Remote Sensing of Environment 125 (2012) 80–91

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Forest biomass estimation from airborne LiDAR data using machine learning approaches

Colin J. Gleason [a], Jungho Im [b,c,*]

[a] Department of Geography, University of California, Los Angeles, CA, USA
[b] School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan, 689-798, South Korea
[c] Department of Environmental Resources Engineering, State University of New York, College of Environmental Science and Forestry, Syracuse, NY 13210, USA

ARTICLE INFO

Article history:
Received 17 August 2011
Received in revised form 9 July 2012
Accepted 11 July 2012
Available online 5 August 2012

Keywords:
Biomass estimation
Lidar remote sensing
Machine learning
Random forest
Cubist
Support vector regression
Linear mixed effects regression
Tree crown delineation

ABSTRACT

During the past decade, procedures for forest biomass quantification from light detection and ranging (LiDAR) data have been improved at a rapid pace. The scope of these methods ranges from simple regression between LiDAR-derived height metrics and biomass to methods including automated tree crown delineation, stochastic simulation, and machine learning approaches. This study compared the effectiveness of four modeling techniques—linear mixed-effects (LME) regression, random forest (RF), support vector regression (SVR), and Cubist—for estimating biomass in moderately dense forest (40–60% canopy closure) at both tree and plot levels. Tree crowns were delineated to provide model estimates of individual tree biomass and investigate the effects of delineation accuracy on biomass modeling. We used our previously developed method (COTH) to delineate tree crowns. Results indicate that biomass estimation accuracy improves when modeled at the plot level and that SVR produced the most accurate biomass model (671 kg RMSE per 380 m^2 plot when forest plots were modeled as a collection of trees). All models provided similar results when estimating biomass at the individual tree level (505, 506, 457, and 502 kg RMSE per tree). We assessed the effect of crown delineation accuracy on biomass estimation by repeating the modeling procedures with manually delineated crowns as inputs. Results indicated that manually delineated crowns did not always produce superior biomass models and that the relationship between crown delineation accuracy and biomass estimation accuracy is complex and needs to be further investigated.

1. Introduction

Quantifying the amount of biomass within a forest stand is necessary for property managers to make informed decisions about the value and use of their forested land. Biomass quantification procedures developed for remotely sensed data, especially from LiDAR data, have been published at a rapid pace with an increasing complexity and variety of techniques (Gleason & Im, 2011). LiDAR data are well suited to biomass estimation, as point clouds generated from forest canopies can accurately depict the physical characteristics of the canopy surface. These physical characteristics are correlated with biomass, and may be regressed against either diameter at breast height (dbh) or biomass to obtain general LiDAR-biomass models (Salas et al., 2010; Zhao et al., 2009). More recently, biomass quantification procedures have moved away from the regression between LiDAR-derived height metrics and biomass and increasingly include methods for automated tree crown delineation, stochastic simulation, and machine learning (Breidenbach et al., 2010; Muinonen et al., 2001; Salas et al., 2010; Vauhkonen et al., 2010).

Biomass quantification models (regardless of method) must model forests in practical units, which is accomplished either by estimating biomass for individual trees or for semi-arbitrary areas of forest (i.e., plots). Identifying treetops is often the first step in locating individual trees, as biomass is strongly correlated with crown width and other crown dimensions that can be derived from treetop position (Popescu, 2007). This requires accurate field data describing the position and height of each tree within a study plot as well as measurements of crown dimensions. Without such field data, treetops may be identified from LiDAR data, often through the process of local maxima filtering. There are numerous methods of local maxima detection involving varying search window sizes based on tree height (Bunting & Lucas, 2006; Chen et al., 2006, 2007; Jang et al., 2008; Kwak et al., 2010; Persson et al., 2002; Popescu, 2007; Popescu & Wynne, 2004; Zhao et al., 2009). Crown dimensions should be measured in the field to assess the accuracy of local maxima filtering. If such field data are not available, crown dimensions may be obtained through interpretation of image data or a LiDAR-generated

* Corresponding author at: School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan, 689–798, South Korea. Tel.: +82 52 217 2824; Department of Environmental Resources Engineering, State University of New York, College of Environmental Science and Forestry, Syracuse, NY 13210, USA. Tel.: +1 315 470 4709.
E-mail addresses: ersgis@unist.ac.kr, imj@esf.edu (J. Im).

0034-4257/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2012.07.006

canopy surface (Chen et al., 2006; Ke et al., 2010; Wang et al., 2004). Accurately obtaining these crown dimensions is critical as these measurements constitute the parameters from which biomass is derived.

The delineation of tree crowns is itself a robust and growing field of study, and crown delineation methods are explained in greater detail in Gleason and Im (2012). Of particular interest are those methods that use the concept of tree crowns as geometric volumes, sometimes called crown/canopy geometric volume (CGV) (Kato et al., 2009). These methods associate LiDAR heights contained within a crown footprint to determine CGV, and produce more accurate representations of volume when forest density is sparse or trees are isolated (Brandtberg, 2007; Breidenbach et al., 2010; Chen et al., 2007; Kato et al., 2009; Kwak et al., 2010). The CGV can extend downward from the canopy surface to either the forest floor or height thresholds that are either species-specific or LiDAR-derived. The concept of a CGV is intuitive in its correlation to biomass, but has not yet been proven a sufficient metric to act as a sole model predictor variable in dense forests. Dense forest conditions often require multiple LiDAR derived variables to estimate biomass, and the accuracy of crown delineation/biomass estimation in such forests tends to be lower for complex canopy surfaces than homogenous forests (Chen et al., 2007).

An alternative to biomass estimation at the individual tree level is biomass estimation at the plot level. Studies employing this technique frequently characterize LiDAR data through statistical descriptions of canopy height, number of LiDAR returns, and ratios of returns. These descriptors can also be used to estimate other forest biophysical parameters (Anderson et al., 2006; Donoghue et al., 2007; Hawbaker et al., 2010; Hyde et al., 2006; Ioki aet al., 2010; Næsset, 2004; Næsset & Økland, 2002; Popescu et al., 2002, 2004; Solberg et al., 2010). Modeling biomass using this approach requires reference biomass data measured at the plot level, which may introduce bias into the modeling procedure. If reference biomass was calculated including snags, woody debris, and understory vegetation, LiDAR first returns may not penetrate denser canopy to a sufficient degree to accurately describe these features (Næsset, 2005).

There are multiple methods used to estimate biomass/tree volume, which are varied in their assumptions and complexity. Sophisticated regression techniques take into account bias and the correlation of predictor variables (e.g., linear mixed effects regression, geographically weighted regression) rather than the somewhat rigid assumptions of ordinary least squares regression (Hudak et al., 2008; Powell et al., 2010; Salas et al., 2010; Yu et al., 2011; Zhao et al., 2009). When comparing statistical regression models, Salas et al. (2010) found that the linear mixed effects (LME) model significantly outperforms geographically weighted regression, ordinary least squares regression, and generalized least squares regression when estimating tree diameter from LiDAR data. Such studies provide evidence that regressing LiDAR-derived variables with field data is an effective method for estimating biomass, yet there is a large set of assumptions and site-specific considerations that must be made for each study. Zhao et al. (2009) also note that scale issues often affect the performance of biomass estimation regression procedure: i.e., models are built to output biomass at a specific plot size and changing this plot size may affect the accuracy of results. To reduce the effects of regression assumptions on plot scale biomass estimation (population assumptions that do not represent the heterogeneity of forest stands), machine learning techniques such as random forest (RF) and most similar neighbor (MSN) may be used (Breidenbach et al., 2010; Muinonen et al., 2001; Vauhkonen et al., 2010).

This study aims at evaluating machine learning approaches—RF, support vector regression (SVR), Cubist ® regression trees—of forest biomass estimation at both individual tree and plot levels using high posting density airborne LiDAR data. Unlike other biomass estimations that estimate tree diameter or volume from LiDAR data and then calculate biomass from this prediction, this study estimates biomass using field-measured biomass to inform the models. Such a choice allows for combined species modeling, as dbh dependent species-specific allometry is applied *a priori*. This modeling also provides more flexibility and may increase the accuracy of estimating deciduous biomass, which is traditionally more difficult to quantify than coniferous biomass. The objectives of this study are to (1) delineate individual trees from airborne LiDAR data, (2) assess impacts of this delineation on biomass estimation, (3) estimate biomass through four different models, LME, RF, SVR and Cubist, and (4) compare the output of these four models for estimating biomass of all trees at the individual tree and the plot level, paying particular attention to the effects of segregating trees on a coniferous/deciduous divide. The different modeling scenarios for plot/tree combinations are referred to as 'schemes,' and are described fully in the methods section.

2. Study area and reference data

2.1. Study area

The study was conducted within the 1700 ha Heiberg Memorial Forest, located in Tully, NY and managed by State University of New York College of Environmental Science and Forestry (SUNY ESF) (Fig. 1). The College maintains continuous forest inventory (CFI) plots within the forest, and these plots were inventoried in summer of 2010, seasonally coincident with an August 10th, 2010 LiDAR collection. Each CFI plot has a radius of 11 m and is located on a 14 chain grid throughout Heiberg Forest. The plots contain coniferous and deciduous trees common to Upstate New York. Table 1 presents a summary of the CFI plots and the field data which were recorded for each tree with a dbh greater than 9.14 cm.

2.2. LiDAR data

Discrete multiple-return LiDAR data for this study was acquired on August 10th, 2010 from an airborne ALS60 sensor (Table 2). Raw laser data was post-processed using the TerraSolid software suite with manual editing by the vendor (Kucera International, Inc.), and resulted in the creation of a canopy point cloud and a bare earth point cloud. Both point clouds were then converted to raster surface data (cell size 0.25 m) in ArcGIS 9.3 using inverse distance weighted (IDW) interpolation, which is a valid surface creation method for LiDAR data with high point density (Popescu et al., 2002). Point density for our study varied across the study site, with an average point density of 12.7 pts/m^2.

2.3. Reference data

Reference data for this study include ground inventory of tree species, dbh, and height. Tree height data was collected in August of 2010 using a Haglof Vertex III hypsometer, and these data were used to visually examine the LiDAR-derived canopy surface data. Leaf area index (LAI) was measured at the plot level using an ACCUPAR LP80 Ceptometer. Forty LAI measurements per plot were averaged to arrive at the LAI used for biomass estimation.

The field team relied on the allometric equations provided by the USDA Forest Service to provide reference biomass levels (Jenkins et al., 2004). These generalized equations were developed for all deciduous and coniferous species in the United States via a thorough canvas of published allometric equations and previous reviews on the subject conducted in 2003 by the same authors. The Jenkins et al. (2003) equations give two parameters that fit their biomass equation, as well as the number of data points used to generate such an equation and the maximum dbh for which the formulation is applicable. These national scale formulations were adopted for this research because the goal was to investigate a transferable model that requires minimal region-specific information, and because the reported root-mean-squared-errors (RMSEs) of the Jenkins et al. (2003) equations provide evidence that

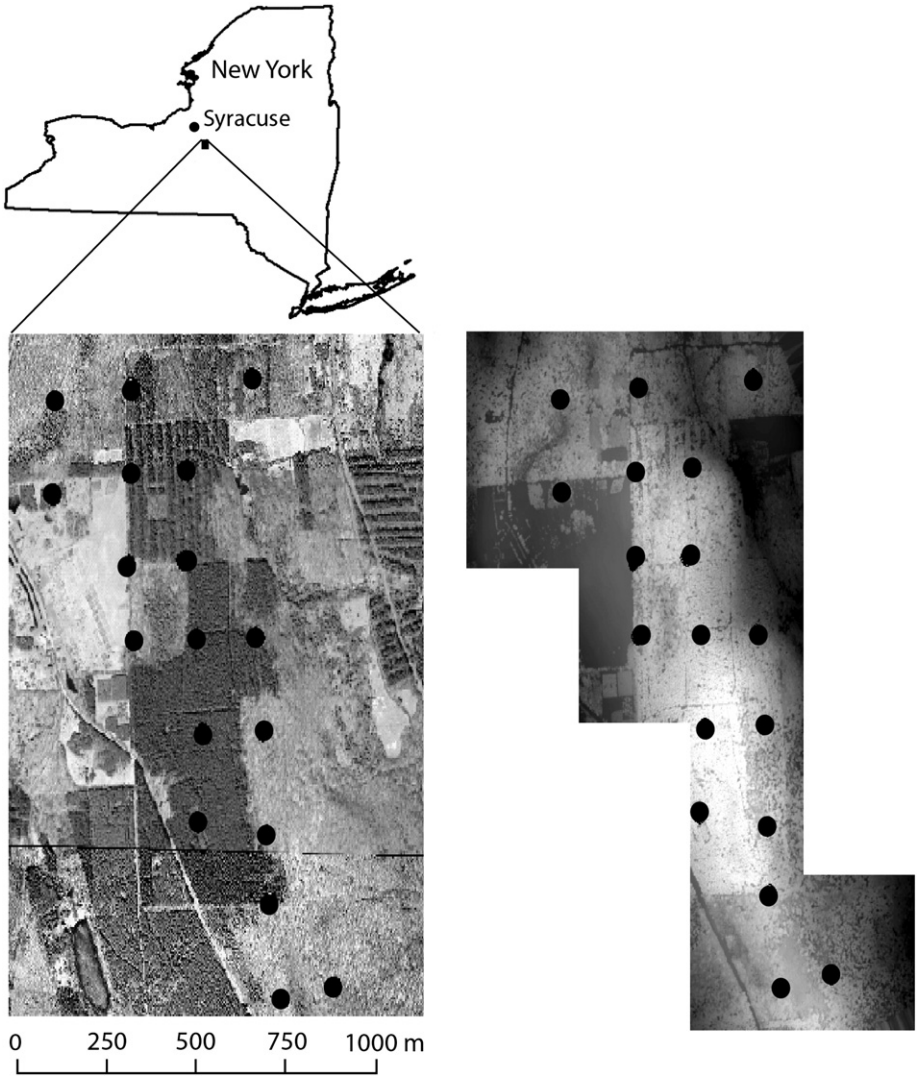

Fig. 1. The study area was located within Heiberg Forest, Tully, NY. The circular CFI plots are maintained by SUNY ESF and are inventoried every 10 years, with the most recent inventory coinciding with the LiDAR collection for this study. The left image is a 2006 aerial orthophoto, New York State GIS Clearinghouse (http://www.nysgis.state.ny.us/) and the right image is the LiDAR-derived canopy surface (i.e., DSM).

the formulations can provide an accurate estimate of biomass. The equations used in Jenkins et al. (2003) have been used to create reference biomass data in other studies that estimate biomass from remotely sensed data (Gonzalez et al., 2010; Popescu, 2007).

Reference crowns were manually delineated from the LiDAR-derived surfaces for each of the plots by three remote sensing scientists with experience in both LiDAR data processing and image interpretation. Each interpreter independently delineated the crowns in each plot and all three crown maps were merged into a final crown surface by mutual agreement. Crown dimensions were difficult to obtain from field data, where tightly closed canopies with low live-crown ratios made discerning the boundaries of each tree crown difficult to determine via either total station or GPS. It is likely that manual delineation from the LiDAR surface resulted in an underestimation of crown size, as crowns viewed from above will exclude the sub-canopy portions of each crown. However, in the absence of field data, such manual delineation is advantageous for the generation of reference crown data (Chen et al., 2006; Ke et al., 2010; Wang et al., 2004). The number of reference crowns did not always coincide with the number of trees recorded in the field, which is the result of two factors. First, the field data reports trees >9.14 cm dbh, and trees in the lower end of that range may not manifest in a meaningful crown area on the LiDAR canopy surface. Second, trees with split trunks or basally sprouting trees produce crowns that are

Table 1
Summary of CFI plots.

	Deciduous	Coniferous
Plots	10	8
Reference crowns	155	225
Mean dbh (cm)	24.9	23.6
Std dev dbh (cm)	4.6	4.2
Mean height (m)	21.65	21.73
Std dev. height (m)	6.3	6.4
Basal area (m^2/ha)	28.7	27.4
Species	Sugar maple, red maple, black cherry, quaking aspen, American beech, white ash, black locust, yellow birch	Norway spruce, scotch pine, red pine, white pine, eastern hemlock, Douglas fir

Table 2
Characteristics of the LiDAR data collection on August 10th, 2010.

Pulse rate	183.8 kHz
FOV	28 deg.
Altitude	487 m
Flying speed	150 knots
Average point density	12.7 pts/m^2

difficult to separate via image interpretation. The latter case does not pose an impediment to biomass estimation, as the crown structure and height for these groups of trees reflects the relative size of each of the component trees. The former case represents those overtopped trees that cannot be identified from the canopy surface. In this case, the trees cannot be located via automated treetop detection, but may be recorded by the LiDAR sensor using multiple returns.

3. Methods

3.1. Overview

Biomass was estimated using four different methods—LME regression, RF regression, SVR, and Cubist regression. For each case, four estimation schemes (comprising two scales) were implemented: (1) plot level biomass estimation using the CFI plot footprint as the basis for generating predictors, (2) plot level estimation using an aggregation of individual trees within a plot to form aggregated plot level variables, (3) estimation at the individual tree level, and (4) estimation at the tree level separately for each phenotype (coniferous or deciduous). These four schemes represent common methods of biomass estimation, and provide additional units of analysis to compare the output of the four estimation models (Gleason & Im, 2011).

To delineate individual trees in schemes 2, 3, and 4, the crown delineation method built from optimized object recognition, treetop detection, and hill climbing (COTH) developed by Gleason and Im (2012) was employed to each study plot, and these delineated crowns formed the basis for generating predictor variables. In scheme 1, the plot boundary was treated as the unit for generating predictor information.

3.2. Tree crown delineation: COTH

Individual tree crowns were delineated using the COTH method (Gleason & Im, 2012), which is a synthesis of genetic algorithm optimized object recognition, treetop identification, and hill climbing to associate crown objects to treetops. The tree height/crown width relationship that determined treetop identification parameters (following Chen et al., 2006) was built from all trees within the study plot. This distinction is important for the transferability of the model, as combining all tree species from the study area builds a more general model of the study site. To mitigate the loss of accuracy that using all plots to generate a width-height relationship introduces, treetop detection was optimized for each plot. To do this, several different levels of statistical confidence, α, were tested for each plot. The width of the prediction interval obtained from a regression (and therefore the treetop search radius) is dependent on α, and setting too high or too low a level of confidence will lead to commission or omission error, respectively. The confidence (α) levels tested for each plot were 0.001, 0.005, 0.01, 0.05, 0.1, 0.15, 0.2, and 0.25. Chen et al. (2006) recommend an α level of 0.1 to reduce commission error, consistent with common statistical practice. The treetop detection provides the basis for tree crown delineation, and tree crown objects are automatically identified using an active contouring model driven by local binary fitting energy (Li et al., 2007). The surface to which this active contouring was applied is a canopy elevation surface generated from the first return points. Lin et al. (2011) used a similar linear interpolation technique as the basis for creating a LiDAR surface to perform active contouring. The parameters in the active contouring model regulate the movement of the contours toward either their interior or their exterior, and a list of such parameters optimized for our study is given in Table 3. The contouring process begins by using initial contours generated from a watershed analysis, which rely on the treetop markers generated via the process described above. This watershed processing is further described in Chen et al. (2006), and results in large crown objects surrounding each treetop. These watershed objects served as initial contours for the active contouring algorithm, which greatly reduced processing time and improved the accuracy of the delineation. The active contouring was optimized by a genetic algorithm (GA) that used the crown area per plot as the fitness function for each evaluation cycle (Gleason & Im, 2012). A GA was chosen because they are easily understood and have been successfully implemented parameter optimization (Fang et al., 2003; Gleason & Im, 2012; Gong et al., 2011; Im et al., 2011; Im et al., 2012a; Tseng et al., 2008).

In the final stage of the COTH process, a hill-climbing algorithm was applied to associate tree crowns with predefined treetops and eliminate spurious crown objects (Ke et al., 2010). The hill climbing is parameterized with a height threshold (here 1 m) that allows the algorithm to overlook most spurious maxima and increases the area of the delineated crowns (Ke et al., 2010). A smaller threshold would constrain the final crowns and a larger one would make crown area approach that of the watershed segmentation method.

3.3. Biomass estimation: variable identification and modeling

The individual trees generated by the COTH process form the basis for biomass estimation in schemes 2, 3, and 4. For each tree, candidate variables derived from height, CGV, and crown area were generated within the delineation footprint. The CGV was calculated as an approximation of the canopy volume per pixel, and uses the COTH delineated footprint and the LiDAR surface therein to form the three-dimensional boundaries of the crown. The bottom of the CGV surface was calculated as the minimum value of the LiDAR surface within the COTH footprint. In other studies, the CGV was calculated by fitting a geometric shape to the point cloud located within a delineated footprint (Breidenbach et al., 2010; Chen et al., 2007; Kato et al., 2009; Kwak et al., 2010). The approach utilized in our study is computationally simpler, and is robust when the LiDAR point density is high as the pixel size representing the LiDAR surface is smaller.

Choosing predictor variables from among these LiDAR derived forest measurements is important to assure the most accurate biomass models. In similar studies, a large list of potential LiDAR-generated covariate candidates were generated, and non-significant variables were eliminated systematically to arrive at the final covariates used for regression (Breidenbach et al., 2010; Hudak et al., 2008; Muinonen et al., 2001; Vauhkonen et al., 2010). In our study, variable reduction was not performed and commonly used predictor variables from previous studies were adopted, as resulting final model variables for most biomass estimations are quite similar regardless of estimation method. The final covariates were generated from the LiDAR surface, and included measures of volume (CGV for 50th, 60th, 70th, and 100th percentiles of height per crown), measures of height (minimum crown height, 70th, and 100th height percentiles per crown), and crown area and diameter. Variables collected at the plot level include LAI and canopy geometric volume (computed as the sum of all individual crowns). These variables are common in biomass estimation studies using LiDAR derived variables, and encompass the final significant variables from Breidenbach et al. (2010) and Vauhkonen et al. (2010) that are possible to calculate from the height surface. Table 4 summarizes the variables used for the four schemes. Biomass was associated with predictor variables wherever a reference tree (and associated biomass) and COTH delineated crown were spatially coincident. In those delineated crowns

Table 3
Five key parameters used in the object recognition algorithm.

Parameter	Effect
λ_1, λ_2	Regulates growth of crown objects
μ	Regulates speed of crown formation
υ	Regulates shape of crown objects
Iterations	Number of iterations

that contained more than one reference tree, biomass was assigned as the sum of the reference trees' biomass. This case represents either reference trees that are overtopped or two reference trees whose crowns were not separated by the COTH process. These LiDAR derived and field measured variables were then fed into the four models for comparison: LME, RF, SVR, and Cubist.

3.3.1. Linear mixed effects (LME) regression

LME modeling was performed in the R statistical software package, and accomplished using the *nlme* add-on package (Pinheiro et al., 2008). LME models are a form of linear regression in which coefficient parameters are the sum of fixed and random effects. In this case, the random effects were allowed to vary spatially, and were parameterized by the user (Pinheiro et al., 2008). In addition, the overall error term is also allowed to vary spatially, which improves prediction between plots. Random error parameterization was characterized per plot for individual tree biomass estimation, which operates under the assumption that each plot in the study site will have a different random error structure and that trees within these plots are not independent of one another (Salas et al., 2010). In effect, this distinction allows the random effects coefficient to vary per tree and per plot.

Salas et al. (2010) have shown that LME models significantly outperform other regression techniques when modeling forest volume, yet this method is not always applicable. LME modeling is not appropriate for variables generated at the plot level, specifically because there are no well-defined groups within which the random effects structure can vary. There are only 18 plots available for this study, and such a sample size does not guarantee statistically powerful results for straightforward regression techniques. For these reasons, LME model results were not generated at the plot level, and only the results modeling individual trees are reported.

Table 4
Summary of the variables used.

Plot level (Schemes 1 and 2)		Tree level (Schemes 3 and 4)	
Designation	Variable	Designation	Variable
VOL	Volume of the canopy within the plot	VOL	Volume of the canopy within the plot
MXCH	Maximum height of the canopy within the plot	CGV	Crown geometric volume (CGV)
MD	Maximum distance from plot center (Scheme 1) Mean of individual crown diameters (Scheme 2)	TH	Tree height
SCA	Area of the plot (Scheme 1) Sum of individual crown areas (Scheme 2)	CD	Crown diameter
CGV50	Crown geometric volume for 50th height percentile of all canopy heights	CA	Crown area
CGV60	Crown geometric volume for 60th height percentile of all canopy heights	CGV50C	CGV for 50th percentile of height per crown
CGV70	Crown geometric volume for 70th height percentile of all canopy heights	CGV60C	CGV for 60th percentile of height per crown
CH70	70th height percentile of all canopy heights	CGV70C	CGV for 70th percentile of height per crown
MNCH	Minimum canopy height	TH70	70th height percentile per crown
LAI	Leaf-area-index (LAI)	MNTH	Minimum crown height
		LAI	Leaf-area-index (LAI)

3.3.2. Random forest (RF)

RF models have been shown to reduce bias and overfitting, work well with random inputs, and in some cases tend to be more accurate than simple regression techniques for biomass estimation (Breiman, 2001; Powell et al., 2010) and have been previously employed in forestry for modeling quantities other than biomass (Falkowski et al., 2009; Hudak et al., 2008; Korpela et al., 2009). When applied to biomass estimation, RF models may be driven by inputs that have been commonly used in regression studies (Breidenbach et al., 2010; Vauhkonen et al., 2010; Yu et al., 2011). RF models have roots in classification and regression trees (CART), which are a series of binary rule-based decisions that dictate how an input is related to its predictor variable. If the error associated with splitting a single rule into multiple rules is lower than using a single rule, the regression tree will 'branch out' and the tree will grow more rules. The decision tree stops growing when the minimum error versus the input data is obtained. The major advantage of such trees is their flexibility as regression trees can accurately describe complex relationships between variables at multiple scales and multiple predictor strengths (Walton, 2008). An aggregation of such trees usually leads to more accurate solutions, and the mechanics of the estimation can be found in Breiman (2001). There is also a heuristic component to RF as the initial sample for each decision tree is chosen randomly, which may lead to different results each time the model is run (Walton, 2008).

The RF model fit for this study was implemented in the *R* statistical software, using the *randomForest* add-on package based on the work of Breiman and Cutler (Liaw & Wiener, 2002). The importance of each node of in RF regression trees is determined by using input data to assess which variable at that node best characterizes the remaining observations (Breidenbach et al., 2010). For the present study, the predictor variables were predetermined rather than discovered using the RF method's ability to determine variable importance. The RF model employed in this study also calculated error and proximity on out-of-bag data. In an out-of-bag method, some of the training data are excluded for each regression tree generation, and the errors for these data can be used to inform the RF of the relative strength and correlation of that tree and ultimately guide the development of the RF (Breiman, 2001).

3.3.3. Support vector regression (SVR)

Support vector machines (SVMs) are machine learning models that have been widely used in remote sensing, and a thorough review of recent applications can be found in Mountrakis, Im, and Ogole (2011). In forestry, SVMs have been successfully applied as a classification tool, providing accurate classification results for a variety of applications (Huang et al., 2008; Kavzoglu & Colkesen, 2009; Kuemmerle et al., 2009). One application of SVMs is their use in regression, although studies using support vector regression (SVR) are not as common in forest biomass estimation. Durbha et al. (2007) successfully implemented SVR to estimate LAI from spectroradiometer data over large agricultural areas, but such data is uncommon in forest biomass studies. As a spatial analog, Ballabio (2009) predicted soil properties in mountainous regions using SVR, showing its ability to detect spatial patterns over varying terrain. SVR holds promise for estimating biomass from airborne LiDAR data, and was thus explored in the present study.

At a conceptual level, SVR operates by assuming that each set of input parameters will have a unique relation to its response variable, and that the grouping and relation of these predictors to one another is sufficient to identify rules that can be used for predicting biomass from a set of inputs. SVR accomplishes this by separating data among multidimensional hyperplanes, assuming that input data are separable in feature space (Mountrakis et al., 2011). Hyperplanes are features that exist in feature space, which is a multidimensional space built from axes that represent each input variable. Each response variable can be located in such space by 'plotting' it according to its predictor variable values, and similar responses are expected to

be spatially grouped in the feature space. The SVR builds hyperplanes to separate these groups, and uses "support vectors" to assign each observation to an appropriately segmented space. The SVR iteratively assigns these hyperplanes amongst the data and then adjusts them based on the errors associated with each, arriving at a solution that converges toward optimal. SVR methods, like RF methods, have been shown to reduce overfitting, which reduces the ability of a model to accurately describe new, unseen data (Mountrakis et al., 2011).

The SVR was implemented using a standalone tool that employed a grid search parameter optimization for a radial basis SVR (Lu et al., in review). The radial basis kernel function was chosen because it has been shown to be effective for forest parameter estimation and has few parameters to define (Huang et al., 2008; Kavzoglu & Colkesen, 2009; Kuemmerle et al., 2009). Grid search is a classic technique for SVR parameter optimization that relies on gradient descent convergence (Chapelle et al., 2002; Huang & Wang, 2006; Keerthi & Lin, 2003; Shevade et al., 2000) and was necessary to achieve meaningful models.

3.3.4. Cubist

Cubist ® (RuleQuest Research, Inc.) is commercially available rule-based software that has been used for mapping and regression studies. Because it is a commercial product, the exact nature of its algorithm is unknown but has been successfully implemented in remote sensing applications. Walton (2008) has shown that Cubist has shorter run times than RF and SVR methods for the same input data, and produces comparable results in a direct comparison for subpixel urban land cover classification. Further discussion of the modeling mechanisms found in Cubist can be found in Walton (2008). Cubist has also been shown to be a viable method for regression, and can be applied to a variety of spatial problems (Im et al., 2009, 2012b; Walker et al., 2007; Yu & Wu, 2006). Blackard et al. (2008) employed Cubist to estimate biomass, and reported moderate to high correlation (correlation reported between 0.39 and 0.92) between reference and modeled data at a national scale. Similarly, Chen et al. (2011) modeled aboveground forest biomass carbon using a Cubist model and reported similar success. Cubist presents succinct rules for modeling given the predictor variables, and discards inputs that it does not consider significant.

It is difficult to identify the exact methods used by Cubist to determine the rules generated to perform regression, but it can be assumed that it is based on regression trees performed with an instance-based model (Walton, 2008). Cubist adopts a modified regression tree system, which creates rule-based predictive multivariate regression models from training data. This differs from typical classification and regression trees (known as CART and embedded in the RF model) in that they contain only a single value at the end of each branch. Cubist allows multivariate linear models to overlap, not requiring them to be mutually exclusive. Thus, Cubist predictions from all trees are averaged to arrive at a final prediction.

3.4. Accuracy assessment

Accuracy assessment for this study is separated into two different metrics for each of the four estimation schemes: training (i.e., calibration) RMSE and leave-one-out cross validation RMSE. Cross validation accuracy is reported in biomass estimation studies when models have few data or the sensitivity of the model to the training data may affect the results (Breidenbach et al., 2010; Ioki et al., 2010; Popescu et al., 2004; Vauhkonen et al., 2010). For the RF and SVR methods, RMSE values are given as the mean of 20 different models (i.e. 20 RFs and 20 SVRs) to control for the variability of these models. The standard deviations of these simulations are also given.

Ideally, data would be portioned into testing and training datasets for each of the four schemes, but this was not possible in this study for several reasons. First, there are 18 plots available for the plot level estimation, and further reducing this number by splitting the data would limit the reliability of the models. Secondly, after associating reference biomass with delineated crowns, there were 136 crown-biomass pairs. While this is a sufficient amount of data for generating a testing/training partition, doing so may have biased the training data based on the coniferous/deciduous divide if a random division did not contain equal representations of each plot. A manual supervision of the division could take this into account, but assuring an even division of tree type may result in spatial bias. This is especially an issue for the machine learning methods, which generally produce more accurate results with larger data volumes. To compare biomass estimation accuracies between plot level schemes and individual tree schemes, the RMSE of each process was compared to the mean biomass for each scheme, resulting in a relative estimate of error.

Crown delineation accuracy is given as two-class classification areal accuracy (crown and non-crown) reasonable for moderately dense forests, and treetop detection accuracy (% identified), consistent with Gleason and Im (2012).

4. Results

Tree crown delineation results using the COTH method are summarized in Fig. 2. The overall accuracy, on average, was 63.6% and tree crown delineations for deciduous plots produced marginally higher accuracy than those for coniferous plots (Fig. 2). Fig. 3 depicts the COTH-delineated crowns for a Norway spruce dominated plot. Confusion matrices for the representative coniferous and deciduous plots with the highest overall accuracy are summarized in Table 5.

Schemes 1 and 2 are plot estimation methods, where biomass is assigned as the sum of individual tree biomass within the delineation footprint. SVR produced the most accurate model for Scheme 1 as evidenced by the ratio of RMSE to mean input biomass (RRMSE). For Scheme 2, RF and Cubist performance remained roughly constant, while SVR showed a significant improvement from 1509 kg RMSE to 671 kg RMSE (a decrease in RRMSE form 18.1% to 13.6%). These SVR models were also stable, with a standard deviation of the 20 models of 109.3 kg and 92.6 kg for schemes 1 and 2, respectively.

By comparison, all models performed markedly worse when estimating biomass at the individual tree level, as evidenced by RRMSE values ranging from 68.1% to 119.6% for schemes 3 and 4 as compared to a range of 13.6% to 34.2% for Schemes 1 and 2 (Table 6). SVR produced the most accurate estimations of biomass for Schemes 3 and 4 with standard deviation of RMSE of 8.36 kg, 24.1 kg, and 10.4 kg for schemes 3, 4c, and 4d. LME, RF, and Cubist performed similarly for scheme 3, but markedly worse than SVR for this scheme. LME modeled biomass poorest when trees were separated by phenotype (scheme 4), and produced the unexpected result of modeling deciduous trees more accurately than coniferous trees. Fig. 4 illustrates the relationship between predicted biomass and field biomass for each estimation method and shows the variety of predictions across schemes and models. RF, SVR, and Cubist calibration and cross validation results were relatively consistent for all schemes while LME results were less stable, especially for scheme 4.

The RF and Cubist models used provide metrics for importance of the variables used in building each model. Fig. 5 illustrates the increase in mean squared error (MSE) in percentage when a variable is held out-of-bag for schemes 2–4 RF. In scheme 2, volume and CGV percentiles were the most important variables while in scheme 3 tree height was the most important. LAI, which was field calculated and not LiDAR derived, was not important for either scheme. Scheme 4 for coniferous trees resulted in a variable importance pattern similar to scheme 3. In scheme 4 for deciduous trees, crown area was very important along with volume and CGV percentiles (50–70), while CGV, crown diameter, and 70th height percentile were least important. The scheme 2 Cubist model only used volume and 70th height percentile and the scheme 4 Cubist model for deciduous trees used

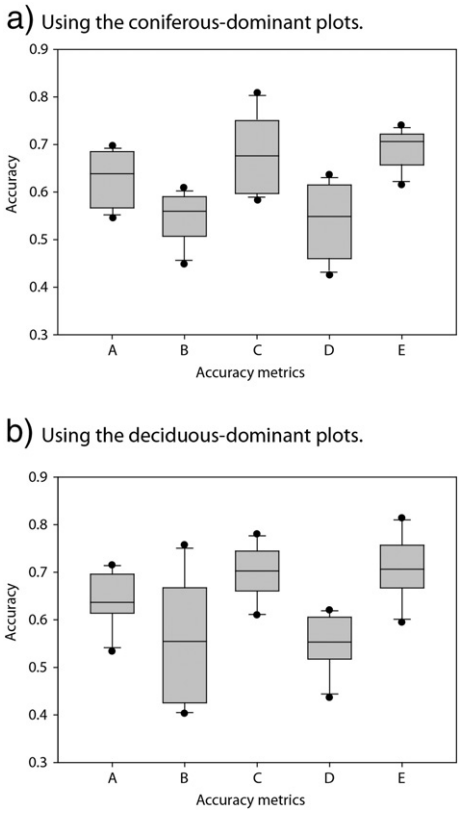

Fig. 2. Box plots of the accuracy metrics based on the COTH results for the 18 plots. The *x*-axis represents overall accuracy (A), user's accuracy for crown (B), user's accuracy for non-crown (C), producer's accuracy for crown (D), and producer's accuracy for non-crown (E).

CGV (not shown). In the scheme 3 Cubist, crown area, CGV 50th percentile, and maximum tree height were useful. There are not many overlaps in the important variables between RF and Cubist models. However, volume appeared an important variable for both RF and Cubist in scheme 2, while tree height was the common important variable in scheme 3.

To show the implications of grouping trees together regardless of phenotype, which increases the flexibility and transferability of the modeling process, Table 7 compares how accurately coniferous and deciduous trees were modeled in scheme 3 and 4. The comparison presented is the difference in model accuracy of trees modeled with scheme 3 and with scheme 4. Scheme 3 trees were separated by phenotype *a posteriori*. RF and Cubist behaved as conjectured, where the deciduous trees were modeled more accurately when included in scheme 3 and the coniferous trees slightly less accurately. However, these changes in accuracy are miniscule when compared to the overall error of schemes 3 and 4, as the maximum gain/loss in accuracy is less than 3% of the RMSE of scheme 3 for either method. SVR showed that coniferous trees were modeled more accurately when included in scheme 3, but the change in accuracy is contained within the standard deviation of the models. LME modeled both coniferous and deciduous trees less accurately within scheme 3, and modeled coniferous trees substantially worse when using that scheme.

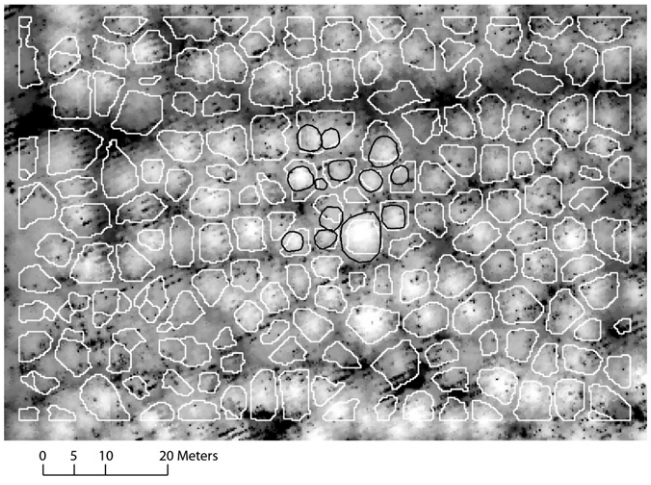

Fig. 3. COTH results for study plot 118, a Norway spruce dominated plot.

Table 5
Summary of confusion matrices of tree crown delineation results for two representative plots.

Plot	Overall	User's accuracy		Producer's accuracy	
		crown	non-crown	crown	non-crown
Coniferous dominant plot	69.1%	50.0%	81.4%	63.6%	71.6%
Deciduous dominant plot	71.5%	75.7%	68.7%	61.1%	81.4%

To assess the impact of tree crown delineation on biomass estimation accuracy, the exact same modeling procedures described above were employed using the manually delineated reference tree crowns to generate estimation variables. Results of this process are given in Table 8, and Fig. 6 shows the differences across schemes and models. For all schemes, LME models built from reference crowns produced superior RRMSE, which was logically expected. However, the other three modeling techniques indicate that switching from COTH to reference crowns does not always produce superior results. Reference crown models were less accurate for RF and Cubist schemes 2, 3, and 4c and for SVR scheme 2. This suggests the non-sequitor that COTH delineated trees are better predictors of biomass than reference crowns, an idea that will be subsequently discussed. SVR results, as in COTH built models, were the most accurate for reference crown derived biomass models.

5. Discussion

5.1. Tree crown delineation and its impact on biomass estimation

Compared to a previous study employing the COTH method (Gleason & Im, 2012), tree crown delineation was not as accurate. This decreased crown delineation accuracy is due to the study conditions, as the previous study employing the COTH method was focused on a single Norway spruce plantation that is simpler to delineate. Despite this decrease in accuracy, the goal of the COTH method was to add flexibility and transferability to crown delineation and biomass modeling, which was achieved. In addition, the overall accuracy of the COTH process was acceptable for biomass modeling, as evidenced by the estimation results achieved in this study. Similar crown delineation studies reported accuracies of 69–81% for treetop detection accuracy in mixed mountainous forest (Lin et al., 2011), and 65–85% for areal agreement of tree crowns with reference data in a Norway Spruce plantation (Ke et al., 2010). Chen et al. (2006) report delineation

Table 6
Cross validation RMSE for each scheme in kg. Numbers in parentheses indicate RMSE relative to mean biomass of input data biomass (%).

Scheme	LME	RF	SVR	Cubist
Scheme 1	NA*	2712 (32.4)	1509 (18.1)	2602 (31.1)
Scheme 2	NA	1593 (32.4)	671 (13.6)	1682 (34.2)
Scheme 3	506 (82.8)	505 (82.7)	457 (74.8)	502 (82.2)
Scheme 4 Coniferous	724 (119.6)	428 (70.8)	412 (68.1)	426 (70.4)
Scheme 4 Deciduous	634 (102.7)	588 (95.2)	508 (82.2)	593 (96.0)

* NA indicates that the model was inappropriate for this scheme.

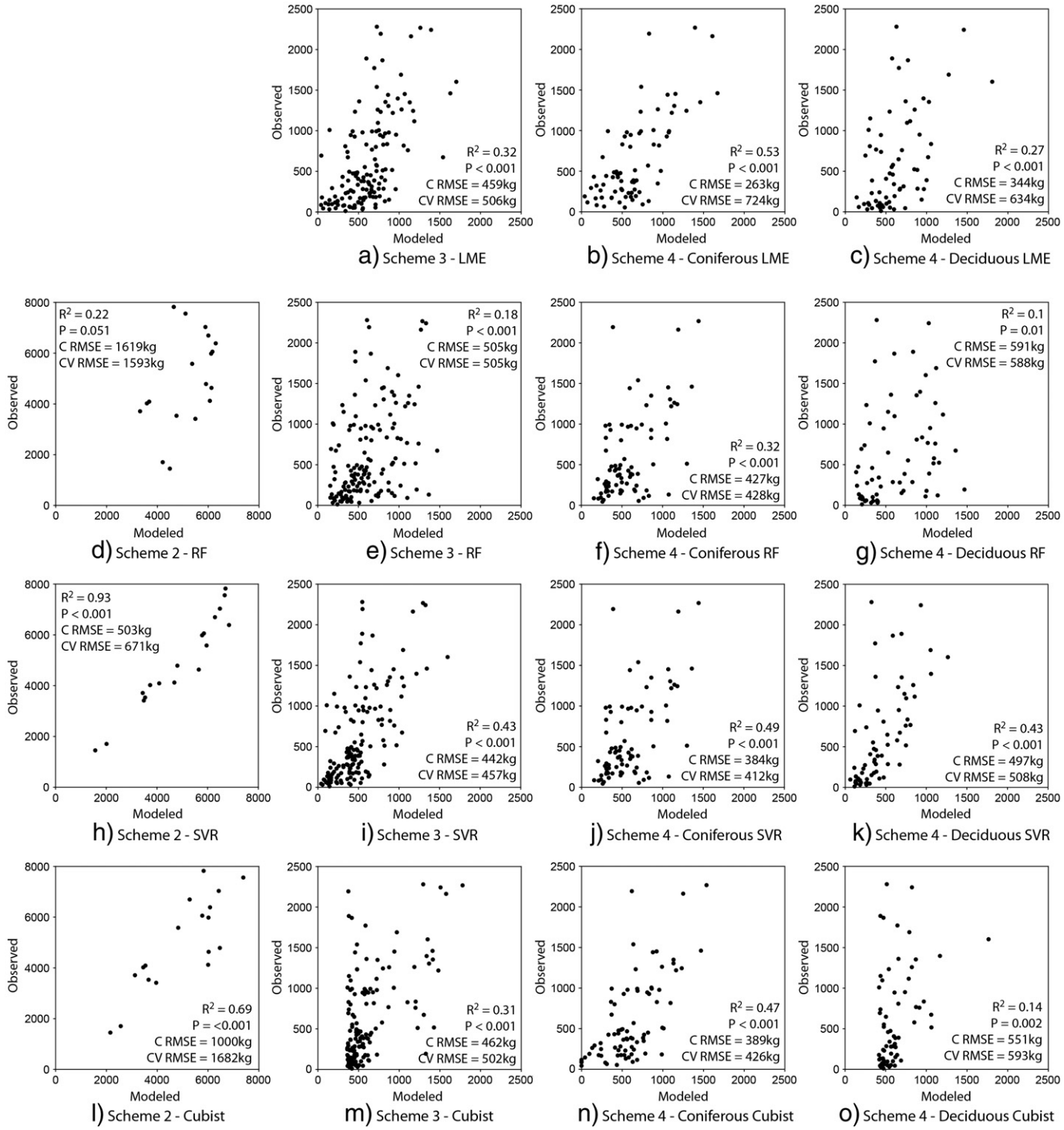

Fig. 4. Predicted values vs. observed values from calibration for schemes 2–4. C RMSE and CV RMSE represent calibration RMSE and cross validation RMSE, respectively.

accuracy using their AATI statistic, which is a ratio of the area tree crowns that spatially coincide (±10%) with reference crowns to total crown area. Their delineation method was found to be 56–64% accurate in savanna woodland using this stringent metric.

The implications of tree crown delineation accuracy on biomass estimation accuracy are shown in Table 8 and Fig. 6.These reference crown results further reinforce the idea that SVR, RF, and Cubist were the most suitable techniques for biomass estimation. SVR produced the most accurate biomass models for both COTH and reference crown models, and RF and Cubist produce models with similar accuracy for both sets of crown data. The exact relation between delineation accuracy and biomass estimation accuracy cannot be assessed within this study, but the results (Fig. 6) present evidence that this relationship is not as straightforward as may be imagined. This is supported by the results obtained from scheme 1, which are not based on delineated crowns. Scheme 1 produced results that are superior to scheme 2 using reference crowns for all models, yet produced equivalent results for RF and Cubist and inferior results for SVR for scheme 2 using COTH crowns. A possible explanation for these results is that COTH delineated crowns are nearly always larger than reference crowns (refer to Fig. 3). These larger crowns gather additional stand information beyond the individual tree, but less

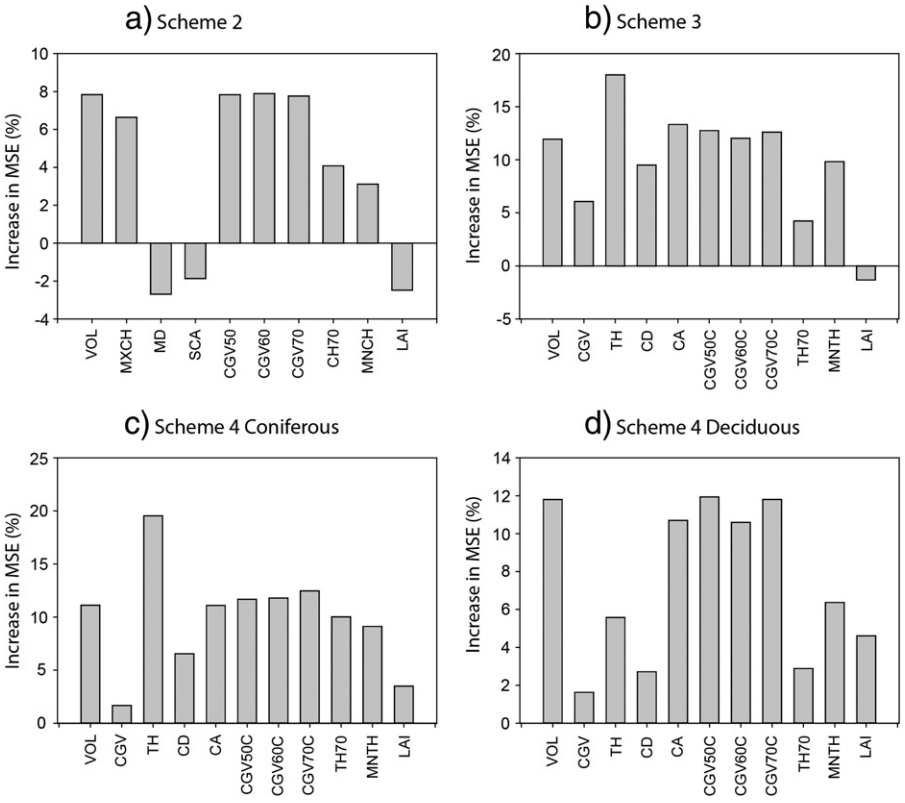

Fig. 5. Importance of the variables used in RF for schemes 2, 3, and 4. The *x*-axis for each plot represents the variables used (refer to Table 4). Increase in MSE (%) was used for the schemes 3 and 4 when the variable was held out-of-bag.

than the entire plot footprint as in scheme 1. This additional forest information may be important for aggregating trees to the plot level, and could explain the differences in the scheme 2 results between reference and COTH crowns.

The varied response of schemes 3 and 4 to reference crown input is another indication of the complex relationship between crown delineation and biomass estimation accuracy. RF and Cubist schemes 3 and 4c were more accurate with COTH crowns, while schemes 2 and 4d for these models were more accurate when using reference crowns. These differences could have arisen from the overlap between COTH and reference crowns with field-located tree positions (which could further conflate the large-crown issue). The accuracy of field recorded tree locations is important for biomass estimation, as crowns were associated with biomass via spatial overlap. Positional error stemming from difficulties in using GPS units underneath dense forest canopies could lead to a mismatch between a delineated crown and its corresponding biomass value that is virtually undetectable. This is a larger issue for schemes 3 and 4 than for scheme 2, but will also cause both commission and omission error when aggregating trees to the plot level.

The effects of crown delineation on scheme 3 were minimal for models other than SVR (Fig. 6), which may be seen as example of the feasibility of combined phenotypic modeling in conjunction with the flexibility of COTH. Scheme 4 results were more affected by COTH accuracy. One possible explanation for this is that coniferous trees were generally found in denser stands. This leads to the possibility of more conglomerated (large) crowns containing multiple biomass points, which could be more easily modeled with COTH, as discussed earlier. In addition, reference crowns were difficult to generate for the coniferous plots located in denser stands, which could have also influenced the results. Scheme 4d is the only scheme in which the reference crowns always produced superior model accuracies, underscoring the difficulties and importance of accurately obtaining crown dimensions for these trees.

5.2. Modeling performance by method and scheme

Tables 6 and 8 show that schemes 1 and 2 do not produce equivalent estimation accuracies, which is consistent with the findings of Zhao et al., 2009. While scheme 2 is certainly affected by crown delineation accuracy, scheme 1 uses no crown information and is therefore a more stable estimation technique. However, as Zhao et al. (2009) have noted, building biomass estimation models around specific plot scales introduces scaling bias, and the somewhat arbitrary choice of plot size can exhibit strong control over the accuracy of the results. Following this argument, it is recommended that plot level analysis be conducted using scheme 2, despite sensitivity to crown delineation. This scheme is less affected by plot size assumptions, and coincides

Table 7
Differences between Schemes 3 and 4 in kg cross validation RMSE. Positive values indicate an increase in error when using Scheme 3 (combined modeling).

	LME	RF	SVR	Cubist
Coniferous	300	21	−8	13
Deciduous	45	−22	10	−1

Table 8
Cross validation RMSE for each scheme in kg using the reference crowns. Scheme 1 is not shown, as it is not based on crown delineation and is unaffected by crown inputs. Numbers in parenthesis indicate RMSE relative to input data mean biomass (%).

Scheme	LME	RF	SVR	CUBIST
Scheme 2	NA[a]	3175 (47.7)	1369 (20.6)	2960 (44.5)
Scheme 3	319 (80.2)	332 (83.5)	230 (57.8)	362 (91.0)
Scheme 4 Coniferous	311 (94.0)	267 (80.7)	211 (63.8)	270 (81.6)
Scheme 4 Deciduous	397 (80.7)	405 (82.3)	366 (74.4)	410 (83.4)

[a] NA indicates that the model was inappropriate for this scheme.

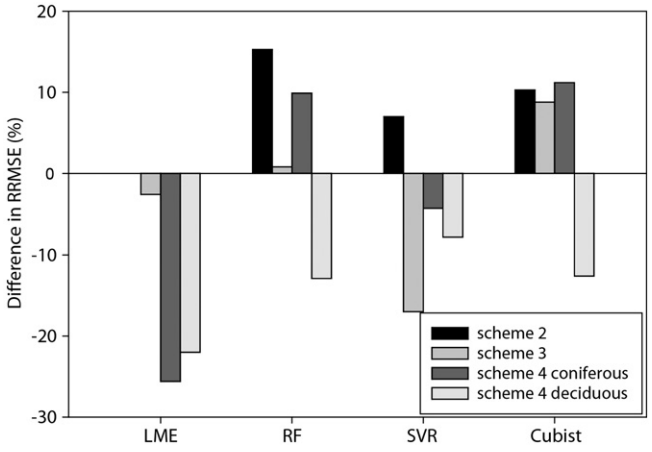

Fig. 6. Difference in RRMSE (%) between biomass modeled using COTH delineated crowns and using reference crowns. Negative values indicate an increase in error using the COTH crowns.

with the 'upscaling' approach described in Breidenbach et al. (2010). As Table 6 shows, RF and Cubist models provide similar results between schemes 1 and 2. SVR scheme 2 using COTH crowns was the most accurate model for any configuration, further affirming scheme 2 as a choice for future modeling.

Schemes 3 and 4 represent methods of estimating biomass that are most closely related to the unit of forest biomass; i.e., the individual tree. As Chen et al. (2007) note, assessing estimation accuracy at this level is not common, with studies often choosing to report accuracy using scheme 2. The RRMSE of schemes 2 is far less than that of schemes 3 and 4, yet the approaches are based upon the same fundamental unit. In schemes 3 and 4, data are more susceptible to outliers and to estimation bias, whereas the averaging and summation processes that aggregated this data into scheme 2 eliminate some of these concerns. As with scheme 2, positional accuracy of tree locations to record biomass important in assuring that crowns and biomass are associated correctly.

Combined modeling affected each of the models differently (Fig. 4). RF and Cubist performed similarly in schemes 3 and 4, showing changes in RRMSE less than 10% between these schemes and showing minimal changes for each phenotype within scheme 3 (Table 7). This suggests that combined modeling is feasible for forests as heterogeneous as those in this study, as there is little change in how accurately either phenotype was modeled between these two schemes using RF and Cubist. This is not the case for LME modeling, where both phenotypes were modeled less accurately in scheme 3 (Table 7), especially coniferous trees. This is due to the parametric nature of LME, which assumes population means and standard deviations to be constant, which is not the case when modeling both forest types together despite controlling for plot to plot error.

Breidenbach et al. (2010) also found RF estimation to produce accurate models for biomass estimation, reporting plot level RRMSE values of 21.62% and 96.55% for coniferous and deciduous plots, respectively. The accuracy of the biomass estimation results achieved in this study (Table 6) is consistent with other published results estimating forest biomass, especially scheme 2 results (Bortolot & Wynne, 2005; Hawbaker et al., 2010; Ioki et al., 2010; Popescu, 2007; Popescu et al., 2004; Vauhkonen et al., 2010). These other studies estimating biomass or volume from airborne LiDAR have reported varying levels of accuracy, including RMSE values from 20% to 500% for both plot and tree level analysis. Such variety is a product of the field conditions on which the estimation was performed; studies conducted in homogenous coniferous stands reported much higher accuracies than those conducted in mixed or deciduous stands.

5.3. Transferability and efficacy of models

RF, SVR, and Cubist models afford a measure of transferability to the study that is absent in regression-based estimates of biomass. The main transferability of the model is obtained from the crown delineation process using COTH, in which all parameters are chosen separately by the GA for each plot. As discussed previously, the biomass estimation process is pseudo- transferrable, meaning that while the methods used are flexible and able to adapt to a variety of forest conditions, it is likely that the rules that are generated by each of the four methods of estimating biomass would not apply to another study site (Foody et al., 2003). However, it is expected that site conditions similar to the ones studied here (i.e., protected, mixed forest in the northeastern United States) would produce quite similar model rules.

The RF and SVR models contain a stochastic element that results in a different biomass model each time they are applied to the same data. The values shown in Table 6 represent the mean of 20 different cross validation RMSEs for each of these models. The standard deviation of these errors was small for RF, ranging from 1.39 kg (scheme 3) to 16.64 kg (scheme 1). These results indicate that the RF models built from the delineated tree crowns are quite stable. The SVR models produced substantially higher RMSE standard deviations than RF, ranging from 8.6 kg (scheme 3) to 109 kg (scheme 1) yet these deviations are all less than 7% of the mean RMSE for each scheme.

The gains in transferability made by modeling trees using these methods are reinforced by their relative accuracies in modeling each tree phenotype. In all three of the flexible models, there were moderate changes in either deciduous or coniferous estimation accuracy for combined modeling (Table 7), suggesting that these models are indeed flexible and are ideal for such modeling. LME modeling, however, remains dependent on particular grouping and parameterization structures, rendering it less transferrable than the other models.

Ultimately, the accuracy of a biomass estimation study is important only as it compares to traditional field estimation made by skilled foresters. Foresters generally develop 'prism plots' to estimate biomass in the field: a technique in which measurements of basal area on trees selected by an optical prism are used to compute biomass and the number of prism plots are determined by stand size, forest heterogeneity, and purpose of the estimation (Ellis, 2011). Desired accuracy for such a plot entails having between 10% and 15% error for biomass estimations. This study only achieves such accuracy when SVR is used with scheme 2. The reported error for this scheme assumes that the allometric equations employed are accurate representations of field biomass, which is likely not the case in this study. The methodologies represented here remain powerful tools for biomass estimation, but only become useful to foresters when fit to allometric equations that are developed specifically for the study forest or fit to destructively sampled biomass. For example, even an allometric equation developed for a particular region may have an error of 10–15% in itself, which would compound with the 13–35% error achieved in the remotely sensed biomass estimation. As such, researchers hoping to develop efficient and accurate biomass estimation models from remotely sensed data should pay particular attention to on-site allometric relationships. Regional scale allometry is suitable for studies attempting to quantify biomass across large areas (e.g. the UN REDD project), yet for local estimations thorough fieldwork is the only means to assure that biomass model accuracy reflects knowledge of forest biomass.

6. Conclusion

This study compared the effectiveness of four recent modeling techniques for estimating biomass in moderately dense forest conditions. Biomass was estimated using four schemes representing different plot

and tree level configurations of the forest, and individual trees were delineated using the COTH algorithm (Gleason & Im, 2012). Results indicate that using SVR provided the most accurate results for all estimation schemes. In addition, the SVR models were quite stable, although less so than RF models. LME modeling produced coherent biomass results that were only competitive with the other models for scheme 3. Given that foresters estimate biomass at the plot level as an aggregation of trees and given the accuracy of scheme 2, it becomes the recommended modeling scheme to emerge from this study. RF and Cubist were remarkably consistent in accuracy and response to crown inputs, showing the same trends in accuracy for phenotype separation and reference crown modeling as well as having a maximum in-scheme RRMSE difference of 7.5%. In choosing between RF and Cubist, one must remember that RF may be implemented through use of transparent freeware, while Cubist is a 'black box' commercial product. RF is a more widely available model, and produced very stable estimates of biomass for this study. SVR produced the most accurate results in this study, yet some obstacles remain in its widespread adoption. While open source SVR models may be commonly found, they have a steep learning curve and are less transparent than the other open source models.

Another goal of this study was transferability: the ability of the model to be flexible and work in a variety of field conditions. This transferability was achieved by modeling biomass directly and by testing the accuracy of combining the reference tree phenotypes. Results indicate that for the three most accurate models, combining coniferous and deciduous trees is a viable method of biomass estimation. This study also found that the accuracy of the crown delineation used to generate predictor variables affects biomass estimation differently depending on the scheme that is employed. Further work is needed to assess the relationship between crown delineation accuracy and biomass estimation for both individual tree and plot level estimates of biomass.

References

Anderson, J., Martin, M. E., Smith, M., Dubayah, R. O., Hofton, M. A., Hyde, P., et al. (2006). The use of waveform lidar to measure northern temperate mixed conifer and deciduous forest structure in New Hampshire. *Remote Sensing of Environment*, *105*, 248–261.

Ballabio, C. (2009). Spatial prediction of soil properties in temperate mountain regions using support vector regression. *Geoderma*, *151*, 338–350.

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.

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

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

Breidenbach, J., Næsset, E., Lien, V., Gobakken, T., & Solberg, S. (2010). Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment*, *114*, 911–924.

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

Bunting, P., & Lucas, R. (2006). The delineation of tree crowns in Australian mixed species forests using hyperspectral Compact Airborne Spectrographic Imager (CASI) data. *Remote Sensing of Environment*, *101*, 230–248.

Chapelle, O., Vapnik, V., Bousquet, O., & Mukherjee, S. (2002). Choosing multiple parameters for support vector machines. *Machine Learning*, *46*, 131–159.

Chen, Q., Gong, P., Baldocchi, D., & Tian, Y. Q. (2007). Estimating basal area and stem volume for individual trees from lidar data. *Photogrammetric Engineering and Remote Sensing*, *73*, 1355–1365.

Chen, Q., Baldocchi, D., Gong, P., & Kelly, M. (2006). Isolating individual trees in a savanna woodland using small footprint lidar data. *Photogrammetric Engineering and Remote Sensing*, *72*, 923–932.

Chen, X., Liu, S., Zhu, Z., Vogelmann, J., Li, Z., & Ohlen, D. (2011). Estimating aboveground forest biomass carbon and fire consumption in the U.S. Utah High Plateaus using data from the Forest Inventory and Analysis Program, Landsat, and LANDFIRE. *Ecological Indicators*, *11*, 140–148.

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

Durbha, S. S., King, R. L., & Younan, N. H. (2007). Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sensing of Environment*, *107*, 348–361.

Ellis, Brian. (2011). Research Assistant, *SUNY College of Environmental Science and Forestry, Department of Forest and Natural Resource Management*. Personal communication. April 2011.

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

Fang, H., Liang, S., & Kuusk, A. (2003). Retrieving leaf area index using a genetic algorithm with a canopy radiative transfer model. *Remote Sensing of Environment*, *85*, 257–270.

Foody, G. M., Boyd, D. S., & Cutler, M. E. J. (2003). Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sensing of Environment*, *85*, 463–474.

Gleason, C. J., & Im, J. (2011). A review of remote sensing of forest biomass and biofuel: options for small scale forests. *GIScience and Remote Sensing*, *48*, 141–170.

Gleason, C. J., & Im, J. (2012). A fusion approach for tree crown delineation from LiDAR data. *Photogrammetric Engineering and Remote Sensing*, *78*, 679–692.

Gong, B., Im, J., & Mountrakis, G. (2011). An artificial immune network approach to multi-sensor land use/land cover classification. *Remote Sensing of Environment*, *115*, 600–614.

Gonzalez, P., Asner, G. P., Battles, J. J., Lefsky, M. A., Waring, K. M., & Palace, M. (2010). Forest carbon densities and uncertainties from Lidar, QuickBird, and field measurements in California. *Remote Sensing of Environment*, *114*, 1561–1575.

Hawbaker, T. J., Gobakken, T., Lesak, A., Trømborg, E., Contrucci, K., & Radeloff, V. (2010). Light detection and ranging-based measures of mixed hardwood forest structure. *Forest Science*, *56*, 313–326.

Huang, C. L., & Wang, C. J. (2006). A GA-based feature selection and parameters optimization for support vector machines. *Expert Systems with Applications*, *31*, 231–240.

Huang, C., Song, K., Kim, S., Townshend, J. R. G., Davis, P., Masek, J. G., et al. (2008). Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sensing of Environment*, *112*, 970–985.

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

Hyde, P., Dubayah, R., Walker, W., Blair, J. B., Hofton, M., & Hunsaker, C. (2006). Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. *Remote Sensing of Environment*, *102*, 63–73.

Im, J., Jensen, J. R., Coleman, M., & Nelson, E. (2009). Hyperspectral remote sensing analysis of short rotation woody crops grown with controlled nutrient and irrigation treatment. *Geocarto International*, *24*, 293–312.

Im, J., Lu, Z., & Jensen, J. R. (2011). A genetic algorithm approach to moving threshold optimization for binary change detection. *Photogrammetric Engineering and Remote Sensing*, *77*, 167–180.

Im, J., Lu, Z., Rhee, J., & Jensen, J. R. (2012a). Fusion of feature selection and optimized immune networks for hyperspectral image classification of urban landscapes. *Geocarto International*, *27*, 373–393.

Im, J., Lu, Z., Rhee, J., & Quackenbush, L. J. (2012b). Impervious surface quantification using a synthesis of artificial immune networks and decision/regression trees from multi-sensor data. *Remote Sensing of Environment*, *117*, 102–113.

Ioki, K., Imanishi, J., Sasaki, T., Morimoto, Y., & Kitada, K. (2010). Estimating stand volume in broad-leaved forest using discrete-return LiDAR: Plot-based approach. *Landscape and Ecological Engineering*, *6*, 29–36.

Jang, J., Payan, V., Viau, A. A., & Devost, A. (2008). The use of airborne lidar for orchard tree inventory. *International Journal of Remote Sensing*, *29*, 1767–1780.

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

Jenkins, J. C., Chojnacky, D. C., Heath, L. S., & Birdsey, R. A. (2004). Comprehensive database of diameter-based biomass regressions for North American tree species. *General technical report NE-319*. Newtown Square, PA: U.S. Department ofAgriculture, Forest Service.

Kato, A., Moskal, L. M., Schiess, P., Swanson, M. E., Calhoun, D., & Stuetzle, W. (2009). Capturing tree crown formation through implicit surface reconstruction using airborne lidar data. *Remote Sensing of Environment*, *113*, 1148–1162.

Kavzoglu, T., & Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, *11*, 352–359.

Ke, Y., Zhang, W., & Quackenbush, L. J. (2010). Active contour and hill climbing for tree crown detection and delineation. *Photogrammetric Engineering and Remote Sensing*, *76*, 1169–1181.

Keerthi, S. S., & Lin, C. J. (2003). Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation*, *15*, 1667–1689.

Korpela, I., Koskinen, M., Vasander, H., Holopainen, M., & Minkkinen, K. (2009). Airborne small-footprint discrete-return LiDAR data in the assessment of boreal mire surface patterns, vegetation, and habitats. *Forest Ecology and Management*, *258*, 1549–1566.

Kuemmerle, T., Chaskovskyy, O., Knorn, J., Radeloff, V. C., Kruhlov, I., Keeton, W. S., et al. (2009). Forest cover change and illegal logging in the Ukrainian Carpathians in the transition period from 1988 to 2007. *Remote Sensing of Environment*, *113*, 1194–1207.

Kwak, D., Lee, W., Cho, H., Lee, S., Son, Y., Kafatos, M., et al. (2010). Estimating stem volume and biomass of *Pinus koraiensis* using LiDAR data. *Journal of Plant Research*, *123*, 421–432.

Li, C., Kao, C. Y., Gore, J. C., & Ding, Z. (2007). Implicit active contours driven by local binary fitting energy. *Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition, 17–22 June, Minneapolis, Minnesota* (pp. 1–7).

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2(3)*. (pp. 18–22) [Available at http://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf.]
Lin, C., Thomson, G., & Lo, C. (2011). A multi-level morphological active contour algorithm for delineating tree crowns in mountainous forest. *Photogrammetric Engineering and Remote Sensing*, *77*, 241–249.
Lu, Z., Im, J., & Hodgson, M. (in review). Building type classification using spatial attributes derived from LiDAR remote sensing data, *ISPRS Journal of Photogrammetry and Remote Sensing*, in review.
Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*, 247–259.
Muinonen, E., Maltamo, M., Hyppänen, H., & Vainikainen, V. (2001). Forest stand characteristics estimation using a most similar neighbor approach and image spatial structure information. *Remote Sensing of Environment*, *78*, 223–228.
Næsset, E. (2004). Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research*, *19*, 164–179.
Næsset, E. (2005). Assessing sensor effects and effects of leaf-off and leaf-on canopy conditions on biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*, *98*, 356–370.
Næsset, E., & Økland, T. (2002). Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment*, *79*, 105–115.
Persson, Å., Holmgren, J., & Söderman, U. (2002). Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing*, *68*, 925–932.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & The R Core Team (2008). *nlme: Linear and nonlinear mixed effects models* (pp. 1–89). [R package version 3].
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: Using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogrammetric Engineering and Remote Sensing*, *70*, 589–604.
Popescu, S. C., Wynne, R. H., & Scrivani, J. A. (2004). Fusion of small-footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, USA. *Forest Science*, *50*, 551–565.
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.
Powell, S. L., Cohen, W. B., Healey, S. P., Kennedy, R. E., Moisen, G. G., Pierce, K. B., et al. (2010). Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment*, *114*, 1053–1068.
Salas, C., Ene, L., Gregoire, T. G., Næsset, E., & Gobakken, T. (2010). Modelling tree diameter from airborne laser scanning derived variables: A comparison of spatial statistical models. *Remote Sensing of Environment*, *114*, 1277–1285.
Shevade, S. K., Keerthi, S. S., Bhattacharyya, C., & Murthy, K. R. K. (2000). Improvements to the SMO algorithm for SVM regression. *IEEE Transactions on Neural Networks*, *11*, 1188–1193.
Solberg, S., Astrup, R., Gobakken, T., Næsset, E., & Weydahl, D. J. (2010). Estimating spruce and pine biomass with interferometric X-band SAR. *Remote Sensing of Environment*, *114*, 2353–2360.
Tseng, M., Chen, S., Hwang, G., & Shen, M. (2008). A genetic algorithm rule-based approach for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *63*, 202–212.
Vauhkonen, J., Korpela, I., Maltamo, M., & Tokola, T. (2010). Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sensing of Environment*, *114*, 1263–1276.
Walker, W. S., Kellndorfer, J. M., LaPoint, E., Hoppus, M., & Westfall, J. (2007). An empirical InSAR-optical fusion approach to mapping vegetation canopy height. *Remote Sensing of Environment*, *109*, 482–499.
Walton, J. T. (2008). Subpixel urban land cover estimation: Comparing cubist, random forests, and support vector regression. *Photogrammetric Engineering and Remote Sensing*, *74*, 1213–1222.
Wang, L., Gong, P., & Biging, G. S. (2004). Individual tree-crown delineation and treetop detection in high-spatial-resolution aerial imagery. *Photogrammetric Engineering and Remote Sensing*, *70*, 351–357.
Yu, D., & Wu, C. (2006). Incorporating remote sensing information in modeling house values: A regression tree approach. *Photogrammetric Engineering and Remote Sensing*, *72*, 129–138.
Yu, X., Hyyppä, J., Vastaranta, M., Holopainen, M., & Viitala, R. (2011). Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*, 28–37.
Zhao, K., Popescu, S., & Nelson, R. (2009). Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment*, *113*, 182–196.