Title
The Use of LiDAR in Multi-Scale Forestry Applications

Permalink
https://escholarship.org/uc/item/7p97c4ng

Author
Su, Yanjun

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, MERCED

The Use of LiDAR in Multi-Scale Forestry Applications

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Environmental Systems

by

Yanjun Su

Committee in charge:

    Professor Qinghua Guo, Chair
    Professor Roger C. Bales, Co-Chair
    Professor Thomas C. Harmon
    Professor Shawn Newsam
    Professor Joshua Viers

2017

Portion of Chapter 2 © 2016 Taylor & Francis Group

Portion of Chapter 3 © 2016 Taylor & Francis Group

Portion of Chapter 4 © 2016 Taylor & Francis Group

Portion of Chapter 5 © 2016 Elsevier and 2016 Multidisciplinary Digital Publishing
Institute

©

Yanjun Su, 2017

The Dissertation of Yanjun Su is approved, and it is acceptable
in quality and form for publication on microfilm and electronically:

Dr. Joshua Viers

Dr. Shawn Newsam

Dr. Thomas C. Harmon

Dr. Roger C. Bales, Co-Chair

Dr. Qinghua Guo, Chair

University of California, Merced

2017

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to express my appreciations to my advisors Professor Qinghua Guo and Professor Roger Bales for their guidance and support in my research. They direct and supervise my research, and help me to develop independent research skills. They are good at motivating students, and are always ready to help whenever I encounter challenges.

I want to thank Professor Thomas Harmon, Professor Shawn Newsam, Professor Joshua Viers, Professor Martha Conklin, Professor Maggi Kelly, and Professor John Battles for their time, constructive comments, and valuable guidance in my research at UC Merced.

I am grateful to my labmates and colleagues, Qin Ma, Dr. Wenkai Li, Dr. Otto Alvarez, Jacob Flanagan, Dr. Bo Wan, Dr. Fang Fang, Dr. Jia Wang, Dr. Laiping Luo, Melissa Thaw, Xiande Meng, Joe Rungee, Jim Roche, and Erin Stacy. I also want to thank for the support of Coty Ventura, Alexis Valle-Arevalo, Crystal Galvan, and other staff at SNRI for their support on my research and life at UC Merced. I would like to further to thank my friends, Youhong Zeng, Shuo Liu, Chengjie Qin, Andrew Zumkehr, Joannee Zumkehr, Tao Ren, Zhijiang Ye, and others, for their support and help on my life at Merced.

Portion of Chapter 2 is a reprint of the material as it appears in (Su Y., Guo Q., Collins B., Fry D., Kelly M., 2016. Forest fuel treatment detection using multi-temporal airborne LiDAR data and high-resolution aerial imagery: A case study at Sierra Nevada Mountains, California. International Journal of Remote Sensing, 37, 3322-3345.). Portion of Chapter 3 is a reprint of the material as it appears in (Su Y., Guo Q., Fry D., Collins B., Kelly M., Flanagan J., Battles J., 2016. A vegetation mapping strategy for conifer forests by combining airborne LiDAR data and aerial imagery. Canadian Journal of Remote Sensing, 42, 1-15.). Portion of Chapter 4 is a reprint of the material as it appears in (Su Y., Ma Q., Guo Q., 2017. Fine-resolution forest tree height estimation across the Sierra Nevada through the integration of spaceborne LiDAR, airborne LiDAR and optical imagery. International Journal of Digital Earth, 10, 307-323.). Portion of Chapter 5 is a reprint of the materials as they appear in (Su Y., Guo Q., Xue B., Hu T., Alvarez O., Tao S., Fang J., 2016. Spatial distribution of forest aboveground biomass in China: Estimation through combination of spaceborne LiDAR, optical imagery, and forest inventory data. Remote Sensing of Environment, 173, 187-199; Hu T., Su Y.#, Xue B., Liu J., Zhao X., Fang J., Guo Q., 2016. Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery, and forest inventory data. Remote Sensing, 8, 565, # indicates co-first author). Permissions to use copyrighted materials in the dissertation have been granted by Taylor & Francis Group, Elsevier, and Multidisciplinary Digital Publishing Institute.

Finally, I want to send extra special thanks to my families, who give me strong support in my life. I would not be where I am today without your love, support, and unwavering encouragement.

Curriculum Vitae

Yanjun Su

School of Engineering
Environmental Systems Graduate Group
University of California Merced, CA 95343
Email: ysu3@ucmerced.edu, suyanjun1987@gmail.com

Education

Ph.D. Candidate in Environmental Systems   May, 2017 (Expected)
   University of California, Merced
   Dissertation title: The use of LiDAR in multi-scale forestry applications.
   Advisor: Dr. Qinghua Guo; Co-advisor: Dr. Roger C. Bales

M.S. in Cartography & Geographic Information Science   July, 2012
   University of Chinese Academy of Sciences
   Thesis title: Extraction of the geo-patterns in the Lower Yellow River changes during
   the last 3000 years.

B.E. in Surveying & Mapping Engineering   July, 2009
   China University of Geosciences, Beijing

Publications

Refereed Journal Articles

Guo Q., **Su Y.**, Liu J., Zhao X., Wu F., Chen L., Hu T., Xu G., Lin G., Zheng Y., Lin Y., Mi X., Fei L., Wang X., 2017. An integrated UAV-borne LiDAR system for 3D habitat mapping in three forest ecosystems across China. *International Journal of Remote Sensing*, 38, 2954-2972.

Zhu J., **Su Y.**, Guo Q., Harmon T.C., 2017. Unsupervised object-based differencing procedure for land-use change detection using very high spatial resolution data. *Photogrammetric Engineering & Remote Sensing*, 83, 225-236.

Su Y., Ma Q., Guo Q., 2017. Fine-resolution forest tree height estimation across the Sierra Nevada through the integration of spaceborne LiDAR, airborne LiDAR and optical imagery. *International Journal of Digital Earth*, 10, 307-323.

Li Y., Guo Q., Tao S., Zheng G., Zhao K., **Su Y.**, 2016. Derivation, validation, and sensitivity analysis of terrestrial laser scanning-based leaf area index. *Canadian Journal of Remote Sensing*, 42, 719-729.

Hu T., **Su Y.**[#], Xue B., Liu J., Zhao X., Fang J., Guo Q., 2016. Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery, and forest inventory data. *Remote Sensing*, 8, 565. ([#] **Co-first author**)

Su Y., Guo Q., Collins B., Fry D., Kelly M., 2016. Forest fuel treatment detection using multi-temporal airborne LiDAR data and high-resolution aerial imagery: A case study at Sierra Nevada Mountains, California. *International Journal of Remote Sensing*, 37, 3322-3345. (**Journal cover page article**)

Zhao X., Guo Q., **Su Y.**, Xue B., 2016. Improved progressive TIN densification filtering algorithm for airborne LiDAR in forested areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117, 79-91.

Su Y., Guo Q., Xue B., Hu T., Alvarez O., Tao S., Fang J., 2016. Spatial distribution of forest aboveground biomass in China: Estimation through combination of spaceborne LiDAR, optical imagery, and forest inventory data. *Remote Sensing of Environment*, 173, 187-199.

Su Y., Guo Q., Fry D., Collins B., Kelly M., Flanagan J., Battles J., 2016. A vegetation mapping strategy for conifer forests by combining airborne LiDAR data and aerial imagery. *Canadian Journal of Remote Sensing*, 42, 1-15.

Su Y., Guo Q., Ma Q., Li W., 2015. SRTM DEM correction in vegetated mountain areas through the integration of spaceborne LiDAR, airborne LiDAR, and optical imagery. *Remote Sensing*, 7, 11202-11225.

Tempel D.J., Gutiérrez R.J., Battles J.J., Fry D.L., **Su Y.**, Guo Q., Reetz M.J., Whitmore S.A., Jones G.M., Collins B.M., Stephens L.L., Kelly M., Berigan W.J., Peery M.Z., 2015. Evaluating short- and long-term impacts of fuels treatments and wildfire on an old-forest species. *Ecosphere*, 12, art261.

Wan B., Guo Q., Fang F., **Su Y.**, Wang R., 2015. Mapping US urban extents from time-series MODIS data using one-class. *Remote Sensing*, 7, 10143-10163.

Tao S., Guo Q., Xu S., **Su Y.**, Li Y., Wu F., 2015. A geometric method for wood–leaf separation
using terrestrial LiDAR data and simulated point cloud. *Photogrammetric Engineering & Remote Sensing*, 81, 767-776.

Tao S., Guo Q., Li L., Xue B., Kelly M., Li W., **Su Y.**, 2014. Airborne LiDAR-derived volume metrics for aboveground biomass estimation: A comparative assessment for conifer stands. *Agricultural and Forest Meteorology*, 198, 24-32.

Su Y., Guo Q., 2014. A practical method for SRTM DEM correction over vegetated mountain areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 216-228.

Wang Y., **Su Y.***, 2013. Influence of solar activity on breaching, overflowing and course shifting events of the lower yellow river in the late Holocene. *Holocene*, 23, 656-666. **(* Corresponding author)**

Wang Y., **Su Y.***, 2011. The geo-pattern of course shifts of the Lower Yellow River. *Journal of Geographical Sciences*, 21, 1019-1036. **(* Corresponding author)**

Manuscripts in Submission

Su Y., Bales R.C., Ma Q., Guo. Q., Nydick K., Ray R., Li W. Relative resiliency of Giant Sequoia groves to multi-year dry periods in a warming climate. *Remote Sensing of Environment*. (In preparation)

Zhao X., **Su Y.** [#], Hu T., Chen L., Gao S., Wang R., Jin S., Guo Q. A global corrected SRTM DEM product over vegetated areas. *Geophysical Research Letters*. ([#] **Co-first author;** submitted)

Li W., Guo Q., Tao S., **Su Y.** A voxel-based radiative transfer model for heterogeneous three-dimensional scenes. *Remote Sensing of Environment*. (Major revision)

Ma Q., **Su Y.**, Guo Q. Responses of tree growths to tree size, competition, and topographic conditions in the Sierra Nevada forests using bi-temporal ALS data. *International Journal of Digital Earth*. (Submitted)

Ma Q., **Su Y.**, Guo Q. Comparison of canopy cover estimations from airborne LiDAR, aerial imagery, and satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. (Major revision)

Li Y., **Su Y.**, Tao S., Zhao K., Xu G. Retrieving the gap fraction, element clumping index, and leaf area index of individual trees using single-scan data from a terrestrial laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing*. (Major revision)

Xue B., Guo Q., Hu T., Wang Y., Tao S., **Su Y.**, Liu J., Zhao X. Evaluation of modeled global carbon dynamics: analysis based on global carbon flux and above-ground biomass data. *Agricultural and Forest Meteorology*. (Submitted)

Xue B., Guo Q., Hu T., Xiao J., Yang Y., Tao S., **Su Y.**, Liu J., Zhao X. Global patterns of woody residence time and its influence on model simulation of aboveground biomass. *Global Biogeochemical Cycles.* (Major revision)

Zhao X., Guo Q., Li W., **Su Y.**, Liu J. Influence of slope, canopy cover, elevation variability, and LiDAR sampling density on several filtering algorithms in forest areas. *International Journal of Remote Sensing*. (Submitted)

Honors & Awards

William A. Fischer Memorial Scholarship, ASPRS, 2017
Travel Award, Sequoia and Kings Canyon Science Symposium, 2016
Travel Fellowship, UC Merced, 2016
Southern California Edison Fellowship, UC Merced, 2015
Summer Fellowship, UC Merced, 2013 and 2015
Bobcat Fellowship, UC Merced, 2014
Merit Student, University of Chinese Academy of Sciences, 2010 and 2011
Excellent Thesis for Bachelor Degree, China University of Geosciences, Beijing, 2009
Outstanding Graduates, China University of Geosciences, Beijing, 2009
The First Prize Scholarship, China University of Geosciences, Beijing, 2007-2009
National Scholarship**,** China University of Geosciences, Beijing, 2007
Merit Student, China University of Geosciences, Beijing, 2007
Excellent Physics Practice, China University of Geosciences, Beijing, 2007
The Runner-up Prize Scholarship, China University of Geosciences, Beijing, 2006

Invited Talks & Posters

Su Y. (2017). The use of LiDAR in multi-scale environmental studies. Louisiana State University, Baton Rouge, Louisiana. (Talk)

Su Y., Hu T., Xue B., Liu J., Zhao X., Fang J., Guo Q. (2016). Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery, and forest inventory data. 2016 AGU Fall Meeting, San Francisco, California. (Poster presentation)

Su Y., (2016). Vulnerability of Giant Sequoia groves to the changing climate and extreme droughts during the last three decades. Sequoia and Kings Canyon Science Symposium, Three Rivers, CA. (Poster presentation)

Su Y., (2016). Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery, and forest inventory data. The 6th International Society of Digital Earth Summit, Beijing, China. (Talk)

Su Y., (2015). The use of GIS and remote sensing techniques in environmental

preservations. The Intersection of Next Generation Technology and Resource Preservation seminar, University of California, Merced. (Talk)

Su Y., Guo Q., Ma Q., Li W. (2015). Fine resolution tree height estimation from LiDAR data and its application in SRTM DEM correction across forests of Sierra Nevada, California, USA. 2015 AGU Fall Meeting, San Francisco, California. (Poster presentation)

Su Y., Guo Q., Collins B., Fry D., Kelly M. (2014). Forest fuel treatment detection using multi-temporal airborne LiDAR data and high resolution aerial imagery ---- A case study at Sierra Nevada Mountains, California. 2014 AGU Fall Meeting, San Francisco, California. (Poster presentation)

Teaching Experience

Instructor. University of California, Merced Aug., 2016-Dec., 2016
  Course name: Spatial Analysis and Modelling (ENGR180)
  Evaluation: 6.41 on average out of 7.0.

Instructor. University of California, Merced Aug., 2015-Dec., 2015
  Course name: Remote Sensing of Environment (ENVE152/ES252)
  Evaluation: 6.54 on average out of 7.0.

Guest Instructor. University of California, Merced Oct., 2015
  Course name: Spatial Analysis and Modelling (ENGR180)
  Topic: The use of remote sensing data in spatial analysis

Graduate Mentor. University of California, Merced Aug., 2014-Dec., 2015
  Mentored two undergraduate students to collect and process airborne LiDAR data.

Teaching Assistant. University of California, Merced Aug., 2013-Dec., 2013
  Course name: Remote Sensing of Environment (ENVE152/ES252)
  Evaluation: 6.83 on average out of 7.0.

Service to Profession

- Peer mentor: Served as a peer mentor for three first-year graduate students in the GRAD-EXCEL Peer Mentorship Program of the University of California, Merced
- Facilitator: Co-organized two seminars and two symposiums for the Natural and Cultural Resources Preservation Interdisciplinary Research Group at the University of California, Merced.
- Reviewer: *International Journal of Digital Earth; ISPRS International Journal of Geo-Information*; *International Journal of Remote Sensing*.

Abstract

The Use of LiDAR in Multi-Scale Forestry Applications

by

Yanjun Su

Doctor of Philosophy in Environmental Systems

University of California, Merced, 2017

Professor Qinghua Guo, Chair
Professor Roger C. Bales, Co-Chair

Forest ecosystems are a significant faction of the Earth's landscape, and accurate estimates of forest structures are important for understanding and predicting how forest ecosystems respond to climate change and human activities. Light detection and ranging (LiDAR) technology, an active remote sensing technology, can penetrate the forest canopy and greatly improve the efficiency and accuracy of mapping forest structures, compared to traditional passive optical remote sensing and radar technologies. However, currently, LiDAR has two major weaknesses, the lack of spectral information and the limited spatial coverage. These weaknesses have limited its accuracy in certain forestry applications (e.g., vegetation mapping) and its application in large-scale forest structure mapping. The aim of research described in this dissertation is to develop data fusion algorithms to address these limitations. In this dissertation, the effectiveness of LiDAR in estimating forest structures and therefore monitoring forest dynamics is first compared with aerial imagery by detecting forest fuel treatment activities at the local scale. Then, a vegetation mapping algorithm is developed based on the fusion of LiDAR data and aerial imagery. This algorithm can automatically determine the optimized number of vegetation units in a forest and take both the vegetation species and vegetation structure characteristics into account in classifying the vegetation types. To extend the use of LiDAR in mapping forest structures in areas without LiDAR coverage, a data fusion algorithm is proposed to map fine-resolution tree height from airborne LiDAR, spaceborne LiDAR, optical imagery and radar data in regional scale. Finally, this dissertation further investigates the methodology to integrate spaceborne LiDAR, optical imagery, radar data and climate surfaces for the purpose of mapping national- to global-scale forest aboveground biomass. The proposed data fusion algorithms and the generated regional to global forest structure parameters will have important applications in ecological and hydrologic studies and forest management.

Chapter 1   Introduction

Forests cover around 30% of the Earth's land surface, and provide diverse ecosystem services, such as converting carbon dioxide to oxygen and biomass, acting as a carbon sink, regulating climate, purifying water, maintaining essential ecological processes and life support systems, and providing habitats for plants and animals (Führer et al., 2000; Lasch et al., 2002; Bales et al., 2011). To improve our understanding of the global forest ecosystem, it is of great significance to quantify three-dimensional (3D) forest structures, such as tree height, aboveground biomass (AGB) and leaf area index (LAI). However, currently, we still lack accurate forest structure estimations at national to global scales. How to map forest structures accurately and therefore monitor forest dynamics from local to global scales is still a question that needs to be addressed.

1.1 Traditional ways to measure forest structure

Traditionally, forest structures are observed based on field measurements. Surveyors use instruments [such as global positioning system (GPS) receiver, diameter tape, range finder, LAI 2000, hemispherical camera] to record the information of each individual trees [such as tree location, diameter at breast height (DBH), tree height, LAI]. These field measurements are highly accurate, but have three major limitations. First, taking field measurements is highly labor-intensive and time-consuming, especially in certain wild areas. Brienen et al. (2015) claimed that the time and labor costs of taking field measurements in wild Amazon forests was from 100 to 200 times of those in developed forest areas. Second, measuring certain forest parameters in the field is destructive to the environment. For example, the harvesting method is the most accurate way to measure forest AGB. However, the removal of trees may have negative impacts on the forest ecosystem, such as destroying the habitats of wildlife (Tempel et al., 2015). Third, field measurements are usually taken at the plot scale, which cannot be used to directly generate spatial continuous forest structure layers. These limitations have largely constrained the use of field measurements in large-scale forest structure mapping.

    The development of remote sensing technology can greatly help to address the limitations of field measurements. Remote sensing technology can provide continuous land surface observations with much higher efficiency and lower cost (Brivio et al., 2002; Jenson, 2009). Currently, passive optical sensors [e.g., Landsat Thematic Mapper (TM) and Moderate Resolution Imaging Spectroradiometer (MODIS) data] and radar sensors [e.g., phased array L-band Synthetic Aperture Radar (SAR)] are two frequently used remote sensing data sources, and there have been numerous studies focusing on how to use these data to map forest structure parameters. For example, Chen et al. (1996) used Landsat TM images to retrieve LAI of boreal conifer forests; Zhang et al. (2014) mapped the forest tree height, LAI and AGB in California using the Landsat TM data; Lu et al. (2005) explored using Landsat TM data to map forest AGB in Brazilian Amazon; Fensholt et al. (2004) evaluated the performance of MODIS in deriving LAI in semi-arid environment; Gao et al. (2008) proposed an algorithm to produce temporally and spatially continuous LAI from MODIS data. However, many studies have demonstrated that these

approaches using traditional passive optical imagery and radar data are fraught with uncertainty issues due to the saturation effect of these sensors. For example, it was found that the saturation points for using passive optical imagery to map forest AGB can range from 15 to 150 Mg/ha depending on different sensors (Baccini et al., 2008; Cartus et al., 2012; Foody et al., 2001; Luckman et al., 1997; Ranson and Sun et al., 1994; Saatchi et al., 2007; Sader et al., 1989; Sun et al., 2002; Toan et al., 1992; Wagner et al., 2003), and the saturation points for SAR are higher than optical imagery, which range from 30 to over 300 Mg/ha based on the use of different frequencies and polarization methods (Huang et al., 2015; Kasischke et al., 2012; Lu, 2006; Mitchard et al., 2009; Myneni et al., 2001; Woodhouse et al., 2012).

1.2 Light detection and ranging (LiDAR)

LiDAR, an active remote sensing technology, can greatly help to solve the above-mentioned problems related to field measurements and traditional remote sensing data. It shoots a focused short-wavelength laser pulse to a targeted land surface object, and measures the range to the object by determining the travel time of the laser pulse. Through integrating GPS and inertial measurement unit information, the obtained range measurement can be converted to accurate 3D locations (Shan and Toth, 2008). Because the use of a focused short-wavelength laser pulse, LiDAR can effectively penetrate the forest canopy and measure the 3D structures of forests. For example, Anderson et al. (2006), Clark et al. (2004), Naesset (1997), Nilsson (1996), Wing et al. (2015), and Zimble et al. (2003) explored the methods to derive forest tree height from LiDAR data at different scales; Morsdorf et al. (2006), Riano et al. (2004), Richardson et al. (2009), and Zhao and Popescu (2009) developed algorithms to estimate LAI from LiDAR data; Bortolot and Wynne (2005), Clark et al. (2011), Nelson et al. (2009), Popescu et al. (2011), and Swatanran et al. (2011) successfully used LiDAR data to estimate forest AGB in areas with various forest types. All these previous studies have shown that LiDAR can help to quantify forest structures with unprecedented accuracy, and therefore improve our understanding on global forest ecosystem.

1.3 Limitations of LiDAR

Although LiDAR can help to accurately quantify 3D forest structures, it currently has two major limitations. First, LiDAR lacks spectral information. Spectral information is highly useful for differentiating land surface objects, and it is the foundation of the passive optical remote sensing. There have been many studies showing that using hyperspectral and multispectral imagery can help to classify land cover types (Aplin, 2004; Franklin, 1986; Green et al., 1994; Mallinis et al., 2008; Townshend et al., 1991) and vegetation species (Adam et al., 2010; Carpenter et al., 1999; Hirano et al., 2003, Li et al., 2005; Xu and Gong, 2007). However, currently, most LiDAR sensors can only shoot and receive energy at a single band (usually the near-infrared band). Although dual-band LiDAR has been developed recently (such as the Optech Titan equipped with three bands), there is nearly no practical use of this kind of sensor currently due to the extremely high cost of these sensors. Moreover, even with the intensity (the spectral information recorded in LiDAR

data) information of the single-band LiDAR, it cannot be used for direct comparison. Due to energy attenuation, multi-path effects and the influence of different incident angles, the intensity information for the same types of land surface objects obtained from different times, flight lines, or locations can be significantly different (Boyd and Hill, 2007; Hasega'wa, 2006; Mazzarini et al., 2002; Song et al., 2002; Wang et al., 2009). Although there have been algorithms developed trying to calibrate the LiDAR intensity information (Jutzi and Gross, 2009; Kaasalainen et al., 2009; Kaasalainen et al., 2011), these methods are usually data specific or sensor specific and the calibrated results are still problematic.

Second, the availability of LiDAR is currently limited. Based on the platform that LiDAR sensors are mounted on, LiDAR can be divided into four categories, terrestrial LiDAR, near surface LiDAR, airborne LiDAR and spaceborne LiDAR. Terrestrial LiDAR can be used to collect LiDAR data with extreme high point density and accuracy (Dassot et al., 2011; Tao et al., 2015; Li et al., 2016). However, like traditional field measurements, collecting terrestrial LiDAR data is highly time-consuming and labor-intensive, rendering it impractical for large-scale studies. Near surface LiDAR, characterized by the UAV (unmanned aerial vehicle) LiDAR system, has become more popular recently due to its increased flexibility and low cost. Many research-based or commercial UAV LiDAR systems have been developed during the last decade (Chisholm et al., 2013; Guo et al., 2017; Lin et al., 2011; Nagai et al., 2009; Wallace et al., 2012). However, UAV-based LiDAR systems can only cover small areas (<10 km^2), and can hardly be used in large-scale studies (Guo et al., 2017). Compared with terrestrial and UAV LiDAR systems, airborne LiDAR can collect data with higher efficiency and enough density to derive 3D forest structures, which makes it one of the most frequently used platforms in regional-scale studies. However, currently, airborne LiDAR data are only available in certain areas of the world due to the high flight mission cost. It is estimated by the U.S. Forest Service that the current price for collecting airborne LiDAR data with a point density of 5 pts/m^2 is around \$2.5 per acre (Hummel et al., 2011), and this price will increase significantly with the increase of the required point density (Jakubowski et al., 2013). This significantly constrains the application of airborne LiDAR data in large-scale (e.g., national- to global-scale) studies. The Geoscience Laser Altimeter System (GLAS) onboard the Ice, Cloud, and land Elevation (ICESat) satellite, as the only available spaceborne LiDAR data currently, can provide global-scale LiDAR measurements with much lower cost. However, its ellipsoidal ground footprints, which are ~65 m in diameter, are spaced at 170 m along the track and at tens of kilometers across tracks (Schutz et al., 2005), which are too sparse to be used directly to generate spatial continuous forest structure estimations in large scale.

These two limitations of LiDAR have largely constrained its application in mapping vegetation types and estimating large-scale forest 3D structures. How to overcome these limitations of LiDAR and use LiDAR to improve the accuracy of vegetation mapping and large-scale forest structure estimations is an important area of research.

1.4 Study objectives

In summary, optical passive remote sensing and radar data have the spectral information and spatial coverage, and LiDAR data have the capability to map 3D forest structures accurately. By taking advantages each type of remotely sensed data, it is possible to solve

the problems of each data type and generate reliable vegetation maps and forest structure products at multiple scales (Baccini et al., 2008; Boudreau et al., 2008; Saatchi et al., 2011; Simard et al., 2011). Therefore, this dissertation aims to answer the following two questions. 1) What are the advantages of LiDAR in estimating forest structures and monitoring forest changes? 2) How to fuse LiDAR data with other remotely sensed datasets to overcome the limitations of LiDAR for different applications at different scales? Specifically, this dissertation has four objectives.

1) Examine the effectiveness of LiDAR for mapping forest structure and detecting forest changes at a local scale. We compared the capability of LiDAR and aerial imagery for detecting forest fuel treatment activities. Moreover, a novel algorithm was proposed to detect and quantify forest fuel treatments directly from bi-temporal LiDAR point cloud data.

2) Develop an algorithm to map vegetation types through the fusion of LiDAR data and optical imagery at a local scale. We proposed a strategy to map forest vegetation composition using both multispectral aerial imagery and airborne LiDAR data. This approach included the use of a Bayesian information criterion algorithm to determine the optimized number of vegetation groups and an unsupervised classification technique and post hoc analysis to map these vegetation groups.

3) Develop a data fusion algorithm to map fine-resolution forest structures at regional scales. In this study, we used forest tree height as an example. The fine-resolution forest tree height distribution in the Sierra Nevada was estimated through a method using field measurements, airborne LiDAR data, spaceborne LiDAR data, optical imagery and radar data.

4) Develop an algorithm to map forest structures at national to global scales. We focused on forest AGB for this study. A data fusion algorithm was proposed to map national- to global-scale forest AGB from field measurements, spaceborne LiDAR data, optical imagery, radar data and climate surfaces. An uncertainty field model was introduced in the algorithm to minimize the influence of the plot locality uncertainty. This algorithm was first tested at the national scale to map the forest AGB distribution across China and then applied to map the global-scale forest AGB distribution.

References

Adam, E., Mutanga, O., & Rugege, D. (2010). Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review. Wetlands Ecology and Management, 18, 281-296.

Andersen, H.-E., McGaughey, R.J., & Reutebuch, S.E. (2005). Estimating forest canopy fuel parameters using LIDAR data. Remote Sensing of Environment, 94, 441-449.

Aplin, P. (2004). Remote sensing: land cover. Progress in physical geography, 28, 283-293.

Baccini, A., Laporte, N., Goetz, S., Sun, M., & Dong, H. (2008). A first map of tropical Africa's above-ground biomass derived from satellite imagery. Environmental Research Letters, 3, 045011.

Bales, R.C., Battles, J.J., Chen, Y., Conklin, M.H., Holst, E., O'Hara, K.L., Saksa, P., & Stewart, W. (2011). Forests and water in the Sierra Nevada: Sierra Nevada watershed ecosystem enhancement project. In, Sierra Nevada Research Institute report. Merced, California, USA.: Sierra Nevada Research Institute.

Bortolot, Z.J., & Wynne, R.H. (2005). Estimating forest biomass using small footprint LiDAR data: An individual tree-based approach that incorporates training data. ISPRS Journal of Photogrammetry and Remote Sensing, 59, 342-360.

Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L., & Kimes, D.S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. Remote Sensing of Environment, 112, 3876-3890.

Boyd, D.S., & Hill, R. (2007). Validation of airborne LiDAR intensity values from a forested landscape using hymap data: preliminary analyses. International Archives of Photogrammetric Engineering and Remote Sensing, 36, 71-76.

Brienen, R.J.W., Phillips, O.L., Feldpausch, T.R., Gloor, E., Baker, T.R., Lloyd, J., Lopez-Gonzalez, G., Monteagudo-Mendoza, A., Malhi, Y., & Lewis, S.L. (2015). Long-term decline of the Amazon carbon sink. Nature, 519, 344-348.

Brivio, P., Colombo, R., Maggi, M., & Tomasoni, R. (2002). Integration of remote sensing data and GIS for accurate mapping of flooded areas. International Journal of Remote Sensing, 23, 429-441.

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

Cartus, O., Santoro, M., & Kellndorfer, J. (2012). Mapping forest aboveground biomass in the Northeastern United States with ALOS PALSAR dual-polarization L-band. Remote Sensing of Environment, 124, 466-478.

Chen, J.M., & Cihlar, J. (1996). Retrieving leaf area index of boreal conifer forests using Landsat TM images. Remote Sensing of Environment, 55, 153-162.

Chisholm, R.A., Cui, J., Lum, S.K.Y., & Chen, B.M. (2013). UAV LiDAR for below-canopy forest surveys. Journal of Unmanned Vehicle Systems, 1, 61-68.

Clark, M.L., & Aide, T.M. (2011). Virtual Interpretation of Earth Web-Interface Tool (VIEW-IT) for Collecting Land-Use/Land-Cover Reference Data. Remote Sensing, 3, 601-620.

Clark, M.L., Clark, D.B., & Roberts, D.A. (2004). Small-footprint LiDAR estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. Remote Sensing of Environment, 91, 68-89.

Dassot, M., Constant, T., & Fournier, M. (2011). The use of terrestrial LiDAR technology in forest science: application fields, benefits and challenges. Annals of Forest Science, 68, 959-974.

Fensholt, R., Sandholt, I., & Rasmussen, M.S. (2004). Evaluation of MODIS LAI, fAPAR and the relation between fAPAR and NDVI in a semi-arid environment using in situ measurements. Remote Sensing of Environment, 91, 490-507.

Franklin, J. (1986). Thematic Mapper analysis of coniferous forest structure and composition. International Journal of Remote Sensing, 7, 1287-1301.

Führer, E. (2000). Forest functions, ecosystem stability and management. Forest Ecology and Management, 132, 29-38.

Gao, F., Morisette, J.T., Wolfe, R.E., Ederer, G., Pedelty, J., Masuoka, E., Myneni, R., Tan, B., & Nightingale, J. (2008). An algorithm to produce temporally and spatially continuous MODIS-LAI time series. IEEE Geoscience and Remote Sensing Letters, 5, 60-64.

Green, K., Kempka, D., & Lackey, L. (1994). Using remote sensing to detect and monitor land-cover and land-use change. Photogrammetric Engineering and Remote Sensing, 60, 331-337.

Guo, Q., Su, Y., Hu, T., Zhao, X., Wu, F., Li, Y., Liu, J., Chen, L., Xu, G., & Lin, G. (2017). An integrated UAV-borne LiDAR system for 3D habitat mapping in three forest ecosystems across China. International Journal of Remote Sensing, 1-19.

Hasega'wa, H. (2006). Evaluations of LiDAR reflectance amplitude sensitivity towards land cover conditions. Bulletin of the Geographical Survey Institute, 53, 43-50.

Hirano, A., Madden, M., & Welch, R. (2003). Hyperspectral image data for mapping wetland vegetation. Wetlands, 23, 436-448.

Huang, W., Sun, G., Ni, W., Zhang, Z., & Dubayah, R. (2015). Sensitivity of multi-source SAR backscatter to changes in forest aboveground biomass. Remote Sensing, 7, 9587-9609.

Hummel, S., Hudak, A., Uebler, E., Falkowski, M., & Megown, K. (2011). A comparison of accuracy and cost of LiDAR versus stand exam data for landscape management on the Malheur National Forest. Journal of Forestry, 109, 267-273.

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

Jensen, J.R. (2009). Remote sensing of the environment: An earth resource perspective (2nd edition). Pearson Education India. Upper Saddle River, NJ.

Jutzi, B., & Gross, H. (2009). Normalization of LiDAR intensity data based on range and surface incidence angle. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 38, 213-218.

Kaasalainen, S., Hyyppa, H., Kukko, A., Litkey, P., Ahokas, E., Hyyppa, J., Lehner, H., Jaakkola, A., Suomalainen, J., & Akujarvi, A. (2009). Radiometric calibration of LIDAR intensity with commercially available reference targets. IEEE Transactions on Geoscience and Remote Sensing, 47, 588-598.

Kaasalainen, S., Jaakkola, A., Kaasalainen, M., Krooks, A., & Kukko, A. (2011). Analysis of incidence angle and distance effects on terrestrial laser scanner intensity: Search for correction methods. Remote Sensing, 3, 2207-2221.

Kasischke, E.S., Melack, J.M., & Dobson, M.C. (1997). The use of imaging radars for ecological applications—a review. Remote Sensing of Environment, 59, 141-156.

Lasch, P., Lindner, M., Erhard, M., Suckow, F., & Wenzel, A. (2002). Regional impact assessment on forest structure and functions under climate change—the Brandenburg case study. Forest Ecology and Management, 162, 73-86.

Le Toan, T., Beaudoin, A., Riom, J., & Guyon, D. (1992). Relating forest biomass to SAR data. IEEE Transactions on Geoscience and Remote Sensing, 30, 403-411.

Li, L., Ustin, S., & Lay, M. (2005). Application of multiple endmember spectral mixture analysis (MESMA) to AVIRIS imagery for coastal salt marsh mapping: a case study in China Camp, CA, USA. International Journal of Remote Sensing, 26, 5193-5207.

Li, Y., Guo, Q., Tao, S., Zheng, G., Zhao, K., Xue, B., & Su, Y. (2016). Derivation, validation, and sensitivity analysis of terrestrial laser scanning-based leaf area index. Canadian Journal of Remote Sensing, 42, 719-729.

Lin, Y., Hyyppa, J., & Jaakkola, A. (2011). Mini-UAV-borne LiDAR for fine-scale mapping. IEEE Geoscience and Remote Sensing Letters, 8, 426-430.

Lu, D. (2005). Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon. International Journal of Remote Sensing, 26, 2509-2525.

Lu, D. (2006). The potential and challenge of remote sensing-based biomass estimation. International Journal of Remote Sensing, 27, 1297-1328.

Luckman, A., Baker, J., Kuplich, T.M., Yanasse, C.d.C.F., & Frery, A.C. (1997). A study of the relationship between radar backscatter and regenerating tropical forest biomass for spaceborne SAR instruments. Remote Sensing of Environment, 60, 1-13.

Mallinis, G., Koutsias, N., Tsakiri-Strati, M., & Karteris, M. (2008). Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a Mediterranean test site. ISPRS Journal of Photogrammetry and Remote Sensing, 63, 237-250.

Mazzarini, F., Pareschi, M.T., Favalli, M., Isola, I., Tarquini, S., & Boschi, E. (2007). Lava flow identification and aging by means of LiDAR intensity: Mount Etna case. Journal of Geophysical Research: Solid Earth, 112, B02201.

Mitchard, E.T., Saatchi, S.S., Woodhouse, I.H., Nangendo, G., Ribeiro, N., Williams, M., Ryan, C.M., Lewis, S.L., Feldpausch, T., & Meir, P. (2009). Using satellite radar backscatter to predict above-ground woody biomass: A consistent relationship across four different African landscapes. Geophysical Research Letters, 36, L23401.

Morsdorf, F., Kotz, B., Meier, E., Itten, K.I., & 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, 50-61.

Myneni, R., Dong, J., Tucker, C., Kaufmann, R., Kauppi, P., Liski, J., Zhou, L., Alexeyev, V., & Hughes, M. (2001). A large carbon sink in the woody biomass of northern forests. Proceedings of the National Academy of Sciences, 98, 14784-14789.

Naesset, E. (1997). Determination of mean tree height of forest stands using airborne laser scanner data. ISPRS Journal of Photogrammetry and Remote Sensing, 52, 49-56.

Nagai, M., Tianen Chen, T., Shibasaki, R., Kumagai, H., & Ahmed, A. (2009). UAV-borne 3D mapping system by multisensor integration. IEEE Transactions on Geoscience and Remote Sensing, 47, 701-708.

Nelson, R., Ranson, K., Sun, G., Kimes, D., Kharuk, V., & Montesano, P. (2009). Estimating Siberian timber volume using MODIS and ICESat/GLAS. Remote Sensing of Environment, 113, 691-701.

Nilsson, M. (1996). Estimation of tree heights and stand volume using an airborne LiDAR system. Remote Sensing of Environment, 56, 1-7.

Popescu, S., Zhao, K., & Neuenschwander, A. (2011). Satellite LiDAR vs. small footprint airborne LiDAR: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. Remote Sensing of Environment, 115, 2786-2797.

Ranson, K., & Sun, G. (1994). Mapping biomass of a northern forest using multi-frequency SAR data. IEEE Transactions on Geoscience and Remote Sensing, 32, 388-396.

Riano, D., Valladares, F., Condes, S., & Chuvieco, E. (2004). Estimation of leaf area index and covered ground from airborne laser scanner (LiDAR) in two contrasting forests. Agricultural and Forest Meteorology, 124, 269-275.

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

Saatchi, S., Halligan, K., Despain, D.G., & Crabtree, R.L. (2007). Estimation of forest fuel load from radar remote sensing. IEEE Transactions on Geoscience and Remote Sensing, 45, 1726-1740.

Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., & Hagen, S. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. Proceedings of the National Academy of Sciences, 108, 9899-9904.

Sader, S.A., Waide, R.B., Lawrence, W.T., & Joyce, A.T. (1989). Tropical forest biomass and successional age class relationships to a vegetation index derived from Landsat TM data. Remote Sensing of Environment, 28, 143IN1159-1156IN2198.

Schutz, B., Zwally, H., Shuman, C., Hancock, D., & DiMarzio, J. (2005). Overview of the ICESat mission. Geophysical Research Letters, 32, L21S01.

Shan, J., & Toth, C.K. (2008). Topographic laser ranging and scanning: principles and processing. CRC press.

Simard, M., Pinto, N., Fisher, J.B., & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne LiDAR. Journal of Geophysical Research: Biogeosciences, 116, G04021.

Song, J.-H., Han, S.-H., Yu, K., & Kim, Y.-I. (2002). Assessing the possibility of land-cover classification using LiDAR intensity data. International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences, 34, 259-262.

Sun, G., Ranson, K., & Kharuk, V. (2002). Radiometric slope correction for forest biomass estimation from SAR data in the Western Sayani Mountains, Siberia. Remote Sensing of Environment, 79, 279-287.

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

Tao, S., Guo, Q., Su, Y., Xu, S., Li, Y., & Wu, F. (2015). A geometric method for wood-leaf separation using terrestrial and simulated LiDAR data. Photogrammetric Engineering & Remote Sensing, 81, 767-776.

Tempel, D.J., Gutiérrez, R., Battles, J.J., Fry, D.L., Su, Y., Guo, Q., Reetz, M.J., Whitmore, S.A., Jones, G.M., & Collins, B.M. (2015). Evaluating short- and long-term impacts of fuels treatments and simulated wildfire on an old-forest species. Ecosphere, 6, 1-18.

Townshend, J., Justice, C., Li, W., Gurney, C., & McManus, J. (1991). Global land cover classification by remote sensing: present capabilities and future possibilities. Remote Sensing of Environment, 35, 243-255.

Wagner, W., Luckman, A., Vietmeier, J., Tansey, K., Balzter, H., Schmullius, C., Davidson, M., Gaveau, D., Gluck, M., & Le Toan, T. (2003). Large-scale mapping of boreal forest in Siberia using ERS tandem coherence and JERS backscatter data. Remote Sensing of Environment, 85, 125-144.

Wallace, L., Lucieer, A., Watson, C., & Turner, D. (2012). Development of a UAV-LiDAR System with Application to Forest Inventory. Remote Sensing, 4, 1519-1543.

Wang, C., & Glenn, N.F. (2009). Integrating LiDAR intensity and elevation data for terrain characterization in a forested area. IEEE Geoscience and Remote Sensing Letters, 6, 463-466.

Wing, B.M., Ritchie, M.W., Boston, K., Cohen, W.B., & Olsen, M.J. (2015). Individual snag detection using neighborhood attribute filtered airborne LiDAR data. Remote Sensing of Environment, 163, 165-179.

Woodhouse, I.H., Mitchard, E.T., Brolly, M., Maniatis, D., & Ryan, C.M. (2012). Radar backscatter is not a direct measure of forest biomass. Nature Climate Change, 2, 556-557.

Xu, B., & Gong, P. (2007). Land-use/land-cover classification with multispectral and hyperspectral EO-1 data. Photogrammetric Engineering & Remote Sensing, 73, 955-965.

Zhang, G., Ganguly, S., Nemani, R.R., White, M.A., Milesi, C., Hashimoto, H., Wang, W., Saatchi, S., Yu, Y., & Myneni, R.B. (2014). Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data. Remote Sensing of Environment, 151, 44-56.

Zhao, K., & Popescu, S. (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, 1628-1645.

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

Chapter 2   Forest fuel treatment detection using multi-temporal airborne LiDAR data and high resolution aerial imagery

Abstract

Treatments to reduce forest fuels are often performed in forests to enhance forest health, regulate stand density, and reduce the risk of wildfires. While commonly employed, there are concerns that these forest fuel treatments (FTs) may have negative impacts on certain wildlife species. Often FTs are planned across large landscapes, but the actual treatment extents can differ from planned extents due to operational constraints and protection of resources (e.g., perennial streams, cultural resources, wildlife habitats). Identifying the actual extent of treated areas is of primary importance to understand the environmental influence of FTs. Light detection and ranging (LiDAR) is a powerful remote sensing tool that can provide accurate measurements of forest structure and has great potential for monitoring forest changes. This study used canopy height model (CHM) and canopy cover (CC) products derived from multi-temporal airborne laser scanning (ALS) data to monitor forest changes following the implementation of landscape-scale FT projects. Our approach involved the combination of a pixel-wise thresholding method and an object-of-interest segmentation method. We also investigated forest change through the use of normalized difference vegetation index and standardized principal component analysis from multi-temporal high resolution aerial imagery. The same FT detection routine was then applied to compare the capability of ALS data and aerial imagery for FT detection. Our results demonstrate that the FT detection using ALS derived CC products produced both the highest total accuracy (93.5%) and kappa coefficient (κ) (0.70), and was more robust at identifying areas with light FTs. The accuracy using ALS derived CHM products (the total accuracy was 91.6%, and the κ was 0.59) was significantly lower than that of the result using ALS derived CC, but was still higher than using aerial imagery. Moreover, we also developed and tested a method to recognize the intensity of FTs directly from pre- and post-treatment ALS point clouds.

Keywords: forest fuel treatment; remote sensing; change detection; LiDAR; aerial imagery

2.1 Introduction

Forests of the Sierra Nevada mountain range in California, USA are extensive, but face increasing risk of wildfire, largely due to fire suppression, fuel buildup, and changes in climate (Stephens et al., 2010). In 2013 alone, the American Fire and the Rim Fire, burned over 111 km^2 of the Tahoe National Forest (located at the northern Sierra Nevada), and over 1,041 km^2 of forest in the Stanislaus National Forest and Yosemite National Park (located at the southern Sierra Nevada), respectively. To reduce the loss from uncharacteristically large and high severity wildfire, the USDA (United States Department of Agriculture) Forest Service (referred as USFS hereafter) and other land management agencies conduct mechanical forest fuel treatments (FT) on areas with high fire risk in the Sierra Nevada. Through the forest thinning and mastication, these mechanical FT activities aim to modify wildland fire behaviors and minimize negative impacts on the forest health

and enhance forest fire suppression capabilities (Collins et al., 2010). However, there have been concerns over potential negative impacts of these fire mitigation treatments on the habitat of protected animals, particularly the California spotted owl (*Strix occidentalis occidentalis*) and the Pacific fisher (*Martes pennant*).

Knowing the actual (as opposed to planned) extent of FTs (note that FTs referred hereafter are mechanical FT activities) is critical for understanding how FTs affect the wildfire risk, wildlife and forest health. The method for reporting completed FTs, which is still being used currently, involves using planned treatment boundaries and then updating them based on field observations. Planned FT boundaries are often geographically distinct from the actual extents due to operational constraints and the protection of resources (e.g., perennial streams, cultural resources, wildlife habitat). Thus, mapping the actual FT extent relies heavily on field observations, which are very labor-intensive and time-consuming. Moreover, currently, the intensity of FTs is commonly reported by FT types (e.g., low, crown, or selection thinning) (Agee and Skinner, 2005) or harvesting methods (e.g., mastication, thinning, or cable thinning). Fundamentally, there remains a lack of accurate methods to quantitatively evaluate both the extent and intensity of FTs.

Remote sensing, which can expand both spatial and temporal scales of land surface observations, provides an alternative way to detect forest structure changes due to FTs. Change detection, defined as the process of identifying differences in the state of an object or a phenomenon by observing it at different times (Singh, 1989), is the most commonly used method to identify land surface changes from multi-temporal passive remote sensing data. Generally, change detection can be divided into three groups: image algebra, transformation, and classification methods (Lu et al., 2004; Singh, 1989). Image algebra methods apply algebraic calculus (e.g., image differencing, regression, ratioing) on multi-temporal remotely sensed data at the pixel level to obtain a change image (Allen and Kupfer, 2000; Lunetta et al., 2006; Patra et al., 2011; Prakash and Gupta, 1998; Sohl, 1999). Transformation methods use statistical procedures [e.g., Principal component analysis (PCA), Kauth-Thomas (KT) transformation, Gramm-Schmidt (GS) transformation, and Chi-square transformation] to minimize the redundant information in multi-temporal data and derive change images (Collins and Woodcock, 1994, 1996; Li and Yeh, 1998; Qin et al., 2013; Ridd and Liu, 1998; Seto e al., 2002). Both the algebra and transformation methods have to determine thresholds to differentiate the changed and unchanged areas (Guerra et al., 1998; Yousif and Ban, 2013). Classification methods identify change areas by comparing the image classification results from multi-temporal remote sensing data (Desclée et al., 2006; Hame et al., 1998; Hao et al., 2014; Li and Xu, 2010; Walter, 2004). However, few studies have mapped FT areas using passive remote sensing data because of the limited ability of these sensors for penetrating the forest canopy (Weishampel et al., 2000). Because FTs typically remove smaller and sub-dominant trees to reduce fuel laddering and use mastication of shrubs to remove surface fuels, the changes in forest surface spectral characteristic may be too slight to be identified with passive remote sensing.

Light Detection and Ranging (LiDAR), an active remote sensing technique, uses a focused short wavelength laser pulse, which can penetrate the forest canopy more effectively. It has been proven that LiDAR can be used to accurately estimate forest parameters, such as tree height (Clark et al., 2004; Hudak et al, 2002; Zimble et al., 2003), leaf area index (LAI) (Morsdorf et al., 2006; Riano et al., 2004; Richardson et al., 2009;

Zhao and Popescu, 2009), and aboveground biomass (Bortolot and Wynne, 2005; Popescu et al., 2011; Zhao et al., 2009). Clark et al. (2004) used the airborne laser scanning (ALS) data to accurately estimate sub-canopy elevation and tree height in a tropical rain forest landscape. Korhonen et al. (2011) found that ALS discrete data can be used to obtain comparable estimations of forest vertical canopy cover, angular canopy closure, and LAI. Andersen et al. (2005) estimated the canopy fuel parameters (e.g. crown fuel weight, crown bulk density, canopy height, and canopy base height) using regression analysis between ALS data derived metrics and field measurements. Kramer et al. (2014) found that the forest percentage cover between 2 and 4 m derived from ALS data is a good indicator of ladder fuels, which are an important forest structural attribute contributing to wildfire hazard. Kane et al. (2013 and 2014) evaluated the effects of wildfire on forest spatial structure at different height strata through the fusion of optical imagery and ALS data, and found that three forest spatial structures (i.e., canopy gap, clump-open, and open) were associated with the fire severity.

Accurate estimations of these forest parameters from ALS data are highly desirable for understanding the composition and structure information of forests and therefore monitoring forest changes (Dubayah et al., 2010). For example, Yu et al. (2004 and 2006) estimated tree growths and monitored harvested trees using canopy height models (CHM) derived from multi-temporal ALS data in a conifer forest. However, to the best of our knowledge, there have been few studies conducted on using multi-temporal ALS to detect FT extents. Moreover, research comparing the capability of ALS data and aerial images for detecting FT extents is still lacking.

The main objective of this paper is to study the capability of ALS data for identifying FTs using change detection routines that combine the pixel-wise thresholding (i.e., algebra or transformation methods) and the object-of-interest (OBI) segmentation (i.e., classification methods). As a comparison, the same change detection routine is also applied on multi-temporal high resolution aerial imagery. Additionally, this study aims to investigate the possibility of quantitatively mapping the FT intensity directly from multi-temporal ALS data.

2.2 Data
2.2.1 Study area

The Last Chance study site (39°07'N, 120°35'W) covers 92.1 km^2 of the Tahoe National Forest, California, USA (Figure 2-1). It is on the southwest side of the Sierra Nevada crest. The elevation ranges from 228 m to 2189 m, and the slope ranges from 0° to near 90°. Over 90% of the study area is covered by vegetation, and the average canopy cover is around 67%. The study site is covered by mixed conifer forests, which are dominated by white fir (*Abies concolor*), ponderosa pine (*Pinus ponderosa*), incense-cedar (*Calocedrus decurrens*), sugar pine (*Pinus lambertiana*), and Douglas-fir (*Pseudotsuga menziesii*) (Su et al., 2016). Within the mixed conifer stands, the major hardwoods are black oak (*Quercus kelloggii*) and canyon live oak (*Quercus chrysolepis*) (Su et al., 2016). The FTs detected in the study area were implemented between 2008 and 2012. The main FTs conducted in this area were mechanical forest thinning, which aimed to reduce the ladder fuels by removing small to mid-sized tree from low and intermediate canopy strata within the treatment areas.

It is important to note these treatments intended to leave larger overstory trees intact and have post-treatment canopy cover between 40% and 50% (Collins et al. 2011).
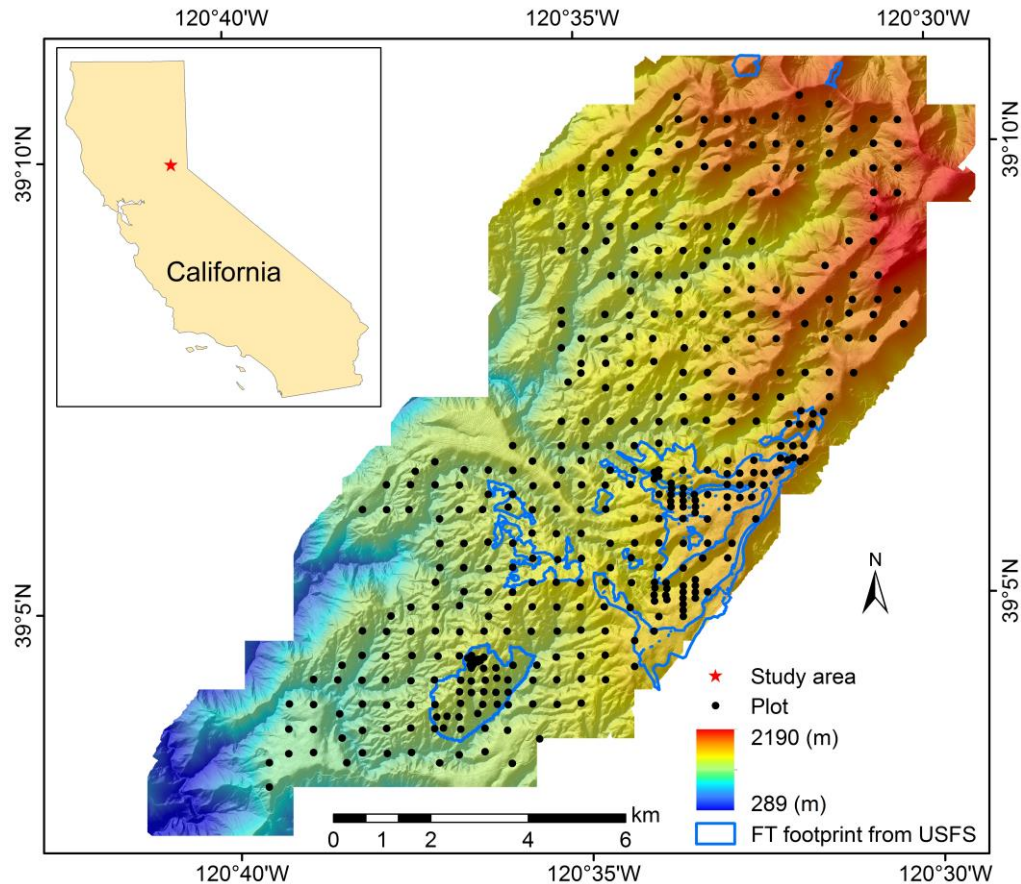
Figure 2-1 The Last Chance study area with the distribution of field plots, terrain elevations, and the proposed forest fuel treatment (FT) footprints from the United States Forest Services (USFS).

2.2.2 Field measurements

408 plots (12.62 m in radius and 500 m^2 in area) were selected in the study site (Figure 2-1). The first plot was randomly chosen and the following plots were placed on 500 m $\times$ 500 m grids. Within specific watersheds, the sampling was intensified to 250 m $\times$ 250 m. When locating each plot by using a Trimble^{TM} GeoXH GPS in the field, the plot center had to be farther than 12.62 m away from any landing or road surfaces. If the plot needed to be offset, we randomly chose one of the four cardinal directions and moved the plot 25 m in that direction. In each plot, the canopy cover (CC) measurement and the treatment condition were collected from the field. The CC was measured by a sight tube on a 5 m by 5 m regular grid centered on the plot center (total of 25 points per plot). These measurements were made twice, first in the summer of 2007 (prior to treatment implementation), and again in 2013 (after treatment implementation). It should be noted that 39 plots were eliminated in the analysis of this study because their treatment conditions

were not collected in the field. Among the remaining 369 plots, 46 plots were recognized as treated. Besides these plot measurements, polygons of the intended treatment areas (created by the USFS) were also used in this study to evaluate the FT detection accuracy (Figure 2-1).

2.2.3 Small footprint ALS data

The pre-treatment ALS data used in this study were acquired in September, 2007, and the post-treatment ALS data were acquired in November, 2012 and August, 2013 (the 2012 flight was stopped by snow). An Optech GEMINI airborne laser terrain mapper (ALTM) from National Center of Airborne Laser Mapping at the University of Houston was used for both the pre- and post-treatment ALS data acquisition. The sensor was operated at 100 kHz with a scanning frequency of 40–60 Hz and a scan angle of 12–14° on either side of nadir. It was mounted on a twin-engine Cessna Skymaster which flew at 600–800 m above the ground. The average swath width of a single pass was around 510 m with over 50% overlap between two adjacent flight lines, and the obtained point density was 9.6 points/m^2 on average. The ALS footprint size is about 15-20 cm in diameter. To make sure the pre- and post-treatment ALS flights aligned together, over 800 ground check points, positioned by ground GPS, were set to calibrate and assess the vertical and horizontal accuracy of the ALS flights. The obtained horizontal accuracy was around 10 cm and the vertical accuracy was from 5 to 35 cm.

Pre-and post-treatment CHM and CC products were calculated from ALS point clouds to detect FTs. To derive CHM products, digital surface model (DSM) and digital elevation model (DEM) products were interpolated from the first returns and ground returns, respectively. The ground returns were identified using the software Terrascan following the standard industrial procedure. The interpolation scheme used was the ordinary kriging algorithm, because it has been proven to be more accurate than other interpolation methods (such as inverse distance weighted and spline) for interpolating DEM and DSM from ALS-derived elevation points (Clark et al., 2004; Lloyd and Atkinson, 2002; Guo et al, 2010). The CHM was calculated by the differences between the DSM and the DEM. Note that CHMs and the corresponding DEMs and DSMs were produced in two different resolutions (1 m and 20 m). The 1 m resolution CHMs were used to generate CC products (in 20 m resolution) using a CHM-based method that showed a good correspondence with field measurements (Lucas et al., 2006). Each of 1 m × 1 m CHM cell with a value above a selected height threshold was coded as 1. A height threshold of 2 m was selected in this study to generate equivalent CC estimations from ALS data with the field measurements. The CC was calculated by the percentage of the number of coded CHM cells within in each 20 m × 20 m grid. The accuracy of the ALS derived pre- and post-treatment CC products were evaluated by the pre- and post-treatment field measurements, respectively.

The obtained 20 m resolution CHMs and CCs were then used in the following FT detection procedures. The differences between the pre- and post-treatment CHMs and CCs were used to represent changes in forest structure, respectively. Because mechanical forest thinning, the main FT type conducted in the study area, usually incorporated the removal of vegetation to reduce the forest fuel load, the CHM and CC were expected to have a drop in treated areas compared to control areas.

2.2.4 High resolution aerial imagery

Aerial imagery (1 m resolution) from the National Agriculture Imagery Program (NAIP) was used in this study for the purpose of comparing the capability of ALS data on FT detection with that of traditional passive remote sensing data. The 2005 color-infrared (CIR) imagery (composed of green band, red band and near-infrared (NIR) band) was used to represent the pre-treatment forest, and the 2012 CIR imagery was selected to represent the post-treatment forest. Both the pre- and post-treatment NAIP data were resampled to a resolution of 20 m using the weighted mean value method to get comparable results with field measurements. The misalignment of digital number (DN) values for time-invariant objects is one of the major error sources for applying change detection techniques on multi-temporal aerial images (Canty and Nielsen, 2008; Lu et al., 2004; Singh, 1989), and therefore radiometric normalization is one prerequisite for performing change detection algorithms on multi-temporal aerial images. To comprehensively compare the FT accuracy from aerial imagery with that from ALS data, two different techniques (NDVI differencing, and standardized PCA) were applied to obtain change images from multi-temporal aerial imagery.

2.2.4.1 Radiometric normalization

We used an automatic iteratively re-weighted multivariate alteration detection (IR-MAD) transformation method developed by Canty and Nielsen (2008) to determine the time-invariant pixels. These time-invariant pixels were then used to normalize the pre- and post-treatment aerial images. The IR-MAD method can calculate the no-change probability of pixels by transforming multi-temporal images into a set of mutually orthogonal difference images (Nielsen, 2007). Only the pixels with no-change probability above the selected threshold (0.95 is a commonly used threshold) were selected as time-invariant pixels. In this study, the pre-treatment aerial image was used as the reference image in the radiometric normalization procedure, and the post-treatment aerial image was used as the target image that needed to be normalized. There were overall 1,071 time-invariant pixels selected. Two-thirds of them (714 pixels) were used as the training pixels to normalize the aerial images, and the remaining 357 pixels were saved to be used as testing pixels. It should be noted that all the following procedures related to aerial images were processed based on the radiometrically normalized aerial images.

2.2.4.2 NDVI differencing method

NDVI is one of most robust and frequently used vegetation indexes in monitoring vegetation status and estimating vegetation parameters (Anyamba and Tucker, 2005; Carlson and Ripley, 1997; Carreiras et al., 2006; Pettorelli et al., 2005), which is defined as,

$$\text{NDVI} = \frac{(\text{DN})_{\text{NIR}} - (\text{DN})_{\text{R}}}{(\text{DN})_{\text{NIR}} + (\text{DN})_{\text{R}}} \tag{2-1}$$

where DN_{NIR} and DN_R are the DN values in the NIR band and red band. In this study, the pre- and post-treatment NDVI were calculated from the normalized aerial images, and the change image was represented by the difference between the pre- and post-treatment NDVI.

2.2.4.3 Standardized PCA method

Standardized PCA, one type of PCA algorithm, requires all input parameters to be standardized before performing PCA algorithm. Eastman and Filk (1993) and Mas (1999) found that it is better for statistical control than non-standardized PCA method and can improve the accuracy of change detection. Thus, in this study, all the six bands from normalized pre- and post-treatment aerial images were first standardized by the following equation,

$$DN_{std} = \frac{(DN)_{norm} - \overline{x}}{\sigma} \quad (2\text{-}2)$$

where DN_{std} is the standardized DN values, DN_{norm} is the original normalized DN values, and $\overline{x}$ and σ are the mean and standard deviation of the corresponding band. Then, both the standardized pre- and post-treatment aerial images were used as the input for PCA transformations. In the standardized PCA routine outputs, the component having the poorest correlation with other components usually highlights the changed information, and therefore was selected as the change image. As shown in Table 2-1, in each row (or column), the R^2 for the relationship between the component represented by that row (or column) and component 4 is the smallest when related to all other components. Therefore, component 4 was selected to be used as the change image for the standardized PCA method in the following FT detection routine.

Table 2-1 Correlations (represented by the R^2) between six components obtained from the standardized principal component analysis (PCA) for normalized pre- and post-treatment aerial images.

	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6
Comp. 1	1.00	0.85	0.93	0.20	0.34	0.36
Comp. 2	0.85	1.00	0.98	0.07	0.64	0.62
Comp. 3	0.93	0.98	1.00	0.14	0.55	0.56
Comp. 4	0.20	0.07	0.14	1.00	0.07	0.27
Comp. 5	0.34	0.64	0.55	0.07	1.00	0.97
Comp.6	0.36	0.62	0.56	0.27	0.97	1.00

2.3 FT detection methods

2.3.1 FT detection strategy

The same FT detection routine, combining pixel-wise thresholding and OBI segmentation, was applied to above mentioned change images (Figure 2-2). An unfiltered pixel-wise FT result was first obtained by using a threshold to differentiate treated and untreated pixels in the change image. In this study, we assumed that the change parameter should be normally distributed, and the variation within the 95% confidence should be recognized as the background information. Thus, μ +/- 1.96σ was used as the threshold to differentiate

the treated and untreated pixels, where μ and σ are the mean and standard deviation of the change image, respectively. It should be noted that the threshold was only selected at one side (either $\mu + 1.96\sigma$ or $\mu - 1.96\sigma$) because FTs only remove trees and can only result in unidirectional movement of selected change parameters.

However, we anticipated that there would be still noise remaining in the unfiltered pixel-wise results. To further remove the noise, the OBI segmentation method was used to filter the pixel-wise result considering the fact that FTs were usually conducted in spatially continuous areas (Zhang et al., 2013). The OBI segmentation was conducted using the "Segment Only Feature Extraction Workflow" module in ENVI software. The mean of a change image within each segmented polygon was then used to determine its treatment attribute. If its value exceeded the threshold ($\mu +/- 1.96\sigma$) used in the pixel-wise thresholding procedure, this polygon was recognized as treated, and *vice versa*. The identified isolated treated polygons with an area smaller than 800 m^2 were further removed in the OBI segmentation workflow. Finally, the detected unfiltered FT pixels within the retained treated polygons were recognized as the final pixel-wise FT extents.
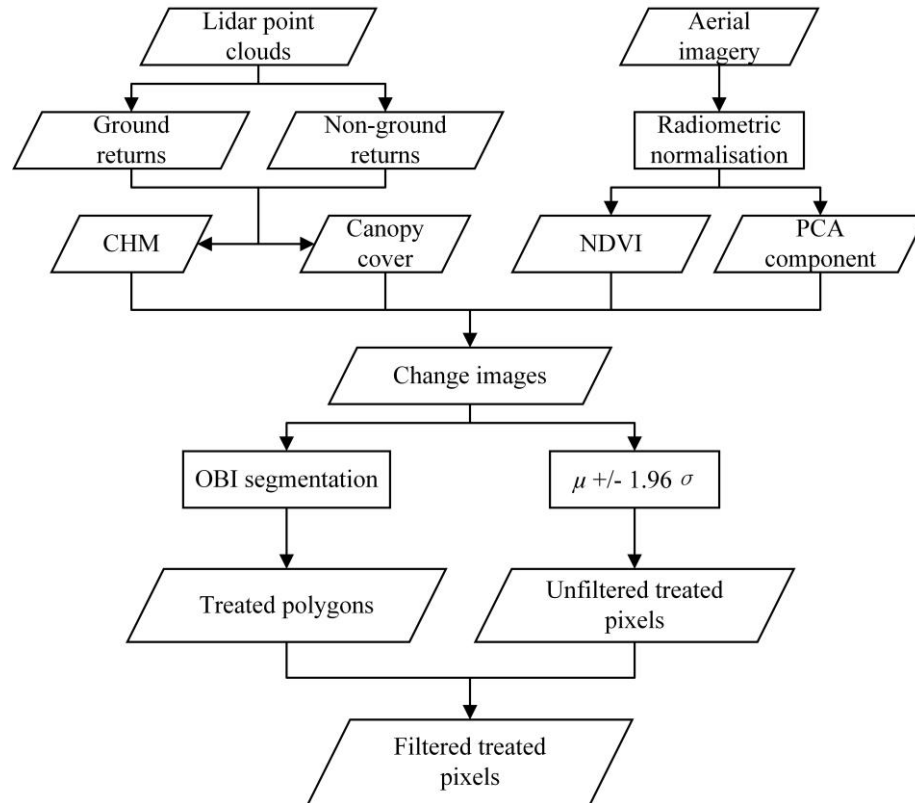
Figure 2-2 The procedure for forest treatment detection that combines a pixel-wise threshold control method and an object-of-interest segmentation method. Note that μ and σ are the mean and standard deviation of the change image.

2.3.2 FT intensity recognition

Within the detected FT extents from the method producing the highest accuracy, we further developed and tested a method to identify the FT intensity directly from multi-temporal ALS point clouds. We hypothesized that FT activities can change the vertical structure of forests and therefore influence the vertical distribution of ALS points. Therefore, we chose

the vertical profile area change between the pre- and post-treatment ALS points within a pixel to represent the FT intensity. The profile area change can be calculated from pre- and post-treatment raw ALS point clouds using the following procedure. 1) Normalize the height of raw ALS points using DEM. 2) Normalize the height of ALS points from Step 1 to 0-1 within each 20 m by 20 m cell. This aimed to reduce the influence of different vegetation heights on the FT intensity recognition result. 3) Sort all ALS points from Step 2 within each 20 m × 20 m pixel ascendingly by height, and draw a height percentile ranking profile for each pixel (Figure 2-3). 4) Compute the area composed by the percentile ranking profile and *x*-axis (i.e. the axis labeled as Point Count in Figure 2-3) for each 20 m × 20 m cell. 5) Calculate the profile area change as the difference between the profile areas of post- and pre-treatment ALS point clouds (i.e., the highlighted gray zones in Figure 2-3).
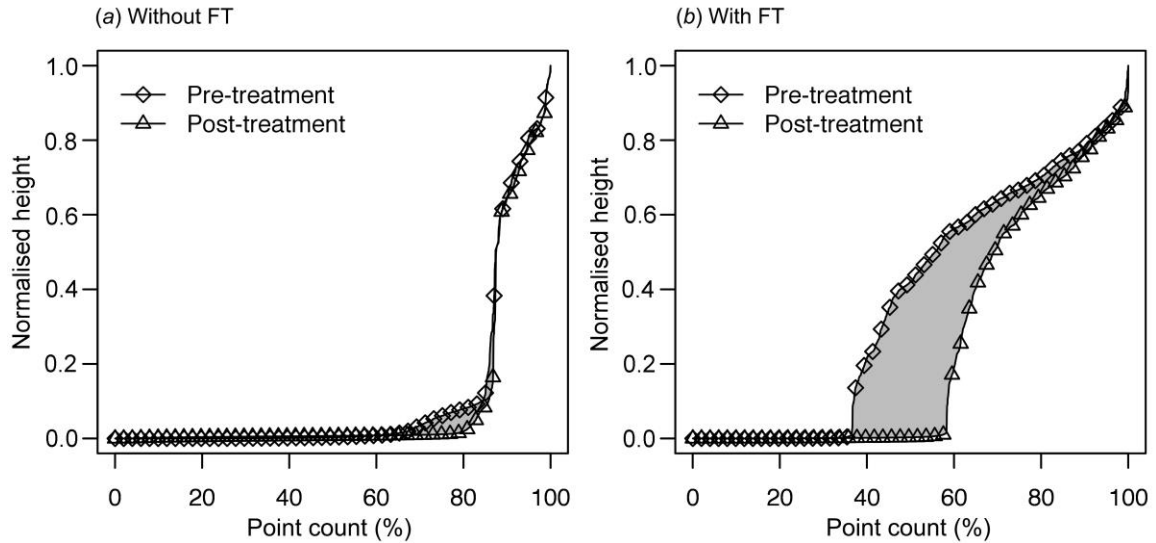
Figure 2-3 Examples of pre- and post-treatment percentile ranking profiles (a) in a 20 by 20 m pixel that has not been treated and (b) in a 20 by 20 m pixel that has been treated.

The profile area change can range from -100% to 100%, theoretically. When a pixel contains no trees in the pre-treatment condition but the post-treatment condition has dense enough trees that prevents the ALS signal from penetrating the forest canopy, its profile area change should be -100%. Conversely, if a pixel contains dense trees in the pre-treatment condition but all are cut down by FT activities, then the profile area change should be 100%. If there are no FT activities in a pixel, the corresponding profile area should be very similar and the profile area change should be close to 0%. Figure 2-3 shows two typical examples (i.e., with and without FT) in the study area. As shown in Figure 2-3(*a*), for a pixel without FT, the pre- and post-treatment profiles are almost identical, and the slight tree growth during the time span of the two ALS flights results in a small negative profile area change (-0.95% for Figure 2-3*a*). In a pixel with FT (Figure 2-3*b*), the number of post-treatment ALS points in upper height levels is significantly increased and the number in lower height levels can be decreased. These changes can lead to relatively large positive profile area change (11.94% for Figure 2-3*b*), and the higher the profile area change is, the higher the FT intensity should be.

2.3.3 Accuracy assessment

Polygons from the USFS that illustrate the intended FT boundary were used to compare with FT detection results of different methods. The proportions of the detected FT areas within the USFS treatment footprints were calculated for all FT detection results. Additionally, plot measurements were also used to evaluate the FT detection accuracy by the total accuracy (τ) and kappa coefficient (κ), which can be denoted as,

$$\tau = \frac{a}{N} \tag{2-3}$$

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{2-4}$$

where a is the number of plots agreeing with the FT detection result, and N is the total number of plots; p_o is the relative observed agreement, and p_e is the hypothetical probability of chance agreement (Jenson, 2005). It should be noted that a 20 m radius buffer was made for each plot to compare with the FT detection result considering the influence of plot size and the mis-registration between the plots and ALS data and aerial images. If the plot was labeled as treated in the field, it would be recognized as true positive (correctly identified by FT detection results) when any part of its buffer overlapped with a treated pixel from FT detection results; otherwise, it would be recognized as false positive (incorrectly identified by FT detection results). If the plot was labeled as untreated in the field, it would be recognized as true negative (correctly rejected by FT detection results) when its buffer was isolated from treated pixels from FT detection results; otherwise, it would be recognized as false negative (incorrectly rejected by FT results).

2.4 Results

2.4.1 ALS derived CC

Both the pre- and post-treatment CC were positively correlated with the field measurements (Figure 2-4). The coefficients of determination (R^2) for the correlations between both the pre- and post-treatment ALS derived CC and field measurements are above 0.5. The R^2 for post-treatment correlation is slightly higher than that for the pre-treatment correlation. The mean difference between the pre-treatment ALS derived CC and field measurements is 13.7%, and that for the post-treatment is 5.5%. The absolute value of coefficient of variation of the root mean square error (CV(RMSE)) for the pre-treatment is slightly lower than the post-treatment. Figure 2-5 shows the change in CC observed from the field measurements. In general, the study area has an 3% increase in CC. However, the changes area widely distributed from -60% to 80%.

    Both the pre- and post-treatment ALS derived CC are systematically higher than field measurements. Over 83% and 63% of the pre- and post-treatment ALS derived CC are higher than the pre- and post-treatment field measurements, respectively. This result is expected since the sight tube can recognize finer canopy gap than the CHM method used in this study. Besides, the ALS derived CC may be more consistent than field measurements. FTs usually involve the selective removal of trees, which can result in a decrease in CC. However, based on field measurements, there are 12 out of 46 treated plots with an increase of in CC. 11 of these plots showed a drop in CC from ALS data, and only

one has an increase in CC (Table 2-2). This phenomenon may have resulted from the insufficient sampling points to measure the CC in the field. Studies have found that the CC obtained using a sight tube with 23 points had a 5.5% underestimation compared with the CC using 195 points (Korhonen et al. 2006), and may be insufficient to capture the variation in forest canopy (Fiala et al., 2006).
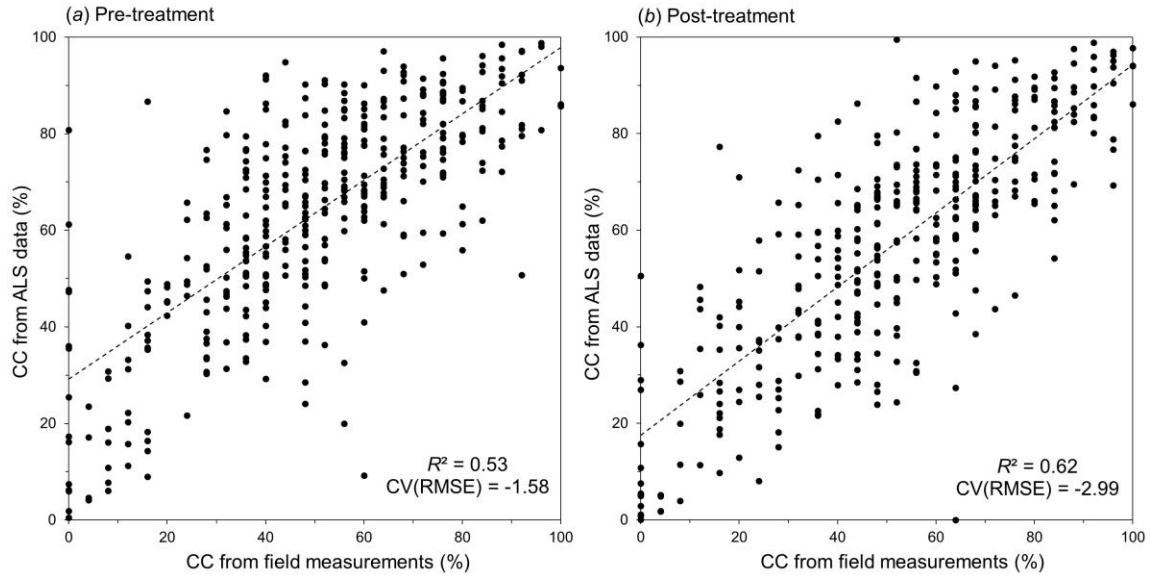
Figure 2-4 Comparison between the ALS derived canopy cover (CC) and field measured CC for both (a) pre-treatment and (b) post-treatment. R^2 represents the coefficient of determination, and CV(RMSE) represents the coefficient of variation of the root mean square error.
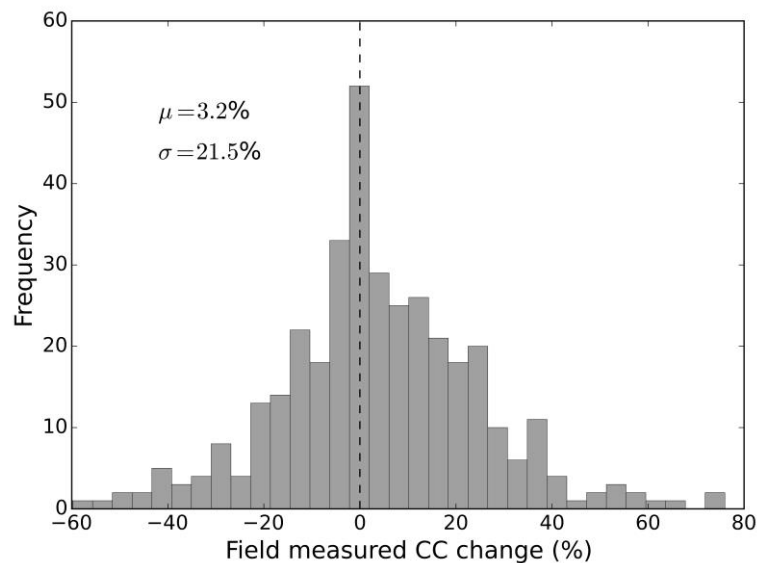
Figure 2-5 Histogram of changes between pre- and post-treatment field measured CC (post-treatment CC minus pre-treatment CC). μ represents the mean difference, and σ represents the standard deviation.

Table 2-2 The comparison between the field measured canopy cover (CC) and ALS derived CC for plots which were identified as treated areas in the field but with an increase in field measured CC.

Plot ID	Field measured CC (%)			ALS derived CC (%)		
	Pre-treatment	Post-treatment	Post-Pre*	Pre-treatment	Post-treatment	Post-Pre*
109	48	56	8	65.1	53.7	-11.4
131	32	68	36	84.7	81.7	-3.0
216	60	64	4	51.6	62.6	11.0
224	48	56	8	51.0	30.5	-20.5
233	52	60	8	78.1	61.2	-16.9
242	56	60	4	68.3	55.0	-13.2
266	24	32	8	62.2	54.6	-7.6
281	28	32	4	33.7	29.9	-3.8
283	20	36	16	45.3	40.67	-4.7
322	36	48	12	72.6	54.8	-17.8
337	40	52	12	71.3	57.9	-13.4
354	52	64	12	57.0	53.7	-3.3

*"Post-Pre" means using the post-treatment CC minus the pre-treatment CC.

2.4.2 Radiometric normalization for aerial images

The built radiometric normalization equation for each band using IR-MAD method is shown in Figure 2-6. The R^2 for green and red bands are both higher than 0.9, and that for the NIR band is slightly lower. The accuracy assessment by the 357 time-invariant testing pixels is shown in Table 2-3. The differences of means and variances between the normalized post- and pre-treatment aerial images are 98% and 70% smaller than those between the original post- and pre-treatment aerial images on average. Moreover, the p-values for the test of equal means and variances between the pre-treatment aerial image and normalized post-treatment aerial image are higher than the significance level of $\alpha=0.05$, which indicates that means and variances between the pre-treatment aerial image and normalized post-treatment aerial image have no significant differences.

Table 2-3 Comparison of means and variances for the 357 time-invariant testing pixels of pre-treatment, post-treatment and normalized post-treatment aerial images. Paired t-test and F-test were used to test equal means and variances between the pre-treatment aerial image and the normalized post-treatment aerial image.

Variables	Green band	Red band	NIR band
Pre-treatment mean	130.6	121.3	158.9
Post-treatment mean	87.6	69.4	123.9
Normalized post-treatment mean	130.1	120.8	159.8
p-value (paired t-test) for means	0.15	0.12	0.11
Pre-treatment variance	211.0	311.1	304.0
Post-treatment variance	290.3	510.9	293.9
Normalized post-treatment variance	182.9	288.6	339.1
p-value (F-test) for variances	0.18	0.48	0.30
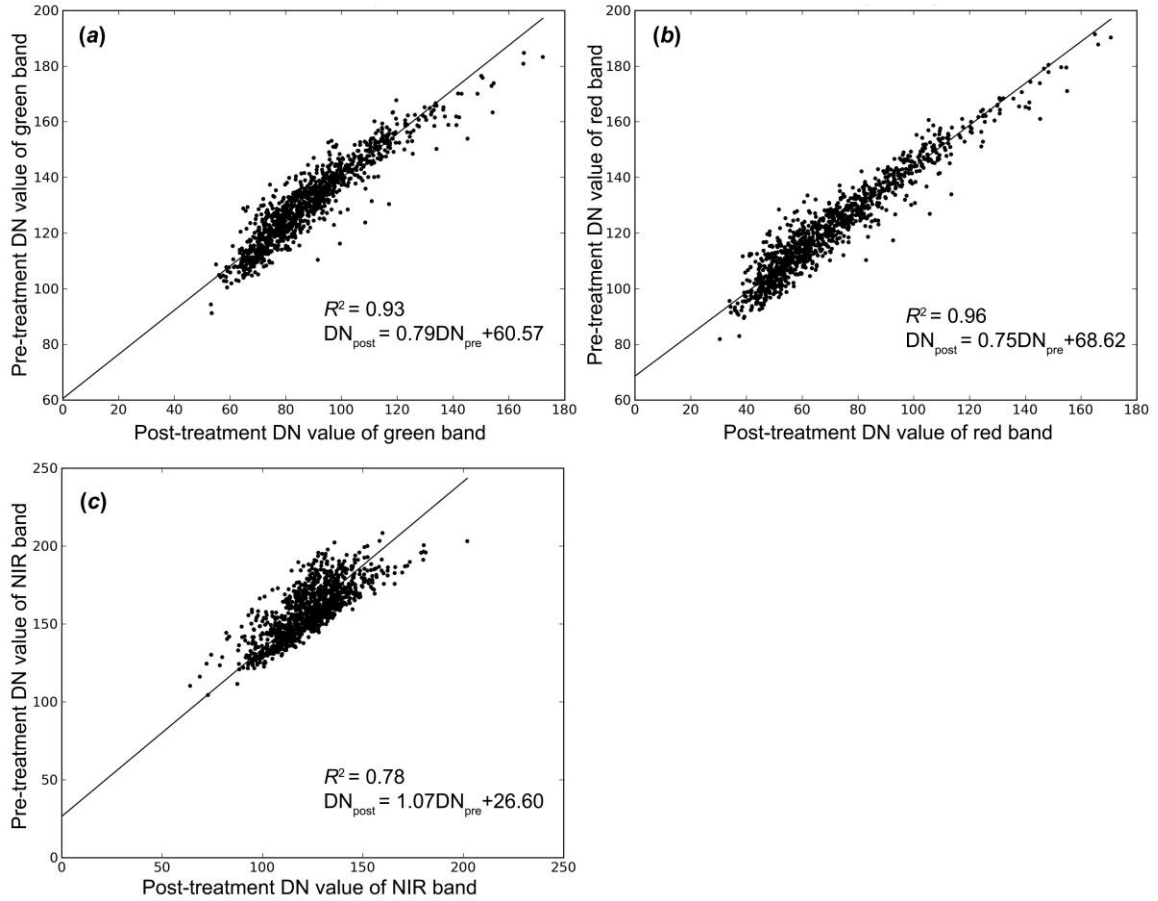
Figure 2-6 Scatter plots for the 714 time-invariant training pixels obtained from IR-MAD normalization method and corresponding simulated radiometric normalization equations for (a) green band, (b) red band, and (c) near infrared (NIR) band of aerial images.

2.4.3 FT detection results

The unfiltered pixel-wise FT detection results are shown in Figure 2-7. To quantitatively describe the noise rate, the detected FT areas outside the USFS FT footprints are temporally considered as "noise", and the noise rate is defined as the ratio of "noise" area to all detected FT area. As can be seen, the noise rate for the FT detection result using CC from ALS data is the lowest at 37%. The noise rates using CHM from ALS data and NDVI and standardized PCA methods from aerial imagery are slightly higher at 41%, 51%, and 41%, respectively. After applying the OBI segmentation routine to filter the pixel-wise results, the noise rate for all FT detection results has a significant decrease (Figure 2-8). Most of the noise from ALS derived CC method and aerial image derived NDVI method are concentrated in the northeastern and southwestern areas, and most of the noise from the standardized PCA method are concentrated in the southwestern areas.

The accuracy of the FT detection was further assessed by plot measurements. The FT detection using the ALS derived CC showed the highest accuracy among all results. The total accuracy for the FT detection result using ALS derived CC is about an average of 4%

higher than the other three methods, and the κ is about an average of 23% higher (Table 2-4). This is further supported by its producer accuracy, defined as the percentage of correct predictions based on field measurements. Its producer accuracy for treated areas is 74%, compared with 59%, 63% and 70% for the CHM method, NDVI method and standardized PCA method, respectively; its producer accuracy for untreated areas is about 96%, compared with 96%, 94%, and 92% for the other three methods (Table 2-4). Moreover, it accurately identified the FT activity in the USFS FT footprints within the rectangles "A", "B", "C" and "D" in Figure 8. All the other three methods detected close to nothing from these areas, except the CHM method within the rectangle "A" (Figure 2-8a). The area within rectangle "E" in Figure 2-8 was only recognized as treated by the method using the ALS derived CC. This area may be true FT detection for the fact that there were plots recognized as treated in it.
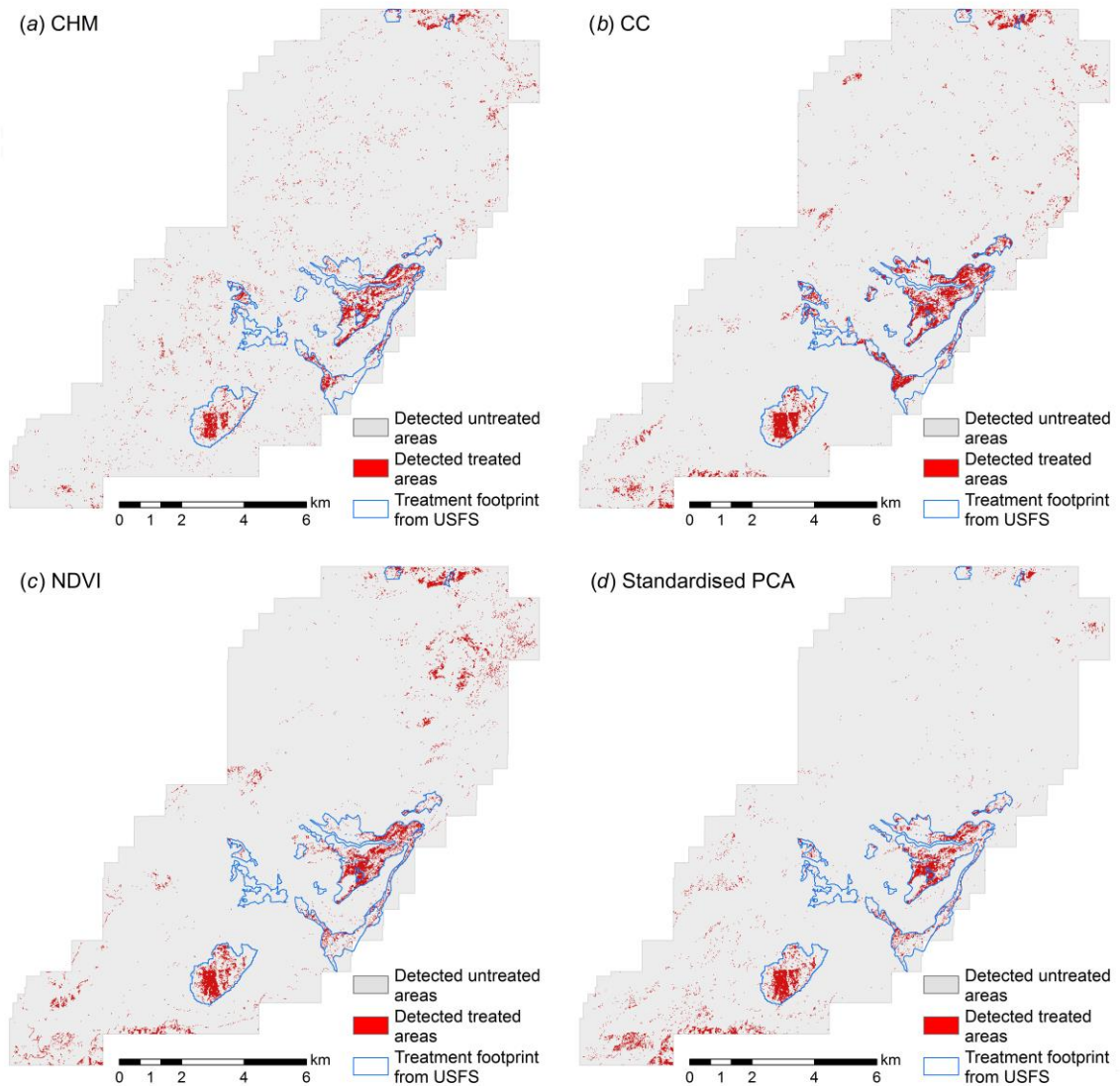
Figure 2-7 Unfiltered pixel-wise FT detection results from the methods using (a) ALS derived canopy height model (CHM), (b) ALS derived CC, (c) aerial imagery derived

normalized difference vegetation index (NDVI), and (d) aerial imagery derived standardized principle component analysis (PCA) change component. The location of the study site is presented in Figure 2-1.
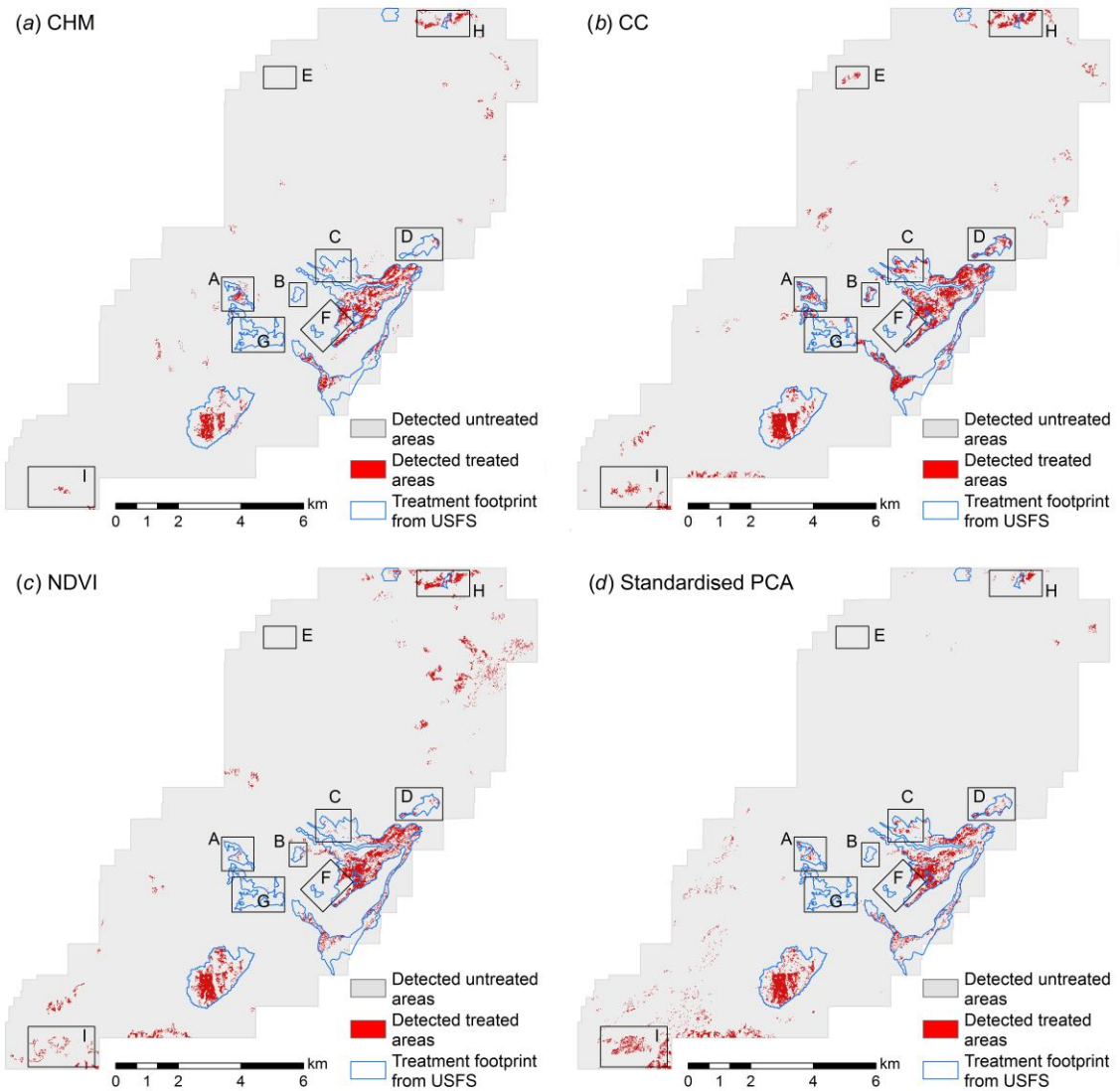
Figure 2-8 Filtered pixel-wise FT detection results from the methods using (a) ALS derived CHM, (b) ALS derived CC, (c) aerial imagery derived NDVI, and (d) aerial imagery derived standardized PCA change component. Areas in black rectangles marked with "A" to "H" are examples for different scenarios. Area "A" can be identified by methods using LiDAR data but not methods using aerial imagery; area "B" to area "D" can only be identified by LiDAR derived CC method; area "E", area "H", and area "I" can be identified by at least one of the four FT detection methods used in this study, but are not within the USFS FT footprints; area "F" and area "G" are within the USFS FT footprints, but cannot be identified by any of the four FT detection methods. The location of the study site is presented in Figure 2-1.

The CHM method from ALS data has slightly better accuracy than the NDVI method and the standardized PCA method using aerial images. Both the total accuracy and κ using the ALS derived CHM method are higher than those two methods using aerial images. The two methods using aerial images provided similar accuracies. Although the total accuracy for the result using the NDVI method is higher than that of the standardized PCA method and its κ is lower, their differences are very small, which are only 0.27% and 0.01, respectively (Table 2-4).

Table 2-4 Forest treatment detection accuracy assessment by comparing with field plot measurements.

		CHM		CC		NDVI*		Standardized PCA	
		Treated	Non-treated	Treated	Non-treated	Treated	Non-treated	Treated	Non-treated
Field data	Treated	27	19	34	12	29	17	32	14
	Non-treated	12	311	12	311	21	302	25	298
Total accuracy (%)		91.6		93.5		89.7		89.4	
Kappa coefficient		0.59		0.70		0.55		0.56	

*Normalized difference vegetation index.

2.4.4 FT intensity distribution
As mentioned in section 2.4.3, the ALS derived CC produced the FT detection result with the highest accuracy. Therefore, we tested the proposed FT intensity recognition method based on the FT extents from ALS derived CC. Figure 2-9 shows the distribution of profile area changes (pre-treatment minus post-treatment) within the FT extents. All profile area changes within FT extents are larger than 0% (Figure 2-9). The intensity of FTs conducted in the southern part of the study area is relatively larger, where the profile area change is over 20% in general. The profile area changes of FTs conducted in the middle and northern parts of the study area are concentrated in the value from 5% to 20%. Based on our plot measurements, the only two plots labeled as cable thinning in the field (indicating the removal of big trees) have profile area change values larger than 15%, and the only two plots labeled as mastication (indicating the removal of surface fuels, e.g. bushes) have profile area change values between 3% and 5%. As to the other 42 treated plots, they were labeled as general forest thinning, and therefore we cannot perform a detailed evaluation to the FT intensity recognition result currently.
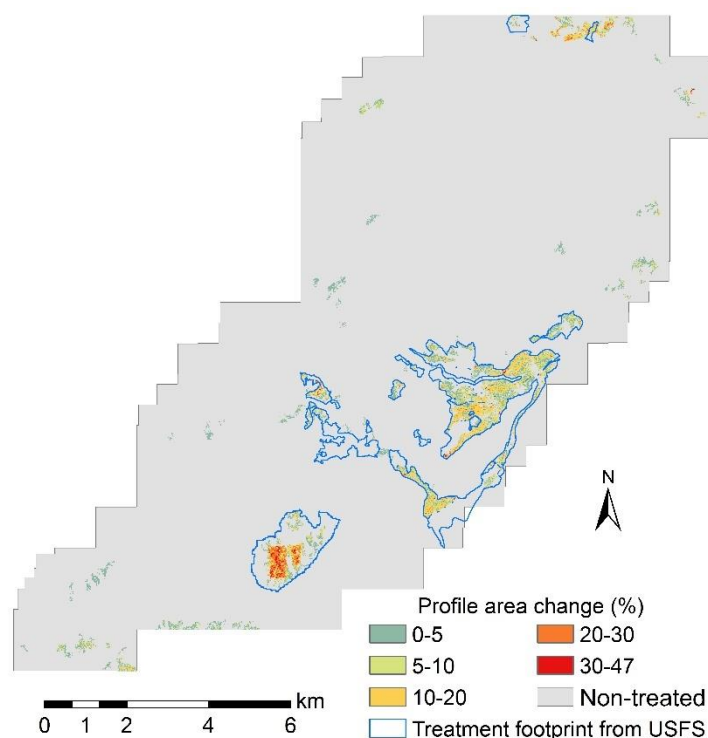
Figure 2-9 The FT severity distribution of the study area indicated by the profile area change in each 20 m by 20 m pixel (pre-treatment profile area minus post-treatment profile area). Note that this result is based on the FT detection result using LiDAR-derived CC method, which has the highest detection accuracy. The location of the study site is presented in Figure 2-1.

2.5 Discussion

In this study, we compared our FT detection results with other published works on using change detection techniques to monitor forest changes or map wildfires, and the accuracies of the FT detection results from all four methods in this study are comparable to these studies. For example, Desclée et al. (2006) monitored the land cover changes including forested areas in Eastern Belgium from SPOT-HRV imagery using a statistical OBI method and obtained an accuracy over 90%; Kennedy et al. (2007) detected forest disturbance dynamics in western Oregon, US from Landsat TM imagery using trajectory-based change detection method and obtained an overall accuracy of 90%; Schroeder et al. (2011) mapped wildfire and clearcut harvest disturbances from Landsat TM imagery in central Saskatchewan and obtained a total accuracy of around 90%. In this study, the total accuracies of the results of all four methods can reach approximately 90%, which are similar to these studies.

Among the four FT detection methods in this study, the FT detection results from the ALS derived CC method achieved the highest accuracy. A possible explanation for this may be related to the possible superiority of the ALS derived CC method in identifying less intensive FTs. Examining an example in areas within rectangle "B" in Figure 2-8,

nearly all removed trees in this area were short-statured trees (usually lower than 5 m in height) and were dispersed. We can clearly see that the removal of these small trees can still lead to a decrease in canopy cover (Figure 2-10*a* and *b*), which may increase the possibility for the ALS derived CC method to effectively detect these FT areas. Moreover, this result is also in consistent with other previously published works. Stephens et al. (2009) found that the CC decreased significantly in forests with mechanical FTs, and Kramer et al. (2014) found that the LiDAR percentage cover between 2 and 4 m was strongly correlated the forest ladder fuels and could be used to distinguish treated from untreated areas. To examine the influence of different height cutoffs on the FT detection results, we performed the same FT detection routine using the ALS derived CC products at 5 m and 10 m cutoffs, respectively (Figure 2-11). The omission rate of the results using a 5 m cutoff was only increased by around 8% compared with the results using a 2 m cutoff. After the cutoff reaching 5 m, the omission rate increased rapidly, and the omission rate of the results using a 10 m cutoff was 100%.
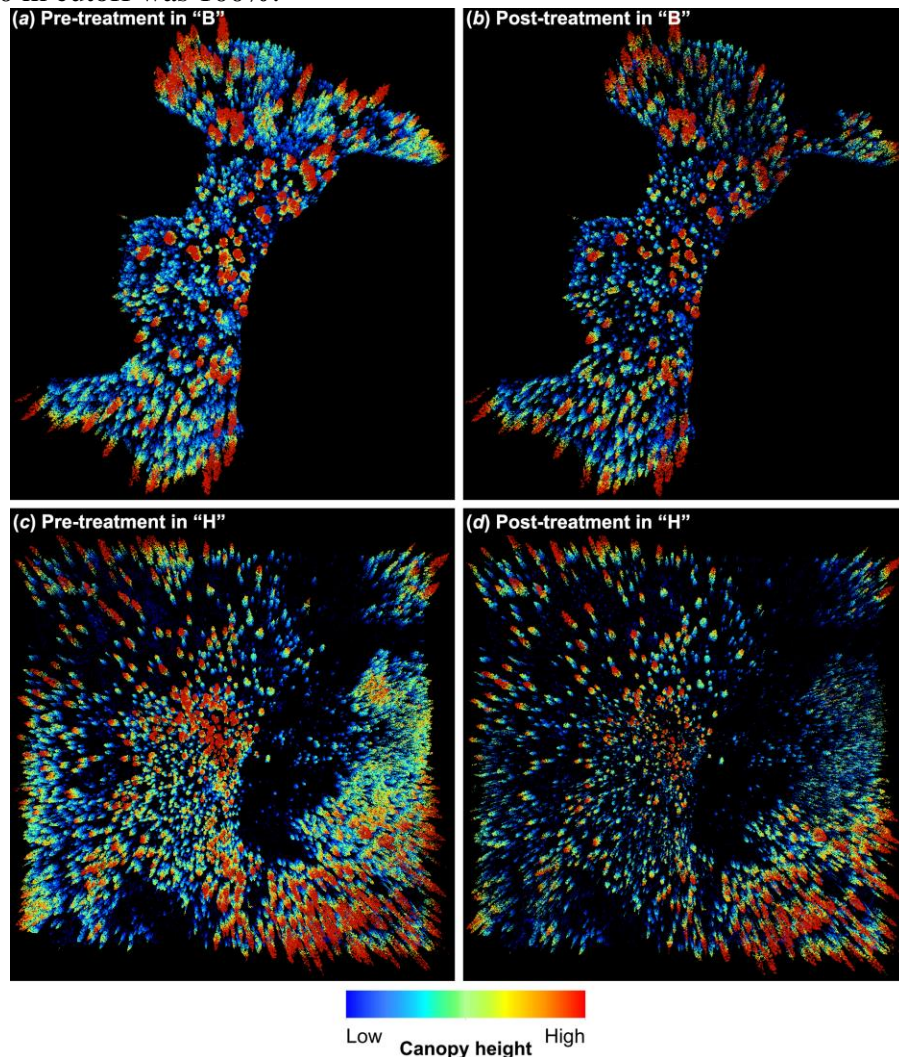
Figure 2-10 Comparison between the pre- and post-treatment three-dimensional point clouds within the rectangle "B" ((a) and (b)) and the rectangle "H" ((c) and (d)) in Figure 2-8.
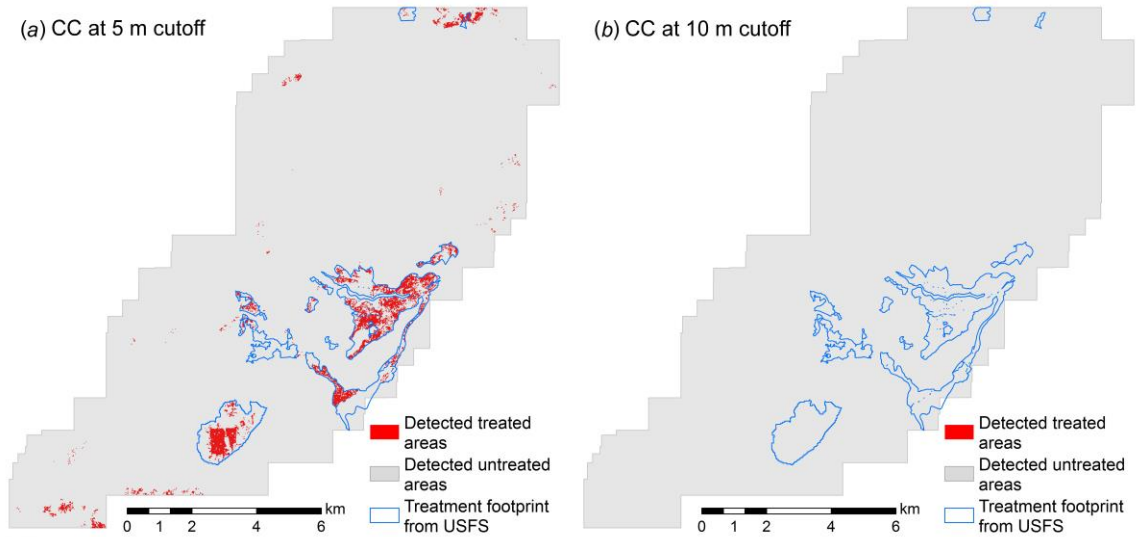
Figure 2-11 Comparison between FT detection results using the ALS derived CC products computed at a height break of (a) 5 m (b) 10 m, respectively.

The accuracy of FT detection from methods using aerial imagery (NDVI and standardized PCA) is generally lower than methods using ALS data. Both of these two methods poorly detected areas with less severe FTs. As mentioned, due to the penetration limitation, the aerial images can reflect the forest surface spectral characteristics, but may not provide accurate forest structure information (Weishampel et al., 2000). This is particularly problematic for detecting change in areas where overstory trees are unaffected, which is common for most FTs (Agee and Skinner 2005). Moreover, both of these two methods using aerial imagery have relatively high false FT detection rates, which are located at the northeast and southwest of the study area, respectively. The high false detection effects were commonly seen in previous studies using NDVI to map forest fires (Remmel and Perera, 2001). One of the possible explanations is the light cloud contamination suppressed the NDVI (Remmel and Perera, 2001). Another possible reason may be related to the tree mortality in those areas, which can result in a significant change in forest surface spectral reflectance (Jenson, 2005; Liu et al., 2006). Dead trees, especially those far away from the road (greater than 60 m), may not be removed from the forest.

Among the so-called noise areas, there are some common areas that were detected by several methods, which are concentrated in the north of the study site (within the rectangle "H") and the south of the study site (within the rectangle "I") (Figure 2-8). Although these areas are not included in the USFS FT footprints, this may not mean that these areas have not been treated. Figure 2-10c and d show the comparison between the pre- and post-treatment point clouds in the sub-area within rectangle "H". We can clearly see that there are trees that have been removed, which suggests that this area may have been treated. There are some areas within the USFS FT footprints but cannot be detected by any of the four methods, e.g. areas within rectangles "F" and "G" in Figure 2-8. It is possible that in these areas the intact overstory trees, which were not removed in the FTs, inhibited our ability to detect the removal of small tress with both ALS data and aerial imagery. There are eight field plots within the USFS FT footprints under these two squares, and six of them

were identified as non-treated in the field. Even for ALS data, the capability for mapping understory shrubs in dense forest stands can be limited (Estornell et al., 2011; Hodgson et al., 2005; Jakubowski et al., 2013a).

The proposed FT intensity recognition method shows the potential to detect the conducted FT types by comparing with a limited number of field measurements. As mentioned, Kramer et al. (2014) found that the LiDAR percentage cover between 2 and 4 m was highly correlated with forest ladder fuels. However, the height breaks or strata may not be consistent for different areas or treatment types. The proposed method directly examines the change in the vertical distribution of ALS point clouds, and may be applicable in different areas. Unfortunately, due to the lack of detailed ground truth data, we cannot map different profile area change values to different FT types in the current study. Moreover, different point cloud densities of different ALS flights may also influence the value of the profile area change. In this study, the pre- and post-treatment ALS data were acquired using the same protocol, and were accurately coregistered together using ground control points. This made the point cloud densities of the pre- and post-treatment flights very close. However, in most cases, the ALS point density from different ALS flights may significantly different with each other. Further study is still needed to identify how different ALS point densities influence the profile area changes.

Recently, the cost of ALS data acquisition has decreased (Tilley et al., 2004) and is comparable or even less expensive than the cost of large-scale field data collection and image data analysis (Hummel et al., 2011; Jakubowski et al., 2013b). For example, Hummel et al. (2011) compared the cost of using ALS data to exam forest structural attributes with the cost of using field measurements, and found that the cost of ALS data (6.31 points/m^2) acquisition and process was \$2.63/ac (acre) (US dollar) compared with \$2.46/ac (US dollar) to collect field measurements. Johansen et al. (2010) evaluated the costs of using ALS data (3.98 point/m^2), QuickBird imagery, and SPOT-5 imagery to study a stream network with a length of 26, 000 km, and the total research costs using these three datasets were \$3.8 M, \$6.4 M, and \$2.6 M (Australian dollar), respectively. Besides, considering the improvement of using ALS derived CC products to detect FTs (an increase of around 23% in κ compared to the methods using aerial imagery) and capability of ALS data on mapping the FT intensity, it may be a better choice for the forest managers to use ALS data to manage the FT activities.

2.6 Conclusions

This study developed a method combing pixel-wise thresholding and OBI segmentation to identify FT extents from ALS derived CC and CHM. A systematic comparison using the same FT detection routine as using ALS data was made to evaluate the capability of ALS data and aerial imagery on detecting FT extents. The results demonstrate the ability of ALS data to penetrate forest canopy, making it a more effective tool than aerial imagery for capturing forest structure change following FTs. Both the ALS derived CHM and CC provided higher accuracy in their FT detection results. The ALS derived CC, which can reflect the vertical structure information of a forest, produced an even better result than the ALS derived CHM method. Its total detection accuracy is over 93%, and its κ is 0.7, which are both the highest among all methods. FT detection results using NDVI and standardized

PCA from multi-temporal aerial imagery produced almost identical total accuracy and κ. Both methods showed relatively limited capacity to detect light FT areas, and had higher false detection rate (recognized untreated areas as treated areas) compared to the methods using ALS derived parameters.

Moreover, this study developed and tested a method to detect FT intensity directly from ALS point clouds. Based on a limited number of plot measurements, the methods show great potential to identify different FT types. However, we cannot make a detailed evaluation on the result due to the lack of ground truth in current study. Besides, how the change of ALS point densities of different ALS flights still needs to be further studied.

References

Agee, J.K., & Skinner, C.N. (2005). Basic principles of forest fuel reduction treatments. Forest Ecology and Management, 211, 83–96.

Allen, T.R., & Kupfer, J.A. (2000). Application of spherical statistics to change vector analysis of Landsat data: Southern Appalachian spruce-fir forests. Remote Sensing of Environment, 74, 482-493.

Andersen, H.E., McGaughey, R.J., & Reutebuch, S.E. (2005). Estimating forest canopy fuel parameters using LiDAR data. Remote Sensing of Environment, 94, 441–449.

Anyamba, A., & Tucker, C. (2005). Analysis of Sahelian vegetation dynamics using NOAA-AVHRR NDVI data from 1981–2003. Journal of Arid Environments, 63, 596-614.

Bortolot, Z.J., & Wynne., R.H. (2005). Estimating forest biomass using small footprint LiDAR data: An individual tree-based approach that incorporates training data. ISPRS Journal of Photogrammetry and Remote Sensing, 59, 342–360.

Canty, M.J., & Nielsen, A.A. (2008). Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation. Remote Sensing of Environment, 112, 1025-1036.

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

Carreiras, J., Pereira, J., & Pereira, J.S. (2006). Estimation of tree canopy cover in evergreen oak woodlands using remote sensing. Forest Ecology and Management, 223, 45-53.

Clark, M.L., Clark, D.B., & Roberts, D.A. (2004). Small-footprint LiDAR estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. Remote Sensing of Environment, 91, 68–89.

Collins, B.M., Stephens, S.L., Moghaddas, J.J., & Battles, J.J. (2010). Challenges and approaches in planning fuel treatments across fire-excluded forested landscapes. Journal of Forestry, 108, 24–31.

Collins, B.M., Stephens, S.L., Roller, G.B., & Battles, J.J. (2011). Simulating fire and forest dynamics for a landscape fuel treatment project in the Sierra Nevada. Forest Science, 57, 77–88.

Collins, J.B., & Woodcock, C.E. (1994). Change detection using the Gramm-Schmidt transformation applied to mapping forest mortality. Remote Sensing of Environment, 50, 267-279.

Collins, J.B., & Woodcock, C.E. (1996). An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data. Remote Sensing of Environment, 56, 66-77.

Desclée, B., Bogaert, P., & Defourny, P. (2006). Forest change detection by statistical object-based method. Remote Sensing of Environment, 102, 1-11.

Dubayah, R., Sheldon, S., Clark, D., Hofton, M., Blair, J., Hurtt, G., & Chazdon, R. (2010). Estimation of tropical forest height and biomass dynamics using LiDAR remote sensing at La Selva, Costa Rica. Journal of Geophysical Research: Biogeosciences, 115, G00E09.

Eastman, J.R., & Filk, M. (1993). Long sequence time series evaluation using standardized principal components. Photogrammetric Engineering & Remote Sensing, 59, 991-996.

Estornell, J., Ruiz, L., Velázquez-Martí, B., & Fernández-Sarría., A. (2011). Estimation of shrub biomass by airborne LiDAR data in small forest stands. Forest Ecology and Management, 262, 1697–1703.

Fiala, A. C.S., Garman, S.L., & Gray, A.N. (2006). Comparison of five canopy cover estimation techniques in the western Oregon Cascades. Forest Ecology and Management, 232, 188–197.

Guerra, F., Puig, H., & Chaume, R. (1998). The forest-savanna dynamics from multi-date Landsat-TM data in Sierra Parima, Venezuela. International Journal of Remote Sensing, 19, 2061-2075.

Guo, Q., Li, W., Yu, H., & Alvarez., O. (2010). Effects of topographic variability and LiDAR sampling density on several DEM interpolation methods. Photogrammetric Engineering & Remote Sensing, 76, 701–712.

Hame, T., Heiler, I., & San Miguel-Ayanz, J. (1998). An unsupervised change detection and recognition system for forestry. International Journal of Remote Sensing, 19, 1079-1099.

Hao, M., Shi, W.Z., Zhang, H., & Li, C. (2014). Unsupervised change detection with expectation-maximization-based level set. IEEE Geoscience and Remote Sensing Letters, 11, 210-214.

Hodgson, M.E., Jensen, J., Raber, G., Tullis, J., Davis, B.A., Thompson, G., & Schuckman, K. (2005). An evaluation of LiDAR-derived elevation and terrain slope in leaf-off conditions. Photogrammetric Engineering & Remote Sensing, 71, 817–823.

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

Hummel, S., Hudak, A.T., Uebler, E.H., Falkowski, M.J., & Megown, K.A. (2011). A comparison of accuracy and cost of LiDAR versus stand exam data for landscape management on the Malheur National Forest. Journal of Forestry, 109, 267–273.

Jakubowksi, M.K., Guo, Q., Collins, B.M., Stephens, S., & Kelly, M. (2013). Predicting surface fuel models and fuel metrics using LiDAR and CIR imagery in a dense, mountainous forest. Photogrammetric Engineering & Remote Sensing, 79, 37–49.

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

Jensen, J. R. (2005). Introductory digital image processing. A remote sensing perspective. Prentice Hall Series in Geographic Information Science. Upper Saddle River, NJ.

Johansen, K., Phinn, S., & Witte, C. (2010). Mapping of riparian zone attributes using discrete return LiDAR, QuickBird and SPOT-5 imagery: Assessing accuracy and costs. Remote Sensing of Environment, 114, 2679-2691.

Kane, V.R., Lutz, J.A., Roberts, S.L., Smith, D.F., McGaughey, R.J., Povak, N.A., & Brooks, M.L. (2013). Landscape-scale effects of fire severity on mixed-conifer and red fir forest structure in Yosemite National Park. Forest Ecology and Management, 287, 17-31.

Kane, V.R., North, M.P., Lutz, J.A., Churchill, D.J., Roberts, S.L., Smith, D.F., McGaughey, R.J., Kane, J.T., & Brooks, M.L. (2014). Assessing fire effects on forest spatial structure using a fusion of Landsat and airborne LiDAR data in Yosemite National Park. Remote Sensing of Environment, 151, 89-101.

Kennedy, R.E., Cohen, W.B., & Schroeder, T.A. (2007). Trajectory-based change detection for automated characterization of forest disturbance dynamics. Remote Sensing of Environment, 110, 370-386.

Korhonen, L., Korhonen, K.T., Rautiainen, M., & Stenberg, P. (2006). Estimation of forest canopy cover: a comparison of field measurement techniques. Silva Fennica, 40, 577.

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, 1065-1080.

Kramer, H.A., Collins, B.M., Kelly, M., & Stephens, S.L. (2014). Quantifying ladder fuels: A new approach using LiDAR. Forests, 5, 1432–1453.

Li, P., & Xu, H. (2010). Land-cover change detection using one-class support vector machine. Photogrammetric Engineering & Remote Sensing, 76, 255-263.

Li, X., & Yeh, A.G.O. (1998). Principal component analysis of stacked multi-temporal images for the monitoring of rapid urban expansion in the Pearl River Delta. International Journal of Remote Sensing, 19, 1501-1518.

Liu, D., Kelly, M., & Gong, P. (2006). A spatial–temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery. Remote Sensing of Environment, 101, 167–180.

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

Lu, D., Mausel, P., Brondizio, E., & Moran, E. (2004). Change detection techniques. International Journal of Remote Sensing, 25, 2365-2401.

Lucas, R.M., Cronin, N., Lee, A., Moghaddam, M., Witte, C., & Tickle, P. (2006). Empirical relationships between AIRSAR backscatter and LiDAR-derived forest biomass, Queensland, Australia. Remote Sensing of Environment, 100, 407-425.

Lunetta, R.S., Knight, J.F., Ediriwickrema, J., Lyon, J.G., & Worthy, L.D. (2006). Land-cover change detection using multi-temporal MODIS NDVI data. Remote Sensing of Environment, 105, 142-154.

Mas, J.F. (1999). Monitoring land-cover changes: a comparison of change detection techniques. International Journal of Remote Sensing, 20, 139-152.

Morsdorf, F., Kotz, B., Meier, E., Itten, K.I., & 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, 50-61.

Nielsen, A.A. (2007). The regularized iteratively reweighted MAD method for change detection in multi-and hyperspectral data. IEEE Transactions on Image Processing, 16, 463-478.

Patra, S., Ghosh, S., & Ghosh, A. (2011). Histogram thresholding for unsupervised change detection of remote sensing images. International Journal of Remote Sensing, 32, 6071-6089.

Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.-M., Tucker, C.J., & Stenseth, N.C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends in Ecology & Evolution, 20, 503-510.

Popescu, S., Zhao, K., & Neuenschwander, A. (2011). Satellite LiDAR vs. small footprint airborne LiDAR: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. Remote Sensing of Environment, 115, 2786-2797.

Prakash, A., & Gupta, R. P. (1998). Land-use mapping and change detection in a coal mining area - a case study in the Jharia Coalfield, India. International Journal of Remote Sensing, 19, 391–410.

Qin, Y., Niu, Z., Chen, F., Li, B., & Ban, Y. (2013). Object-based land cover change detection for cross-sensor images. International Journal of Remote Sensing, 34, 6723-6737.

Remmel, T.K., & Perera, A.H. (2001). Fire mapping in a northern boreal forest: assessing AVHRR/NDVI methods of change detection. Forest Ecology and Management, 152, 119-129.

Riano, D., Valladares, F., Condes, S., & Chuvieco, E. (2004). Estimation of leaf area index and covered ground from airborne laser scanner (LiDAR) in two contrasting forests. Agricultural and Forest Meteorology, 124, 269-275.

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

Ridd, M.K., & Liu, J.J. (1998). A comparison of four algorithms for change detection in an urban environment. Remote Sensing of Environment, 63, 95-100.

Schroeder, T.A., Wulder, M.A., Healey, S.P., & Moisen, G.G. (2011). Mapping wildfire and clearcut harvest disturbances in boreal forests with Landsat time series data. Remote Sensing of Environment, 115, 1421-1433.

Seto, K.C., Woodcock, C.E., Song, C., Huang, X., Lu, J., & Kaufmann, R.K. (2002). Monitoring land-use change in the Pearl River Delta using Landsat TM. International Journal of Remote Sensing, 23, 1985-2004.

Singh, A. (1989). Review Article Digital change detection techniques using remotely-sensed data. International Journal of Remote Sensing, 10, 989-1003.

Sohl, T.L. (1999). Change analysis in the United Arab Emirates: An investigation of techniques. Photogrammetric Engineering & Remote Sensing, 65, 475-484.

Stephens, S.L., Millar, C.I., & Collins, B.M. (2010). Operational approaches to managing forests of the future in Mediterranean regions within a context of changing climates. Environmental Research Letters, 5, 024003.

Stephens, S.L., Moghaddas, J.J., Edminster, C., Fiedler, C.E., Haase, S., Harrington, M., Keeley, J.E., Knapp, E.E., McIver, J.D., Metlen, K., & Youngblood, A. (2009). Fire

treatment effects on vegetation structure, fuels, and potential fire severity in western U.S. forests. Ecological Applications, 19, 305-320.

Su, Y., Guo, Q., Fry, D.L., Collins, B.M., Kelly, M., Flanagan, J.P., & Battles, J.J. (2016). A vegetation mapping strategy for conifer forests by combining airborne LiDAR data and aerial imagery. Canadian Journal of Remote Sensing, 42, 1-15.

Tilley, B.K., Munn, I.A., Evans, D.L., Parker, R.C., & Roberts, S.D. (2004). Cost considerations of using LiDAR for timber inventory. Paper presented at the Southern Forest Economics Workshop, St. Augustine, FL, March 14–16.

Walter, V. (2004). Object-based classification of remote sensing data for change detection. ISPRS Journal of Photogrammetry and Remote Sensing, 58, 225–238.

Weishampel, J.F., Blair, J.B., Knox, R., Dubayah, R., & Clark, D. (2000). Volumetric LiDAR return patterns from an old-growth tropical rainforest canopy. International Journal of Remote Sensing, 21, 409-415.

Yousif, O., & Ban, Y.F. (2013). Improving urban change detection from multitemporal SAR images using PCA-NLM. IEEE Transactions on Geoscience and Remote Sensing, 51, 2032-2041.

Yu, X., Hyyppä, J., Kukko, A., Maltamo, M., & Kaartinen, H. (2004). Change detection techniques for canopy height growth measurements using airborne laser scanner data. Photogrammetric Engineering & Remote Sensing, 72, 1339-1348.

Yu, X., Hyyppa, J., Kukko, A., Maltamo, M., & Kaartinen, H. (2006). Change detection techniques for canopy height growth measurements using airborne laser scanner data. Photogrammetric Engineering & Remote Sensing, 72, 1339.

Zhang, X., Xiao, P., & Feng, X. (2013). Impervious surface extraction from high-resolution satellite image using pixel-and object-based hybrid analysis. International Journal of Remote Sensing, 34, 4449–4465.

Zhao, K., & Popescu. S. (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, 1628–1645.

Zhao, K., Popescu, S., & Nelson, R. (2009). "LiDAR remote sensing of forest biomass: a scale invariant estimation approach using airborne lasers. Remote Sensing of Environment, 113, 182–196.

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

Chapter 3   A vegetation mapping strategy for conifer forests by combining airborne LiDAR data and aerial imagery

Abstract

Accurate vegetation mapping is critical for natural resources management, ecological analysis, and hydrological modeling, among other tasks. Remotely sensed multispectral and hyperspectral imageries have proved to be valuable inputs to the vegetation mapping process, but they can provide limited vegetation structure characteristics, which are critical for differentiating vegetation communities in compositionally homogeneous forests. Light detection and ranging (LiDAR) can accurately measure the forest vertical and horizontal structures, and provide a great opportunity for solving this problem. This study introduces a strategy using both multispectral aerial imagery and LiDAR data to map vegetation composition and structure over large spatial scales. Our approach included the use of a Bayesian Information Criterion algorithm to determine the optimized number of vegetation groups within mixed-conifer forests in two study areas in the Sierra Nevada, California, and an unsupervised classification technique and post-hoc analysis to map these vegetation groups across both study areas. The results show that the proposed strategy can recognize four and seven vegetation groups at the two study areas, respectively. Each vegetation group has its unique vegetation structure characteristics or vegetation species composition. The overall accuracy and kappa coefficient of the vegetation mapping results are over 78% and 0.64 for both study sites.

Keywords: Vegetation mapping; LiDAR; aerial imagery; forest structure; mixed conifer forest.

3.1 Introduction

Vegetation mapping is the process of characterizing vegetation units across a landscape from measured environmental parameters (Franklin 1995, Pedrotti 2012). Typically these units convey information about the dominant plant species present and the morphological structure of the vegetation (e.g., a mesic hardwood or a high-elevation meadow). Accurate and up-to-date vegetation maps are critical for managers and scientists, because they serve a range of functions in natural resource management (e.g., forest inventory, timber harvest, wildfire risk control, wildlife protection), ecological and hydrological modelling, and climate change studies (Chuvieco and Congalton 1988, Talbot and Markon 1988, Daly et al. 1994, Stephens 1998, Pearce et al. 2001, Mermoz et al. 2005, Alvarez et al. 2013). Traditional methods for vegetation mapping usually rely on field surveys, literature reviews, aerial photography interpretation, and collateral and ancillary data analysis (Pedrotti 2012). However, these methods are expensive and time-consuming. Consequently, vegetation maps produced by the traditional approach reflect past conditions when released and are not updated frequently (Daly et al. 1994).

Remote sensing has proved to be a powerful tool in vegetation mapping by employing image classification techniques. Multispectral remote sensing imagery such as Landsat, SPOT, MODIS, AVHRR, IKONOS, and QuickBird are among of the most commonly used. For example, Franklin (1986) used the Landsat Thematic Mapper (TM) simulator data to

discriminate the composition of conifer forests in the Klamath Mountains in northern California. Carpenter et al. (1999) produced a lifeform map for the Sierra Nevada in California from Landsat TM data by applying the ARTMAP neural network method. Liu et al. (2006) mapped the distribution of forest disease, sudden oak death, in northern California from two-year images obtained by Airborne Data Acquisition and Registration system. Mallinis et al. (2008) used object-based classification method to delineate vegetation polygons in a conifer forest from Quickbird imagery. Wang et al. (2004) combined pixel-based and object-based classification methods to map the different mangrove canopy types along the Caribbean coast of Panama. Zhang et al. (2003) and Knight et al. (2006) monitored vegetation to produce phenology-based land cover maps from MODIS data. Besides multispectral data, hyperspectral imagery is another frequently used data type in vegetation mapping (Hirano et al. 2003, Li et al. 2005). The use of hyperspectral data can produce more finely classified vegetation mapping results than multispectral data (Xu and Gong 2007, Adam et al. 2010), because hyperspectral sensors are designed to collect data from hundreds of continuous spectral channels compared with multispectral sensors with broad wavelength intervals.

All these studies using both multispectral and hyperspectral imagery usually only focus on either mapping the land cover type or mapping the vegetation composition. Examining the detailed structure characteristics in forests has rarely been considered because the limited penetration capability for multispectral and hyperspectral data. However, this information also plays a very important role in many ecological studies. For example, Lindenmayer et al. (2000) advocated that forest structure-based parameters can impact biodiversity, and should be taken into account in forest managements. Zielinski et al. (2006) and García-Feced et al. (2011) demonstrated that forest structure information was critical for mapping the habitat of Pacific fisher (*Pekania pennanti*) and California spotted owl (*Strix occidentalis occidentalis*). Graham et al. (2004), Agee and Skinner (2005) and Peterson et al. (2005) all pointed to the important role that forest structure has on wildfire behavior, and argued that modifying forest structure through forest treatment may be necessary to reduce fire risk in many dry conifer forest types. Developing methods to integrate structure information into the process of vegetation mapping is an important area of research.

Light detection and ranging (LiDAR), an active remote sensing technique, can accurately measure the three-dimensional distribution of surface objects (Lefsky et al. 2002). The focused and narrow laser beam used by LiDAR sensors has a strong penetration capability in forest areas (Lim et al. 2003, Jensen 2009, Su and Guo 2014). It has been well documented that LiDAR data can be used to derive highly reliable forest structure parameters such as tree height (Nilsson 1996, Andersen et al. 2006, Su et al. 2015), canopy cover (Lim et al. 2003, Korhonen et al. 2011), leaf area index (Riaño et al. 2004, Jensen et al. 2008), stand volume (Nilsson 1996, Naesset 1997), and tree diameter (Popescu 2007, Huang et al. 2011). The capacity to resolve forest structure parameters provides a great opportunity for developing vegetation mapping strategies (Kramer et al. 2014). Donoghue et al. (2007) and Heinzel and Koch (2011) explored the possibility of identifying tree species mixtures from parameters derived from LiDAR data. Ørka et al. (2009) and Kim et al. (2009) used LiDAR intensity data to differentiate broadleaf and needleleaf trees. Reitberger et al. (2008) used full-waveform LiDAR data to classify deciduous and

coniferous trees. Holmgren and Persson (2004) identified individual tree species, including Norway spruce (*Picea abies L. Karst*), Scots pine (*Pinus sylvestris L.*), and deciduous tree, by analyzing individual crown shape and rich tree structure parameters derived from LiDAR data. However, due to the lack of forest canopy spectral information, the accuracy of tree species classification from LiDAR data is limited in complex vegetation conditions.

The integration of LiDAR data and multispectral/hyperspectral imagery has been used to address the limitation of using only LiDAR data in vegetation mapping. For example, Cho et al. (2012), Colgan et al. (2012) and Naidoo et al. (2012) mapped tree species compositions in African savannas through the combination of LiDAR data and hyperspectral data using maximum likelihood, random forest, and support vector machine classifiers, respectively; Dalponte et al. (2012) and Hill and Thomson (2005) classified tree species compositions of broadleaf and coniferous mixed forests through the fusion of spectral and LiDAR data; Holmgren et al. (2008) and Koukoulas and Blackburn (2005) used maximum likelihood classifier to identify individual tree species from LiDAR derived structure parameters and multispectral information in deciduous and coniferous forests, respectively. It has been reported that the integration of LiDAR data and optical imagery can increase the vegetation composition classification accuracy by 16-20% in rangelands, compared to using only LiDAR data or optical imagery (Bork and Su 2007). However, most of these studies on mapping vegetation units are still mainly focusing on classifying vegetated from non-vegetated areas or detecting differences in species composition. Forest structure characteristics, which can be estimated by statistical imputation methods that incorporate field measurements with LiDAR data and optical imagery (Falkowski et al. 2010, Hummel at al. 2011, Wallerman and Holmgren 2007), are rarely considered in classification systems.

The objective of this study is to develop and test a new strategy to map vegetation communities in two mixed conifer forests by considering both the dominant tree species composition and vegetation structure characteristics. Multispectral aerial imagery and airborne LiDAR data were integrated, along with a robust network of systematically established field plots in the vegetation mapping process. An unsupervised classification scheme using an automatic cluster determination algorithm based on Bayesian Information Criterion (BIC) and k-means classification was applied to the fused data to map the vegetation, and a post-hoc analysis based on field measurements was used to interpret the ecological properties for each vegetation unit.

3.2 Materials and methods
3.2.1 Study areas
Our two forest study sites are located in the Sierra Nevada mountain range, California, USA (Figure 3-1). The northern site, Last Chance, covers an area of 92.1 km^2, and the southern site, Sugar Pine, covers an area of 72.8 km^2. The elevation ranges from 280 m to 2190 m for the Last Chance site and from 500 m to 2650 m for the Sugar Pine site, and the average elevation for both study sites is over 1500m. Trees common to the Sierran mixed conifer and true fir forests dominate the vegetation cover at both sites. The major species present include: ponderosa pine (*Pinus ponderosa*), incense-cedar (*Calocedrus decurrens*), sugar pine (*Pinus lambertiana*), white fir (*Abies concolor*), California red fir (*Abies*

magnifica), and Douglas-fir (*Pseudotsuga menziesii*). Within the mixed conifer stands, the major hardwoods are black oak (*Quercus kelloggii*) and canyon live oak (*Quercus chrysolepis*). Forest cover is relatively homogeneous at both the study sites, but the Last Chance site has more heterogeneity than the Sugar Pine site.
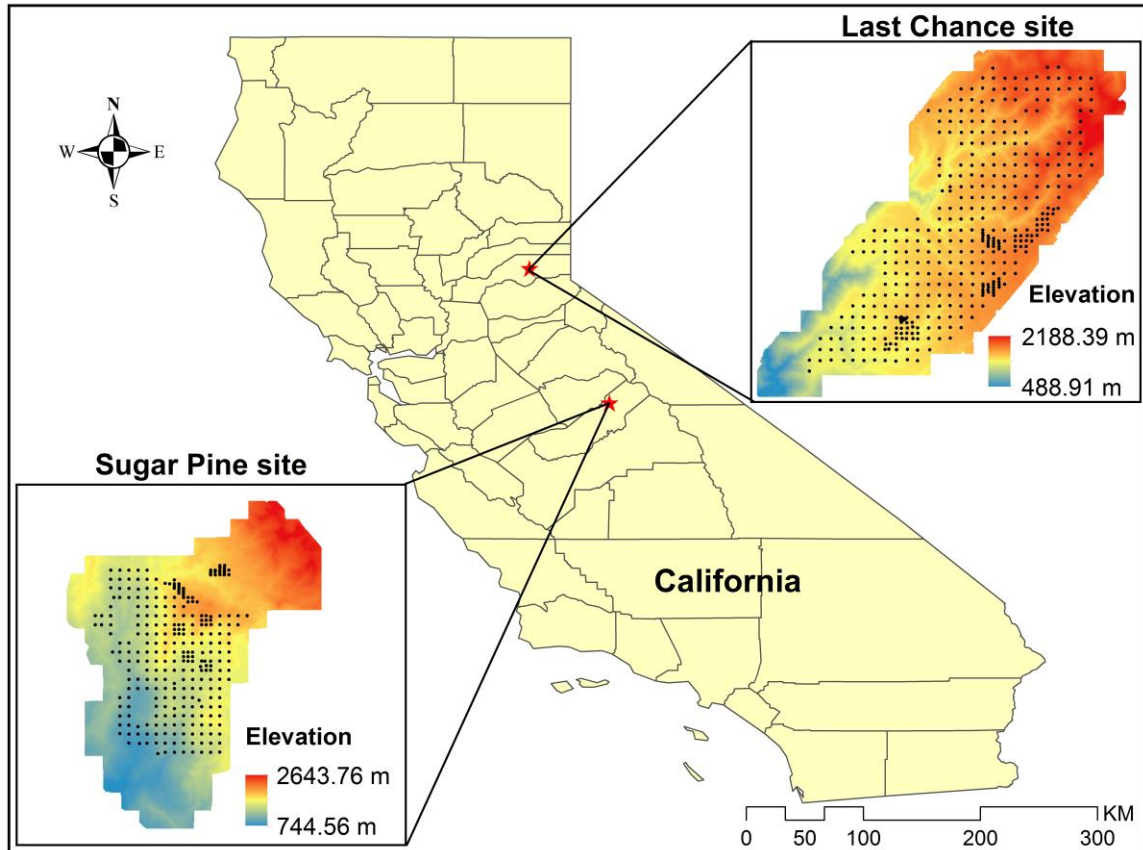
Figure 3-1 The geo-locations and terrain information of the Last Chance and Sugar Pine study sites with the distribution of field plots.

3.2.2 Field measurements

Plot measurements (12.62 m in radius and 500 m^2 in area) were taken in the summer of 2007 and 2008 (Figure 3-1). The same plot selection procedure was applied to determine the location of 372 and 268 evenly distributed plots in Last Chance site and Sugar Pine site, respectively. A random point was first chosen to be used as the center of the first plot in each study site. Then, this plot center was taken as a seed point to build a grid on a 500 m spacing in the four cardinal directions, and the following plot centers were placed on the intersections of the grid. Within watersheds for specific research purposes (e.g., studying hydrological responses to forest fuel treatments), the sampling was intensified to a 250 m by 250 m grid. The position of each plot center in the field was located using a Trimble$^{\text{TM}}$ GeoXH GPS. If there were any landing or road surfaces within the plot footprint, the plot center was randomly moved by 25 m in one of the four cardinal directions.

Within each plot, field measurements on individual live trees included: tree species, tree height, diameter at breast height (DBH, breast height = 1.37 m) and height to live crown base. Trees were defined as individuals at least 5 cm in DBH. Moreover, the plot level canopy cover was measured using a sight tube with 25 sampling points. The plot level Lorey's height and total basal area were calculated from field measurements and used in the vegetation mapping process in this study, which can be calculated from the following equations.

$$LH_z = \frac{\sum_{i=1}^{n_z} BA_i \times H_i}{\sum_{i=1}^{n_z} BA_i} \qquad (3\text{-}1)$$

$$TBA_z = \sum_{i=1}^{n_z} BA_i \qquad (3\text{-}2)$$

where LH_z and TBA_z represent the Lorey's height and total basal area of the zth plot, and BA_i and H_i are the basal area and tree height of the ith tree in the zth plot.

3.2.3 LiDAR data

Small footprint airborne LiDAR data covering the Sugar Pine site and Last Chance site were acquired in September, 2007 and September, 2008 using an Optech GEMINI airborne laser terrain mapper (ALTM) from the National Center of Airborne Laser Mapping at the University of Houston. It was mounted on a twin-engine Cessna Skymaster, and was flown at 600-700 m above the ground. The ALTM sensor was operated at 100 kHz with a scanning frequency of 40-60 Hz and a total scan angle of 24–28°. The average swath width of a single pass was around 510 m, and the overlap between two adjoining swaths was 65% of the swath width. The point density was 6–10 points/m^2, and positioning accuracy was about 10 cm horizontally and 10-15 cm vertically.

Overall, there are 13 layers derived from the raw LiDAR point cloud for both study sites, including the canopy height model (CHM), canopy cover, and 11 canopy quantile metrics. The CHM was calculated by the difference between the LiDAR derived digital elevation model (DEM) and digital surface model (DSM), which were interpolated from the LiDAR ground returns and LiDAR first returns, respectively. The interpolation algorithm used in this study was ordinary kriging, which has been proved to be more accurate than other schemes (e.g., inverse distance weighted or spline) for interpolating DEM and DSM from LiDAR derived elevation points (Lloyd and Atkinson 2002, Clark et al. 2004, Guo et al. 2010).

The canopy cover was calculated by a CHM based method, a reliable and consistent approach for estimating canopy cover from LiDAR data (Lucas et al. 2006). First, a fine resolution CHM (1×1 m^2) was calculated from LiDAR point cloud using the above-mentioned algorithm, and the pixels above a selected height threshold were coded as 1 or 0 otherwise. The height threshold was set as 2 m in this study to match field-based canopy cover measurements. Then, this coded CHM was used to overlap with a 20×20 m^2 grid, and the canopy cover was calculated as the percentage of the number of coded CHM pixels

with a value of 1 to the total number of coded CHM pixels within each 20×20 m^2 grid. The final canopy cover layer was produced in 20 m resolution to roughly match the scale of field plots.

Canopy quantile metrics, representing the height below *X%* of the LiDAR point cloud, are one of most frequently used LiDAR products to estimate the forest parameters that cannot be obtained directly from LiDAR point cloud, e.g. diameter at breast height and biomass (Lim and Treitz 2004, Thomas et al. 2006). In this study, 11 quantile metrics, including 0%, 1%, 5%, 10%, 25%, 50%, 75%, 90%, 95%, 99%, and 100%, were calculated in 20 m resolution directly from the LiDAR point cloud.

3.2.4 Aerial imagery
The 2005 National Agriculture Imagery Program (NAIP) color-infrared (CIR) aerial imagery in 1×1 m^2 resolution (composed by green band, red band and near-infrared (NIR) band) are used in the vegetation mapping procedure of this study. The NAIP program is run by the Farm Service of US Department of Agriculture (USDA) for the purpose of making high-resolution digital orthographies available to maintain common land units. All NAIP images were taken under permitted weather condition, and followed the specification of no more than 10% cloud cover per quarter quad tile. The Aerial Photography Field Office has adjusted and balanced the dynamic range of each image tile to full range of Digital Number (DN) value (0-255), and orthorectified each image file using National Elevation Dataset before releasing the data (Hart and Veblen 2015). To ensure the NAIP imagery co-registered with LiDAR data, we georeferenced the NAIP imagery using over 20 correspondence points for each study site selected from NAIP imagery and LiDAR derived products (i.e., DEMs and CHMs).

In addition to the three spectral bands, seven texture layers (including mean, variance, homogeneity, contrast, dissimilarity, entropy and second moment) were extracted from each spectral band using the gray-level co-occurrence matrix (GLCM) filtering method. GLCM is defined over an image to be the distribution of co-occurring values at a given offset (*Δx*, *Δy*) (Haralick et al. 1973, Anys et al. 1994, Soh and Tsatsoulis 1999), which can be mathematically described as:

$$GLCM_{\Delta x, \Delta y}(i, j) = \sum_{p=1}^{m} \sum_{q=1}^{n} \begin{cases} 1, & \text{if } I(p, q) = i \text{ and } I(p + \Delta x, q + \Delta y) = j \\ 0, & \text{otherwise} \end{cases} \tag{3-3}$$

where (*i*, *j*) is one DN values combination of the image *I* at the given offset (*Δx*, *Δy*), (*p*, *q*) are the spatial position indexes in the image *I*, and (*m*, *n*) are the number of rows and columns of the image *I*. The offset (*Δx*, *Δy*) is determined by the angular relation between the neighboring pixels and spatial resolution of the image. The texture parameters for the corresponding GLCM can be calculated using equations provided by Haralick et al. (1973), and will not be discussed in detail here. In this study, a 3×3 moving window was used to generate GLCMs and calculate corresponding texture parameters for each cell. To match the spatial scale of the field plots and LiDAR products, the NAIP imagery and obtained texture layers were resampled to the resolution of 20×20 m^2 using the weighted mean value method (Jakubowksi et al. 2013). All of the following vegetation mapping procedures used the resampled NAIP imagery and texture layers.

3.2.5 Vegetation mapping strategy

There are overall 24 aerial imagery derived features (including the spectral bands and derived texture layers) and 13 LiDAR derived features initially available for this analysis. This large number of potential input layers for vegetation mapping could negatively influence the results given the likelihood of redundant information captured by the layers. Many algorithms have been developed to reduce the dimensionality of an input dataset, e.g., Principal Component Analysis (PCA), Linear Discriminant Analysis, Correspondence Analysis, and Detrended Correspondence Analysis. As one of the most commonly used techniques, the PCA algorithm has been proven to be effective at removing redundant information in remotely sensed data (Mutlu et al. 2008; Pohl and Van Genderen 1998). Therefore, in this study, the standardized PCA method was first applied separately to the aerial imagery derived and LiDAR derived features (Figure 3-2). The first three PCA components from aerial imagery derived features and the first three components from LiDAR derived features were combined as the input for vegetation mapping strategy. An unsupervised classification strategy and post-hoc analysis integrated with field measurements was then applied on the six PCA components to define vegetation groups and delineate the boundaries of different groups (Figure 3-2). The detailed descriptions for the unsupervised classification strategy and post-hoc analysis are provided below.
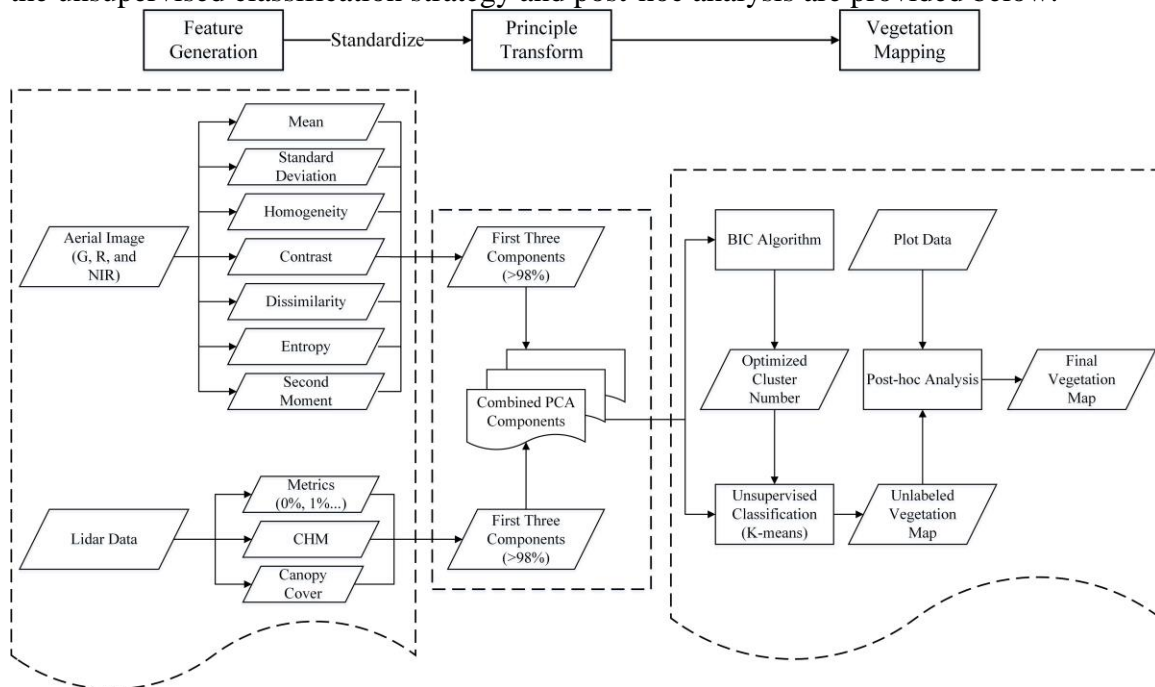
Figure 3-2 Procedure for the vegetation mapping strategy used in this study.

3.2.5.1 Unsupervised classification strategy

The specific number and character of vegetation groups within a particular forest are usually unknown prior to the vegetation mapping process. Thus, one of the main challenges for vegetation mapping is to identify distinct vegetation groups and delineate boundaries among groups. In this study, an automatic cluster number determination algorithm based

on BIC, developed by Chiu et al. (2001), was combined with k-means unsupervised classification to initially map the vegetation. BIC is a robust measure for model selection among a finite set of models, which is defined as:

$$BIC_k = -2l_k + r_k \log n \tag{3-4}$$

where k is the cluster number, l_k is the classification likelihood function, r_k is the number of independent parameters, and n is the number of observations.

To obtain the optimized cluster number, a large maximum cluster number was first defined. In this study we used the hierarchical cluster analysis of species composition (linkage method = Ward's; distance measure = Euclidean) following the method described in McCune et al. (2002) to determine the maximum number of vegetation groups in both study sites. BIC values for all possible cluster numbers (from one to the defined maximum cluster number) were then calculated. With these BIC values, the optimized number of clusters was determined in two steps. First, the initial value of the cluster number was estimated. Let $dBIC(k)$ be the change of BIC values from two adjacent cluster numbers ($dBIC(k) = BIC_k - BIC_{k-1}$), and $rBIC(k)$ be the ratio of BIC from k clusters and BIC from only one cluster ($rBIC(k) = BIC_k / BIC_1$). If the $dBIC(2)$ was larger than 0, the initial cluster number was set as one; otherwise, the initial cluster number was set equal to the number of clusters where $rBIC(k)$ was smaller than 0.04 for the first time. Second, if the initial cluster number was one, the final cluster number was set as one; otherwise, the ratio change in log-likelihood distance was further used to optimize the cluster number. Let $R(k)$ be the ratio of log-likelihood distances (d_k) from two adjacent cluster numbers ($R(k) = d_k / d_{k-1}$). The ratio of change in log-likelihood was computed as $R(k_1) / R(k_2)$, where k_1 and k_2 were the cluster numbers of the two largest $R(k)$ smaller than the obtained initial cluster number. If the ratio of change was larger than 1.15, the final cluster number was set equal to k_1; otherwise, it was set equal to the maximum value between k_1 and k_2. It should be noted that all the thresholds used in the BIC algorithm were determined by statistical experiments by Chiu et al (2001).

With the optimized cluster number, we used k-means clustering algorithm to delineate the boundary of different vegetation types. K-means divides observations into a pre-defined number of clusters and each observation belongs to the cluster with the nearest mean (Hartigan 1975), which can be mathematically described as:

$$\arg \min \sum_{i=1}^{k_{BIC}} \sum_{x_j \in S_i} \left\| x_j - u_i \right\|^2 \tag{3-5}$$

where k_{BIC} is the pre-defined number of clusters, x_j is the jth observation vector, S_i is the ith set of observation vectors, and μ_i is the mean point of the ith set. In this study, the maximum iterations for k-means unsupervised classification was set to 10, and the change threshold of the mean points was set to 5%.

3.2.5.2 Post-hoc analysis

Field measurements were used to describe the dominant tree species composition and forest structure characteristics. The unsupervised vegetation group for each plot was extracted by overlapping the plot location with unsupervised classification result. Then, for all plots

belonging to the same unsupervised classification group, we analyzed their dominant tree species and forest structure characteristics measured from the field. The dominant tree species were defined by the proportions of different tree species weighted by basal area, and the forest structure characteristics were defined by the plot level basal area, Lorey's height and canopy cover. Finally, these plot-derived dominant tree species information and forest structure characteristics were used to determine the property of each unsupervised classification group. It should be noted that approximately two-thirds of the plots (273 in Last Chance and 177 in Sugar Pine) were randomly selected and used to define vegetation group properties. The other plots were reserved to validate the vegetation mapping result.

3.2.6 Accuracy assessment

PCA ordination analysis, one type of multivariate analysis that can depict species relationships in low-dimensional space (Gauch 1982), was used to evaluate the capability of proposed vegetation mapping strategy on differentiating tree species. It has been widely used as a complement to other data clustering techniques that help identify repeatable vegetation patterns and discontinuities in species composition (Lepš and Šmilauer, 2003). In this study, relative species abundance for ordination analysis was represented by basal area (i.e., the ratio of basal area for each tree species to the total basal area of all trees at a plot). Moreover, the permutation test, a type of robust non-parametric statistical significance test (Nichols and Holmes 2002), was used to evaluate the capability of the proposed vegetation mapping strategy on recognizing different structure characteristics, since the field measured forest structure parameters are not normally distributed based on Shapiro-Wilk test ($\alpha=0.05$) (Table 3-1).

Table 3-1 Tests of normality for forest structure parameters using Shapiro-Wilk test.

	Last Chance site			Sugar Pine site		
	Statistic	df	Sig.	Statistic	df	Sig.
Lorey's Height	0.630	370	0.000	0.988	268	0.030
Basal Area	0.489	370	0.000	0.941	268	0.000
Canopy Cover	0.988	370	0.003	0.048	268	0.000

In addition, the total accuracy (*TA*) and kappa coefficient (κ) were also calculated for the purpose of evaluating vegetation mapping results, which can be denoted as:

$$TA = \frac{a}{N} \tag{3-6}$$

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \tag{3-7}$$

where *a* is the number of plots whose vegetation group agree with the vegetation mapping result, *N* is the total number of plots used for accuracy assessment; *Pr(a)* is the relative observed agreement, and *Pr(e)* is the hypothetical probability of chance agreement. The 95% confidence interval for the *TA* was calculated using the method provided by Foody (2009). About one-third of the plot measurements at each study site were used to calculate *TA* and κ. The vegetation group assignments for these test plots were determined by the

minimum Mahalanobis distance between these plots and the center of each vegetation group. The parameters used for calculating the Mahalanobis distance include the three forest structure parameters and the coordinates on the primary and secondary axes from the ordination analysis. The center for each vegetation group was calculated by the means of plots used to name vegetation groups. To minimize the influence of the different scales of parameters, all parameters were normalized before calculating the Mahalanobis distance.

3.3 Results

3.3.1 Optimized cluster number determination

In this study, the hierarchical cluster analysis result showed that there was never any support for more than 8 vegetation classes at either study site. Thus as a conservative starting point, we approximately doubled the estimate from preliminary results (i.e. 15 vegetation classes) and set it as the upper limit of the BIC cluster number determination algorithm. As shown in Table 3-2, all *dBIC* values for the Last Chance site were smaller than zero, and the cluster number was 14 when the *rBIC* was smaller than 0.04 for the first time. The initial cluster number was set as 14 for the Last Chance site. When the cluster number was smaller than 14, the two largest *R(k)* values were from results having two clusters and seven clusters. Due to the fact that the ratio between these two *R(k)* was smaller than 1.15, the final optimized cluster number for the Last Chance site was set to seven. Similarly, the final optimized cluster number for the Sugar Pine site was set to four. It should be noted that the initial cluster number for the Sugar Pine site was set to 15 (i.e., the pre-defined maximum cluster number) because all the *rBIC* values were larger than 0.04.

3.3.2 Vegetation mapping results

The vegetation mapping results for the Last Chance and Sugar Pine sites are shown in Fig. 3. Both sites were dominated by Sierran mixed conifer trees. Specifically, 56% of the Last Chance site was classified as the Mature Mixed Conifer Forest, 19% as Young Mixed Conifer Forest, and 12.6% as Mixed Conifer Woodland. The Young Mixed Conifer Forest was mainly scattered within the Mature Mixed Conifer Forest (Fig. 3-3a). Pine- and open true fir-dominated forest types were less abundant, covering 7.3% and 3.7% of the study area, respectively (Table 3-3). These forest types were mainly found at the north end of the study site, and their coverage increased with elevation (Fig. 1 and Fig. 3-3a). The proportion of the Low and High Shrub types were very small, both around 0.6%. At Sugar Pine, the Mature Mixed Conifer Forest again was the most common type occupying 57.1% of the landscape (Fig. 3-3b). Closed-canopy Mixed Conifer Forest was the next most common type, at 25.9% of area, with the greatest concentration in the middle of the study site. The Pine-Cedar Woodland and Open Pine-Oak Woodland were distributed at the southeast and northwest of the study site, occupying 13.8% and 3.2%, respectively.

Table 3-2 The optimized cluster number determination results using Bayesian information criterion (BIC) algorithm for the Last Chance and Sugar Pine study sites.

	Last Chance site					Sugar Pine site			
k	BIC	dBIC[a]	rBIC[b]	R(k)[c]	k	BIC	dBIC[a]	rBIC[b]	R(k)[c]
1	1390210.558				1	1478604.199			
2	1132790.048	-257420.510	1.000	1.942	2	1216658.927	-261945.272	1.000	2.154
3	1000327.759	-132462.289	.515	1.386	3	1095132.435	-121526.492	.464	1.177
4	904805.634	-95522.126	.371	1.236	4	991913.123	-103219.312	.394	1.920
5	827541.566	-77264.068	.300	1.139	5	938229.015	-53684.108	.205	1.055
6	759733.224	-67808.342	.263	1.502	6	887345.487	-50883.528	.194	1.140
7	714639.554	-45093.670	.175	1.774	7	842710.399	-44635.088	.170	1.237
8	689292.817	-25346.737	.098	1.139	8	806647.821	-36062.578	.138	1.461
9	667049.072	-22243.744	.086	1.119	9	782010.413	-24637.409	.094	1.047
10	647181.368	-19867.704	.077	1.318	10	758479.133	-23531.280	.090	1.029
11	632149.235	-15032.134	.058	1.165	11	735610.756	-22868.377	.087	1.009
12	619263.081	-12886.154	.050	1.098	12	712946.401	-22664.355	.087	1.008
13	607542.700	-11720.381	.046	1.407	13	690455.682	-22490.719	.086	1.321
14	599259.290	-8283.410	.032	1.162	14	673467.151	-16988.531	.065	1.058
15	592152.613	-7106.676	.028	1.014	15	657416.013	-16051.138	.061	1.005

[a]The changes (dBIC) are from the previous number of clusters in the table.
[b]The ratios of changes (rBIC) are relative to the change for the two cluster solution.
[c]The ratios of distance measures (R(k)) are based on the current number of clusters against the previous number of clusters.

Table 3-3 Forest structure parameters and dominant tree species for each vegetation group obtained from the k-means unsupervised classification procedure. The dominant tree species are evaluated by the relative basal area of each tree species. Note that certain tree species with too small relative basal areas for all groups (<1%) were not included in the table.

Group ID	Vegetation Type	Basal Area (m²/ha)	Lorey's Height (m)	Canopy Cover (%)	Dominant Tree Species[b] Relative Basal Area (%)								
					ABCO	ABMA	CADE	PILA	PIMO	PIPO	PSME	QUKE	LO
					Last Chance site								
G1	Low Shrub	N/A[a]	N/A[a]	N/A[a]	Manzanita (*Arctostaphylos spp.)*								
G2	High Shrub	N/A[a]	N/A[a]	N/A[a]	Manzanita (*Arctostaphylos spp.)*								
G3	Open True Fir	4.0	10.1	9.2	69	19	0	0	1	11	0	0	0
G4	Pine Woodland	11.2	13.2	21.8	15	5	0	22	0	41	17	0	0
G5	Mixed Conifer Woodland	20.3	15.6	36.4	44	2	8	5	0	21	18	3	0
G6	Young Mixed Conifer Forest	24.7	18.5	46.1	24	1	8	18	0	26	21	1	1
G7	Mature Mixed Conifer Forest	48.3	26.3	61.5	34	4	6	18	0	12	22	3	0
					Sugar Pine site								
G1	Open Pine-Oak Woodland	11.4	12.2	14.7	0	0	0	3	0	72	0	24	0
G2	Pine-Cedar Woodland	19.8	17.6	38.1	11	1	20	11	0	30	0	10	17
G3	Mature Mixed Conifer Forest	47.3	25.3	66.8	26	1	28	8	0	19	0	8	10
G4	Closed-canopy Mixed Conifer	68.0	32.4	74.6	40	1	29	13	0	9	0	5	2

[a]N/A means the value is not available for corresponding blank.
[b]Species code: ABCO, white fir (*Abies concolor*); ABMA, California red fir (*Abies magnifica*); CADE, incense-cedar (*Calocedrus decurrens*); PILA, sugar pine (*Pinus lambertiana*); PIMO, western white pine (*Pinus monticola*); PIPO, ponderosa pine (*Pinus ponderosa*); PSME, Douglas-fir (*Pseudotsuga menziesii*); QUKE, black oak (*Quercus kelloggii*); LO, canyon live oak (*Quercus chrysolepis*).
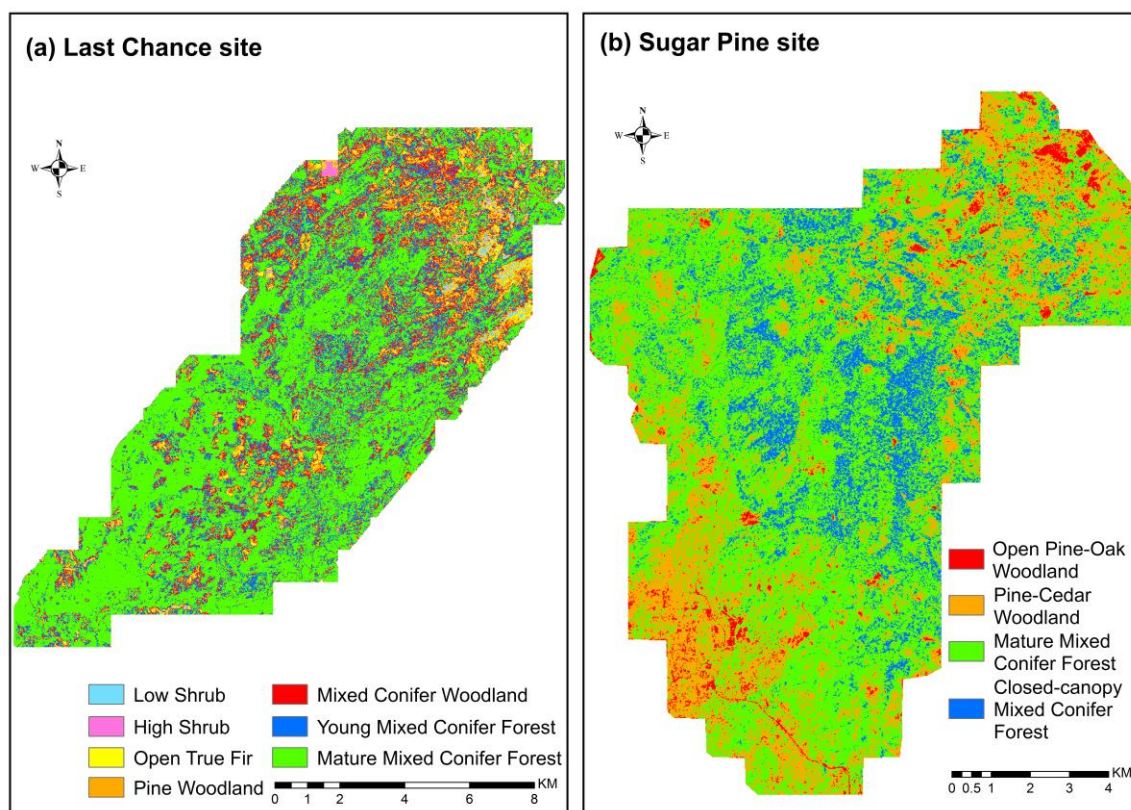
Figure 3-3 Labeled vegetation mapping results for the two study sites.

The forest vertical structure information and dominant tree species composition for each vegetation group are shown in Table 3-3. Naming conventions for the unsupervised groups were based on the dominant tree species (Table 3-3). If the tree species composition for two vegetation groups was similar, the name recognizes the differences in the forest structures. For example, at the Last Chance site, composition of the dominant tree species for Young Mixed Conifer Forest and Mature Mixed Conifer Forest are similar, but the Mature Mixed Conifer Forest has larger, taller trees and greater canopy cover (Table 3-3). Note there is no tree information for groups identified as Low Shrub and High Shrub, because there no trees were measured with a DBH of 5 cm or greater in these groups.

The capability of the proposed vegetation mapping strategy to differentiate among dominant species was evaluated by ordination analysis. In Figure 3-4, the first two axes for both study sites represent over 50% information of all data. The tree species composition between vegetation groups differ greatly with each other at the Last Chance site (Figure 3-4a). Although the tree species composition of Young Mixed Conifer Forest and Mature Mixed Conifer Forest are similar (Table 3-3), the proportion of white fir for the Mature Mixed Conifer Forest is larger than Young Mixed Conifer Forest, and that for ponderosa pine is smaller (Table 3-3). At the Sugar Pine site, the proportion of black oak trees for Open Pine-Oak Woodland is higher than other three vegetation groups, which makes it unique among all four vegetation groups (Figure 3-4b). The tree species compositions for the other three vegetation groups are similar, especially the Mature Mixed Conifer Forest and Closed-canopy Mixed Conifer Forest. The proportion of white fir and California red

fir for the Pine-Cedar Woodland is relatively smaller, compared to the Mature Mixed Conifer Forest and Closed-canopy Mixed Conifer Forest.
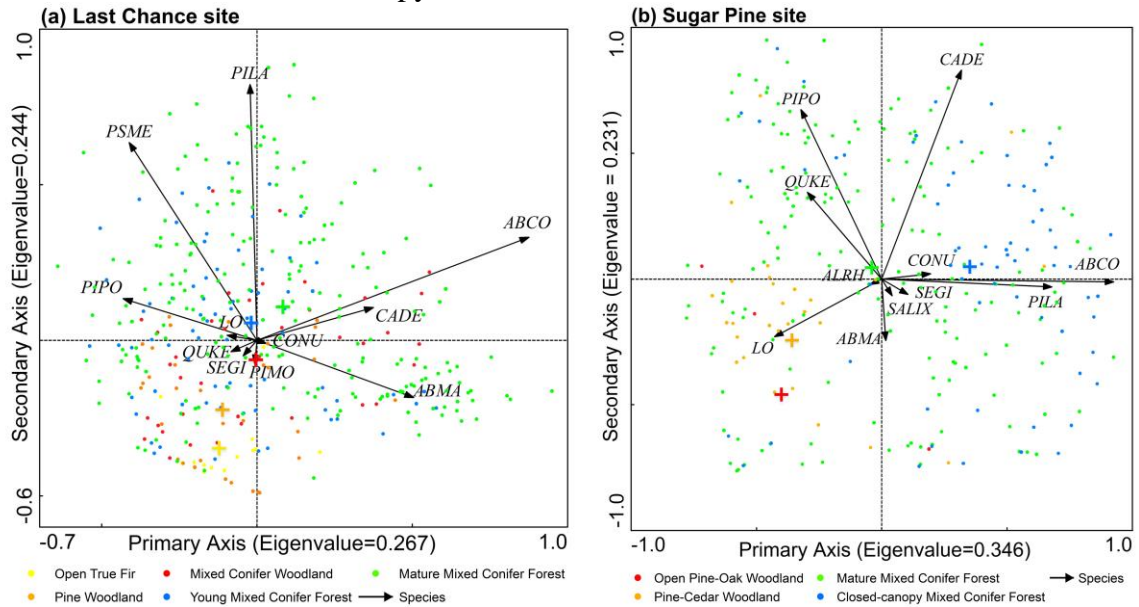
Figure 3-4 Ordination analysis results for the Last Chance and Sugar Pine sites. The "+" symbol in each color represents the centroid of the vegetation group represented by the corresponding color in each figure. Species code: ABCO, white fir (*Abies concolor*); ABMA, California red fir (*Abies magnifica*); ALRH, white alder (*Alnus rhombifolia*); CADE, incense-cedar (*Calocedrus decurrens*); CONU, mountain dogwood; LO, canyon live oak (*Quercus chrysolepis*); PILA, sugar pine (*Pinus lambertiana*); PIMO, western white pine (*Pinus monticola*); PIPO, ponderosa pine (*Pinus ponderosa*); PSME, Douglas-fir (*Pseudotsuga menziesii*); QUKE, black oak (*Quercus kelloggii*); SALIX, peachleaf willow (*Salix amygdaloides*); SEGI, giant sequoia (*Sequoiadendron giganteum*).

The capability of the proposed vegetation mapping strategy to differentiate the forest vertical structure characteristics was examined by permutation test under the null hypothesis that the means of vegetation vertical structure parameters between vegetation groups have no difference. Since there were no forest structure parameters for the plots within the Low Shrub and High Shrub groups at Last Chance, these two groups were excluded in the permutation test. At the Last Chance site, this null hypothesis is rejected for differences in parameters between all vegetation groups ($\alpha<0.05$), except the difference of Lorey's height between Open True Fir and Pine Woodland and that between Pine Woodland and Mixed Conifer Woodland (Table 3-4). For differences in Lorey's height between these two group combinations, the null hypothesis can still be rejected at the significant level of $\alpha=0.10$. At the Sugar Pine site, the variation in vegetation structure parameters among groups is not as pronounced as the Last Chance site. The vegetation parameters for the Closed-canopy Mixed Conifer Forest are the most distinct. The p-values for the differences in all three parameters between the Closed-canopy Mixed Conifer Forest and the other three vegetation groups are all smaller than 0.05 except for the difference in canopy cover with Mature Mixed Conifer Forest. The basal area and Lorey's height of the Mature Mixed Conifer Forest are significantly different from all other groups ($\alpha<0.05$).

However, its canopy cover has no significant difference from all other vegetation groups. The differences in all three parameters between Open Pine-Oak Woodland and Pine-Cedar Woodland are not significant.

The accuracy of the vegetation mapping results was evaluated by the independent plot measurements (Table 3-5). As can been seen, the overall accuracies of the vegetation mapping results are around 80% with a 95% confidence interval of ~8% for both study sites, and kappa coefficients are higher than 0.65. At Last Chance site, the commission errors and omission errors for most vegetation groups are lower than 20%, except the commission errors for the Mixed Conifer Woodland and Young Mixed Conifer Forest and the omission error for the Mixed Conifer Woodland. At Sugar Pine site, all commission and omission errors are lower than 30%, except the omission error for the Pine-Cedar Woodland. The omission rate of the Pine-Cedar Woodland is as high as 41%, and six out of seven omitted Pine-Cedar Woodland plots were misclassified as Mature Mixed Conifer Forest.

3.4 Discussion

Remote sensing technology has been shown to be extremely helpful for mapping and monitoring vegetation over large spatial scales (Xie et al. 2008). However, choosing a classification system that comprehensively capture vegetation community composition and structure is still a major challenge for vegetation mapping from remotely sensed data (Rapp et al. 2005). Traditionally, the number of vegetation units and/or the properties of vegetation units within a forest were pre-defined by the prior knowledge of experts from previous experience or field sampling data (Bork and Su 2007, Carpenter et al. 1999, Naidoo et al. 2012). However, this may lead to biased or inconsistent classification systems across regions and may not result in optimal breaks among different vegetation communities. Heinzel and Koch (2011) found that the accuracy of vegetation mapping can increase from 57% to 91% with corresponding decreases in the number of vegetation classes from six to two. It is critical to determine the optimal number of groups that balances the value of recognizing differences in vegetation structure and composition with the reliability of identifying these differences.

By combining the LiDAR data and high-resolution aerial image, this study used a novel automatic cluster number determination algorithm and k-means unsupervised classification to define an optimized classification system. The classification of each vegetation group was determined by fully considering both the vegetation structure characteristics and dominant tree species composition. The results at both study sites show that the proposed vegetation mapping strategy can differentiate vegetation groups by vegetation structure parameters or dominant species composition or both (Figure 3-3). At the Last Chance site, the small differences in the relative abundance of the common tree species were captured along with steep gradients in structure (Figure 3-4a, Table 3-3, and Table 3-4). Although the tree species composition for the Young Mixed Conifer Forest and Mature Mixed Conifer Forest were very similar, trees in Mature Mixed Conifer Forest were considerably larger than in Young Mixed Conifer Forest (Table 3-3). Similarly for the Low Shrub and High Shrub groups, which were both dominated by manzanita (*Arctostaphylos* spp.), the latter was about 30 cm higher on average than the former based on the LiDAR

Table 3-4 The p-values of permutation test for the differences in forest structure parameters between different vegetation groups in the Last Chance and Sugar Pine study sites.

	Basal Area					Lorey's Height					Canopy Cover				
	G3[a]	G4[a]	G5[a]	G6[a]	G7[a]	G3[a]	G4[a]	G5[a]	G6[a]	G7[a]	G3[a]	G4[a]	G5[a]	G6[a]	G7[a]
	Last Chance site														
G3[a]	1.000	0.004	0.001	0.000	0.000	1.000	0.066	0.009	0.001	0.000	1.000	0.000	0.000	0.000	0.000
G4[a]	0.004	1.000	0.002	0.000	0.000	0.066	1.000	0.078	0.001	0.000	0.000	1.000	0.000	0.000	0.000
G5[a]	0.001	0.002	1.000	0.013	0.000	0.009	0.078	1.000	0.037	0.000	0.000	0.000	1.000	0.001	0.000
G6[a]	0.000	0.000	0.013	1.000	0.000	0.001	0.001	0.037	1.000	0.000	0.000	0.000	0.001	1.000	0.000
G7[a]	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000
	G1[a]	G2[a]	G3[a]	G4[a]	N/A[b]	G1[a]	G2[a]	G3[a]	G4[a]	N/A[b]	G3[a]	G4[a]	G5[a]	G6[a]	G7[a]
	Sugar Pine site														
G1[a]	1.000	0.920	0.109	0.013	N/A[b]	1.000	0.806	0.245	0.014	N/A[b]	1.000	0.403	0.684	0.000	N/A[b]
G2[a]	0.920	1.000	0.000	0.000	N/A[b]	0.806	1.000	0.019	0.000	N/A[b]	0.403	1.000	0.613	0.000	N/A[b]
G3[a]	0.109	0.000	1.000	0.000	N/A[b]	0.245	0.019	1.000	0.000	N/A[b]	0.684	0.613	1.000	0.334	N/A[b]
G4[a]	0.013	0.000	0.000	1.000	N/A[b]	0.014	0.000	0.000	1.000	N/A[b]	0.000	0.000	0.334	1.000	N/A[b]

[a]G3 to G7 and G1 to G4 for the Last Chance site and Sugar Pine site represent the corresponding vegetation group listed in Table 3-3.
[b]N/A means values is not available for corresponding blank.

Table 3-5 The confusion matrices and accuracy assessments for the vegetation mapping results of Last Chance site and Sugar Pine site.

| | Last Chance Site | | | | | | | | | Sugar Pine site | | | | | |
| | Reference | | | | | | | Commission | Kappa | Reference | | | | Commission | Kappa |
Predicted	G1[a]	G2[a]	G3[a]	G4[a]	G5[a]	G6[a]	G7[a]	Error (%)	Coefficient	G1[a]	G2[a]	G3[a]	G4[a]	Error (%)	Coefficient
G1[a]	1	0	0	0	0	0	0	0.0		1	0	0	0	0	
G2[a]	0	1	0	0	0	0	0	0.0		0	10	1	0	9.1	
G3[a]	0	0	4	0	0	0	0	0.0		0	6	42	5	20.7	
G4[a]	0	0	0	11	2	0	0	15.4		0	1	6	18	28.0	
G5[a]	0	0	0	2	5	0	3	50	0.70	N/A[b]	N/A[b]	N/A[b]	N/A[b]	N/A[b]	0.64
G6[a]	0	0	0	0	3	12	6	42.9		N/A[b]	N/A[b]	N/A[b]	N/A[b]	N/A[b]	
G7[a]	0	0	0	0	1	3	46	8.0		N/A[b]	N/A[b]	N/A[b]	N/A[b]	N/A[b]	
Omission Error (%)	0.0	0.0	0.0	15.4	54.5	20.0	16.3	N/A[b]		0.0	41.2	14.3	21.7	N/A[b]	
Overall Accuracy (%)	80.0±7.9 (95% confidence interval)									78.9±8.3 (95% confidence interval)					

[a]G1 to G7 and G1 to G4 for the Last Chance site and Sugar Pine site represent the corresponding vegetation group listed in Table 3-3.
[b]N/A means values is not available for corresponding blank.

derived CHM. At Sugar Pine, the unsupervised classification clearly detected the pine-oak vegetation type from the matrix of mixed conifer forests (Figure 3-4b) as well as the structural gradient present (Table 3-3, Table 3-4).

Forest structure information, which has been difficult to incorporate in previous vegetation mapping strategy, is an important factor that has influence on various ecological applications (Peterson et al. 2005, Zielinski et al. 2006), and should be used in the procedure of developing vegetation maps for forest managements (Lindenmayer et al. 2000). This is particularly true in more compositionally homogeneous forests. In these forest, traditional vegetation mapping methods which rely on passive remote sensing data might miss the underlying structural differences within the forest. By including LiDAR data, the proposed vegetation mapping strategy can detect differences in vegetation vertical structure characteristics that in turn inform the assessment of wildlife habitat suitability, wildfire hazard, and water yield. For example, the Sugar Pine site is dominated by three vegetation types, Pine-Cedar Woodland, Mature Mixed Conifer Forest, and Closed-canopy Mixed Canopy Forest (Figure 3-3b), that have similar tree species composition (Table 3-3). Without considering forest vertical structure characteristics from LiDAR data, these three vegetation groups might be classified only as one larger group.

The field measurements of species composition and plot-level forest structure support the results obtained by the unsupervised classification strategy. The proposed vegetation mapping strategy can produce sufficiently high overall accuracies (nearly 80% in both cases) and kappa coefficients (over 0.64 in both sites) for most applications where the vegetation map provides the essential classification and scaling information. Moreover, the overall accuracy and kappa coefficient obtained from the proposed vegetation mapping strategy are comparable to most previous supervised vegetation mapping strategies integrating LiDAR data and multispectral imagery (Bork and Su 2007, Dalponte et al 2012, Cho et al. 2012).

Although the commission and omission errors for certain vegetation groups were high, they may be caused by misregistration between plot measurements and remotely sensed data (LiDAR data and aerial imagery). The plot locations were measured using a GPS in the field. Although it can produce centimeter level positioning accuracy in most cases, the blocking effect of forest canopy may reduce the GPS positioning accuracy significantly (Sigrist et al. 1999). The possible positioning error may lead to poor co-registration with remotely sensed data. Particularly, this misregistration may have a pronounced effect on the commission and omission errors of vegetation groups that do not cluster together. For example, the Young Mixed Conifer Forest in the Last Chance site had both a relatively high commission error and omission error. Instead of aggregating together, Young Mixed Conifer Forest was mainly scattered within Mature Mixed Conifer Forest (Figure 3-3a). A commission error of 66.7% for Young Mixed Conifer Forest was due to the misclassification as Mature Mixed Conifer Forest.

The quality of NAIP aerial imagery may be another factor that influences the vegetation mapping accuracy. As known, there is non-linear color balancing effect existed in the NAIP imagery due to the dynamic range of different image tiles and different data acquiring time (Hart and Veblen 2015). Moreover, the absolute horizontal accuracy for the NAIP imagery is around 6 m at a 95% confidence level (USDA Farm Service Agency, 2015). Although this study has tried to reduce the influence of misregistration between

NAIP imagery and LiDAR products by matching correspondence points, it still cannot be totally eliminated. Further study is still needed to address how the non-linear color balancing effect and horizontal accuracy influence the vegetation mapping accuracy. Moreover, it has been frequently reported that hyperspectral data outperformed multispectral data in recognizing plant species (Adam et al. 2010, Xu and Gong 2007), and there has been studies showing that the integration of hyperspectral data and LiDAR data can produce more accurate vegetation maps than the integration of multispectral data and LiDAR data (Dalponte et al 2012).

3.5 Conclusions

This study proposed a vegetation mapping strategy through the combination of multispectral aerial imagery and LiDAR data. Both the vegetation structure and composition information were taken into consideration of the determination of classification system. The BIC algorithm was used to automatically optimize the number of vegetation units within two mixed-conifer forests, and the property of each vegetation group was identified by post-hoc analysis based on field measurements. The results show that the proposed vegetation mapping strategy is a robust method to map vegetation in mixed-conifer forests with a sufficient high accuracy. The overall accuracy and kappa coefficient are over 78% and 0.64 for both study sites. Each identified vegetation group can be differentiated from others by vegetation structure parameters or dominant species composition or both. The obtained vegetation maps have the potential to considerably improve the identification of critical habitat for species of concern (e.g., Pacific fisher and California spotted owl), as well as identifying wildfire risk through characterizing ladder fuels (Kramer et al. 2014).

References

Adam, E., Mutanga, O., & Rugege, D. (2010). Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review. Wetlands Ecology and Management, 18, 281-296.

Agee, J.K., & Skinner, C.N. (2005). Basic principles of forest fuel reduction treatments. Forest Ecology and Management, 211, 83-96.

Alvarez, O., Guo, Q., Klinger, R.C., Li, W., & Doherty, P. (2013). Comparison of elevation and remote sensing derived products as auxiliary data for climate surface interpolation. International Journal of Climatology, 34, 2258-2268.

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, 355-366.

Anys, H., Bannari, A., He, D., & Morin, D. (1994). Texture analysis for the mapping of urban areas using airborne MEIS-II images. In Proceedings of the First International Airborne Remote Sensing Conference and Exhibition, Strasbourg, France, 231-245.

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

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

Chiu, T., Fang, D., Chen, J., Wang, Y., & Jeris, C. (2001). A robust and scalable clustering algorithm for mixed type attributes in large database environment. In Proceedings of the seventh ACM SIGKDD International Conference On Knowledge Discovery and Data Mining. ACM, 263-268.

Cho, M.A., Mathieu, R., Asner, G.P., Naidoo, L., van Aardt, J., Ramoelo, A., Debba, P., Wessels, K., Main, R., & Smit, I.P. (2012). Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system. Remote Sensing of Environment, 125, 214-226.

Chuvieco, E., & Congalton, R.G. (1988). Mapping and inventory of forest fires from digital processing of TM data. Geocarto International, 3, 41-53.

Clark, M.L., Clark, D.B., & Roberts, D.A. (2004). Small-footprint LiDAR estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. Remote Sensing of Environment, 91, 68-89.

Colgan, M.S., Baldeck, C.A., Féret, J.-B., & Asner, G.P. (2012). Mapping savanna tree species at ecosystem scales using support vector machine classification and BRDF correction on airborne hyperspectral and LiDAR data. Remote Sensing, 4, 3462-3480.

Daly, C., Neilson, R.P., & Phillips, D.L. (1994). A statistical-topographic model for mapping climatological precipitation over mountainous terrain. Journal of Applied Meteorology, 33, 140-158.

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.

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

Falkowski, M.J., Hudak, A.T., Crookston, N.L., Gessler, P.E., Uebler, E.H., & Smith, A.M. (2010). Landscape-scale parameterization of a tree-level forest growth model: a k-nearest neighbor imputation approach incorporating LiDAR data. Canadian Journal of Forest Research, 40, 184-199.

Foody, G.M. (2009). Sample size determination for image classification accuracy assessment and comparison. International Journal of Remote Sensing, 30, 5273-5291.

Franklin, J. (1986). Thematic Mapper analysis of coniferous forest structure and composition. International Journal of Remote Sensing, 7, 1287-1301.

Franklin, J. (1995). Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography, 19, 474-499.

García-Feced, C., Tempel, D.J., & Kelly, M. (2011). LiDAR as a tool to characterize wildlife habitat: California spotted owl nesting habitat as an example. Journal of Forestry, 109, 436-443.

Gauch, H.G. (1982). Multivariate analysis in community ecology. Cambridge University Press. New York, NY.

Graham, R.T., McCaffrey, S., & Jain, T.B. (2004). Science basis for changing forest structure to modify wildfire behavior and severity, General Technical Report, RMRS-120. USDA Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado, USA, 43.

Guo, Q., Li, W., Yu, H., & Alvarez, O. (2010). Effects of topographic variability and LiDAR sampling density on several DEM interpolation methods. Photogrammetric Engineering & Remote Sensing, 76, 701-712.

Haralick, R.M., Shanmugam, K., & Dinstein, I.H. (1973). Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics, SMC-3, 610-621.

Hart, S.J., & Veblen, T.T. (2015). Detection of spruce beetle-induced tree mortality using high-and medium-resolution remotely sensed imagery. Remote Sensing of Environment, 168, 134-145.

Hartigan, J.A. (1975). Clustering algorithms. John Wiley & Sons.

Heinzel, J., & Koch, B. (2011). Exploring full-waveform LiDAR parameters for tree species classification. International Journal of Applied Earth Observation and Geoinformation, 13, 152-160.

Hill, R., & Thomson, A. (2005). Mapping woodland species composition and structure using airborne spectral and LiDAR data. International Journal of Remote Sensing, 26, 3763-3779.

Hirano, A., Madden, M., & Welch, R. (2003). Hyperspectral image data for mapping wetland vegetation. Wetlands, 23, 436-448.

Holmgren, J., & Persson, Å. (2004). Identifying species of individual trees using airborne laser scanner. Remote Sensing of Environment, 90, 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, 1537-1552.

Huang, H., Li, Z., Gong, P., Cheng, X., Clinton, N., Cao, C., Ni, W., & Wang, L. (2011). Automated methods for measuring DBH and tree heights with a commercial scanning LiDAR. Photogrammetric Engineering & Remote Sensing, 77, 219-227.

Hummel, S., Hudak, A., Uebler, E., Falkowski, M., & Megown, K. (2011). A comparison of accuracy and cost of LiDAR versus stand exam data for landscape management on the Malheur National Forest. Journal of Forestry, 109, 267-273.

Jakubowksi, M.K., Guo, Q., Collins, B., Stephens, S., & Kelly, M. (2013). Predicting surface fuel models and fuel metrics using LiDAR and CIR imagery in a dense, mountainous forest. Photogrammetric Engineering & Remote Sensing, 79, 37-49.

Jensen, J.L., Humes, K.S., Vierling, L.A., & Hudak, A.T. (2008). Discrete return LiDAR-based prediction of leaf area index in two conifer forests. Remote Sensing of Environment, 112, 3947-3957.

Jensen, J.R., 2009. Remote sensing of the environment: An earth resource perspective (2nd edition). Pearson Education India. Upper Saddle River, NJ.

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

Knight, J.F., Lunetta, R.S., Ediriwickrema, J., & Khorram, S. (2006). Regional scale land cover characterization using MODIS-NDVI 250 m multi-temporal imagery: A phenology-based approach. GIScience & Remote Sensing, 43, 1-23.

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, 1065-1080.

Koukoulas, S., & Blackburn, G.A. (2005). Mapping individual tree location, height and species in broadleaved deciduous forest using airborne LIDAR and multi-spectral remotely sensed data. International Journal of Remote Sensing, 26, 431-455.

Kramer, H. A., Collins, B.M., Stephens, S.L., & Kelly, M. (2014). Quantifying ladder fuels: a new approach using LiDAR. Forests, 5, 1432-1453.

Lefsky, M.A., Cohen, W.B., Parker, G.G., & Harding, D.J. (2002). LiDAR Remote Sensing for Ecosystem Studies LiDAR, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. BioScience, 52, 19-30.

Lepš, J., & Šmilauer, P. (2003). Multivariate analysis of ecological data using CANOCO. Cambridge university press. Cambridge, UK.

Li, L., Ustin, S., & Lay, M. (2005). Application of multiple endmember spectral mixture analysis (MESMA) to AVIRIS imagery for coastal salt marsh mapping: a case study in China Camp, CA, USA. International Journal of Remote Sensing, 26, 5193-5207.

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

Lim, K.S., & Treitz, P.M. (2004). Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. Scandinavian Journal of Forest Research, 19, 558-570.

Lindenmayer, D.B., Margules, C.R., & Botkin, D.B. (2000). Indicators of biodiversity for ecologically sustainable forest management. Conservation biology, 14, 941-950.

Liu, D., Kelly, M., & Gong, P. (2006). A spatial–temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery. Remote Sensing of Environment, 101, 167-180.

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

Lucas, R.M., Cronin, N., Lee, A., Moghaddam, M., Witte, C., & Tickle, P. (2006). Empirical relationships between AIRSAR backscatter and LiDAR-derived forest biomass, Queensland, Australia. Remote Sensing of Environment, 100, 407-425.

Mallinis, G., Koutsias, N., Tsakiri-Strati, M., & Karteris, M. (2008). Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a Mediterranean test site. ISPRS Journal of Photogrammetry and Remote Sensing, 63, 237-250.

Mermoz, M., Kitzberger, T., & Veblen, T.T. (2005). Landscape influences on occurrence and spread of wildfires in Patagonian forests and shrublands. Ecology, 86, 2705-2715.

McCune, B., Grace, J. B., & Urban, D. L (2002). Analysis of ecological communities (vol. 28). MjM software design. Gleneden Beach, OR.

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

Naesset, E. (1997). Estimating timber volume of forest stands using airborne laser scanner data. Remote Sensing of Environment, 61, 246-253.

Naidoo, L., Cho, M., Mathieu, R., & Asner, G. (2012). Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment. ISPRS Journal of Photogrammetry and Remote Sensing, 69, 167-179.

Nichols, T.E., & Holmes, A.P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. Human Brain Mapping, 15, 1-25.

Nilsson, M. (1996). Estimation of tree heights and stand volume using an airborne LiDAR system. Remote Sensing of Environment, 56, 1-7.

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, 1163-1174.

Pearce, J., Cherry, K., & Whish, G. (2001). Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution. Journal of Applied Ecology, 38, 412-424.

Pedrotti, F. (2012). Plant and Vegetation Mapping. Springer, New York, USA.

Peterson, D.L., Johnson, M.C., Agee, J.K., Jain, T.B., McKenzie, D., & Reinhardt, E.D. (2005). Forest structure and fire hazard in dry forests of the western United States, General Technical Report, PNWGTR-268. USDA Forest Service, Pacific Northwest Research Station, Portland, Oregon, USA.

Pohl, C., & Van Genderen, J.L. (1998). Review article multisensor image fusion in remote sensing: concepts, methods and applications. International Journal of Remote Sensing, 19, 823-854.

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

Rapp, J., Wang, D., Capen, D., Thompson, E., & Lautzenheiser, T. (2005). Evaluating error in using the National Vegetation Classification System for ecological community mapping in northern New England, USA. Natural Areas Journal, 25, 46-54.

Reitberger, J., Krzystek, P., & Stilla, U. (2008). Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. International Journal of Remote Sensing, 29, 1407-1431.

Riaño, D., Valladares, F., Condés, S., & Chuvieco, E. (2004). Estimation of leaf area index and covered ground from airborne laser scanner (LiDAR) in two contrasting forests. Agricultural and Forest Meteorology, 124, 269-275.

Sigrist, P., Coppin, P., & Hermy, M. (1999). Impact of forest canopy on quality and accuracy of GPS measurements. International Journal of Remote Sensing, 20, 3595-3610.

Soh, L.-K., & Tsatsoulis, C. (1999). Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. IEEE Transactions on Geoscience and Remote Sensing, 37, 780-795.

Stephens, S.L. (1998). Evaluation of the effects of silvicultural and fuels treatments on potential fire behaviour in Sierra Nevada mixed-conifer forests. Forest Ecology and Management, 105, 21-35.

Su, Y., & Guo, Q. (2014). A practical method for SRTM DEM correction over vegetated mountain areas. ISPRS Journal of Photogrammetry and Remote Sensing, 87, 216-228.

Su, Y., Guo, Q., Ma, Q., & Li, W. (2015). SRTM DEM Correction in Vegetated Mountain Areas through the Integration of Spaceborne LiDAR, Airborne LiDAR, and Optical Imagery. Remote Sensing, 7, 11202-11225.

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

Talbot, S., & Markon, C. (1988). Intermediate-scale vegetation mapping of Innoko National Wildlife Refuge, Alaska using Landsat MSS digital data. Photogrammetric engineering & remote sensing, 54, 377-383.

Thomas, V., Treitz, P., McCaughey, J., & Morrison, I. (2006). Mapping stand-level forest biophysical variables for a mixedwood boreal forest using LiDAR: an examination of scanning density. Canadian Journal of Forest Research, 36, 34-47.

USDA Farm Service Agency, 2015. NAIP Imagery (online). Accessed 3 May 2015. Available from http://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/index.

Wallerman, J., & Holmgren, J. (2007). Estimating field-plot data of forest stands using airborne laser scanning and SPOT HRG data. Remote Sensing of Environment, 110, 501-508.

Walter, V. (2004). Object-based classification of remote sensing data for change detection. ISPRS Journal of photogrammetry and remote sensing, 58, 225-238.

Wang, L., Sousa, W., & Gong, P. (2004). Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery. International Journal of Remote Sensing, 25, No. 24, 5655-5668.

Xie, Y., Sha, Z., & Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review. Journal of Plant Ecology, 1, 9-23.

Xu, B., & Gong, P. (2007). Land-use/land-cover classification with multispectral and hyperspectral EO-1 data. Photogrammetric Engineering and Remote Sensing, 73, 955-965.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C., Gao, F., Reed, B.C., & Huete, A. (2003). Monitoring vegetation phenology using MODIS. Remote Sensing of Environment, 84, 471-475.

Zielinski, W.J., Truex, R.L., Dunk, J.R., & Gaman, T. (2006). Using forest inventory data to assess fisher resting habitat suitability in California. Ecological Applications, 16, 1010-1025.

Chapter 4   Fine-resolution forest tree height estimation across the Sierra Nevada

Abstract

Forests of the Sierra Nevada (SN) mountain range are valuable natural heritages for the region and the country, and tree height is an important forest structure parameter for understanding the SN forest ecosystem. There is still a need in the accurate estimation of wall-to-wall SN tree height distribution at fine spatial resolution. In this study, we presented a method to map wall-to-wall forest tree height (defined as Lorey's height) across the SN at 70 m resolution by fusing multi-source datasets, including over 1600 in-situ tree height measurements and over 1600 km^2 airborne light detection and ranging (LiDAR) data. Accurate tree height estimates within these airborne LiDAR boundaries were first computed based on in-situ measurements, and then these airborne LiDAR derived tree heights were used as reference data to estimate tree heights at Geoscience Laser Altimeter System (GLAS) footprints. Finally, the random forest algorithm was used to model the SN tree height from these GLAS tree heights, optical imagery, topographic data, and climate data. The results show that our fine-resolution SN tree height product has a good correspondence with field measurements. The coefficient of determination between them is 0.60, and the root-mean-squared error is 5.45 m.

Keywords: Tree height; Sierra Nevada; LiDAR; Integration; Fine-resolution.

4.1 Introduction

Forests of the Sierra Nevada (SN) mountain range, covering an area of 63 100 km^2, are valuable natural heritages for the region and the country. As one of the most diverse temperate conifer forests on Earth, the SN serves a series of ecosystem functions for the region and even the United States (US), e.g. regulating functions by maintaining essential ecological processes and life support systems, supporting functions by providing habitats for wild plants and animals in the region, provisioning functions by supplying provisions of natural resources, and cultural functions by offering life fulfillment opportunities and cognitive developments (Bales et al., 2011). In recent decades, forests of the SN have been extensively exploited and managed to deal with the increasing wildfire risk induced by fire suppression, fuel build up, and climate change (Stephens, Millar, and Collins, 2010). These forest management activities have brought growing concerns that they may destroy wildlife habitats, especially the habitats of endangered species (Berigan et al., 2012; Matthews et al., 2013). Tree height, an important forest structure parameter, plays a crucial role in understanding the ecosystem of a forest and therefore modelling the wildlife habitat within the forest (Martinuzzi et al., 2009; Vierling et al., 2008).

Traditionally, forest tree heights are randomly sampled by taking accurate in-situ measurements (Naesset, 1997; Næsset and Økland, 2002). Numerous efforts from US Forest Service (USFS) have been devoted to obtaining ground measured tree height in the SN forests, e.g. the Forest Inventory and Analysis (FIA) Program and the Sierra Nevada Adaptive Managements Program (SNAMP). However, these ground inventories are usually taken at the plot scale, and therefore cannot be used to estimate wall-to-wall tree

height distribution alone. Moreover, taking ground measurements is highly time-consuming, labor-intensive and costly, and may be impossible for certain forested areas due to the inaccessibility. Optical remote sensing can obtain spatially continuous land surface observations at reasonable costs and feasible efforts, which provides an indirect way to map tree height (Donoghue and Watt, 2006; Hall et al., 2006; Kellndorfer et al., 2012; Lu et al., 2004; Zhang et al., 2014). Nevertheless, due to the limited penetration ability through the forest canopy, optical sensors have strong saturation effects in dense forested areas, which makes the retrieved tree height estimations from optical sensors usually fraught with uncertainty issues (Donoghue and Watt, 2006; McCombs et al., 2003; Su et al., 2015).

Light detection and ranging (LiDAR), an active remote sensing technique, can penetrate the forest canopy effectively due to the use of a focused wavelength laser pulse (Popescu, Wynne, and Nelson, 2002; Coops et al., 2007; Su and Guo, 2014). It has been proven that it can be used to obtain three-dimensional structure information of forests (Morsdorf et al., 2006; Nelson et al., 1984; Popescu et al., 2011; Zhao et al., 2012; Zhao and Popescu, 2009). The airborne LiDAR system, one of the most frequently used LiDAR platforms, has been widely used to estimate tree height in various forest types with high accuracy (Andersen et al., 2006; Clark et al., 2004; Naesset, 1997; Nilsson, 1996; Wing et al., 2015). However, currently, there is no spatial wall-to-wall coverage of airborne LiDAR data across the SN because of the high cost of flight missions to acquire airborne LiDAR data. The Geoscience Laser Altimeter System (GLAS) onboard the Ice, Cloud, and land Elevation Satellite (ICESat), the only available spaceborne LiDAR data in global scale, provides an alternative data source for estimating forest tree height. For example, Duncanson et al. (2010), Lefsky et al. (2005) and Lefsky et al. (2007) demonstrated the feasibility of using GLAS full-waveform LiDAR metrics to calculate forest tree heights. However, the footprints of GLAS data are spaced at 170 m along tracks and tens of kilometres across tracks, which are too sparse to be used to generate spatial continuous fine-resolution tree height products.

Recently, by taking advantages of GLAS data, airborne LiDAR data and optical imagery, there have been studies focusing on estimating forest tree heights through the integration of these data. For example, Leckie et al. (2003) and Suárez et al. (2005) developed methods to estimate tree heights from optical imagery and airborne LiDAR data in a Douglas-fir (*Pseudotsuga menziesii*) forest and spruce (*Picea*) forest, respectively; Lefsky (2010) segmented the global forests into over 4.4 million patches using Moderate Resolution Imaging Spectroradiometer (MODIS) data, and estimated the forest tree height for each forest patch using GLAS data; Simard et al. (2011) estimated the forest canopy height (RH100) values from GLAS full-waveform records, and extrapolated the GLAS RH100 estimations to global scale based on forest types, tree cover data, elevation data and climate data. However, current forest tree height products covering the SN forests (e.g., tree height products from Lefsky (2010) and Simard et al. (2011)) are still too coarse to be used for evaluating the influence of forest management on wildlife behaviours.

This study aims to develop a method to estimate the tree height distribution across the SN at fine resolution through the combination of ICESat/GLAS data, airborne LiDAR, optical imagery, and other ancillary datasets (e.g., vegetation map, climate data and topographic data). To address this objective, we collected over 1600 in-situ plot

measurements and over 1500 km^2 airborne LiDAR data across the SN forests. A procedure integrating the step-wise regression, multiple-linear regression and regression tree methods was developed to calculate tree heights within airborne LiDAR footprints, compute tree heights at the GLAS footprints, and extrapolate tree height estimations from GLAS footprints to the whole SN forests. The resulting SN tree height product can be downloaded from the web (http://guolabLiDAR.com/SNtreeheight).

4.2 Data and Methodology
4.2.1 Study area
The SN is a mountain range located in the western US, which runs ~640 km from north to south and ~110 km from west to east (Figure 4-1a). The elevation increases gradually from the northwest to the southeast, and the mean elevation is over 1500 m; the slope, with a mean of 12°, can reach up to 81°, and is generally higher in the southern SN than in the northern SN. The lower elevation western SN foothills are covered by Mediterranean forests, woodlands and shrub woodlands, and higher elevation mountainous areas are temperate coniferous forests. The region is composed of six sub-regions, namely North Sierra, North Central Sierra, Central Sierra, South Central Sierra, South Sierra, and East Sierra (Figure 4-1).
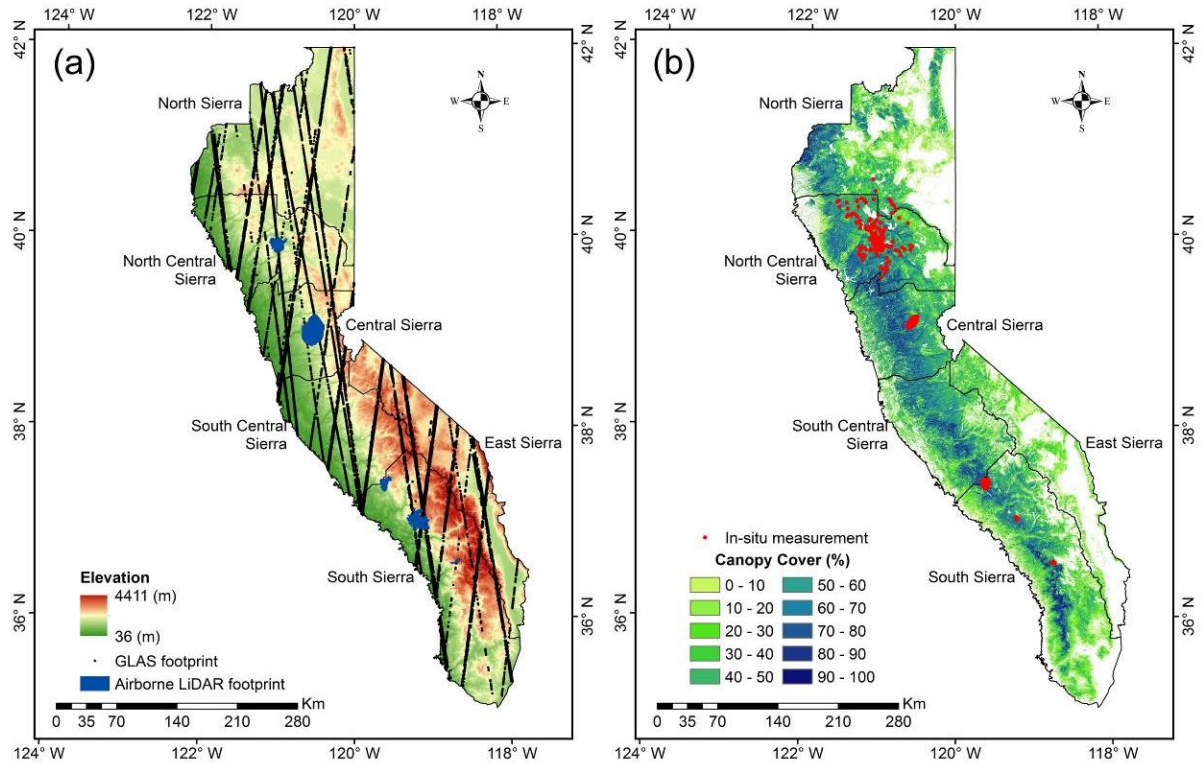
Figure 4-1 (a) The location and elevation of the Sierra Nevada (SN), and the distributions of Geoscience Laser Altimeter System (GLAS) footprints and airborne light detection and ranging (LiDAR) data within the SN. (b) The distributions of canopy cover and in-situ measurements within the SN.

4.2.2 In-situ measurements

Field observed tree heights are fundamental for calculating and validating tree height estimations from remotely sensed data. In this study, we collected 1671 plot measurements from SNAMP, Critical Zone Observations (CZO), and US Forest Service (USFS), covering an area of over 6000 km^2 (Figure 4-1b). All plot measurements from SNAMP and CZO were taken under the same protocol in the growing season of 2012 and 2013. Each plot, 12.62 m in radius (~500 m^2 in area), was randomly selected in the corresponding sites of these projects. The height and diameter at breast height (DBH) for each individual tree with a DBH >5 cm were collected in the field. The plot measurements from USFS were taken under the national FIA protocol during the growing season between 2004 and 2009 (Woodall and Monleon-Moscardo, 2008). These plots were originally set to study the behaviour of the California spotted owl (*Strix occidentalis occidentalis*), an endangered wildlife species in the SN. Each plot location was randomly selected and had to be potentially suitable for the owl to nest or forage (Su et al., 2015). Overall, there were 1008 FIA subplot-level plots (7.32 m in radius) collected in this study, and 596 of those plots were within the airborne LiDAR footprint (Table 4-1). The tree height and DBH for each individual tree with a DBH lager than 12.7 cm were recorded in the field.

Table 4-1 Collected field tree height measurements and airborne light detection and ranging (LiDAR) data in this study.

Data source	Airborne LiDAR		Field measurements			
	Year	Coverage (km^2)	Year	Number of plots within LiDAR footprint	Number of plots outside of LiDAR footprint	Plot size* (m)
SNAMP	2012&2013	850	2013	606	0	12.62
CZO	2009	58	2009	57	0	12.62
USFS	2009-2012	676	2004-2009	596	412	7.32

* The size of circular plot is presented by the length of radius.

4.2.3 Spaceborne LiDAR data and pre-processing

GLAS on-board the NASA ICESat, which was first launched in January 2003 and retired in February 2010, provided global-scale full-waveform LiDAR measurements. Its footprints, in a nominal diameter of 65 m, were spaced at 170 m along tracks and tens of kilometres across tracks (Schutz et al., 2005). In this study we collected all available GLA01 and GLA06 products between 2003 and 2009 within the SN from the ICESat/GLAS data pool (http://nsidc.org/data/icesat/order.html). The GLA01 product provided the full-waveform measurements, and the GLA06 product provided the geolocation and data quality for each full-waveform record. The GLAS footprint samples within the SN were further filtered to ensure that 1) they were taken under cloud-free conditions; 2) they had no saturation effect; 3) they had high signal to noise ratio (i.e., >50); and 4) they were not significantly higher (i.e., 100 m) than the land surface presented by the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM). Overall, there are 27 214 GLAS full-waveform samples retained within the SN (Figure 4-1). For each GLAS sample, we calculated three parameters from the full-waveform information,

namely waveform extent, leading edge extent, and trailing edge extent. These three parameters have been proved to be highly correlated to canopy height, canopy height variability, and terrain slope, respectively (Baghdadi et al., 2014; Bourgine and Baghdadi, 2005; Lefsky, 2010; Lefsky et al., 2007). Definitions of these three parameters can be found in Lefsky et al. (2007), and will not be discussed in detail here.

4.2.4 Airborne LiDAR data and pre-processing

In this study, the airborne LiDAR data, covering around 1600 km^2, were collected from the SNAMP, CZO, and USFS (Figure 4-1). These airborne LiDAR data were all acquired during the growing season during the years of 2009-2013 (Table 4-1). The average point density for each airborne LiDAR footprint ranges from 4 pts/m^2 to 12 pts/m^2. Within each airborne LiDAR footprint, a 1 m resolution DEM was first interpolated from the LiDAR ground returns using ordinary kriging, since it outperformed other interpolation schemes (e.g., inverse distance weighted and spline) for generating DEM from LiDAR derived elevation points (Clark et al., 2004; Guo et al., 2010). Then, the elevation of each LiDAR point was normalized by the obtained DEMs. Finally, canopy quantile metrics, representing the height below X% of the LiDAR point cloud, were generated from these normalized LiDAR point clouds. Canopy quantile metrics are frequently used to compute forest structure parameters that cannot be measured directly from LiDAR point cloud (Lim and Treitz, 2004; Thomas et al., 2006). In this study, 11 canopy quantile metrics, including 0%, 1%, 5%, 25%, 50%, 75%, 90%, 95%, 99%, and 100%, were generated for each LiDAR footprint at 20 m resolution, which is roughly correspondent to the plot size.

4.2.5 Wall-to-wall tree height estimation method

As can be seen in Figure 4-2, a three-step procedure was used to estimate the wall-to-wall tree height across the SN: 1) estimate tree height within the airborne LiDAR footprint using the step-wise regression method; 2) build regression model to estimate tree heights at GLAS footprints; and 3) extrapolate the tree height estimations from GLAS footprints to the whole SN forests using a regression tree method. It should be noted that tree height mentioned hereafter refers to Lorey's height (HL), which can be calculated from in-situ individual tree height (H) and basal area (B) measurements,

$$H_L = \frac{\sum_{i=1}^{n} B_i \times H_i}{\sum_{i=1}^{n} B_i} \tag{4-1}$$

where i is the ith tree within each plot, and n is the total number of trees recorded for each plot.
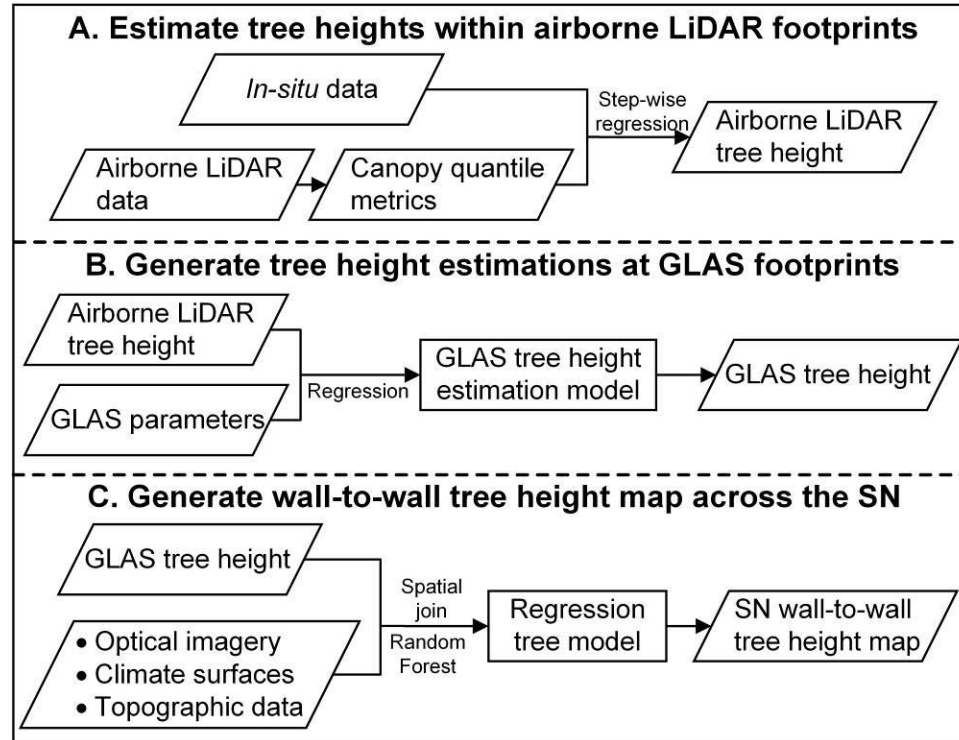
Figure 4-2 Flow chart of the proposed SN tree height estimation procedure.

Since the in-situ measurements can hardly be spatially coincident with GLAS footprints directly, the airborne LiDAR derived tree heights were used as the media to link ground measurements with GLAS data (Tang et al., 2014). Previous studies have shown that multiple linear step-wise regression method is an effective method to estimate Lorey's height from airborne LiDAR canopy quantile metrics and inventory data (Andersen et al., 2005; Hall et al., 2005). In this study, within each airborne LiDAR footprint, two-thirds of the corresponding plot measurements were randomly selected, and used as training data to build the regression model for estimating tree height at the airborne LiDAR footprints. These airborne LiDAR derived tree height estimations were then used to spatially link GLAS measurements, and therefore built a multiple linear regression model to estimate tree height from GLAS parameters (Lefsky et al., 2007). This regression model was applied to all GLAS data to calculate tree heights at GLAS footprints. It should be noted that all airborne LiDAR derived tree heights were resampled from 20 m to 70 m resolution using the weighted mean value method in this step (Jakubowksi et al., 2013) so that they can roughly match the footprint size of GLAS measurements.

The regression tree method Random Forest (RF) was employed to extrapolate the tree height estimations from GLAS footprints to the whole SN forests. Saatchi et al. (2001) suggested that the normality of co-variables used in the forest parameter prediction models is often violated when complex environmental and ecological parameters are introduced. RF is a formalized non-parametric machine learning algorithm, which does not require the assumption to be made regarding the normality of covariables (Breiman, 2001). Moreover, previous studies have shown that the RF algorithm is robust in modelling forest parameters

(e.g., tree height and aboveground biomass) from GLAS data and other remotely sensed imagery (Baccini et al., 2008; Simard et al., 2011; Su et al., 2016; Wilkes et al., 2015).

In this study, 14 ancillary datasets were included in the RF regression tree model, primarily representing vegetation, climate and topographic conditions (Table 4-2). All available Landsat TM Land Surface Reflectance images covering the SN from the growing season of 2009 were collected. These images were further visually examined to make sure the cloud and snow coverage was lower than 10%. The maximum value composite (MVC) was then calculated from those retained Landsat TM images, and the six spectral bands of the MVC (the thermal band was excluded) were used in the RF regression model. In addition, we also computed the normalized vegetation difference index (NDVI) from the MVC and included it in the RF regression tree model. All covariables were interpolated into 70 m resolution using bilinear interpolation method to match the GLAS footprint size. Note that the slope product was generated from the interpolated SRTM DEM product using ESRI^{TM} ArcMap software.

Table 4-2 Ancillary datasets used in the Random Forest regression tree procedure to map forest tree height across the Sierra Nevada.

Dataset	Year	Resolution	Data Source
Landsat TM Land Surface Reflectance product	2009	30 m	US Geological Survey
NDVI calculated from Landsat TM images	2009	30 m	N/A*
Tree canopy cover	2011	30 m	National Land Cover Database
Annual Mean Temperature	1950-2010	70 m	Alvarez et al. (2014)
Annual Temperature Seasonality	1950-2010	70 m	Alvarez et al. (2014)
Annual Total Precipitation	1950-2010	70 m	Alvarez et al. (2014)
Annual Temperature Seasonality	1950-2010	70 m	Alvarez et al. (2014)
SRTM Elevation dataset	2000	30 m	US Geological Survey
Slope calculated from SRTM data	2000	70 m	N/A*

* "N/A" represents the corresponding dataset is calculated from other collected dataset.

The RF extrapolation procedure was implemented using the randomForest R package (Liaw and Wiener, 2002), which included both the classification and regression functions. There are two primary user defined parameters, number of trees and number of variables tried at each split (Liaw and Wiener, 2002). In this study, these two parameters were determined by manual examination. The number of trees and number of variables tried at each split were increased by one iteratively, and the values that produced the minimum mean-squared error were selected to extrapolate the GLAS tree height measurements. Here, based on manual iterative examination, 500 "RF trees" were included and four variables were tried at each split. The built RF regression tree model was used to map the forest tree height across the SN. This initial SN tree height result was further masked by tree canopy cover product to obtain the final SN tree height product (Figure 4-2). In areas without tree coverage (tree canopy cover = 0%), tree height values were set to 0 m.

The large amount of covariables used in the extrapolation process may result in the RF algorithm overfitting the model. To reduce this possibility, we further examined the importance of all collected variables, which was evaluated by the percentage increase in the mean-squared error (Liaw and Wiener, 2002). The larger the value was, the more important the corresponding variable was. As can be seen in Figure 4-3, all Landsat TM MVC spectral bands (except band 4) had relatively lower importance to the RF regression model. Therefore, these five bands were not used in the above mentioned RF regression model.
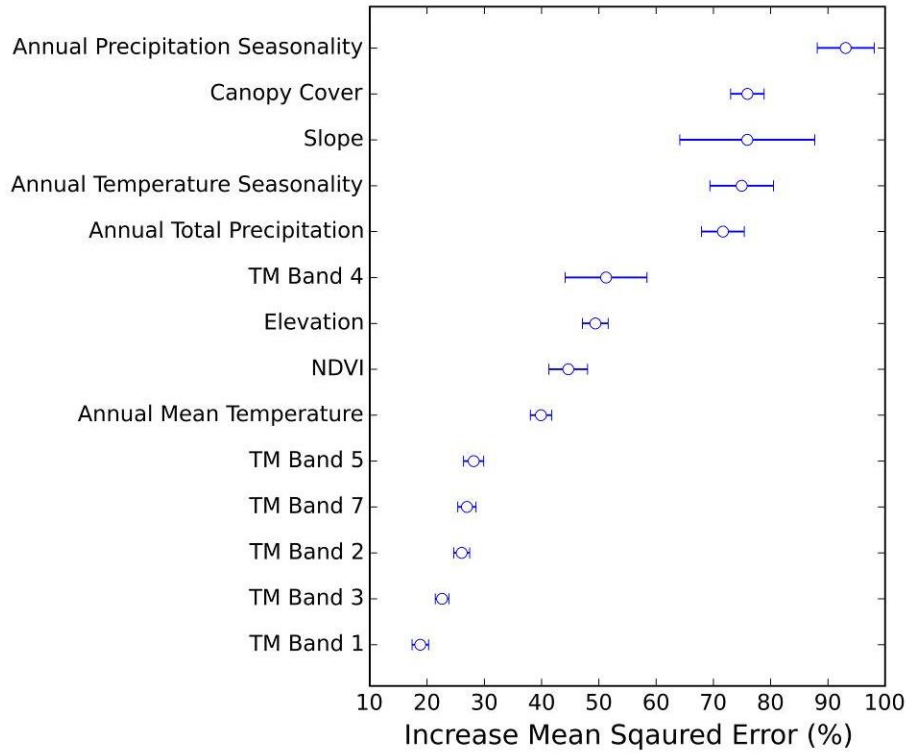
Figure 4-3 Importance of variables, denoted by percentage increase of mean squared error, for the forest tree height estimation Random Forest model.

4.2.6 Accuracy assessment

The accuracy of the estimated wall-to-wall SN tree height product was evaluated using the collected plot measurements, airborne LiDAR derived tree height, and GLAS derived tree height, respectively. Two commonly used statistical parameters, adjusted coefficient of determination (denoted as R^2) and root-mean-squared error (RMSE), were calculated from the following equations,

$$R^2 = 1 - \frac{(n-1)\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{(n-p)\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (4-2)$$

$$RMSE = \sqrt{\dfrac{\sum\limits_{i=1}^{n}(x_i - \hat{x}_i)^2}{n - p}} \qquad (4\text{-}3)$$

where n is the number of samples, p is the degree of freedom (i.e., number of parameters), x_i is the ith original value, and $\hat{x}_i$ is the ith estimated value. Moreover, the test statistics were performed on all regression analyses at the significance level (α) of 0.01.

To further evaluate the accuracy of our product, this study collected other available tree height products covering the SN and compared our result with these products. Overall, we found five published works from Simard et al. (2011), Los et al. (2012), Lefsky (2010), Kellndorfer et al. (2012) and Zhang et al. (2014) covering the SN. Simard et al. (2011) estimated the global forest canopy height (RH100) distribution at 1 km resolution from GLAS data, MODIS data, climate surfaces, and topographic data; Los et al. (2012) estimated the forest tree height between 60° N and 60° S by decomposing GLAS waveforms and aggregated them into a spatial continuous layer at 0.5°; Lefsky (2010) segmented the global forests into over 4.4 million patches using MODIS data, and estimated the tree height value for each patch from GLAS data; Kellndorfer et al. (2012) estimated the US nation-wide tree height (Lorey's height) distribution at 30 m resolution using FIA plot measurements, SRTM data, and three-season Landsat ETM+ data; Zhang et al. (2014) estimated the tree height distribution in California at 30 m resolution using GLAS data and Landsat TM derived leaf area index data. Unfortunately, the tree height products from Lefsky (2010) and Zhang et al. (2014) are not available to us. The data downloading link provided by Lefsky (2010) in the paper is not accessible anymore. Therefore, the current study only compared our result with the products from Simard et al. (2011), Los et al. (2012), and Kellndorfer et al. (2012). To match the spatial resolution of our result and products from Simard et al. (2011) and Los et al. (2012), we resampled our SN tree height map into 1 km resolution and 0.5° resolution respectively by averaging values of all 70 m cells within each coarser resolution cell; to match the spatial resolution with the product from Kellndorfer et al. (2012), we resampled their product into 70 m resolution using a similar procedure.

4.3 Results
4.3.1 Tree height estimations within airborne LiDAR footprints
The Lorey's height distribution within each airborne LiDAR footprint was estimated using the step-wise regression method from corresponding plot measurements and canopy quantile metrics. To evaluate the accuracy of airborne LiDAR derived tree heights, one third of the plot measurements within each airborne LiDAR footprint were retained from the regression analysis. Figure 4-4 shows the accuracy assessment result of airborne LiDAR derived tree heights using all evaluation plot samples. As can be seen, airborne LiDAR derived tree heights show good agreements with field measurements. The R^2 is higher than 0.7 and the RMSE is around 4.7 m. However, it still tends to slightly overestimate the tree heights at areas with low trees, and underestimate tree heights at areas with relatively tall trees.
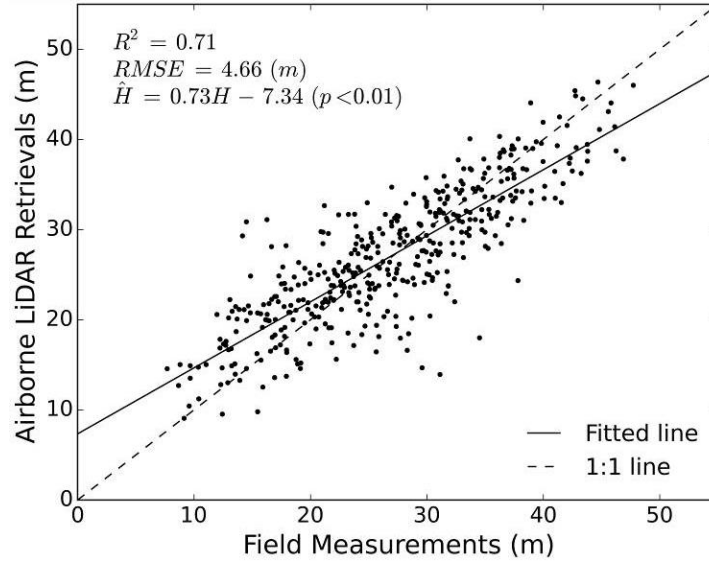
Figure 4-4 Evaluation of the airborne LiDAR retrieved Lorey's height using the validation plots. Note that R^2 represents the adjusted coefficient of determination, *RMSE* represents the root-mean-square error, *H* represents the Lorey's height calculated from field measurements, and $\hat{H}$ represents the airborne LiDAR retrieved Lorey's height.

4.3.2 Tree height estimations at GLAS footprints

By linking GLAS parameters with airborne LiDAR derived tree heights, this study estimated forest tree heights at GLAS footprints using the step-wise regression method. The waveform extent, leading edge extent, trailing edge extent, and slope were used to build the regression model. The following model combining waveform extent, leading edge extent and trailing edge extents was selected by the stepwise regression method,

$$LH_{GLAS} = -0.163WE + 0.404LE - 10.820\tan(slope) + 14.808 \quad (p < 0.01) \quad \text{(4-4)}$$

where LH_{GLAS} is the estimated Lorey's height at GLAS footprint, *WE* is the waveform extent, and *LE* is the leading edge extent. The R^2 for the regression model is 0.71, and all coefficients and the constant in the model are statistically significant. This regression model was used to compute tree heights for GLAS footprints outside of airborne LiDAR boundaries.

Figure 4-5 shows the distributions of GLAS estimated tree heights against different topography, vegetation, and climate conditions. As can be seen from Figure 4-5a and b, topographic parameters (slope and elevation) have similar relationships with GLAS tree heights. The GLAS tree height first increases with both slope and elevation, and then decreases after reaching certain values (40° for slope and 3000 m for elevation respectively). Canopy cover and NDVI share a similar positive influence on GLAS tree heights (Figure 4-5c and d). The higher the canopy cover or NDVI is, the higher the tree height is. There is one exception for the influence of NDVI on GLAS tree height. After NDVI reaches 0.9, the GLAS tree height has a slight decrease compared to previous NDVI group. Annual total precipitation and annual precipitation seasonality have weak positive

correlations with GLAS tree heights (Figure 4-5e and f). Once the annual total precipitation reaches 1000 mm or the annual precipitation seasonality reaches 70% (except the group with values larger than 90%), the average GLAS tree height become stable. The exception for the group with an annual precipitation seasonality larger than 90% may be caused by the relatively small number of GLAS footprints within this group (Appendix A, Figure A1). Annual mean temperature is negatively correlated with the GLAS tree height (Figure 4-5g), and annual temperature seasonality has no correlation with the GLAS tree height (Figure 4-5h).

4.3.3 Extrapolated wall-to-wall SN tree height product
The wall-to-wall 70 m resolution forest tree height map across the SN is shown in Figure 4-6. Overall, the built RF regression tree model can explain 62.63% of the variation in GLAS tree heights. The average tree height of the whole SN is 14.86 m and the standard deviation is 11.11 m. For forested areas (with a canopy cover larger than 0%), the average tree height is 21.88 m and the standard deviation is 5.32 m.

To evaluate the estimated SN tree height product, we compared the obtained product with plot measurements, airborne LiDAR derived tree height, and GLAS tree height, respectively. As shown in Figure 4-7a, the estimated SN tree height product is in good correspondence with field measurements (R^2=0.60, RMSE=5.45m). Moreover, although the fitted line is close to the 1:1 line, the estimated SN tree height product tends to overestimate tree heights in areas with relatively low trees (<33 m), and the lower the tree is, the more pronounced the overestimation effect is. In areas with relatively higher trees, the estimated SN tree height product tends to slightly underestimate tree heights. A pixel level comparison between the estimated SN tree height product and the airborne LiDAR derived tree heights was shown in Figure 4-7b. The difference between the estimated tree height and airborne LiDAR derived tree height follows a normal distribution with a mean of -5.17 m and a standard deviation of 10.24 m. Around 61% of differences between them are within the range of ±10 m, and around 34% of those are within the range of ±5 m. In general, the estimated SN tree height product tends to be lower than the airborne LiDAR derived tree heights. Around 72% pixels of the airborne LiDAR derived tree heights are higher than the SN tree height product.

The accuracy of the modelled SN tree height product was further evaluated by comparing it to GLAS tree heights. Table 4-3 shows the statistics of differences between the estimated SN tree height product and GLAS tree heights for each sub-region. The differences range approximately between -10 m and 10 m (except the North Central Sierra and South Sierra regions), and the average differences for all sub-regions are near 0 m. The standard deviation of differences is lower than 2.4 m for all sub-regions, except the South Sierra region, which is the highest among all sub-regions. Moreover, we further examine the difference between the estimated SN tree height product and GLAS tree heights within 0.25° cells (Figure 4-8). The model RMSE for each cell was calculated from all footprint level GLAS tree heights within the cell. It should be noted that there were five cells were not covered by any GLAS footprints, and their model RMSE values were interpolated from their neighbouring cells using bilinear interpolation method. As shown in Figure 4-8, the model RMSE ranges from 0.3 m to 15.6 m, and about a half of the 0.25°cells are smaller

than 6.5m. Most of cells with large RMSE values are concentrated in the East Sierra and South Sierra regions.
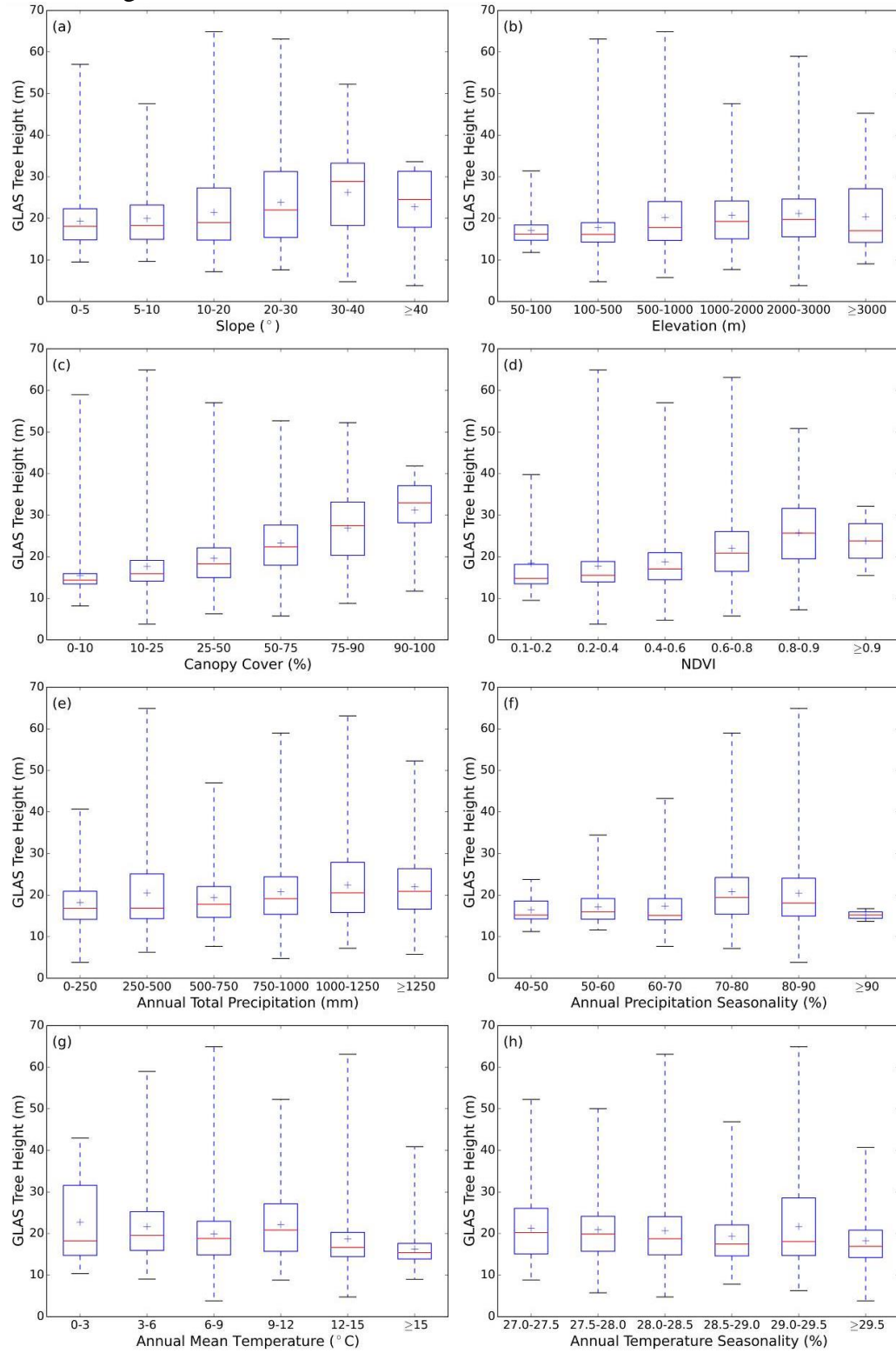
Figure 4-5 Boxplots of GLAS derived tree height corresponding to different terrain, vegetation, and climate parameters. The blue "+" indicates the mean GLAS derived tree height of corresponding group.
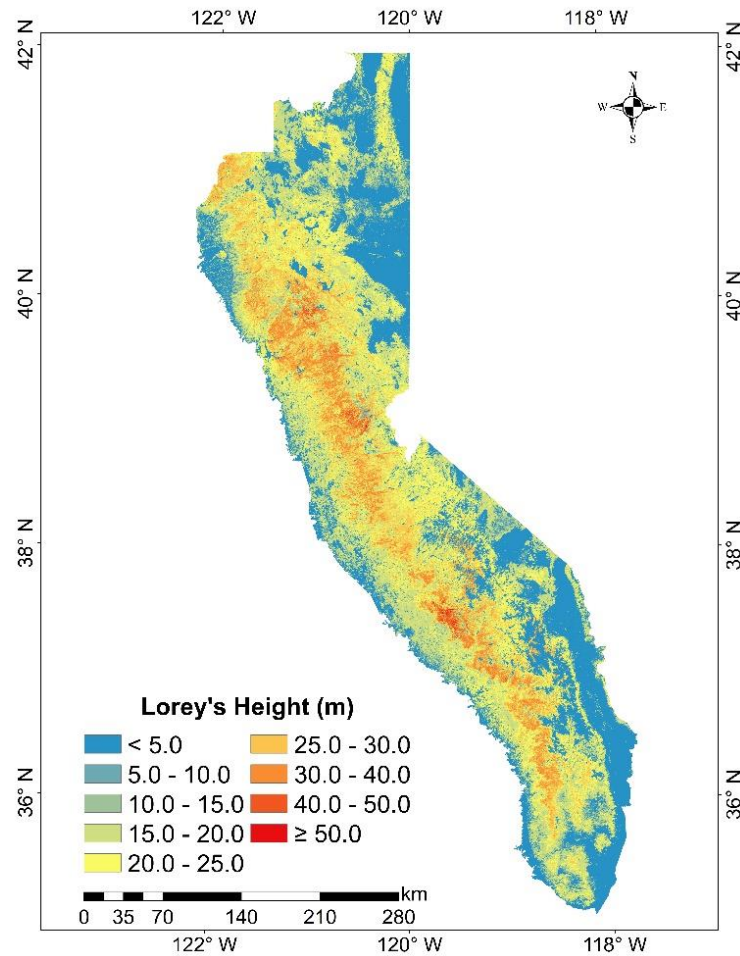
Figure 4-6 Estimated forest tree height (Lorey's height) distribution across the SN.
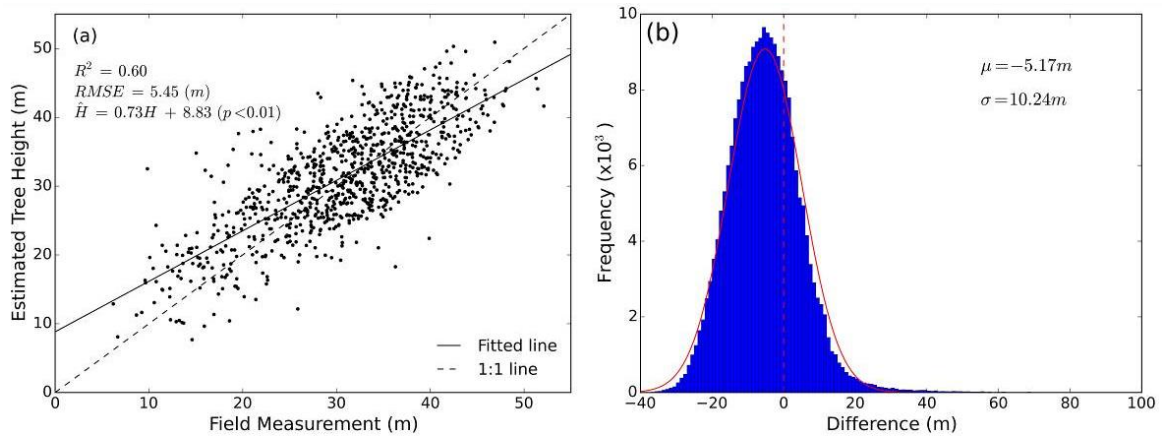
Figure 4-7 (a) Scatter plot between estimated tree heights ($\hat{H}$) and field measured tree heights (H). (b) Histogram of differences between estimated tree height and airborne

LiDAR derived tree height (estimated tree height minus airborne LiDAR derived tree height). Note that μ and σ represent the mean and the standard deviation of the differences, respectively.

Table 4-3 Statistics for differences between GLAS-derived tree height and estimated forest tree height products at sub-region scale.

Sub-region	Number of GLAS Shots	Minimum (m)	Maximum (m)	Average (m)	Standard Deviation (m)
North Sierra	6320	-7.98	10.69	-0.06	1.33
North Central Sierra	4251	-12.63	20.43	-0.10	2.37
Central Sierra	3155	-11.57	11.45	-0.14	2.22
South Central Sierra	4412	-9.09	13.83	-0.07	1.94
South Sierra	5284	-19.47	36.80	0.07	4.03
East Sierra	3792	-8.58	11.16	0.00	1.40
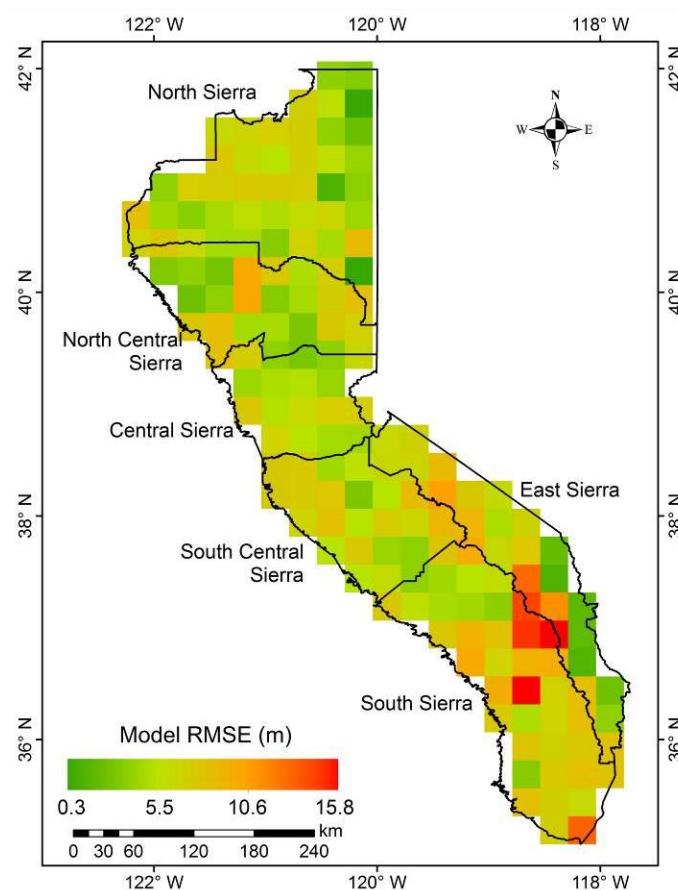
Figure 4-8 RMSE of the estimated forest tree height with respect to that calculated from GLAS shots in 0.25° cells.

4.3.4 Comparison with published products

Figure 4-9 shows the pixel-level differences between our estimated SN tree height result and the products from Simard et al. (2011), Los et al. (2012), and Kellndorfer et al. (2012). The mean difference between our result and the product from Simard et al. (2011) is -1.1 m, and the standard deviation is 9.0 m. The differences for over 77% of the 1 km cells are within ±10 m. Due to the coarse resolution of the result from Los et al. (2012), there are only 41 cells within the SN (Figure 4-9b). In general, their result tends be taller than ours in areas with low trees, and be lower than ours in areas with tall trees. The differences for 16 out of the 41 cells are within the range from -5 m to 5 m, and over a half are within the range from -10 m to 10 m. The tree height product from Kellndorfer et al. (2012) also tends to be lower than our product (Figure 4-9c). The mean difference between our product and the product from Kellndorfer et al. (2012) is 3.6 m, and the standard deviation is 7.3 m. Over 81% pixels from our SN tree height product have larger values than the product from Kellndorfer et al. (2012).
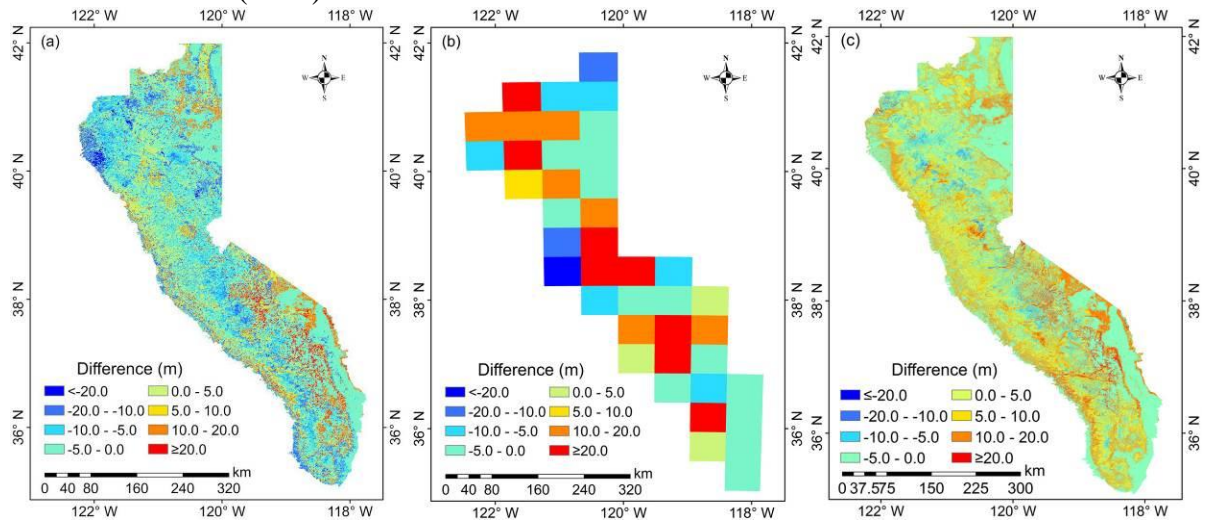
Figure 4-9 Pixel level comparison between our estimated SN tree height product and tree height products from (a) Simard et al. (2011), (b) Los et al. (2012), and (c) Kellndorfer et al. (2012). The pixel-level differences are calculated by our SN tree height product minus other products.

4.4 Discussion

This study developed a procedure to map the fine-resolution tree height distribution across the SN through the combination of ground inventory, airborne LiDAR, and GLAS data. The footprint level GLAS tree heights estimated from GLAS parameters and airborne LiDAR derived tree heights provided the foundation to build the RF regression tree model, and therefore estimate wall-to-wall tree height distribution from ancillary datasets. As mentioned in section 4.3.2, different topographic, vegetation, and climate parameters have different influences on the SN tree height distribution (Figure 4-5). With the increase of both slope and elevation, the GLAS tree height increases first and then decreases after they reach to certain thresholds (Figure 4-5a and b). This may be caused by the tree species distribution pattern in the SN. The SN oak woodlands are generally distributed in foothill

areas with low elevation (<1200 m) and slope values, and the height of oak trees are usually lower than 15m; pine forests and mixed conifer forest mainly spread out in areas with steep slope and elevation from 1200 m to around 2200 m, and individual tree height is usually taller than 15m; in areas with elevation higher than ~2300m, the conditions for tree growth become harsher and the main vegetation group is upper montane forest (Barbour and Billings, 2000).Similar to previous studies, we found that NDVI had saturation effect when being used to model forest canopy parameters, because of the limited penetration capability of optical sensors (Lu, 2006; Lu et al., 2004; Wang et al., 2005). Moreover, both the annual total precipitation and annual precipitation seasonality have a positive influence on the SN tree height, but the influence may become saturated when values reach certain thresholds. The exception for the group with an annual precipitation seasonality higher than 90% may be caused by the fact that too few GLAS footprints fall in this group. Annual mean temperature has a unique negative correlation with the SN tree height. This may be caused by the reason that the temperature decreases with the increase of elevation, and most tall coniferous trees are distributed in relatively high elevation areas in the SN (Barbour and Billings, 2000).

Although our SN tree height product shows good spatial correspondence with the other three published products, there are still differences. In general, our result shows more variations in tree height than the product from Simard et al. (2011). In areas with relatively steeper slopes (>15°), tree height from our result tends to be taller than that from Simard et al. (2011). One possible reason for this is that they did not include terrain slope in their tree height mapping procedure, and the GLAS waveform parameters are highly influenced by the terrain slope (Lefsky et al., 2007). Along the eastern foothills of the SN, our result also tends to be higher than the tree height from Simard et al. (2011) (Figure 4-9a). Moreover, although both our SN tree height product and the product from Kellndorfer et al. (2012) are Lorey's height, our tree height product is systematically higher. This may be caused by the fact that Kellndorfer et al. (2012) only used the Landsat ETM+ data and SRTM data to map the forest tree height. As mentioned, model-based tree height measurements from optical sensors usually suffer from the saturation effects, especially in areas with high values (Donoghue and Watt, 2006; McCombs et al., 2003; Su et al., 2015).

Several issues may have complicated the comparison analysis in this study. First of all, the definition of tree height among these three products are different. The tree height mentioned in this study refers to the Lorey's height. However, the tree height from Simard et al. (2011) is defined by the GLAS waveform shape parameter RH100, and that from Los et al. (2012) is calculated by decomposing the GLAS waveform, which are both expected to be higher than Lorey's height. This may be also one of the reasons that the differences for around 60% of the 1 km cells between our product and the product from Simard et al. (2011) are negative. Moreover, the spatial resolution of these three products are significantly different. The resampling process may further increase the uncertainty in the comparison results.

The accuracy of the estimated SN tree height product was evaluated by ground plot measurements, airborne LiDAR derived tree heights, and footprint level GLAS tree heights, respectively. These accuracy assessment results indicate that our SN tree height product can be used to reflect the spatial distribution of canopy height in the SN. Although the RMSE of the product is around 5.5 m by comparing with ground measurements, and may

even reach over 10 m by comparing with airborne LiDAR tree height at pixel level, the map from this study provides one of the best options to understand and explore forest tree height variation across the SN at fine resolution. However, this study still has limitations. Firstly, the time span of the data used in this study is very wide. The GLAS data were acquired between 2003 and 2009, and the airborne LiDAR data and ground measurements were taken from 2004 to 2013. Wensel, Meerschaert, and Biging (1987) found that without human interference, the conifer trees in the northern California can increase around 3-5 m every five years. Therefore, the influence of tree growth during the time span of airborne LiDAR data and in-situ measurements may not be neglectable. Moreover, forest changes brought by both nature incidents (e.g. wildfires) and human activities (e.g. forest fuel treatment, deforestation, and reforestation) may also result in significant changes in tree heights (Su et al., 2016). Secondly, the current accuracy assessment using independent plot measurements may be biased due to the mismatch between the plot size and the cell size of the estimated product. In the current study, we do not have access to in-situ measurements with bigger plot size, and we cannot produce the tree height map with finer resolution due to the limitation of the GLAS plot size. This issue can be addressed by collecting more airborne LiDAR data in the Sierra Nevada. With enough airborne LiDAR data, we can directly derive the tree height without using GLAS data, which allows us to map the SN tree height with finer (30 m) resolution. Finally, this study has not performed a fine-resolution uncertainty analysis on the obtained result using error propagation method. Future study is still needed to analyze how the uncertainty from different prediction variables influence the tree height estimation result.

4.5 Conclusions

This study presented a method to map fine-resolution forest tree height across the SN by combining plot measurements, airborne LiDAR data, GLAS data, optical imagery, topographic data, and climate surfaces. Over 1600 in-situ field measurements and airborne LiDAR data covering around 1600 km^2 were collected to address this mission. The resulting SN tree height map at 70 m resolution is available at the web (http://guolabLiDAR.com/SNtreeheight). Across the SN, the average tree height is 14.86 m and the standard deviation is 11.11 m; for only forested areas, the average tree height is 21.88 m and the standard deviation is 5.32 m. The estimated SN tree height product shows good agreements with in-situ plot measurements (R^2=0.60, RMSE=5.45m). The mean difference between the estimated SN tree height product and airborne LiDAR derived tree heights is -5.17 m, and the estimated product tends to be lower than airborne LiDAR derived tree heights. This fine resolution SN tree height map can be used to reveal the canopy vertical structure characteristics of the SN forests.

References

Alvarez, O., Guo, Q., Klinger, R.C., Li, W., & Doherty, P. (2014). Comparison of elevation and remote sensing derived products as auxiliary data for climate surface interpolation. International Journal of Climatology, 34, 2258-2268.

Andersen, H.-E., McGaughey, R.J., & Reutebuch, S.E. (2005). Estimating forest canopy fuel parameters using LiDAR data. Remote Sensing of Environment, 94, 441-449.

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, 355-366.

Baccini, A., Laporte, N., Goetz, S., Sun, M., & Dong, H. (2008). A first map of tropical Africa's above-ground biomass derived from satellite imagery. Environmental Research Letters, 3, 045011.

Baghdadi, N., Le Maire, G., Fayad, I., Bailly, J.S., Nouvellon, Y., Lemos, C., & Hakamada, R. (2014). Testing different methods of forest height and aboveground biomass estimations from ICESat/GLAS data in Eucalyptus plantations in Brazil. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7, 290-299.

Bales, R.C., Battles, J.J., Chen, Y., Conklin, M.H., Holst, E., O'Hara, K.L., Saksa, P., & Stewart, W. (2011). Forests and water in the Sierra Nevada: Sierra Nevada watershed ecosystem enhancement project. In, Sierra Nevada Research Institute report. Merced, California, USA.: Sierra Nevada Research Institute.

Barbour, M.G., & Billings, W.D. (2000). North American terrestrial vegetation. Cambridge University Press.

Berigan, W.J., Gutiérrez, R., & Tempel, D.J. (2012). Evaluating the efficacy of protected habitat areas for the California spotted owl using long-term monitoring data. Journal of Forestry, 110, 299-303.

Bourgine, B., & Baghdadi, N. (2005). Assessment of C-band SRTM DEM in a dense equatorial forest zone. Comptes Rendus Geoscience, 337, 1225-1234.

Breiman, L. (2001). Random Forests. Machine learning, 45, 5–32.

Clark, M.L., Clark, D.B., & Roberts, D.A. (2004). Small-footprint LiDAR estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. Remote Sensing of Environment, 91, 68-89.

Coops, N.C., Hilker, T., Wulder, M.A., St-Onge, B., Newnham, G., Siggins, A., & Trofymow, J.T. (2007). Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. Trees, 21, 295-310.

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

Duncanson, L., Niemann, K., & Wulder, M. (2010). Estimating forest canopy height and terrain relief from GLAS waveform metrics. Remote Sensing of Environment, 114, 138-154.

Guo, Q., Li, W., Yu, H., & Alvarez, O. (2010). Effects of topographic variability and LiDAR sampling density on several DEM interpolation methods. Photogrammetric Engineering & Remote Sensing, 76, 701-712.

Hall, R., Skakun, R., Arsenault, E., & Case, B. (2006). Modeling forest stand structure attributes using Landsat ETM+ data: Application to mapping of aboveground biomass and stand volume. Forest Ecology and Management, 225, 378-390.

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

Jakubowksi, M.K., Guo, Q., Collins, B., Stephens, S., & Kelly, M. (2013). Predicting surface fuel models and fuel metrics using LiDAR and CIR imagery in a dense, mountainous forest. Photogrammetric Engineering & Remote Sensing, 79, 37-49.

Kellndorfer, J., Walker, W., LaPoint, E., Bishop, J., Cormier, T., Fiske, G., Hoppus, M., Kirsch, K., & Westfall, J. NACP Aboveground Biomass and Carbon Baseline Data (NBCD 2000), USA, 2000, Dataset; 2012. In

Leckie, D., Gougeon, F., Hill, D., Quinn, R., Armstrong, L., & Shreenan, R. (2003). Combined high-density LiDAR and multispectral imagery for individual tree crown analysis. Canadian Journal of Remote Sensing, 29, 633-649.

Lefsky, M.A. (2010). A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. Geophysical Research Letters, 37, L15401.

Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Del Bom Espirito-Santo, F., Hunter, M.O., & de Oliveira, R. (2005). Estimates of forest canopy height and aboveground biomass using ICESat. Geophysical Research Letters, 32

Lefsky, M.A., Keller, M., Pang, Y., De Camargo, P.B., & Hunter, M.O. (2007). Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. Journal of Applied Remote Sensing, 1, 013537.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2, 18-22.

Lim, K.S., & Treitz, P.M. (2004). Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. Scandinavian Journal of Forest Research, 19, 558-570.

Los, S., Rosette, J., Kljun, N., North, P., Chasmer, L., Suárez, J., Hopkinson, C., Hill, R., Gorsel, E.v., & Mahoney, C. (2012). Vegetation height and cover fraction between 60 S and 60 N from ICESat GLAS data. Geoscientific Model Development, 5, 413-432.

Lu, D. (2006). The potential and challenge of remote sensing-based biomass estimation. International Journal of Remote Sensing, 27, 1297-1328.

Lu, D., Mausel, P., Brondızio, E., & Moran, E. (2004). Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. Forest Ecology and Management, 198, 149-167.

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

Matthews, S.M., Higley, J.M., Finn, J.T., Rennie, K.M., Thompson, C.M., Purcell, K.L., Sweitzer, R.A., Haire, S.L., Sievert, P.R., & Fuller, T.K. (2013). An evaluation of a weaning index for wild fishers (Pekania [Martes] pennanti) in California. Journal of Mammalogy, 94, 1161-1168.

McCombs, J.W., Roberts, S.D., & Evans, D.L. (2003). Influence of fusing LiDAR and multispectral imagery on remotely sensed estimates of stand density and mean tree height in a managed loblolly pine plantation. Forest Science, 49, 457-466.

Morsdorf, F., Kötz, B., Meier, E., Itten, K., & Allgöwer, B. (2006). Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. Remote Sensing of Environment, 104, 50-61.

Naesset, E. (1997). Determination of mean tree height of forest stands using airborne laser scanner data. ISPRS Journal of Photogrammetry and Remote Sensing, 52, 49-56.

Næsset, E., & Økland, T. (2002). Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. Remote Sensing of Environment, 79, 105-115.

Nelson, R., Krabill, W., & MacLean, G. (1984). Determining forest canopy characteristics using airborne laser data. Remote Sensing of Environment, 15, 201-212.

Nilsson, M. (1996). Estimation of tree heights and stand volume using an airborne LiDAR system. Remote Sensing of Environment, 56, 1-7.

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

Popescu, S.C., Zhao, K., Neuenschwander, A., & Lin, C. (2011). Satellite LiDAR vs. small footprint airborne LiDAR: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. Remote Sensing of Environment, 115, 2786-2797.

Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., & Hagen, S. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. Proceedings of the National Academy of Sciences, 108, 9899-9904.

Schutz, B., Zwally, H., Shuman, C., Hancock, D., & DiMarzio, J. (2005). Overview of the ICESat mission. Geophysical Research Letters, 32

Simard, M., Pinto, N., Fisher, J.B., & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne LiDAR. Journal of Geophysical Research: Biogeosciences, 116, G04021.

Stephens, S.L., Millar, C.I., & Collins, B.M. (2010). Operational approaches to managing forests of the future in Mediterranean regions within a context of changing climates. Environmental Research Letters, 5, 024003.

Su, Y., & Guo Q. (2014). A practical method for SRTM DEM correction over vegetated mountain areas. ISPRS Journal of Photogrammetry and Remote Sensing, 87, 216-228.

Su, Y., Guo, Q., Xue, B., Hu, T., Alvarez, O., Tao, S., & Fang, J. (2016). Spatial distribution of forest aboveground biomass in China: Estimation through combination of spaceborne LiDAR, optical imagery, and forest inventory data. Remote Sensing of Environment, 173, 187-199.

Su, Y., Guo, Q., Fry, D., Collins, B., Kelly, M., Flanagan, J., & Battles, J. (2016). A vegetation mapping strategy for conifer forests by combining airborne LiDAR data and aerial imagery. Canadian Journal of Remote Sensing, 42, 1-15.

Su, Y., Guo, Q., Ma, Q., & Li, W. (2015). SRTM DEM correction in vegetated mountain areas through the integration of spaceborne LiDAR, airborne LiDAR, and optical imagery. Remote Sensing, 7, 11202-11225.

Suárez, J.C., Ontiveros, C., Smith, S., & Snape, S. (2005). Use of airborne LiDAR and aerial photography in the estimation of individual tree heights in forestry. Computers & Geosciences, 31, 253-262.

Tang, H., Brolly, M., Zhao, F., Strahler, A.H., Schaaf, C.L., Ganguly, S., Zhang, G., & Dubayah, R. (2014). Deriving and validating Leaf Area Index (LAI) at multiple spatial

scales through LiDAR remote sensing: A case study in Sierra National Forest, CA. Remote Sensing of Environment, 143, 131-141.

Thomas, V., Treitz, P., McCaughey, J., & Morrison, I. (2006). Mapping stand-level forest biophysical variables for a mixedwood boreal forest using LiDAR: an examination of scanning density. Canadian Journal of Forest Research, 36, 34-47.

Vierling, K.T., Vierling, L.A., Gould, W.A., Martinuzzi, S., & Clawges, R.M. (2008). LiDAR: shedding new light on habitat characterization and modeling. Frontiers in Ecology and the Environment, 6, 90-98.

Wang, Q., Adiku, S., Tenhunen, J., & Granier, A. (2005). On the relationship of NDVI with leaf area index in a deciduous forest site. Remote Sensing of Environment, 94, 244-255.

Wensel, L.C., Meerschaert, W.J., & Biging, G.S. (1987). Tree height and diameter growth models for northern California conifers. Agriculture and Natural Resources Publ., University of Calif.

Wilkes, P., Jones, S., Suarez, L., Mellor, A., Woodgate, W., Soto-Berelov, M., Haywood, A., & Skidmore, A. (2015). Mapping forest canopy height across large areas by upscaling ALS estimates with freely available satellite data. Remote Sensing, 7, 12563.

Wing, B.M., Ritchie, M.W., Boston, K., Cohen, W.B., & Olsen, M.J. (2015). Individual snag detection using neighborhood attribute filtered airborne LiDAR data. Remote Sensing of Environment, 163, 165-179.

Woodall, C., & Monleon-Moscardo, V.J. (2008). Sampling protocol, estimation, and analysis procedures for the down woody materials indicator of the FIA program. United States Department of Agriculture, Forest Service, Northern Research Station.

Zhang, G., Ganguly, S., Nemani, R.R., White, M.A., Milesi, C., Hashimoto, H., Wang, W., Saatchi, S., Yu, Y., & Myneni, R.B. (2014). Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data. Remote Sensing of Environment, 151, 44-56.

Zhao, F., Guo, Q., & Kelly, M. (2012). Allometric equation choice impacts LiDAR-based forest biomass estimates: A case study from the Sierra National Forest, CA. Agricultural and Forest Meteorology, 165, 64-72.

Zhao, K., & Popescu, S. (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, 1628-1645.

Chapter 5   Forest aboveground biomass estimation at national to global scales

Abstract

The global forest ecosystem, which acts as a large carbon sink, plays an important role in modeling the global carbon balance. An accurate estimation of the total forest carbon stock in the aboveground biomass (AGB) is therefore necessary for improving our understanding of carbon dynamics, especially against the background of global climate change. However, because of limitations in forest AGB mapping methods and the availability of ground inventory data, there is still a lack in global wall-to-wall forest AGB estimation map. In this study, we developed an AGB mapping method using a combination of these ground inventory data, Geoscience Laser Altimeter System (GLAS)/Ice, Cloud, and Land Elevation Satellite (ICESat) data, optical imagery, climate surfaces, and topographic data. An uncertainty field model was introduced into the forest AGB mapping procedure to minimize the influence of plot location uncertainty. To test the feasibility of the proposed algorithm, we first conducted a pilot study in China. Over 8000 ground inventory records were collected from published literatures to estimate the forest AGB of China. Our nationwide wall-to-wall forest AGB mapping results show that the forest AGB density in China is 120 Mg/ha on average, with a standard deviation of 61 Mg/ha. Evaluation with an independent ground inventory dataset showed that our proposed method can accurately map wall-to-wall forest AGB across a large landscape. The adjusted coefficient of determination (R^2) and root-mean-square error (RMSE) between our predicted results and the validation dataset are 0.75 and 42.39 Mg/ha, respectively. Then, we applied this method to global scale, and generated the first global-scale forest AGB map at 1-km resolution. The results show that the global forest AGB density is 210.09 Mg/ha on average, with a standard deviation of 109.31 Mg/ha. The R^2 and RMSE between our predicted results and the validation plots were 0.56 and 87.53 Mg/ha, respectively. This new method and the resulting global wall-to-wall forest AGB map will help to improve the accuracy of carbon dynamic predictions.

Keywords: forest aboveground biomass, GLAS/ICESat, LiDAR, ground inventory

5.1 Introduction

Global forest ecosystems provide a large carbon sink, which plays an import role in the global carbon balance (Pan et al., 2011). Overall, forest ecosystems cover ~30% of the land surface, accounting for ~75% of terrestrial gross primary production and ~80% of global plant biomass (Beer et al., 2010; Bonan, 2008; Kindermann et al., 2008). It is therefore necessary to accurately estimate the current distribution of, and temporal variations in, the forest carbon stock in aboveground biomass (AGB) to obtain a clearer understanding of carbon dynamics against the background of global climate change (Galbraith et al., 2010; Keith et al., 2009; Pan et al., 2011).

Generally, forest AGB can be estimated through three available methods: model-based simulations (Bergh et al., 1998; White et al., 2000), measurements from traditional ground inventories (Botkin and Simpson, 1990; Botkin et al., 1993; Fang et al., 1998), and

retrievals from remote-sensing datasets (Ghasemi et al., 2011; Mitchard et al., 2011a). Model-based simulation methods usually provide forest AGB estimations from local to global scales based on model inputs (e.g., radiation, climate surfaces, and elevations) instead of the actual forest AGB distribution (Iverson et al., 1994; Lu, 2006). Traditional forest inventory methods (e.g., direct harvest methods and indirect allometric modelling methods) can provide reliable information on biomass at local or regional scales (Fang et al., 2001; Malhi et al., 2002). There have been many national forest inventory programs conducted across the world to provide accurate information for forest management, such as the National Forest Inventories program conducted in the Nordic countries (Tomppo et al, 2010). However, taking ground measurements is labor intensive and expensive when used for large areas, and is time consuming for nationwide forest survey (Houghton, 2005). For example, in U.S. and China, nationwide forest inventories provide reliable information on forest AGB, but these inventories are usually conducted on a five-year cycle at the nationwide scale, and require extensive human and economic resources.

Compared with the forest inventory approach, remote-sensing techniques significantly improve the efficiency of forest AGB mapping in large areas and areas that are difficult to access (Lu et al., 2005; Powell et al., 2010). By linking with ground inventory data, forest AGB can be estimated from remote sensing datasets using statistical models. Typically, passive optical remote sensing [e.g., Moderate Resolution Imaging Spectroradiometer (MODIS) and Landsat Thematic Mapper (TM)] and radar techniques [e.g., phased array L-band synthetic aperture radar (PALSAR) and Shuttle Radar Topography Mission (SRTM)] have become primary data sources for estimating forest AGB, because of their availability (Dong et al., 2003; Ghasemi et al., 2011; Koch, 2010; Mitchard et al., 2011a; Rauste, 2005; Soenen et al., 2010). However, the retrieved AGB values using these approaches are usually fraught with saturation effects, because of their limited penetration in vegetated areas (Baccini et al., 2008; Luckman et al., 1997). The saturation points for optical remote sensing range from 15 to 70 Mg/ha, and those for radar range from 30 to over 300 Mg/ha based on the use of different wavelengths (Lu, 2006; Mitchard et al., 2009, Myneni et al., 2001; Sader et al., 1989; Woodhouse et al., 2012).

Another alternative method, light detection and ranging (LiDAR), an active remote sensing technique, can penetrate the forest canopy effectively, because it uses a focused short-wavelength laser pulse (Su and Guo, 2014). This technique has shown great potential in mapping forest AGB by providing accurate estimates of tree metrics such as tree height, which are closely linked to AGB (Boudreau et al., 2008; Clark et al., 2004; Lefsky et al., 2005; Popescu et al., 2011). LiDAR does not suffer from saturation effects, even at high AGB levels, and therefore far exceeds the capabilities of optical passive and radar remote sensing in mapping forest AGB (Clark et al., 2011; Nelson et al., 2009; Swatantran et al., 2011). However, neither of the two major LiDAR platforms, airborne LiDAR and spaceborne LiDAR, provides complete global coverage of land surfaces. Currently, airborne LiDAR data are only available for certain small areas across the world. The Geoscience Laser Altimeter System (GLAS), onboard the Ice, Cloud, and Land Elevation Satellite (ICESat) is the only available spaceborne LiDAR system, and its footprints (with a nominal diameter of ~65 m) are spaced at 170 m along tracks and tens of kilometers across tracks.

By taking advantages of each type of data, reliable regional or even global AGB products can be obtained by combining multisource datasets. For example, Boudreau et al. (2008) explored the possibility of using GLAS data, Landsat-derived land-cover maps, and SRTM data to map the forest AGB distribution in Quebec, Canada at the patch level. Baccini et al. (2008) mapped the forest AGB distribution in Africa at a resolution of 1 km by combining GLAS and MODIS data. Mitchard et al. (2011b) and Saatchi et al. (2011) mapped the forest AGB distributions in tropical areas using GLAS and PALSAR data, and GLAS, MODIS, and radar data, respectively. Zhang et al. (2014c) mapped the AGB in northeastern China by integrating field data, GLAS data, and MODIS products using regression models. Neigh et al. (2013) estimated the total AGB for the entire circumboreal forest biome by incorporating ground inventory, airborne LiDAR data and spaceborne LiDAR data. These pioneering studies prove the validity of GLAS-derived metrics in the regional-level estimation of forest AGB. However, to the best of our knowledge, few published studies have explored the global forest AGB distribution using spaceborne LiDAR. One of the main obstacles is the lack of ground inventory data with which to build linkages with the GLAS-derived metrics across the country.

In this study, we developed a forest AGB estimation procedure using a combination of ground inventory data, spaceborne LiDAR, optical imagery, climate surfaces, and topographic data. Over 10, 000 plot measurements were collected from the published literature, and used to map the global forest AGB at a 1 km resolution. The plot location was usually recorded at a 1–10 km accuracy, and therefore the influence of the plot location uncertainty was further considered in the forest AGB mapping procedure. The wall-to-wall maps of forest AGB are downloadable from http://guolabLiDAR.com.

5.2 A pilot study in estimating forest AGB distribution of China

To develop the forest AGB mapping algorithm and test its feasibility, we first conducted a pilot study in China. The forest area of China is among the top five globally and covers 20.36% of the country's total area (State Forestry Administration of China, 2013). In addition, because of persistent afforestation and reforestation efforts (e.g., Grain for Green Program, GGP) since the1950s, the country's forest area has increased by 0.73×10^6 km^2 (~60%) since the 1970s (State Forestry Administration of China, 2013). The forest terrestrial ecosystem of China is a large carbon sink and contributes significantly to national and global carbon storage (Fang et al., 2001; Piao et al., 2009). Recent research has highlighted the substantial increase in carbon stock as a result of afforestation and reforestation (Fang et al., 2014; Liu et al., 2014; Xiao, 2014). Modeling results show that sequestration of an additional 110.45 Tg of carbon is expected by the 2020s, as a result of the GGP (Liu et al., 2014). Substantial changes in forest areas also result in large spatial and temporal variations in nationwide carbon stocks (Guo et al., 2013). However, it is still a challenge to estimate the carbon stocks in AGB, because of the lack of efficient and large-scale practical methods.

5.2.1 Data

5.2.1.1 Forest inventory data

Forest ground inventory data are fundamental in retrieving AGB from remotely sensed datasets. In this study, we collected 8429 records of plot measurements (including both research plots and forest inventory plots) from previously published papers. The collected plot measurements covered both plantations and natural forests. The geolocation of each individual plot record and its corresponding attributes (including stand origin, measurement method, data published year, and AGB value) were included in the supplemental materials. Since these in-situ plot measurements were collected from various data sources and were taken under different standards, this study further used four filtering criteria to ensure their quality: (1) they should be georeferenced; (2) they should be larger than 0.05 ha; (3) they should have been measured after 2000; and (4) they should not have been surveyed using harvesting methods. Records with the same geolocations but measured at different years were further averaged. Finally, 1065 plot samples were retained for the following forest AGB mapping procedures (Figure 5-1).
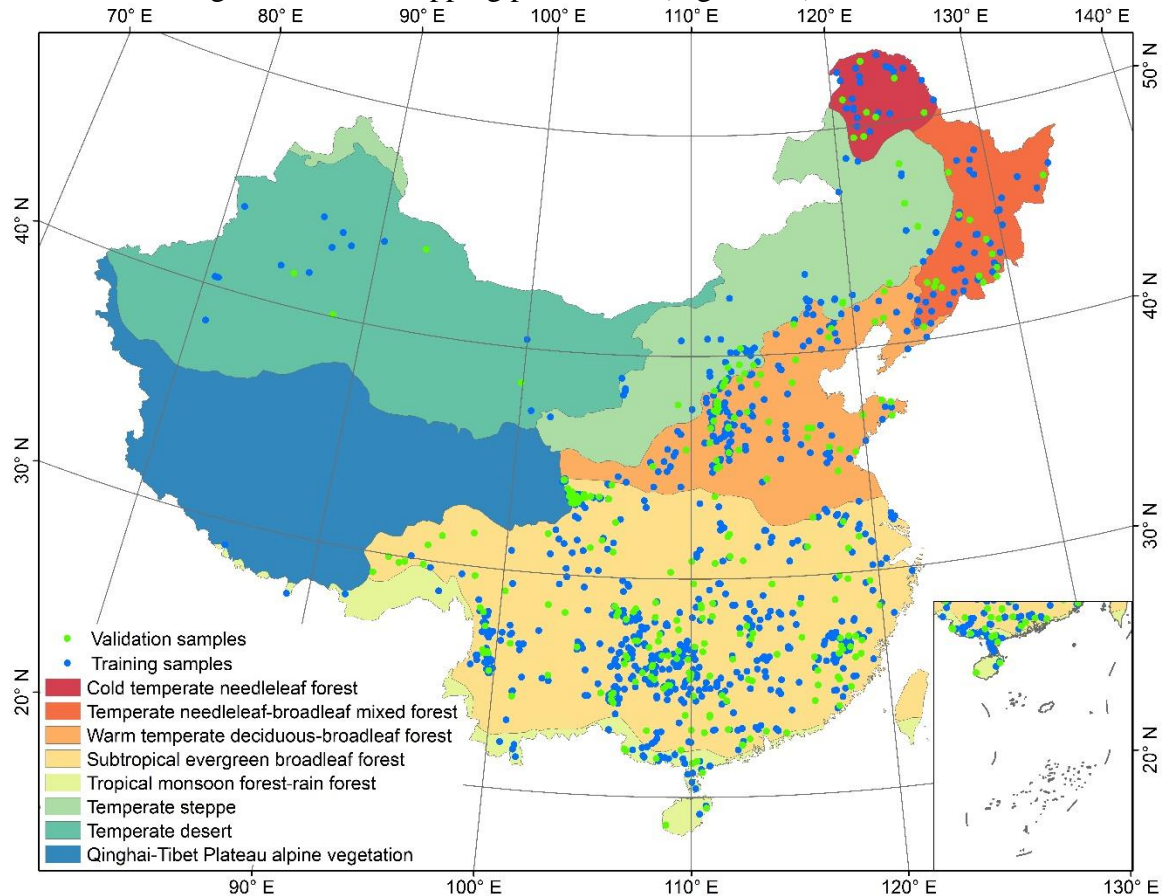
Figure 5-1 Geolocations of collected ground inventory data and corresponding vegetation zones from Hou (2001). The collected ground inventory data were divided into validation samples (green dots) and training samples (blue dots), at a ratio of 1:3.

5.2.1.2 ICESat GLAS laser altimetry data

The NASA (National Aeronautics and Space Administration) GLAS instrument, onboard the ICESat satellite, was launched on January 12, 2003; it was designed to have a 183-day ground track repeating cycle. It was equipped with three laser sensors, L1–L3, and 1064 nm laser pulses operated at 20 Hz were used to record the returned full-waveform altimetry data within ~65 m ellipsoidal footprints. The footprints were spaced at 170 m along the track and tens of kilometers across tracks (Schutz et al., 2005). The returned waveform over land had 544 range bins with 15 cm or 60 cm intervals, corresponding to a vertical range of 81.6 m or 150 m above the ground.

Since most of the collected plot measurements were obtained around 2004, we selected the GLAS data during the operating periods L2B, L2C, and L3A (from February 17, 2004 to November 09, 2004) for further AGB mapping procedures. Three GLAS products, GLA01, GLA06, and GLA14, were collected from the ICESat/GLAS data pool (http://nsidc.org/data/icesat/order.html); they provided full-waveform information, geolocation and data quality information, and surface elevation information, respectively. These three products were linked together using the unique ID and shot time of each laser pulse. For the GLAS footprints within the study area, four further filtering criteria were used to ensure the quality of GLAS measurements: 1) they should be taken under cloud free conditions; 2) they should have no saturation effects; 3) they should have high signal to noise ratios (i.e., >50); and 4) they should not be significantly higher (i.e., < 100 m) than the land surface elevation denoted by the SRTM data.
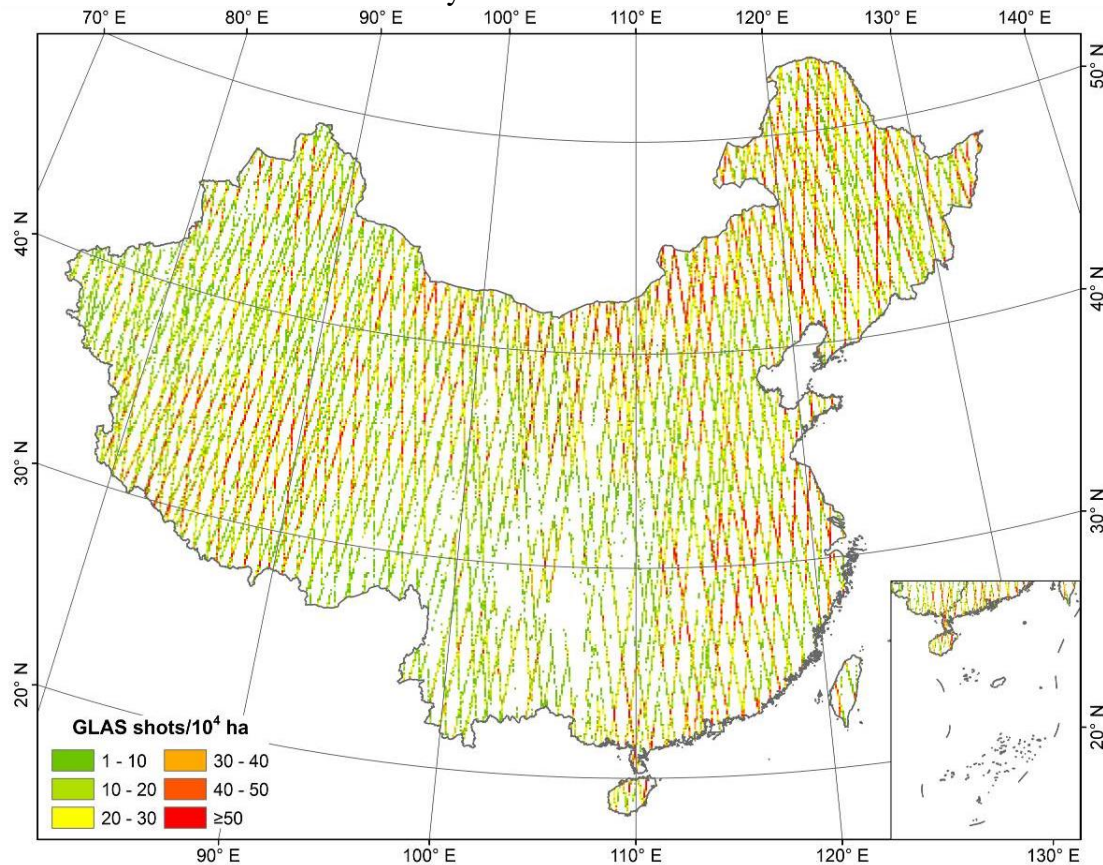
Figure 5-2 Densities of filtered GLAS shots (per 10^4 ha).

The final GLAS dataset contained 629 075 full-waveform records across China. The sampling density was generally higher in flat and forested areas (Figure 5-2). For each retained GLAS record, three parameters were derived from its full-waveform information, i.e., the waveform extent, leading edge extent, and trailing edge extent, which have been shown to be highly correlated with canopy height, canopy height variability, and terrain slope, respectively (Boudreau et al., 2008; Lefsky et al., 2007; Lefsky, 2010; Su et al., 2015). The definitions of these three parameters can be found in Lefsky et al. (2007), and will not be discussed in detail here.

5.2.1.3 MODIS-Terra MOD13A2 NDVI data

The MODIS-Terra MOD13A2 data are calculated from the MODIS atmospherically corrected bi-directional surface reflectance product, and are provided globally at a 1 km resolution every 16 d. Water bodies, clouds, aerosols, and cloud shadows are masked in the product. In this study, we collected time-series MOD13A2 data during the growing season of the year 2004 (from May to September). The use of cumulative normalized difference vegetation index (NDVI) from time-series NDVI can increase the AGB estimation accuracy compared with the use of NDVI data from a single time period (Li et al., 2015). Therefore, we computed the cumulative NDVI from the sum of all collected MOD13A2 data, and used it as a predictor in the AGB mapping procedure.

5.2.1.4 Topographic data

SRTM, a joint mission conducted by NASA and the National Geospatial-Intelligence Agency, provides a digital elevation model (DEM) product that covers 99.97% of the Earth's land surface, from 56° S to 60° N (U.S. Geological Survey, 2013). Its designed accuracy is 20 m horizontally and 16 m vertically. In this study, we used the second version of the SRTM data for China at a resolution of 3 arcsec (often quoted as 90 m), which show well-defined water bodies and coastlines, and the absence of spikes and wells (U.S. Geological Survey, 2013). To be consistent with other datasets, the SRTM DEM was resampled to 1 km resolution for further AGB mapping procedures. The slope product (denoted by tangent values of slope) was also calculated from the resampled SRTM DEM.

5.2.1.5 Climate surfaces

Four climate surfaces, namely annual mean temperature, annual temperature seasonality, annual total precipitation, and annual precipitation seasonality between 1950 and 2000 were calculated at 1 km resolution using data from 2888 weather stations (1512 stations with temperature records and 1376 stations with precipitation records) across China. These weather station data were obtained from the Food and Agriculture Organization FAOCLIM 2.0 (http://www.fao.org/nr/climpag/pub/en1102_en.asp) and the Global Historical Climate Network Dataset (GHCN) version 2 (Peterson and Vose, 1997). All the weather station data were manually checked to remove as many outliers or human errors as possible.

To generate these four climate surfaces, monthly mean temperature and monthly total precipitation layers from 1950 to 2000 were interpolated from the weather station data

using the thin plate spline algorithm, with the resampled 1 km resolution SRTM DEM as the covariable (Alvarez et al., 2014). Then, for each year, the mean temperature and total precipitation were directly calculated from the monthly layers, and the temperature seasonality and precipitation seasonality were computed from the following equation (Xu & Hutchinson, 2011):

$$Seasonality = 100 \times SD_{monthly} / Mean_{monthly} \tag{5-1}$$

where *Seasonality* represents the temperature seasonality or precipitation seasonality of the corresponding year, and $SD_{monthly}$ and $Mean_{monthly}$ are the standard deviation and mean of the monthly temperature (in Kelvin) or monthly precipitation (in millimeters) for the corresponding year. Finally, the annual mean temperature, annual temperature seasonality, annual total precipitation, and annual precipitation seasonality were derived from the average of the corresponding yearly products.

5.2.1.6 Auxiliary data

In addition to the above-mentioned data, two auxiliary datasets, i.e., a 1:1 000 000 vegetation map of China (Hou, 2001) and a 1 km land-use map of China at a 1:100 000 scale from the year 2000 land-use database (Liu et al., 2002), were used in this study. The vegetation map divided China into eight vegetation zones, based on the dominant vegetation type (Figure 5-1). The land-use map classified the land cover into six major categories (i.e., cropland, forest, grass land, water bodies, artificial areas, and bare earth) and 23 subgroups, based on the interpretation of Landsat TM images. Within each subgroup, the percentage of corresponding land-use type aggregated from the 30 m Landsat TM images was provided. In this study, we were particularly interested in two subgroups in the forest category, the closed-forested area group and the open-forested area group.

5.2.2 Methodology
5.2.2.1 Extrapolation of GLAS-derived parameters

The GLAS-derived parameters, i.e., waveform extent, leading edge extent, and trailing edge extent, normally do not have direct biological meanings (Lefsky et al., 2007). Additional forest parameter data derived from airborne LiDAR or field plot measurements are usually required to convert GLAS parameters to biologically meaningful parameters (e.g., canopy height and AGB) (Boudreau et al., 2008; Lefsky, 2010; Saatchi et al., 2011). However, airborne LiDAR data are only available for a few small areas, because of limitations in terms of flight mission cost and mission duration. Moreover, in our cases, the probability of the collected plot measurements overlapping with GLAS footprints was too small, because of the huge gap between two adjacent GLAS tracks (Figure 5-2), and the considerable uncertainty within the plot location further increased the difficulty of matching the GLAS footprints with plot measurements.

In this study, to relate the GLAS parameters to plot measurements, we extrapolated the three GLAS parameters into spatial continuous layers using a regression method. GLAS footprints that did not fall in the closed-forested area group or the open-forested area group from the land-use map (Liu et al., 2002) were excluded to minimize the influence of non-forested areas on the extrapolation results. In total, 202 298 GLAS footprints were retained

for the subsequent extrapolation process. Furthermore, because all the collected raster datasets were at (or resampled to) 1 km resolution, we aggregated the retained GLAS footprints into 87 893 1 km pixels, and the averages of the three GLAS-derived parameters within each pixel were used in the final GLAS parameter extrapolation procedure.

The regression tree method Random Forest (RF) (Breiman, 2001) was used to extrapolate the three GLAS-derived parameters. RF, a formalized non-parametric machine-learning algorithm, has been successfully used in areas such as biomass mapping (Baccini et al., 2008) and niche modeling (Prasad et al., 2006). Simard et al. (2011) showed that RF was robust in extrapolating RH100, one of the GLAS-derived parameters, to a spatial continuous layer. One of the main advantages of RF is that it does not require an assumption to be made regarding the normality of covariables (Breiman, 2001), which is often violated when complex ecological systems and environmental variables are introduced (Saatchi et al., 2011). Also, RF can minimize the within-group variance at the expense of a small increase in the bias and overcome the overfitting habit of decision tree algorithms (e.g., Classification and Regression Tree algorithm) (Breiman, 2001; Friedman et al., 2001), because of the use of its unique tree "bagging" algorithm, which selects a random subset of covariables at each candidate split.

Seven ancillary predictors were used in the RF regression tree modeling process, namely cumulative NDVI, elevation, slope, and the four climate surfaces. Moreover, each vegetation zone from Hou (2001) was represented by a unique numerical identification number and fed into the RF regression tree model. The RF extrapolation method was implemented using the randomForest R package (Liaw and Wiener, 2002), which includes both classification and regression functions. In this study, 500 "RF trees" were included and four variables were tried at each split, based on manually iterative examination.

5.2.2.2 Forest AGB estimation

The 1065 plot measurements were randomly divided into a training dataset and validation dataset, at a ratio of 3:1 (Figure 5-1). The 799 training plot measurements along with the three GLAS metrics and the seven collected predictors were used to estimate forest AGB using the RF regression tree algorithm (Figure 5-3). As in the GLAS extrapolation procedure, the regression RF forests were built by relating the training plot measurements to the prediction variables. Similar to the GLAS parameter extrapolation process, 500 RF trees and four prediction variables tried at each split were selected based on manually iterative examination.
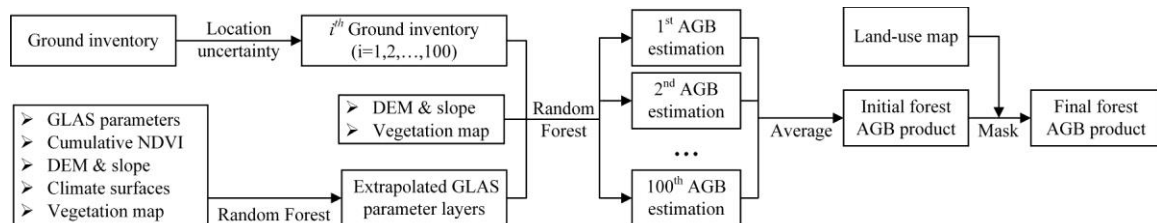

Figure 5-3 Scheme for estimating forest AGB distribution in China from multisource remotely sensed data and ground inventory data.

To improve the computational efficiency and remove redundant information, we explored the importance of all prediction metrics and determined the optimal variables for forest AGB mapping. The percentage increase in the mean-squared error (%IncMSE) and the increase in node purity (IncNodePurity) were calculated to evaluate the variable importance (Liaw and Wiener, 2002). The larger the %IncMSE and the IncNodePurity of a variable are, the more important this variable is. As can be seen from Figure 5-4, regardless of whether the evaluation was performed using the %IncMSE or the IncNodePurity, the absence of the three GLAS-derived metrics and topographic data significantly increased the mean-squared error and node purity, indicating a decrease in the prediction accuracy of the built RF regression tree model. The three GLAS-derived metrics and topographic data were therefore used as the prediction variables for mapping the AGB distribution. Note that considering the different characteristics of trees from different vegetation zones (e.g., tree species, tree size and tree density), the vegetation zone data was also used in the AGB mapping process.
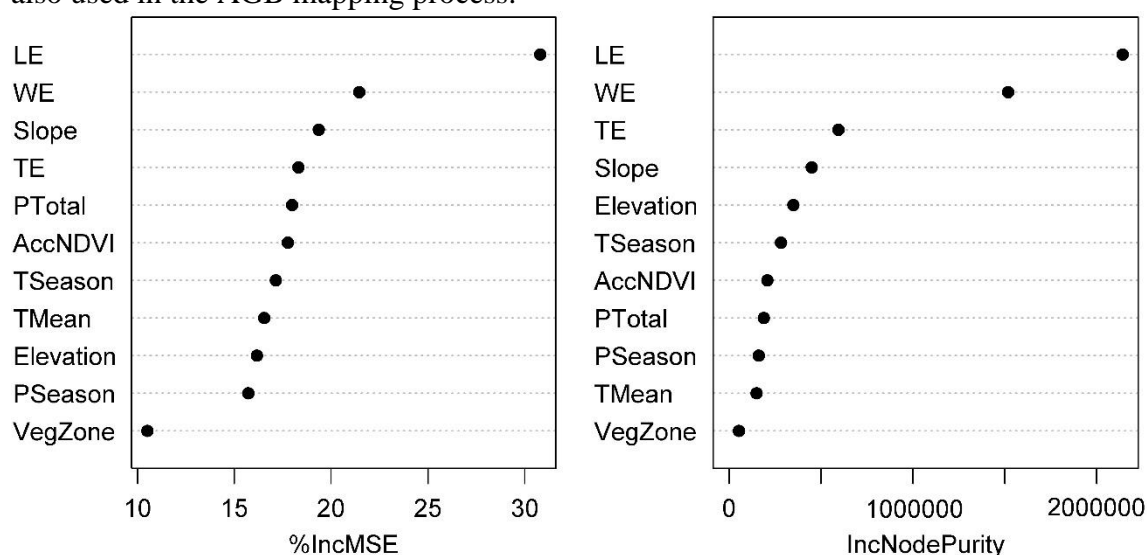
Figure 5-4 Importance of variables, denoted by percentage increase of mean-squared error (%IncMSE) (left) and the increase in node purity (IncNodePurity) (right), for AGB estimation RF model built from training dataset without considering plot location uncertainty. LE, WE, TE, PTotal, PSeason, TMean, TSeason, AccNDVI, and VegZone represent the leading edge extent, waveform extent, trailing edge extent, annual total precipitation, annual precipitation seasonality, annual mean temperature, annual temperature seasonality, cumulative NDVI, and vegetation zone, respectively.

Additionally, when relating the ground inventory data to the predictors, the plot location uncertainties were too large to be neglected. Without considering the surveying accuracy, most of the latitudes and longitudes given by the inventory data were accurate to 0.01° (corresponding to ~1 km), and some of them were only accurate to 0.1° (corresponding to ~10 km). These huge plot location uncertainties could result in significant differences in the corresponding values of predictor variables, and therefore influence the forest AGB estimation result. To minimize the influence of plot location uncertainty, we introduced an uncertainty field model (Guo et al., 2008) into the RF AGB mapping procedure (Figure 5-3). This method hypothesized that the real plot center was

randomly located within a circular buffer zone of the provided plot location, and the radius of the buffer was determined by the corresponding plot location uncertainty. In this study, by assuming that the plot location cannot be more than 1 or 10 km (determined by the accuracy of the given plot location) away from the given location in any direction, we created a 1 or 10 km buffer around each plot. Then 100 sets of different ground inventory data were randomly generated within the buffers.

The AGB mapping procedure based on the RF regression tree algorithm was performed for each set of ground inventory data with a location uncertainty (Figure 5-3). We generated 100 predicted forest AGB layers using this process, and the average of these forest AGB layers was taken as the initial forest AGB estimation result. The final forest AGB mapping result was computed from the initial forest AGB estimation result by setting the non-forested pixels at 0 Mg/ha. In this study, if a 1 km pixel was not covered by either of the two forested groups from the land-use map (i.e., the closed-forested area group and the open-forested area group), it was treated as a non-forested pixel, and vice versa. It should be noted that the RF prediction process for each run was operated separately on each vegetation zone using the same RF regression tree model obtained from the corresponding run, considering the computation efficiency. The Qinghai-Tibet Plateau alpine vegetation and the temperate steppe vegetation zones were merged together and treated as one zone in this process, which were both dominated by grassland and shrubs and had relatively small forested areas.

5.2.2.3 Accuracy assessment

The accuracy of the estimated forest AGB of China was evaluated at the plot level, pixel level, and vegetation zone level. The plot-level accuracy was assessed by directly comparing the estimated result with an independent validation ground inventory dataset (Figure 5-3). The adjusted coefficient of determination (R^2) and root-mean-square error (RMSE) were computed using the following equations:

$$R^2 = 1 - \frac{(n-1)\sum_{i=1}^{n}(B_i - \hat{B}_i)^2}{(n-2)\sum_{i=1}^{n}(B_i - \bar{B})^2} \tag{5-2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(B_i - \hat{B}_i)^2}{n-2}} \tag{5-3}$$

where B_i is the observed AGB from the validation plot, $\hat{B}_i$ is the modeled AGB at the plot location from the wall-to-wall AGB map, $\bar{B}$ is the average AGB of all validation plots, and n is the number of validation plots.

The pixel-level uncertainty was only evaluated by the uncertainty induced by the plot location ($\varepsilon_{location}$), which was calculated from the standard deviation of 100 AGB estimation iterations:

$$\varepsilon_{location} = \sqrt{\frac{\sum_{j=1}^{100}(\hat{B}_i - \overline{\hat{B}})^2}{100}}$$
(5-4)

where $\hat{B}_j$ is the modeled AGB at the jth iteration and $\overline{\hat{B}}$ is the average of all modeled AGBs. Finally, the uncertainty at the vegetation zone scale was estimated by the RMSE between the estimated AGB and the independent validation plot inventory data for each vegetation zone.
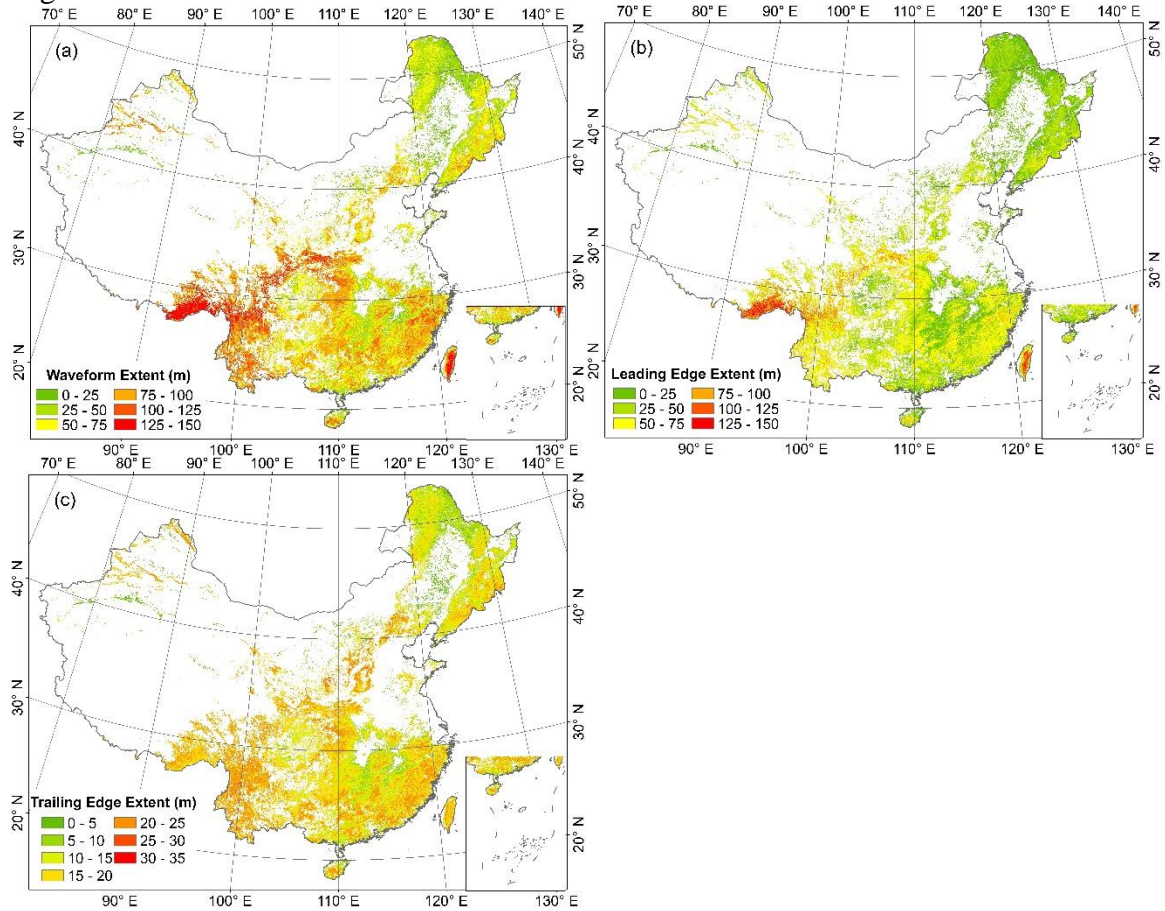
Figure 5-5 Distributions of extrapolated GLAS-derived full-waveform parameters, namely (a) waveform extent, (b) leading edge extent, and (c) trailing edge extent. Note that areas without closed-forested area and open-forested area coverage have been excluded from the maps.

5.2.3 Results

5.2.3.1 Extrapolated GLAS parameters

The GLAS-derived waveform extent, leading edge extent, and trailing edge extent were extrapolated to spatially continuous layers using the RF regression tree method (Figure 5-5). Overall, the built RF regression models can explain around 57%, 52%, and 46% of the

variances in the waveform extent, leading edge extent, and trailing edge extent, respectively. The root-mean-square residuals for the extrapolated waveform extent, leading edge extent, and trailing edge extent are 24 m, 20 m, and 6 m, respectively. The waveform extent and leading edge extent share similar spatial patterns. Values in southern China are generally higher than those in northern China. In the tropical monsoon forest–rain forest located in southern Tibet, both the waveform extent and leading edge extent have significantly higher values than in other areas. Compared with the waveform extent and the leading edge extent, the trailing edge extent is more evenly distributed across China.

5.2.3.2 Forest AGB estimation in China

The wall-to-wall forest AGB map is shown in Figure 5-6 and can be downloaded from http://faculty.ucmerced.edu/qguo/carbon. The RF regression tree model built from plot measurements without considering location uncertainties can explain 72% of variances in forest AGB and the root-mean-squared residual is 40.12 Mg/ha. Overall, the forest AGB density is higher in southern China than in northern China (Figure 5-6). The average forest AGB density for the whole China (without considering non-forested areas) is about 120 Mg/ha, and the standard deviation is 61 Mg/ha.
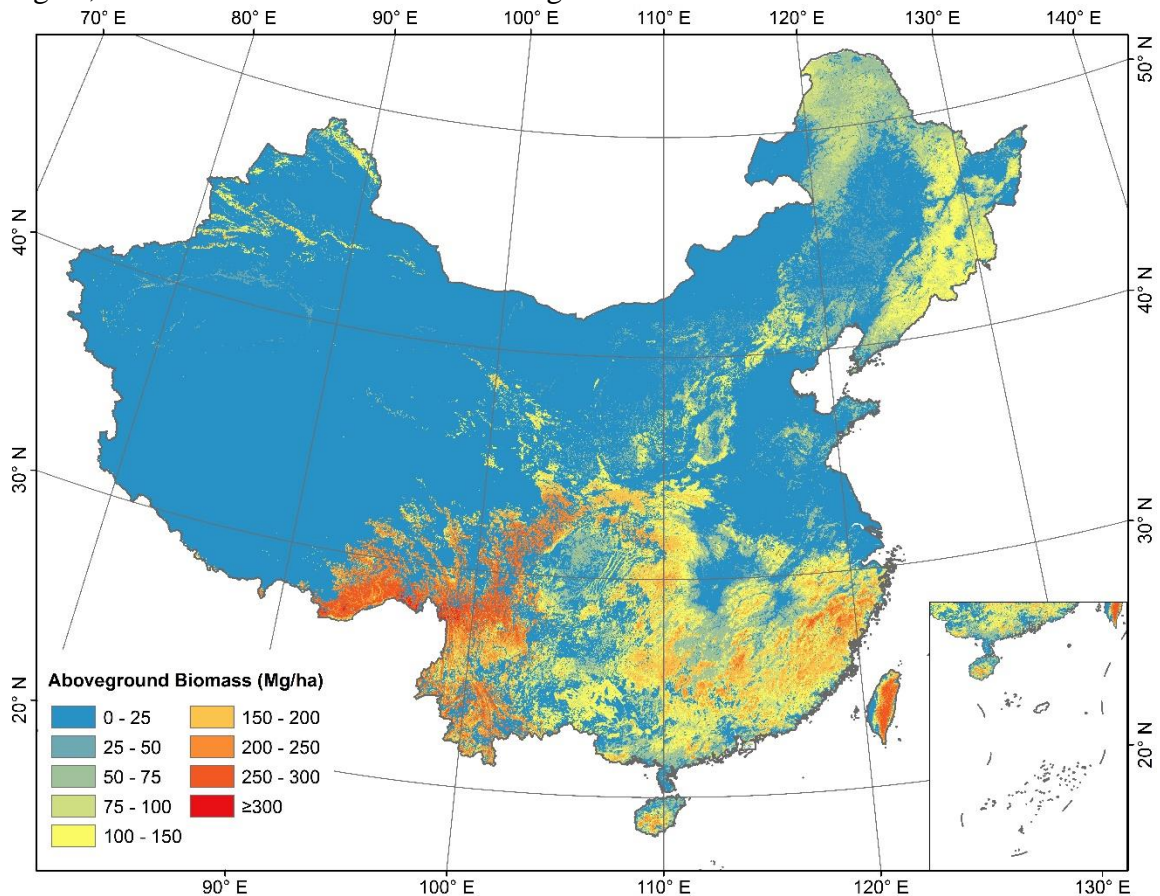
Figure 5-6 Wall-to-wall map of forest AGB distribution in China.

Northern China is mainly occupied by the cold temperate needleleaf forest, temperate needleleaf–broadleaf mixed forest, warm temperate deciduous broadleaf forest, temperate steppe, and temperate desert (Figure 5-1). The mean AGB density for the cold temperate needleleaf forest is around 75 Mg/ha, and is more homogeneously distributed. Over 50% of such forested areas have an AGB density range of 70–80 Mg/ha (Figure 5-7). The warm temperate needleleaf–broadleaf mixed forest has an average AGB density of around 90 Mg/ha, and over a half of such forested areas have an AGB density range of 80–120 Mg/ha. Although the warm temperate deciduous broadleaf forest is in the south of the temperate needleleaf–broadleaf mixed forest, its average AGB density is smaller, and close to that of the cold temperate needleleaf forest (Figure 5-7). This may be due to the fact that these areas have been heavily disturbed by human activities, and most of these forests are plantations instead of natural forests. The average AGB density for the temperate steppe is the lowest among all vegetation zones (~65 Mg/ha) (Figure 5-7), and its proportion of forest is less than 9%. The average AGB density of the temperate desert is about 100 Mg/ha, and over 50% of this area is within the range 50 to 150 Mg/ha. Most forests in this vegetation zone are distributed within the Junggar Basin (Figure 5-6).
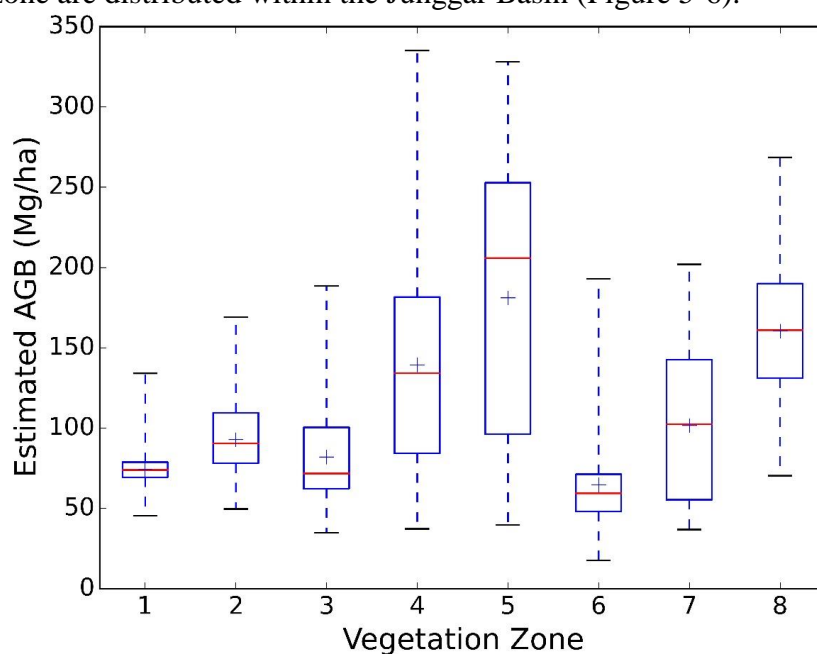
Figure 5-7 Boxplot of estimated forest AGB in each vegetation zone. The blue "+" indicates the mean forest AGB of corresponding vegetation zone. Numbers 1–8 on x axis represent cold temperate needleleaf forest, temperate needleleaf–broadleaf mixed forest, warm temperate deciduous broadleaf forest, subtropical evergreen broadleaf forest, tropical monsoon forest–rain forest, temperate steppe, temperate desert, and Qinghai-Tibet Plateau alpine vegetation, respectively.

Southern China is mainly covered by subtropical evergreen broadleaf forest, tropical monsoon forest–rain forest, and Qinghai-Tibet Plateau alpine vegetation. All these vegetation zones have significantly higher average AGB densities than those of the vegetation zones in northern China (p<0.0001) (Figure 5-7). Specifically, the average AGB

density for the subtropical evergreen broadleaf forest is about 140 Mg/ha, and the maximum AGB density can reach over 330 Mg/ha. The tropical monsoon forest–rain forest has the highest mean forest AGB density among all vegetation zones, and has a broader range of AGB density distributions (Figure 5-7). Its median AGB density is over 200 Mg/ha, and about half of the area is within the AGB density range 100–250 Mg/ha. The average AGB density for the Qinghai-Tibet Plateau alpine vegetation zone is about 150 Mg/ha, and is mainly contributed by forested areas in the most southeastern part of the Qinghai-Tibet Plateau (Figure 5-6).

The accuracy of the wall-to-wall forest AGB estimation in China was evaluated using 266 independent validation plot measurements. As shown in Figure 5-8, the R^2 between the predicted and field-measured AGB is 0.75 and the RMSE is 42.39 Mg/ha. Although the slope and intercept for the correlation between estimated and field measured AGB are larger than one and smaller than zero, the fitted line is still very close to the 1:1 line (Figure 5-8). The negative intercept suggests that the proposed AGB estimation method tends to slightly overestimate AGB densities at areas with low values (<87 Mg/ha). After the forest AGB density reaching 87 Mg/ha, the proposed AGB estimation method tends to underestimate the forest AGB density, and this underestimation effect becomes more pronounced with the increase of forest AGB density.

$$R^2 = 0.75$$
$$RMSE = 42.39\ Mg/ha$$
$$B = 1.21\hat{B} - 18.27\ (p < 0.01)$$

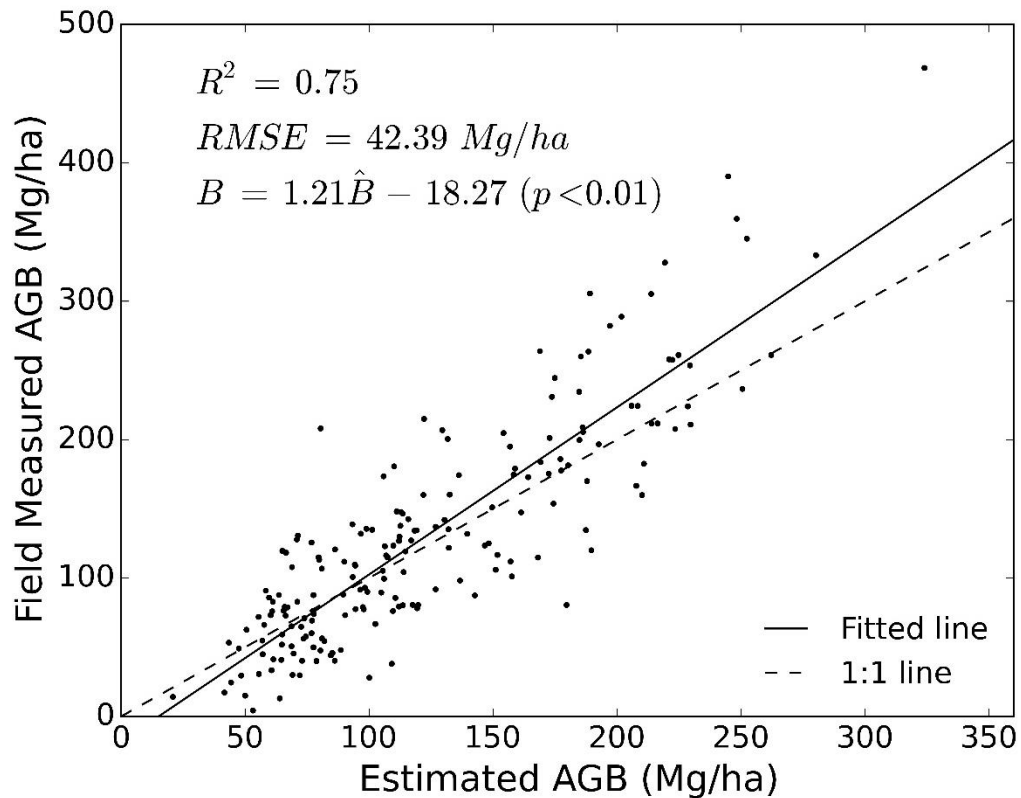Figure 5-8 Distribution of modeled forest AGB uncertainty introduced by plot location uncertainty at pixel level. White areas indicate areas with neither closed-forested area nor open-forested area land-use cover.

The estimated mean AGB density for each vegetation zone was further compared with that calculated from all 1065 plot measurements in each vegetation zone. As the data in Table 1 show, the differences between the predicted mean AGB density and plot-measured mean AGB density for most vegetation zones are smaller than 10 Mg/ha. The temperate desert has the largest difference, i.e., ~20 Mg/ha. The RMSE for each vegetation zone, calculated from the validation plot measurements, is shown in Table 5-1. As can be seen, the RMSEs for most vegetation zones are within the range 23–33 Mg/ha. The subtropical evergreen broadleaf forest and Qinghai-Tibet Plateau alpine vegetation have relatively large RMSEs (over 45 Mg/ha). The high RMSE for the Qinghai-Tibet alpine vegetation may be caused by the small number of validation plots (only three) located in this vegetation zone. The temperate desert has the smallest RMSE (~14 Mg/ha) among all vegetation zones.

Table 5-1 Comparison of mean forest AGB values from wall-to-wall predicted map and all plot measurements in each vegetation zone. Note that the RMSE was only calculated based on the validation plot dataset.

Vegetation zone	Predicted mean AGB (Mg/ha)	Plot mean AGB (Mg/ha)	RMSE (Mg/ha)
Cold temperate needleleaf forest	72.75	74.48	32.81
Temperate needleleaf–broadleaf mixed forest	104.02	92.79	26.08
Warm temperate deciduous broadleaf forest	72.29	82.12	32.56
Subtropical evergreen broadleaf forest	133.22	139.43	45.86
Tropical monsoon forest–rain forest	191.43	180.90	23.30
Temperate steppe	56.11	64.70	26.20
Temperate desert	81.45	101.63	13.86
Qinghai-Tibet Plateau alpine vegetation	168.20	160.42	50.12

5.2.4 Discussion
5.2.4.1 Linking ground plots and GLAS data
In this study, we collected over 8000 records of forest AGB field measurements and developed a procedure to estimate wall-to-wall forest AGB distributions in China, using a combination of plot measurements, GLAS data, optical imagery, climate surfaces, and topographic data. Because the GLAS data are spatially discontinuous, how to link plot measurements to GLAS footprints is one of the key issues in estimating forest AGB distribution. Generally, there are three methods for addressing this issue: 1) direct link based on the geolocation (Saatchi et al., 2011; Zhang et al., 2014c); 2) use of airborne LiDAR data as a medium (Boudreau et al., 2008); and 3) extrapolation of GLAS parameters (Zhang et al., 2014b).

Direct linkage between the ground inventory and GLAS measurements at the same geolocation is the most intuitive matching method. However, in our cases, about 80% of the plots are over 1 km away from their closest GLAS footprint. The discrepancy between their locations means that they can hardly be matched. Instead of direct linkage, airborne LiDAR has been used as the medium between ground inventory and GLAS measurements. Airborne LiDAR has proved to be capable of estimating forest AGB based on plot measurements (Boudreau et al., 2008). The use of airborne LiDAR data obtained by

selectively flying in portions of the study area enables the forest AGB within the footprints to be accurately estimated and therefore used as ground truth data to estimate the AGB within the GLAS footprints. However, currently, the availability of airborne LiDAR data in China is very limited because of the high flight mission costs.

Instead of trying to link the plot AGB and GLAS parameters at the GLAS footprint level, Zhang et al. (2014a) first extrapolated the GLAS-derived vegetation heights to continuous layers using a TM-derived leaf area index (LAI) product, and then estimated the AGB distribution from the vegetation height in California, USA at a 30 m resolution. The TM-derived LAI was the only predictor used to generate the continuous vegetation height, therefore their final AGB estimation relied heavily on the accuracy of the LAI product. However, previous studies have shown that the TM-derived LAI data are fraught with uncertainty issues because of the spectral saturation effect (Lu et al., 2004; Wang et al., 2005). This was also indicated by their results, which showed that the estimated forest AGB had a large uncertainty, ranging from 40 to 150 Mg/ha (Zhang et al., 2014a). Simard et al. (2011) showed the feasibility of deriving vegetation height by extrapolating RH100, a GLAS-derived parameter, to a global continuous layer using an RF regression tree algorithm. This greatly improved the potential of GLAS data in the estimation of global AGB. However, it has been reported that RH100 underestimates the canopy height (Sun et al., 2008) and careful calibration may be needed for further forest AGB estimation (Zhang et al., 2014a).
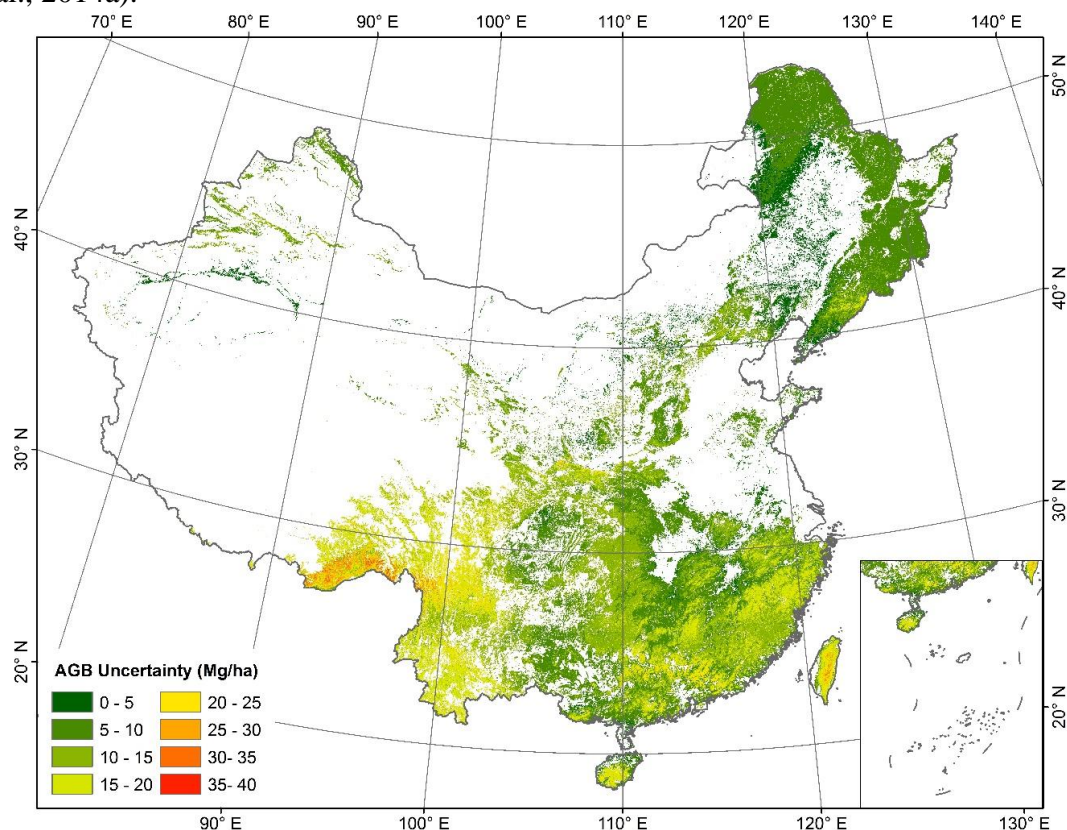

Figure 5-9 Distribution of modeled forest AGB uncertainty introduced by plot location uncertainty at pixel level. White areas indicate areas with neither closed-forested area nor open-forested area land-use cover.

In this study, we extrapolated the GLAS-derived waveform extent, leading edge extent, and trailing edge extent to spatially continuous products, using an RF regression tree model similar to that used by Simard et al. (2011). These three parameters have been successfully used to estimate the vegetation height and AGB at the GLAS footprint level (Boudreau et al., 2008; Lefsky et al., 2005). In this study, these three extrapolated GLAS metrics along with topographic data can explain over 72% variances of the forest AGB, and the forest AGB distribution can be mapped with high accuracy by linking the plot-measured AGB to them (Figure 5-8). This shows the feasibility of using extrapolated GLAS full-waveform parameters to estimate forest AGB.

Moreover, the plot location uncertainty has rarely been considered when linking plot measurements to GLAS parameters. For the plot measurements collected from the literature in particular, the plot location uncertainty can be as high as 10 to 15 km, without considering the actual surveying uncertainty. This huge location uncertainty can result in significantly different correspondences between field-measured AGB and predictors. In this study, as a result of introducing plot location uncertainty into the forest AGB modeling procedure, R^2 between the predicted AGB and plot validation dataset increased from 0.64 to 0.75, and the RMSE decreased from 50 to 42 Mg/ha. At the pixel level, the absolute value of the uncertainty introduced by the plot location increased with forest AGB density, and contributed to around 10% of the final AGB estimation (Figure 5-9). In the tropical monsoon forest–rain forest in the most southeastern part of Tibet, the uncertainty brought by the plot location can reach nearly 40 Mg/ha.

5.2.4.2 Influence of different forest definitions on total forest AGB estimation
Usually, forest extent is defined as an area with a tree cover higher than a fractional cover threshold. However, the tree cover threshold used to define forest extent varies with time and place. Tree covers of 10%, 25%, and 30% are the most frequently used thresholds used by individual parties to the United Nations Framework Convention on Climate Change. This discrepancy in the definition of a forest may result in significantly different total forest areas, and therefore influence the estimation of total biomass carbon stock in the forest.
Table 5-2 shows the total forest areas and total AGB estimations for different forest definitions. As can be seen, the total forest area for each vegetation zone decreases significantly with increasing tree cover threshold used to define the forest extent. The total forest area from the original Landsat TM land-use map is about 7% larger than the total forest area when a 30% tree cover threshold is applied on the Landsat TM land-use map to define the forest extent. This difference in total forest area can lead to a drop of ~5% in the total AGB estimation for China, which equals about 1000 million tons of forest biomass. The drop effect of the total forest AGB is greater in climate zones with relatively small total forest areas. For example, the temperate desert and Qinghai-Tibet Plateau alpine vegetation have a very small proportion of forest cover, and differences in the definitions of forest extent can result in a decrease of over 16% in the total forest AGB estimation.

5.2.4.3 Comparison of estimated forest AGB and published results
We compared our nationwide forest AGB map with previously published AGB mapping results covering different areas of China to further evaluate our result. For northeastern China, Zhang et al. (2014c) estimated that the average forest AGB density was 83.50 Mg/ha, and the total forest AGB carbon stock was 1.55 Pg. Our results indicate that the average forest AGB density in northeastern China is ~88.50 Mg/ha, and the total forest AGB carbon stock is 1.5 Pg (using a ratio of 50% to convert forest AGB to forest AGB carbon stock) (Saatchi et al., 2011). This total AGB carbon stock estimation is also consistent with the value of 1.4–1.6 Pg of carbon obtained by simulations using the TRIPLEX 1.0 model (Peng et al., 2009). The provincial average forest AGB density from our results is also close to that estimated by Guo et al. (2013) for northeastern China, which was estimated based on the national forest inventory data of China. The average forest AGB densities for Inner Mongolia, Liaoning, Jilin, and Heilongjiang provinces from our wall-to-wall AGB map are 67, 77, 93, and 81 Mg/ha, respectively, and those from their results are 77, 63, 119, and 85 Mg/ha. However, these values are significantly higher than those reported by Thurner et al. (2014), namely 27, 25, 40 and 33 Mg/ha, respectively. This may be caused by the growing stock volume product retrieved from synthetic aperture radar data, which was used as the predictor in their AGB estimation process. This product was mainly calibrated based on plot measurements in Russia, which might make it less representative for forests in northern China.

Table 5-2 Comparison of forest area and total forest AGB for each climate zone under different forest definitions. Vegetation zones 1–8 represent cold temperate needleleaf forest, temperate needleleaf–broadleaf mixed forest, warm temperate deciduous broadleaf forest, subtropical evergreen broadleaf forest, tropical monsoon forest–rain forest, temperate steppe, temperate desert, and Qinghai-Tibet Plateau alpine vegetation, respectively.

| Vegetation zone | 0%[a] | | 10%[b] | | 25%[b] | | 30%[c] | |
	Forest area (M km^2)	Total AGB (M ton)	Forest area (M km^2)	Total AGB (M ton)	Forest area (M km^2)	Total AGB (M ton)	Forest area (M km^2)	Total AGB (M ton)
1	0.14	926.54	0.14	923.17	0.13	907.54	0.13	899.02
2	0.22	2053.42	0.22	2047.99	0.22	2024.97	0.21	2012.64
3	0.11	796.07	0.11	783.83	0.10	743.50	0.10	725.25
4	1.00	11416.54	0.99	11322.78	0.96	10944.46	0.94	10750.21
5	0.12	2123.52	0.12	2112.93	0.12	2074.80	0.12	2056.13
6	0.08	472.55	0.08	461.35	0.07	426.29	0.07	411.64
7	0.03	205.42	0.03	199.73	0.02	179.61	0.02	170.59
8	0.01	127.97	0.01	124.52	0.01	112.05	0.01	106.39
Total	1.71	18122.03	1.70	17976.30	1.63	17413.21	1.60	17131.86

For southern China, Saatchi et al. (2011) estimated the forest AGB distribution in tropical regions using a combination of GLAS, MODIS, and radar data, which included forests of China lower than 40° N. Their forest AGB estimation ranges match our nationwide wall-to-wall forest AGB map as well, except in the case of the Qinghai-Tibet

Plateau. For the Qinghai-Tibet Plateau, the majority of our estimates are 0 Mg/ha, but their result ranges around 20-80 Mg/ha. This difference may be caused by the different land-use maps used by these two studies. In the land-use map we used, most of the Qinghai-Tibet Plateau was classified as non-forest groups, and our corresponding AGB values were set at 0 Mg/ha.
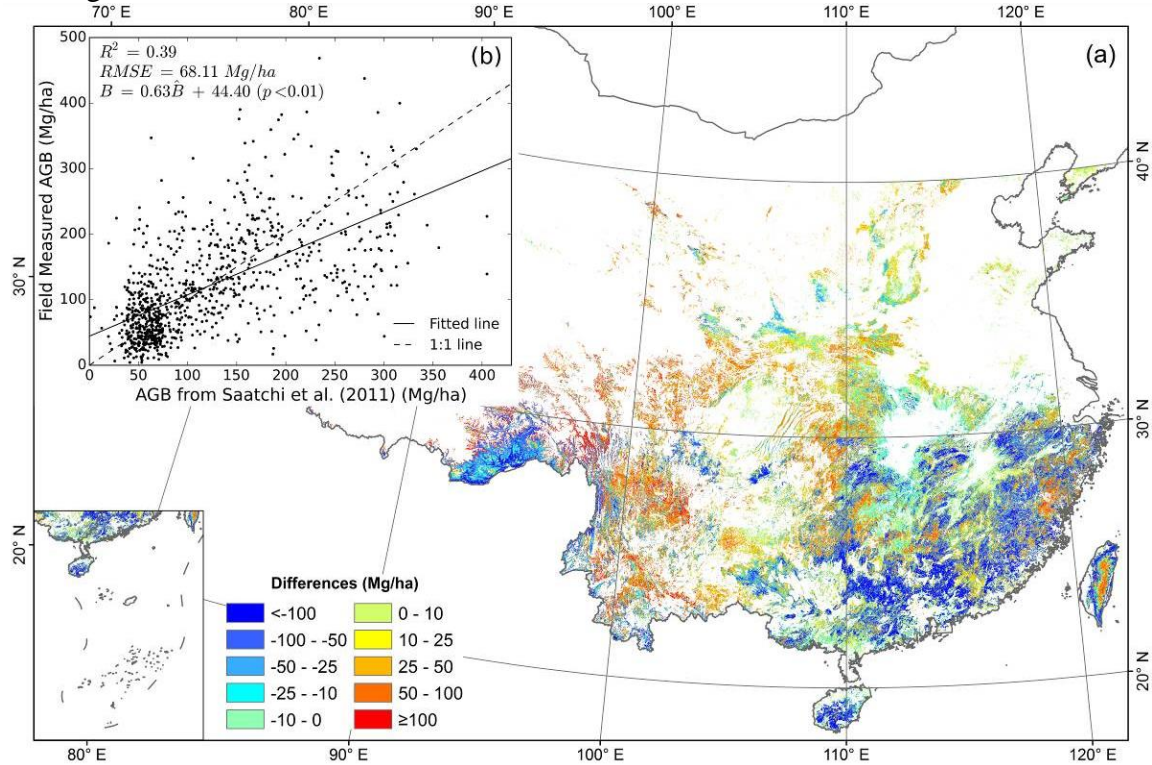

Figure 5-10 (a) Pixel-level differences between the estimated forest AGB from this study and the forest AGB from Saatchi et al. (2011) [our estimated AGB minus the forest AGB from Saatchi et al. (2011)]. (b) Scatter plot between the field measured AGB and the forest AGB from Saatchi et al. (2011). Note that R2 represents the adjusted coefficient of determination, RMSE represents the root-mean-square error, B represents the field measured AGB, and   represents the estimated AGB from Saatchi et al. (2011).

Besides the regional scale comparison, we also performed a pixel-level comparison between our estimated forest AGB map and that from Saatchi et al. (2011) in southern China (Figure 5-10a). It should be noted that the forest AGB map from Saatchi et al. (2011) was masked by the land-use map from Liu et al. (2002) before the comparison. Similar to our forest AGB map, the non-forested areas were set at 0 Mg/ha. The average difference between our and their forest AGB results is -7.23 Mg/ha. The absolute values of differences for 60% of the pixels are smaller than 50 Mg/ha, and for 36% of the pixels are smaller than 25 Mg/ha. As can be seen in Figure 5-10a, our forest AGB result tends to be lower than their result in southeastern China, and higher than their result in Yunnan, Chongqing, Sichuan and Guizhou provinces. In the most southeastern part of the Qinghai-Tibet Plateau, the majority of their forest AGB densities are slightly higher (10-50 Mg/ha) than our estimates. These differences are possibly caused by the fact that their forest AGB result was estimated based on plots from tropical areas across the world and were less

representative of forest conditions in China. As shown in Fig10b, we evaluated their AGB mapping result by our collected field measurements. Overall, their result shows a satisfying accuracy; the R^2 and the RMSE are around 0.4 and 68 Mg/ha respectively. However, their result inclines to underestimate forest AGB in areas with low values, and overestimate it in areas with high values (Figure 5-10b). In southeastern China, although this region belongs to the subtropical evergreen broadleaf vegetation zone, it has been heavily disturbed by human activities, and the proportion of plantations is relatively higher. The possible lack of plot measurements from these plantation areas may lead to overestimations of forest AGB and therefore make their result higher than our forest AGB map.

5.2.4.4 Limitations of the current study

Although the resulting nationwide wall-to-wall forest AGB map of this study shows good accuracy by comparing with both independent field measurements and other published products, there are still limitations in the methodology of the current study. The three extrapolated GLAS parameters along with topographic information can be used to explain over 72% of variations in forest AGB, and therefore accurately map the wall-to-wall forest AGB distribution of China. However, the extrapolated GLAS parameters were imputed based on the regression analysis using selected ancillary datasets (i.e., cumulative NDVI, climate surfaces, and topographic data). Moreover, the spatial resolution of the original GLAS parameters and the extrapolated GLAS parameters are also different. These differences may lead the biophysical meanings of these three GLAS full-waveform parameters to become unclear. Although using spatial interpolation techniques (e.g. ordinary kriging and thin plate spline) to extrapolate GLAS parameters can help to preserve their biophysical meanings, the distribution of GLAS footprints is not random but concentrated around the ground tracks, which can result in significant strip effects to the extrapolated GLAS parameters.

Further, the extrapolation of GLAS parameters may even bring new error sources to the forest AGB estimation step. However, in the current study, the uncertainties induced by the extrapolated GLAS parameters as well as other prediction variables were not considered and evaluated in the forest AGB estimation procedure. The pixel level uncertainty can be evaluated in more detail by introducing a numerical error propagation model based on Monte Carlo simulation (Zhang et al., 2014b). A systematic evaluation of how uncertainties from different sources influence the final forest AGB distribution result will be conducted in a future study.

Additionally, the influence of forest growth on the AGB estimation result was not considered in this study. The plot measurements collected in this study were mainly taken between 2000 and 2010. The AGB increase brought by the forest growth might be non-negligible for certain plots, and result in mismatches between these plot measurements and remotely sensed covariables. If we only chose plot measurements from 2004 (the same year with remotely sensed datasets), the number of plot measurements would be too small to build a reliable forest AGB estimation model. In this study, the average of plot measurements with the same geolocations but measured at different years were used to represent the forest AGB condition at that plot. This may be helpful to partially reduce the

influence of forest growth on the final AGB estimation. However, collecting more ground inventory data for each calendar year is still needed to fully address this issue.

5.3 Global forest AGB distribution estimation

The pilot study in previous section shows the feasibility of the proposed algorithm in mapping large-scale forest AGB. Therefore, we further extended this method to global scale and mapped the global-scale forest AGB distribution. There are overall 4090 records of plot measurements from papers published between 1990 and 2013, and 3348 of them were retained for the forest AGB mapping procedure (Figure 5-11; Appendix B, Table B1 and Table B2).
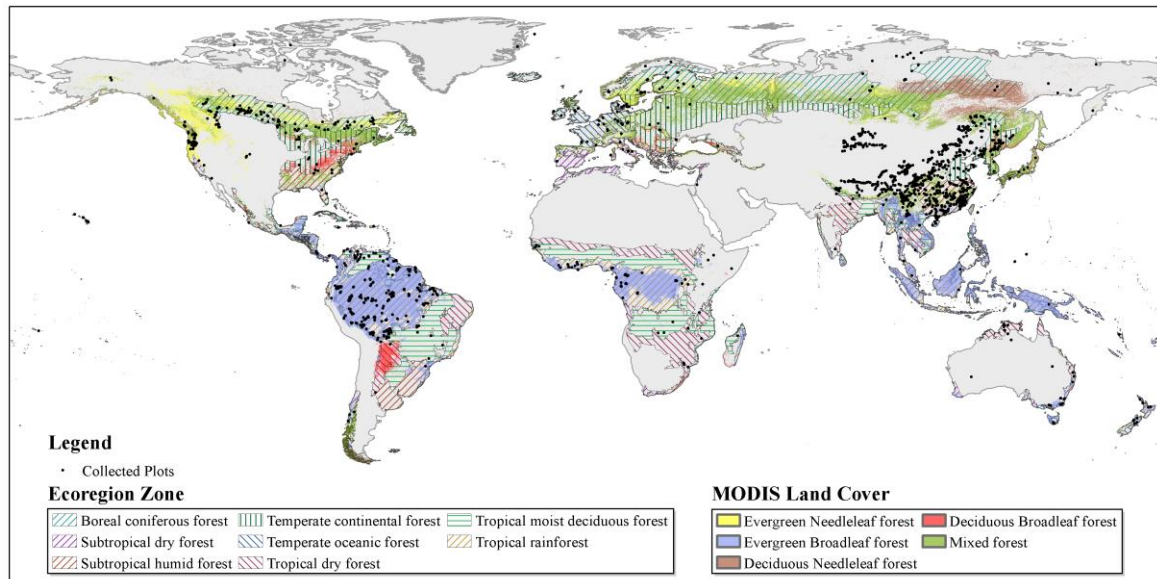

Figure 5-11 Geolocations of the collected ground inventory data. The background maps are the ecoregion zone map and land cover map from MODIS 2004.

The resulting global wall-to-wall forest AGB map is shown in Figure 5-12a. The RF regression tree model built from plot measurements without considering plot location uncertainty explained 61.21% of the variances in forest AGB and the root-mean-squared residual was 99.30 Mg/ha. The final global forest AGB result indicated that the mean AGB density for global forests was 210.09 Mg/ha with a standard deviation of 109.31 Mg/ha. The forest AGB density gradually increased from the boreal forest area to the tropical forest area (Figure 5-12a). The statistics of the national and continental forest AGB are shown in Table B3 and Table B4 (Appendix B). The mean uncertainty introduced by the plot location was 14.53 Mg/ha with a range of 3.75–127.69 Mg/ha (Figure 5-12b), and it accounted for lower than 10% of the AGB estimations in most regions (Figure 5-12c). Forests distributed in the European, North American, and Amazon areas generally had relatively low uncertainty values (<30 Mg/ha). In Southeast Asia, the influence of plot location uncertainty became larger, resulting in a forest AGB uncertainty of over 50 Mg/ha. The AGB estimation uncertainty introduced by the plot location was the highest on the western coast of North America and central Papua New Guinea, reaching values of 100 Mg/ha.
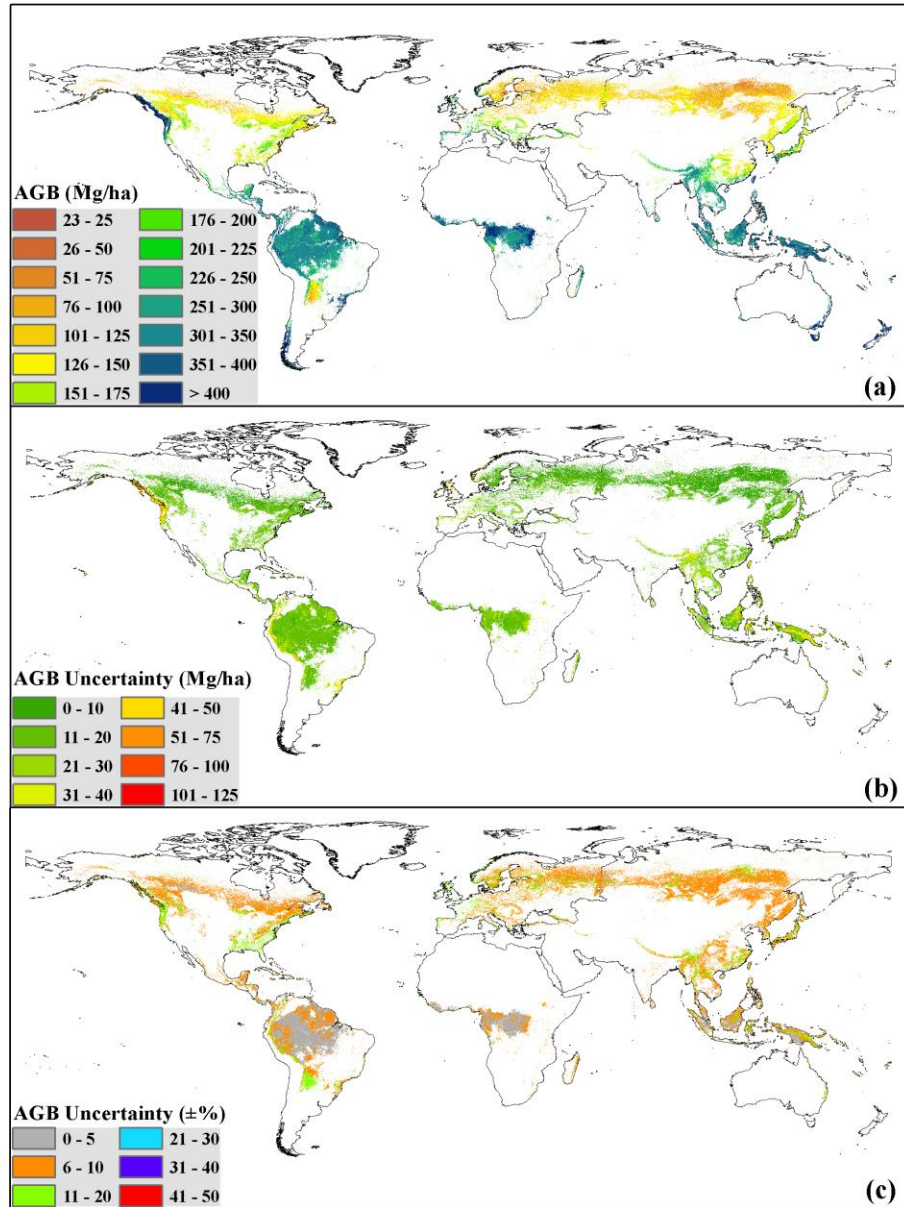
Figure 5-12 The derived wall-to-wall map of global forest AGB in this study; (b) the absolute uncertainty induced by plot location uncertainty (estimated as the standard deviation of the 100 RF run results); and (c) the relative uncertainty induced by plot location uncertainty. The study region was bounded at 80° N and 58° S and from longitude -180° to 180°.

Around one third of plot measurements were retained from the AGB mapping process to evaluate the forest AGB estimation result at the plot scale. As can be seen in Figure 5-13, the estimated forest AGB map showed a good agreement with the field-measured AGB. The fitted line between the observed values and predicted values was close to the 1:1 line. The statistical parameters also showed that the accuracy of the estimated global forest AGB

map was satisfactory. The R^2 between the observed and predicted AGB was 0.56 and the RMSE was 87.53 Mg/ha. However, the global forest AGB map had some limitations. It was inclined to overestimate forest AGB density in areas with low AGB density (<221 Mg/ha), and underestimate forest AGB density in areas with relatively high AGB density (>221 Mg/ha).
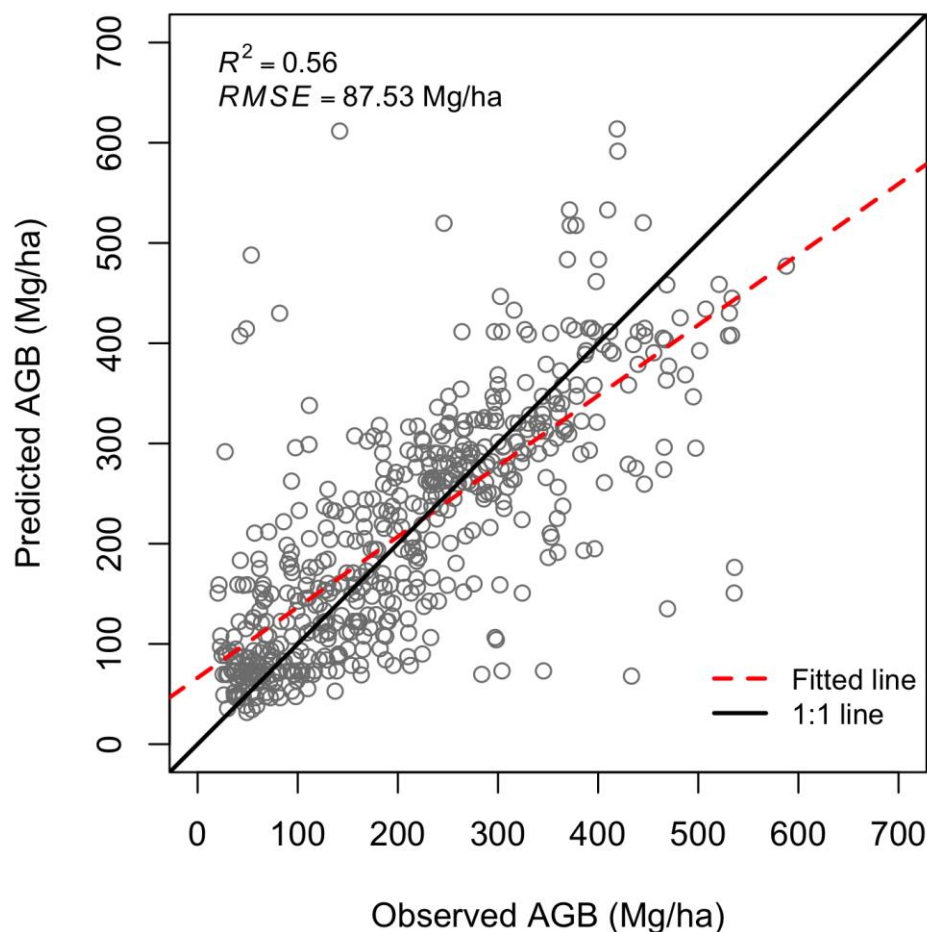
Figure 5-13 Evaluation of predicted forest AGB using validation ground inventory dataset at the plot level. R^2 represents the adjusted coefficient of determination, RMSE represents the root-mean-square error.

At the ecoregion zone level, the forest AGB map was compared with the AGB density values of the eight ecological zones in different continents provided by the Intergovernmental Panel on Climate Change (IPCC). As shown in Figure 5-14, the R^2 and RMSE between the IPCC suggested values and the predicted AGB were 0.38 and 101.21 Mg/ha, respectively. However, there was a clear outlier where the predicted forest AGB density for the North America temperate oceanic forest was significantly lower than the value provided by the IPCC. When we removed this data point, the R^2 increased to 0.56 and the RMSE decreased to 82.38 Mg/ha, which were very close to the plot-level validation results.
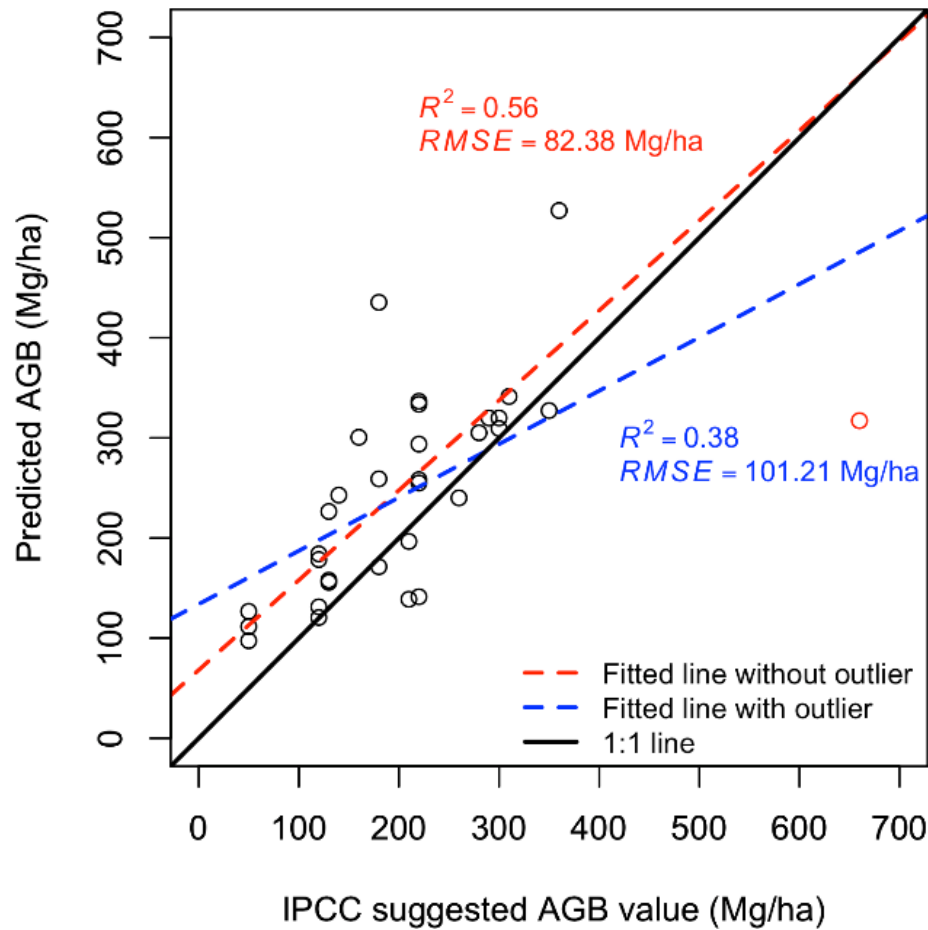
Figure 5-14 Comparison of the predicted forest AGB with IPCC suggested values at the ecological zone level. The outlier is the temperate oceanic forest in North America. R^2 represents the adjusted coefficient of determination, and RMSE represents the root-mean-square error.

Moreover, we further compared our global wall-to-wall forest AGB map with other eight regional forest AGB maps (Appdix C, Figure C1-C8) (Avitabile et al., 2015; Baccini et al., 2012; Blackard et al., 2008; Margolis et al., 2015; Neigh et al., 2015; Reusch & Gibbs, 2008; Saatchi et al., 2011; Thurner et al., 2014). Overall, our product corresponded well with the products generated based on LiDAR data, especially in tropical areas. Products generated using traditional optical passive remote sensing and radar data intended to underestimate the forest AGB density.

5.4 Conclusions

In this study, we developed a method to estimate forest AGB distribution through the combination of multi-source remote sensing data sets. The proposed method is firstly tested at the national scale of China. The results show that the proposed method can be used to accurately estimate forest AGB density in a large scale. The R^2 and RMSE between the

predicted map and independent field measurements are 0.75 and 42.39 Mg/ha, respectively. The plot location uncertainty can induce up to 10% uncertainty in the final forest AGB estimation, and generally the higher the forest AGB density is, the higher uncertainty induced by the plot location uncertainty. By considering the plot location uncertainty into the AGB estimation model, the forest AGB estimation accuracy can be significantly improved.

The high accuracy of the resulting national-scale forest AGB map proves the feasibility of the proposed method, and therefore we further use this method in the global scale. The resulting global-scale forest AGB map shows that the global forest AGB density is around 210 Mg/ha on average with a standard deviation of 109 Mg/ha. This global product is evaluated at the plot level and the result show that our product has a good correspondence with the independent plot measurements (R^2=0.56, RMSE=87.53 Mg/ha). At the ecoregion level, values recommended by the IPCC are used to compare with this new map which also show a good correspondence (R^2=0.38, RMSE=101.21 Mg/ha with one outlier; R^2=0.56, RMSE=82.38 Mg/ha after removing one outlier). This new global forest AGB map is highly beneficial to global carbon assessments and biogeochemical modellings. Moreover, it can also help to quantify changes in forest biomass caused by deforestation or afforestation.

References

Alvarez, O., Guo, Q., Klinger, R.C., Li, W. & Doherty, P. (2014). Comparison of elevation and remote sensing derived products as auxiliary data for climate surface interpolation. International Journal of Climatology, 34, 2258-2268.

Avitabile, V., Herold, M., Heuvelink, G.B.M., Lewis, S.L., Phillips, O.L., Asner, G.P., Armston, J., Asthon, P., Banin, L.F., Bayol, N., Berry, N. J., Boeckx, P., de Jong, B. H. J., DeVries, B., Girardin, C. A. J., Kearsley, E., Lindsell, J. A., Lopez-Gonzalez, G., Lucas, R., Malhi, Y., Morel, A., Mitchard, E. T. A., Nagy, L., Qie, L., Quinones, M. J., Ryan, C. M., Ferry, S. J. W., Sunderland, T., Laurin, G. V., Gatti, R. C., Valentini, R., Verbeeck, H., Wijaya, A., & Willcock, S. (2016). An integrated pan-tropical biomass map using multiple reference datasets. Global Change Biology, 22, 1406–1420.

Baccini, A., Laporte, N., Goetz, S., Sun, M. & Dong, H. (2008). A first map of tropical Africa's above-ground biomass derived from satellite imagery. Environmental Research Letters, 3, 045011.

Baccini, A., Goetz, S., Walker, W., Laporte, N., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P., Dubayah, R., & Friedl, M. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. Nature Climate Change, 2, 182–185.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M.A., Baldocchi, D. & Bonan, G.B. (2010). Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. Science, 329, 834-838.

Bergh, J., McMurtrie, R.E. & Linder, S. (1998). Climatic factors controlling the productivity of Norway spruce: a model-based analysis. Forest Ecology and Management, 110, 127-139.

Blackard, J.A., Finco, M.V., Helmer, E.H., & Holden, G.R. (2008). Mapping US forest biomass using nationwide forest inventory data and moderate resolution information. Remote Sensing of Environment, 112, 1658–1677.

Bonan, G.B. (2008). Forests and climate change: forcings, feedbacks, and the climate benefits of forests. Science, 320, 1444-1449.

Botkin, D.B. & Simpson, L.G. (1990). Biomass of the North American boreal forest: A step toward accurate global measures. Biogeochemistry, 9, 161-174.

Botkin, D.B., Simpson, L.G. & Nisbet, R.A. (1993). Biomass and carbon storage of the North American deciduous forest. Biogeochemistry, 20, 1-17.

Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L. & Kimes, D.S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. Remote Sensing of Environment, 112, 3876-3890.

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

Clark, M.L., Clark, D.B. & Roberts, D.A. (2004). Small-footprint LiDAR estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. Remote Sensing of Environment, 91, 68-89.

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, 2931-2942.

Dong, J., Kaufmann, R.K., Myneni, R.B., Tucker, C.J., Kauppi, P.E., Liski, J., Buermann, W., Alexeyev, V. & Hughes, M.K. (2003). Remote sensing estimates of boreal and temperate forest woody biomass: carbon pools, sources, and sinks. Remote Sensing of Environment, 84, 393-410.

Fang, J., Wang, G.G., Liu, G. & Xu, S. (1998). Forest biomass of China: an estimate based on the biomass-volume relationship. Ecological Applications, 8, 1084-1091.

Fang, J., Chen, A., Peng, C., Zhao, S. & Ci, L. (2001). Changes in forest biomass carbon storage in China between 1949 and 1998. Science, 292, 2320-2322.

Fang, J., Guo, Z., Hu, H., Kato, T., Muraoka, H. & Son, Y. (2014). Forest biomass carbon sinks in East Asia, with special reference to the relative contributions of forest expansion and forest growth. Global Change Biology, 20, 2019-2030.

Friedman, J., Hastie, T. & Tibshirani, R. (2001). The elements of statistical learning. Springer, Berlin: Springer series in statistics.

Galbraith, D., Levy, P.E., Sitch, S., Huntingford, C., Cox, P., Williams, M. & Meir, P. (2010). Multiple mechanisms of Amazonian forest biomass losses in three dynamic global vegetation models under climate change. New Phytologist, 187, 647-665.

Ghasemi, N., Sahebi, M.R. & Mohammadzadeh, A. (2011). A review on biomass estimation methods using synthetic aperture radar data. International Journal of Geomatics and Geosciences, 1, 776-788.

Guo, Q., Liu, Y. & Wieczorek, J. (2008). Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach. International Journal of Geographical Information Science, 22, 1067-1090.

Guo, Z., Hu, H., Li, P., Li, N. & Fang, J. (2013). Spatio-temporal changes in biomass carbon sinks in China's forests from 1977 to 2008. Science China Life Sciences, 56, 661-671.

Hou, X. (2001). Vegetation atlas of China. Chinese Academy of Science, The Editorial Board of Vegetation Map of China. Scientific Press, Beijing, 113-124.

Houghton, R. (2005). Aboveground forest biomass and the global carbon balance. Global Change Biology, 11, 945-958.

Iverson, L.R., Brown, S., Prasad, A., Mitasova, H., Gillespie, A.J. & Lugo, A.E. (1994). Use of GIS for estimating potential and actual forest biomass for continental South and Southeast Asia. Effects of Land-Use Change on Atmospheric CO2 Concentrations (ed. by V.H. Dale), pp. 67-116. Springer-Verlag, New York.

Keith, H., Mackey, B.G. & Lindenmayer, D.B. (2009). Re-evaluation of forest biomass carbon stocks and lessons from the world's most carbon-dense forests. Proceedings of the National Academy of Sciences, 106, 11635-11640.

Kindermann, G.E., McCallum, I., Fritz, S. & Obersteiner, M. (2008). A global forest growing stock, biomass and carbon map based on FAO statistics. Silva Fennica, 42, 387-396.

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

Lefsky, M.A. (2010). A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. Geophysical Research Letters, 37, L15401.

Lefsky, M.A., Keller, M., Pang, Y., De Camargo, P.B. & Hunter, M.O. (2007). Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. Journal of Applied Remote Sensing, 1, 013537.

Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Del Bom Espirito-Santo, F., Hunter, M.O. & de Oliveira, R. (2005). Estimates of forest canopy height and aboveground biomass using ICESat. Geophysical Research Letters, 32, L22S02.

Li, L., Guo, Q., Tao, S., Kelly, M. & Xu, G. (2015). LiDAR with multi-temporal MODIS provide a means to upscale predictions of forest biomass. ISPRS Journal of Photogrammetry and Remote Sensing, 102, 198-208.

Liaw, A. & Wiener, M. (2002) Classification and regression by randomForest. R News, 2, 18-22.

Liu, D., Chen, Y., Cai, W., Dong, W., Xiao, J., Chen, J., Zhang, H., Xia, J. & Yuan, W. (2014). The contribution of China's Grain to Green Program to carbon sequestration. Landscape Ecology, 29, 1675-1688.

Liu, J., Liu, M., Deng, X., Zhuang, D., Zhang, Z. & Luo, D. (2002). The land use and land cover change database and its relative studies in China. Journal of Geographical Sciences, 12, 275-282.

Lu, D. (2005). Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon. International Journal of Remote Sensing, 26, 2509-2525.

Lu, D. (2006). The potential and challenge of remote sensing‐based biomass estimation. International Journal of Remote Sensing, 27, 1297-1328.

Lu, D., Mausel, P., Brondızio, E. & Moran, E. (2004). Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. Forest Ecology and Management, 198, 149-167.

Luckman, A., Baker, J., Kuplich, T.M., Yanasse, C.d.C.F. & Frery, A.C. (1997). A study of the relationship between radar backscatter and regenerating tropical forest biomass for spaceborne SAR instruments. Remote Sensing of Environment, 60, 1-13.

Malhi, Y., Phillips, O., Lloyd, J., Baker, T., Wright, J., Almeida, S., Arroyo, L., Frederiksen, T., Grace, J. & Higuchi, N. (2002). An international network to monitor the structure, composition and dynamics of Amazonian forests (RAINFOR). Journal of Vegetation Science, 13, 439-450.

Margolis, H.A., Nelson, R.F., Montesano, P.M., Beaudoin, A., Sun, G., Andersen, H.-E., & Wulder, M.A. (2015). Combining satellite LiDAR, airborne LiDAR, and ground plots to estimate the amount and distribution of aboveground biomass in the boreal forest of North America. Canadian Journal of Forest Research, 45, 838–855.

Mitchard, E., Saatchi, S., Lewis, S., Feldpausch, T., Woodhouse, I., Sonké, B., Rowland, C. & Meir, P. (2011a). Measuring biomass changes due to woody encroachment and deforestation/degradation in a forest–savanna boundary region of central Africa using multi-temporal L-band radar backscatter. Remote Sensing of Environment, 115, 2861-2873.

Mitchard, E., Saatchi, S., Woodhouse, I., Nangendo, G., Ribeiro, N., Williams, M., Ryan, C., Lewis, S., Feldpausch, T. & Meir, P. (2009). Using satellite radar backscatter to predict above‐ground woody biomass: A consistent relationship across four different African landscapes. Geophysical Research Letters, 36, L23401.

Mitchard, E., Saatchi, S., White, L., Abernethy, K., Jeffery, K., Lewis, S., Collins, M., Lefsky, M., Leal, M. & Woodhouse, I. (2011b). Mapping tropical forest biomass with radar and spaceborne LiDAR: overcoming problems of high biomass and persistent cloud. Biogeosciences Discussions, 8, 8781.

Myneni, R., Dong, J., Tucker, C., Kaufmann, R., Kauppi, P., Liski, J., Zhou, L., Alexeyev, V. & Hughes, M. (2001). A large carbon sink in the woody biomass of northern forests. Proceedings of the National Academy of Sciences, 98, 14784-14789.

Neigh, C.S., Nelson, R.F., Ranson, K.J., Margolis, H.A., Montesano, P.M., Sun, G., Kharuk, V., Næsset, E., Wulder, M.A., & Andersen, H.-E. (2013). Taking stock of circumboreal forest carbon with ground measurements, airborne and spaceborne LiDAR. Remote Sensing of Environment, 137, 274-287.

Neigh, C.S., Nelson, R.F., Ranson, K.J., Margolis, H., Montesano, P.M., Sun, G., Kharuk, V., Naesset, E., Wulder, M.A., & Anderson, H. (2015). LiDAR-based biomass estimates, boreal forest biome, Eurasia, 2005–2006; ORNL Distributed Active Archive Center: Oak Ridge, Tennessee, USA.

Nelson, R., Ranson, K., Sun, G., Kimes, D., Kharuk, V., & Montesano, P. (2009). Estimating Siberian timber volume using MODIS and ICESat/GLAS. Remote Sensing of Environment, 113, 691-701.

Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A., Phillips, O.L., Shvidenko, A., Lewis, S.L. & Canadell, J.G. (2011). A large and persistent carbon sink in the world's forests. Science, 333, 988-993.

Peng, C., Zhou, X., Zhao, S., Wang, X., Zhu, B., Piao, S. & Fang, J. (2009). Quantifying the response of forest carbon balance to future climate change in Northeastern China: model validation and prediction. Global and Planetary Change, 66, 179-194.

Peterson, T.C. & Vose, R.S. (1997). An overview of the Global Historical Climatology Network temperature database. Bulletin of the American Meteorological Society, 78, 2837-2849.

Piao, S., Fang, J., Ciais, P., Peylin, P., Huang, Y., Sitch, S. & Wang, T. (2009). The carbon balance of terrestrial ecosystems in China. Nature, 458, 1009-1013.

Popescu, S.C., Zhao, K., Neuenschwander, A. & Lin, C. (2011). Satellite LiDAR vs. small footprint airborne LiDAR: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. Remote Sensing of Environment, 115, 2786-2797.

Powell, S.L., Cohen, W.B., Healey, S.P., Kennedy, R.E., Moisen, G.G., Pierce, K.B. & Ohmann, J.L. (2010). Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. Remote Sensing of Environment, 114, 1053-1068.

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

Rauste, Y. (2005). Multi-temporal JERS SAR data in boreal forest biomass mapping. Remote Sensing of Environment, 97, 263-275.

Ruesch, A., & Gibbs, H.K. (2008). New IPCC Tier-1 Global Biomass Carbon Map for the Year 2000; Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory: Oak Ridge, Tennessee, USA.

Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L. & Hagen, S. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. Proceedings of the National Academy of Sciences, 108, 9899-9904.

Sader, S.A., Waide, R.B., Lawrence, W.T. & Joyce, A.T. (1989). Tropical forest biomass and successional age class relationships to a vegetation index derived from Landsat TM data. Remote Sensing of Environment, 28, 143-198.

Schutz, B., Zwally, H., Shuman, C., Hancock, D. & DiMarzio, J. (2005). Overview of the ICESat mission. Geophysical Research Letters, 32, L21S01.

Simard, M., Pinto, N., Fisher, J.B. & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne LiDAR. Journal of Geophysical Research: Biogeosciences, 116, G04021.

Soenen, S.A., Peddle, D.R., Hall, R.J., Coburn, C.A. & Hall, F.G. (2010). Estimating aboveground forest biomass from canopy reflectance model inversion in mountainous terrain. Remote Sensing of Environment, 114, 1325-1337.

State Forestry Administration of China (2013). 2012 China Forestry Statistical Yearbook. China Forestry Publishing House, Beijing.

Su, Y. & Guo, Q. (2014). A practical method for SRTM DEM correction over vegetated mountain areas. ISPRS Journal of Photogrammetry and Remote Sensing, 87, 216-228.

Su, Y., Guo, Q., Ma, Q., & Li, W. (2015). SRTM DEM correction in vegetated mountain areas through the integration of spaceborne LiDAR, airborne LiDAR, and optical imagery. Remote Sensing, 7, 11202-11225.

Sun, G., Ranson, K., Masek, J., Guo, Z., Pang, Y., Fu, A. & Wang, D. (2008). Estimation of Tree Height and Forest Biomass from GLAS Data (< Special Issue> Silvilaser). Journal of Forest Planning, 13, 157-164.

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

Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., Shvidenko, A., Kompter, E., Ahrens, B. & Levick, S.R. (2014). Carbon stock and density of northern boreal and temperate forests. Global Ecology and Biogeography, 23, 297-310.

Tomppo, E., Gschwantner, T., Lawrence, M., & McRoberts, R.E. (2010). National forest inventories: Pathways for Common Reporting. New York: Springer.

U.S. Geological Survey (2013). Shuttle Radar Topography Mission: http://www2.jpl.nasa.gov/srtm/ (accessed 07 April, 2015).

Wang, Q., Adiku, S., Tenhunen, J. & Granier, A. (2005). On the relationship of NDVI with leaf area index in a deciduous forest site. Remote Sensing of Environment, 94, 244-255.

White, J., Coops, N. & Scott, N. (2000). Estimates of New Zealand forest and scrub biomass from the 3-PG model. Ecological Modelling, 131, 175-190.

Woodhouse, I.H., Mitchard, E.T., Brolly, M., Maniatis, D. & Ryan, C.M. (2012). Radar backscatter is not a'direct measure'of forest biomass. Nature Climate Change, 2, 556-557.

Xiao, J. (2014). Satellite evidence for significant biophysical consequences of the "Grain for Green" Program on the Loess Plateau in China. Journal of Geophysical Research: Biogeosciences, 119, 2261-2275.

Xu, T., & Hutchinson, M. (2011). ANUCLIM version 6.1 user guide. The Australian National University, Fenner School of Environment and Society, Canberra.

Zhang, C., Ju, W., Chen, J.M., Li, D., Wang, X., Fan, W., Li, M. & Zan, M. (2014a). Mapping forest stand age in China using remotely sensed forest height and observation data. Journal of Geophysical Research: Biogeosciences, 119, 1163-1179.

Zhang, G., Ganguly, S., Nemani, R.R., White, M.A., Milesi, C., Hashimoto, H., Wang, W., Saatchi, S., Yu, Y. & Myneni, R.B. (2014b). Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data. Remote Sensing of Environment, 151, 44-56.

Zhang, Y., Liang, S. & Sun, G. (2014c). Forest biomass mapping of northeastern China using GLAS and MODIS data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7, 140-152.

Chapter 6　　Conclusions

This dissertation shows that light detection and ranging (LiDAR) demonstrates a strong capability to penetrate forest canopy, making it a more effective tool than aerial imagery for capturing forest structure changes following forest fuel treatments. All change detection methods using LiDAR-derived forest structure parameters outperform the methods using aerial imagery, especially in areas with light forest fuel treatment activities. The method using LiDAR-derived canopy cover reaches an overall accuracy of 93.5% and a kappa coefficient (a statistic measuring inter-rater agreement) of 0.7, which are over 5% and 25% higher than the methods using aerial imagery. Moreover, through a direct comparison of bi-temporal LiDAR point clouds, this dissertation presents a new algorithm to quantify the severity of forest structure changes caused by forest fuel treatment activities, well beyond what can be done using methods based on aerial imagery. The systematic evaluation and the developed algorithm for detecting and quantifying forest fuel treatments provide a scalable approach to accurately monitor changes in mixed conifer forests.

The weakness of LiDAR in lacking spectral information can be partially compensated by fusing LiDAR data with multispectral aerial imagery. This dissertation focuses on mapping vegetation types at the regional scale through the fusion of LiDAR data and aerial imagery. The developed vegetation mapping strategy can classify vegetation units by considering both vegetation species information and forest structure characteristics. The Bayesian Information Criterion method is integrated into the algorithm, which can automatically determine the optimized number of vegetation groups. This proposed algorithm is shown to be robust in two mixed conifer forests in the Sierra Nevada, California. The overall accuracy and the kappa coefficient are over 78% and 0.64 for both study sites, which are comparable to traditional methods that only relied on multispectral imagery. Moreover, the proposed algorithm has a great advantage for mapping vegetation types in relative homogeneous forests. Comparing to methods using only multispectral imagery, the fusion of LiDAR and multispectral imagery can help to recognize more vegetation variances in an area by considering both vegetation structure characteristics and dominant species composition.

The fusion of multi-source remote-sensing data can also help to address the limitation related to the LiDAR spatial availability. The proposed algorithm of integrating field measurements, airborne LiDAR data, spaceborne LiDAR data, optical imagery and radar data can be successfully used to map fine-resolution tree height in a large area without full LiDAR coverage. In this dissertation, this algorithm is used to generate a tree height product across the whole Sierra Nevada at 70 m resolution. The estimated Sierra Nevada tree height product shows a good agreement with in-situ plot measurements. The coefficient of determination (R^2) is 0.60, and the root-mean-squared error (RMSE) is 5.45 m. The resolution of the generated Sierra Nevada tree height product is much finer than previously published tree height products in the Sierra Nevada (>1 km usually), and has a comparable accuracy with them. Moreover, it can greatly help to address the saturation effect of the tree height product derived from multispectral imagery only.

Through the fusion of spaceborne LiDAR data with other remotely sensed datasets, LiDAR can be used to increase the accuracy of national- to global-scale forest structure estimations at a relatively coarser resolution. This dissertation presents an algorithm to fuse field measurements, spaceborne LiDAR data, optical imagery, radar data and climate surfaces to map the spatial distribution of aboveground biomass at national to global scales. The integrated uncertainty field model can greatly help to reduce the influence of plot location uncertainty on the prediction results. Compared to the results without using the uncertainty field model, R^2 of the estimated forest aboveground biomass with the uncertainty field model increases by 17%, and the RMSE decreases by 16%. The proposed algorithm is successfully used to generate the first national-scale forest aboveground biomass map of China and the first global-scale forest aboveground biomass map at 1-km resolution. Both maps show good correspondences with independent field measurements (China: R^2=0.75, RMSE=42.39 Mg/ha; Global: R^2=0.56, RMSE=82.38Mh/ha). Moreover, both maps are less influenced by the saturation effect of optical imagery, compared to other published regional forest aboveground biomass maps. This new global forest AGB map is highly beneficial to improve our understanding on global carbon cycle and improve the accuracy of global biogeochemical modellings.

The proposed algorithms in this dissertation address the limitations of LiDAR data in forestry applications. However, this dissertation only uses tree height and AGB as examples in the regional- to global-scale forest structure estimations. In the future, we need to further extend these studies by generating other important forest structure parameters, such as canopy cover, leaf area index, tree age, and tree density. Moreover, the resolution of the generated forest AGB product at national- to global-scale is relatively coarser compared to the regional tree height product. With the increasing availability of the airborne LiDAR data across the world, future studies on generating global-scale fine-resolution forest structure estimations need to be conducted.
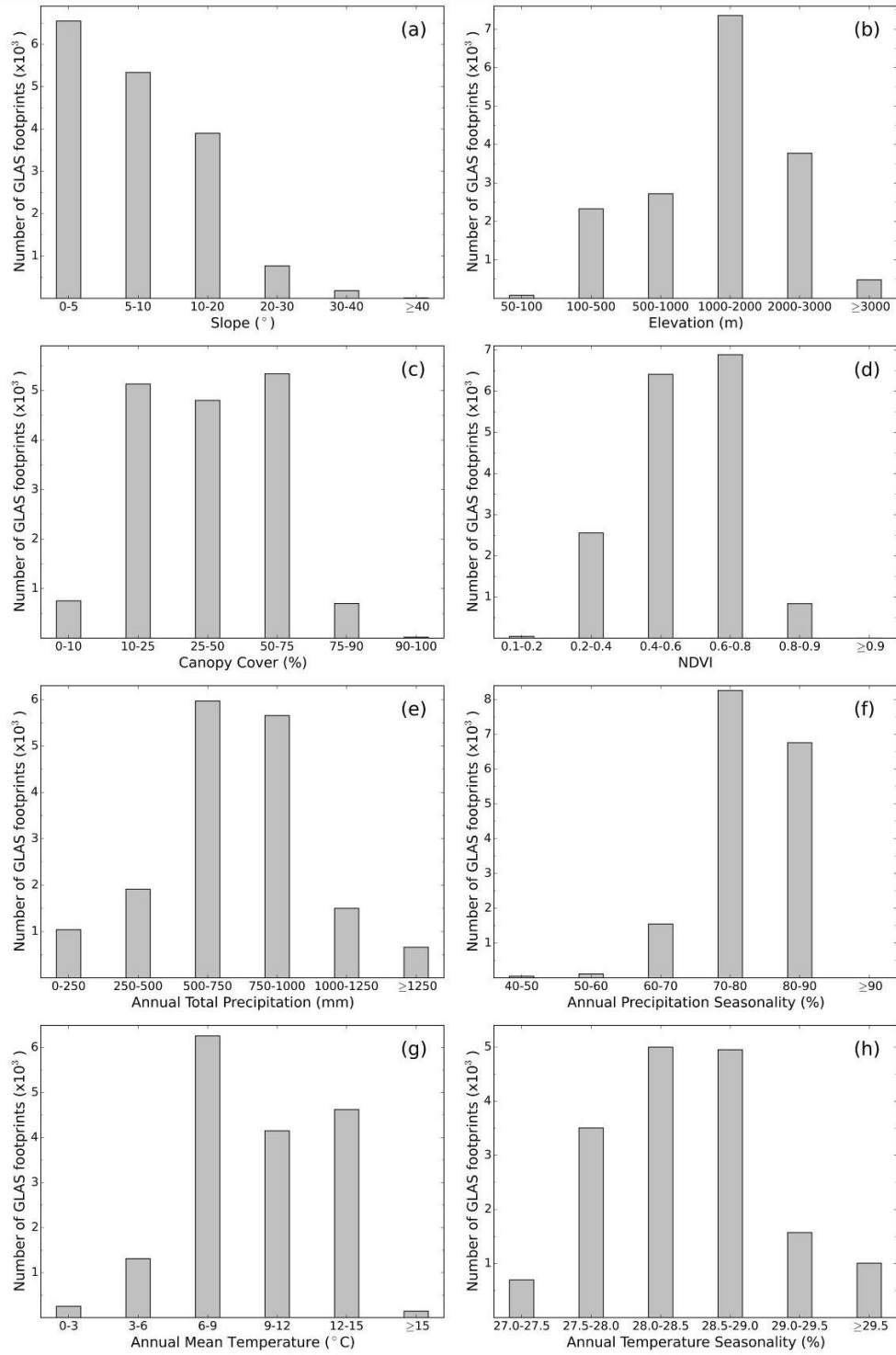
Appendix A

Figure A1 Numbers of GLAS footprints within different terrain, vegetation, and climate parameter groups (corresponding to Figure 4-5).

Appendix B

Table B1 The main literature provided the AGB using allometric equation.

ID	Author	Year	Title	Journal	Plots
1	Keeling and Phillips	2007	The global relationship between forest productivity and biomass.	Global Ecology and Biogeography	55
2	Liu, et al	2013	How temperature, precipitation and stand age control the biomass carbon density of global mature forests	Global Ecology and Biogeography	160
3	Malhi, et al.	2006	The regional variation of aboveground live biomass in old-growth Amazonian forests	Global Change Biology	227
4	Ma, et al	2012	Regional drought-induced reduction in the biomass carbon sink of Canada's boreal forests	PNAS	155
5	Mitchard, et al	2014	Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites	Global Ecology and Biogeography	413

Table B2 The distribution of plots in each continents and ecoregions.

Continent	Original	Filtered
Africa	157	118
Asia	2486	1932
Australia	68	49
Europe	163	127
North America	368	317
South America	814	787
Oceania	34	18
Total	4090	3348
Ecoregion	**Original**	**Filtered**
Boreal coniferous forest	244	219
Boreal mountain system	57	31
Boreal tundra woodland	19	5
Polar	15	1
Subtropical desert	36	27
Subtropical dry forest	14	10
Subtropical humid forest	976	899
Subtropical mountain system	274	238
Temperate continental forest	269	257
Temperate desert	50	9
Temperate mountain system	838	521
Temperate oceanic forest	87	80
Temperate steppe	90	32
Tropical desert	1	0
Tropical dry forest	41	31
Tropical moist deciduous forest	154	116
Tropical mountain system	59	51
Tropical rainforest	824	812
Tropical shrubland	10	7
Other	32	2
Total	4090	3348

Table B3 The estimated national-level forest AGB density and biomass stock from the generated global forest AGB map.

Country	Mean AGB, Mg/ha	Forest area, Mha	Total AGB, Pg
Angola	221.03 ± 42.44	2.8	0.63
Argentina	163.14 ± 95.93	16.39	3.33
Australia	415.85 ± 131.69	20.28	8.68
Austria	165.56 ± 25.48	3.96	0.66
Belgium	205.02 ± 54.55	0.67	0.14
Bangladesh	297.96 ± 33.34	0.45	0.14
Bulgaria	168.08 ± 22.79	2.68	0.45
Bosnia and Herzegovina	181.86 ± 22.16	1.99	0.36
Belarus	104.72 ± 10.28	7.59	0.79
Belize	302.78 ± 24.54	1.27	0.39
Bolivia	255.94 ± 59.79	38.66	10.45
Brazil	306.79 ± 36.1	317.34	97.44
Brunei Darussalam	279.71 ± 33.71	0.42	0.12
Bhutan	277.98 ± 36.84	2.45	0.68
Central African Republic	355.58 ± 36.77	4.32	1.54
Canada	141.38 ± 64.68	268.81	38.26
Switzerland	203.56 ± 74.53	1.17	0.23
Chile	410.6 ± 104.67	21.11	8.81
China	160.74 ± 45.16	101.34	16.41
Cameroon	373.62 ± 49.06	16.86	6.36
Democratic Republic of the Congo	342.01 ± 49.17	103.67	35.52
Republic of Congo	325.16 ± 52.88	17.72	5.85
Colombia	322.76 ± 53.71	59.34	19.33
Costa Rica	346.5 ± 52.41	1.99	0.7
Cuba	279.08 ± 56.83	1.08	0.31
Czech Republic	142.47 ± 20.76	2.56	0.37
Germany	146.04 ± 44.15	10.95	1.61
Dominican Republic	339.69 ± 39.43	0.66	0.22
Ecuador	328.16 ± 54.17	12.2	4.05
Spain	265.65 ± 67.1	4.2	1.11
Estonia	99.17 ± 8.07	2.53	0.25
Ethiopia	307.51 ± 54.37	2.36	0.75
Finland	89.42 ± 7.99	15.51	1.37
Fiji	316.05 ± 19.23	0.99	0.31
France	261.66 ± 85.86	14.92	4.05
Gabon	354.3 ± 56.66	15.59	5.55
United Kingdom	265.09 ± 87.3	3.22	0.87
Georgia	187.09 ± 27.24	2.21	0.42
Ghana	343.08 ± 27.22	1.42	0.48
Guinea	325.04 ± 31.96	0.51	0.17
Equatorial Guinea	382.23 ± 31.71	1.85	0.7
Greece	163.55 ± 20.46	1.31	0.21
Guatemala	290.15 ± 34.62	3.72	1.08
Guyana	333.05 ± 30.02	15.12	5.02
Honduras	312.28 ± 34.71	3.64	1.14
Croatia	164.64 ± 30.4	1.83	0.3
Hungary	137.06 ± 25.52	0.98	0.13
Indonesia	328.25 ± 43.39	102.79	34.04
India	253.38 ± 61.34	19.46	5.25
Ireland	246 ± 80.9	0.79	0.18

Table B3 (continued)

Country	Mean AGB, Mg/ha	Forest area, Mha	Total AGB, Pg
Iran	151.72 ± 17.77	1.19	0.18
Italy	188.72 ± 44.85	5.16	0.98
Jamaica	323.08 ± 23.02	0.37	0.12
Japan	175.78 ± 41.69	22.83	4.03
Kazakhstan	125.04 ± 26.21	1.02	0.13
Kenya	334.59 ± 107.12	0.55	0.2
Cambodia	279.47 ± 38.84	4.05	1.15
Republic of Korea	141.72 ± 9.57	4.16	0.59
Lao PDR	276.6 ± 26.71	13.42	3.73
Liberia	350.27 ± 31.01	4.27	1.5
Sri Lanka	279.46 ± 56.08	1.44	0.4
Lithuania	101.31 ± 11.39	1.89	0.19
Latvia	102.98 ± 9.67	3.31	0.34
Madagascar	239.17 ± 66.35	6.34	1.62
Mexico	243.66 ± 50.25	17.83	4.51
Myanmar	289.05 ± 40.36	24.65	7.25
Mongolia	106.86 ± 11.96	2.01	0.21
Malaysia	322.33 ± 35.03	20.64	6.69
New Caledonia	316.87 ± 35.14	0.66	0.21
Nigeria	322.46 ± 45.55	3.59	1.16
Nicaragua	329.52 ± 33.37	3.31	1.1
Norway	210.15 ± 116.76	9.04	1.95
Nepal	230.36 ± 40.42	3.16	0.74
New Zealand	488.69 ± 107.94	12.24	6.07
Panama	335.61 ± 30.78	3.82	1.29
Peru	316.49 ± 34.14	66.13	20.84
Philippines	336.33 ± 29.98	8.97	3.03
Papua New Guinea	357.15 ± 46.76	32.32	11.55
Poland	108.36 ± 23.93	8.78	0.95
Dem. Rep. Korea	143.6 ± 9.1	5.41	0.78
Portugal	253.7 ± 56.38	0.7	0.18
Paraguay	163.04 ± 67.7	6.83	1.28
Romania	163.32 ± 22.88	6.15	1.01
Russian Federation	110.14 ± 23.48	530.48	59.87
Solomon Islands	336.73 ± 25.59	2.13	0.71
Sierra Leone	297.04 ± 32.75	1.4	0.42
Serbia	177.91 ± 24.6	1.61	0.29
Suriname	339.81 ± 26.28	11.31	3.84
Slovakia	170.73 ± 23.31	2.22	0.38
Slovenia	173.12 ± 22.76	1.33	0.23
Sweden	108.11 ± 16.39	26.32	2.82
Thailand	279.58 ± 41.75	7.21	2.05
Turkey	175.62 ± 33.7	4.39	0.77
Taiwan	370.45 ± 79.48	1.76	0.67
Tanzania	231.22 ± 58.5	1.17	0.29
Uganda	391.6 ± 63.5	0.96	0.38
Ukraine	124.65 ± 30.26	7	0.88
United States	180.96 ± 92.44	150.61	27.71
Venezuela	318.49 ± 31.54	36.52	11.63
Vietnam	277.58 ± 34.41	9.35	2.62
Vanuatu	325.07 ± 23.98	0.71	0.23
South Africa	256.31 ± 93.44	1.45	0.4

Table B4 The estimated continental-level forest AGB density and biomass stock from the generated global forest AGB map.

| Continent	Mean AGB, Mg/ha	Forest area, Mha	Total AGB, Pg
Africa	333.34 ± 63.80	191.0	64.65
Asia	172.28 ± 94.75	762.2	143.14
Australia	415.66 ± 131.75	20.3	8.69
North America	166.48 ± 84.97	459.1	77.46
Oceania	424.30 ± 114.03	21.9	9.30
South America	301.68 ± 67.43	608.6	188.68
Europe	132.97 ± 50.70	310.1	40.83

Appendix C
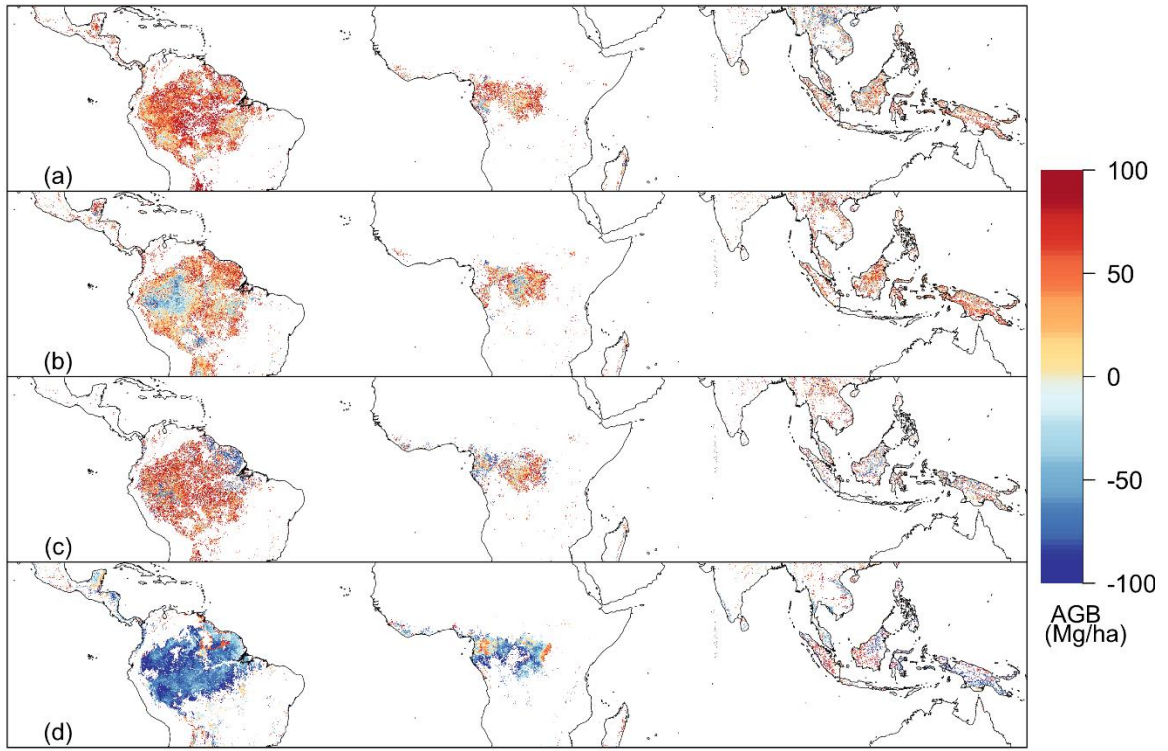
Figure C1 Differences between our global wall-to-wall forest AGB product and the products from (a) Saatchi et al. (2011), (b) Baccini et al. (2012), Avitabile et al. (2015), and (d) Reusch & Gibbs (2008) in pan-tropical forest areas.
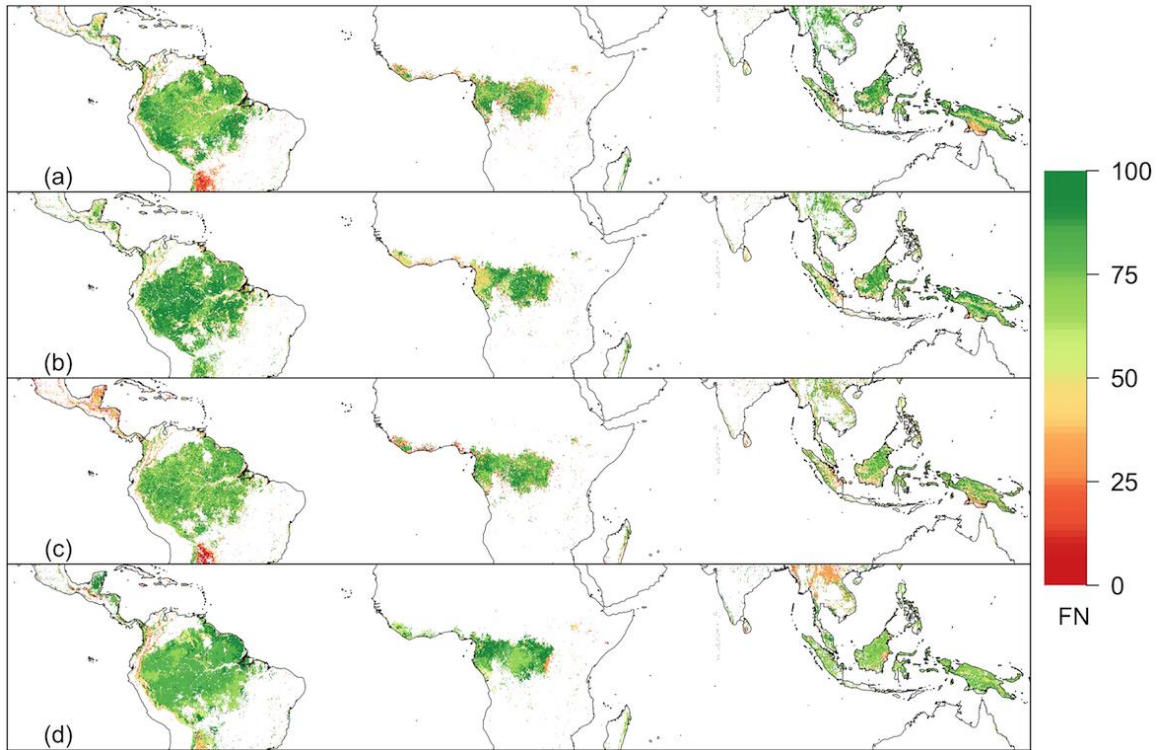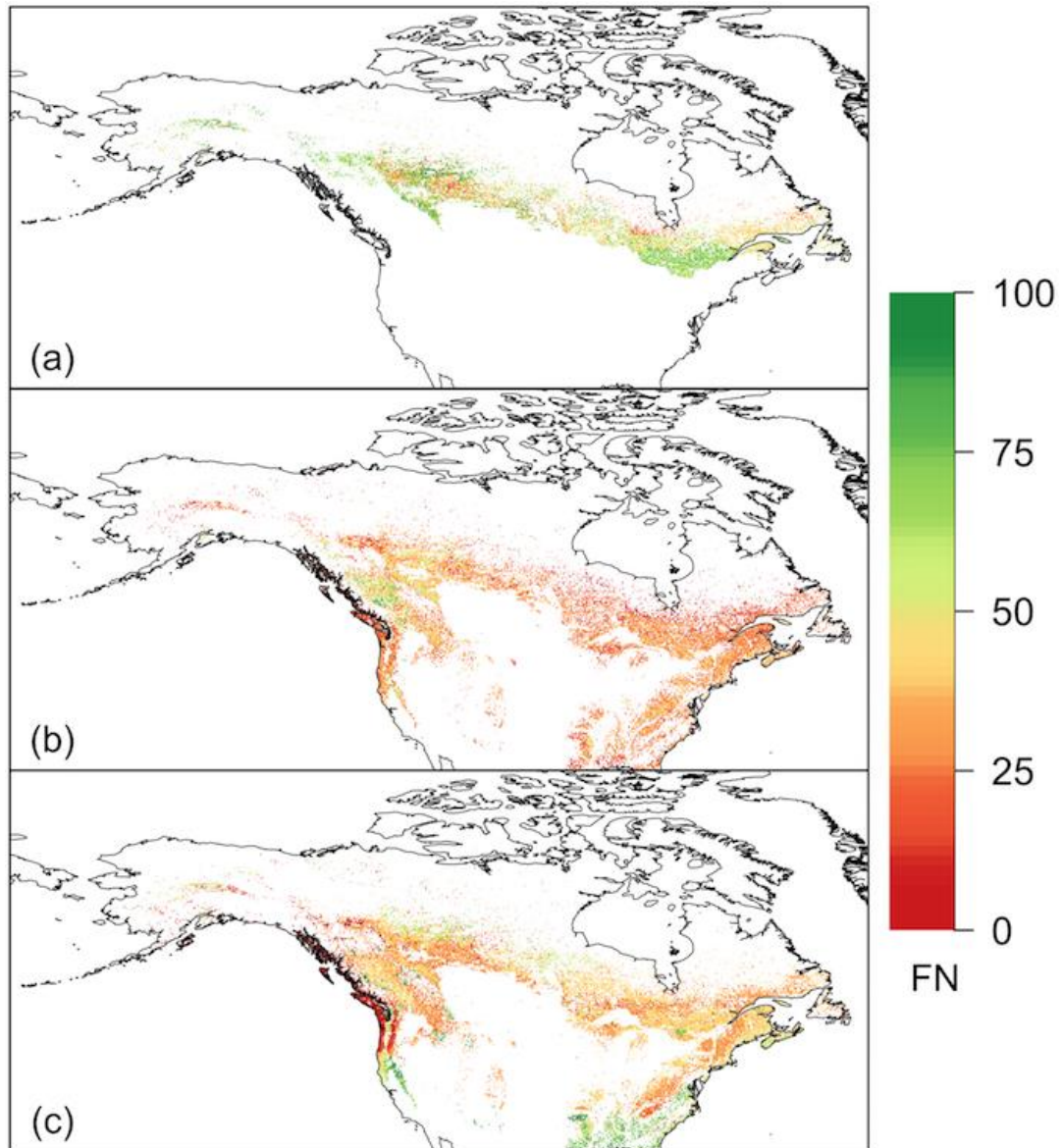
Figure C2 Spatial similarities between our global wall-to-wall forest AGB product and the products from (a) Saatchi et al. (2011), (b) Baccini et al. (2012), Avitabile et al. (2015), and (d) Reusch & Gibbs (2008) in pan-tropical forest areas. The spatial similarity is defined by the Fuzzy Numerical index, which ranges from 0 (fully distinct maps) to 100 (fully identical maps).
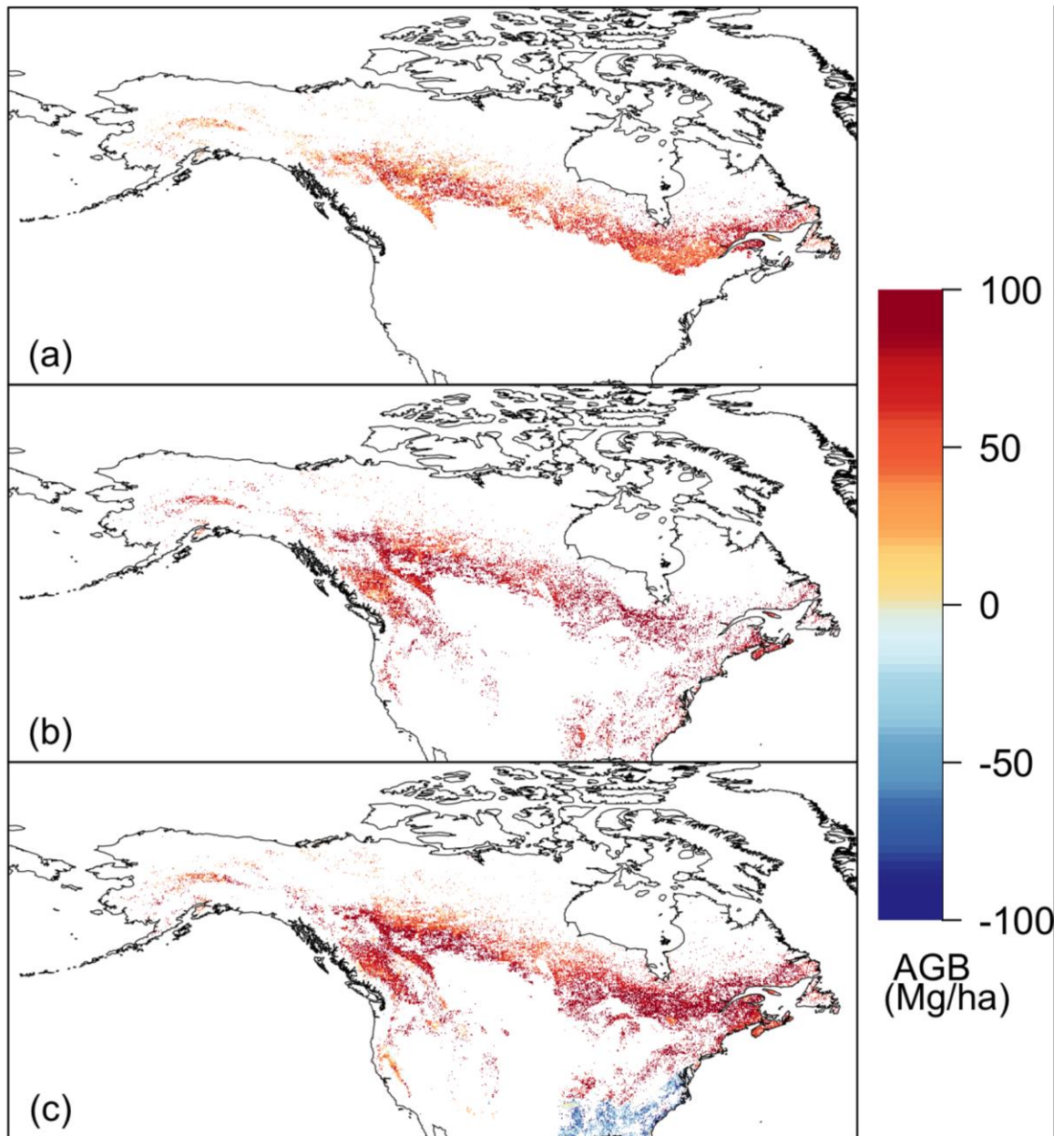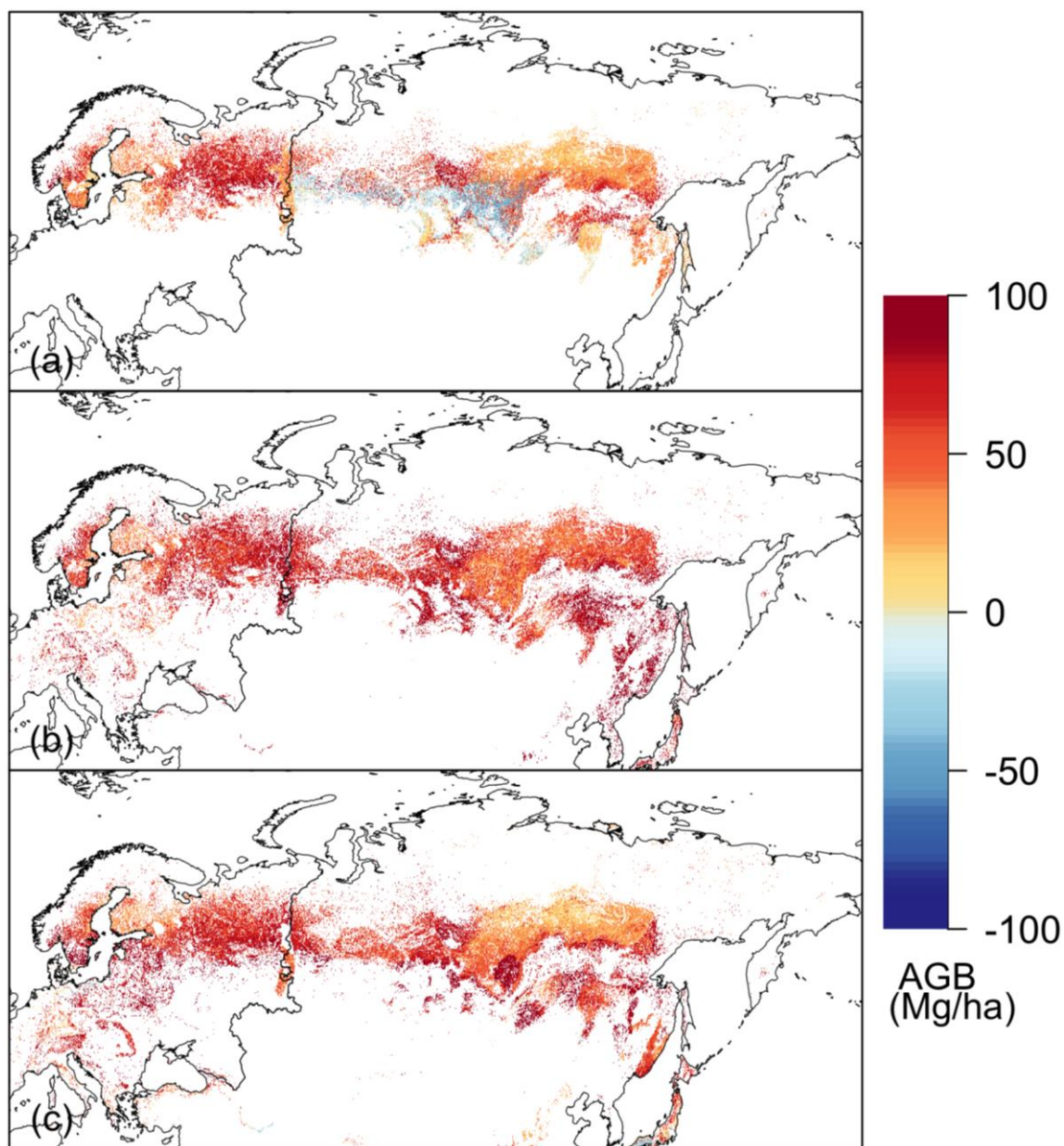
Figure C3 Differences between our global wall-to-wall forest AGB product and the products from (a) Margolis et al. (2015), (b) Thurner et al. (2014), and (c) Reusch & Gibbs (2008) in the North America.

Figure C4 Spatial similarities between our global wall-to-wall forest AGB product and the products from (a) Margolis et al. (2015), (b) Thurner et al. (2014), and (c) Reusch & Gibbs (2008) in the North America. The spatial similarity is defined by the Fuzzy Numerical index, which ranges from 0 (fully distinct maps) to 100 (fully identical maps).

Figure C5 Differences between our global wall-to-wall forest AGB product and the products from (a) Neigh et al. (2015), (b) Thurner et al. (2014), and (c) Reusch & Gibbs (2008) in the northern Eurasia.
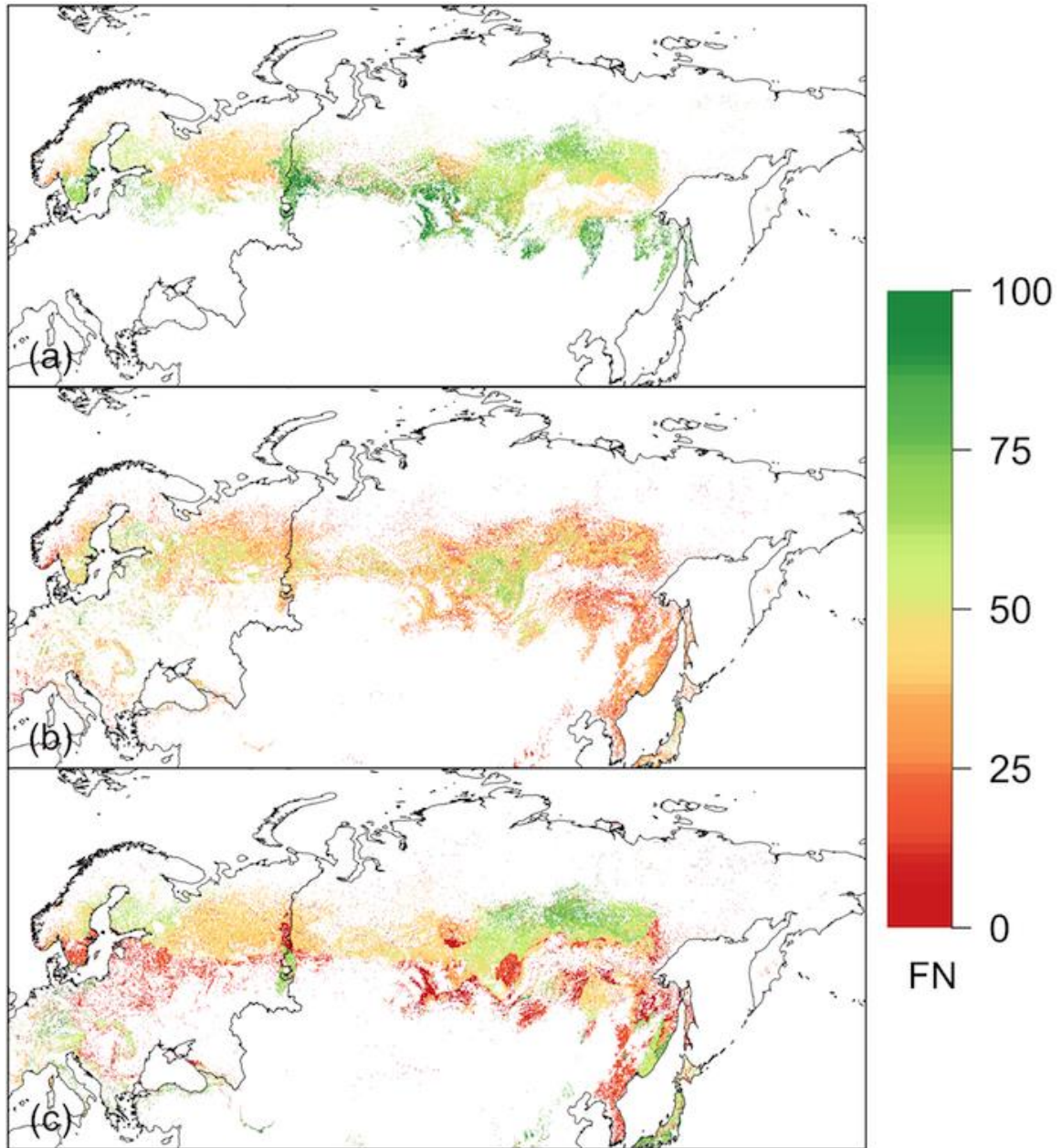
Figure C6 Spatial similarities between our global wall-to-wall forest AGB product and the products from (a) Neigh et al. (2015), (b) Thurner et al. (2014), and (c) Reusch & Gibbs (2008) in the northern Eurasia. The spatial similarity is defined by the Fuzzy Numerical index, which ranges from 0 (fully distinct maps) to 100 (fully identical maps).
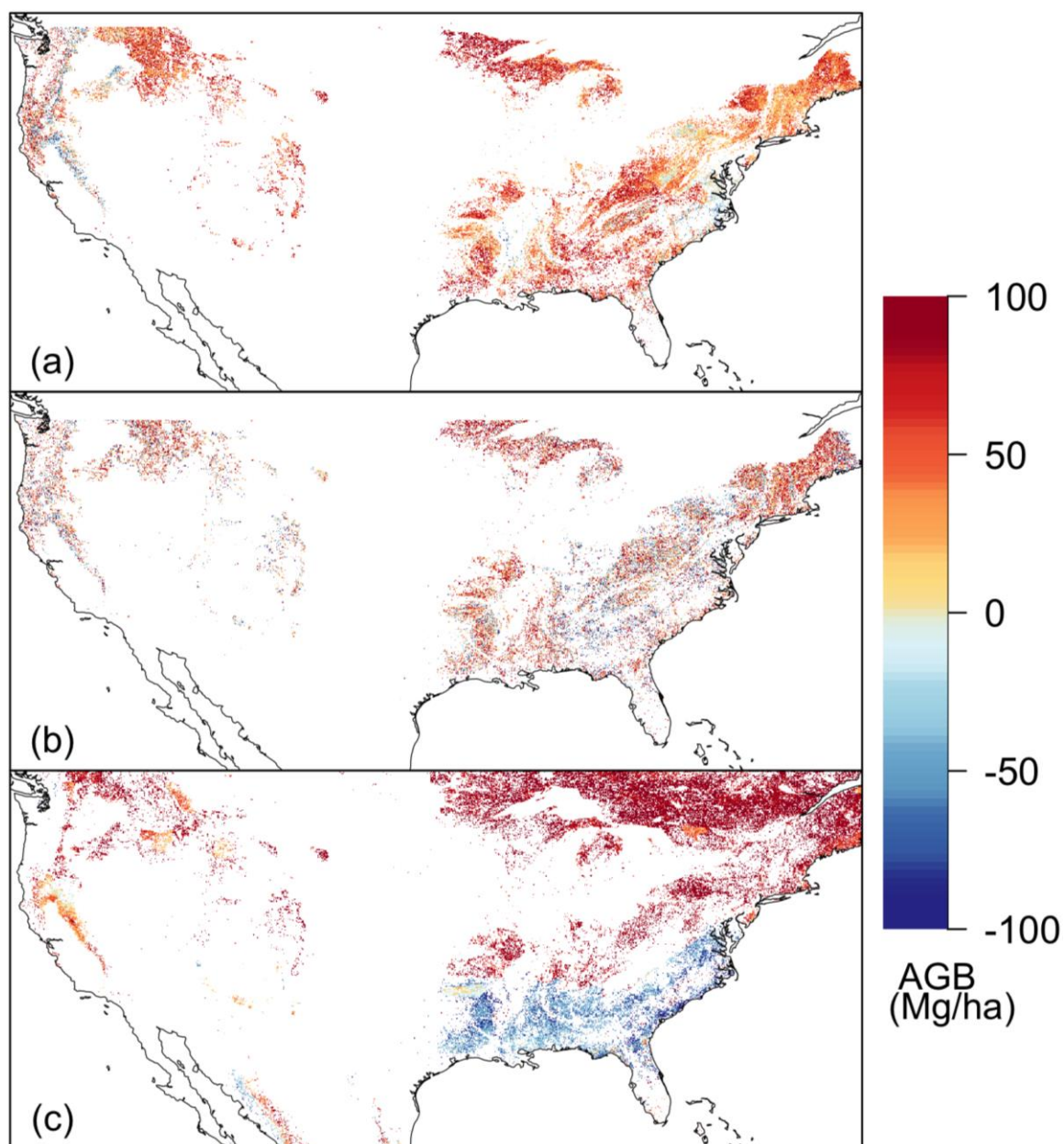
Figure C7 Differences between our global wall-to-wall forest AGB product and the products from (a) Blackard et al. (2008), (b) Saatchi et al. (2005), and (c) Reusch & Gibbs (2008) in the U.S.
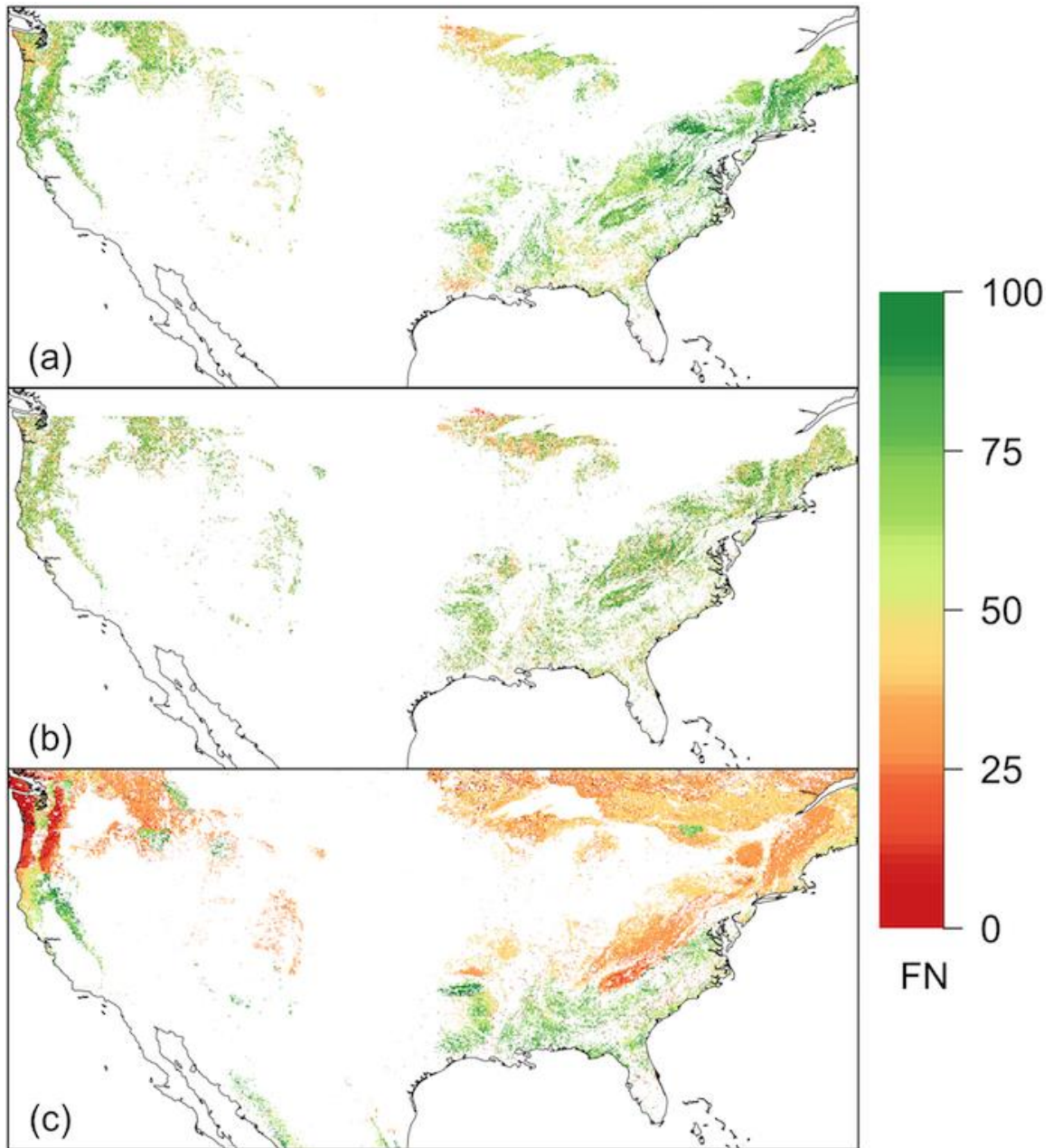
Figure C8 Spatial similarities between our global wall-to-wall forest AGB product and the products from (a) Blackard et al. (2008), (b) Saatchi et al. (2005), and (c) Reusch & Gibbs (2008) in the U.S. The spatial similarity is defined by the Fuzzy Numerical index, which ranges from 0 (fully distinct maps) to 100 (fully identical maps).