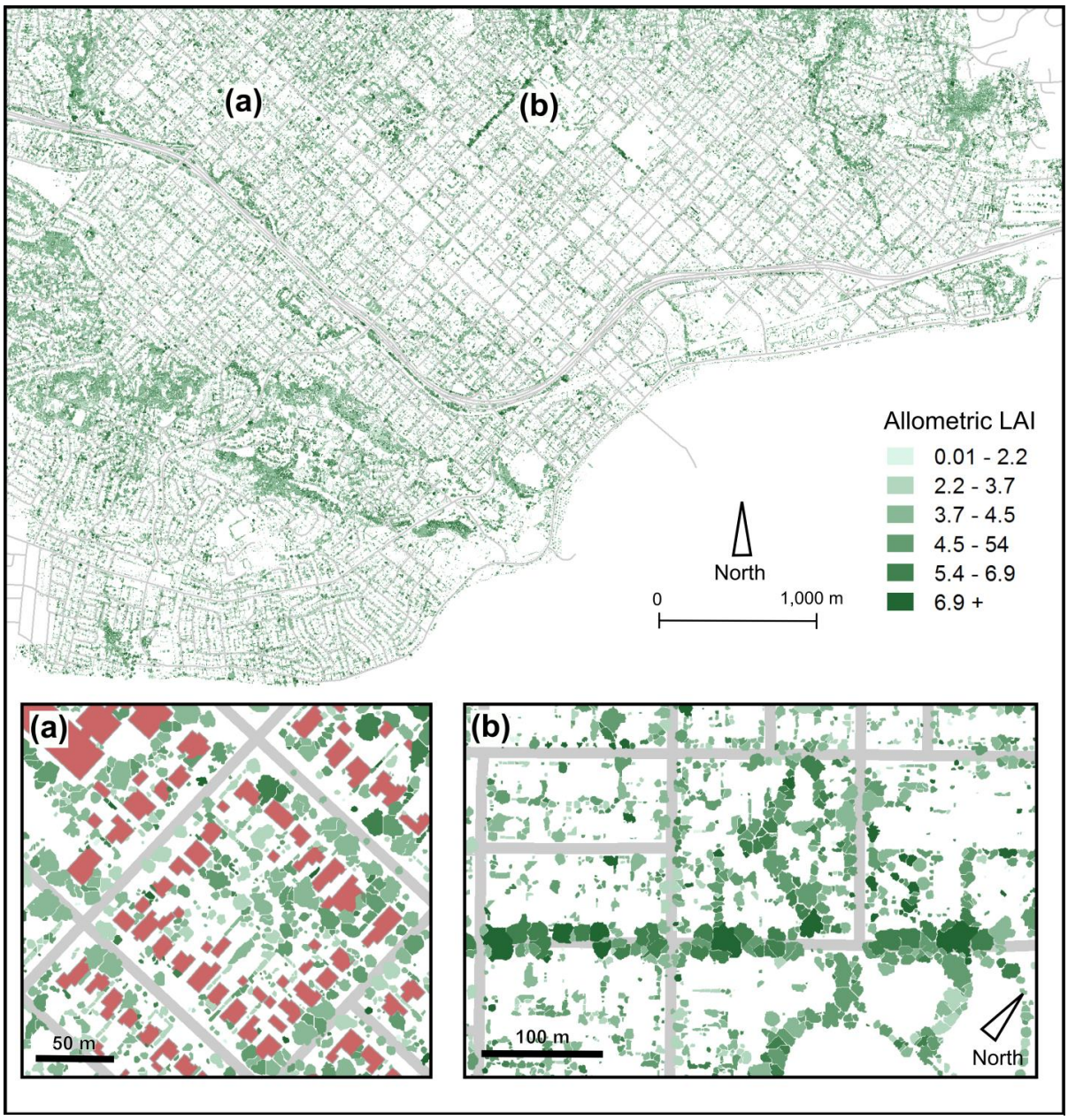

Figure 4.3: The main map shows LAI calculated using allometric methods at the crown scale. (A) Example of fine scale spatial relationship between buildings and canopy. (B) Example of canopy overhanging impervious surface.

underpinnings could produce similar results increases confidence that the final map has low model-driven error. The most likely source of error in the map is due to extrapolation of the regression models beyond the parameters of the field sampled data. The model was trained on a relatively small set of well-defined crowns but applied

to automatically-delineated crown segments whose dimensions were frequently more varied and irregular. Measurement error was deemed negligible for both field estimates and lidar estimates. In the field, the mean absolute error of height measurements was estimated to be <1m through a series of re-measurements. For the lidar data on flat ground, the vertical measurement uncertainty was 0.02 m and Alonzo et al. (2015) showed a strong relationship between field-measured and lidar-measured tree crown length (height minus crown base height).

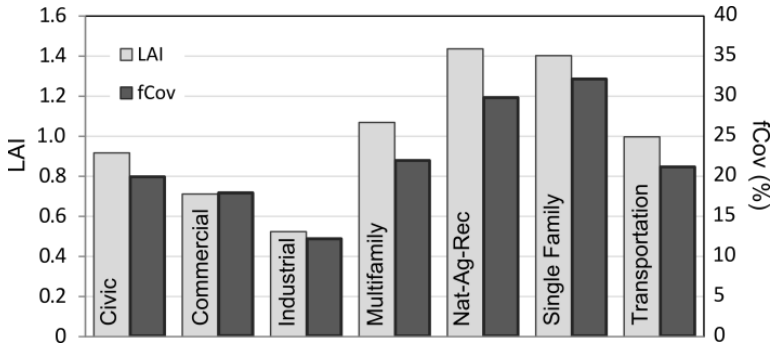

Figure 4.4: Leaf area index (LAI) and canopy fractional cover by land use

Leaf area index is the primary biophysical control on the magnitude of ecosystem services provided by trees in urban areas. However, the ultimate social value of benefits such as air pollution reduction, avoided stormwater runoff, and building energy-use modification are highly dependent on the location and context within the urban environment. With a mapped result we are able to report on, for example, LAI by land use (**Fig 4.4**) and LAI relative to climate-controlled structures (**Fig 4.5** and **Fig 4.3a**). In the Santa Barbara study area, we used the "transportation" land-use class to represent roads and parking lots. The LAI of this class is 0.99, which derives from a fractional cover of 0.22 and an average canopy LAI of 4.48.

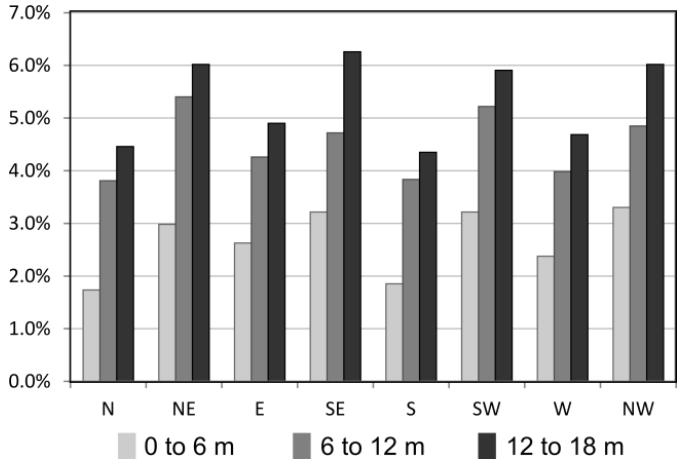

Figure 4.5: Average distance and direction of canopy from buildings as a percentage of all canopy within 18 m of buildings (Simpson et al. 2002).

Canopy overhanging roads and parking lots is particularly relevant from both an air pollution reduction and stormwater runoff management perspectives (Nowak et al., 2014; Wang et al., 2008). Air pollution removal can be high along roadways due to high pollution, however it should also be noted that canopy can trap ground-level emissions thus producing a net increase in local concentrations (Gromke & Ruck, 2009; Wania et al., 2012) Canopy located optimally near buildings can, through direct shading and evapotranspirative cooling, reduce energy use and consequently, CO_2 emissions from power plants. In the summertime, to maximize the cooling benefits of trees, the preferred locations are the east, south, and west sides of buildings in order to shade structures (Simpson, 2002). In this study, we found that the most canopy was found to the northeast side of buildings (14.3%), but very similar amounts were found to the NW, SW, and SE (**Fig 4.5**). It is likely that these directions are more heavily planted compared to the cardinal directions because of the predominant NW-SE layout of Santa Barbara's street grid. The latter set of results could be combined with lidar-extracted

building information and perhaps utility data to more completely model the spatial variation in energy use throughout a city.

4.4.4 Carbon storage mapping

We developed two models for dry-weight carbon (kg), one with separate equations for each leaf type and one for all leaf types pooled together. Using stepwise regression on all possible combinations of the 28 lidar structural measurements the leaf type equations were as follows:

$$\ln(\mathrm{C_b}) = \ln(0.10 + 1.31\mathrm{H_b} + 1.63\mathrm{W_b}) \quad (3.6)$$

$$\ln(\mathrm{C_n}) = \ln(-1.30 + 3.36\mathrm{H_n} - 2.55\mathrm{W_n}) \quad (3.7)$$

$$\ln(\mathrm{C_p}) = \ln(-2.27 + 2.82\mathrm{H_p} - 0.58\mathrm{W_p}) \quad (3.8)$$

The subscripts *b*, *n*, and *p* denote broadleaf, needleleaf, and palm leaf types respectively. *H* represents a height variable and *W* represents a width variable (both in meters). Different specific height and width variables maximized the predictive power of the models for each leaf type. H_b was the median height of lidar returns within a crown while H_n and H_p were both represented by the simple maximum tree height. W_b was the width of the tree at mean canopy height, W_n was the ratio of crown width to height at 75th percentile height, and W_p was the ratio of crown height to width and the median height of the lidar returns within the crown. The pooled equation took the form of:

$$\ln(\mathrm{C_{pool}}) = \ln(0.09 + 1.12\mathrm{H_{pool}} + 1.86\mathrm{W_{pool}}) \quad (3.9)$$

where H_{pool} and W_{pool} were represented by the same height and width variables as the broadleaf model, thus, the two equations only differ in their coefficients.

In model formulation, the separated model yielded a better explanatory relationship between the lidar predictors and the i-Tree response (R^2_{sep} = 0.74, R^2_{pool}=0.69) as well as lower mean absolute deviation in kilograms (MAD_{sep}= 30 kg C, MAD_{pool}= 60 kg C). However, the R^2 values became roughly equal when tested using holdout cross-validation. This suggests that the separated models were overfit, perhaps due to limited training data for palms and needleleaf species. Both models were then applied to all crown segments that intersected with one of the 89 i-Tree plots. Neither was clearly superior with respect to all of R^2, bias, MAD, and RMSE (**Fig 4.7**). The separated model resulted in a MAD (240 kg C) which was 18% lower than the pooled MAD (290 kg C) and a slope of 1.03 compared with a pooled model slope of 0.91. However, the pooled model resulted in an RMSE (1020 kg C) which was 13% lower than the separated model RMSE (1170 kg C) and an R^2 of 0.75 compared to a separated

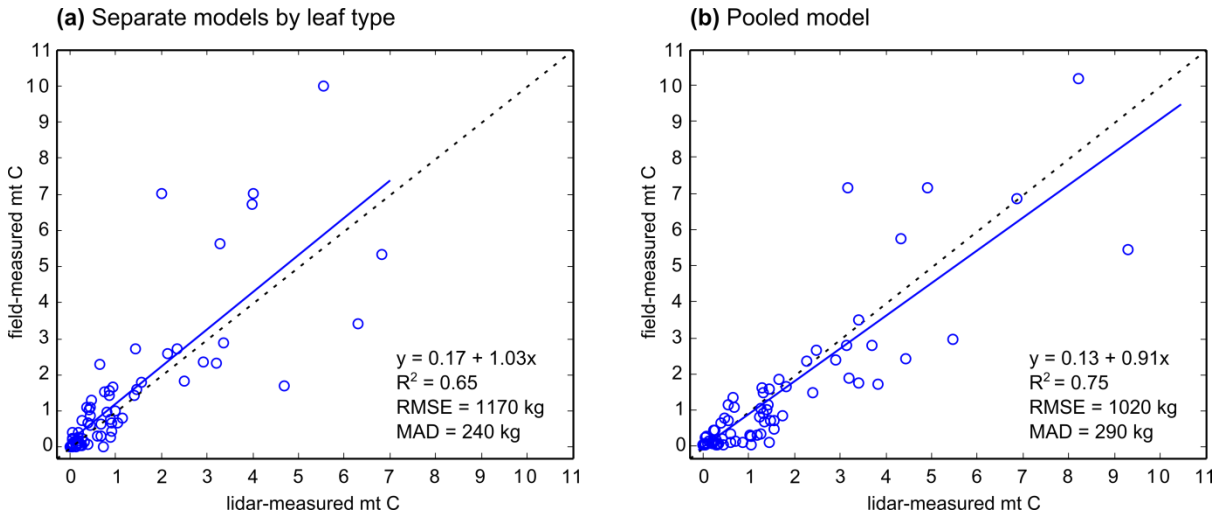

Figure 4.7: Predicting field-measured dry-weight carbon using lidar. (a) The plot-level result using one allometric equation for each leaf type (broadleaf, needleleaf, palm). (b) The same but with only one allometric equation for all trees. Dashed line is 1:1 and blue line is best linear fit.

model R^2 of 0.65. We believe these results suggest that the separated model was more accurate at more of the lower fractional cover plots, while the pooled model was more accurate at high cover, high biomass plots.

The variables and coefficients selected for each leaf type model and the pooled model shed light on the difficulty of constructing a universal, urban-tree biomass equation. For example, palms and needleleaf trees were best characterized by simple maximum height measurements but broadleaf height was represented by the median height of lidar returns in the crown. This possibly indicates that variability in broadleaf crown biomass is related to variability in the internal density of leaves and branches. Also, bole dimensions (e.g., dbh) are also important variables in estimating tree biomass and carbon. Previous studies have used a similar lidar metrics (e.g., height of median energy), to successfully estimate biomass in tropical (Drake et al., 2002) and temperate (Muss et al., 2011) forest settings. Palm trees, on the other hand, can be assumed more homogeneous in terms of their internal structure. It is also clear that palm biomass cannot be based strictly on height (e.g., consider the tall, spindly *Washingtonia robusta*). This is evidenced in the log-form equation by a complex compensation for height contribution to biomass by reduction in biomass with increased width and a negative intercept term.

The relative performance of the pooled and the separated models can also be examined vis-à-vis variables and coefficients. The variables in the pooled model were the same as those chosen for the broadleaf model confirming the dominance of broadleaf trees such as *Quercus agrifolia* and *Platanus racemosa* in the Southern California landscape. The coefficients differed mainly in that the pooled model placed

greater emphasis on width (β_{pool}=1.86, β_b=1.63) than height (β_{pool}=1.12, β_b=1.31). This difference likely occurs because, at least, palm trees require a model emphasizing crown width. This difference likely results in the better performance of the pooled model on plots with very large broadleaf trees. The biomass of these trees (e.g., a 25 m wide *Platanus racemosa*) was underestimated using the broadleaf model that was more driven by height. This example highlights a situation where a pixel-based biomass estimate may prove superior to a crown-based estimate. Raciti et al. (2014) successfully estimated carbon storage in Boston, Massachusetts' broadleaf-dominated urban forest using only lidar data gridded into a canopy height model. Nevertheless, there is clear degradation of the pooled model performance on plots with palm and needleleaf trees. This is most evident on plots with tall, *Washingtonia robusta*, whose biomass is vastly overestimated by the pooled model. Given the goal of creating a carbon storage map that maximizes accuracy at the most locations (rather than total carbon storage for the city), all further analysis was conducted using the model separated by leaf type.

The total amount of carbon stored in the study area's trees is 50,991 metric tons (t), based on aggregated crown-scale lidar estimates. The modeled carbon storage based on field sampling was 55,900 t. The most carbon per unit area and per unit canopy area was stored in Natural-Agricultural-Recreation land uses (**Fig 4.8b**). This pattern was expected because these areas are less developed and have coherent stands of large-canopy native species such as *Quercus agrifolia* and *Platanus racemosa* (**Fig 4.8a**). The least carbon per unit area was found in the industrial zone due to a combination of low stem counts and preferential planting of small canopy trees. However, in industrial zones the amount of carbon per unit canopy area remained high.

This may be driven by high modeled carbon density for palm trees which are more prevalent in these zones.

Citywide, the lidar estimate equates to a carbon storage density of 22.4 t C ha^{-1}. This value is below the national average urban value of 25.1 t C ha^{-1} reported in Nowak

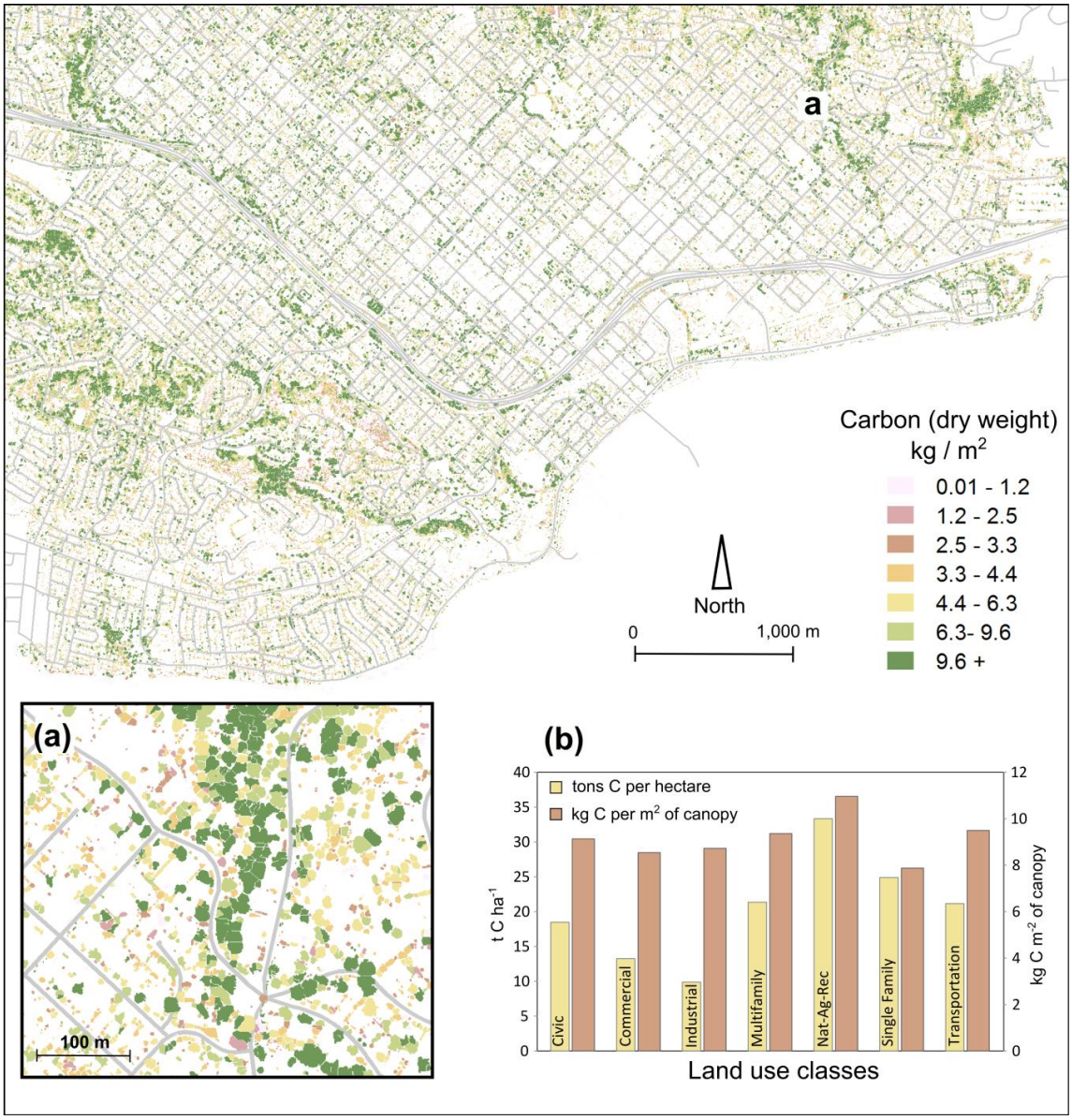

Figure 4.8: Main map shows dry-weight carbon per m^2 at the crown scale. (a) Zoom in on riparian corridor. (b) Carbon by land use

and Crane (2002), as would be expected given Santa Barbara's semi-arid climate. In our study area, the single family residential land-use class had a carbon storage density of 24.9 t C ha^{-1} aligning it more closely with the national estimate. However, this residential estimate is well below the 32.8 t C ha^{-1} estimate for residential land use in the naturally-forested setting of Boston (Raciti et al., 2014). Hutyra et al. (2011) estimated terrestrial carbon stocks along an urban to rural gradient in Seattle, Washington, creating a map of biomass based on a regional land cover map and mean biomass values for the grouped cover classes of heavy, medium, and low urban, and mixed and coniferous forest. They noted a clear trend of increased biomass per hectare with decreased urbanization intensity. We obtained similar results albeit at finer spatial resolution and thematic specificity: Carbon storage is lowest in industrial areas, and increases through commercial, multi-family residential, single family residential, to a high value in Natural-Agricultural-Recreation land uses (**Fig 4.8b**).

The above results are largely a function of simple variation in canopy cover and are not good indicators of carbon density per unit canopy area. When translated to this carbon density metric, Hutyra et al. (2011), for example, found 3.3, 7.1, and 11.6 kg C m^{-2} tree canopy on high, medium, and low intensity urban land uses respectively. Nowak et al. (2013b) aggregated carbon density rates from throughout the United States finding a mean of 7.69 kg C m^{-2} and a range of 3.14 to 14.1 kg C m^{-2}. They note that there is less variability among cities in this metric than a C-per-unit-land-area metric that is heavily influenced by fractional cover. In the downtown Santa Barbara study area the mean carbon density was 8.94 kg C m^{-2} , ranging narrowly from single family residential

on the low end (7.88 kg C m^{-2}) to Natural-Agricultural-Recreation on the high end (11.0 kg C m^{-2}, **Fig 4.8b**).

The citywide standard error of the plot-based estimate was 9,899 t (17.7%). This value represented sampling error only and did not account for the uncertainty of the biomass equations. The uncertainty associated with applying various biomass equations has been shown to have a variability up to 60 percent for population estimates (McHale et al., 2009). The lidar estimates have no sampling error but they do internalize the uncertainties of the i-Tree model related to application of biomass equations (i.e., the lidar models assume no error in the carbon estimates from i-Tree).

Additional uncertainty was introduced to the lidar estimates through generation of leaf-type models using linear regression. First, misclassification of crowns to the leaf-type level led to selection of the incorrect allometric equation. For example, the carbon storage of one correctly segmented juniper that was misclassified as a broadleaf was estimated by i-Tree to be 180 kg but only 20 kg in the lidar model. Second, generalization to the leaf-type level reduced sensitivity to species-level biomass variation. The carbon storage of the most prevalent species, *Quercus agrifolia* was consistently underestimated by the lidar model because the leaf type equation was trained on many broadleaf species whose crown heights and widths related to lower biomass. Nevertheless, it was shown above that the pooled model was likely adequate for large-scale estimation of biomass, especially in areas with high broadleaf fractional cover. This result opens the door for future accounting of urban biomass even where hyperspectral data for leaf type classification are not available.

4.5 Summary and Conclusions

Remote sensing methods can provide detailed information about the spatial variation in urban forest structure and function across the city. The resultant maps enhance our ability to understand spatially varying ecological, public health, social, economic, and management processes. The primary goal of this research was to estimate urban forest structure and function at the individual tree crown scale for every tree in an urban area using airborne remote sensing. Building on the results of Alonzo et al. (2014), we created a species map that included every tree in our 22 km^2 downtown Santa Barbara study area. Sampling error was eliminated from this product and the quantity disagreement born of classification error was substantially less than the standard error from plot-based estimates. Moreover, we showed that the spatial patterns of native species found in natural stands and introduced species planted in highly managed areas could be well represented. Visualization of these species distribution patterns at fine spatial resolution will be of immediate use to managers seeking to prepare for pests, target invasive species, or maximize their urban forest's ecosystem services (Alberti, 1999, 2005; Manning, 2008).

Second, we mapped leaf area index for each tree in the study area using lidar measurements of crowns and an allometric approach, following the methods in Alonzo et al. (2015). The uncertainty of the lidar-based estimate for each plot was considered to be on par with the field-based approach based on good alignment with both field-based results and results using a lidar proxy for gap fraction inversion. Again, sampling error is not present in the final, citywide map. We believe that, in the future, this spatial product will allow for refined modeling of stormwater runoff reduction and building

energy-use modification. Spatially explicit LAI facilitated mapping of air pollution reduction benefits at a very fine resolution compared to what has previously been available. Future research incorporating distributed measurements of pollution emissions would greatly improve the utility of the maps presented in this research.

Finally, we mapped urban tree carbon storage using lidar measurements of crown structure such as height and width, with allometric models. Results from lidar models separated by leaf type corresponded well to field-based estimates on the more urban plots (i.e., low fractional cover, low biomass) while results from the pooled model proved superior for plots in naturally-occurring, high biomass stands. While our remote sensing based maps do not have the sampling error of a field inventory, they do have uncertainties stemming from the application of the biomass equations used in i-Tree, remote sensing classification to the leaf type level, and extrapolation of the stepwise regression model beyond the values encountered in the field-measured training data. Still, the final, spatially explicit product offers finer grain insight into the capacity for urban areas to store carbon and how urbanization patterns mediate this process. A map like this could also be of immediate use for emergency managers: knowing the spatial distribution of tree biomass can improve planning for clean-up after storm events.

This research demonstrates the capability for wall-to-wall, spatial estimation of the urban forest structure and function using airborne remote sensing. Maps of urban forest structure extend plot sampling results significantly by allowing for (1) measurement of difficult-to-access areas (e.g., private property, steep slopes); (2) analysis of the spatial relationships between tree canopy and other spatial urban variables such as population, economic activity, or health; (3) more frequent repetition

of measurement and finer-grain (spatial and temporal) capacity for change detection; and (4) data consistency that is minimally dependent on field personnel. While there will always be a need for field-based urban ecosystem analysis, remote sensing data products greatly increase the breadth of scientific questions that can be asked as well as the specificity, reliability, and extensiveness with which those questions might be answered.

4.6 Acknowledgments

This research was funded by the first author's EPA-STAR Fellowship (FP-91768801-0), the National Science Foundation (BCS-0948914), and the Garden Club of America Fellowship in Urban Forestry. Special thanks to Robert Hoehn, Alexis Ellis, and Satoshi Hirabayashi for their work customizing, running, and communicating i-Tree Eco sub-models.

Chapter 5

Conclusions

5.1 Summary of research

The overall goal of this dissertation was to create a spatially explicit urban forest inventory using advanced, airborne remote sensing data and techniques. A suite of maps showing such structural attributes as tree species, leaf area index, and carbon storage will allow for more targeted urban forest management as well as improved ability to spatially match forest ecosystem services to the humans who will benefit. I have shown in three chapters how we can use hyperspectral imagery and lidar to identify tree species at the individual crown scale, how we can use lidar to measure LAI in two theoretically distinct ways, and how these techniques can be scaled for citywide analysis. The following paragraphs will review the key findings of each of these chapters and the subsequent sections will provide thoughts on future directions and applications of this research.

In **Chapter 2** I sought to improve species and leaf-type mapping in the urban forest. I selected 29 common species that dominate the canopy in Santa Barbara, California and classified them using canonical discriminant analysis (CDA) on combined hyperspectral and high point-density lidar data. I achieved a species-level accuracy among trained species of 83.4%. I mapped the *entire* set of sample crowns, including ~70 less common species, to the leaf-type level with 87.9% accuracy. I believe this study demonstrates the potential for separating highly overlapping species classes

using data fusion at the crown-object level. In an immediate, operational sense, the techniques described in this chapter are likely applicable with high accuracy for discriminating among urban vegetation growth forms (e.g. herbs, shrubs, trees) where simple structural metrics could vastly improve separability when combined with either multi- or hyperspectral data.

AVIRIS data were most important for species classification and lidar data only improved overall classification accuracy by 4.2 pp. However, I showed that small crowns and crowns with unique morphological characteristics were more apt to be correctly labeled with the inclusion of structural data. In line with previous research, species discrimination in this study was bolstered by lidar variables pertaining to tree height, crown morphology, and perhaps the internal arrangement of leaves and branches. Further, I showed that classification following automated crown segmentation was more accurate than a pixel-level result and the diminution in accuracy introduced from segmentation error was quite small.

In **Chapter 3,** I demonstrated the potential for mapping Leaf Area Index (LAI) in a heterogeneous urban environment using two theoretically distinct methods. I first showed strong agreement between effective LAI (LAI_e) estimated from a laser penetration metric (LPM) and LAI_e measured in the field using hemispherical photography ($r^2 = 0.82$). In order to quantify the relationship between the two methods, I developed novel methods to correct for off-nadir pulse angles and plot-level clumping in a structurally diverse and discontinuous canopy. While I initially made use of a lidar dataset with very high point density (22 pts m^{-2}), I found that the results could be reproduced at decimated point densities down to 2 pts m^{-2}. This indicates that cities

may have success implementing LPM methodology for calculating LAI_e using existing data.

Secondly, I showed that lidar-derived structural metrics such as height, crown base height, and crown segment area can be used as inputs to existing allometric equations for prediction of LAI. This result was compared against LAI allometrically estimated from field measurements of individual trees and yielded an r^2 of 0.84 when formulating separate models for each leaf type. A key difference in the mapped outputs between the methods was the spatial resolution. I found that a map produced using LPMs must have a pixel size large enough to allow for lidar beam penetration to ground, even under dense canopy. An allometric map has no theoretical lower bound on the size of the output crown object. However, maps generated using allometric equations are subject to an unknown amount of error associated with use of coefficients not developed for the trees or site in question. Thus, while this work demonstrates that LAI can be mapped at citywide scales, it is still important that the practitioner be aware of the trade-offs inherent in each of the methods. The ability to map LAI across large urban areas offers new potential to constrain estimates from hydrological and atmospheric models and better understand the spatial distribution of urban ecosystem services at increasingly fine scale.

The primary goal of the research communicated in **Chapter 4** was to scale the methods developed in the previous chapters such that they could be applied and evaluated for the entire Santa Barbara study area. Building on the results from **Chapter 2**, I created a wall-to-wall species map for 22 km^2 of downtown Santa Barbara, California. Sampling error was eliminated from this product and the quantity

disagreement born of classification error was substantially less than the standard error from plot-based estimates. Moreover, I showed that the spatial patterns of native species found in natural stands and introduced species planted in the urban core could be well represented. Visualization of these species distribution patterns at fine spatial resolution will be of immediate use to managers seeking to prepare for pests, target invasive species, or maximize their urban forest's ecosystem services.

Second, I mapped leaf area index for each tree in the study area using lidar measurements of crowns and an allometric approach, following the methods demonstrated in **Chapter 3**. The uncertainty of the lidar-based estimate for each plot was considered to be on par with the field-based approach based on good alignment with both field-based results and results using a lidar proxy for gap fraction inversion. Again, sampling error is not present in the final, citywide map. I believe that, in the future, this spatial product will allow for refined modeling of stormwater runoff reduction and building energy-use modification. Spatially explicit LAI plus block-level population information facilitated mapping of air pollution reduction benefits at a very fine resolution compared to what has previously been available. Future research incorporating distributed measurements of pollution emissions would greatly improve the utility of the maps presented in this research.

Finally, I mapped urban tree carbon storage using lidar measurements of crown structure such as height and width, with allometric models. Results from lidar models separated by leaf type corresponded well to field-based estimates on the more urban plots (i.e., low fractional cover, low biomass) while results from a pooled model proved superior for plots in naturally-occurring, high biomass stands. While my remote sensing

based maps do not have the sampling error of a field inventory, they do have uncertainties stemming from the application of the biomass equations used in i-Tree Eco, remote sensing classification to the leaf type level, and extrapolation of the stepwise regression model beyond the values encountered in the field-measured training data. Still, the final, spatially explicit product offers finer grain insight into the capacity for urban areas to sequester carbon and how urbanization patterns mediate this process.

5.2 Applications and future directions

This research demonstrates the capability for wall-to-wall, spatial expression of the urban forest structure and function using airborne remote sensing. Maps of urban forest structure extend plot sampling results significantly by allowing for (1) measurement of difficult-to-access areas (e.g., private property, steep slopes); (2) analysis of the spatial relationships between tree canopy and other spatial urban variables such as population, economic activity, or health; (3) more frequent repetition of measurement and finer-grain (spatial and temporal) capacity for change detection; and (4) data consistency that is minimally dependent on field personnel. It was always my intent with this dissertation to stay on the "applied side" of remote sensing science. Thus, it is not difficult to envision specific uses of the methods I have produced. The following paragraphs will highlight several research areas that may benefit from the use of these tools as well as ways in which these methods, that have thus far only been tested in Santa Barbara, could become more widely available.

5.2.1 Improved characterization of city-scale ecosystem processes

A host of satellite-based studies have evaluated the spatial distribution of the urban heat island relative to biophysical descriptors such as fractional cover of green vegetation or leaf area index (Gillies & Temesgen, 2004; Lu & Weng, 2006; Weng, 2009). These studies largely use moderate resolution imagery and spectral and temperature unmixing approaches to model the correlations between impervious surface and higher LSTs and vegetation and lower LSTs. Some have postulated that the uncertainty in the vegetation-LST relationship is due to either canopy stress or the presence of differing underlying substrate materials (Deng & Wu, 2013; Roberts et al., 2012).

Trees alter the local thermal environment through a combination of structural shading and evapotranspiration (ET). Conversely, canopy temperatures are mediated by the temperature of the underlying substrate and surrounding, vertical structures. The scale over which tree-driven cooling takes place varies as a function of time of day, season, canopy water status, canopy structure, and meteorological variables. The distance decay of the green-space cooling effect has been shown to be on the order of several hundred meters for large, wooded parks (e.g., Hamada & Ohta, 2010) with the directionality partially determined by advection (Oke, 1989). To date, there is little research at the fine scale coinciding with the manner in which trees are arranged in densely-populated settings.

In order to deconstruct the urban heat island, we need to characterize the fine scale variation in surface composition and 3-D structure across the urban realm. The methods developed in this dissertation allow for creation of these maps using hyperspectral imagery and lidar data. Future work could leverage the Goddard LiDAR,

Hyperspectral, and Thermal Airborne Imager(G-LiHT) sensor package which provides the ability to simultaneously capture not only hyperspectral and lidar data but ~1 m scale directional brightness temperature data (Cook et al., 2013). With this co-acquired dataset, one could potentially characterize the directional anisotropy of the thermal emissions based on surface microstructure and macro-scale facet (e.g., building wall) orientation.

Disaggregating urban temperatures will also improve the specificity with which we can model urban air pollution removal. Pollutant dry deposition on leaves is a function of LAI, meteorological conditions (e.g., temperature), and pollution concentrations (Nowak et al., 2014). However, the air pollution reduction results from plot-based studies and from the research presented in this dissertation rely largely on limited point measurements of pollution and weather. Thus, the *potential* reductions in NO_2, O_3, CO, SO_2, and PM2.5 can be mapped but the map correlates exactly with the map of LAI. In order to introduce appropriate variation based on actual pollution concentrations, disaggregated modeled, or measured emissions datasets are needed. Distributed emissions modeling has been undertaken in urban areas based on either land use regression (Hoek et al., 2008) or traffic intensity information (Beelen et al., 2008). These outputs could be combined with models of pollution dispersion along roadways (e.g., Gromke & Ruck, 2009), distributed temperature, and the relevant canopy information at the crown scale to greatly enhance the utility of this product.

5.2.2 Urban surface dynamics and larger scale atmospheric and hydrological modeling

Fine-scale variation in urban surface composition and 3-D vegetation structure can play an important role in the larger scale climate and hydrological models (Oleson et al., 2008; Tague & Band, 2004; Yang et al., 2013). At a watershed scale, the Physically Based Analytical Spatial Air Temperature and Humidity (PASATH) model distributes air temperature and humidity estimates based on gridded partitioning of surface energy fluxes (Yang et al., 2013). Tests of this model have shown high accuracy over impervious surfaces but divergence from measured air temperature over vegetation. There are potential gains through incorporation of high resolution maps of LAI and other urban surface parameters (e.g., albedo, emissivity) to constrain flux partitioning based on more precise characterization of urban surface composition. This should lead to improved estimates of urban temperature especially over areas with higher vegetation cover. Also at the watershed scale there is a call to better constrain the spatial relationships between canopy leaf type, leaf area, and the nature of the underlying surface. It has been shown that increased canopy in cities can reduce stormwater runoff but these relationships could be significantly refined with better inputs of deciduousness, leaf morphology, and substrate material (Xiao & McPherson, 2002).

On a global scale, climate models are beginning to include urban parameterizations for their land surface sub-models. For example in the Community Land Model (Oleson et al., 2008) there is now an urban component that allows for subgrid surface heterogeneity though only in the form of a simple canyon model (Oleson et al., 2010).

Using fused hyperspectral and lidar classification and structural estimation methods at an object scale has enormous potential to better constrain key surface parameters such as albedo (currently averaged by facet type) and to add vegetation function to the currently barren model landscape. Incorporation of VNIR-SWIR hyperspectral along with LST data has been explored as a constraint to urban surface energy balance schemes but thus far the use of lidar to characterize, for example, the heat trapping capacity of urban canyons, has been minimal (Xu et al., 2008).

5.2.3 Thoughts on the future of ecosystem structure measurements

The research published in this dissertation can immediately be employed in the modeling contexts discussed above. However, to take a step back, there is still considerable room for growth in urban vegetation characterization using either fused hyperspectral imagery and lidar, or more widely-available, less expensive datasets.

There is work in progress that expands some of the methods in this dissertation to apply to a broader typology of urban green infrastructure[2]. Whereas I only discriminated among 29 tree species, it is hoped that we can soon segment the urban landscape further based on simple object (spectral, structural) properties as well as compound object spatial context. Examples of simple object typologies that await detailed characterization in urban areas include: hedges, extensive green roofs, and lawn. Examples of compound objects whose spatial context may drive variation in ecosystem service provision include tree rows along median strips, trees with understory shrubs, or intensively cultivated green roofs. While the simple objects may

[2] Funding source for research in progress: Department of Earth and Environmental Science Division Forest, Nature and Landscapes, University KU Leuven, Belgium (Belspo), Urban Ecosystem Analysis supported by Remote Sensing UrbanEARS

be largely delineated using spectral information, the compound object delineation will require a set of rules also incorporating shape parameters and spatial relationships with neighbors. This multilevel object-oriented approach will represent an important step forward in linking high dimension, historically lower spatial resolution spectral information with object-oriented approaches originally developed for low dimension, hyperspatial data (e.g., IKONOS, GeoEYE).

It is also critical, in terms of broadening the impact of this work, that we explore the value of more widely available remote sensing datasets. While it is the case that cities like New York and Los Angeles have been imaged by both hyperspectral and lidar sensors, updates to the hyperspectral coverage is irregular at best. And smaller cities may have only lidar or no access to either dataset. I believe that some applied contributions of my work do not necessitate hyperspectral data but they do require at least, lidar and high resolution multispectral imagery. For instance, I showed that lidar pulse densities decimated to 2 pts m^{-2} can still lead to serviceable estimates of urban LAI. While species identification at an actionable accuracy is not yet possible with only multispectral data (e.g., WorldView-2; Pu & Landry, 2012), classification to the *leaf-type* level could be feasible with the inclusion of lidar measurements. At present, this level of classification is sufficient to parameterize some ecosystem models (e.g., i-Tree Eco).

The number of airborne sensors that can simultaneously acquire hyperspectral, lidar, and even thermal data is on the rise (Asner et al., 2007; Cook et al., 2013; Kampe, 2010). However, they are predominately tasked with collecting data over natural ecological study sites. Given that >80% of all Americans and >50% of all humans live in

cities (US Census Bureau, 2010; UN-DESA, 2014), it is my hope that a platform with multiple, high resolution sensors can be dedicated to urban inquiry in the near future.

Appendix A

Supplementary material for the publication "Urban tree species mapping using hyperspectral and lidar data fusion"

Authors: Michael Alonzo, Bodo Bookhagen, and Dar A. Roberts

The following items offer additional information and clarification beyond what could be included in the printed manuscript. All sections are referred to explicitly from the main document.

Matlab code: Code (*.m files) may be accessed via the links but is also available in printed form at the end of this appendix.

A.1 Cumulative canopy distributions as a function of species in three cities

- We aim to map 29 of the Santa Barbara species that contribute heavily to citywide canopy cover based on a combination of i-Tree Eco data collected from 105 plots and a geospatial database of publically managed trees from the City of Santa Barbara.
- We originally proposed to map 30 species. This number was based on visual analysis of **Fig. A.1** below (and **Fig. 1.3** in the main text) which shows the marginal increase in canopy cover by species. For instance, in Santa Barbara, *Quercus agrifolia* is estimated to be the largest single contributor of canopy cover with a 16% contribution to the total. It seemed that the gains in total canopy mapped for each species after the 30th would be small.
- Ultimately we chose to map 29 species, 20 of which were in the top 30 based on i-Tree Eco estimates. We excluded some species from the top 30 (e.g. *Grevillea robusta, Quercus suber*) because they are not actually common citywide. We excluded others (e.g. *Heteromeles arbutifolia*, *Myoporum laetum, Gleditsia triacanthos*) because they were generally understory trees.
- This project could be structured similarly in a place like Washington, DC given similar species-to-canopy-cover relationships. In a desert city like Los Angeles though, there are no ubiquitous, dominant canopy species. The two largest contributors, *Jacaranda mimosifolia* and *Magnolia grandiflora*, together only contribute 14% of total canopy. Thus, more species must be identified to arrive at the same 80% threshold.

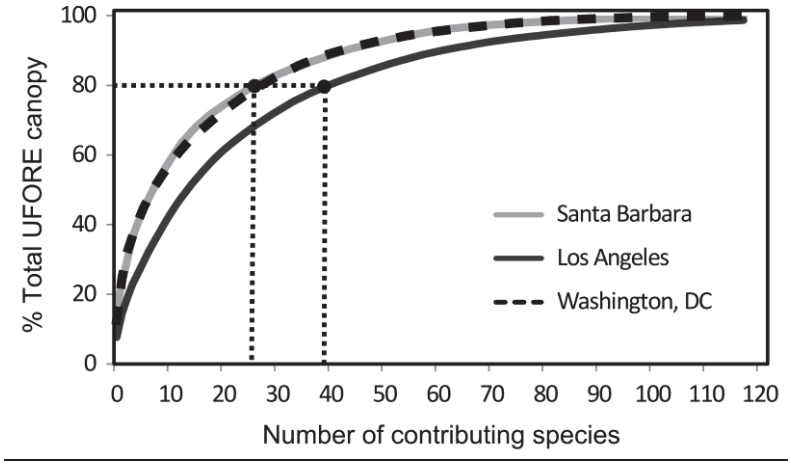

Figure A.1: Marginal contribution to overall canopy cover by species based on i-Tree Eco studies in Santa Barbara (2012), Washington, DC (Casey Trees, 2010), and Los Angeles (Clarke et al., 2013). The dotted lines indicate the number of species required to reach 80% canopy cover.

A.2: Lidar precision assessment and Matlab code

- The Matlab code (*Ground precision DR.m*) calculates the vertical precision of lidar returns. Lidar returns were sampled from smooth (but not necessarily flat) locations such as tennis courts, streets, and parking lots. Multivariate regression was run using UTM northings and eastings as the independent variables and elevation as the dependent variable. A plane was fit to the results and the root mean square error (RMSE) of the residuals (in meters) was calculated.
- Results are reported by sample polygon (which may include multiple flight lines) and by flight line (which may include multiple sample polygons)
    - Mean vertical error by polygon = 0.028 m
    - Standard deviation of vertical error by polygon = 0.016 m
    - Mean vertical error by flight line = 0.034 m
    - Standard deviation of vertical error by polygon = 0.023 m

A.3: Segmentation details and Matlab code

- We have linked to a Matlab script (*watershed for full scene.m*) that acts as a control file for the watershed segmentation and canopy maxima model (CMM) calculations underpinning this research For more background on watershed segmentation for tree delineation, see Chen et al. (2006).
- The *watershed for full scene.m* script calls several other custom functions used to:

- Execute watershed segmentation on a distance transformed binary canopy image (*watershedshape.m*)
- Generate a CMM using a window size that varies based on the empirical relationship shown below in **figure A.2**. (*cmmWrap.m*, *max_filter_variable_window.m*)
- Execute watershed segmentation on the CMM (*watershedsurface1.m*)

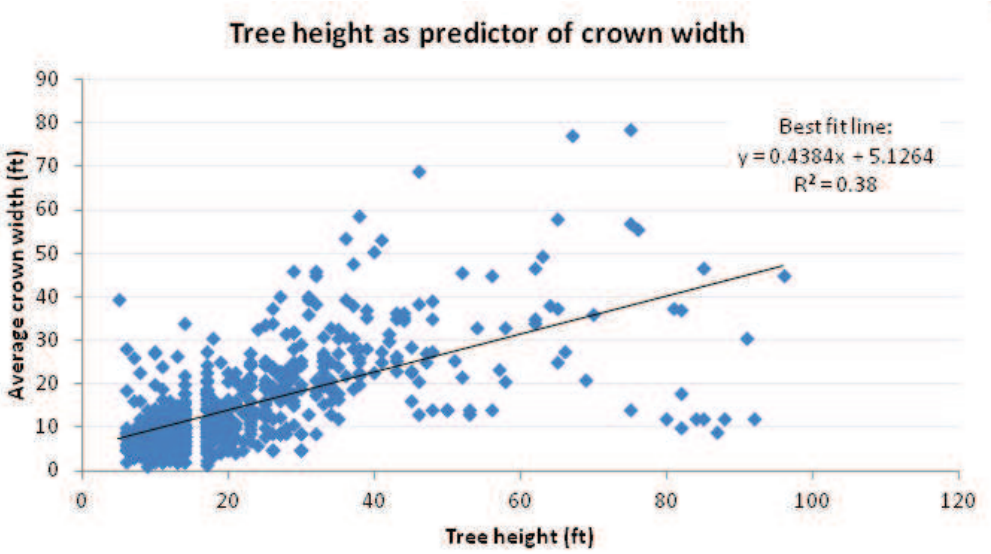

Figure A.2: Relationship between field-measured tree heights and crown widths. N = 612 trees from 105 randomly distributed plots in downtown Santa Barbara.

A.4: Structural variable computation and Matlab code (G = Goal, M = method)

- Maximum tree height (*h_1*)
  - **G:** Record the height above ground of the highest lidar return in the tree crown
  - **M**: Record height at 95th percentile height of returns in segment. This slightly underestimates height in unobstructed cases but filters out intruding returns from power lines or other overlapping trees.
- Median height of returns in crown (*h_2*)
  - **G:** This should be an indication of crown structure and porosity. One can envision a porous crown (e.g. PLRA) where many returns penetrate deeply having a lower *h_2* value than a crown of similar height with a very dense upper canopy (e.g. FIMI). Could be related to LAI.
  - **M:** Index all returns falling below *h_1* and above *h_5*. Calculate median height.

- Crown surface heights (*h_3, h_4*)
    - **G**: Assess height along surface of crown only. Height values will be lower than max tree height values as a function of crown closure and grid size selected for analysis.
    - **M**: Create grids of user-determined pixel sizes and overlay on top of crowns. In each grid square extract the height of highest return. This method was inspired by Kim et al. (2009). Matlab code: *surfacecalcs.m*
- Crown base height (*h_5*):
    - **G:** Record the height above ground of the lowest branches containing foliage.
    - **M:** First, a window was moved vertically through the crown from median height downward. When an empirically-defined lower bound on the number of points in the window was crossed, an initial CBH was established. If this CBH estimate was merely an area of low point density in the crown, the CBH was relocated downward to the 5th percentile height of returns. The percentile portion of the algorithm was inspired by Reitberger et al. (2009) though they were operating under leaf-off conditions.
- Crown width at various heights in crown (*w_1, w_2, w_3, w_4*)
    - **G:** Estimate widths at various heights as a measure of crown morphology.
    - **M**: The various percentile heights will already have been calculated. At each height determine a vertical crown slice size that is appropriate for your lidar dataset's point density (e.g. 1 m) and search outward, computing average radius of that slice in the x,y domain. Refer to the linked Matlab code: *widthatheight.m*
- Ratio of crown length to tree height (*hw_rat_1*)
    - **G:** Account for correlation between crown length and tree height. Separate species with different crown sizes (e.g. WARO and EUGL are similarly tall but have very different crown base heights)
    - **M**: Simple ratio: *(h_1 – h_5) / (h_1)*
- Ratios of crown height to width at selected heights in the crown (*hw_rat_2, hw_rat _3, hw_rat_4, hw_rat _5*)
    - **G**: Offer a normalized version of (*w_1, w_2,w_3, w_4*)
    - **M**: Simple ratios. For example, *hw_rat_2 = h_2 / w_1*
- Ratio of N-S width to E-W width (*hw_rat_6*)
    - **G:** Measurement of shape of crown projection area
    - **M:** Extents of each individual segment in the N-S and E-W directions.
- Average intensity above median height (*int_1)*

    - **G**: Measure uncalibrated return intensity in portion of crown above *h_2*. This may relate to internal crown structure in terms of distribution of woody vs. foliar surfaces.
    - **M**: Index pulse intensity values above *h_2* and take simple mean.
- Average intensity below median height (*int_2)*
    - **G**: Measure uncalibrated return intensity in portion of crown below *h_2*. This may relate to internal crown structure in terms of distribution of woody vs. foliar surfaces.
    - **M**: Index pulse intensity values below *h_2* and take simple mean.
- Crown surface intensities (*int_3, int_4)*
    - **G**: Assess intensity values along surface of crown only. Surface intensity values may more directly relate to foliage reflectance.
    - **M**: Create grids of user-determined pixel sizes and overlay on top of crowns. In each grid square extract intensity of highest return. This method was inspired by Kim et al. (2009). Matlab code: *surfacecalcs.m*
- Crown surface intensity / overall average crown intensity (*int_dist_1*)
    - **G**: Crown surface returns are likely functions of foliage reflectance. Normalize these values by overall intensity in the crown.
    - **M**: See above for crown surface calculations
- Skewness of intensity distribution through crown (*int_dist_2*)
    - **G**: Evaluate shape of distribution of intensity values throughout crown as indicator of internal crown structure.
    - **M**: Index all points in crown and compute skewness on their uncalibrated intensity values.
- Surface intensity (0.25 m) / surface intensity (1 m) (*int_dist_3*)
    - **G:** In dense upper crowns this ratio may be close to 1. In sparse crowns the 0.25 m grid may access returns somewhat lower in the crown.
    - **M**: Simple ratio: *int_3 = int_1 / int_2*
- Return intensity above median crown height / below median crown height (*int_dist_4*)
    - **G:** Normalize the above intensity metrics to account for larger spatial scale variability in the uncalibrated return intensities.
    - **M**: See above for crown surface calculations.
- Surface heights (0.25 m) / surface heights (1 m) (*cp_1*)
    - **G:** In dense upper crowns this ratio may be close to 1. In sparse crowns the 0.25 m grid may access returns somewhat lower in the crown.
    - **M:** See above for crown surface calculations.
- Mean crown height - median height of returns) / crown height (*cp_2*)

    - **G:** Measure the distribution of foliar and woody elements in the crown. If the median height of points is higher than the mean height, it may indicate a dense upper crown.
    - **M:** [(h_1+h_5)/2 – h_3] / h_1
- Counts of returns at various heights in crowns divided by the width at each height (*cp_3*, *cp_4, cp_5*)
    - **G**: Assess crown porosity by counting the number of points at different heights in the crown and normalizing by crown width at that height.
    - **M**: Use pre-established height percentiles and width calculations: simple count of points in a 0.5 m vertical disc.

Table A.1: Correlation matrix for all 28 structural metrics. The FFS-selected metrics for the watershed segments are in bold. Their correlations are outlined. Lower correlations are in green and higher correlations are in red.

	h_1	**h_2**	h_3	h_4	h_5	**w_1**	w_2	w_3	w_4	hw_rat_1	**hw_rat_2**	hw_rat_3	hw_rat_4	hw_rat_5	hw_rat_6	int_1	**int_2**	**int_3**	int_4	int_dist_1	int_dist_2	int_dist_3	int_dist_4	**cp_1**	cp_2	**cp_3**	cp_4	cp_5
h_1	1.00																											
h_2	0.96	1.00																										
h_3	0.97	1.00	1.00																									
h_4	0.97	0.99	1.00	1.00																								
h_5	0.54	0.58	0.58	0.56	1.00																							
w_1	0.25	0.18	0.18	0.17	0.27	1.00																						
w_2	0.10	0.04	0.05	0.03	0.26	0.95	1.00																					
w_3	0.18	0.15	0.15	0.14	0.26	0.95	0.94	1.00																				
w_4	0.26	0.25	0.24	0.23	0.28	0.92	0.88	0.94	1.00																			
hw_rat_1	0.49	0.38	0.39	0.41	0.37	0.07	0.07	0.02	0.05	1.00																		
hw_rat_2	0.25	0.28	0.28	0.28	0.06	0.22	0.25	0.22	0.15	0.16	1.00																	
hw_rat_3	0.25	0.23	0.23	0.24	0.06	0.23	0.28	0.25	0.24	0.17	0.15	1.00																
hw_rat_4	0.14	0.18	0.17	0.17	0.01	0.03	0.07	0.08	0.02	0.10	0.08	0.09	1.00															
hw_rat_5	0.22	0.26	0.26	0.25	0.02	0.12	0.21	0.17	0.09	0.17	0.08	0.12	0.21	1.00														
hw_rat_6	0.04	0.04	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.01	0.00	0.05	0.03	0.03	1.00													
int_1	0.07	0.10	0.09	0.07	0.10	0.04	0.03	0.05	0.04	0.03	0.11	0.01	0.06	0.07	0.01	1.00												
int_2	0.09	0.17	0.15	0.13	0.07	0.21	0.26	0.25	0.17	0.02	0.23	0.11	0.17	0.23	0.01	0.69	1.00											
int_3	0.01	0.05	0.04	0.02	0.10	0.11	0.09	0.11	0.09	0.10	0.12	0.04	0.06	0.08	0.01	0.88	0.84	1.00										
int_4	0.03	0.07	0.06	0.04	0.09	0.08	0.06	0.08	0.07	0.08	0.10	0.04	0.08	0.07	0.00	0.83	0.69	0.88	1.00									
int_dist_1	0.19	0.21	0.22	0.22	0.06	0.12	0.19	0.14	0.08	0.17	0.13	0.06	0.10	0.16	0.00	0.21	0.08	0.26	0.37	1.00								
int_dist_2	0.05	0.05	0.05	0.05	0.03	0.11	0.10	0.11	0.12	0.04	0.01	0.03	0.01	0.02	0.01	0.08	0.06	0.10	0.08	0.32	1.00							
int_dist_3	0.09	0.09	0.09	0.08	0.04	0.06	0.07	0.06	0.04	0.06	0.00	0.00	0.01	0.03	0.04	0.19	0.04	0.10	0.39	0.47	0.02	1.00						
int_dist_4	0.06	0.00	0.01	0.01	0.06	0.20	0.25	0.22	0.15	0.07	0.10	0.12	0.10	0.14	0.01	0.45	0.29	0.10	0.24	0.34	0.02	0.33	1.00					
cp_1	0.26	0.28	0.28	0.27	0.23	0.21	0.17	0.22	0.24	0.07	0.02	0.01	0.01	0.01	0.05	0.04	0.05	0.02	0.05	0.05	0.01	0.09	0.11	1.00				
cp_2	0.32	0.51	0.47	0.45	0.12	0.06	0.09	0.07	0.15	0.19	0.21	0.03	0.14	0.20	0.00	0.12	0.21	0.13	0.14	0.13	0.05	0.08	0.04	0.19	1.00			
cp_3	0.03	0.08	0.07	0.04	0.20	0.56	0.59	0.63	0.61	0.09	0.12	0.14	0.04	0.09	0.01	0.12	0.04	0.06	0.07	0.13	0.11	0.04	0.19	0.34	0.28	1.00		
cp_4	0.33	0.33	0.33	0.34	0.07	0.40	0.45	0.37	0.33	0.46	0.18	0.25	0.09	0.33	0.00	0.02	0.08	0.02	0.05	0.13	0.01	0.11	0.00	0.18	0.29	0.38	1.00	
cp_5	0.11	0.03	0.05	0.08	0.18	0.48	0.50	0.51	0.52	0.30	0.10	0.21	0.00	0.07	0.01	0.06	0.04	0.07	0.02	0.06	0.06	0.10	0.01	0.34	0.17	0.70	0.67	1.00

A.5: Spectral Band Correlation structure

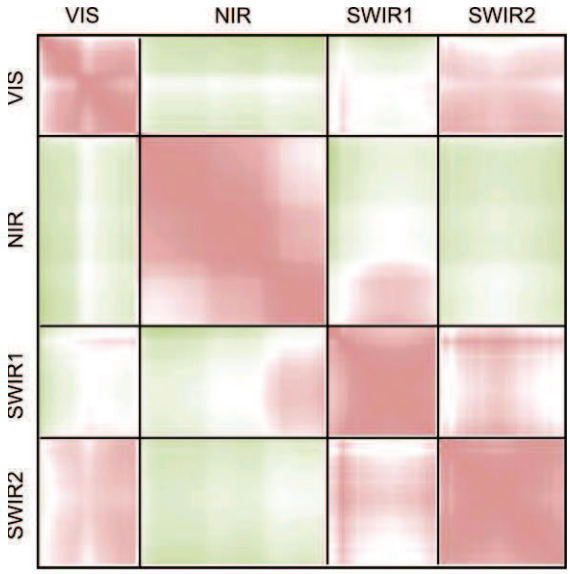

Figure A.3: Correlation among 178 AVIRIS bands. Higher correlation is in red and lower correlation is in green. VIS = Visible bands (395—734 nm), NIR = Near Infrared (744—1313 nm), SWIR1 = Shortwave Infrared 1 (1443—1802 nm), SWIR2 = Shortwave Infrared 2 (2018—2425 nm)

- Note the generally high correlation among bands in the same spectral region but lower correlation among green peak bands (the middle of the VIS box in Fig S7) and the rest of the VIS and the SWIR2.
- Note that there is generally high correlation among the VIS, SWIR1, and SWIR2 regions. This may provide partial explanation as to why there were many bands selected from the VIS region but fewer from either of the shortwave infrared regions.

References for Appendix A:

Casey Trees (2010). *i-Tree Ecosystem Analysis: Washington, DC*. URL: http://caseytrees.org/wp-content/uploads/2012/02/report-2010-01-ufore2009.pdf

Chen, Q., Baldocchi, D., Gong, P., & Kelly, M. (2006). Isolating Individual Trees in a Savanna Woodland Using Small Footprint Lidar Data. *Photogrammetric Engineering & Remote Sensing*, 72(8), 923–932.

Clarke, L.W., Jenerette, D.G., & Davila, A. (2013). The luxury of vegetation and the legacy of tree biodiversity in Los Angeles, CA. *Landscape and Urban Planning*, 116, 48—59.

Digabel, H., & Lantuéjoul, C. (1978). Iterative algorithms. In Proc. 2nd European Symp. Quantitative Analysis of Microstructures in Material Science, Biology and Medicine (Vol. 19, No. 7, p. 8). Stuttgart, West Germany: Riederer Verlag.

The following pages contain printed Matlab code related to Chapter 2:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Validation script for ground precision by flightline and by
polygon

%Inputs: A text file for las pts extracted from each validation
polygon
%with xyzpc attributes
%
%Ouput: RMSE by flight line and RMSE by polygon
%
%Parameters to set: The percentage of points to sample (in large
datasets the
%regression will not work). Default = 20% (0.2)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% percent of dataset to include in precision calculation
subsample = 0.4;

%storage location of text files containing lidar point info in 5
xyzpc
%format
base_path = '..\ground_precision_val\';

%get list of text files in the base dir
dir_txt = dir(strcat(base_path,'*.txt'));

%construct an error matrix where there is one row for each
flightline represented in
%each polygon. Compute mean squared error using multivariate
regression.
full_error_mat = [];

%loop through each text file (polygon)
for i = 1:size(dir_txt,1)

    fileName = dir_txt(i).name;

    %check if text file has data
    if dir_txt(i).bytes > 0
        M_full = dlmread(strcat(base_path,fileName));

        %need to subsample to maintain full rank matrix
        M_rand = rand(length(M_full),1);
        rand_idx = find(M_rand < subsample); %keep x% of data
        M = M_full(rand_idx,:);

        %find unique flight lines in this polygon
        lines = unique(M(:,4));
        %loop through each flight line represented within each
```

```
        polygon
        for j=1:length(lines)
            line_idx = find(M(:,4)==lines(j));
            constants = ones(length(line_idx),1);
            X=horzcat(constants, M(line_idx,1:2));
            z=M(line_idx,3);
            [b,bint,r] = regress(z,X); %r is vector of residuals
            error_mat(j,1) = i;
            error_mat(j,2) = lines(j);
            %need this count to weight average polygon error
            error_mat(j,3) = length(line_idx);
            %calculate RMSE
            error_mat(j,4) = sqrt(mean(r.^2));
            error_mat(j,5) = std(r);
        end
    full_error_mat = vertcat(full_error_mat,error_mat);
    clear error_mat

    end

end


%error aggregated by flightline (across polygons)
f_lines = unique(full_error_mat(:,2));
for i = 1:length(f_lines);
    line_idx = find(full_error_mat(:,2)==f_lines(i));
    poly_line_count = full_error_mat(line_idx,3);
    poly_line_err = full_error_mat(line_idx,4);
    %average error weighted by number of points per flight line
    (using inner
    %product)
    line_err(i,2) = (poly_line_count'*poly_line_err)/sum(
    poly_line_count);
    line_err(i,1) = f_lines(i);
end

%error aggregated by polygon (across flightlines)
polys = unique(full_error_mat(:,1));
for i = 1:length(polys);
    poly_idx = find(full_error_mat(:,1)==polys(i));
    poly_line_count = full_error_mat(poly_idx,3);
    poly_line_err = full_error_mat(poly_idx,4);
    %average error weighted by number of points per flight line
    (using
    %inner product)
    poly_err(i,2) = (poly_line_count'*poly_line_err)/sum(
    poly_line_count);
    poly_err(i,1) = polys(i);
```

```
end

%report line and poly error
line_err
poly_err


%visualize the fit between the points and a fit plane
scatter3(X(:,2),X(:,3),z,30,'filled')
hold on
x1fit = min(X(:,2)):0.1:max(X(:,2));
x2fit = min(X(:,3)):0.1:max(X(:,3));
[X1FIT,X2FIT] = meshgrid(x1fit,x2fit);
YFIT = b(1) + b(2)*X1FIT + b(3)*X2FIT;
mesh(X1FIT,X2FIT,YFIT)
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Control file for all watershed-segmentation related processes
%
%NOTE:  This code is for reference purposes only. You should not
assume
%that this will work for your specific imagery. Feel free to edit
as
%needed. Please cite Alonzo et al. (2014).
%
%Required functions (beyond standard Matlab image processing
functionality)
    %watershedshape.m
    %cmmWrap.m
    %max_filter_variable_window.m
    %watershedsurface1.m
    %multicore suite:
    http://www.mathworks.com/matlabcentral/fileexchange/13775-multi
    core-parallel-processing-on-multiple-cores
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%
%Input imagery:  Assumes gridded layers with some previous
classification
%(for instance: point cloud classification with gridded export from
%lastools)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%


%High vegetation geotiff with spatial reference
addpath('..\surface_models');
[veg_high,R_veg] = geotiffread(
'..\surface_models\veg_high_fill_sev_full.tif');
veg_high = double(veg_high);

%High unclassified imagery (along with separate tfw worldfile)
file_data= '..\surface_models\uncl_high_full_scene.tif';
uncl_high = double(imread(file_data));
worldFileName = getworldfilename(file_data);
uncl_high_tfw = worldfileread(worldFileName, 'planar', size(
uncl_high));

%Read in building data with separate tfw worldfile
file_data= '..\surface_models\bldg_elev_full_scene.tif';
```

```
bldg_elev = double(imread(file_data));
worldFileName = getworldfilename(file_data);
bldg_elev_tfw = worldfileread(worldFileName, 'planar', size(
bldg_elev));
%pad the building layer b/c the size is not the same as the
others (though
%the alignment is fine)
bldg_elev_pad = zeros(size(veg_high));
bldg_elev_pad(1:size(bldg_elev,1),1:size(bldg_elev,2)) = bldg_elev;
bldg_elev = bldg_elev_pad;

%bring in canopy class (binary image) with worldfile
file_data= '..\surface_models\canopy_class_full.tif';
canopy_class = logical(imread(file_data));
worldFileName = getworldfilename(file_data);
canopy_class_tfw = worldfileread(worldFileName, 'planar', size(
canopy_class));

%This section logically combines high veg, unclassified imagery,
and buildings that
%exist within the existing canopy classification. Purpose: use
the binary
%canopy layer as the gold standard and reclassify all layers
based on it.
veg_high_log = logical(veg_high);
veg_high_canopy_log = veg_high_log.*canopy_class;
uncl_high_log = logical(uncl_high);
bldg_elev_log = logical(bldg_elev);
bldg_elev_canopy_log = bldg_elev_log.*canopy_class;
bldg_elev_canopy = bldg_elev.*bldg_elev_canopy_log;
just_uncl_log = ~veg_high_log.*(uncl_high_log.*canopy_class);
just_uncl = just_uncl_log.*uncl_high;
just_bldg_log = ~(veg_high_log | uncl_high_log).*
bldg_elev_canopy_log;
just_bldg = just_bldg_log.*bldg_elev;
veg_high_canopy = veg_high.*veg_high_canopy_log;
veg_high = veg_high_canopy + just_uncl + just_bldg; %final canopy
layer

%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%
%basic morphological operations to clean binary canopy layer
% (took about 2 min to run on Santa Barbara scene)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%

%convert to binary image
```

```
veg_high_bw = im2bw(veg_high,0.05);

%morphological opening and closing
veg_high_O = bwmorph(veg_high_bw,'open');
veg_high_C = bwmorph(veg_high_O,'close');

%further cleaning
veg_high_T = bwmorph(veg_high_C,'thicken',1);
veg_high_F = bwmorph(veg_high_T,'fill');
veg_high_F = bwmorph(veg_high_F,'majority');

%a final elimination of canopy less than 16 px (=1sq meter at
0.25 m res)
veg_high_O2 = bwareaopen(veg_high_F,16);

%Thickening required because both watershed-shape and
%watershed-surf progressively thin segments because segment
boundaries have
%1-pixel thickness
veg_high_O2 = bwmorph(veg_high_O2,'thicken',2);

%cleaned up canopy height model (CHM)
veg_high_clean = medfilt2((veg_high_O2.*veg_high),[3 3]);
%binary image (for use in watershedshape and future cleaning)
veg_high_bin = veg_high_O2;


%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%
%Watershed segmentation baed on distance transform
%Required functions:
%   watershedshape.m
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%

%w-shape (took about 13 min to run on Santa Barbara scene)
[w_shape,w_shape_lines] = watershedshape(...
        veg_high_bin,50);

%Display results
% shape_over = imoverlay(mat2gray(veg_high), w_shape_lines,[.3 1
.3]);
% imtool(shape_over);

%clear stuff because memory starts to get tight
clear veg_high_FULL
clear veg_high_C
clear veg_high_T
```

```
clear veg_high_O

%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%
%Prepare segments for processing to CMM and watershed seg

%The main purpose of this code block is to put every segment (as
created
%from watershedshape) into its own cell so that it can be
processed to CMM
%indpendent of the rest of the image. This allows for a vast
reduction in
%processing time for two reasons:  1. CMM generation is only
happening
%where canopy exists (<40% of most urban scenes) and 2.
Parallelization
%using multicore.m is now easy to implement and allows for an
arbitrary
%decrease in processing time depending on how many cores are at
your
%disposal.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%

%pad the entire output of watershedshape to avoid border issues
w_shape_pad = padarray(w_shape, [10 10]);
w_shape_pad = bwlabel(w_shape_pad); %ensure this is label-style
veg_high_clean_pad = padarray(veg_high_clean, [10 10]);

%sort and count all of the watershedshape segments as labeled by
bwlabel
seg_ids = sortrows(unique(w_shape_pad));
seg_ids = seg_ids(2:end); %exclude zero
n = length(seg_ids); %-1 to exclude zero

%access area, bounding box extents, and bb ul corner coords
seg_props = regionprops(w_shape_pad,'Area','BoundingBox');

%This loop does two things:
%1. Clip segments from the full image based on bounding box,
%2. eliminate vegetation from other segments that has leaked
%   into the bounding box of the segment of interest,

   for i = 1:n
    %extract image subset
        m_ul(i) = floor(seg_props(seg_ids(i)).BoundingBox(2)) - 3;
         %"-3" = segment padding
```

```
        n_ul(i) = floor(seg_props(seg_ids(i)).BoundingBox(1)) - 3;
        m_ll(i) = m_ul(i) + seg_props(seg_ids(i)).BoundingBox(4) +
         6; %"+6" = segment padding
        n_ll(i) = n_ul(i) + seg_props(seg_ids(i)).BoundingBox(3) +
         6;

        %cut out the clean image and the logical sub to exclude
        any other
        %segments that are leaking into the bounding box. Veg
        from other
        %segments leaking in will cause problems when calculating
        CMM.
        veg_high_clean_pad_logical_sub{i} = w_shape_pad(m_ul(i):
        m_ll(i),n_ul(i):n_ll(i))==seg_ids(i);
        veg_high_clean_pad_sub{i} = veg_high_clean_pad(m_ul(i):
        m_ll(i),n_ul(i):n_ll(i)).*...
            veg_high_clean_pad_logical_sub{i};
    end


    %This loop fills the parameterCell in preparation for parallel
    %processing of CMM generation. The nested loop holds each
    segment while
    %the outer loop holds groups of segments (number of groups =
    subdiv).
    %Purpose: CMM parallelization works much better via multicore
    if the
    %packets sent out to slaves are large.
        count = 0;
        subdiv = 36;  %number of packets that will be divvied to
        slaves
        sub_size = floor(length(veg_high_clean_pad_sub)/subdiv);
        sub_remainder = length(veg_high_clean_pad_sub)-(sub_size *
         subdiv);
        for i = 1:subdiv

            for j = 1:length(veg_high_clean_pad_sub)/subdiv
                chm_sub{j} = veg_high_clean_pad_sub{count+j};
            end

            count = count + j;
            parameterCell{i} = {chm_sub};

            clear chm_sub

        end

%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%%%%%%%%%%
%Run the parallel process for creating the per-segment CMM
%Parallelization needs:
% Open an additional matlab window for each additional slave
process up to
% number-of-available-cores - 1
% run startmulticoreslave.m on each

%Required functions:
%   cmmWrap.m (if you're grouping segments in larger sets)
%   max_filter_variable_window.m
%   the suite of multicoremaster functions

%DO NOT RUN THIS BLOCK IF YOU ALREADY HAVE A CMM CREATED
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%

%parallelize the processing of the CMM
%fprintf('About to start CMM for: %u\n',plots)
tic
cmm_out = startmulticoremaster(@cmmWrap,...
    parameterCell);
toc_cmm = toc;
%fprintf('Completed CMM for: %u\n',plots)

%cell storing packages of processed CMMs
cmm_storage = cmm_out;

%IF YOU USED cmmWrap: NEED TO BREAK SEGS OUT OF THEIR packages
cmm_break = [];
for i = 1:length(cmm_out)
    cmm_break = vertcat(cmm_break, cmm_out{i});
end
cmm_out = cmm_break; %reassign for downstream processing


%define output images
w_surf_plot_pad = zeros(size(veg_high_clean_pad));
cmm_plot_pad = zeros(size(veg_high_clean_pad));

%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%
%Run the CMM-based watershed segmentation
%Required functions:
%   watershedsurface1.m
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%
```

```
%gather data about segments to check to see whether further
segmentation should occur
stats_bw = regionprops(logical(w_shape_pad),'eccentricity','area');
stats_veg_high = regionprops(logical(w_shape_pad), ...
veg_high_clean_pad,'MeanIntensity');

%Run the CMM-based watershed segmentation and restitch CMM and
final
%segments back into full images

for i=1:n

    %SWAP COMMENTS WHEN ACCESSING PRE-CREATED CMM
    %reduce speckle of cmm with median filter
    cmm_med3 = medfilt2(cmm_out{i}, [3 3]);
    %cmm_med3 = medfilt2(cmm_storage{i}, [3 3]);

    %output should not have grown due to max filtering;
    re-constrain
    cmm_constrained{i} = cmm_med3.*veg_high_clean_pad_logical_sub{
    i};

   %only send segment for further segmentation if it's large and
   not very
   %circular. These particular values were determined empirically
   based on
   %a sampling of validation segments.
   if stats_bw(i).Area < 1000 && stats_bw(i).Eccentricity < 0.7
        %more_seg(i) = 0;
        w_surf{i} = logical(cmm_constrained{i});
    else
        more_seg(i) = 1;
        [w_surf{i},w_surf_lines{i},w_surf_maxes{i}] = ...
        watershedsurface1(cmm_constrained{i},more_seg(i));
    end

    %more_seg(i) = 1;

    %iteratively restitch the segments into the full scene
    w_surf_plot_pad(m_ul(i):m_ll(i),n_ul(i):n_ll(i)) = ...
        w_surf_plot_pad(m_ul(i):m_ll(i),n_ul(i):n_ll(i)) + ...
        double(w_surf{i});

    cmm_plot_pad(m_ul(i):m_ll(i),n_ul(i):n_ll(i)) = ...
        cmm_plot_pad(m_ul(i):m_ll(i),n_ul(i):n_ll(i)) + ...
        cmm_constrained{i};
end
```

```
%ditch padding on image
cmm_plot = cmm_plot_pad(11:size(cmm_plot_pad,1)-10,...
        11:size(cmm_plot_pad,2)-10);
%ditch padding on image
w_surf_plot = w_surf_plot_pad(11:size(w_surf_plot_pad,1)-10,...
        11:size(w_surf_plot_pad,2)-10);



 %%
 %export segments

export_dir = '..\full_scene\';

%w-shape output
%NOTE: export the solid polygon and NOT the outline
fileName = horzcat('w-shape_full_scene_130812.tif');
geotiffwrite(horzcat(export_dir,fileName),logical(w_shape),R_veg,
...
    'CoordRefSysCode','EPSG:26911');

%w-surf output
%NOTE: export the solid polygon and NOT the outline
fileName = horzcat('w-surf_full_scene_130812.tif');
geotiffwrite(horzcat(export_dir,fileName),logical(w_surf_plot),
R_veg,...
    'CoordRefSysCode','EPSG:26911');

%cmm output
fileName = horzcat('cmm_lin_130812.tif');
geotiffwrite(horzcat(export_dir,fileName),cmm_plot,R_veg,...
    'CoordRefSysCode','EPSG:26911');


%output the veg_high_clean layer
fileName = horzcat('veg_high_clean_130812.tif');
geotiffwrite(horzcat(export_dir,fileName),veg_high_clean,R_veg,...
    'CoordRefSysCode','EPSG:26911');

%output the veg_high_bin binary layer
fileName = horzcat('veg_high_bin_130812.tif');
geotiffwrite(horzcat(export_dir,fileName),veg_high_bin,R_veg,...
    'CoordRefSysCode','EPSG:26911');
```

```
function [L_fore, L_lines] = watershedshape(bw_image,
exclusion_param)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%
%Purpose: Segment a binary image based on inverse distance
transform

%Inputs: a binary image and an exclusion parameter to mask out
areas that are too far
%from any edges to be reliably segmented.
%   For example: If there are large swaths of closed-canopy
forest, there
%   is no basis on which to conduct shape-based segmentation.
Allowing the
%   process to run in these areas frequently results in meaningless
%   segments.
%   Values:  I use "50" which means that areas >50 pixels from an
edge will
%   be masked out of this analysis. This number is based on image
%   resolution and the scale of vegetated areas in the study area.

%Outputs:
%   L_fore: foreground watershed segments
%   L_lines: the watershed boundary lines

%NOTE:  This code is for reference purposes only. You should not
assume
%that this will work for your specific imagery. Feel free to edit
as
%needed. Please cite Alonzo et al. (2014).
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%

%check input arguments
if nargin < 2
    exclusion_param = 50;
elseif nargin < 1
    error('Too few arguments for watershedshape.');
end

%shape-based watershed
D = -bwdist(~bw_image);
include = D > -exclusion_param; %don't use watershedshape on
forested stands

%find mins of the complemented inverse dist image and mark them
D_min = imextendedmin(D,1).*include;
D_imp = imimposemin(D,D_min);
```

```
%watershed transform
D_imp(D==0)=-Inf;
L = watershed(D_imp);

%cut out the background segments and clean tiny segments
L_fore = immultiply(L,bw_image);
L_fore_bw_clean = bwareaopen(L_fore,16);
L_fore = immultiply(L_fore,L_fore_bw_clean);

%generate outlines from L (for display purposes)
L_lines = L==0;


return
```

```
function cmm_wrapper = cmmWrap(chm_package)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%This code allows for large groups of segments to be sent in bulk
to
%available cores using multicore parallelization.

%Input format: an mx1 cell each containing a segment (double)

%This code calls max_filter_variable_window.m
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    cmm_wrapper = cell(length(chm_package),1);
    num = length(chm_package);

    for seg_prog = 1:length(chm_package)

        %each segment is sent for CMM processing here
        cmm_wrapper{seg_prog} = max_filter_variable_window(
        chm_package{seg_prog});

        %status updates printed
        fprintf('Seg per: %u\n',(seg_prog/num)*100)

    end

return
```

```
function cmm = max_filter_variable_window(chm,seg_prog)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%This code accepts a canopy height model of any size as input (and
%optionally a parameter to indicate computation progress).

%It is recommended that the input chm here is relatively small
(i.e. run on
%individual segments rather than a full scene). Performance,
partly because
%this needs to run on every pixel and partly because of the use of
%imdilate, is not optimal.

%Output: a canopy maxima model of the same dims as the input CHM.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


%get dimensions of incoming chm
m_dim = size(chm,1);
n_dim = size(chm,2);
bw_im = zeros(size(chm));
cmm = zeros(size(chm));

%set the model image upon which to base the variable window
%user needs to determine the way in which this will be
calculated. Below I
%show a linear example and a 2nd degree polynomial based the
relationship
%between tree height and crown width in Santa Barbara using field
data.
win = 0.56  * chm + 1.82; %linear
%win = -0.0062*(chm.^2) + 0.9171*chm + 0.159; %poly2

%convert to a "radius image". Each pixel in win_rad contains the
value of
%the search radius to be used to create the variable window.
win_rad = (round(win./2)).*4;


%Purpose of this loop:
%Iterate through image using linear indexing (outperformed
coordinate
%indexing significantly). Create a variably sized mask based on
the radius
%image at that address. Use the mask to grab the neighborhood max
value
%and assign to CMM at the address.
for j = 1:m_dim
    for k=1:n_dim
```

```
        %find linear position of seed px
        lin_idx = sub2ind(size(chm),j,k);
        bw_im(lin_idx) = 1;

        %set the window size using a structured element
        %Note: setting SE using "ones()" is much faster than
        creating a
        %true disc shape using strel but probably less accurate
        if win_rad(lin_idx) > 0
            SE = ones(win_rad(lin_idx),win_rad(lin_idx));
        else
            SE = ones(1,1);
        end

        %dilate to create a search neighborhood within which the
        maximum
        %value will be found
        bw_th = imdilate(bw_im,SE);

        %grab the max chm value that is within the SE and assign
        %it to the CMM at the address in question
        cmm(lin_idx) = max(chm(bw_th==1));

        %reset bw_im
        bw_im(lin_idx)=0;
    end

end

%print progress if input parameter provided
if nargin == 2
    fprintf('Seg: %u\n',seg_prog)
end

return
```

```
function [w_surf, w_surf_lines] = ...
    watershedsurface1(image,more_seg, shape_result)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%
%Purpose:
    %This function takes a surface model and optionally
    boundaries that
    %have already been defined by a previous segmentation (e.g.
    shape-based
    %watershed) and computes the 2.5D watershed transformation.
    It returns
    %the label matrix (w_surf),and the combined lines from this
    watershed
    %and the input segments (w_surf_lines).

%Inputs:
    %image = the surface model to be segmented
    %more_seg = flag. 0 = do not actually run segmentation
    %shape_result = lines from previous segmentation. Not really
    relevant
    %when running this code on individual segments but this code
    can be run
    %on larger images too. Then shape_result will allow for
    retention of
    %original boundaries in addition to the new boundaries.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%

    %adaptive smoothing
    image = wiener2(image,[3 3],1);

    %%set up outlines and layers
    veg_high_bw = im2bw(image,0.05);

    %set defaults for optional arguments
    if nargin < 3
        %if no previous segments supplied the boundary of the
        binary image
        %will be considered the shape segment
        shape_result = ~bwperim(veg_high_bw,8);
    elseif nargin < 1
        error('Too few arguments for watershedsurface1.');
    end

    %if more_seg flag is 0 then this segmentation will not occur
    if more_seg == 0;
        w_surf = bwlabel(veg_high_bw);
```

```
        w_surf_lines = ~shape_result;
        return
    end


    %place boundaries from shape-based watershed on image
    L_0 = shape_result == 0;
    image(L_0) = 0;


    %find mins
    veg_comp = imcomplement(image);
    veg_min = imregionalmin(veg_comp);

    %clean up the mins
    veg_min_F = imfill(veg_min,'holes');
    veg_min_O = bwareaopen(veg_min_F,2);
    veg_min_S = bwmorph(veg_min_O,'shrink',Inf);


    %impose mins
    veg_imp = imimposemin(veg_comp,veg_min_S);
    veg_imp(~shape_result) = Inf;
    %set the background stuff to neg inf
    veg_imp(veg_imp>0 & veg_imp < 10000)=-Inf;
    %only find mins where -Inf has been imposed
    veg_imp = imhmin(veg_imp,100);

    %compute watershed
    veg_L = watershed(veg_imp,8);
%     figure, imagesc(veg_L);

    %cut out the background segments and clean tiny segments
    w_surf = immultiply(veg_L,veg_high_bw);
    L_fore_bw_clean = bwareaopen(w_surf,16);
    w_surf = immultiply(w_surf,L_fore_bw_clean);


    %segment boundaries
    w_surf_lines = veg_L == 0;



return
```

```
function [surf_int_mean, surf_ht_mean, surf_E, surf_N, surf_H,
surf_I] = surfacecalcs(tree_E,tree_N,tree_H, ...
    tree_I,max_crown_ht, cbh, grid_sz)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%
    %Purpose:
    %This function takes descriptions of extents of point-cloud
    subset
    %(generally a tree crown or crown segment) and returns the
    height and
    %intensity of the highest points along the crown surface in a
    user-defined grid size.
    %This is useful for understanding the crown closure of a
    canopy.


    % Required arguments:
    % tree_E = (m,1) vector of eastings
    % tree_N = (m,1) vector of northings (of same points)
    % tree_H = (m,1) vector of heights (of same points)
    % tree_I = (m,1) vector of intensities (of same points)
    % max_crown_ht = max height of crown in question
    % cbh = crown base height of crown in question
    %
    % Optional args:
    % grid_sz = size of grid over which to extract ht/int info.
    Default
    %           size is 0.5 m.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%

%set defaults for optional arguments
    if nargin < 7
        grid_sz = 0.5;
    elseif nargin < 6
        error('Too few arguments for surfacecalcs. Shameful.');
    end

%establish grid
    bins_EW = min(tree_E):grid_sz:max(tree_E);
    bins_NS = min(tree_N):grid_sz:max(tree_N);


%iterate through bins
  for j=1:length(bins_EW)-1  %E-W slices
      easting_range = bins_EW(j:j+1);
```

```
    idx_EW = find(tree_E >= easting_range(1) & tree_E <
    easting_range(2));

    for k=1:length(bins_NS)-1  %N-S bins in each E-W slice

        northing_range = bins_NS(k:k+1);
        idx_NS_EW = find(tree_N(idx_EW) >= northing_range(1) &
        tree_N(idx_EW)...
            < northing_range(2));

        %finds the highest point in the northing by easting bin
        that is
        %inside the crown max and min
        idx_z = find(tree_H(idx_EW(idx_NS_EW))<= max_crown_ht &
        tree_H(idx_EW(idx_NS_EW))>= ...
            cbh & tree_H(idx_EW(idx_NS_EW))== max(tree_H(idx_EW(
            idx_NS_EW))));

       %if the idx_z is not empty then grab the intensity
       value(s) for
       %the highest point(s)
        if ~isempty(idx_z)
          surf_holder(k,j) = mean(tree_I(idx_EW(idx_NS_EW(idx_z
          ))));  %needs "mean" b/c there can be several same max
          heights
          surf_holder_max(k,j) = max(tree_I(idx_EW(idx_NS_EW(
          idx_z))));

          surf_H_holder(k,j) = mean(tree_H(idx_EW(idx_NS_EW(
          idx_z))));
          %COULD ADD IN SIMILAR VARIABLES TO HOLD HEIGHT
          VARIATION

          %store northings and eastings of highest point
          surf_E_holder(k,j) = mean(tree_E(idx_EW(idx_NS_EW(
          idx_z))));
          surf_N_holder(k,j) = mean(tree_N(idx_EW(idx_NS_EW(
          idx_z))));

        else
           surf_holder(k,j) = 0;

           surf_H_holder(k,j) = 0;
           surf_E_holder(k,j) = 0;
           surf_N_holder(k,j) = 0;
        end
        %b=b+1;
    end
end
```

```
        surf_idx = find(surf_holder>0);

         %surf_int_pattern{i} = surf_holder; %saving for texture
         analysis
        surf_int_mean = mean2(surf_holder(surf_idx));
        surf_ht_mean = mean2(surf_H_holder(surf_idx));

        %report out location of high points
        dimE = size(surf_E_holder,1);
        dimN = size(surf_E_holder,2);
        surf_E = reshape(surf_E_holder,(dimE*dimN),1);
        surf_N = reshape(surf_N_holder,(dimE*dimN),1);
        surf_H = reshape(surf_H_holder,(dimE*dimN),1);
        surf_I = reshape(surf_holder,(dimE*dimN),1);

end
```

```
function [N,S,E,W, diam, count] = widthatheight(tree_E,tree_N,
tree_H, ht_in_crown, slice_sz, ...
    bin_size)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%
    %Purpose:
    %This function takes a the northings, eastings, and height
    triplets for
    %a crown or crown segment and computes the width of the crown
    at
    %user-defined heights.
    %This is helpful in quantifying tree form.

    % Required arguments:
    % tree_E = (m,1) vector of eastings
    % tree_N = (m,1) vector of northings (of same points)
    % tree_H = (m,1) vector of heights (of same points)
    % ht_in_crown = height at which you want to find width
    %
    % Optional args:
    % slice_sz = slice size (half the total up-down search
    distance in which to
    %     locate points (above and below ht_in_crown), default =
    0.5 m
    % bin_size = bin size over which to aggregate points, default
    = 0.25 m
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%


    %set defaults for optional arguments
    if nargin <=5
        bin_size = 0.25;
    elseif nargin <= 4
        slice_sz = 0.5;
    elseif nargin <=3
        error('Too few arguments for widthatheight.');
    end


 %  index points within slice_sz of the ht_in_crown
    idx_height = find(tree_H > ht_in_crown-slice_sz & tree_H <
    ht_in_crown+slice_sz);
    count = length(idx_height);

    if ~isempty(idx_height)
        %subsets of points near median height
        tree_E_sub = tree_E(idx_height);
```

```
tree_N_sub = tree_N(idx_height);

%binning EW
bins_EW = min(tree_E_sub):bin_size:max(tree_E_sub);

if length(bins_EW) >= 3
    [avg,num,std]=bindata(tree_E_sub,tree_N_sub,bins_EW);
    nan_idx = find(isnan(num));
    EW_num = num; EW_num(nan_idx) = 0;

    %find the east and west extents of crown at ht_in_crown
    furthest_east_idx = find(cumsum(EW_num)>(0.99*sum(
    EW_num)),1,'first');
    E = (bin_size * furthest_east_idx) + min(tree_E_sub);
    furthest_west_idx = find(cumsum(EW_num)>(0.01*sum(
    EW_num)),1,'first');
    W = (bin_size * furthest_west_idx) + min(tree_E_sub);
else
    warning('Cannot create viable EW line from number of
    points');
    E = min(bins_EW);
    W = max(bins_EW);
end


%binning NS
bins_NS = min(tree_N_sub):bin_size:max(tree_N_sub);

if length(bins_NS) >=3
    [avg,num,std]=bindata(tree_N_sub,tree_N_sub,bins_NS);
    nan_idx = find(isnan(num));
    NS_num = num; NS_num(nan_idx) = 0; %#ok<FNDSB>

    %NS_avg = avg; NS_avg(nan_idx) = 0;
    %NS_std = std; NS_std(nan_idx) = 0;

    %find the north and south extents of crown at
    ht_in_crown
    furthest_north_idx = find(cumsum(NS_num)>(0.99*sum(
    NS_num)),1,'first');
    N = (bin_size * furthest_north_idx) + min(tree_N_sub);
    furthest_south_idx = find(cumsum(NS_num)>(0.01*sum(
    NS_num)),1,'first');
    S = (bin_size * furthest_south_idx) + min(tree_N_sub);
else
    warning('Cannot create viable NS line from number of
    points');
    N = min(bins_NS);
    S = max(bins_NS);
```

```
    end
else
    %if there are no points in this slice assign a zero length
    %line to the midpoint of the segment
     warning('Cannot create viable NS line from number of
     points');
        N = (min(tree_N)+max(tree_N))/2;
        S = N;
        E = (min(tree_E)+max(tree_E))/2;
        W = E;
end

%compute average diam at the given height
diam = ((N-S)+(E-W))/2;
```

Appendix B

Matlab code for the publication: "Mapping urban forest leaf area index with airborne lidar using penetration metrics and allometry"

Authors: Michael Alonzo, Bodo Bookhagen, Joseph P. McFadden, Alex Sun and Dar A. Roberts

Explanation of the purpose and functionality of the code can be found within each file printed on the following pages.

```
%%
%PURPOSE:
%This script should be use to evaluate the effects of changes in
lidar
%pulse angle on estimates of single-tree effective LAI. The user
may
%control a number of parameters in the following cell. LAIe
estimates are
%produced using a ratio of pulses that pass through canopy to
total number
%of pulses intersecting canopy. A correction factor is introduced
that is
%based on the ratio of the expected path length of pulses through
canopy at a give pulse
%angle to the expected path length of hypothetical nadir-view
pulses.

%Secondarily, this simulator outputs a "clumping ratio" which
sheds light
%on how much plot-level gap fraction estimates will be impacted
by clumping
%of foliage into spatially disaggregated trees. The clumping
ratio will be
%close to 1 when the leaf area for a tree on a plot is
distributed homogeneously
%across the entire plot.

%Please contact Mike Alonzo (mike.alonzo@geog.ucsb.edu) with
questions
%about this code or citation information.

%must have access to functions in:
%geom2d, geom3d, and kdtree2 (these are all available at:
%http://geog.ucsb.edu/~malonzo/matlab_code.html)


%%
%USER INPUTS

%Enter parameters for tree and study plot
numLeaves = 10000; %number of leaves
leafrad = 0.03; %leaf radius in meters (circular leaves)
tree_ht = 12; %tree height
tree_r = 3; %crown radius
cbh = 2; %crown base height
plot_rad = 15; %plot radius
ext_coef=0.5; %extinction coefficient (default at 0.5 to coincide
with spherical LAD)
```

```
%sensor parameters
%scan angle set must ALWAYS include 0.
scan_angles = [0 5 10 15 20 25 30 35]; %in degrees from nadir.
pointSpace = 0.25; %beam spacing on ground (m)
sensorAGL = 1000; %sensor altitude (m above ground)
%Azimuth. There is no reason to favor one azimuth over another
but with any given
%randomization of leaves, running the model at different azimuths
can lead
%to significantly different estimates due to clumping.
az = 180;

%%
%Parameterize tree and ground from user inputs

%Define height of cylindrical component of tree
if (tree_ht-tree_r-cbh)>0
    cyl_ht = tree_ht-tree_r-cbh;
else
    cyl_ht = 0;
end

%construct two random distributions for leaves to give trees a
%reasonable shape: 1. A hemispherical hat, 2. A cylindrical base
%Partition the respective areas of the hemisphere and cylinder so
that
%leaves can be distributed appropriately
hemi_vol = ((4/3)*pi*tree_r^3)*0.5;
cyl_vol = (pi*tree_r^2)*cyl_ht;
hemi_rat = hemi_vol/(hemi_vol+cyl_vol);
cyl_rat = 1-hemi_rat;
hemi_leaves = round(numLeaves*hemi_rat);
cyl_leaves = numLeaves - hemi_leaves;

%construct the rounded crown to sit on top of the cylinder
R_hemi = tree_r*sqrt(rand(hemi_leaves,1));
Az_hemi = 2*pi*rand(hemi_leaves,1);
El_hemi = (pi/2)*rand(hemi_leaves,1);
[Xsph, Ysph, Zsph] = sph2cart(Az_hemi,El_hemi,R_hemi);
Zsph = Zsph + (cyl_ht + cbh);

%construct the cylinder
R_cyl = tree_r*sqrt(rand(cyl_leaves,1));
Az_cyl = 2*pi*rand(cyl_leaves,1);
[Xcyl, Ycyl] = pol2cart(Az_cyl,R_cyl);
Zcyl =(rand(cyl_leaves,1)*cyl_ht)+cbh;

%full tree construction (cap the cylinder with the hemisphere)
X=[];Y=[]; Z=[];
```

```
X = vertcat(Xsph,Xcyl);
Y = vertcat(Ysph,Ycyl);
Z = vertcat(Zsph,Zcyl);

%show tree
figure, hold on, grid on
plot3(X,Y,Z,'.r')
axis equal
view(32,32)

%spherical leaf angle distribution with mean at around 57 deg
leaf_angle1 = 90-rad2deg(asin(rand(numLeaves,1)));

%azimuthal distribution between 0 and 360 (uniform)
leaf_az = rad2deg(rand(numLeaves,1)*(2*pi()));

%calculate tree and plot LAI based on input params
tree_cpa = sum(pi()*(tree_r^2)) %tree crown projection area
plot_cpa = pi()*(plot_rad^2) %plot area
LA = ((leafrad^2)*pi())*numLeaves; %total tree leaf area
LAI = LA/plot_cpa %LAI distributed across entire plot
tree_LAI = LA/tree_cpa %LAI just of tree

%Prepare beam locations on ground based on point spacing
X_g_vec = -1*plot_rad:pointSpace:plot_rad*1;
Y_g_vec = X_g_vec;
[X_g, Y_g] = meshgrid(X_g_vec, Y_g_vec); %ground returns (or:
possible ground returns)
Z_g = zeros(size(X_g,1),size(X_g,2));

%%
%Generate leaves (aka circles cut out of planes)
%generate three points for each leaf to define plane
tic
leaf_points = []; leaf_planes = []; leaf_circle = [];
for i = 1:numLeaves
    leaf_circle{i} = [X(i) Y(i) Z(i) leafrad  leaf_angle1(i)
    leaf_az(i) leaf_az(i)];

    %get three points on circle to define plane (set apart by 115
    %degrees -- number doesn't really matter)
    for k = 1:3
        circ_pts(k,:) = circle3dPoint(leaf_circle{i}, k*115);
    end

    %three points for each leaf and create plane
    leaf_points{i} = circ_pts;
    leaf_planes{i} = createPlane(leaf_points{i});
```

```
end
toc_leaf_circs = toc


%%
%this loop runs through the LAI calculations for the tree by scan
angle. A
%scan angle of 0 must always be included for calculation of nadir
expected
%path length.


for theta = 1:length(scan_angles)
%%
    %construct a new kd tree for spatial indexing of leaves
    clear tree
    tree = kdtree_build([X Y Z]);

    %sensor parameters
    scanAngle = scan_angles(theta)
    %r is the radial displacement of the sensor with respect to
    the center of
    %the tree (0,0)
    r = tan(deg2rad(scanAngle))*sensorAGL;
    [x_sens, y_sens] = pol2cart(deg2rad(az),r); %x,y sensor
    position
    sensCoord = [x_sens y_sens sensorAGL];

    %generate convex hull for the set of leaf points
    dt_leaves = DelaunayTri([X Y Z]);
    leaves_ch = convexHull(dt_leaves);

    %check for intersection with crown by each pulse.
    %There is likely significant room for performance improvement
    here.
    counter = 1; epl = []; pulse_line = []; plR1 = [];
    parameterCell = [];
    int_holder = [];
    for i = 1:numel(X_g) %number of pulses hitting ground

        %create line from each point on ground to sensor
        check_line = createLine3d([X_g(i) Y_g(i) Z_g(i)],sensCoord
        );
        %check if check_line intersects the convex hull of leaves
        [intersect_true, ~, ~ ] = intersectLineMesh3d(check_line,[
        X Y Z], leaves_ch);

        %only proceed if check_line intersects the convex hull
        if not(isempty(intersect_true))
```

```
%top intersection is always 2nd point after sorting
int_points{counter} = sortrows(intersect_true,3);

%discretize the beam into some number of points
(default = 100)
%this is necessary for a kd ball query
dx = linspace(int_points{counter}(1,1),int_points{
counter}(2,1),100);
dy = linspace(int_points{counter}(1,2),int_points{
counter}(2,2),100);
dz = linspace(int_points{counter}(1,3),int_points{
counter}(2,3),100);
sample_line{i} = [dx' dy' dz'];

%prepare for kd tree ball query
max_path = sqrt(tree_ht^2 + (2*tree_r)^2);
ball_space = max_path / 100;
%ball_rad needs to be large enough so the cylinder
sampled
%from ANY point along the pulse never has a radius
less than
%leafrad.
ball_rad = max(2*ball_space,2*leafrad);
include_leaves_idx = [];

%run kd ball query for each point along beam (dx,dy,dz)
%index leaves that are captured by ball of given radius
for kd = 1:length(dx)
    qpoint = sample_line{i}(kd,:);
    [idxs, dists] = kdtree_ball_query( tree, qpoint,
    ball_rad);
    include_leaves_idx = vertcat(include_leaves_idx,
    idxs);
end

%needs to be unique leaves!
leaves_in_question = leaf_planes(unique(
include_leaves_idx));
numLeaves_in_bins = length(leaves_in_question);

%check for and return intersections with each leaf
pulse_line{counter} = check_line;
[int_holder{counter}] = leafContacts(pulse_line{
counter},...
    leaves_in_question, numLeaves_in_bins, X(unique(
    include_leaves_idx)),...
    Y(unique(include_leaves_idx)), Z(unique(
    include_leaves_idx)), leafrad);
```

```
        counter = counter +1;
    end
end

%number of beams that intersected convex hull
beams(theta) = counter-1;

%setting up loop to calculate expected path length and actual
path
%length of beam through crown (convex hull)
contacts_per_beam=[]; pulse_death_pt =[];
for counter = 1:beams(theta) %num beams intersecting convex
hull

    %find the highest leaf intersection and log that as the
    initial leaf
    %contact
    max_ht_leaf = max(int_holder{counter}(:,3));
    if max_ht_leaf>0
        high_leaf_idx = find(int_holder{counter}(:,3)==
        max_ht_leaf);
        pulse_death_pt{counter} = int_holder{counter}(
        high_leaf_idx,:);
    else
        pulse_death_pt{counter} = [];
    end

    %compute path length and expected path length
    if size(int_points{counter},1)>1
        %compute the expected path length (epl = top
        intersection to bottom intersection)
        epl(counter) = distancePoints3d(int_points{counter}(
        end,:),int_points{counter}(1,:));

        %if we know that the pulse "died" in the crown
        calculate actual
        %path length. This is more useful for future crown
        density
        %simulation.
        if not(isempty(pulse_death_pt{counter}))
            plR1(counter) = distancePoints3d(int_points{
            counter}(end,:), pulse_death_pt{counter});
        else
            plR1(counter) = epl(counter);
        end
    else
        %fill in zeros if there weren't two points of
        intersection with the solid
```

```
        epl(counter) = 0; %expected path length through crown
        plR1(counter) = 0; %actual distance traveled prior to
        leaf intersection
    end

end

%store epl and plR1 values for each theta
epl_stor{theta} = epl;
plR1_stor{theta} = plR1;

%calc nadir_epl from epl at scan angle of 0
if scan_angles(theta) == 0
    nadir_epl = median(epl);
    nadir_epl_count = length(epl);
end


%generate secant correction (no correction for nadir view)
%this correction is for canopy only (doesn't account for number
%of pulses intersecting canopy)
sec_cor(theta) = median(epl)/nadir_epl

%this correction is for direct plot method (accounts for
number of intersections)
num_pulses_in_plot = plot_cpa*(1/pointSpace)^2;
epl_dir(theta) = mean(vertcat(epl',zeros(int64(
num_pulses_in_plot-length(epl)),1)))
if scan_angles(theta) == 0
    nad_epl_dir = epl_dir(theta);
end
sec_cor_dir(theta) = epl_dir(theta)/nad_epl_dir;

%count of beams with an expected path length
epl_count(theta) = length(epl);
%median epl
epl_med(theta) = median(epl);
%number of times there is a leaf contact / no contact in the
canopy
contacts(theta) = length(find(epl>plR1));
no_contacts(theta) = length(find(epl<=plR1));
%gap fraction based on number of no contacts
gap_frac(theta) = (no_contacts(theta)/length(epl))

%compute gap frac for plot LAI (direct method)
gap_frac_dir(theta) = (num_pulses_in_plot-contacts(theta))/...
    num_pulses_in_plot;

%estimate of canopy LAI uncorrected
```

```
    LAI_est(theta) = -log(gap_frac(theta))/ext_coef;

    %apply secant correction
    LAI_est_cor_lai(theta) = -log(gap_frac(theta))/(sec_cor(theta
    )*ext_coef);

    %outputs by scan angle
    %corrected plot LAI by fCov method (canopy LAI modified by
    plot fractional cover)
    plot_LAI_est_using_fCov_corr(theta) = LAI_est_cor_lai(theta) *
     (tree_cpa/plot_cpa)
    %uncorrected version of fcov method estimate
    plot_LAI_est_using_fCov_Uncorr(theta) = LAI_est(theta) * (
    tree_cpa/plot_cpa)
    %corrected plot LAI using direct method (grap frac from all
    plot
    %pulses)
    plot_LAI_direct_cor(theta) = -log(gap_frac_dir(theta))/(
    ext_coef*sec_cor_dir(theta))
    %uncorrected direct plot method estimate
    plot_LAI_direct_Uncor(theta) = -log(gap_frac_dir(theta))/(
    ext_coef)
    %how evenly are leaves spread across plot?
    point_spread_ratio(theta) = tree_cpa/plot_cpa
    %true plot LAI based on user-entered params
    true_plot_LAI(theta) = LAI
    %correction for clumping fCov method output divided by direct
    method
    %output
    clmp_mult(theta) = plot_LAI_est_using_fCov_corr(theta) /
    plot_LAI_direct_cor(theta)


end


%%
%graphical display of the most recent simulation

figure, hold on, grid on, box on
    for i = 1:200:length(X)
        drawCircle3d([X(i) Y(i) Z(i) leafrad  leaf_angle1(i)
        leaf_az(i)], 'LineWidth', 2, 'Color', 'g');
    end
    ground_plot_idx = find(rand(numel(X_g),1)>0.85);
    plot3(X_g(ground_plot_idx), Y_g(ground_plot_idx), Z_g(
    ground_plot_idx),'.r')
    trisurf(leaves_ch,X, Y, Z, 'FaceColor', 'yellow')
    alpha(0.2)
```

```
rand_lines_idx = find(rand(length(pulse_line),1)>0.98);
for j = 1:length(rand_lines_idx)
    drawLine3d(pulse_line{rand_lines_idx(j)},'LineWidth',1.5)
end

%add plot circle
numPoints=100; %Number of points making up the circle
%Define circle in polar coordinates (angle and radius)
theta=linspace(0,2*pi,numPoints); %100 evenly spaced points
between 0 and 2pi
rho=ones(1,numPoints)*plot_rad; %Radius should be 1 for all
100 points
%Convert polar coordinates to Cartesian for plotting
[Xc,Yc] = pol2cart(theta,rho);
%Plot a red circle
plot(Xc,Yc,'g-','linewidth',3);

axis equal
```

```
function [int_holder] = leafContacts(pulse_line, leaf_planes,
numLeaves, X, Y, Z, leafrad)

%PURPOSE: Check for intersections between a beam (3d line) and a
leaf plane
%in 3d. Store number of intersections.

%REQUIRED INPUTS: pulse_line (equation of 3d line representing
beam in
%question), leaf_planes (the leaf planes selected by ball query),
numLeaves
%(number of selected leaves), XYZ (position of each selected
leaf), leafrad
%(leaf radius)

if nargin < 7
    error('You need all 7 variables to run leafContacts')
end

%holder for leaf intersections
int_holder=zeros(numLeaves,3);

%loop through each leaf testing for beam intersection
 for j = 1:numLeaves

    %first, where does the beam intersect the infinite leaf plane
    line_plane_int = intersectLinePlane(pulse_line,leaf_planes{j});

    %check if point on the plane is in the circle by checking
    %whether distance between the points is less that radius of
    %leaf
    if distancePoints3d([X(j) Y(j) Z(j)], line_plane_int)<leafrad
        int_holder(j,:) = line_plane_int;
    end


 end
```

References:

Akbari, H. (2002). Shade trees reduce building energy use and CO2 emissions from power plants. *Environmental Pollution*, *116*, 119–126.

Alberti, M. (1999). Urban patterns and environmental performance: what do we know? *Journal of Planning Education and Research*.

Alberti, M. (2005). The Effects of Urban Patterns on Ecosystem Function. *International Regional Science Review*, *28*(2), 168–192.

Alonzo, M., Bookhagen, B., Mcfadden, J. P., Sun, A., & Roberts, D. A. (2015). Mapping urban forest leaf area index with airborne lidar using penetration metrics and allometry. *Remote Sensing of Environment*, *162*, 141–153.

Alonzo, M., Bookhagen, B., & Roberts, D. A. (2014). Urban tree species mapping using hyperspectral and lidar data fusion. *Remote Sensing of Environment*, *148*, 70–83.

Alonzo, M., Roth, K. L., & Roberts, D. A. (2013). Identifying Santa Barbara's urban tree species from AVIRIS imagery using canonical discriminant analysis. *Remote Sensing Letters*, *4*(5), 513–521.

Andersen, H.-E., Reutebuch, S. E., & McGaughey, R. J. (2006). A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods. *Canadian Journal of Remote Sensing*, *32*(5), 355–366.

Anderson, J., Plourde, L., Martin, M., Braswell, B., Smith, M., Dubayah, R., Hofton, M., Blair, J. (2008). Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sensing of Environment*, 112(4), 1856–1870.

Asner, G. P. (1998). Biophysical and Biochemical Sources of Variability in Canopy Reflectance. *Remote Sensing of Environment*, *64*(3), 234–253.

Asner, G. P., Knapp, D. E., Kennedy-Bowdoin, T., Jones, M. O., Martin, R. E., Boardman, J., & Field, C. B. (2007). Carnegie Airborne Observatory: in-flight fusion of hyperspectral imaging and waveform light detection and ranging for three-dimensional studies of ecosystems. *Journal of Applied Remote Sensing*, *1*(1), 013536.

Asner, G. P., & Martin, R. E. (2008). Spectral and chemical analysis of tropical forests: Scaling from leaf to canopy levels. *Remote Sensing of Environment*, *112*(10), 3958–3970.

Baldocchi, D. D., Hicks, B., & Camara, P. (1987). A canopy stomatal resistance model for gaseous deposition to vegetated surfaces. *Atmospheric Environment (1967)*, *21*(1), 91–101.

Beelen, R., Hoek, G., van den Brandt, P. a, Goldbohm, R. A., Fischer, P., Schouten, L. J., … Brunekreef, B. (2008). Long-term effects of traffic-related air pollution on mortality in a Dutch cohort (NLCS-AIR study). *Environmental Health Perspectives*, *116*(2), 196–202.

Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., & Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3-4), 239–258.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), 2–16.

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

Boschetti, M., Boschetti, L., Oliveri, S., Casati, L., & Canova, I. (2007). Tree species mapping with airborne hyperspectral MIVIS data: the Ticino Park study case. *International Journal of Remote Sensing*, 28, 1251–1261.

Brandtberg, T. (2007). Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(5), 325–340.

Bréda, N. J. J. (2003). Ground-based measurements of leaf area index: a review of methods, instruments and current controversies. *Journal of Experimental Botany*, *54*(392), 2403–17.

Carlson, T., & Ripley, D. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment*, *62*(3), 241–252.

Casey Trees (2010). i-Tree Ecosystem Analysis: Washington, DC. URL: http://caseytrees.org/wp-content/uploads/2012/02/report-2010-01-ufore2009.pdf

Castro-Esau, K., Sanchez-Azofeifa, G., Rivard, B., Wright, S. J., & Quesada, M. (2006). Variability in leaf optical properties of Mesoamerican trees and the potential for species classification. *American Journal of Botany*, 93(4), 517–530.

Chapman, L. (2007). Potential applications of near infra-red hemispherical imagery in forest environments. *Agricultural and Forest Meteorology*, *143*(1-2), 151–156.

Chen, J. M., & Black, T. (1991). Measuring leaf area index of plant canopies with branch architecture. *Agricultural and Forest Meteorology*, *57*(1953), 1–12.

Chen, J. M., & Black, T. (1992). Defining leaf area index for non-flat leaves. *Plant, Cell & Environment*, *15*, 421–429.

Chen, J. M., Rich, P. M., Gower, S. T., Norman, J. M., & Plummer, S. (1997). Leaf area index of boreal forests: Theory, techniques, and measurements. *Journal of Geophysical Research*, *102*(D24), 29429.

Chen, Q., Baldocchi, D., Gong, P., & Kelly, M. (2006). Isolating Individual Trees in a Savanna Woodland Using Small Footprint Lidar Data. *Photogrammetric Engineering & Remote Sensing*, *72*(8), 923–932.

Clark, M. L., Roberts, D. A., & Clark, D. B. (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, *96*(3-4), 375–398.

Clark, M. L., Roberts, D. A., Ewel, J. J., & Clark, D. B. (2011). Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. *Remote Sensing of Environment*, 115(11), 2931–2942

Clarke, L.W., Jenerette, D.G., & Davila, A. (2013). The luxury of vegetation and the legacy of tree biodiversity in Los Angeles, CA. *Landscape and Urban Planning*, 116, 48—59.

Cochrane, M. A. (2000). Using vegetation reflectance variability for species level classification of hyperspectral data. *International Journal of Remote Sensing*, *21*(10), 2075–2087.

Congalton, R. G. (1991). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment*, *37*, 35–46.

Cook, B. D., Corp, L. a., Nelson, R. F., Middleton, E. M., Morton, D. C., McCorkel, J. T., … Montesano, P. M. (2013). NASA goddard's LiDAR, hyperspectral and thermal (G-LiHT) airborne imager. *Remote Sensing*, *5*, 4045–4066.

Dalponte, M., Bruzzone, L., & Gianelle, D. (2008). Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5), 1416–1427.

Dalponte, M., Bruzzone, L., & Gianelle, D. (2012). Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sensing of Environment*, 123, 258–270.

Dalponte, M., Ørka, H. O., Ene, L. T., Gobakken, T., & Næsset, E. (2014). Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. *Remote Sensing of Environment*, 140, 306–317.

Deng, C., & Wu, C. (2013). Examining the impacts of urban biophysical compositions on surface urban heat island: A spectral unmixing and thermal mixing approach. *Remote Sensing of Environment*, *131*, 262–274.

Dennison, P. E., & Roberts, D. A. (2003). The effects of vegetation phenology on endmember selection and species mapping in southern California chaparral. *Remote Sensing of Environment*, 87(2), 295-309.

Digabel, H., & Lantuéjoul, C. (1978). Iterative algorithms. In *Proc. 2nd European Symp. Quantitative Analysis of Microstructures in Material Science, Biology and Medicine* (Vol. 19, No. 7, p. 8). Stuttgart, West Germany: Riederer Verlag.

Drake, J. B., Dubayah, R. O., Clark, D. B., Knox, R. G., Blair, J. B., Hofton, M. a., … Prince, S. (2002). Estimation of tropical forest structural characteristics, using large-footprint lidar. *Remote Sensing of Environment*, *79*(2-3), 305–319.

Duda, R. O., and Hart, P.E. (1973). *Pattern classification and scene analysis*. New York: John Wiley and Sons, Inc.

Dwyer, J. F., Nowak, D. J., Noble, M. H., & Sisinni, S. M. (2000). *Connecting People With Ecosystems in the 21st Century: An Assessment of Our Nation's Urban Forests*. *General Technical Report PNW-GTR-490*.

Edson, C., & Wing, M. G. (2011). Airborne Light Detection and Ranging (LiDAR) for Individual Tree Stem Location, Height, and Biomass Measurements. *Remote Sensing*, *3*(12), 2494–2528.

Escobedo, F. J., & Nowak, D. J. (2009). Spatial heterogeneity and air pollution removal by an urban forest. *Landscape and Urban Planning*, *90*(3-4), 102–110.

Falster, D. S., & Westoby, M. (2003). Leaf size and angle vary widely across species: what consequences for light interception? *New Phytologist*, *158*(3), 509–525.

Ferraz, A., Bretar, F., Jacquemoud, S., Gonçalves, G., Pereira, L., Tomé, M., & Soares, P. (2012). 3-D mapping of a multi-layered Mediterranean forest using ALS data. *Remote Sensing of Environment*, *121*, 210–223.

Franke, J., Roberts, D. a., Halligan, K., & Menz, G. (2009). Hierarchical Multiple Endmember Spectral Mixture Analysis (MESMA) of hyperspectral imagery for urban environments. *Remote Sensing of Environment*, 113(8), 1712–1723.

Fraser, E. D. G., & Kenney, a. (2000). Cultural factors and landscape history affecting perceptions of the urban forest. *The Journal of Arboriculture*, *26*(2), 107–113.

Gates, D. M., Keegan, H. J., Schleter, J. C., & Weidner, V. R. (1965). Spectral Properties of Plants. *Applied Optics*, *4*(1), 11.

Georgi, J. N., & Dimitriou, D. (2010). The contribution of urban green spaces to the improvement of environment in cities: Case study of Chania, Greece. *Building and Environment*, *45*(6), 1401–1414.
Georgi, N. J., & Zafiriadis, K. (2006). The impact of park trees on microclimate in urban areas. *Urban Ecosystems*, *9*(May), 195–209.
Gillies, R. R., & Temesgen, B. (2004). Biophysical variables at the land surface. In D. A. Quattrochi & J. C. Luvall (Eds.), *Thermal Remote Sensing in Land Surface Processes*. Boca Raton, FL: CRC Press.
Goodfellow, J.W. (1995). Engineering and Construction Alternatives to Line Clearance Tree Work. *Journal of Arboriculture*. 21(1):41-49.
Gower, S., Kucharik, C., & Norman, J. (1999). Direct and indirect estimation of leaf area index, fapar, and net primary production of terrestrial ecosystems. *Remote Sensing of Environment*, *4257*(99).
Green, R. O., Eastwood, M. L., Sarture, C., Chrien, T. G., Aronsson, M., Chippendale, B., … Williams, O. (1998). Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sensing of Environment*, *65*(3), 227–248.
Grimm, N. B., Faeth, S. H., Golubiewski, N. E., Redman, C. L., Wu, J., Bai, X., & Briggs, J. M. (2008). Global change and the ecology of cities. *Science (New York, N.Y.)*, *319*(5864), 756–60.
Grimmond, C., Blackett, M., Best, M., Barlow, J., Baik, J., Belcher, S., … Zhang, N. (2010). The international urban energy balance models comparison project: first results from phase 1. *Journal of Applied Meteorology and Climatology*, *49*, 1268–1291.
Gromke, C., & Ruck, B. (2009). On the impact of trees on dispersion processes of traffic emissions in street canyons. *Boundary-Layer Meteorology*, *131*(1), 19–34.
Gurrola, L. D., Keller, E. a, Chen, J. H., Owen, L. a, & Spencer, J. Q. (2014). Tectonic geomorphology of marine terraces; Santa Barbara fold belt, California. *Geological Society of America Bulletin*, *126*(1/2), 219–233.
Hamada, S., & Ohta, T. (2010). Seasonal variations in the cooling effect of urban green areas on surrounding urban areas. *Urban Forestry & Urban Greening*, *9*(1), 15–24.
Hamlin, L., Green, R. O., Mouroulis, P., Eastwood, M., Wilson, D., Dudik, M., & Paine, C. (2011). Imaging spectrometer science measurements for terrestrial ecology: AVIRIS and new developments. *Aerospace Conference, 2011 IEEE*.
Hardin, P. J., & Jensen, R. R. (2007). The effect of urban leaf area on summertime urban surface kinetic temperatures: A Terre Haute case study. *Urban Forestry & Urban Greening*, *6*(2), 63–72.
Heinzel, J., & Koch, B. (2011). Exploring full-waveform LiDAR parameters for tree species classification. *International Journal of Applied Earth Observation and Geoinformation*, 13(1), 152–160.
Herold, M., Roberts, D. A., Gardner, M. E., & Dennison, P. E. (2004). Spectrometry for urban area remote sensing -- Development and analysis of a spectral library from 350 to 2400 nm. *Remote Sensing of Environment*, *91*, 304–319.
Heynen, N., Perkins, H., & Roy, P. (2006). The Political Ecology of Uneven Urban Green Space: The Impact of Political Economy on Race and Ethnicity in Producing Environmental Inequality in Milwaukee. *Urban Affairs Review*, *42*(1), 3–25.

Hirabayashi, S., Kroll, C. N., & Nowak, D. J. (2011). Component-based development and sensitivity analyses of an air pollutant dry deposition model. *Environmental Modelling & Software*, *26*(6), 804–816.
Hirabayashi, S., Kroll, C. N., & Nowak, D. J. (2012). Development of a distributed air pollutant dry deposition modeling framework. *Environmental Pollution*, *171*, 9–17.
Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., & Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, *42*(33), 7561–7578.
Hoffbeck, J., & Landgrebe, D. (1996). Classification of remote sensing images having high spectral resolution. *Remote Sensing of Environment*, 57, 119–126.
Holmgren, J., Nilsson, M., & Olsson, H. (2003). Simulating the effects of lidar scanning angle for estimation of mean tree height and canopy closure. *Canadian Journal of Remote Sensing*, *29*(5), 623–632.
Holmgren, J., & Persson, A. (2004). Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment*, *90*(4), 415–423.
Holmgren, J., Persson, Å., & Söderman, U. (2008). Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. *International Journal of Remote Sensing*, 29(5), 1537–1552.
Hopkinson, C., & Chasmer, L. (2009). Testing LiDAR models of fractional cover across multiple forest ecozones. *Remote Sensing of Environment*, *113*(1), 275–288.
Hopkinson, C., Lovell, J., Chasmer, L., Jupp, D., Kljun, N., Gorsel, E. Van, & van Gorsel, E. (2013). Integrating terrestrial and airborne lidar to calibrate a 3D canopy model of effective leaf area index. *Remote Sensing of Environment*, *136*, 301–314.
Hutyra, L. R., Yoon, B., & Alberti, M. (2011). Terrestrial carbon stocks across a gradient of urbanization: a study of the Seattle, WA region. *Global Change Biology*, *17*(2), 783–797.
Jenerette, G.D., Weller Clarke, L., Avolio, M.L., Pataki, D.E., Gillespie, T.W., Pincetl, S., Nowak, D.J., Hutyra, L.R., McHale, M., McFadden, J.P., & Alonzo, M. Climate filters and trait choices shape urban tree biodiversity: Beyond biome matching and homogenization. *In preparation.*
Jensen, J., & Cowen, D. (1999). Remote sensing of urban suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering and Remote Sensing*, *65*(5), 611–622.
Jensen, R. R., Hardin, P. J., Bekker, M., Farnes, D. S., Lulla, V., & Hardin, A. J. (2009). Modeling urban leaf area index with AISA+ hyperspectral data. *Applied Geography*, *29*(3), 320–332.
Jensen, R. R., Hardin, P. J., & Hardin, A. J. (2012). Estimating Urban Leaf Area Index (LAI) of Individual Trees with Hyperspectral Data. *Photogrammetric Engineering & Remote Sensing*, *78*(5), 495–504.
Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., & Baret, F. (2004). Review of methods for in situ leaf area index determination. *Agricultural and Forest Meteorology*, *121*(1-2), 19–35.
Jonckheere, I., Nackaerts, K., Muys, B., & Coppin, P. (2005). Assessment of automatic gap fraction estimation of forests from digital hemispherical photography. *Agricultural and Forest Meteorology*, *132*(1-2), 96–114.

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(12), 2841–2852.

Kalkstein, L. S., & Greene, J. S. (1997). An evaluation of climate/mortality relationships in large U.S. cities and the possible impacts of a climate change. *Environmental Health Perspectives*, *105*(1), 84–93.

Kampe, T. U. (2010). NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. *Journal of Applied Remote Sensing*, *4*(1), 043510.

Kim, S., Hinckley, T. M., & Briggs, D. J. (2011). Classifying individual tree genera using stepwise cluster analysis based on height and intensity metrics derived from airborne laser scanner data. *Remote Sensing of Environment*, *115*(12), 3329–3342.

Kim, S., Mcgaughey, R. J., Andersen, H., & 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*(8), 1575–1586.

Klecka, W. (1980), *Discriminant Analysis*. Beverly Hills: Sage Publications.

Koetz, B., Morsdorf, F., van der Linder, S., Curt, T., & Allgöwer, 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.

Kokaly, R. F., Asner, G. P., Ollinger, S. V, Martin, M. E., & Wessman, C. A. (2009). Remote Sensing of Environment Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sensing of Environment*, *113*, S78–S91.

Korhonen, L., Korpela, I., Heiskanen, J., & Maltamo, M. (2011). Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. *Remote Sensing of Environment*, *115*(4), 1065–1080.

Kuo, F. E., & Sullivan, W. C. (2001). Environment and Crime in the Inner City: Does Vegetation Reduce Crime? *Environment and Behavior*, *33*(3), 343–367.

Laćan, I., & McBride, J. R. (2008). Pest Vulnerability Matrix (PVM): A graphic model for assessing the interaction between tree species diversity and urban forest susceptibility to insects and diseases. *Urban Forestry & Urban Greening*, *7*(4), 291–300.

Latifi, H., Fassnacht, F., & Koch, B. (2012). Forest structure modeling with combined airborne hyperspectral and LiDAR data. *Remote Sensing of Environment*, 121, 10–25.

Lee, J. H., & Bang, K. W. (2000). Characterization of Urban Stormwater Runoff. *Water Resources*, *34*(6), 1773–1780.

Lim, K., Treitz, P., Wulder, M., St-Onge, B., & Flood, M. (2003). LiDAR remote sensing of forest structure. *Progress in Physical Geography*, 27(1), 88–106.

Lovell, J., Jupp, D., Culvenor, D. S., & Coops, N. C. (2003). Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. *Canadian Journal of Remote Sensing*, *29*(5), 607–622.

Lu, D., & Weng, Q. (2006). Spectral mixture analysis of ASTER images for examining the relationship between urban thermal features and biophysical descriptors in Indianapolis, Indiana, USA. *Remote Sensing of Environment*, *104*(2), 157–167.

Lucas, R. M., Lee, A. C., & Bunting, P. J. (2008). Retrieving forest biomass through integration of CASI and LiDAR data. *International Journal of Remote Sensing*, 29(5), 1553–1577.
Lyytimäki, J., Petersen, L. K., Normander, B., & Bezák, P. (2008). Nature as a nuisance? Ecosystem services and disservices to urban lifestyle. *Environmental Sciences*, *5*(3), 161–172.
Lyytimäki, J., & Sipilä, M. (2009). Hopping on one leg – The challenge of ecosystem disservices for urban green management. *Urban Forestry & Urban Greening*, *8*(4), 309–315.
MacFaden, S. W., O'Neil-Dunne, J., Royar, A. R., Lu, J. W. T., & Rundle, A. G. (2012). High-resolution tree canopy mapping for New York City using LIDAR and object-based image analysis. *Journal of Applied Remote Sensing*, *6*(1), 063567–1.
Manning, W. J. (2008). Plants in urban ecosystems: Essential role of urban forests in urban metabolism and succession toward sustainability. *The International Journal of Sustainable Development and World Ecology*, *15*(4), 362–370.
Martin, M. E., Newman, S. D., Aber, J. D., & Congalton, R. G. (1998). Determining forest species composition using high spectral resolution remote sensing data. *Remote Sensing of Environment*, 65, 249–254.
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(12), 3770–3774.
McCarthy, H. R., & Pataki, D. E. (2010). Drivers of variability in water use of native and non-native urban trees in the greater Los Angeles area. *Urban Ecosystems*, *13*(4), 393–414.
McGee, J. A., Day, S. D., Wynne, R. H., & White, M. B. (2012). Using geospatial tools to assess the urban tree canopy: Decision support for local governments. *Journal of Forestry*, 110(5), 275-286.
McHale, M. R., Burke, I. C., Lefsky, M., Peper, P. J., & McPherson, E. G. (2009). Urban forest biomass estimates: is it important to use allometric relationships developed specifically for urban trees? *Urban Ecosystems*, *12*(1), 95–113.
McPherson, E. G., & Simpson, J. R. (1999a). *Carbon Dioxide Reduction Through Urban Forestry*.
McPherson, E. G., & Simpson, J. R. (1999b). *Carbon Dioxide Reduction Through Urban Forestry : Guidelines for Professional and Volunteer Tree Planters*.
McPherson, E. G., Simpson, J. R., Xiao, Q., & Wu, C. (2011). Million trees Los Angeles canopy cover and benefit assessment. *Landscape and Urban Planning*, *99*(1), 40–50.
Morani, A., Nowak, D., Hirabayashi, S., Guidolotti, G., Medori, M., Muzzini, V., … Calfapietra, C. (2014). Comparing i-Tree modeled ozone deposition with field measurements in a periurban Mediterranean forest. *Environmental Pollution*, *195*, 202–209.
Morsdorf, F., Frey, O., Meier, E., & Itten, K. (2008). Assessment of the influence of flying altitude and scan angle on biophysical vegetation products derived from airborne laser scanning. *International Journal of Remote Sensing*, *29*(5), 1387–1406.
Morsdorf, F., Kotz, B., Meier, E., Itten, K., & Allgower, B. (2006). Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sensing of Environment*, *104*(1), 50–61.

Muss, J. D., Mladenoff, D. J., & Townsend, P. a. (2011). A pseudo-waveform technique to assess forest structure using discrete lidar data. *Remote Sensing of Environment*, *115*(3), 824–835.

Myint, S. W., Brazel, A., Okin, G., & Buyantuyev, A. (2010). Combined Effects of Impervious Surface and Vegetation Cover on Air Temperature Variations in a Rapidly Expanding Desert City. *GIScience & Remote Sensing*, *47*(3), 301–320.

Myint, S. W., Gober, P., Brazel, A., Grossman-Clarke, S., & Weng, Q. (2011). Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, *115*(5), 1145–1161.

Næsset, E., & Gobakken, T. (2008). Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment*, 112(6), 3079–3090.

Nicholls, R. J., & Cazenave, A. (2010). Sea-level rise and its impact on coastal zones. *Science*, *328*, 1517–1520.

Nowak, D. J. (1996). Estimating Leaf Area and Leaf Biomass of Open-Grown Deciduous Urban Trees. *Forest Science*, *42*(4).

Nowak, D. J. (2012). Contrasting natural regeneration and tree planting in fourteen North American cities. *Urban Forestry and Urban Greening*, *11*(4), 374–382.

Nowak, D. J., & Crane, D. E. (2000). The Urban Forest Effects (UFORE) Model: quantifying urban forest structure and functions. *Integrated Tools Proceedings*.

Nowak, D. J., & Crane, D. E. (2002). Carbon storage and sequestration by urban trees in the USA. *Environmental Pollution*, *116*(3), 381–389.

Nowak, D. J., Crane, D. E., & Stevens, J. C. (2006a). Air pollution removal by urban trees and shrubs in the United States. *Urban Forestry & Urban Greening*, *4*, 115–123.

Nowak, D. J., Crane, D. E., Stevens, J., Hoehn, R. E., Walton, J., & Bond, J. (2008a). A ground-based method of assessing urban forest structure and ecosystem services. *Arboriculture and Urban Forestry*, *34*(6), 347–358.

Nowak, D. J., Greenfield, E. J., Hoehn, R. E., & Lapoint, E. (2013a). Carbon storage and sequestration by trees in urban and community areas of the United States. *Environmental Pollution*, *178*, 229–236.

Nowak, D. J., Hirabayashi, S., Bodine, A., & Greenfield, E. (2014). Tree and forest effects on air quality and human health in the United States. *Environmental Pollution*, *193*, 119–129.

Nowak, D. J., Hoehn, R. E., Bodine, A. R., Greenfield, E. J., & O'Neil-Dunne, J. (2013b). Urban forest structure, ecosystem services and change in Syracuse, NY. *Urban Ecosystems*.

Nowak, D. J., Hoehn, R. E., Crane, D. E., Clarke, L. W., & Davila, A. (2010). *Assessing Urban Forest Effects and Values: Los Angeles' Urban Forest*.

Nowak, D. J., Hoehn, R. E., Crane, D. E., Stevens, J. C., & Walton, J. T. (2006b). *Assessing Urban Forest Effects and Values: Washington, D.C.'s Urban Forest*.

Nowak, D. J., Walton, J., Stevens, J., Crane, D. E., & Hoehn, R. E. (2008b). Effect of plot and sample size on timing and precision of urban forest assessments. *Arboriculture and Urban Forestry*, *34*(6), 386–390.

Oke, T. (1982). The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, *108*(455), 1–24.

Oke, T. (1989). The micrometeorology of the urban forest. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *324*(1223), 335–349.
Oleson, K. W., Bonan, G. B., Feddema, J., Vertenstein, M., & Grimmond, C. S. B. (2008). An Urban Parameterization for a Global Climate Model. Part I: Formulation and Evaluation for Two Cities. *Journal of Applied Meteorology and Climatology*, *47*(4), 1038–1060.
Oleson, K. W., Lawrence, D. M., Gordon, B., Flanner, M. G., Kluzek, E., Peter, J., … Zeng, X. (2010). *Technical Description of version 4 . 0 of the Community Land Model (CLM)*.
Ørka, H. O., Næsset, E., & Bollandsås, O. M. (2009). Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sensing of Environment*, 113(6), 1163–1174.
Ørka, H. O., Gobakken, T., Næsset, E., Ene, L., & Lien, V. (2012). Simultaneously acquired airborne laser scanning and multispectral imagery for individual tree species identification. *Canadian Journal of Remote Sensing*, 38(02), 125-138.
Osmond, P. (2009). Application of near-infrared hemispherical photography to estimate leaf area index of urban vegetation. In *International Conference on Urban Climate* (pp. 1–4).
Pekin, B., & Macfarlane, C. (2009). Measurement of Crown Cover and Leaf Area Index Using Digital Cover Photography and Its Application to Remote Sensing. *Remote Sensing*, *1*(4), 1298–1320.
Peper, P. J., & McPherson, E. G. (2003). Evaluation of four methods for estimating leaf area of isolated trees. *Urban Forestry & Urban Greening*, *2*(1), 19–29.
Peters, E. B., & McFadden, J. P. (2010). Influence of seasonality and vegetation type on suburban microclimates. *Urban Ecosystems*, *13*(4), 443–460.
Pillsbury, N. H., Reimer, J. L., & Thompson, R. (1998). *Tree volume equations for fifteen urban species in California*. Urban Forest Ecosystems Institute, California Polytechnic State University.
Pohl, C., & Van Genderen, J. (1998). Review article Multisensor image fusion in remote sensing: concepts, methods and applications. *International Journal of Remote Sensing*, *19*(5), 37–41.
Popescu, S. C. (2007). Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy*, *31*(9), 646–655.
Popescu, S., Wynne, R., & Nelson, R. (2003). Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, 29(5), 564–577.
Popescu, S. C., & Zhao, K. (2008). A voxel-based lidar method for estimating crown base height for deciduous and pine trees. *Remote Sensing of Environment*, *112*(3), 767–781.
Price, J. C. (1994). How Unique Are Spectral Signatures? *Remote Sensing of Environment*, *49*(3), 181–186.
Pu, R. (2009). Broadleaf species recognition with in situ hyperspectral data. *International Journal of Remote Sensing*, *30*(11), 2759–2779.
Pu, R., & Liu, D. (2011). Segmented canonical discriminant analysis of in situ hyperspectral data for identifying 13 urban tree species. *International Journal of Remote Sensing*, 32(8), 2207–2226.

Pu, R., & Landry, S. (2012). A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. *Remote Sensing of Environment*, *124*, 516–533.

Raciti, S. M., Hutyra, L. R., & Newell, J. D. (2014). Mapping carbon storage in urban trees with multi-source remote sensing data: Relationships between biomass, land use, and demographics in Boston neighborhoods. *Science of The Total Environment*, *500-501*, 72–83.

Reitberger, J., Schnörr, C., Krzystek, P., & Stilla, U. (2009). 3D segmentation of single trees exploiting full waveform LIDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(6), 561–574.

Richardson, J. J., & Moskal, L. M. (2014a). Uncertainty in urban forest canopy assessment : Lessons from Seattle, WA, USA. *Urban Forestry & Urban Greening*, *13*, 152–157.

Richardson, J. J., & Moskal, L. M. (2014b). Uncertainty in urban forest canopy assessment: Lessons from Seattle, WA, USA. *Urban Forestry & Urban Greening*, *13*(1), 152–157.

Richardson, J. J., Moskal, L. M., & Kim, S.-H. (2009a). Modeling approaches to estimate effective leaf area index from aerial discrete-return LIDAR. *Agricultural and Forest Meteorology*, *149*(6-7), 1152–1160.

Richardson, J. J., Moskal, L. M., & Kim, S.-H. (2009b). Modeling approaches to estimate effective leaf area index from aerial discrete-return LIDAR. *Agricultural and Forest Meteorology*, *149*(6-7), 1152–1160.

Richter, R., & Schlaepfer, D. (2002). Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: atmospheric/topographic correction. *International Journal of Remote Sensing*, *23*(13), 37–41.

Roberts, D. A., Dennison, P. E., Peterson, S., Sweeney, S., & Rechel, J. (2006). Evaluation of Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and Moderate Resolution Imaging Spectrometer (MODIS) measures of live fuel moisture and fuel condition in a shrubland ecosystem in southern California. *Journal of Geophysical Research*, *111*(G4), 1–16.

Roberts, D. A., Gardner, M., Church, R., Ustin, S., Scheer, G., & Green, R. O. (1998). Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment*, 65(3), 267–279.

Roberts, D. A., Quattrochi, D. A., Hulley, G. C., Hook, S. J., & Green, R. O. (2012). Synergies between VSWIR and TIR data for the urban environment: An evaluation of the potential for the Hyperspectral Infrared Imager (HyspIRI) Decadal Survey mission. *Remote Sensing of Environment*, *117*, 83–101.

Roberts, D. A., Ustin, S. L., Ogunjemiyo, S., Greenberg, J., Dobrowski, S. Z., Chen, J. M., & Hinckley, T. M. (2004). Spectral and Structural Measures of Northwest Forest Vegetation at Leaf to Landscape Scales. *Ecosystems*, *7*(5), 545–562.

Rouse, J. W., Haas, R. S., Schell, J. A., and Deering, D. W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. *Proceedings, 3rd ERTS Symposium*, 1, 48–62.

Roy, S., Byrne, J., & Pickering, C. (2012). A systematic quantitative review of urban tree benefits, costs, and assessment methods across cities in different climatic zones. *Urban Forestry & Urban Greening*, *11*(4), 351–363.

Ryu, Y., Nilson, T., Kobayashi, H., Sonnentag, O., Law, B. E., & Baldocchi, D. (2010a). On the correct estimation of effective leaf area index: Does it reveal information on clumping effects? *Agricultural and Forest Meteorology*, *150*(3), 463–472.

Ryu, Y., Sonnentag, O., Nilson, T., Vargas, R., Kobayashi, H., Wenk, R., & Baldocchi, D. D. (2010b). How to quantify tree leaf area index in an open savanna ecosystem: A multi-instrument and multi-model approach. *Agricultural and Forest Meteorology*, *150*(1), 63–76.

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*(8), 1646–1656.

Sanders, R. A. (1984). Some determinants of urban forest structure. *Urban Ecology*, *8*(1-2), 13–27.

Santamour, F. S. (1990). Trees for urban planting: Diversity, uniformity, and common sense. In *Proceedings of the Seventh Conference of the Metropolitan Tree Improvement Alliance (METRIA)* (pp. 57–65).

Schneider, A., Friedl, M. A., & Potere, D. (2010). Mapping global urban areas using MODIS 500-m data: New methods and datasets based on "urban ecoregions." *Remote Sensing of Environment*, *114*(8), 1733–1746.

Schroeder, H. W., Green, T. L., & Howe, T. J. (2003). Community tree programs in Illinois, U.S.: A statewide survey and assessment. *Journal of Arboriculture*, *29*(4), 218–224.

Schwarz, K., Fragkias, M., Boone, C. G., Zhou, W., McHale, M., Grove, J. M., … Cadenasso, M. L. (2015). Trees Grow on Money: Urban Tree Canopy Cover and Environmental Justice. *Plos One*, *10*(4), e0122051.

Shrestha, R., & Wynne, R. H. (2012). Estimating Biophysical Parameters of Individual Trees in an Urban Environment Using Small Footprint Discrete-Return Imaging Lidar. *Remote Sensing*, 4(12), 484–508.

Simpson, J. (2002). Improved estimates of tree-shade effects on residential energy use. *Energy and Buildings*, *34*(10), 1067–1076.

Solberg, S., Brunner, A., Hanssen, K. H., Lange, H., Næsset, E., Rautiainen, M., & Stenberg, P. (2009a). Mapping LAI in a Norway spruce forest using airborne laser scanning. *Remote Sensing of Environment*, *113*(11), 2317–2327.

Solberg, S., Brunner, A., Hanssen, K. H., Lange, H., Næsset, E., Rautiainen, M., & Stenberg, P. (2009b). Mapping LAI in a Norway spruce forest using airborne laser scanning. *Remote Sensing of Environment*, *113*(11), 2317–2327.

Solberg, S., Næsset, E., Hanssen, K. H., & Christiansen, E. (2006). Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. *Remote Sensing of Environment*, *102*(3-4), 364–376.

Sprugel, D. G. (1983). Correcting for bias in log-transformed allometric equations. *Ecology*.

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(11), 2917–2930.

Tague, C. L., & Band, L. E. (2004). RHESSys : Regional Hydro-Ecologic Simulation System — An Object- Oriented Approach to Spatially Distributed Modeling of Carbon , Water , and Nutrient Cycling. *Earth Interactions*, *8*(19).

Tang, H., Dubayah, R., Swatantran, A., Hofton, M., Sheldon, S., Clark, D. B., & Blair, B. (2012). Retrieval of vertical LAI profiles over tropical rain forests using waveform lidar at La Selva, Costa Rica. *Remote Sensing of Environment*, 124, 242–250.

Toomey, M., Roberts, D. A., & Nelson, B. (2009). The influence of epiphylls on remote sensing of humid forests. *Remote Sensing of Environment*, *113*(8), 1787–1798.

Tzoulas, K., Korpela, K., Venn, S., Yli-Pelkonen, V., Kaźmierczak, A., Niemela, J., & James, P. (2007). Promoting ecosystem and human health in urban areas using Green Infrastructure: A literature review. *Landscape and Urban Planning*, *81*(3), 167–178.

Ulrich, R. S. (1984). View Through Surgery Window May Influence Recovery from. *Science*, *224*(420).

UN-Habitat. (2011). *Cities and climate change : global report on human settlements*. United Nations Human Settlements Programme.

United States Census (2010).

United Nations, Department of Economic and Social Affairs, Population Division. (2014). World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352).

Urban, J. (1992). Bringing order to the technical dysfunction within the urban forest. *Journal of Arboriculture*, *18*(2), 85–90.

US Climate Data. (2015). www.usclimatedata.com.

Ustin, S. L., Gitelson, A. A., Jacquemoud, S., Schaepman, M., Asner, G. P., Gamon, J. a., & Zarco-Tejada, P. (2009). Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sensing of Environment*, *113*, S67–S77.

van Aardt, J., & Wynne, R. H. (2001). Spectral Separability among Six Southern Tree Species. *Photogrammetric Engineering & Remote Sensing*, 67(12), 1367–1375.

van Aardt, J., & Wynne, R. (2007). Examining pine spectral separability using hyperspectral data from an airborne sensor: An extension of field-based results. *International Journal of Remote*.

Van Gardingen, P. R., Jackson, G. E., Hernandez-Daumas, S., Russell, G., & Sharp, L. (1999). Leaf area index estimates obtained for clumped canopies using hemispherical photography. *Agricultural and Forest Meteorology*, *94*(3-4), 243–257.

Vaughn, N. R., Moskal, L. M., & Turnblom, E. C. (2012). Tree Species Detection Accuracies Using Discrete Point Lidar and Airborne Waveform Lidar. *Remote Sensing*, 4(2), 377–403.

Voogt, J. A. (2002). Urban heat island. In I. Douglas (Ed.), *Causes and consequences of global environmental change*, Vol. 3. (pp. 660–666) Chichester: John Wiley & Sons, Ltd.

Voss, M., & Sugumaran, R. (2008). The Seasonal Effect on Tree Species Classification in an Urban Environment Using Hyperspectral Data, LiDAR, and an Object-Oriented Approach. *Sensors*, 8(5), 3020–3036.

Wang, J., Endreny, T. a., & Nowak, D. J. (2008). Mechanistic Simulation of Tree Effects in an Urban Water Balance Model. *JAWRA Journal of the American Water Resources Association*, *44*(1), 75–85.

Wania, A., Bruse, M., Blond, N., & Weber, C. (2012). Analysing the influence of different street vegetation on traffic-induced particle dispersion using microscale simulations. *Journal of Environmental Management*, *94*(1), 91–101.

Welch, R. (1982). Spatial resolution requirements for urban studies. *International Journal of Remote Sensing*, *3*(2), 1982.

Weng, Q. (2009). Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(4), 335–344.
White, R. H., & Zipperer, W. C. (2010). Testing and classification of individual plants for fire behaviour: plant selection for the wildland–urban interface. *International Journal of Wildland Fire*, *19*(2), 213.
Wolf, K. L. (2005). Trees in the Small City Retail Business District : Comparing Resident and Visitor Perceptions. *Journal of Forestry*, *103*(8).
Woodcock, C., & Strahler, a. (1987). The factor of scale in remote sensing. *Remote Sensing of Environment*, *21*(3), 311–332.
Xiao, Q., McPherson, E., Ustin, S., & Grismer, M. E. (2000). A new approach to modeling tree rainfall interception. *Journal of Geophysical Research*, *105*(D23), 29173–29188.
Xiao, Q., & McPherson, E. G. (2002). Rainfall interception by Santa Monica's municipal urban forest. *Urban Ecosystems*, *6*, 291–302.
Xiao, Q., & McPherson, E. G. (2011). Rainfall interception of three trees in Oakland, California. *Urban Ecosystems*, *14*(4), 755–769.
Xiao, Q., Ustin, S. L., & McPherson, E. G. (2004). Using AVIRIS data and multiple-masking techniques to map urban forest tree species. *International Journal of Remote Sensing*, *25*(24), 5637–5654.
Xu, W., Wooster, M., & Grimmond, C. (2008). Modelling of urban sensible heat flux at multiple spatial scales: A demonstration using airborne hyperspectral imagery of Shanghai and a temperature–emissivity separation approach. *Remote Sensing of Environment*, *112*(9), 3493–3510.
Yang, Y., Endreny, T. a., & Nowak, D. J. (2013). A physically based analytical spatial air temperature and humidity model. *Journal of Geophysical Research: Atmospheres*, *118*(18), 10,449–10,463.
Yang, C., Everitt, J. H., Fletcher, R. S., Jensen, R. R., & Mausel, P. W. (2009). Evaluating AISA + Hyperspectral Imagery for Mapping Black Mangrove along the South Texas Gulf Coast. *Photogrammetric Engineering & Remote Sensing*, 75(4), 425–435.
Yao, W., Krzystek, P., & Heurich, M. (2012). Tree species classification and estimation of stem volume and DBH based on single tree extraction by exploiting airborne full-waveform LiDAR data. *Remote Sensing of Environment*, 123, 368–380.
Youngentob, K. N., Roberts, D. a., Held, A. A.., Dennison, P. E., Jia, X., & Lindenmayer, D. B. (2011). Mapping two Eucalyptus subgenera using multiple endmember spectral mixture analysis and continuum-removed imaging spectrometry data. *Remote Sensing of Environment*, 115(5), 1115–1128.
Yu, B., Ostland, M., Gong, P., & Pu, R. (1999). Penalized discriminant analysis of in situ hyperspectral data for conifer species recognition. *IEEE Transactions on Geoscience and Remote Sensing*, 37(5), 2569–2577.
Zhang, C., & Qiu, F. (2012). Mapping Individual Tree Species in an Urban Forest Using Airborne Lidar Data and Hyperspectral Imagery. *Photogrammetric Engineering & Remote Sensing*, *78*(10), 1079–1087.
Zhang, Y., Chen, J. M., & Miller, J. R. (2005). Determining digital hemispherical photograph exposure for leaf area index estimation. *Agricultural and Forest Meteorology*, *133*(1-4), 166–181.

Zhao, G., & Maclean, A. (2000). A comparison of canonical discriminant analysis and principal component analysis for spectral transformation. *Photogrammetric Engineering & Remote Sensing*, 66(7), 841–847.

Zhao, K., & Popescu, S. C. (2009). Lidar-based mapping of leaf area index and its use for validating GLOBCARBON satellite LAI product in a temperate forest of the southern USA. *Remote Sensing of Environment*, *113*(8), 1628–1645.