INTEGRATION OF LIDAR REMOTE SENSING FROM MULTIPLE PLATFORMS TO ASSESS VEGETATION BIOPHYSICAL PARAMETERS

A Dissertation

by

NIAN-WEI KU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Sorin C. Popescu
Committee Members,	Marian Eriksson
	X. Ben Wu
	Anthony Filippi
Head of Department,	Kathleen Kavanagh

May 2018

Major Subject: Ecosystem Science and Management

ABSTRACT

This research concentrates on using multiple platforms of lidar remote sensing for assessing vegetation biophysical parameters. Airborne and spaceborne light detection and ranging (lidar) (i.e., ICESat) remote sensing can characterize the three-dimensional structure of vegetation and therefore can provide useful information for assessing forest and rangeland woody plant biomass. The objectives of this research are 1) developing robust methods using airborne lidar and multispectral data to generate a local woody plant biomass map in northern Texas, 2) investigating the accuracy of existing global forest canopy height maps using airborne lidar data in multiple ecoregions in the southern United States, and 3) upscaling local forest aboveground biomass estimates to regional scale in an ecoregion. This research integrates statistical methods and remote sensing techniques to develop the procedure for building the regional forest aboveground biomass map. First, this research results in an approach for employing both airborne lidar and multispectral data with statistical methods to create a local scale woody plant aboveground biomass map in northern Texas. Then, the validation and calibration of the global forest canopy height map (GCHM) are used throughout rangelands and forests in the southern United States. A calibrated global forest canopy height map (cGCHM) serves as a primary data source for upscaling the forest aboveground biomass map from the local- to regional-scale in the South Central Plains ecoregion. In summary, the research utilized lidar data which was collected from multiple platforms to estimate aboveground biomass at multiple scales.

DEDICATION

To my parents and brother.

ACKNOWLEDGEMENTS

I sincerely appreciate my advisor, Dr. Sorin C. Popescu, for guiding me patiently and helpfully in my research. I truly thank my committee, Dr. Marian Eriksson, Dr. X. Ben Wu, and Dr. Anthony Filippi, for their constructive comments and support throughout the course of this research.

Thanks also go to my colleagues of Lidar Applications for the Study of Ecosystems with Remote Sensing Laboratory (LASERS) and friends, as well as the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my mother, father, and brother for their encouragement, patience, and love.

CONTRIBUTORS AND FUNDING SOURCES

This research work was supervised by a dissertation committee consisting of Professors Sorin C. Popescu, Marian Eriksson, and X. Ben Wu of the Department of Ecosystem Science and Management, and Professor Anthony Filippi of the Department of Geography.

All work for the dissertation was completed independently by the student.

This work was supported by the ICESat-2 SDT program of the National Aeronautics and Space Administration (NASA) under Grant Number NNX15AD02G and the RAPID Response program of NASA under Grant NNX14AN99G. The student support provided as Teaching Assistantships by the Department of Ecosystem Science and Management at Texas A&M University.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND BACKGROUND

The research employs remotely sensed data to assess vegetation biophysical parameters on rangelands and forests. Remote sensing technology enables us to collect measurements across large extents and estimate the vegetation biophysical parameters while minimizing or avoiding the need for destructive sampling. Remotely sensed data have been utilized to study, among other things, forest diseases and insect prevention (Bhattacharya and Chattopadhyay, 2013, Olsson et al., 2012), forest fuel and wild fire observation (Chuvieco et al., 2010), forest carbon storage and aboveground biomass estimation (Watts et al., 2009, Anaya et al., 2009), and encroachment of woody plants on rangelands (Mohamed et al., 2011). This research utilize remotely sensed data to estimate the woody plant aboveground biomass of forest and rangeland vegetation in the southern United States

Remote sensing technology has two main categories based on the type of sensor used to collect data. Passive remote sensing measures solar energy which is reflected by or emitted by objects of interest. For example, multispectral remote sensing instruments receive and record the intensity of the visible wavelengths which is reflected or emitted from the objects. On the other hand, active remote sensing instruments produce and emit energy. That energy is reflected by objects, returned back to the instrument which records attributes of the returned energy. An example of an active remote sensing system

is lidar, which transmits a low-power laser pulse to the objects and records the energy returned from each pulse, along with positional information, and the time of flight. This information can be used to infer the distance of objects from the platform. In this research we integrate both passive (multispectral imagery) and active (lidar data) remote sensing technologies for estimating forest aboveground biomass in this study.

Mainly, this research concentrates on the applications of lidar remote sensing. Lidar remote sensing technology has been utilized for forest measurements (e.g., tree height) since the 1980s. Nelson et al. (1984) operated an airborne laser altimeter system to investigate the profile and explore the structure of forests. Recently, lidar remote sensing data has been incorporated into various forest and rangeland studies in order to derive tree characteristics (Hilbert et al., 2011), estimate aboveground vegetation biomass (Lefsky et al., 2002a, Ni-Meister et al., 2010), explore forest structure (Morsdorf et al., 2009), assess forest carbon storage (Asner et al., 2012), and identify individual trees on rangelands (Sankey and Bond, 2011).

This research addressed three issues concerning (1) the suitability of certain statistical methods for aboveground woody biomass estimation, (2) the validation and recalibration of a global canopy height model (GCHM), and (3) the feasibility of mapping regional forest aboveground biomass using data obtained from the combination of multiplatform lidar and multispectral imaging systems. The results of our investigations into these issues are reported in Chapters II–IV, respectively, of this dissertation. In the past, most of the research into lidar-based remote sensing of aboveground biomass has relied on lidar-only-sensed (LOS) information (Nelson et al.,

1988, Simard et al., 2006, Hawbaker et al., 2009, Lefsky et al., 1999a, Naesset, 2011). In Chapter II spatially explicit estimation of aboveground woody biomass on rangelands in northern Texas from airborne lidar-and-multispectral (LAMS) data is considered. While lidar data are used to compute various height measures, multispectral data have often been used to generate vegetation indices which have, in turn, been related to biophysical characteristics of the vegetation (Lefsky et al., 1999a, Jensen, 2007, Gartzia et al., 2014, Asner et al., 2015). Spatially explicit analyses allow for the mapping of predicted values. In addition to the considering LAMS, three different statistical methods stepwise regression, the least absolute shrinkage and selection operator (LASSO), and random forests (Random Forests™) for analyzing the data were compared. These three statistical methods have been applied to many remote sensing studies (Pal, 2005, Verbesselt et al., 2009, Tian et al., 2012, Zandler et al., 2015). Besides, the woody plant aboveground biomass estimation was calculated by three statistical methods which are the stepwise regression, the least absolute shrinkage and selection operator (LASSO), and random forests. Thus, Chapter II investigates the appropriate combination of lidar and multispectral remote sensing variables with proper statistical methods for developing the local woody plant aboveground biomass map.

Chapter III documents the results of our validation/recalibration effort for the southern United States. In this research, we validate and calibrate Simard et al. (2011) GCHM primarily. Lefsky (2010) created the first GCHM in 2010, and Simard et al. generated another GCHM in 2011. Both GCHMs integrated the data of the Geoscience Laser Altimeter System (GLAS) which was the sole instrument on ICESat (Ice, Cloud,

and land Elevation Satellite) as their primary remotely sensed data. Lefsky also combined The Moderate Resolution Imaging Spectroradiometer images (MODIS) and GLAS data using mining approach to generate a continuous GCHM. Simard et al. (2011) used Shuttle Radar Topography Mission (STRM), GLAS, climatic and other ancillary data with the regression tree method Random Forest to generate the GCHM. Then Simard et al. (2011) compared their GCHM to Lefsky's and concluded that their GCHM covered a larger forested area than Lefsky's because Simard's GCHM involved mosaic crops, open forest, and saline flooded forests. Therefore, Simard's map reveals not only the tall trees in forests but also the short woody plants on the ground (Simard et al., 2011b). Furthermore, Bolton et al. (2013) investigated the accuracy of Lefsky and Simard's GCHM over Canada with airborne lidar data. Therefore, we are interested in validating the GCHMs with airborne lidar remote sensing data on rangelands and forests in the southern United States.

The validation results from Chapter III indicated that maps produced from the selected GCHM would benefit from a recalibration of the model. We document our recalibration effort in Chapter IV where we produce a new forest aboveground biomass map of South Central Plains ecoregion. The random forests statistical method was used to develop the forest aboveground biomass map. We employed the airborne lidar data to create local forest aboveground biomass maps as the response variable. Then cGCHMs, the MODIS images, vegetation indices, and canopy cover data were used as the prediction variables. Finally, we compared our forest aboveground biomass map of

South Central Plains to United States Forest Services' (USFS) forest aboveground biomass map (Blackard et al., 2008) at South Central Plains to investigate differences.

CHAPTER II

A COMPARISON OF MULTIPLE METHODS FOR MAPPING LOCAL-SCALE WOODY PLANT ABOVEGROUND BIOMASS WITH REMOTELY SENSED DATA

II.1. Introduction

The map of woody plant aboveground biomass allows a better understanding of carbon stocks and fluxes in forest or rangeland ecosystems. Thus, the accuracy of the estimation of woody plant aboveground biomass plays a critical role. Conventionally, logging is the convenient method to provide reliable and accurate estimation of dried woody plant aboveground biomass. However, the treatment is destructive and strongly disturbs local vegetation ecosystems (Vashum and Jayakumar, 2012). Therefore, remote sensing technology provides a non-destructive method to investigate the woody plant aboveground biomass estimation and limit the negative impacts of destructive methods to prevent disturbances in the forest and rangeland ecosystems (Robert et al., 2014, Calders et al., 2015, Galidaki et al., 2016).

Based on the approaches to retrieve the remotely sensed data, the remote sensing technically divides into passive and active remote sensing. Multispectral remote sensing (i.e., aerial photography and satellite imagery) passively records the spectral reflectance from the Earth's surface. The intensity of the spectral reflectance displays the spectral features of the objects on the ground. The multispectral remote sensing can collect data from local to global scales. The technique is broadly employed in the studies of vegetation distribution (Ansley et al., 2001, Hilker et al., 2014, Dardel et al., 2014) and

the land cover and land use classification (Jin et al., 2013, Hansen et al., 2013). Moreover, multispectral remote sensing data derives vegetation indices by combining different spectral bands. These vegetation indices are used to calculate critical biophysical indicators for extracting and modeling biophysical variables of vegetation (Jensen, 2007). The normalized difference vegetation index (NDVI), one of the vegetation indices, has been known for enhancing the sensitivity of live vegetation analysis and related to several measurable vegetation biophysical parameters, such as leaf area index (LAI) (Bannari et al., 1995, Carlson and Ripley, 1997). Moreover, NDVI has significant relationships with vegetation greenness for improving land cover classification (Anderson et al., 1993, DeFries and Townshend, 1994) and with vegetation biomass studies (Myneni et al., 2001, Cho et al., 2007, Raynolds et al., 2012). Besides, Zheng et al. (2004) discovered the aboveground biomass of pine forest had a strong correlation to the corrected normalized difference vegetation index (NDVIc). Casady et al. (2013) reported that NDVI highly correlated to the shrub biomass in the Sonoran and Mojave Desert with both satellite and ground observation data. Kushwaha et al. (2014) noticed the growing stock and woody biomass, including invasive shrubs, were significantly correlated to NDVI. Therefore, the multispectral dataset and NDVI are derived and computed from the National Agriculture Imagery Program (NAIP) imagery for estimating local scale woody plant aboveground biomass.

Unlike multispectral remote sensing, light detection and ranging (lidar) is an active remote sensing technique that generates three-dimensional point cloud data of target objects. The lidar point cloud data is capable of explicitly depicting the vertical

structure of vegetation and deriving vegetation biophysical parameters, such as tree height, canopy height, diameter at breast height (DBH), and tree crown width (Lefsky et al., 1999b, 2002b). The lidar systems are widely utilized in multiple vegetation ecology and ecosystem studies on various platforms (ex. terrestrial, airborne, mobile, and spaceborne). For instance, Riaño et al. (2007) combined airborne lidar data and color infrared images to identify the height and vegetation index of shrubs, but the height of shrubs was underestimated by the airborne lidar system. Estornell et al. (2011) demonstrated that the airborne lidar systems offer reliable shrub biomass estimation with sufficient point density. Colgan et al. (2012) also flew over South African savanna areas with an airborne lidar system and established a lidar-biomass regression for the savanna biomass estimation. Moreover, Ku et al. (2012) employed a terrestrial lidar system to develop lidar-biomass regressions for the available mesquite tree biomass models at a plot-level area in the northern Texas rangelands. Although lidar systems have proven their capacity for many woody plant aboveground biomass studies, considering the spatial scale of our research, we employed an airborne lidar system to estimate the local scale woody plant aboveground biomass on rangelands.

The woody plant aboveground biomass estimation was calculated by three statistical methods, which are the stepwise regression, the least absolute shrinkage and selection operator (LASSO), and random forests. In practice, the stepwise regression is the most commonly used method to derive the woody plant aboveground biomass (Ku et al., 2012, Tian et al., 2012, Mutanga et al., 2012, Zandler et al., 2015). This regression method selects prediction variables based on statistical criteria like the Akaike

information criterion (AIC) or Bayesian information criterion (BIC). Moreover, three approaches, forward selection, backward elimination, and bidirectional elimination (the combination of forward selection and backward elimination), are utilized to conduct the variable selection. The multiple and adjusted R-squared values are used to analyze the strength of the relationship. However, multicollinearity commonly happens in stepwise regression, so the variance inflation factor (VIF) is used to eliminate problematical variables.

The LASSO regression method is similar to the stepwise regression for predictive variable selection but applies the regularization to improve the prediction accuracy and statistical model interpretation (Tibshirani, 1996). The method uses the constrained form, sometimes called the penalized form, to determine a smaller variance of estimates and subset of prediction variables. The constrained form shrinks the nonzero coefficients and sets others to zero while retaining the significant features of subset selection and ridge regression. Because of the abilities of variable selection and regularization, LASSO is employed in many remote sensing research studies (Verbesselt et al., 2009, Kankare et al., 2013, Kantola et al., 2010). For instance, Vastaranta et al. (2012) selected LASSO to be the primary statistical method to find the best model and avoid over-fitting problems. Furthermore, Zandler et al. (2015) concluded that LASSO had the best performance for quantifying dwarf shrub biomass compared to stepwise regression and random forests methods. Therefore, LASSO is employed to create woody plant aboveground biomass maps and investigate the accuracy of the maps at the local scale.

The last statistical method introduced to this research is a machine learning method, random forests (Breiman, 2001). The method is proven to produce accurate predictions without overfitting the data and has been applied to many ecological studies (Pal, 2005, Cutler et al., 2007, Prasad et al., 2006). Random forests consists of a combination of decision trees, and each decision tree votes for the most popular class to classify training data**.** At the beginning of the procedure of random forests, the procedure utilizes the bootstrap aggregating method (Bagging) to generate the training data by randomly drawing with replacement samples (Breiman, 1996). Other data values which are not selected as training data are called out-of-bag data. The training data is randomly selected and employed as predictors to find the best split at each decision tree node. After the training data selection, the training data is drawn to construct numerous decision trees (500 to 2000 trees) and classified by the most popular voted class from all the decision trees in the forest. A large number of trees limits the generalization error without overfitting and increases the accuracy of the prediction. The out-of-bag data is not utilized in the fitting of the decision trees but is used to calculate an unbiased error rate and variable importance, or cross-validation. Hence, the probabilities of training data for the different classes are evaluated by the proportions of out-of-bag samples in each class.

Also, random forests is considered a "black box" since it is difficult to examine the individual decision trees separately. Fortunately, the variable importance aids to interpret the results of random forests. However, variable importance is a complicated and challenging concept to define in general, because the importance of a variable

depends on the interaction with other variables. Therefore, the random forests method investigates how much prediction error increases to define the importance of a variable when the out-of-bag data of that variable is randomly permuted while all others are left unchanged (Liaw and Wiener, 2002). Moreover, the result of variable importance can also compare relative importance among all predictor variables.

In this research, we attempted to calculate the honey mesquite (*Prosopis glandulosa*) aboveground biomass with the novel LASSO regression and random forests methods. Furthermore, we generated the mesquite aboveground biomass at a very high spatial resolution (1 m^2). Therefore, this research expects to develop the mesquite aboveground biomass maps with novel statistical methods at a high spatial resolution.

Three statistical methods were employed with three groups of remotely sensed data inputs; 1) multispectral data or NAIP imageries, 2) airborne lidar point cloud data, and 3) the combination of NAIP and airborne lidar point cloud data. Given the combination of statistical methods and remotely sensed data, the study produced equations of mesquite tree aboveground biomass estimation. After the comparison, the suitable statistical method was chosen to create a local-scale mesquite tree aboveground biomass map. The specific objectives of the research are to 1) investigate the mesquite tree aboveground biomass estimation with the three statistical methods and remotely sensed data, and then 2) generate the mesquite tree aboveground biomass map at a local scale with the most suitable combination of the statistical method and remotely sensed data.

II.2. Materials and Methods

II.2.1. Study area and field measurements

The study area is the Smith-Walker research unit (34˚02' N, 99˚14' W) of Texas A&M AgriLife Research and Extension Center at Vernon, Texas, USA (Figure 1). The research unit includes approximately 223 hectares (550 acres) of cultivated land and 684 hectares (1690 acres) of rangeland. Texas Rollin Plain tall grasses (Texas wintergrass (*Nassella leucotricha*), buffalograss (*Bouteloua dactyloides*), vine mesquite (*Hopia obtusa*), and dropseeds (*Sporobolus indicus*)) and woody plants (honey mesquite (*Prosopis glandulosa*) and shinnery oak (*Quercus havardii*)) dominate the rangeland of the research unit. The tall grasses are the background vegetation with the height ranging from 20 cm to 180 cm and the mesquite tree dominant patches spread throughout the research unit. The mesquite tree patches are varied in size, where small patches contain only one or two trees, but large patches are comprised of hundreds of trees. The majority of the mature mesquite trees are located in the northern portion of the research unit while the young and new regrowth are found throughout.

The mesquite tree aboveground biomass field data was collected during the leaf-off season in December 2008, March 2009, and December 2009 from 25 study plots in the Smith-Walker research unit. Each study plot was 5 m wide and 20 m long within a similar age and height distributions of the mesquite tree. The study plots were chosen by the age of the mesquite trees. Hence, the study plots had from regrowth to mature mesquite trees and the mesquite tree height range from 1 m in regrowth study plot to 6 m in the mature study plot. The study plot shape and dimensions were designed to

efficiently field measurements and estimate standing mesquite tree biomass with an allometric equation (Equation 1) which relies on the basal stem diameters. The study plot size was also considered a right balance between a larger plot that would be time-consuming for field measurement and a smaller plot that would be too small to capture sufficient represented mesquite trees. The allometric equation relates total tree mass to basal stem diameter at 5 cm to 15 cm height above ground (Ansley et al., 2010). The plot-level aboveground biomass was calculated as the sum of total tree mass per plot and expressed as kg/m^2.

Equation 1

$$Y = 0.34\, x^{1.73}$$

Where *Y* represents the total tree mass (kg) and the *x* is Basal stem diameter (cm). Afterward, the unit of mesquite tree aboveground biomass field data was converted to Mg/ha for the future calculation. The calculation of mesquite tree aboveground biomass field data at each1 m^2 pixel was described in the "d. Mesquite tree aboveground biomass estimation" section.

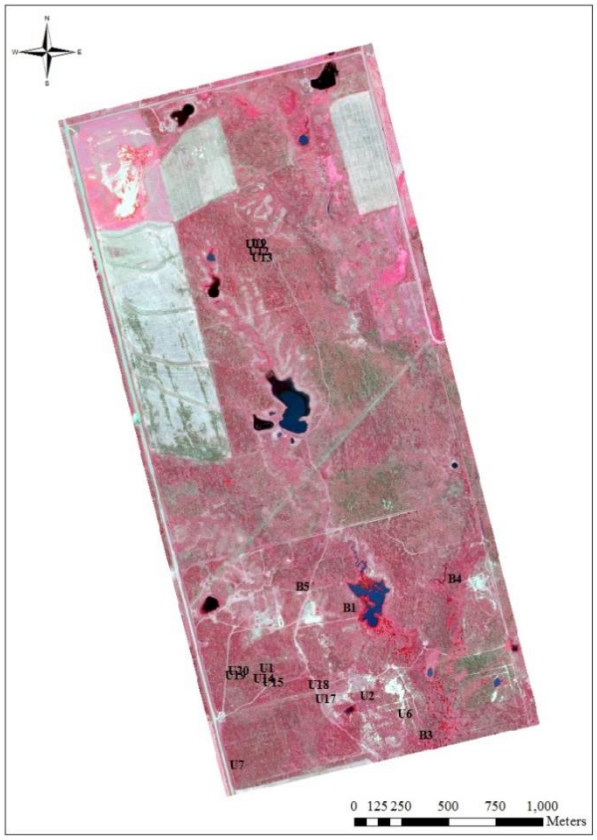

Figure 1. The locations of all mesquite tree study plots were shown on the Smith-Walker research unit color-infrared aerial imageries. The plot labels are displayed in black. The plot labels collide together due to the scale of map

II.2.2. NAIP imagery and NDVI calculation

Two NAIP images, southeastern Lockett and southwestern Boggy Creek, were requested from the Texas Natural Resources Information System (TNRIS). The NAIP imagery was a four-band data including three visible bands (blue, green, and red) and one near-infrared band (NIR) at 1 m spatial resolution. Both NAIP images were acquired during agricultural growing seasons (leaf-on) in 2010 and rectified in the Universal Transverse Mercator (UTM) coordinate system, North American Datum of 1983 (NAD 83) in zone 14.

All multispectral variables for predicting the mesquite tree aboveground biomass, including blue, green, red, and NIR bands, were extracted from the NAIP imagery. Furthermore, the red and NIR bands were employed to generate the NDVI layer of the Smith-Walker research unit based on the following equation:

Equation 2

$$\mathrm{NDVI} = (\mathrm{NIR} - \mathrm{Red})/(\mathrm{NIR} + \mathrm{Red})$$

In Equation 2, the NIR represents the near-infrared band, and the Red represents the red band of the NAIP imagery.

II.2.3. Airborne lidar data collection, preprocessing, and lidar metrics extraction

The airborne lidar data was collected by the Riegl 560 airborne laser scanner during the leaf-off season in November 2010. The plane flew at an average altitude of 600 m above ground level, and the average speed was 203.72 km/h (110 knots). A total of six flight lines covered the Smith-Walker research unit, and the swath width of each flight line is about 693 m with 50% forward overlap in a northeast to southwest direction. The scan angles were ±30° (cumulative 60°) from nadir, and the point density was approximately 3 to 4 points/m^2. The laser pulse rate and lidar system scan rate was 150 kHz and 90 Hz, respectively.

The raw airborne lidar point cloud data included the terrain elevation, vegetation heights, and redundant (i.e., power line or power tower points) or noise (i.e., points beneath the ground or in the sky) points. Thus, preprocessing is essential to remove unnecessary information and retain useful information. First, the ground points were identified and used to generate a 1 m^2 gridded digital elevation model (DEM). Then we acquired relative heights above ground by subtracting the terrain elevation from the raw

lidar point cloud data. Next, the lidar point cloud data above 7 m and below 1 m relative heights were removed to avoid noise, errors, tall grass, and non-vegetated structures. Subsequently, the final lidar point cloud data created a 1 m^2 gridded canopy height model (CHM). Meanwhile, the pixel-based lidar point cloud data of the mesquite tree study plots were extracted from the final lidar point cloud data.

Table 1. The list of the NAIP imagery and airborne lidar metric variables

1) Red band	18) 5th percentile value
2) Green band	19) 10th percentile value
3) Blue band	20) 20th percentile value
4) Near Infrared band	21) 25th percentile value
5) NDVI	22) 30th percentile value
6) Canopy Height Model	23) 40th percentile value
7) Minimum height\Minimum value	24) 50th percentile value
8) Maximum height\Maximum value	25) 60th percentile value
9) Mean height\Mean value	26) 70th percentile value
10) Mode height (The most count of returns)\mode value	27) 75th percentile value
	28) 80th percentile value
11) Standard deviation	29) 90th percentile value
12) Variance	30) 95th percentile value
13) Coefficient of variation	31) 99th percentile value
14) Interquartile range (IQR)	32) Generalized means for the 2nd (Elevation quadratic mean) power p=2
15) Skewness computed*	
16) Kurtosis computed**	33) Generalized means for the 3rd (Elevation cubic mean) power p=3
17) 1st percentile value	

* The Skewness computed indicates asymmetric distribution.
** The Kurtosis computed indicates the tail size of the data distribution.

The lidar metrics at 1 m2 grids was derived by the FUSION/LDV version 3.60+ software of the US Department of Agriculture, Forest Service (McGaughey, 2016). The mesquite tree study plots had 413 pixels. However, the lidar metrics would not be computed if the numbers of points at each pixel were less than 4 points. Thus, only 330 pixels of the field mesquite tree study pixels had the lidar metrics. The multispectral

variables of NAIP imagery and lidar metrics were the prediction variables for the mesquite tree aboveground biomass estimation and listed in Table 1.

II.2.4. Mesquite tree aboveground biomass estimation

The mesquite tree aboveground biomass in a 5 x 20 m field plot was allocated to each 1 x 1 m pixel corresponding to each 1 m^2 NAIP pixel. The biomass allocation was weighted by the maximum height in each pixel to compensate for the unequal distribution of biomass throughout the 5 x 20 m field plot. Thus, pixels with higher maximum height had more mesquite tree aboveground biomass than the pixels with lower maximum height in the same mesquite tree aboveground biomass field data plot. The calculation of the mesquite tree aboveground biomass at each NAIP pixel is showed in Equation 3.

Equation 3

$$\boldsymbol{Biomass_{pixel} = Biomass_{plot} \times \left(\frac{Height_{max}}{\sum_{1}^{m} Height_{max}} \right)}$$

Where $Biomass_{pixel}$ (Mg/ha) represents the mesquite tree aboveground biomass at each 1 m^2 NAIP pixel. $Biomass_{plot}$ (Mg/ha) represents the mesquite tree aboveground biomass at each study plot. $Height_{max}$ (m) represents the lidar metric maximum height at each pixel. m represents the numbers of pixels of a study plot. Finally, the mesquite tree aboveground biomass at each 1 m^2 NAIP pixel was utilized as the response variable for developing the local scale mesquite tree aboveground biomass maps.

The three statistical methods, stepwise regression, least absolute shrinkage and selection operator (LASSO), and random forests were employed to develop the mesquite

tree aboveground biomass estimation models to generate the mesquite tree aboveground biomass maps. The study utilized the R programming language to process all three statistical methods with “glmnet” (Friedman et al., 2010) and “ModelMap” (Freeman and Frescino, 2016) packages. The data of prediction and response variables were split into 80% training and 20% testing data by simple random sampling.

The stepwise regression method selected the prediction variables with the bidirectional elimination as well as the AIC criteria in this research. After the primary prediction variables were determined, the variance inflation factor (VIF) was used to investigate multicollinearity effects and remove problematical prediction variables when VIF was higher than 10. The surviving prediction variables would build the final mesquite tree aboveground biomass estimation equation. Similarly, the LASSO method used the “glmnet” package and set the one standard error rule to generate the most regularized model in this research. After both regression methods established the mesquite tree aboveground biomass equations, the equations were applied in ArcGIS to build the mesquite tree aboveground biomass maps. The testing data was utilized to validate the accuracy of the mesquite tree aboveground biomass maps made by stepwise regression and LASSO, respectively. The random forests method utilized the same training data and executed the ModelMap package. The ModelMap package not only built the random forests model to estimation mesquite tree aboveground biomass but also created the mesquite tree aboveground biomass maps. Thus, all prediction variables were converted to layers and applied to build the mesquite tree aboveground biomass map. The random forests method found the prediction variable importance and built the

mesquite tree aboveground biomass model. The out-of-bag (OOB) data of the random forests method was utilized to investigate the accuracy of the mesquite tree aboveground biomass map. To understand the accuracies of these mesquite tree aboveground biomass estimation equations and models, we investigated the coefficient of determination (R^2) of every equation and model. Besides, for evaluating the accuracy of the mesquite tree aboveground biomass maps, the mean square errors (MSE) were calculated for the mesquite tree aboveground biomass maps' validation.

The multispectral variables of the NAIP imagery and lidar metrics were assigned to three configurations of prediction variables: (1) multispectral variables of the NAIP imagery, (2) lidar metrics, and (3) a combined data of multispectral variables and lidar metrics. Nonetheless, we found the regression between the response and prediction variables is not linear. Hence, the logarithmic transformation was applied to all three statistical methods to improve the accuracy of mesquite tree aboveground biomass estimation. Moreover, Felker et al. (1982) applied the logarithmic transformation to estimate woody plant aboveground biomass and found the high R-squared value (0.9). After processing the combinations of three statistical methods and three configurations of prediction variable data, the comparison of the results reveals the most feasible statistical method and prediction variable for creating mesquite tree aboveground biomass map. Figure 2 showed the procedures for generating the local scale mesquite tree aboveground biomass maps of this research.

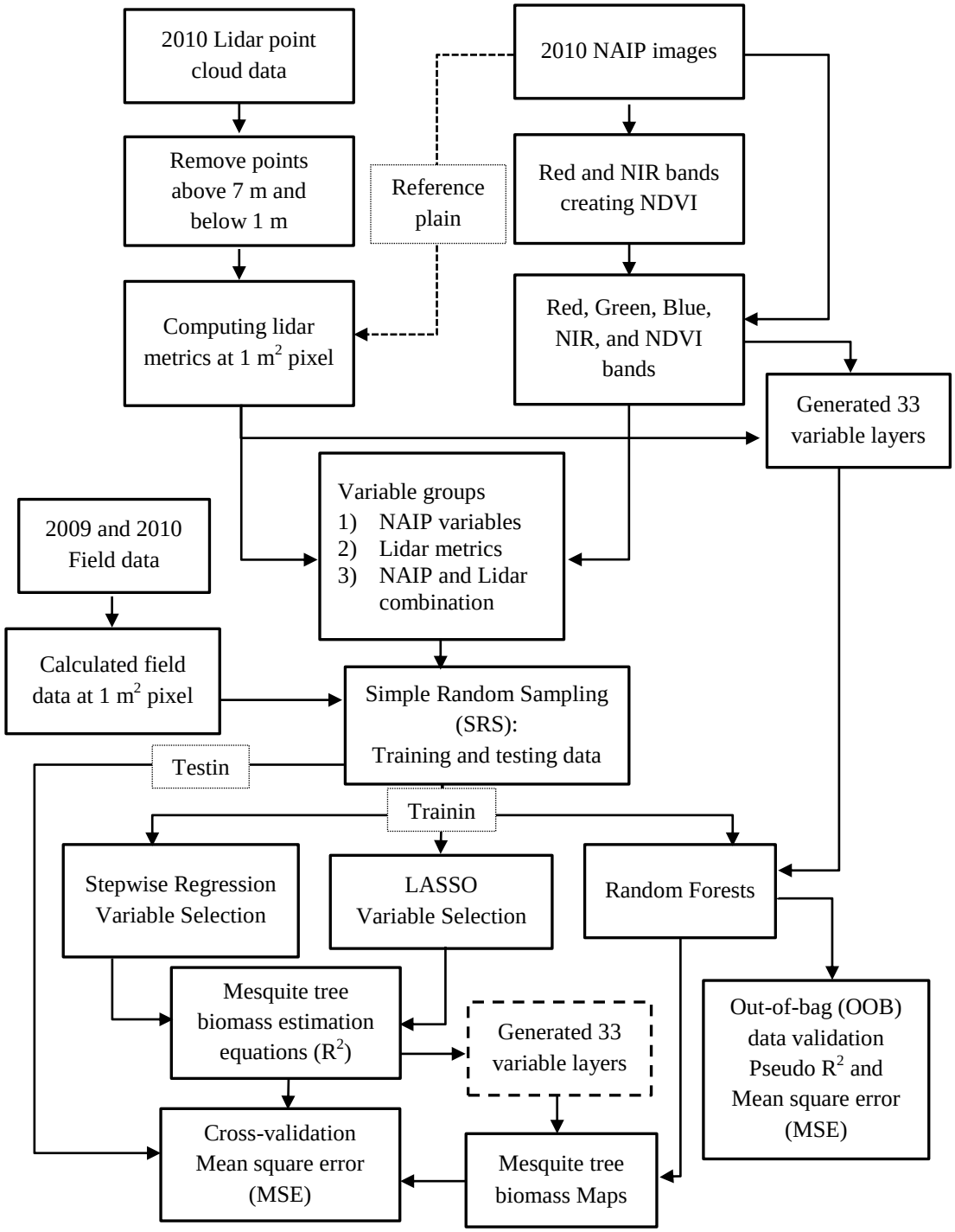

Figure 2. The flowchart of generating local scale mesquite tree aboveground biomass maps

II.3. Results

II.3.1. The mesquite tree aboveground biomass estimation equations and models

The three statistical methods result in 12 mesquite tree aboveground biomass estimation equations from stepwise regression and LASSO methods and 6 random forests models for the mesquite tree aboveground biomass estimation after computing with the three configurations of prediction variable data. The mesquite tree aboveground biomass is only estimated and mapped by the best mesquite tree aboveground biomass equation or model. However, the rest of mesquite aboveground biomass equations and models still show in the results.

Table 2. The adjusted and multiple coefficients of determination for stepwise regression in mesquite tree aboveground biomass estimation

Variable group	Logarithm transformation	Selected variable	Coefficient	Adj. R^2	R^2
NAIP	No	Intercept	4.38	0.04	0.05
		Red	-0.04		
	Yes	Intercept	1.65	0.06	0.06
		Red	-0.02		
Lidar	No	Intercept	-0.37	0.12	0.13
		CHM	-0.19		
		Minimum	0.68		
		Standard deviation	1.71		
	Yes	Intercept	-0.29	0.08	0.08
		CHM	0.18		
NAIP & Lidar	No	Intercept	1.39	0.14	0.15
		Red	-0.02		
		CHM	-0.21		
		Minimum	0.65		
		Standard deviation	1.61		
	Yes	Intercept	1.64	0.06	0.07
		Red	-0.02		

The stepwise regression method shows the results with and without logarithmic transformation. The results without logarithmic transformation poorly establish the mesquite tree aboveground biomass estimation equations when the variables came from only the NAIP data groups (Table 2). When the logarithmic transformation is applied to the NAIP data, the R^2 increases but not significantly. By contrast, the lidar and combined NAIP and lidar data groups present better R^2 without the logarithmic transformation than with the logarithm transformation. However, regardless of the logarithm transformation, the results of all data appear to show poor R^2 values. Overall, the best performance is the combined NAIP and Lidar data without logarithm transformation in the stepwise regression method.

Table 3. The adjusted and multiple coefficients of determination for LASSO in mesquite tree aboveground biomass estimation

Variable group	Logarithm transformation	Selected variable	Coefficient	Adj. R^2	R^2
NAIP	No	Intercept	1.98	0.00	0.00
	Yes	Intercept	1.49	0.05	0.06
		Red	-0.01		
		Green	-0.00		
Lidar	No	Intercept	1.98	0.00	0.00
	Yes	Intercept	-0.32	0.17	0.18
		Maximum	0.14		
		Standard deviation	0.09		
NAIP & Lidar	No	Intercept	1.98	0.00	0.00
	Yes	Intercept	-0.47	0.25	0.27
		Red	-0.01		
		Blue	0.01		
		Maximum	0.20		
		Standard deviation	0.21		
		Variance	0.00		
		Kurtosis	-0.01		

The results of the LASSO method show a different perspective from the stepwise regression method (Table 3). The logarithm transformation is necessary because configurations of prediction variable data apparently appeared to display no relationship between the field measurement data and remotely sensed data. After applying the logarithm transformation, the weak correlations are found between the field measurement data and remotely sensed data. The combined NAIP and lidar data has the highest R^2 while the logarithm transformation was applying.

Table 4. The pseudo coefficient of determination for random forests in mesquite tree aboveground biomass estimation

Variable group	Logarithm transformation	Pseudo R^2
NAIP	No	0.06
	Yes	0.09
Lidar	No	0.04
	Yes	0.16
NAIP & Lidar	No	0.11
	Yes	0.37

Random forests, the last statistical method, was also implemented with and without logarithm transformation and displayed the pseudo coefficient of determination (pseudo R^2) (Table 4). The logarithmic transformation slightly improved the result of NAIP data and significantly increasing the pseudo R^2values in lidar metric and combined NAIP and lidar group. Particularly, the combined NAIP and lidar data group with logarithmic transformation had the best performance (pseudo-R^2 = 0.37). Therefore, after comparing to the rest of our results, the combined NAIP and lidar data group with logarithmic transformation of random forests methods is considered the best mesquite tree aboveground biomass model for building the mesquite tree aboveground biomass map.

The random forests method does not show the intercepts and coefficients like stepwise and LASSO regression. So, the variable importance (Figure 3) exhibits the importance of variables for building a random forests model to estimate mesquite tree aboveground biomass. Moreover, the pseudo R^2 (Table 4) has shown that the combined NAIP and lidar data with logarithmic transformation performed the best result, so we exhibit the figure of the variable importance of the combined NAIP and lidar data with logarithmic transformation (Figure 3). The figure shows that the NDVI had the highest percentage of increased MSE (Figure 3 left) and generated the most numbers of decision trees (Figure 3 right). Furthermore, the variable importance found that the variables involving NAIP data are more important than the lidar metrics. However, the lidar metrics still have the maximum height and 99^{th} percentile height rank in the 5^{th} and 6^{th} importance.

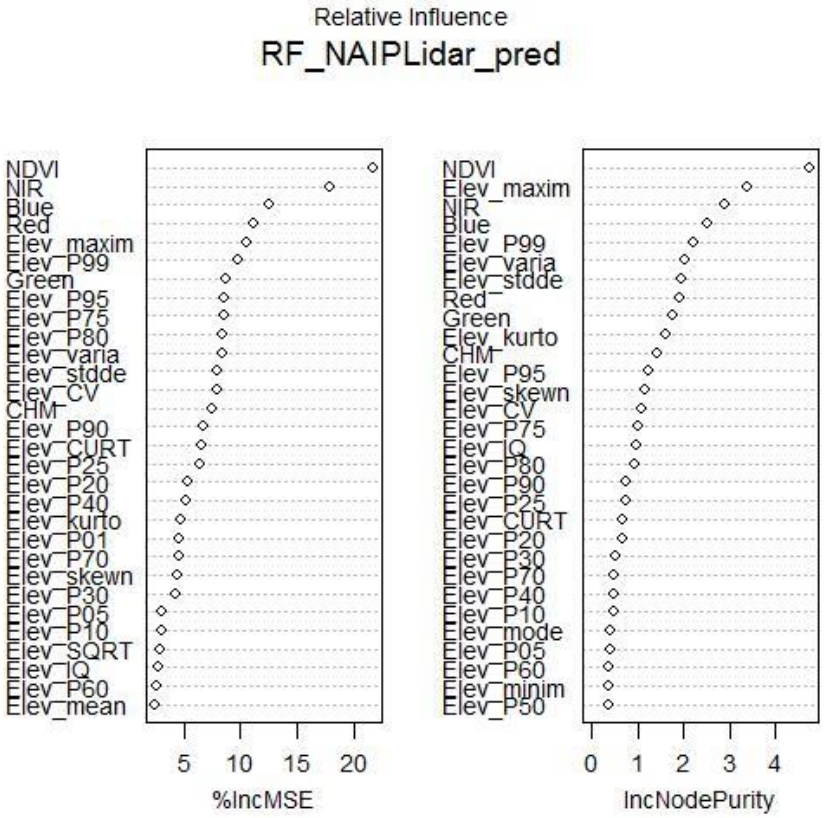

Figure 3. The variable importance of the NAIP and lidar combined data with logarithm transformation.

II.3.2. The mesquite tree aboveground biomass map and validation

The equations and models of mesquite tree aboveground biomass estimation are applied to build the mesquite tree aboveground biomass maps. The results of the mesquite tree aboveground biomass map validation are assigned to with and without logarithmic transformation (Table 5). When the mesquite tree aboveground biomass maps are calculated by stepwise regression and LASSO methods without logarithmic transformation, the MSEs are smaller than those with logarithmic transformation. However, while applying the logarithmic transformation on the random forests method, the MSEs are significantly smaller than the Random Forest results of validation without

logarithmic transformation. Furthermore, the smallest validation of the whole research is performed by the Random Forest method with the combined NAIP and lidar data with logarithmic transformation. Ultimately, the final mesquite tree aboveground biomass map is displayed in Figure 4. The figure merely showed and zoomed into a portion of the mesquite tree aboveground biomass map, because the size of the pixel (1 m x 1m) is too small to exhibit the full distribution of mesquite tree areas in the map.

Table 5. The validation of mesquite tree aboveground biomass estimation

Variable group	Logarithm transformation	Stepwise MSE (Mg/ha)	LASSO MSE (Mg/ha)	Random forests MSE (Mg/ha)
NAIP	No	1.47	1.59	3.22
Lidar		1.32	1.59	3.26
NAIP & Lidar		1.28	1.59	3.03
NAIP	Yes	1.70	1.82	1.12
Lidar		1.62	1.69	1.11
NAIP & Lidar		1.69	1.59	1.08

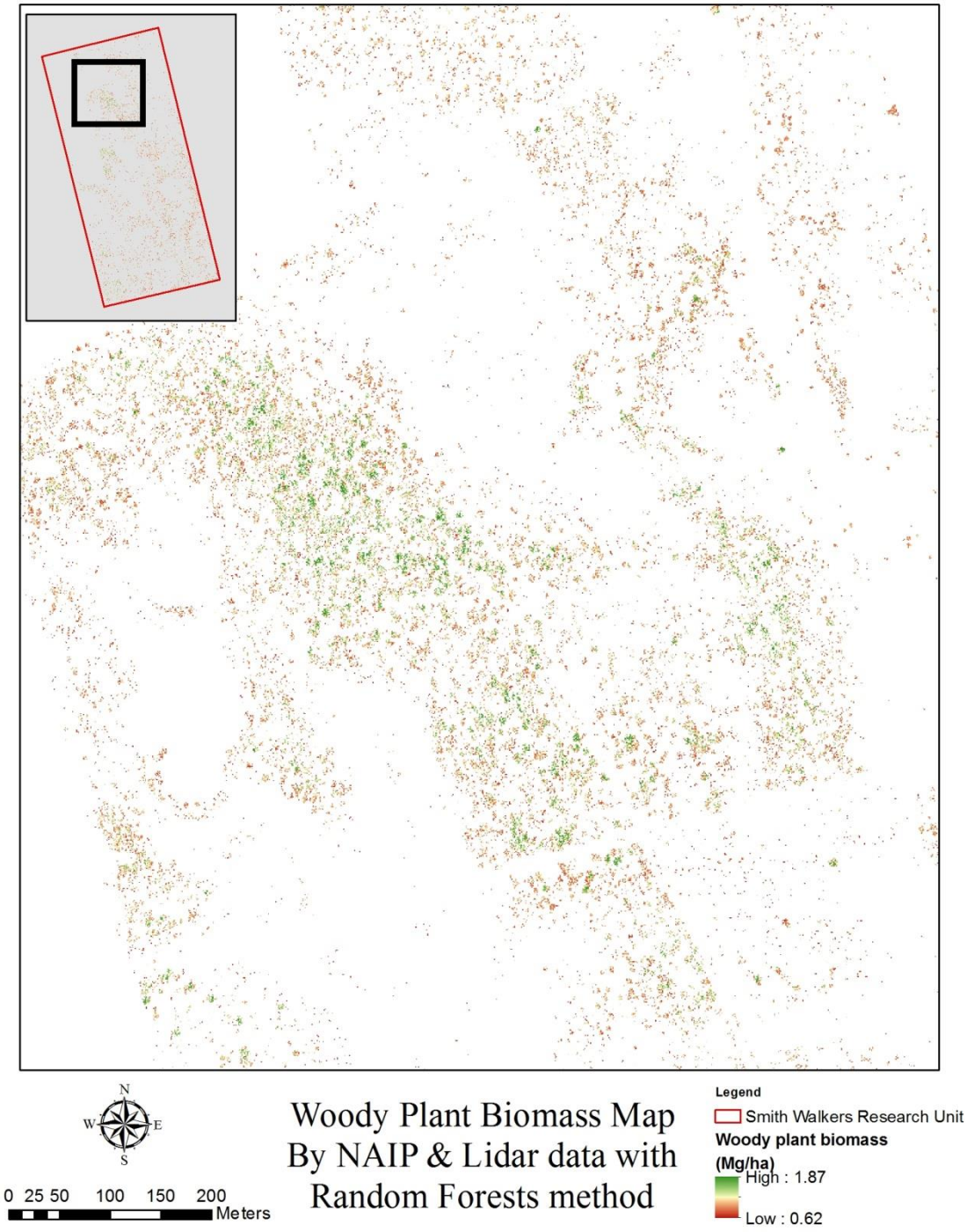

Figure 4. A portion of mesquite tree aboveground biomass map of Smith Walker Research Unit. The black box in the small map shows the location of the biomass map. The green color represents high mesquite tree aboveground biomass, and the red color represents low mesquite tree aboveground biomass at a pixel.

II.4. Discussions

The results show that the mesquite tree aboveground biomass map with random forests method, the combined NAIP and lidar data and the logarithmic transformation generates the most accurate mesquite tree aboveground biomass map (MSE is 1.08 Mg/ha, and R^2 is 0.37) in this research study. Riegel et al. (2013) estimated vegetation aboveground biomass in a mixed forested wetland with ordinary least squares multiple linear regression models and obtained similar adj-R^2 (0.37) with combined both NAIP and lidar data. They mentioned that the limit of lidar footprint samples contributes to the low R^2 in their study. In our research, the R^2 value also shows that the relationship between the remotely sensed data and field measurement data is not strong. The remotely sensed data was anticipated to have a proper correlation with the field mesquite tree aboveground biomass data. Nonetheless, the uncertainties of remotely sensed data (such as the point density and the position of Smith Walk Research Unit in the NAIP images) and the procedure of removing lidar point cloud data below 1 m that dramatically reduced the points at each pixel. Thus, these uncertainties constrain the statistical methods to build a good relationship between the remotely sensed data and the field measurement data and resulted in the low R^2 values. Therefore, we considered the small MSE value to indicate the final mesquite tree aboveground biomass map is reasonable.

One interesting finding is that the combined NAIP and lidar data in each statistical method produced better R^2 value than the other two variable groups (the NAIP variables and lidar metrics, separately). Particularly, the MSE of combined NAIP and

lidar data displays the best result in the random forests method. The lidar metrics show better mesquite tree aboveground biomass estimation than the NAIP variables, but the low point density at every pixel constrains the abilities of lidar remote sensing to offer sufficient points for depicting the structure of mesquite trees. Surprisingly, when combining the NAIP variables and lidar metrics, the R^2 values of equations and models are significantly improved regardless of the logarithmic transformation. The combined NAIP and lidar data provide adequate active and passive remotely sensed information to aid those statistical methods for better developing the equations and models for the mesquite tree aboveground biomass estimation. Thus, we believe the combined NAIP and lidar data provide adequate variables for mesquite tree aboveground biomass estimation.

The results of variable selection from these three statistical methods showed that lidar metrics were favored by the stepwise regression and LASSO methods. However, the random forests method indicated NAIP data was more important than lidar metrics. The stepwise regression selected the red band, CHM, minimum height, and standard deviation without the logarithmic transformation in the combined NAIP and lidar data group. The LASSO method selected the red band, blue band, maximum height, standard deviation height, variance height, and kurtosis height with the logarithmic transformation in the combined NAIP and lidar data group. The CHM and maximum height were critical factors for estimating tree aboveground biomass. The minimum height, standard deviation height, and variance height, and kurtosis were considered as the factors to describe the distribution of tree heights in a group of mesquite trees. The

blue and green bands were selected from NAIP data as well as combined NAIP and lidar data groups in the LASSO method. Nonetheless, both blue and green bands had very small coefficients (-0.003 and 0.008). Moreover, the variance height had a very small coefficient (0.004) from combined NAIP and lidar data group in the LASSO method as well. On the other hand, the random forests method considered NDVI, NIR band, blue band, red band, and maximum height were the top five important variables in the permutation. Furthermore, NDVI, maximum height, NIR band, Blue band, and 99^{th} percentile height were the top five important variables in the random forests model. The results of variable selection of these three statistical methods revealed that stepwise regression and LASSO prefer lidar metrics than NAIP data, but the random forests preferred NAIP data more.

The stepwise regression and LASSO methods both recognized the red band an important variable while in NAIP data and combined NAIP and lidar data groups. In fact, it has been proven that the red edge portion of the electromagnetic spectrum (>700 nm) can detect the canopy chlorophyll content and leaf area index (LAI). Additionally, several studies mentioned that the red band has a significant relationship with vegetation. Todd et al. (1998) found that the red band is sensitive to biomass variations for green vegetation. Zhang et al. (2014) used the red band to differentiate vegetated and non-vegetated pixels according to the red spectral reflectance. Although the red band of the NAIP images is between 604 to 664 nm which is a shorter wavelength compared to the red edge portion, the NAIP images provide enough energy of red spectrum for detecting the mesquite trees. Therefore, the stepwise regression and LASSO methods found a

significant relationship between the red band and the mesquite tree aboveground biomass.

Our research utilized a pixel-level processing concept to process the remotely sensed data at each pixel. Riegel et al. (2013) described a similar small plot concept to collect the vegetation measurements from 10 x 10 m plots and found the relationship between the plot-level vegetation aboveground biomass and remotely sensed data, including lidar and NAIP imageries. Similarly, we conducted all NAIP and lidar training and testing data to derive and calculate biomass at 1 m spatial resolution which is the spatial resolution of the NAIP imagery. Each pixel has its field measurement and remotely sensed data. Then we selected the samples by each pixel to build the mesquite tree aboveground biomass estimation equations and models. Although the very high spatial resolution causes fewer points at each pixel, the combined NAIP and lidar data still provides adequate information for developing mesquite tree aboveground biomass equations and models at such a high spatial resolution. Given this very high spatial specificity, as opposed to commonly plot-level aboveground biomass assessment, our results for scaling up to local estimations are encouraged with the similar R^2 and low MSE values.

II.5. Conclusion

Our study found the best fitting method and remotely sensed data combination for the mesquite tree aboveground biomass estimation on rangelands after comparing the results of combined statistical methods and remotely sensed data. Ultimately, the random forests method with combined NAIP and lidar data provided the best result for

estimating the mesquite tree aboveground biomass at high spatial resolution. Not only did the approach select appropriate remotely sensed variables, but also built an accurate mesquite tree aboveground biomass model. Moreover, our study upscaled the mesquite tree aboveground biomass from pixel- and plot-level to create a local-scale mesquite tree aboveground biomass map. The procedure to build the biomass map offered a novel spatial explicit approach for the future biomass map development.

Future research could use point cloud data with higher point density per square meters, whether lidar- or photogrammetric-derived point cloud data, and multispectral imagery to estimate vegetation aboveground biomass from local to regional level. For local-scale, future investigations could make use of structure-from-motion (SfM) point cloud data acquired affordably through the use of unmanned aerial systems (UAS). Low altitude and cost UAS is one convenient technique for landowners to investigate their properties in rangelands and forests.

CHAPTER III

GLOBAL FOREST CANOPY HEIGHT MAP VALIDATION AND CALIBRATION FOR THE POTENTIAL OF FOREST BIOMASS ESTIMATION IN THE SOUTHERN UNITED STATES

III.1. Introduction

The canopy height model (CHM) is the product of lidar remote sensing used to characterize vegetation heights above ground level for forest measurement applications or ecosystem studies (Popescu et al., 2002, Mustonen et al., 2008). Particularly, the CHM can provide individual tree measurements and biophysical variable information for forest aboveground biomass estimation at a plot or local scale. Thus, the CHM is also a critical remotely sensed variable to assess the carbon storage in forests. However, the majority of the CHMs covers merely local scale areas and is difficult to acquire for larger scale (i.e., national or continental scale) coverage. Therefore, to retrieve canopy information for large extents and to estimate forest aboveground biomass, there is a clear need for the large-scale forest canopy height map.

The global forest canopy height map (GCHM) is one of the global products that could be used for estimating forest aboveground biomass at large-scale. Recently, three global forest canopy height maps have been generated based on the Geoscience Laser Altimeter System (GLAS) and other data sources and methodological approaches. First, Lefsky (2010) generated a continuous global canopy height map by combining GLAS data from the sole instrument on ICESat (Ice, Cloud, and land Elevation Satellite) and

MODIS images with the Cubist data mining method. The global canopy height map is the first to reveal the distribution of forest canopy height globally. Saatchi et al. (2011) investigated the tropical forest carbon stock across three continents based on Lefsky's GCHM. Then Simard et al. (2011b) integrated SRTM, GLAS, climatic and other ancillary data with the random forests method to create the global forest canopy height map. The second global forest canopy height map displayed even more massive forest canopy coverage than the first one (Bolton et al., 2013). Moreover, Los et al. (2012) processed GLAS, SRTM, and MODIS data to create a coarse resolution global vegetation height and vegetation cover fraction product data sets as well. Bevan et al. (2014) applied Los's global vegetation height map to investigate the response of vegetation after the 2003 European drought. Those global vegetation height maps all considered GLAS data as the primary remotely sensed source because of the capabilities for directly measured forest canopy height. Furthermore, the 60 m in diameter footprint on the ground offered essential forest biophysical and structure information at a large scale area.

Nevertheless, Simard et al. (2011b) claimed their GCHM covered a larger area of forests because of involving mosaic crops, open forest, and saline flooded forests. So, Simard's GCHM displays not only the tall trees in forests but also the short woody plants on the ground. Besides, Bolton et al. (2013) have initially investigated the agreement of Lefsky and Simard's GCHM over Canada with airborne lidar data. Their investigation found that Simard's GCHM has better agreement with airborne lidar data because of the removal of the slope-affected GLAS waveforms. Additionally, Los's

GCHM has a very coarse spatial resolution (0.5° x 0.5°) and makes airborne lidar data difficult to use for validation and calibration. Therefore, we considered Simard's GCHM is the appropriate map for the validation and calibration with airborne lidar remote sensing data on rangelands and forests in the southern United States.

Simard et al. (2011b) conducted the field validations for both GLAS estimates of canopy height footprints and the resultant 1 km^2 pixel from their GCHM. Specifically, the 66 field measurements were collected from the FLUXNET La Thuile database (Baldocchi, 2008, Baldocchi et al., 2001) globally. The field validation data included the terrestrial or airborne lidar, laser rangefinder/clinometer, and the manual tree height measurement by tape measure. Only three field validation sites of the GCHM (2 sites in Florida and 1 site in Mississippi) were located in the southern United States. In our calibration study, the Simard's GCHM was regressed against airborne lidar data. The calibrated GCHM (cGCHM) is expected to improve the accuracy of mapping forest aboveground biomass at large scale in future studies, such as for estimating canopy biophysical parameters, including biomass.

This research offers more field measurements than Simard's validations of the southern United States where Simard lacked the field measurements. Moreover, we calibrated the GCHM of the southern United States in order to use the calibrated GCHM for improving the future large-scale forest aboveground biomass estimation in Chapter IV. Thus, this research is establishing a methodology for conducting validation and calibration of Simard's GCHM, or any other potential global map of vegetation heights, for improving potential future applications for large-scale mapping of forest biophysical

parameters. The specific objectives of the proposed study are to 1) validate Simard's global forest canopy height map with airborne lidar metrics on rangelands and forests, and 2) calibrate the global forest canopy height map with the appropriate lidar metric for future forest aboveground biomass estimation.

III.2. Materials and Methods

III.2.1. Study area

The southern United States which is defined by Census Regions and Divisions of the United States of the United States Census Bureau (2018) is the interest of the research. The southern United States includes 16 states and 1 federal district which are Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia, and District of Columbia. Most of the states belong to the humid subtropical climate, and some regions fall in the tropical (southern Florida), and arid (western Texas and western Oklahoma) climate according to Köppen climate classification. The average temperature of the southern United States ranges from 21 °C to 32 °C in summer and 40 °C to 16 °C in winter. Furthermore, the average annual precipitations are between 500 mm to 2000 mm across this regional.

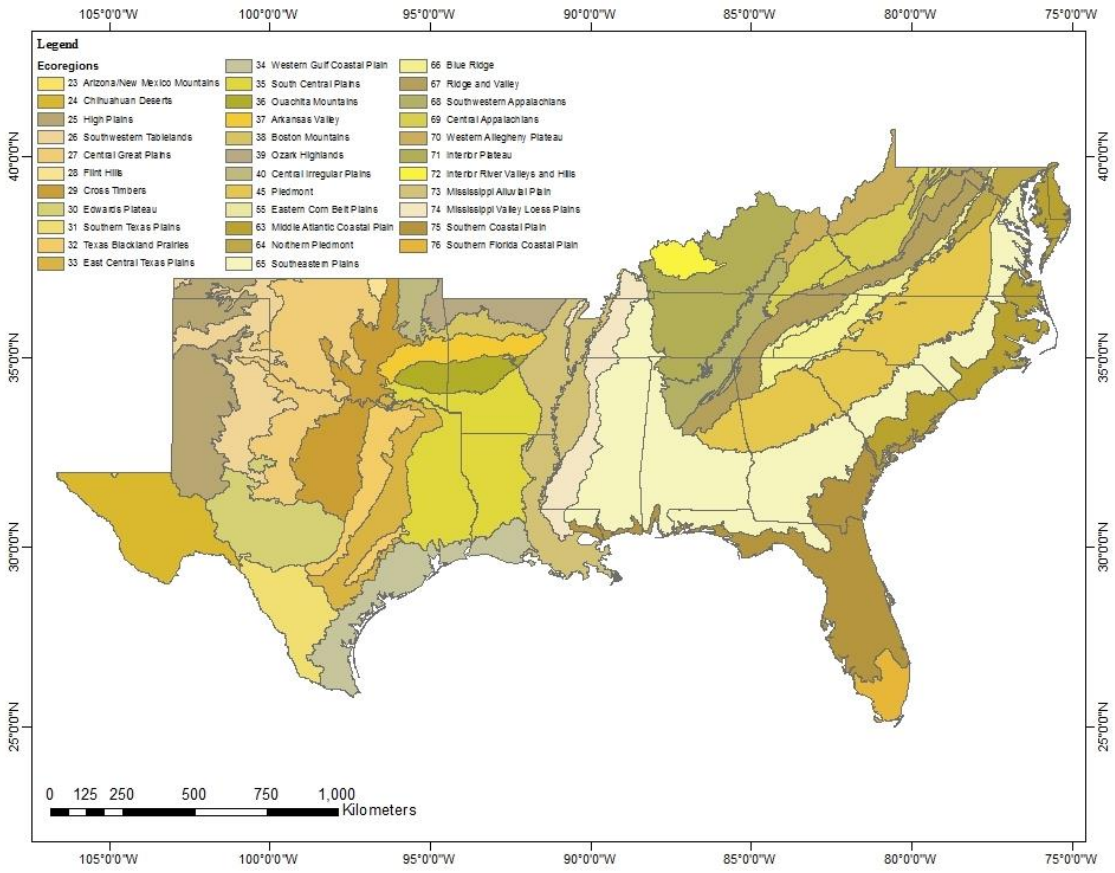

Figure 5. The EPA level III ecoregions in the southern United States (Ecoregions, 2018)

Environment Protection Agency (EPA) had published four different hierarchical levels of ecoregions (Omernik and Griffith, 2014). The range was from level I with 12 ecoregions to level IV with 967 ecoregions in the conterminous U. S. (Ecoregions, 2018). The ecoregions were defined based on similarity in the mosaic of biotic, abiotic, terrestrial, and aquatic ecosystem components. Hence, the appropriate ecoregion levels were chosen for validating the GCHM of the southern United States. Level I (3 ecoregions) and level II (4 ecoregions) ecoregions were eliminated because the ecoregions did not appear to have large extents or significant vegetation canopy (arid,

semi-arid, subtropical, and tropical) in the southern United States. Moreover, the level IV ecoregion had too many small area components when considering the spatial scale of the southern United States. Thus, the level III ecoregions (Table 6) were selected within the southern United States (Figure 5). However, the airborne lidar point cloud data was not available in every EPA level III ecoregion, but included 9 of these ecoregions to validate and calibrate the CGHM with airborne lidar point cloud data. These 9 ecoregions included 1575 GCHM samples.

Table 6. The number of GCHM samples with airborne lidar point cloud data in each ecoregion

Name*	Abbreviation	Number of GCHM samples
Unclassified		2
27 Central Great Plains	CGP	33
30 Edwards Plateau	EP	9
63 Middle Atlantic Coastal Plain	MACP	186
73 Mississippi Alluvial Plain	MAP	189
74 Mississippi Valley Loess Plains	MVLP	240
35 South Central Plains	SCEP	239
65 Southeastern Plains	SEP	161
75 Southern Coastal Plain	SCOP	454
76 Southern Florida Coastal Plain	SFCP	62

*The name of EPA level III ecoregion in the southern United States

III.2.2. The global forest canopy height map

The global forest canopy height map was derived by Simard et al. (2011b) (Figure 6) and was selected for our validation and calibration because of the proper spatial resolution, the large forest area that it covers, and its public availability. The global forest canopy height map can be downloaded from the NASA Jet Propulsion Laboratory website (Simard et al., 2011a). The primary data used to generate the global forest canopy height map was the GLAS-derived estimate of canopy height, RH100, which was the distance between the beginning of laser pulse echo and the corresponding

location of the ground peak (Harding and Carabajal, 2005, Sun et al., 2008, Boudreau et al., 2008). The GLAS data was GLA 14 land product version of 2005 data. Additionally, the GCHM generation integrated climate data, i.e. the annual mean temperature and temperature seasonality data as well as annual mean precipitation and precipitation seasonality of Tropical Rainfall Measurement Mission (TRMM) from the Worldclim dataset, the SRTM elevation, MODIS tree cover data (MOD44 map), and protection status of UN World Database on Protected Areas. Then the primary data and the ancillary data were input to the random forests statistical method to create the global forest canopy height map.

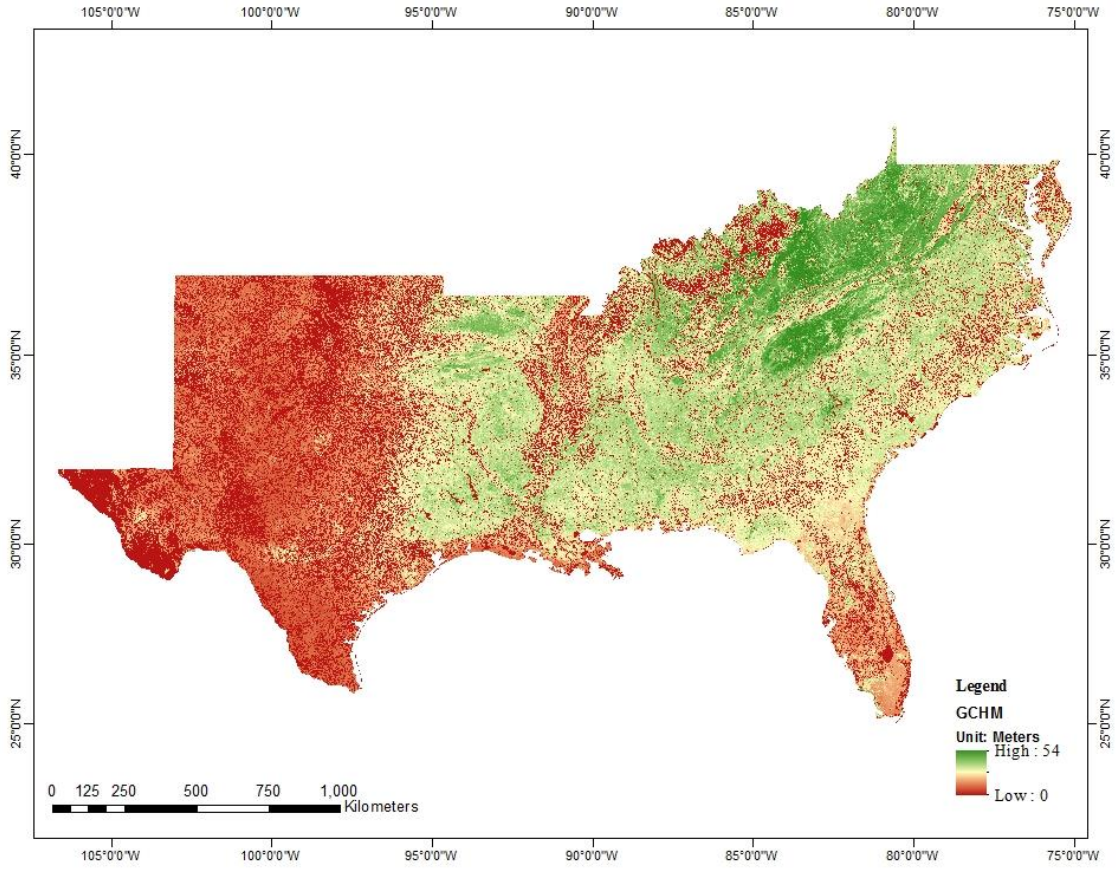

Figure 6. The global forest canopy height map of the southern United States made by Simard et al. (2011b)

III.2.3. Airborne lidar point cloud data

The airborne lidar point cloud data was collected between 2010 and 2012 and acquired from (1) NASA's Goddard's LiDAR, Hyperspectral & Thermal Image (G-LiHT) program; (2) National Ecological Observatory Network's (NEON) prototype data sharing program; (3) NSF OpenTopography program; and (4) Lidar application for the study of ecosystem with remote sensing laboratory (LASERS) in the Department of Ecosystem Science and Management at Texas A&M University. The airborne lidar point

cloud data were spread out in the southern United States, including the states of Florida, Maryland, North Carolina, Texas, and Virginia (Table 7). Also, to provide accurate samples of airborne lidar point data to validate and calibrate the global wall-to-wall forest canopy height map, the airborne lidar point cloud data had completed coverage of their scanned areas as well (Table 6).

Table 7. A list data source of airborne lidar point cloud data

Data Source	Location	States	Month	Year
TAMU LASERS Lab	Vernon	Texas	November	2010
	Big Sandy Creek	Texas	November	2010
	Huntsville	Texas	November	2010
NEON	Ordway-Swisher Biological Station	Florida	September	2010
	Donaldson plantation	Florida	September	2010
NSF OpenTopography	Tuscaloosa	Alabama	December	2010
	The Meeman-Shelby lineament	Arkansas-Tennessee	July to September	2010
	The Reelfoot scarp	Arkansas-Tennessee	July to September	2010
	Apopka	Florida	June	2011
	Bald Point	Florida	September	2010
	Merritt Island	Florida	June	2008
	South Florida Everglades	Florida	November	2012
	Charleston	South Carolina	February	2010
	Canyon Lake Gorge	Texas	October	2010
NASA G-LiHT	Bowie	Maryland	June	2012
	Parker Track	Virginia	July	2011
	Patuxent Refuge	Maryland	June	2012
	Perquimans	North Carolina	July	2011
	Smithsonian Environmental Research Center	Maryland	October	2011
	Wallops Flight Facility	Virginia	September	2011

III.2.4. Airborne lidar point cloud data preprocessing

Most of the airborne lidar point cloud data were raw data with unnecessary information, redundant points, and inaccurate points. Thus, the airborne lidar point cloud data were preprocessed before acquiring airborne lidar metrics. The airborne lidar point cloud data was subtracted from the 3 m^2 spatial resolution digital elevation model (DEM) to remove the terrain height (Hodgson, 2003). Then the urban, cropland and water areas were visually identified and removed from the airborne lidar point cloud data. The distribution of the airborne point cloud data aided to distinguish urban areas, water body, croplands, and grasslands. The remaining airborne lidar point cloud data was assumed to cover woody plants and forests. After removing the significant non-woody plant airborne lidar point cloud data, the inaccurate and redundant points below the ground or higher than the tallest woody plants were removed. In addition, power and communication towers in the forest areas were eliminated as well. The inaccurate, as well as power and communication towers points were removed manually. Ultimately, the airborne lidar point cloud data was processed to 4 different categories based on the height of different vegetation for investigating the potential influence of small vegetation point cloud data in forests, such as understory woody plants, shrubs, bushes, and grass. The 4 height categories of point cloud data were 1) point cloud data greater than 0 m, 2) greater than 1 m to eliminate open grounds and short grass in forests, 3) greater than 3m to trim the tall grass and small bushes off, and 4) greater than 5m to cut the shrubs and bushes off in forests.

III.2.5. Derived airborne lidar metrics at the pixel-based global forest canopy height map

The outlines of every airborne lidar covered areas were co-registered with the GCHM. Those GCHM pixels which had more than 50% of areas covered by the airborne lidar data were selected. Because the spatial resolution of GCHM was about 1 km^2, every selected pixel included sufficient amount of airborne lidar point cloud data for computing the airborne lidar metrics. The airborne lidar metrics were maximum, mean, mode, 50^{th} percentile, 75^{th} percentile, 90^{th} percentile, 95^{th} percentile, and 99^{th} percentile heights (Table 8) and computing in every height category (0, 1, 3, and 5 m). However, due to the short vegetation and woody plants being removed at specific height categories, some pixels would encounter no airborne points. Therefore, these pixels would be omitted from further analysis.

Table 8. The list of airborne lidar metrics from airborne lidar point cloud data

1. Maximum Height
2. Mean Height
3. 50^{th} percentile Height
4. 75^{th} percentile Height
5. 90^{th} percentile Height
6. 95^{th} percentile Height
7. 99^{th} percentile Height
8. Mode Height

III.2.6. Methods of validation and calibration

The validation and calibration of GCHM were processed to the ecoregions with the available airborne lidar point cloud data. The ecoregions without the available airborne lidar point cloud data would not be validated and calibrated. Initially, the procedure of validation was applied to the ecoregion with the available airborne lidar

metrics and the entire southern United States GCHM in different height categories. The airborne lidar metrics was split into 80% training and 20% testing datasets. The training data was employed to validate the GCHM, and the testing data was employed to validate the calibrated GCHM (cGCHM). The bias (Equation 4) and root mean square error (RMSE) (Equation 5) were used to validate the GCHM.

Equation 4

$$\boldsymbol{Bias_j = \frac{\sum_{i=1}^{n}\left(GCHM_{ij} - Metric_{ij}\right)}{n}}$$

Equation 5

$$\boldsymbol{RMSE_j = \sqrt{\frac{\sum_{i=1}^{n}\left(GCHM_{ij} - Metric_{ij}\right)^2}{n}}}$$

Where $GCHM_{ij}$ denotes the pixel *i* of GCHM in the *j* ecoregion, $Metric_{ij}$ denotes the airborne lidar metrics of the correspond pixel *i* of the GCHM in the *j* ecoregion, $Bias_j$ denotes the average of total bias in the *j* ecoregion, and $RMSE_j$ denotes the RMSE in the *j* ecoregion.

After the validation of GCHM, the appropriate lidar metric would be selected to calibrate the GCHM. The appropriate lidar metric selection was not only determined by the smallest RMSE and bias, but also referred to several previous studies (McGaughey et al., 2010, Drake et al., 2002, Bolton et al., 2013, Means et al., 2000, St-Onge et al., 2008, Alexander et al., 2014). The calibration regressed the selected lidar metric against the GCHM to acquire the calibration equation of every ecoregion and entire GCHM (Equation 6).

Equation 6

$$\boldsymbol{Metric_{ij} = \alpha + \beta \cdot GCHM_{ij}}$$

Where $GCHM_{ij}$ denotes the pixel i of GCHM in the j ecoregion, $Metric_{ij}$ denotes the airborne lidar metrics of the correspond pixel i of the GCHM in the j ecoregion, α denotes the intercept and β denotes the coefficient.

The coefficient of determination (R^2) was used for examining the discrepancies of canopy height between the airborne lidar metrics and GCHM. Sequentially, the calibration equations were employed to calibrate each ecoregion and entire GCHM. The testing data then validated the cGCHM for the accuracy and displayed the validation results with RMSE and bias as well. All validation and calibration procedures were implemented in the R programming language. ArcGIS was utilized with the calibration equations to generate cGCHMs. The flowchart (**Error! Not a valid bookmark self-reference.**) shows the general concept of GCHM validation and calibration at one of the ecoregions.

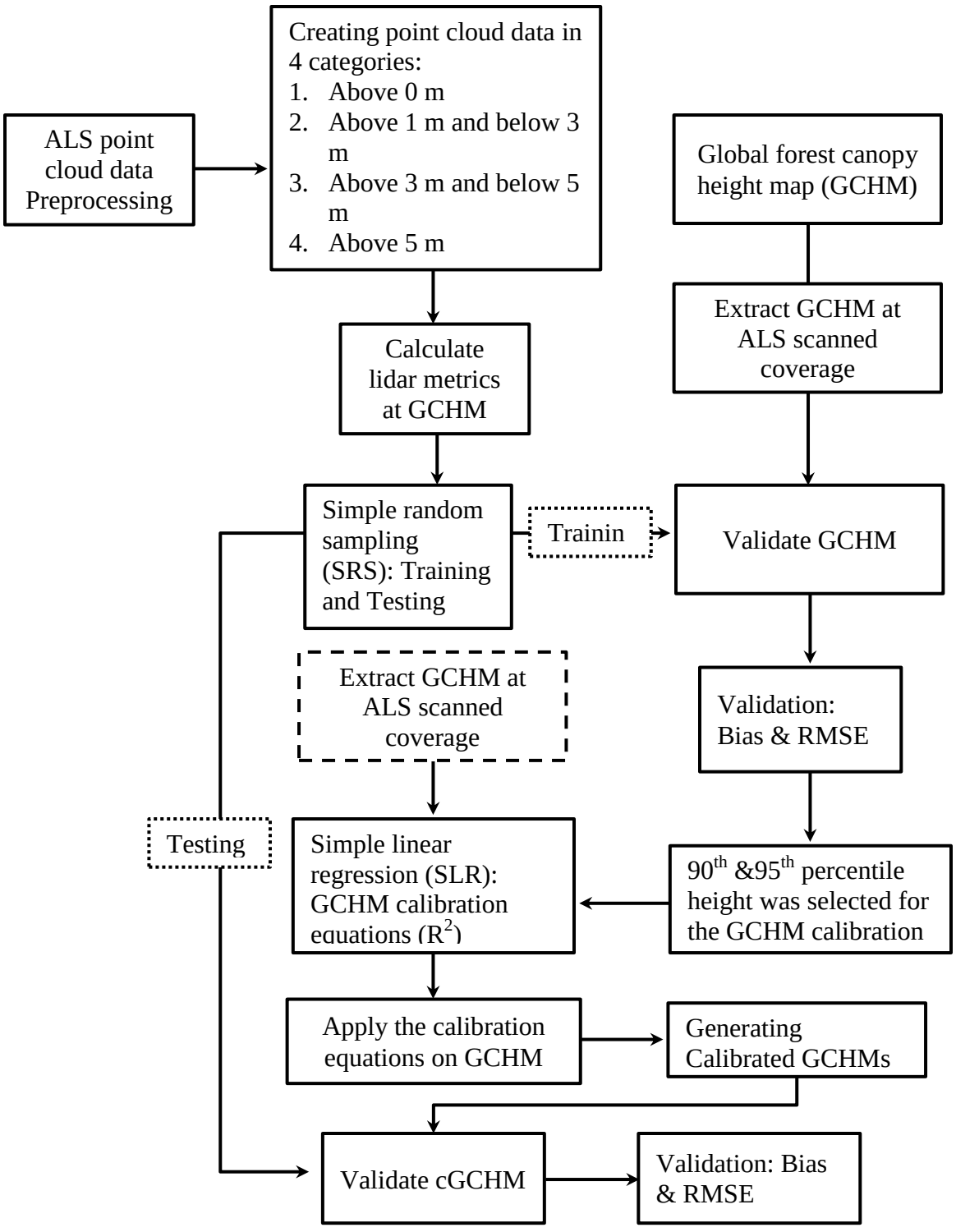

Figure 7. The flowchart of the general concept for validating and calibrating GCHM

III.3. Results

III.3.1. The results of the GCHM validation

The results of the GCHM validation display the bias (Figure 8) and RMSE (Figure 9) between the GCHM and airborne lidar metrics (the max, mean, mode, 50^{th}, 75^{th}, 90^{th}, 95^{th}, and 99^{th} percentile heights). The bias and RMSE are computed in each ecoregion at 4 height categories. After the comparison of bias and RMSE by the airborne lidar metrics in the ecoregion, the nearest zero bias and smallest RMSE in each ecoregion is highlighted with bold and underlined. In Table 9, MACP, MAP, MVLP, SCEP, SEP, SCOP, and Total Points have the nearest zero bias at the 90^{th} percentile height at 0 m height category. CGP has the nearest zero bias at the 99^{th} percentile height, EP has the nearest zero bias at mean height, and SFCP has the nearest zero bias at max height at the 0 m height category. In summary, the majority ecoregions (7 of 10 ecoregions) have the nearest zero bias at the 90^{th} percentile height at 0 m height category (Figure 8A).

At the 1 m height category, MACP, MVLP, SEP, SCOP, and Total Points have the nearest zero bias at the 75^{th} percentile height. CGP has the nearest zero bias at the 95 percentile height. EP has the nearest zero bias at the mode height. MAP has the nearest zero bias at the 50^{th} percentile height. SCEP has the nearest zero height at the 90^{th} percentile height. SFCP has the nearest zero bias at the max height. Thus, the majority ecoregions (5 of 10 ecoregions) have the nearest zero bias at the 75^{th} percentile height at the 1 m height category (Figure 8B).

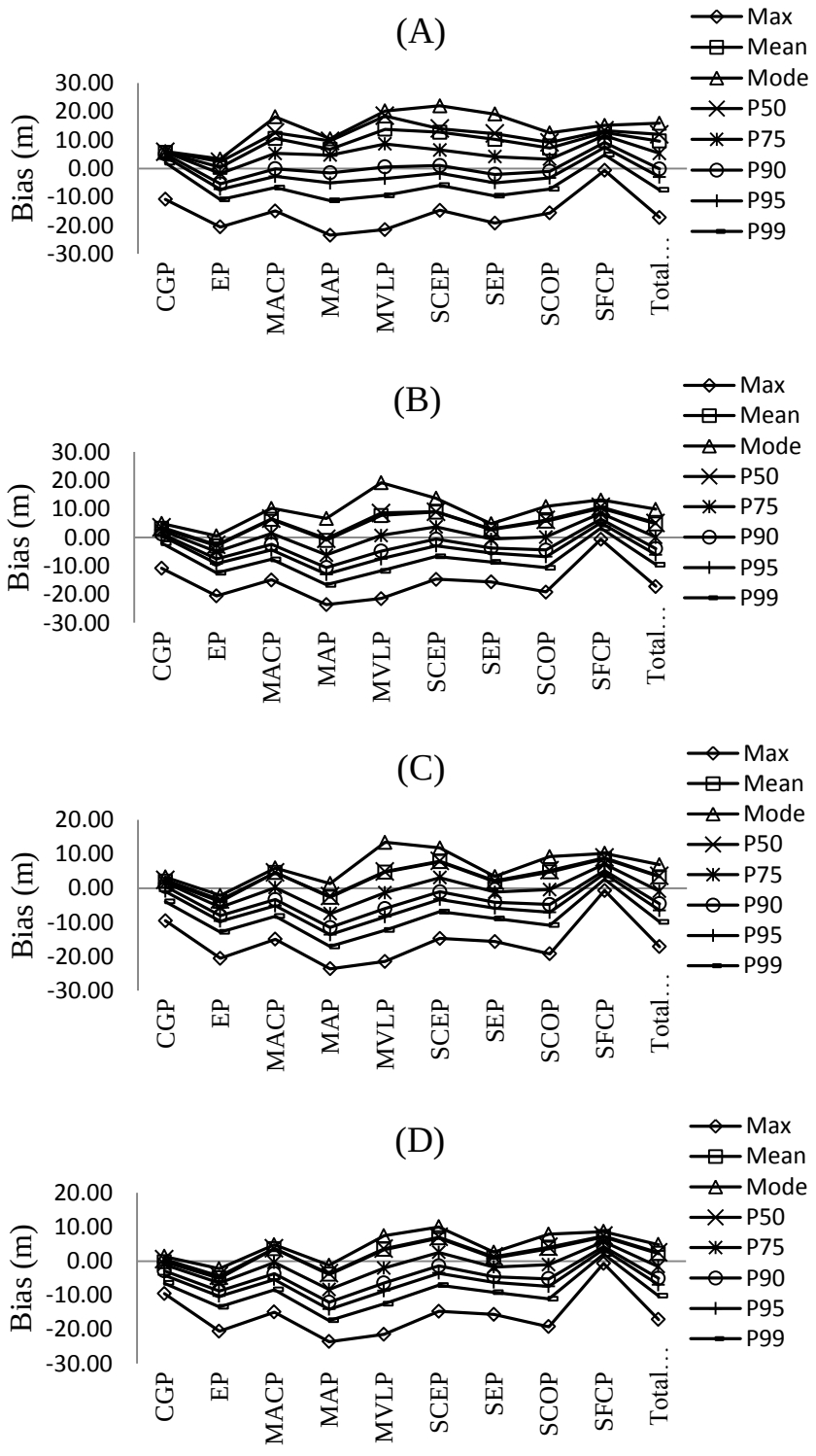

Figure 8. The bias results of GCHM validation: (A) 0 m, (B) 1 m, (C) 3 m, and (D) 5 m height category. Total… represents Total Points.

At the 3 m height category, MACP, MVLP, SEP, SCOP, and Total Points have the nearest zero bias at the 75th percentile height. CGP and SCEP have the nearest zero bias at the 90th percentile height. EP and MAP have the nearest zero bias at the mode height. SFCP has the nearest zero bias at the max height. Totally, the majority ecoregions (5 of 10 ecoregions) have the nearest zero bias at the 75th percentile height at the 3 m height category (Figure 8C).

Table 9. The bias results of validation in every ecoregion and entire GCHM (Total point)

		Bias							
Ecoregion	Height	Max	Mean	Mode	P50	P75	P90	P95	P99
CGP	0 m	-10.85	5.59	5.92	5.84	5.75	5.03	4.06	**1.98**
EP		-20.54	**0.33**	3.29	2.34	-1.74	-5.41	-7.35	-10.96
MACP		-14.92	10.50	18.10	12.53	5.28	**-0.14**	-2.78	-6.75
MAP		-23.45	6.70	10.40	9.65	4.70	**-1.57**	-5.10	-11.27
MVLP		-21.45	13.71	20.17	18.43	8.61	**0.50**	-3.51	-9.44
SCEP		-14.69	12.67	21.96	13.96	6.48	**1.02**	-1.72	-5.91
SEP		-19.19	10.19	19.06	12.21	4.13	**-2.08**	-5.07	-9.60
SCOP		-15.63	7.14	12.46	9.11	3.23	**-0.99**	-3.33	-7.20
SFCP		**-0.68**	12.63	15.10	13.31	11.52	9.09	7.38	4.80
Total Points		-17.20	9.60	15.86	11.94	5.30	**-0.17**	-3.04	-7.58
CGP	1 m	-10.85	3.17	4.89	3.52	2.44	1.26	**0.54**	-2.05
EP		-20.54	-3.01	**0.53**	-2.64	-5.06	-7.60	-9.24	-12.48
MACP		-14.92	6.30	10.16	6.50	**1.41**	-2.48	-4.46	-7.74
MAP		-23.60	-1.05	6.61	**-0.51**	-6.18	-10.63	-12.95	-16.66
MVLP		-21.45	7.70	19.17	8.54	**0.69**	-4.80	-7.50	-11.73
SCEP		-14.69	8.85	13.71	9.01	3.69	**-0.65**	-2.96	-6.69
SEP		-15.63	2.73	4.77	2.91	**-0.63**	-3.71	-5.50	-8.66
SCOP		-19.19	5.65	10.84	6.09	**0.16**	-4.42	-6.80	-10.67
SFCP		**-0.68**	10.26	13.11	10.68	8.43	6.36	5.10	3.22
Total Points		-17.20	4.70	9.85	5.07	**0.01**	-3.98	-6.08	-9.59

* Total Points combines all ecoregion airborne lidar point cloud data

Table 9. Continued.

		Bias							
Ecoregion	Height	Max	Mean	Mode	P50	P75	P90	P95	P99
CGP	3 m	-9.51	1.98	3.27	2.39	1.56	**0.47**	-0.94	-3.93
EP		-20.54	-3.93	**-2.15**	-3.37	-5.61	-8.05	-9.64	-12.85
MACP		-14.92	4.57	5.91	4.55	**0.15**	-3.34	-5.13	-8.18
MAP		-23.59	-2.71	**1.34**	-2.40	-7.49	-11.50	-13.64	-17.09
MVLP		-21.45	4.65	13.40	4.85	**-1.32**	-5.98	-8.38	-12.30
SCEP		-14.69	7.71	11.78	7.88	3.07	**-1.06**	-3.28	-6.89
SEP		-15.63	1.87	3.39	2.13	**-1.16**	-4.11	-5.82	-8.89
SCOP		-19.19	4.69	9.20	5.15	**-0.40**	-4.77	-7.07	-10.84
SFCP		**-0.68**	8.56	10.17	8.88	7.02	5.15	4.19	2.59
Total Points		-17.10	3.28	6.84	3.53	**-0.93**	-4.56	-6.54	-9.88
CGP	5 m	-9.51	**-0.22**	1.37	0.49	-1.02	-2.88	-4.11	-6.46
EP		-20.54	-5.03	**-2.25**	-4.39	-6.45	-8.76	-10.29	-13.44
MACP		-14.92	3.55	4.74	3.66	**-0.44**	-3.70	-5.40	-8.39
MAP		-23.59	-3.95	**-1.25**	-3.64	-8.32	-12.08	-14.10	-17.37
MVLP		-21.45	3.34	7.39	3.57	**-2.06**	-6.43	-8.73	-12.53
SCEP		-14.69	6.69	10.00	6.97	2.52	**-1.44**	-3.56	-7.07
SEP		-15.63	**0.91**	2.57	1.27	-1.77	-4.55	-6.21	-9.17
SCOP		-19.19	3.61	7.85	4.06	**-1.08**	-5.20	-7.41	-11.06
SFCP		**-0.68**	7.07	8.64	7.33	5.65	4.21	3.41	2.03
Total Points		-17.10	2.16	4.81	2.46	**-1.64**	-5.07	-6.97	-10.18

* Total Points combines all ecoregion airborne lidar point cloud data

At the 5 m height category, MACP, MVLP, SCOP, and Total Points have the nearest zero bias at the 75th percentile height. CGP and SEP have the nearest zero bias at the mean height. EP and MAP have the nearest zero bias at the mode height. SCEP has the nearest zero bias at the 90th percentile height. SFCP has the nearest zero bias at the max height. Finally, the majority ecoregions (4 of 10 ecoregions) have the nearest zero bias at the 75th percentile height at the 5 m height category (Figure 8D).

In the RMSE table (Table 10), MACP, MVLP, SCEP, SEP, SCOP, and Total Points have the smallest RMSE at the 90th percentile height at the 0 m height category. CGP has the smallest RMSE at the 95th percentile height. EP has the smallest RMSE at

the mean height. MAP has the smallest RMSE at the 75th percentile height. SFCP has the smallest RMSE at the max height. Therefore, the majority ecoregions (6 of 10 ecoregions) have the smallest RMSE at the 90th percentile height at the 0 m height category (Figure 9A).

At the 1 m height category, MACP, MVLP, SEP, SCOP, and Total Points have the smallest RMSE at the 75th percentile height. CGP has the smallest RMSE at the 95th percentile height. EP has the smallest RMSE at the mode height. MAP has the smallest RMSE at the mean height. SCEP has the smallest height at the 90 percentile height. SCOP has the smallest RMSE at the max height. Thus, the majority ecoregions (5 of 10 ecoregions) have the smallest RMSE at the 75th percentile height at the 1 m height category (Figure 9B).

At the 3 m height category, MACP, MVLP, SEP, SCOP, and Total Points have the smallest RMSE at the 75th percentile height. CGP and SCEP have the smallest RMSE at the 90th percentile height. EP has the smallest RMSE at the mode height. MAP has the smallest RMSE at the 50th percentile height. SFCP has the smallest RMSE at the max height. In summary, the majority ecoregions (5 of 10 ecoregions) have the smallest RMSE at the 75th percentile height at the 3 m height category (Figure 9C).

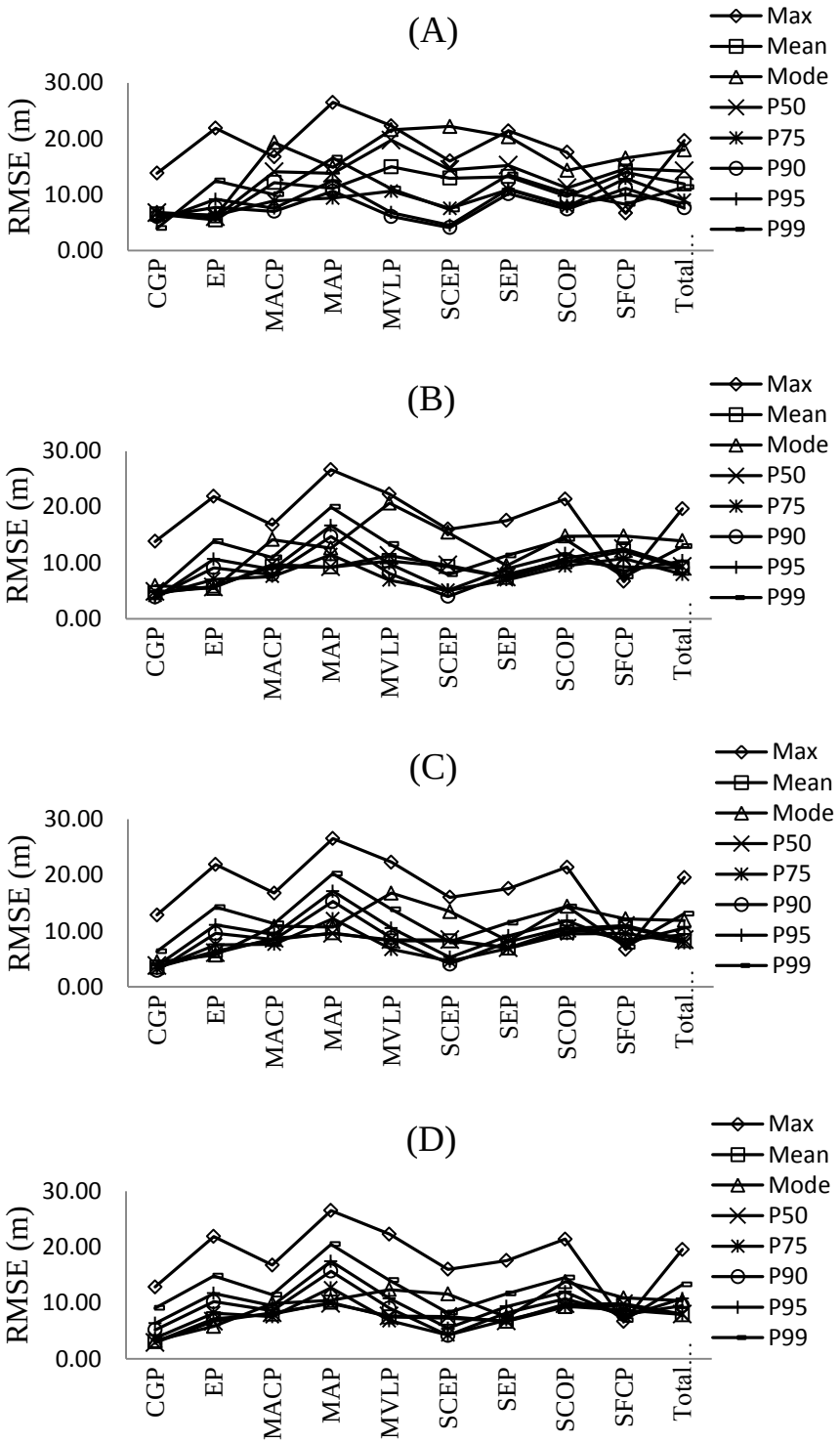

Figure 9. The RMSE results of GCHM validation: (A) 0 m, (B) 1 m, (C) 3 m, and (D) 5 m height category. Total… represents Total Points.

At the 5 m height category, SEP, SCOP, and Total Points have the smallest RMSE at the mean height. CGP and MAP have the smallest RMSE at the 50th percentile height. EP has the smallest RMSE at the mode height. MACP and MVLP have the smallest RMSE at the 75th percentile height. SCEP has the smallest RMSE at the 90th percentile height. SFCP has the smallest height at the max height. Finally, the majority ecoregions (3 of 10 ecoregions) have the smallest RMSE at the mean height (Figure 9D).

Table 10.The RMSE results of validation in every ecoregion and entire GCHM (Total point)

		RMSE							
Ecoregion	Height	Max	Mean	Mode	P50	P75	P90	P95	P99
CGP	0 m	13.84	6.51	6.82	6.74	6.66	6.06	**5.31**	3.98
EP		21.89	**5.48**	6.20	6.01	6.32	7.70	9.17	12.50
MACP		16.76	12.15	19.30	14.10	8.90	**7.01**	7.56	10.03
MAP		26.50	11.15	14.89	13.80	**9.39**	10.62	12.55	16.57
MVLP		22.32	14.99	21.62	19.73	10.62	**6.08**	6.78	11.04
SCEP		16.03	12.95	22.19	14.46	7.54	**4.11**	4.53	7.31
SEP		21.42	13.22	20.34	15.22	10.91	**10.16**	11.01	13.55
SCOP		17.59	9.66	14.34	11.22	7.86	**7.40**	8.03	10.27
SFCP		**6.66**	13.96	16.56	14.67	12.92	11.14	10.04	8.30
Total Points		19.66	11.91	18.00	14.24	9.00	**7.63**	8.47	11.34
CGP	1 m	13.84	4.61	5.93	4.82	4.16	3.89	**3.71**	4.63
EP		21.89	5.80	**5.42**	5.62	7.05	9.13	10.67	13.87
MACP		16.76	9.30	14.16	9.73	**7.58**	7.89	8.85	10.96
MAP		26.68	**9.24**	12.64	9.26	11.47	14.71	16.65	19.98
MVLP		22.32	10.38	20.69	11.28	**6.90**	8.10	9.87	13.36
SCEP		16.03	9.33	15.46	9.58	5.04	**3.99**	5.09	7.92
SEP		17.59	7.34	9.49	7.42	**6.94**	7.93	8.98	11.30
SCOP		21.42	10.13	14.78	10.82	**9.45**	10.47	11.65	14.24
SFCP		**6.66**	12.09	14.82	12.52	10.80	9.28	8.54	7.50
Total Points		19.68	8.99	13.95	9.33	**7.94**	9.12	10.36	12.94

* Total Points combines all ecoregion airborne lidar point cloud data

Table 10. Continued.

		RMSE							
Ecoregion	Height	Max	Mean	Mode	P50	P75	P90	P95	P99
CGP	3 m	12.85	3.49	4.43	3.78	3.29	**2.92**	3.77	6.33
EP		21.89	6.40	**5.71**	6.06	7.51	9.55	11.07	14.25
MACP		16.76	8.48	10.83	8.67	**7.62**	8.42	9.36	11.33
MAP		26.53	9.63	10.69	**9.52**	12.19	15.27	17.09	20.24
MVLP		22.32	8.20	16.77	8.41	**6.69**	8.75	10.51	13.86
SCEP		16.03	8.21	13.47	8.41	4.59	**4.06**	5.27	8.08
SEP		17.59	6.92	8.19	6.99	**6.88**	8.05	9.13	11.44
SCOP		21.42	9.78	14.43	10.42	**9.47**	10.63	11.82	14.38
SFCP		**6.66**	10.64	12.16	10.94	9.56	8.42	7.88	7.13
Total Points		19.56	8.24	11.88	8.42	**7.88**	9.29	10.54	13.07
CGP	5 m	12.85	3.02	3.22	**2.90**	3.69	5.15	6.34	9.08
EP		21.89	7.15	**5.80**	6.70	8.19	10.23	11.71	14.84
MACP		16.76	7.98	9.92	8.22	**7.55**	8.59	9.51	11.45
MAP		26.53	10.07	10.43	**9.88**	12.69	15.68	17.45	20.48
MVLP		22.32	7.49	12.37	7.64	**6.81**	9.06	10.79	14.07
SCEP		16.03	7.26	11.60	7.55	4.26	**4.19**	5.47	8.24
SEP		17.59	**6.69**	7.46	6.73	6.94	8.25	9.35	11.66
SCOP		21.42	**9.29**	13.80	9.82	9.41	10.76	11.98	14.52
SFCP		**6.66**	9.57	10.90	9.83	8.73	7.89	7.47	6.91
Total Points		19.56	**7.86**	10.42	8.00	7.97	9.52	10.78	13.28

* Total Points combines all ecoregion airborne lidar point cloud data

Our results indicated that 90th percentile at 0 m height category has the minimal bias and RMSE in most ecoregions (7 of 10 ecoregions). Our research is in agreement with the Means et al. (2000) results, which predicted the forest stand height with 90th percentile height. However, many previous studies (McGaughey et al., 2010, Drake et al., 2002, St-Onge et al., 2008, Alexander et al., 2014) recommend that 95th percentile heights airborne lidar metric, with or without ground points, is highly correlated to the field tree heights, and represent the field tree height. Therefore, to have comprehensive

results, the airborne lidar metrics of 90^{th} and 95^{th} percentile heights at 0 m height category are both applied to calibrate the GCHM.

III.3.2. The parameters for GCHM calibration

After the GCHM validation with airborne lidar metrics, the 90^{th} and 95^{th} percentile heights at 0 m height category created the simple linear regression equations with the GCHM ecoregions in order to calibrate the GCHM. The equations of the GCHM calibration (Table 11) employ the airborne lidar metrics as the dependent variable and the GCHM as the independent variable. In the Table 11, two equations of the GCHM ecoregions (CGP and EP) display high p-value and the coefficient of determinations are very low in the both 90^{th} and 95^{th} percentile height calibrations. The other equations of the GCHM ecoregions, including the entire southern United States, have shown the range of R^2 values from 0.00 to 0.53 and the range of adjusted R^2 values from -0.20 to 0.52 in the 90^{th} percentile height. Moreover, the range of R^2 values is from 0.02 to 0.52 and the range of adjusted R^2 values is from -0.18 to 0.52 in the 95^{th} percentile height. Among the equations of the GCHM ecoregions, the MVLP ecoregion has the highest R^2 and adjusted R^2 in the both 90^{th} and 95^{th} percentile heights. Figure 10 shows the corrected GCHM (cGCHM) of the southern United States by the 90^{th} and 95^{th} percentile height lidar metrics without water bodies and the original GCHM with water bodies. In Figure 10, the maximum height of Simard's GCHM is much higher than our cGCHMs. Moreover, the cGCHM by the 90^{th} percentile height corrected appears lower minimum height than the cGCHM by the 95^{th} percentile height corrected. The both

cGCHMs are validated by the airborne lidar metrics again to understand how the procedure of calibration improves the accuracy of the cGCHMs.

Table 11. The parameter, coefficient of determination, and p-value of GCHM calibration at 0 m height category

Ecoregion	Percentile Height	Intercept	Coefficient	R^2	Adj R^2	p-value
CGP	90	0.70	0.03	0.02	-0.03	0.54
EP		8.63	0.02	0.00	-0.20	0.90
MACP		12.79	0.30	0.13	0.12	0.00
MAP		6.47	0.53	0.27	0.27	0.00
MVLP		3.69	0.79	0.53	0.52	0.00
SCEP		4.38	0.75	0.28	0.27	0.00
SCOP		10.11	0.27	0.12	0.12	0.00
SEP		24.00	-0.15	0.03	0.03	0.04
SFCP		0.03	0.40	0.23	0.21	0.00
TOTAL		7.00	0.57	0.35	0.34	0.00
CGP	95	1.33	0.09	0.04	0.00	0.33
EP		10.43	0.06	0.02	-0.18	0.77
MACP		16.87	0.23	0.09	0.08	0.00
MAP		10.05	0.52	0.23	0.22	0.00
MVLP		8.91	0.73	0.52	0.52	0.00
SCEP		8.20	0.71	0.23	0.23	0.00
SCOP		12.63	0.25	0.11	0.11	0.00
SEP		27.09	-0.16	0.04	0.03	0.02
SFCP		1.86	0.39	0.19	0.18	0.00
TOTAL		9.92	0.57	0.32	0.32	0.00

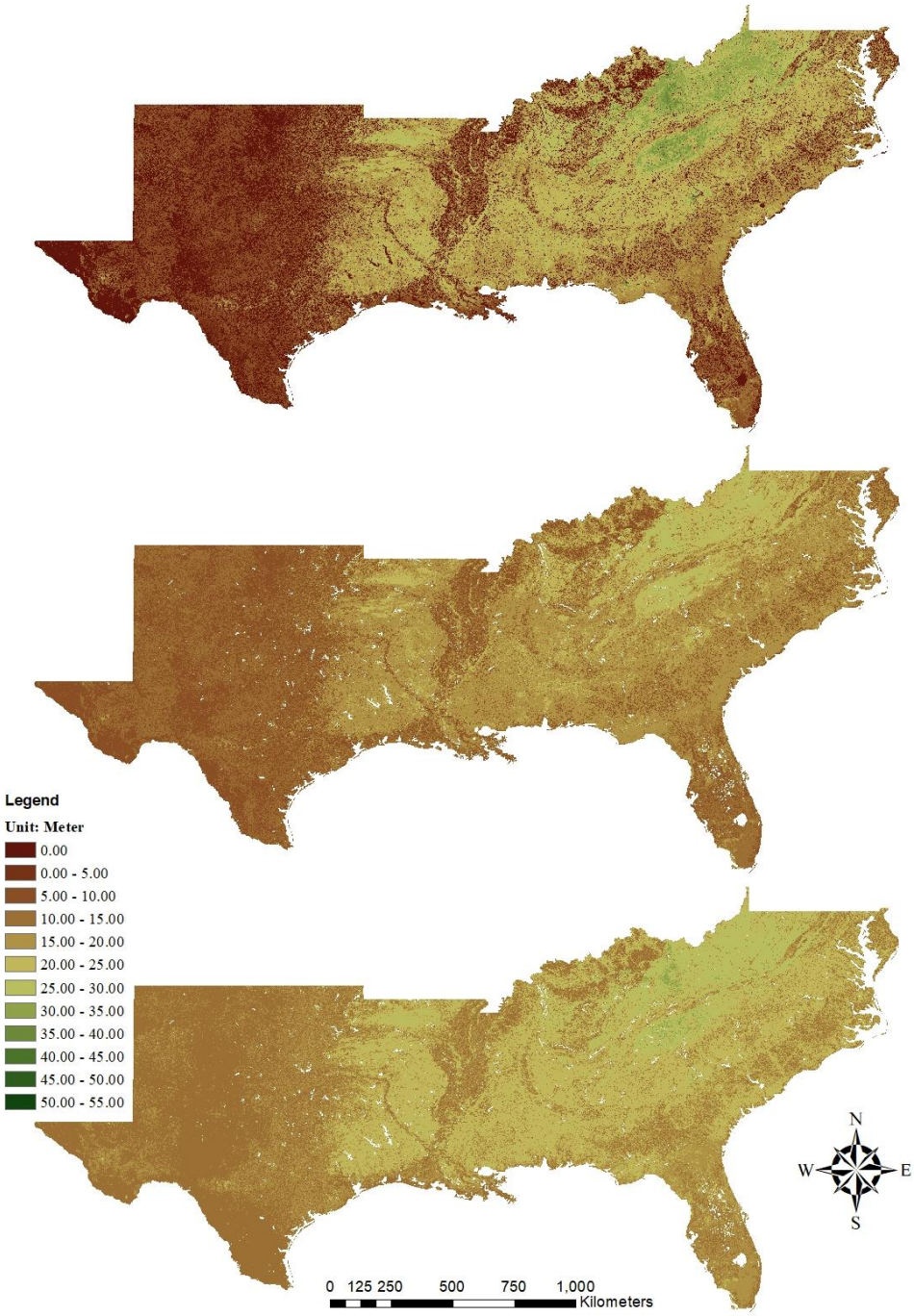

Figure 10. The global forest canopy height map and corrected global forest canopy height maps in the southern United States. The global forest canopy height map by Simard et al. (2011b) (Top), the 90th percentile height calibrated global forest canopy height map without water bodies (Middle), the 95th percentile height calibrated global forest canopy height map without water bodies (Bottom).

III.3.3. 90^{th} and 95^{th} percentile cGCHM validation

The new cGCHMs of the southern United States are validated with the testing data from the airborne lidar metrics which the training data were employed to validate the GCHM. Every ecoregion are calibrated by the 90^{th} and 95^{th} percentile heights respectively and validated by the testing data. The results of the cGCHMs validation display bias (Figure 11) and RMSE (Figure 12) between the cGCHMs and airborne lidar metrics (the max, mean, mode, 50^{th}, 75^{th}, 90^{th}, 95^{th}, and 99^{th} percentile heights). The nearest zero bias and smallest RMSE among each ecoregion are highlighted with bold and underline. In Table 12, EP, MACP, MVLP, SCEP, SCOP, and Total Points have the nearest zero bias at the 90^{th} percentile height in the calibration by the 90^{th} percentile height. CGP, MAP, and SEP have the nearest zero bias at the 95^{th} percentile height. SFCP has the nearest zero bias at the 75^{th} percentile height. The nearest zero bias ranges -1.73 m to -0.27 m and the majority ecoregions (6 of 10 ecoregions) have the nearest zero bias at the 90^{th} percentile height in the calibration by the 90^{th} percentile height (Figure 11A).

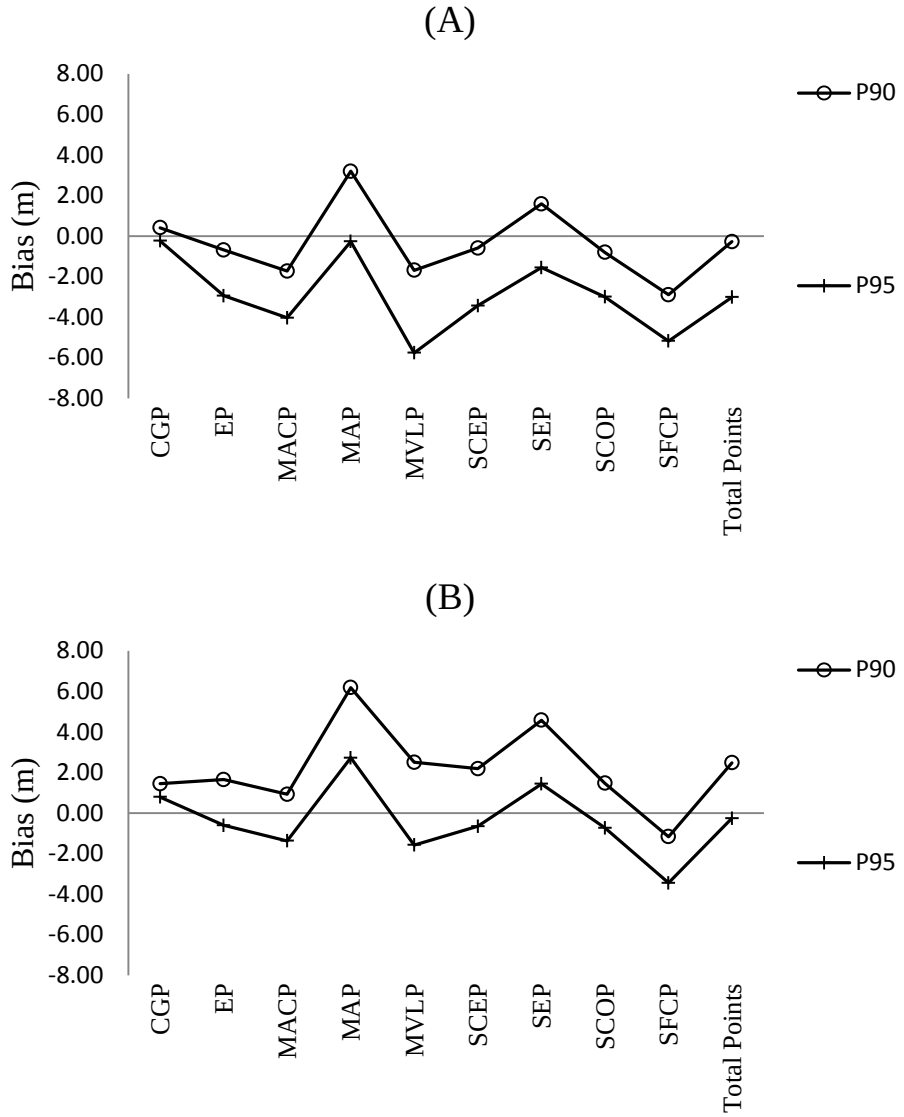

Figure 11. The bias results of cGCHM validation: the cGCHMs were calibrated by (A) 90th and (B) 95th percentile heights

In the calibration by the 95th percentile height, CGP, EP, MAP, MVLP, SCEP, SEP, SCOP, and Total Points have the nearest zero bias at the 95th percentile height. MACP and SFCP have the nearest zero bias at the 90th percentile height. The nearest zero bias ranges -1.57 m to 1.44 m and the majority ecoregions (8 of 10 ecoregions)

have the nearest zero bias at the 95th percentile height in the calibration by the 95th percentile height (Figure 11B).

Table 12. The bias validation results of cGCHM

		Bias							
Ecoregion	Percentile Height	Max	Mean	Mode	P50	P75	P90	P95	P99
CGP	90	-13.55	0.67	0.92	0.84	0.77	0.42	**-0.24**	-2.52
EP		-19.24	5.53	8.87	7.70	3.13	**-0.68**	-2.94	-6.81
MACP		-15.66	9.80	18.30	11.71	3.69	**-1.73**	-4.03	-7.49
MAP		-20.94	7.28	9.26	9.09	6.68	3.19	**-0.27**	-7.11
MVLP		-24.01	11.39	17.34	16.00	7.15	**-1.68**	-5.75	-11.68
SCEP		-15.79	11.46	20.46	13.24	5.17	**-0.59**	-3.43	-7.72
SEP		-15.80	13.51	21.24	16.18	8.09	1.59	**-1.55**	-6.10
SCOP		-14.57	7.53	12.98	9.71	3.52	**-0.80**	-3.00	-6.71
SFCP		-13.18	2.63	5.35	3.89	**2.06**	-2.89	-5.17	-7.72
Total Points		-16.55	9.58	15.84	11.95	5.11	**-0.27**	-3.01	-7.31
CGP	95	-12.53	1.70	1.95	1.87	1.79	1.44	**0.79**	-1.49
EP		-16.91	7.86	11.20	10.04	5.46	1.65	**-0.61**	-4.47
MACP		-13.00	12.46	20.95	14.37	6.35	**0.93**	-1.37	-4.83
MAP		-17.95	10.27	12.25	12.08	9.67	6.18	**2.72**	-4.12
MVLP		-19.83	15.57	21.52	20.19	11.33	2.50	**-1.57**	-7.49
SCEP		-13.01	14.24	23.24	16.02	7.95	2.19	**-0.65**	-4.94
SEP		-12.81	16.50	24.23	19.17	11.08	4.58	**1.44**	-3.12
SCOP		-12.31	9.80	15.24	11.97	5.79	1.47	**-0.73**	-4.44
SFCP		-11.46	4.36	7.08	5.61	3.79	**-1.17**	-3.45	-6.00
Total Points		-13.80	12.33	18.59	14.71	7.86	2.48	**-0.26**	-4.55

In Table 13, CGP, EP, MACP, MAP, MVLP, SCEP, SCOP, and Total Points have the smallest RMSE at the 90th percentile height in the calibration by the 90th percentile height. SEP has the smallest RMSE at the 95th percentile height. SFCP has the smallest RMSE at the mean height. The smallest RMSE ranges from 0.66 m to 7.27 m and the majority ecoregions (8 of 10 ecoregions) have the smallest RMSE at 90th percentile height in the calibration by the 90th percentile height (Figure 12A).

In the calibration by the 95th percentile height, CGP MAP, MVLP, SCEP, SEP, SCOP, and Total Points have the smallest RMSE at the 95th percentile height. EP and MACP have the smallest RMSE at the 90th percentile height. SFCP has the smallest RMSE at the mean height. The smallest RMSE ranges from 1.25 m to 8.60 m and the majority ecoregions (7 of 10 ecoregions) have the smallest RMSE at 95th percentile height in the calibration by the 95th percentile height (Figure 12B).

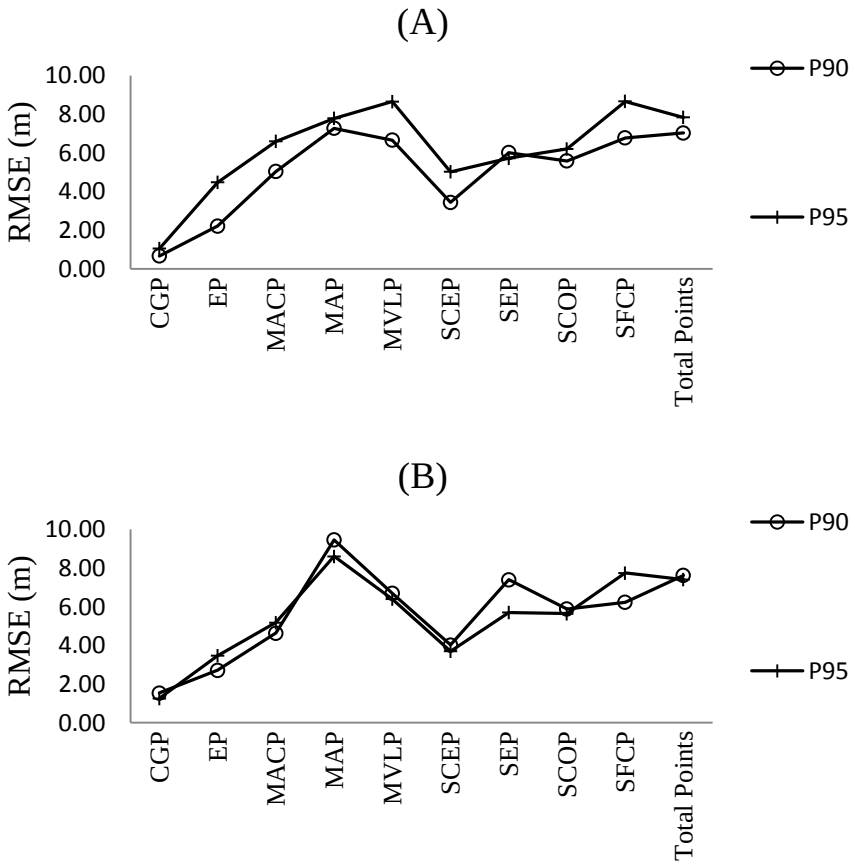

Figure 12. The RMSE results of cGCHM validation: the cGCHMs were calibrated by (A) 90th and (B) 95th percentile heights

Table 13. The RMSE validation results of cGCHM

RMSE									
Ecoregion	Percentile Height	Max	Mean	Mode	P50	P75	P90	P95	P99
CGP	90	15.88	0.69	0.93	0.85	0.77	**0.66**	1.04	2.98
EP		19.33	5.53	8.87	7.75	3.13	**2.20**	4.47	8.51
MACP		16.56	10.29	18.43	12.59	6.25	**5.03**	6.59	9.55
MAP		23.11	8.67	11.11	10.87	8.18	**7.27**	7.78	12.63
MVLP		24.90	12.94	18.88	17.44	9.30	**6.66**	8.65	13.30
SCEP		16.39	11.72	20.66	13.77	6.06	**3.43**	5.02	8.78
SEP		16.31	14.08	21.26	17.11	10.19	6.01	**5.72**	8.00
SCOP		15.85	8.40	13.33	10.66	6.62	**5.58**	6.20	8.57
SFCP		14.82	**3.72**	5.60	4.71	4.26	6.77	8.66	10.43
Total Points		18.43	10.75	16.75	13.23	8.10	**7.03**	7.83	10.43
CGP	95	14.96	1.71	1.97	1.88	1.81	1.53	**1.25**	2.10
EP		17.01	7.87	11.20	10.07	5.46	**2.70**	3.46	6.83
MACP		13.96	12.81	21.02	15.08	8.11	**4.63**	5.17	7.45
MAP		20.28	11.74	14.11	13.89	11.18	9.46	**8.60**	11.30
MVLP		20.76	16.55	22.60	21.17	12.65	6.68	**6.39**	9.60
SCEP		13.72	14.43	23.40	16.45	8.55	4.02	**3.69**	6.45
SEP		13.43	16.97	24.25	19.96	12.70	7.39	**5.69**	6.04
SCOP		13.86	10.58	15.60	12.84	8.19	5.88	**5.64**	7.06
SFCP		13.30	**5.08**	7.26	6.21	5.31	6.22	7.74	9.22
Total Points		16.06	13.37	19.47	15.86	10.20	7.62	**7.40**	8.85

In summary, the results of the cGCHM validation indicate that the calibration improve the accuracy of GCHM. The nearest zero bias of GCHM validation ranges -2.08 m to 1.02 m at the 90th percentile height in the 0 m category. Nevertheless, the nearest zero bias of cGCHM validation ranges -1.73 m to -0.27 m at the 90th percentile height in calibration by the 90th percentile height. Moreover, the smallest RMSE of GCHM validation ranges 4.11 m to 10.16 m at 90th percentile height in the 0 m category. After the calibration, the smallest RMSE of cGCHM validation ranges 0.66 m to 7.27 m.

Therefore, both bias and RMSE results show that the cGCHM has better accuracy than the GCHM.

III.4. Discussions

From the results of GCHM validation, the 90th percentile height at 0 m height category had the nearest zero bias and smallest RMSE with the GCHM. Similarly, Means et al. (2000) found that the 90th percentile height correctly predicted the forest height with airborne lidar point cloud data. Nevertheless, several research groups have mentioned using the 95th percentile height as the feasible lidar metrics to represent the forest canopy height of the field measurement (McGaughey et al., 2010, Drake et al., 2002, St-Onge et al., 2008, Alexander et al., 2014). Thus, Bolton et al. (2013) used the 95th percentile height to investigate the agreement between the GCHM of Canadian forest and airborne lidar point cloud data above 2 m. In our study, the airborne point cloud data was conducted to different height categories (0, 1, 3, and 5 m), and the 95th percentile height rarely demonstrated nearest zero bias and smallest RMSE in any height conditions. Moreover, the GCHM represented the estimated forest canopy height with 1 km^2 spatial resolution. GLAS and other climate data calculated the estimated forest canopy heights through the GLAS maximum height (RH100). Meanwhile, the procedure of creating the GCHM did not remove the remotely sensed information below specific height category. Thus, the GCHM underestimated the forest canopy height, because the GCHM included the understory vegetation, grass, and climate information when estimating the forest canopy heights.

The coefficient of determinations (R^2) of the GCHM calibration equations appears to describe weak or moderate relationships. The R^2 of GCHM with 90^{th} percentile height is between 0.00 to 0.53 and the R^2 of GCHM with 95^{th} percentile heights is between 0.02 to 0.52. Bolton et al. (2013) described similar weak and moderate correlations (0.18 ~ 0.61) between Simard's GCHM and airborne lidar data as well. Weak correlations between the airborne lidar metrics and the GCHM can arise from the following reason: the airborne lidar metrics were derived at 1 km^2 pixel, but the GCHM was the estimated forest canopy height by using GLAS and ancillary data. Nonetheless, the validation of cGCHM significantly reduced bias and RMSE comparing to the validation of GCHM. Moreover, the calibration equations of 90^{th} and 95^{th} percentile heights shorten the difference between the airborne lidar metrics and the cGCHMs. Therefore, the GCHM calibration equations improved the results of cGCHM validation despite low R^2s.

The comparison of Bolton's research and ours differ in the coverage of the airborne lidar point cloud data. Bolton et al. (2013) retrieved the airborne lidar data on average 700 m width flight lines, whereas we downloaded and collected the airborne lidar point cloud data wall-to-wall. Bolton's airborne lidar point cloud data flight lines only cover a partial GCHM pixel, so the airborne lidar point cloud data had the opportunity to over- or under-estimate the forest canopy height. In contrast, the wall-to-wall airborne lidar data offered more accurate characterization of the forest canopy height in a GCHM pixel.

Building GCHMs from the integration of the ICESat GLAS and other ancillary data allow the estimation of global forest aboveground biomass. The global forest canopy height information of GCHMs was crucial for estimating the forest aboveground biomass in previous studies (Lefsky et al., 2002b) and for investigating drought and climate change effects on vegetation (Bevan et al., 2014). However, ICESat GLAS failed in 2009 and was retired in 2010. Therefore, the GLAS data of its use for deriving GCHMs suggest the need for a successor mission to provide spaceborne lidar data with global coverage. Fortunately, the ICESat-2 and Global Ecosystem Dynamics Investigations Lidar (GEDI) missions are set to launch in 2018 and 2019, respectively. The newest photon lidar system, Advanced Topographic Laser Altimeter System (ATLAS), would board on ICESat-2. Furthermore, GEDI would be installed on the International Space Station (ISS). Both lidar systems were anticipated to provide high-resolution lidar data, therefore improving the accuracy of future GCHMs.

III.5. Conclusions

The results of GCHM validation found that GLAS and other ancillary data have underestimated the GCHM. The 90^{th} percentile height of airborne lidar metrics at the 0 m height category verified that the GCHM is affected by short vegetation heights as well. In addition, the results of GCHM validation bring up differences among the ecoregions and created difficulties to validate the GCHM with only one airborne lidar metric. Therefore, the GCHM validation and calibration are highly necessary for improving maps of vegetation biophysical parameters and each ecoregion should employ the most feasible airborne lidar metric for validation.

The results of GCHM calibration encouraged us to consider more specific calibration efforts to improve the accuracy of GCHM. The cGCHMs were considered to provide more accurate forest canopy height data than the GCHM though the R^2 values for validation/calibration are low. However, the results of cGCHM validation indicated that the airborne lidar metrics are useful and significant variables to improve the height discrepancies in the original GCHM. Besides, GCHM has the potential to be used as benchmark map for global forest aboveground biomass estimations and our investigations paved the methodology for large scale validation and calibration studies with potential usefulness for future satellite missions and their vegetation products, such as ICESat-2 and GEDI.

CHAPTER IV

REGIONAL SCALE FOREST ABOVEGROUND BIOMASS ESTIMATION OF SOUTH CENTRAL PLAINS WITH CALIBRATED GLOBAL FOREST CANOPY HEIGHT MAP

IV.1. Introduction

Forests contain 80% of Earth's terrestrial biomass (Kindermann et al., 2008) and are the largest terrestrial carbon sinks (Houghton, 2002) though ocean has been considered the largest carbon sink on Earth. Also, the estimation of terrestrial forest aboveground biomass is critical to understand the global carbon balance (Houghton, 2005). Not only human activities but also natural disasters directly influence the terrestrial forest aboveground biomass. Notably, the deforestation and intense urbanization with increasing human population imperatively threaten natural terrestrial ecosystems and the forest aboveground biomass. Therefore, the estimation and monitoring of forest aboveground biomass with effective methods offers the quantitative means to investigate the dynamics of terrestrial forest carbon.

United States Forest Service (USFS) published a contiguous U.S. forest biomass map (Blackard et al., 2008), and Woods Hole Research Center (WHRC) built a national biomass and carbon dataset (Kellndorfer et al., 2012) as well. The U.S. forest biomass map was derived by using field forest inventory data of the USFS Forest Inventory and Analysis program (FIA), Moderate Resolution Imaging Spectrometer (MODIS)-derived images, and National Land Cover Dataset (NLCD). Comparably, the WHRC national

biomass and carbon dataset combined FIA field forest inventory data with Interferometric synthetic-aperture radar (InSAR) data of the Shuttle Radar Topography Mission (SRTM) and the remotely sensed data of enhancing thematic mapper plus (ETM+) on Landsat 7. However, either the MODIS-derived images or the NLCD data of USFS forest biomass map merely offered the passive spectral and two-dimension information which overlooked measurements of forest structures. Also, although the Woods Hole's biomass data incorporated the InSAR data with the ETM+ images to provide the vegetation canopy height information, the remotely sensed data are dated to before 2002. Therefore, there is a clear need for an updated and accurate national scale forest biomass map.

The global forest canopy height map (GCHM) which was produced using spaceborne lidar and spectral remotely sensed data provided height information for large-scale forest aboveground biomass estimation. The lidar data was able to depict the structure of the forest and forest biophysical parameters, such as tree height, canopy height, diameter at breast height (DBH), and tree crown width (Lefsky et al., 1999b, 2002b). Thus, three global forest canopy height maps have been published (Lefsky, 2010, Simard et al., 2011b, Los et al., 2012). The primary data of the GCHMs was the Geoscience Laser Altimeter System (GLAS) which was the sole instrument on Ice, Cloud, and land Elevation Satellite (ICESat) and other ancillary data depended on the demands of the developers. Simard et al. (2011b) found more substantial coverage of forests than the other two GCHMs because the mosaic crops, open forest, and saline flooded forests were also considered as woody plant areas. However, Simard's GCHM

underestimated the forest canopy height which was proven in Chapter III research. Therefore, the calibrated GCHM (cGCHM) from Simard's GCHM was selected for estimating the regional forest aboveground biomass.

Due to the growing amount of remotely sensed data, machine learning algorithms were introduced to rapidly and conveniently process large remotely sensed data. Pal and Mather (2005) concluded that support vector machines produced better classification accuracy than conventional classifiers. Moreover, Duro et al. (2012) found no significant preference between the machine learning algorithms, but the pixel-based processing required less time than object-based classification. Besides, Blackard et al. (2008) utilized the tree-based Cubist approach to build the forest biomass model. In this research, we applied a machine learning algorithm, random forests (Breiman, 2001), to create the regional forest aboveground biomass map.

Random forests provide accurate predictions without overfitting the data (Cutler et al., 2007, Prasad et al., 2006). The machine learning algorithm comprised of the combinations of decision trees to classify the training data. Furthermore, random forests utilizes the bootstrap aggregating method (Bagging) randomly drew with replacement samples to generate the training data (Breiman, 1996). After Bagging selected the training data, the original data which was not selected would be placed as the out-of-bag data (OOB). The training data employed as predictors and built the predictive model according to the response variable. The OOB data was used to calculate an unbiased error rate and variable importance, or cross-validation. Ultimately, random forests

investigates the increases of the prediction error to determine the variable importance (Liaw and Wiener, 2002).

This research is the first application of the cGCHM to generate the regional scale forest aboveground biomass map. Furthermore, the new regional scale forest aboveground biomass map was compared to the existing forest aboveground biomass map at the same ecoregion. The new regional scale forest aboveground biomass map was created by using the cGCHM, which is a lidar product, but the existing forest aboveground biomass map was created using optical remotely sensed data and traditional field inventory. Thus, we anticipate that the cGCHM is able to improve the accuracy of large-scale forest aboveground biomass maps.

This research generated a regional forest aboveground biomass from local scale forest aboveground biomass maps in the South Central Plains ecoregion. The local scale forest aboveground biomass maps were generated from 2010 airborne lidar data in east Texas. The objectives of the research are to 1) generate a regional forest aboveground biomass map of the South Central Plain ecoregion, and 2) validate and compare the regional forest aboveground biomass map with the USFS forest aboveground biomass map.

IV.2. Materials and Methods

IV.2.1. Study area

The South Central Plains (Figure 13) ecoregion, as defined by Environment Protection Agency (EPA), was selected to be the study area of this research. The land surface of the ecoregion is an irregular plain with the dominant vegetation species, oaks,

hickories, and pines (Omernik, 1987). The ecoregions' land cover types are composed of woodlands and forests with croplands as well as pasture. The forests and woodlands were mostly used for grazing by livestock. The primary soil type of the ecoregion is moist ultisols. Moreover, the ecoregion covers areas across multiple states, including southwest Arkansas, west Louisiana, a small part southeast Oklahoma, and partial east Texas.

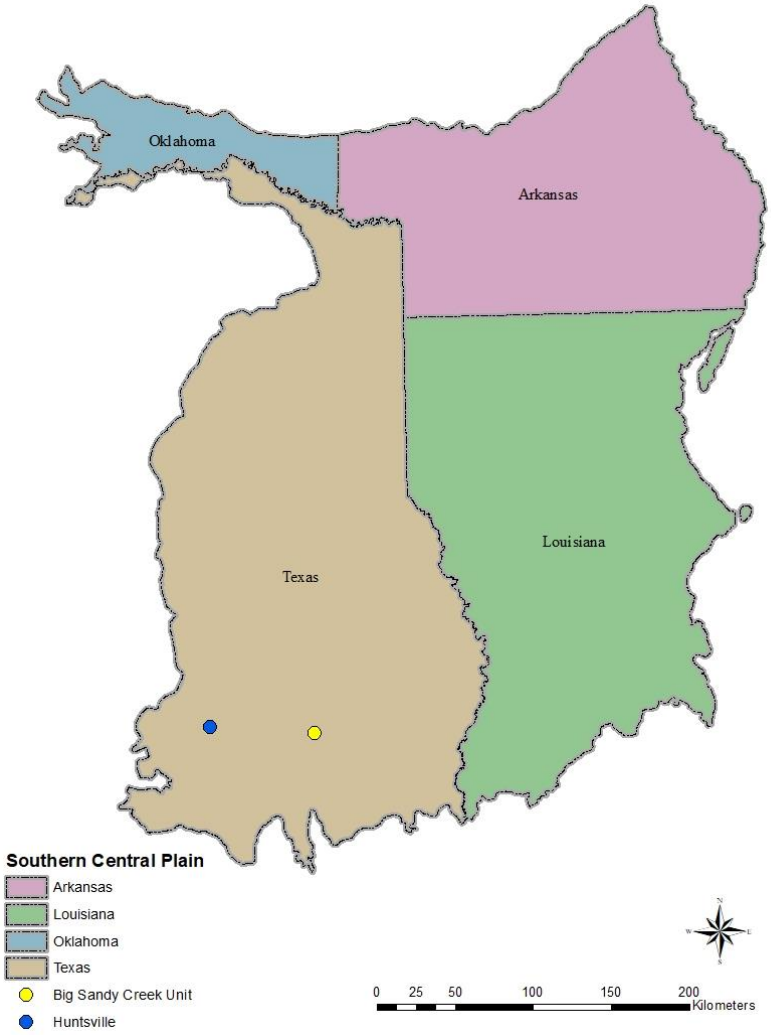

Figure 13. The area of South Central Plains (Ecoregions, 2018)

IV.2.2. The cGCHM of local scale forest aboveground biomass estimation

The airborne lidar point cloud data was provided by the Lidar Applications for the Study of Ecosystems with Remote Sensing Laboratory (LASERS) in the Department of Ecosystem Science and Management at Texas A&M University, College Station, Texas, USA. The locations of the airborne lidar data were Sam Houston National Forest near Huntsville, TX and Big Sandy Creek unit near Livingston, TX. The data was acquired by the Optech 3100EA airborne laser scanner system in November 2010. Both locations are in the South Central Plains ecoregion. The point density of the airborne lidar data was approximately 3 to 4 points/m^2. The laser pulse rate and lidar system scan rate were 150 kHz and 90 Hz, respectively.

The airborne lidar data was used to estimate the local scale forest aboveground biomass maps, along with National Agriculture Imagery Program (NAIP) images, and field tree measurements. To estimate the forest aboveground biomass, the airborne lidar data was converted to a 1 m spatial resolution canopy height model (CHM). We used the methodology described in Popescu (2007) to derive individual tree biomass using airborne lidar data and field measurements, for both pine and deciduous trees. Meanwhile, the NAIP images were classified by maximum likelihood classification to separate pine and deciduous trees. Then the classified pine and deciduous maps were aligned with the CHM to extract the pine and deciduous CHMs. After the extractions, TreeVaW (Popescu et al., 2004) was applied on the CHMs to map individual tree locations and measure the crown size. However, the diameter at breast height (DBH) of each tree was unknown. Hence, the predictive DBH (pDBH) was calculated based on the

regression of tree crowns and the field tree DBH measurements. Subsequently, the Jenkins's national level allometric equations (2003) were employed to compute each tree aboveground biomass according to the pDBH. After acquiring all individual tree aboveground biomass, the individual tree aboveground biomass was aggregated into each pixel of the cGCHM. Finally, the cGCHM and the local scale forest aboveground biomass maps at San Houston National Forest and Big Sandy Creek Unit were used for estimating regional forest aboveground biomass.

IV.2.3. The prediction variables and filters

IV.2.3.1. Corrected global forest canopy height map (cGCHM)

The cGCHMs were calibrated by the 90^{th} and 95^{th} percentile heights of airborne lidar data in the previous chapter. The GCHM (Simard et al., 2011b) was generated using multiple data sources, including 2005 ICESat GLAS, 2005 MODIS MOD44B, 2000 SRTM GTOPO elevation, 2010 protection status and auxiliary data (annual mean precipitation, precipitation seasonality, and annual mean temperature). The previous chapter found that the GCHM underestimated the forest canopy height which was similar to the 90^{th} percentile height. Thus, the corresponding airborne lidar metrics aided to improve the calibrated GCHM. Additionally, the 95^{th} percentile height was considered to represent the real canopy heights (McGaughey et al., 2010, Drake et al., 2002, St-Onge et al., 2008, Alexander et al., 2014). Therefore, both the 90^{th} and 95^{th} percentile height lidar metrics were applied to calibrate the GCHM for the South Central Plains ecoregion, as well as create 90^{th} and 95^{th} percentile height cGCHMs for providing more accurate estimations of the forest aboveground ground biomass.

IV.2.3.2. 2010 MODIS monthly vegetation indices and near-infrared images

MODIS is a valuable instrument aboard the Terra and Aqua satellites for atmosphere, land and ocean observations. The instrument is acquiring images of the entire Earth's surface every 1 to 2 days in 36 spectral bands. Moreover, the spatial resolutions of MODIS are 250 m, 500 m, and 1 km. Thus, MODIS images are valuable remotely sensed data for the regional forest aboveground biomass estimation. The MODIS remotely sensed data used in our study are two vegetation indices and the near-infrared image (NIR). The vegetation indices and NIR image were acquired from 2010 MODIS MOD13A3 and 2010 MODIS MOD13A3NIR reflectance (Band 2). The two vegetation indices were normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI). The 2010 MODIS MOD13A3 product has 16 days 250 m NDVI, 16 day and monthly 1 km NDVI and EVI, and 16 day and monthly 25 km NDVI and EVI for different spatial and temporal application. Therefore, the monthly 1 km 2010 MODIS MOD13A3 product was used because the product matched the spatial resolution of cGCHM and the temporal resolution was proper for understanding the forest dynamics monthly.

IV.2.3.3. Forest canopy cover data and mask data

The forest canopy cover data was acquired from the LANDFIRE program (LF) and National Land Cover Database 2011 (NLCD2011). The LF program is a shared program between USDA FS, Wildland Fire Leadership Council (WFLC) of United States Department of Interior. At the time of completing this study, the NLCD 2011 was

the most recent national land cover product created by the Multi-Resolution Land Characteristics (MRLC) Consortium. The LF program offered 2010 and 2012 forest canopy covers and NLCD data provided 2011 canopy cover. All three forest canopy cover data sets were at 30 m spatial resolution. Thus, all of the forest canopy cover data had to be resampled to match the spatial resolution of the cGCHM. Besides, NLCD 2011 and the National Hydrography Dataset (NHD) were applied to mask non-forest and non-woodland pixels. The response and prediction variables are shown in Table 14.

Table 14. The list of response and prediction variables

Variable	Name	Source
Response	Sam Houston National Forest biomass map Big Sandy Creek Unit biomass map	Airborne lidar
Prediction	cGCHM 90PH (0, 1, 3, and 5 m height conditions) cGCHM 95PH (0, 1, 3, and 5 m height conditions)	Calibrated GCHM
	Enhance vegetation index (EVI) (January - December) Normalized difference vegetation index (NDVI) (January - December) Near Infrared images (NIR) (January - December)	MODIS
	2010 vegetation canopy cover 2012 vegetation canopy cover	LANDFIRE
	2011 vegetation canopy cover	NLCD

IV.2.4. The estimation of regional forest aboveground biomass process

The process for developing the regional forest aboveground biomass is shown in Figure 14. The process was assembled in three phases, 1) response and prediction variables preprocessing 2) modeling and mapping regional forest aboveground biomass

estimation and 3) the regional forest aboveground biomass map post-processing. The first phase was to create and modify the response and prediction variables. The response variable was the local scale forest aboveground biomass in the cGCHM pixels. The Sam Houston National Forest and Big Sandy Creek Unit forest aboveground biomass maps were aligned to the cGCHM. Because the spatial resolution of forest aboveground biomass maps were 1 m^2 and the cGCHM was about 1 km^2, the local scale biomass maps were aggregated into the cGCHM 1-km pixels. The local scale forest aboveground biomass maps were covered by 83 cGCHM pixels. Thus, the 83 cGCHM pixels were served as the response variable for estimating regional forest aboveground biomass. Furthermore, the prediction variables, NIR, NDVI, EVI, and canopy covers, were extracted and processed to spatially match the same pixels of cGCHM. Moreover, the layers of prediction variables were extracted from the South Central Plains ecoregion as well. Finally, the response and prediction variables were split into 80% training and 20% testing data by simple random sampling.

The second phase utilized the random forests algorithm to model and mapped the regional forest aboveground biomass map. After the response and prediction variables were prepared, the ModelMap package of the R programming language employed the response and prediction variables to build the random forests model to estimate the forest aboveground biomass. Subsequently, ModelMap made use of the layers of prediction variables to build the forest aboveground biomass map of the South Central Plain ecoregion based on the model of forest aboveground biomass. Meanwhile, the prediction variable importance and OOB data were utilized to understand the prediction

variables and to validate the forest aboveground biomass map. The MSE and pseudo-R-squared were obtained after the OOB validation.

The third phase was resampling and extraction of the regional forest aboveground biomass map to compare it with the USFS forest aboveground biomass map. The spatial resolution of the USFS forest aboveground biomass map was 250 m, so the regional forest aboveground biomass map was converted from 1 km to 250 m spatial resolution through resampling. The forest and woody plant areas were extracted from the regional forest aboveground biomass map depending upon NLCD 2011 data. Moreover, the water bodies were removed based on the national hydrology database (NHD) as well. The number of pixels, areas, total biomass, maximum biomass, minimum biomass, and standard deviation from both forest aboveground biomass maps were computed to understand the differences between our map and the USFS biomass map.

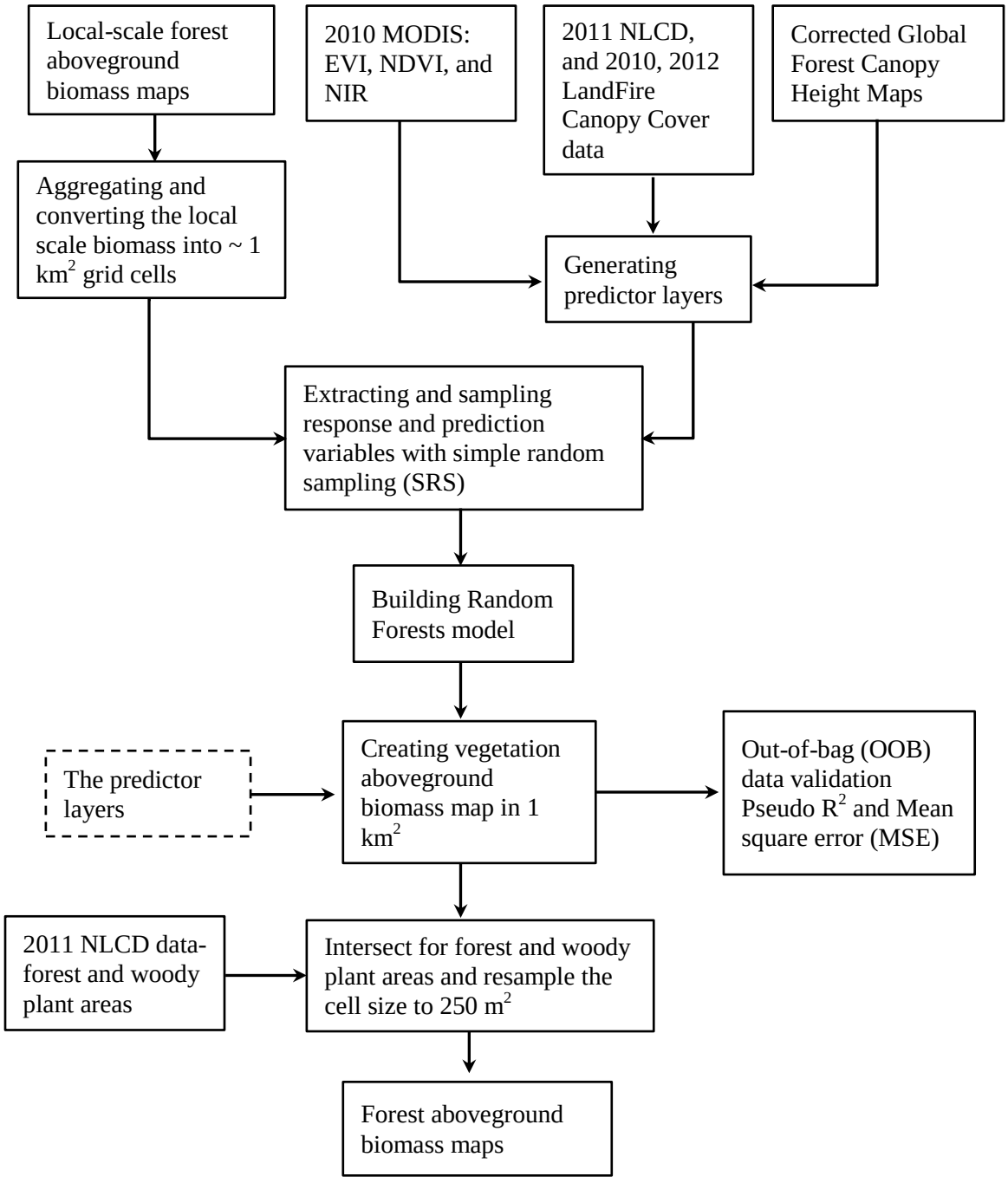

Figure 14. The flowchart of the forest aboveground biomass map development

IV.3. Results

IV.3.1. The South Central Plains forest aboveground biomass map

Both of our and USFS forest aboveground biomass maps of South Central Plains are shown in Figure 15, with blank areas within the ecoregion indicating non-woody plant areas. Our forest aboveground biomass map (Figure 15 left) displays more areas with forest aboveground biomass on the north and west South Central Plains. Notably, the northwest areas of South Central Plains in the USFS forest aboveground biomass map appear to have less forest aboveground biomass than our forest aboveground biomass map. Moreover, the northeast areas of South Central Plains on our forest aboveground biomass map appear to have significantly more forest aboveground biomass along the streams. Comprehensively, our forest aboveground biomass map exhibits larger forest aboveground biomass areas than the USFS forest aboveground biomass map.

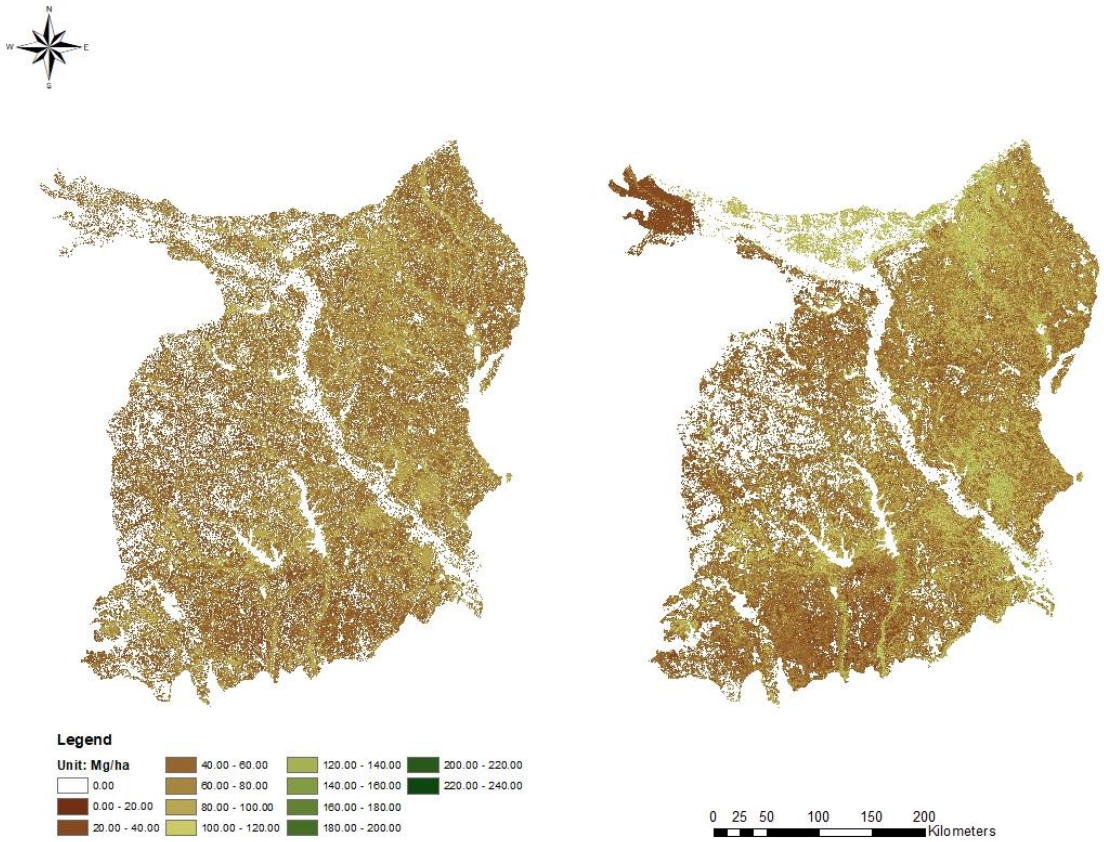

Figure 15. The forest aboveground biomass maps in South Central Plain ecoregion. The final forest aboveground biomass map (Left). The forest aboveground biomass map from USFS (Blackard et al., 2008) (Right).

IV.3.2. The comparison of forest aboveground biomass maps in South Central Plains

The comparison of both USFS and our final forest aboveground biomass maps is displayed in Table 15. Apparently, the coefficient of determination (R^2) of our forest aboveground biomass map is much higher than the USFS forest aboveground biomass. Statistically, according to the descriptive statistic in Table 15, our forest aboveground biomass map displays larger forest aboveground biomass areas than USFS forest aboveground biomass map. In fact, the differences of the forest aboveground biomass

areas are also observed by visualizing the maps. However, our total and maximum forest aboveground biomass is lower than USFS's. Additionally, the standard deviation of our forest aboveground biomass map is smaller than USFS forest aboveground biomass map.

Table 15. The comparison of USFS forest aboveground biomass map and the final forest aboveground biomass map

	USFS biomass map	Final biomass map
R^2	0.06*	0.34
RMSE (Mg/ha)	N/A	27.85
Spatial resolution (m)	250.00	250.00
Number of pixels	4373746.00	4406398.00
Area (km^2)	273359.13	275399.88
Total biomass (Mg/ha)	127905656.49	106587371.15
Minimum biomass (Mg/ha)	0.00	21.91
Maximum biomass (Mg/ha)	226.09	116.82
Standard deviation	36.75	15.96

* The mapping zone 37 (Blackard et al., 2008)

IV.4. Discussions

The cGCHM represents the canopy height in 1 km^2grid cells, but the canopy height values are not indicative of the forest types. Thus, the ecoregion was critical to aid the forest type information for the forest aboveground biomass estimation. In this research, the South Central Plains ecoregion is covered by mixed forests with diverse tree species and varying tree heights. The dominant species in the South Central Plains were oaks, hickories, and pines.

Despite the fact that our forest aboveground biomass map showed higher R^2 value than USFS forest aboveground biomass map, the R^2 value of our forest aboveground biomass map was still lower than 0.5. Several strategies for the forest

aboveground biomass estimation were considered to improve the accuracy of the forest aboveground biomass map. First, the forest aboveground biomass was influenced by many environmental factors. In this research, we selected the vegetation indices, canopy height, near-infrared image, and canopy cover percentage to estimate the forest aboveground biomass. However, the climate, soil, hydrological and other ancillary data are influencing forest growth as well. Thus, more suitable prediction variables could assist random forests to estimate the forest aboveground biomass in the proper environment condition. Second, the improvement of cGCHM accuracy was critical because the canopy height directly affected the accuracy of forest aboveground biomass. The cGCHM was obtained after the calibration of Simard's GCHM (Simard et al., 2011b). Simard et al. built the GCHM with ICESat GLAS data, a mission that was retired in 2010. Therefore, newer spaceborne lidar data, like the data from ICESat-2 and Global Ecosystem Dynamics Investigation (GEDI), are expected to generate updated and more accurate GCHM. Third, the availability of a local forest aboveground biomass map generated with airborne lidar data was an important factor in our methodology for upscaling the biomass values to larger extents. Our research areas only covered two local biomass maps that are representative for East Texas pine and deciduous species, but the sampling process with random forests is therefore localized and not spread throughout the South Central Plains. Therefore, more maps estimating local-scale forest aboveground biomass spread in our region of interest could aid in improving the accuracy of deriving regional forest aboveground biomass maps.

The comparison between the USFS forest aboveground biomass map and our forest aboveground biomass map showed that our map was more accurate, though the R^2 was low. The USFS forest aboveground biomass map used passive remotely sensed data, MODIS, as the primary data, and did not involve any active remotely sensed data with canopy height information. Although optical remotely sensed data are well known to classify conifer and deciduous trees (Key et al., 2001, Haala and Brenner, 1999), the tree physical measurement information (height, DBH, and crown width) is needed for more accurate aboveground biomass estimation. Our forest aboveground biomass map not only used the lidar data for the response variable but also included cGCHM to estimate the forest aboveground biomass. cGCHM provided the canopy height parameter which was critical for improving the estimates of the forest aboveground biomass.

In addition, the USFS maximum forest aboveground biomass (226.09 Mg/ha) was much higher than our maximum forest aboveground biomass (116.82 Mg/ha). Moreover, USFS forest aboveground biomass map displayed the minimum forest aboveground biomass as 0 Mg/ha, but our minimum forest aboveground biomass was 21.91 Mg/ha. The USFS maximum forest aboveground biomass was almost two times larger than ours. Brown et al. (1999) found that the hardwood forest biomass was lower than 100 Mg/ha in Iowa, Missouri, Oklahoma, and Texas. Additionally, Schroeder et al. (1997) studied the temperate broadleaf forest aboveground biomass in the U.S. and indicated that the aboveground biomass of both oak-hickory and maple-beech-birch forests were between 28 to 200 Mg/ha. The maximum broadleaf forest aboveground biomass was lower than 200 Mg/ha as well. Therefore, our forest aboveground biomass

map was more reasonable than USFS forest aboveground biomass map. Nevertheless, USFS forest aboveground biomass map had high correlation coefficients (0.73) in the Pacific Northwest regions (Blackard et al., 2008). Temperate rainforest covered the Pacific Northwest regions with ocean climate in coastal areas and Alpine climate in the high mountains. The most forest types were conifer forest in Pacific Northwest regions whereas the forest types were a mixed forest in South Central Plains. Thus, the USFS forest aboveground biomass map apparently had better estimation in the conifer forests. Furthermore, the USFS estimated the forest aboveground biomass nationally.

IV.5. Conclusions

This research concluded that the random forests method for building our forest aboveground biomass map shows stable and reasonable forest aboveground biomass estimation. Although the R^2 of our forest aboveground biomass map was low, the amount of forest aboveground biomass was close to previous studies. In addition, cGCHM was able to build a regional scale forest aboveground biomass map. Therefore, the random forests method and cGCHM could create a reliable regional forest aboveground biomass map.

In the future, ICESat-2 and GEDI will offer better spaceborne lidar data for developing more accurate GCHM than ICESat GLAS. Although GCHM is necessary to be calibrated, we expect that a new cGCHM will assist to create more accurate regional-scale and larger-scale forest aboveground biomass maps. Accurate forest aboveground biomass maps would improve estimation and monitoring of terrestrial carbon storage

CHAPTER V

CONCLUSIONS

Chapter II found the best fitting method and remotely sensed data combination for the woody plant aboveground biomass estimation on rangelands. The random forests method with combined NAIP and lidar data performed the best result for adequately estimating the woody plant aboveground biomass. Moreover, the pixel-based of local woody plant aboveground biomass map offered a novel spatial explicit approach for the future biomass map development.

Chapter III research showed that GLAS and other ancillary data had underestimated Simard's GCHM because of GCHM including short vegetation heights. Also, the research also concluded that the 95^{th} percentile height among the airborne lidar metrics was not the best lidar metric to validate tree height in various ecoregions. Therefore, the GCHM validation is mandatory, and each ecoregion should be validated with the feasible airborne lidar metric. Furthermore, the GCHM calibration was necessary for the future forest and carbon cycle studies.

Chapter IV research concluded that the random forests method is the suitable statistical method to build reasonable forest aboveground biomass maps. Besides, the research has proven that cGCHM is an appropriate spaceborne lidar product to generate a regional scale forest aboveground biomass map. Therefore, the random forests method associate with GCC indeed creates a reliable regional forest aboveground biomass map.

In the future, ICESat-2 and GEDI are anticipated to create updated and accurate GCHMs. Such global canopy height products are expected to estimate regional and larger-scale forest aboveground biomass more accurately and improve estimation and monitoring of terrestrial carbon storage.

REFERENCES

Alexander, C., Bocher, P.K., Arge, L., & Svenning, J. (2014). Regional-scale mapping of tree cover, height and main phenological tree types using airborne laser scanning data. *Remote Sensing of Environment, 147*, 156

Anantha M. Prasad, Louis R. Iverson, & Andy Liaw. (2006). Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems, 9*, 181-199

Anaya, J.A., Chuvieco, E., & Palacios-Orueta, A. (2009). Aboveground biomass assessment in Colombia: A remote sensing approach. *Forest Ecology and Management, 257*, 1237-1246

Anderson, G.L., Hanson, J.D., & Haas, R.H. (1993). Evaluating landsat thematic mapper derived vegetation indices for estimating above-ground biomass on semiarid rangelands. *Remote Sensing of Environment, 45*, 165-175

Ansley, R.J., Pinchak, W.E., Teague, W.R., Kramp, B.A., Jones, D.L., & Barnett. K. (2010). Integrated Grazing and Prescribed Fire Restoration Strategies in a Mesquite Savanna: II. Fire Behavior and Mesquite Landscape Cover Responses. *Rangeland Ecology & Management, 63*, 286-297

Ansley, R.J., Wu, X.B., & Kramp, B.A. (2001). Observation: Long-Term Increases in Mesquite Canopy Cover in a North Texas Savanna. *Journal of Range Management, 54*, 171-176

Asner, G.P., Mascaro, J., Muller-Landau, H.C., Vieilledent, G., Vaudry, R., Rasamoelina, M., et al. (2012). A universal airborne LiDAR approach for tropical forest carbon mapping. *Oecologia, 168,* 1147-1160

Baldocchi, D. (2008). TURNER REVIEW No. 15. 'Breathing' of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Australian Journal of Botany, 56,* 1-26

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., et al. (2001). FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem–Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. *Bulletin of the American Meteorological Society, 82,* 2415-2434

Bannari, A., Morin, D., Bonn, F., & Huete, A.R. (1995). A review of vegetation indices. *Remote Sensing Reviews, 13,* 95-120

Bethany A. Bradley, & John F. Mustard. (2006). Characterizing the Landscape Dynamics of an Invasive Plant and Risk of Invasion Using Remote Sensing. *Ecological Applications, 16,* 1132-1147

Bevan, S.L., Los, S.O., & North, P.R.J. (2014). Response of vegetation to the 2003 European drought was mitigated by height. *Biogeosciences, 11,* 2897-2908

Bhattacharya, B.K., & Chattopadhyay, C. (2013). A multi-stage tracking for mustard rot disease combining surface meteorology and satellite remote sensing. *Computers and Electronics in Agriculture, 90,* 35-44

Blackard, J.A., Finco, M.V., Helmer, E.H., Holden, G.R., Hoppus, M.L., Jacobs, D.M., et al. (2008). Mapping U.S. forest biomass using nationwide forest inventory data

and moderate resolution information. *Remote Sensing of Environment, 112*, 1658-1677

Bolton, D.K., Coops, N.C., & Wulder, M.A. (2013). Investigating the agreement between global canopy height maps and airborne Lidar derived height estimates over Canada. *Canadian Journal of Remote Sensing, 39*, S151

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. (1996). Bagging predictors. *Machine Learning, 24*, 123-140

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

Brown, S.L., Schroeder, P., & Kern, J.S. (1999). Spatial distribution of biomass in forests of the eastern USA. *Forest Ecology and Management, 123*, 81-90

Calders, K., Newnham, G., Burt, A., Murphy, S., Raumonen, P., Herold, M., et al. (2015). Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods in Ecology and Evolution, 6*, 198-208

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

Casady, G.M, VanLeeuwen, W.JD, & Reed, B.C. (2013). Estimating Winter Annual Biomass in the Sonoran and Mojave Deserts with Satellite- and Ground-Based Observations. *Remote Sensing, 5*, 909-926

Census Regions and Divisions of the United States. , *accessed on March 22, 2018*

Cho, M.A., Skidmore, A., Corsi, F., van Wieren, S.E., & Sobhan, I. (2007). Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observations and Geoinformation, 9*, 414-424

Chuvieco, E., Aguado, I., Yebra, M., Nieto, H., Salas, J., Martín, M.P., et al. (2010). Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecological Modelling, 221*, 46-58

Colgan, M.S., Asner, G.P., Levick, S.R., Martin, R.E., & Chadwick, O.A. (2012). Topo-edaphic controls over woody plant biomass in South African savannas. *Biogeosciences, 9*, 1809-1821

Coombs, J., Hall, D.O., & Long, S.P. (2014). *Techniques in Bioproductivity and Photosynthesis: Pergamon International Library of Science, Technology, Engineering and Social Studies.* : Pergamon

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., et al. (2007). Random Forests for Classification in Ecology. *Ecology, 88*, 2783-2792

Dardel, C., Kergoat, L., Hiernaux, P., Mougin, E., Grippa, M., & Tucker, C.J. (2014). Re-greening Sahel: 30years of remote sensing data and field observations (Mali, Niger). *Remote Sensing of Environment, 140*, 350-364

DeFries, R.S., & Townshend, J.R.G. (1994). NDVI-derived land cover classifications at a global scale. *International Journal of Remote Sensing, 15*, 3567-3586

Dockray, M., Horler, D.N.H., & Barber, J. (1983). The red edge of plant leaf reflectance. *International Journal of Remote Sensing, 4*, 273-288

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

Duro, D.C., Franklin, S.E., & Dubé, M.G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment, 118*, 259-272

Ecoregions | Ecosystem Research | US EPA. , *accessed on March 22, 2018*

Estornell, J., Ruiz, L.A., 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

Felker, P., Clark, P.R., Osborn, J.F., & Cannell, G.H. (1982). Biomass Estimation in a Young Stand of Mesquite (Prosopis spp.), Ironwood (Olneya tesota), Palo Verde (Cercidium floridium, and Parkinsonia aculeata), and Leucaena (Leucaena leucocephala). *Journal of Range Management, 35*, 87-89

Filella, I., & Penuelas, J. (1994). The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International Journal of Remote Sensing, 15*, 1459-1470

Freeman, E.A., Frescino, T.S., & Moisen, G.G. (2016). ModelMap: an R Package for Model Creation and Map Production. *R Package Version*, 12

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software, 33*

Galidaki, G., Zianis, D., Gitas, I., Radoglou, K., Karathanassi, V., Tsakiri-Strati, M., et al. (2017). Vegetation biomass estimation with remote sensing: focus on forest and other wooded land over the Mediterranean ecosystem. *International Journal of Remote Sensing, 38,* 1940-1966

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

Gartzia, M., Alados, C.L., & Pérez-Cabello, F. (2014). Assessment of the effects of biophysical and anthropogenic factors on woody plant encroachment in dense and sparse mountain grasslands based on remote sensing data. *Progress in Physical Geography, 38,* 201-217

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

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., et al. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science, 342,* 850-853

Harding, D.J. & Carabajal, C.C. (2005). ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters, 32,* L21S10

Hawbaker, T.J., Keuler, N.S., Lesak, A.A., Gobakken, T., Contrucci, K., & Radeloff, V.C. (2009). Improved estimates of forest vegetation structure and biomass with

a LiDAR-optimized sampling design. *Journal of Geophysical Research - Biogeosciences, 114*, n/a

Hilbert, C., Schmullius, C., & Zink, M. (Oct 24, 2011). Analysis of ICESat/GLAS waveform data for characterizing forests in a hilly region. *Proceedings of SPIE, 8286*, 828628-8286210

Hilker, T., Lyapustin, A.I., Tucker, C.J., Hall, F.G., Myneni, R.B., Wang, Y., et al. (2014). Vegetation dynamics and rainfall sensitivity of the Amazon. *Proceedings of the National Academy of Sciences of the United States of America, 111*, 16041-16046

Hodgson, M.E., Jensen, J.R., Schmidt, L., Schill, S., & Davis, B. (2003). An evaluation of LIDAR- and IFSAR-derived digital elevation models in leaf-on conditions with USGS Level 1 and Level 2 DEMs. *Remote Sensing of Environment, 84*, 295-308

Houghton, R.A. (2002). Terrestrial carbon sinks--uncertain explanations. *Biologist (London, England), 49*, 155

Houghton, R.A. (2005). Aboveground Forest Biomass and the Global Carbon Balance. *Global Change Biology, 11*, 945-958

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

Jensen, J.R. (2007). *Remote sensing of the environment.* Upper Saddle River, NJ: Pearson, Prentice Hall

Jin, S., Yang, L., Danielson, P., Homer, C., Fry, J., & Xian, G. (2013). A comprehensive change detection method for updating the National Land Cover Database to circa 2011. *Remote Sensing of Environment, 132,* 159-175

Kankare, V., Vastaranta, M., Holopainen, M., Räty, M., Yu, X., Hyyppä, J., et al. (2013). Retrieval of Forest Aboveground Biomass and Stem Volume with Airborne Scanning LiDAR. *Remote Sensing, 5*, 2257-2274

Kantola, T., Vastaranta, M., Yu, X., Lyytikainen-Saarenmaa, P., Holopainen, M., Talvitie, M., et al. (2010). Classification of Defoliated Trees Using Tree-Level Airborne Laser Scanning Data Combined with Aerial Images. *Remote Sensing, 2*, 2665-2679

Kellndorfer, J., Walker, W., Kirsch, K., Fiske, G., Bishop, J., Lapoint, L., et al. (2012). NACP Aboveground Biomass and Carbon Baseline Data (NBCD 2000), U.S.A., 2000.

Key, T., Warner, T.A., McGraw, J.B., & Fajvan, M.A. (2001). A Comparison of Multispectral and Multitemporal Information in High Spatial Resolution Imagery for Classification of Individual Tree Species in a Temperate Hardwood Forest. *Remote Sensing of Environment, 75*, 100-112

Kindermann, G., 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

Ku, N., Popescu, S.C., Ansley, R.J., Perotto-Baldivieso, H.L., & Filippi, A.M. (2012). Assessment of available rangeland woody plant biomass with a terrestrial LIDAR system. *Photogrammetric Engineering & Remote Sensing, 78*, 349-361

Kushwaha, S., Nandy, S., & Gupta, M. (2014). Growing stock and woody biomass assessment in Asola-Bhatti Wildlife Sanctuary, Delhi, India. *Environmental Monitoring and Assessment, 186*, 5911-5920

Lefsky, M. A., Cohen, W.B., Harding, D.J., Parker, G.G., Acker, S.A., & Gower, S.T. (2002a). Lidar Remote Sensing of Above-Ground Biomass in Three Biomes. *Global Ecology and Biogeography, 11*, 393-399

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

Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A., & Harding, D. (1999a). Lidar Remote Sensing of the Canopy Structure and Biophysical Properties of Douglas-Fir Western Hemlock Forests. *Remote Sensing of Environment, 70*, 339-361

Lefsky, M.A., Cohen, W.B., Parker, G.G. & Harding. D.J., (2002b). Lidar Remote Sensing for Ecosystem Studies. *BioScience, 52*, 19-30

Lefsky, M.A., Harding, D., Cohen, W.B., Parker, G., & Shugart, H.H. (1999b). Surface Lidar Remote Sensing of Basal Area and Biomass in Deciduous Forests of Eastern Maryland, USA. *Remote Sensing of Environment, 67*, 83-98

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

Los, S.O., Rosette, J.A.B., Kljun, N., North, P.R.J., Chasmer, L., Suárez, J.C., et al. (2012). Vegetation height and cover fraction between 60° S and 60° N from ICESat GLAS data. *Geoscientific Model Development, 5*, 413-432

McGaughey, R.J., Bakker, J.D., Gersonde, R.F., Kane, V.R., Lutz, J.A., & Franklin, J.F. (2010). Comparisons between field- and LiDAR-based measures of stand structural complexity. *Canadian Journal of Forest Research, 40*, 761-773

Means, J.E., Acker, S.A., Fitt, B.J., Renslow, M., Emerson, L., & Hendrix, C.J. (2000). Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering and Remote Sensing, 66*, 1367-1372

Mohamed, A.H., Holechek, J.L., Bailey, D.W., Campbell, C.L., & DeMers, M.N. (2011). Mesquite encroachment impact on southern New Mexico rangelands: remote sensing and geographic information systems approach. *Journal of Applied Remote Sensing, 5*, 053514

Morsdorf, F., Nichol, C., Malthus, T., & Woodhouse, I.H. (2009). Assessing forest structural and physiological information content of multi-spectral LiDAR waveforms by radiative transfer modelling. *Remote Sensing of Environment, 113*, 2152-2163

Mustonen, J., Packalén, P., & Kangas, A. (2008). Automatic segmentation of forest stands using a canopy height model and aerial photography. *Scandinavian Journal of Forest Research, 23*, 534-545

Mutanga, O., Adam, E., & Cho, M.A. (2012). High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observations and Geoinformation, 18*, 399

Næsset, E. (2011). Estimating above-ground biomass in young forests with airborne laser scanning. *International Journal of Remote Sensing, 32*, 473-501

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

Nelson, R., Krabill, W., & Tonelli, J. (1988). Estimating forest biomass and volume using airborne laser data. *Remote Sensing of Environment, 24*, 247-267

Ni-Meister, W., Lee, S., Strahler, A.H., Woodcock, C.E., Schaaf, C., Yao, T., et al. (2010). Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. *Journal of Geophysical Research: Biogeosciences, 115*, n/a

Olsson, P., Jönsson, A.M., & Eklundh, L. (2012). A new invasive insect in Sweden – Physokermes inopinatus: Tracing forest damage with satellite based remote sensing. *Forest Ecology and Management, 285*, 29-37

Omernik, J., & Griffith, G. (2014). Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework. *Environmental Management, 54*, 1249-1266

Omernik, J.M. (1987). Ecoregions of the Conterminous United States. *Annals of the Association of American Geographers, 77,* 118-125

Pal, M., & Mather, P.M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing, 26,* 1007-1011

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

Popescu, S.C., & Wynne, R.H. (2004). Seeing the trees in the forest. *Photogrammetric Engineering & Remote Sensing, 70,* 589-604

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

Myneni, R.B., Dong, J., Tucker, C.J., Kaufmann, R.K., Kauppi, P.E., Liski, J., Zhou, L., Alexeyev, V. & Hughes, M.K. (2001). A Large Carbon Sink in the Woody Biomass of Northern Forests. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 14784-14789

Raynolds, M.K., Walker, D.A., Epstein, H.E., Pinzon, J.E., & Tucker, C.J. (2012). A new estimate of tundra-biome phytomass from trans-Arctic field data and AVHRR NDVI. *Remote Sensing Letters, 3,* 403-411

Riaño, D., Chuvieco, E., Ustin, S.L., Salas, J., Rodríguez-Pérez, J.R., Ribeiro, L.M., et al. (2007). Estimation of shrub height for fuel-type mapping combining airborne LiDAR and simultaneous color infrared ortho imaging. *International Journal of Wildland Fire, 16,* 341-348

Riegel, J.B., Bernhardt, E., & Swenson, J. (2013). Estimating Above-Ground Carbon Biomass in a Newly Restored Coastal Plain Wetland Using Remote Sensing. *PLoS One, 8*, e68251

Sankey, T.T., & Bond, P. (2011). LiDAR-Based Classification of Sagebrush Community Types. *Rangeland Ecology & Management, 64*, 92-98

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. & Petrova, S., et al. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 9899-9904

Schroeder, P., Brown, S., Mo, J., Birdsey, R., & Cieszewski, C. (1997). Biomass estimation for temperate broadleaf forests of the United States using inventory data. *Forest Science, 43*, 424-434

Simard, M., Pinto, N., Fisher, J.B., & Baccini, A. (2011a). Global 1km Forest Canopy Height. *, accessed on March 22, 2018*

Simard, M., Pinto, N., Fisher, J.B., & Baccini, A. (2011b). Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research, 116*

Simard, M., Zhang, K., Rivera-Monroy, V.H., Ross, M.S., Ruiz, P.L., Castañeda-Moya, E., et al. (2006). Mapping height and biomass of mangrove forests in Everglades National Park with SRTM elevation data. *Photogrammetric Engineering & Remote Sensing, 72*, 299-311

St-Onge, B., Vega, C., Fournier, R.A., & Hu, Y. (2008). Mapping canopy height using a combination of digital stereo-photogrammetry and lidar. *International Journal of Remote Sensing, 29*, 3343-3364

Sun, G., Ranson, K.J., Kimes, D.S., Blair, J.B., & Kovacs, K. (2008). Forest vertical structure from GLAS: An evaluation using LVIS and SRTM data. *Remote Sensing of Environment, 112*, 107-117

Thenkabail, P.D., Prasad S. (2015). *Land Resources Monitoring, Modeling, and Mapping with Remote Sensing.* Florence: CRC Press

Tian, X., Su, Z., Chen, E., & Li, Z. (2012). Estimation of forest above-ground biomass using multi-parameter remote sensing data over a cold arid area. *International journal of applied earth observation and geoinformation, 14*, 160-168

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological), 58*, 267-288

Todd, S.W., Hoffer, R.M., & Milchunas, D.G. (1998). Biomass estimation on grazed and ungrazed rangelands using spectral indices. *International Journal of Remote Sensing, 19*, 427-438

Vashum, K.T., & Jayakumar, S. (2012). Methods to estimate above-ground biomass and carbon stock in natural forests-a review. *J.Ecosyst.Ecogr, 2*, 1-7

Vastaranta, M., Kankare, V., Holopainen, M., Yu, X., Hyyppä, J., & Hyyppä, H. (2012). Combination of individual tree detection and area-based approach in imputation of forest variables using airborne laser data. *ISPRS Journal of Photogrammetry and Remote Sensing, 67*, 73-79

Verbesselt, J., Robinson, A., Stone, C., & Culvenor, D. (2009). Forecasting tree mortality using change metrics derived from MODIS satellite data. *Forest Ecology and Management, 258,* 1166-1173

Watts, J.D., Lawrence, R.L., Miller, P.R., & Montagne, C. (2009). Monitoring of cropland practices for carbon sequestration purposes in north central Montana by Landsat remote sensing. *Remote Sensing of Environment, 113,* 1843-1852

Zandler, H., Brenning, A., & Samimi, C. (2015). Quantifying dwarf shrub biomass in an arid environment: comparing empirical methods in a high dimensional setting. *Remote Sensing of Environment, 158,* 140-155

Zhang, G., Ganguly, S., Nemani, R.R., White, M.A., Milesi, C., Hashimoto, H., et al. (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

Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moine, J., et al. (2004). Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote Sensing of Environment, 93,* 402-411