

# EVALUATION OF ALTERNATIVE METHODS FOR USING LIDAR TO PREDICT ABOVEGROUND BIOMASS IN MIXED SPECIES AND STRUCTURALLY COMPLEX FORESTS IN NORTHEASTERN NORTH AMERICA

REI HAYASHI<sup>1</sup>, JOHN A. KERSHAW, JR.<sup>2</sup>, AARON WEISKITTEL<sup>1</sup>

<sup>1</sup>*University of Maine, School of Forest Resources, Orono, ME 04469 USA*

<sup>2</sup>*University of New Brunswick, Faculty of Forestry and Environmental Management, Fredericton, NB Canada E3B 5A3*

**ABSTRACT.** Light detection and ranging (LiDAR) has become a common means for predicting key forest structural attributes, but comparisons of alternative statistical methods and the spatial extent of LiDAR metrics extraction on independent datasets have been minimal. The primary objective of this study was to assess the performance of local and non-local LiDAR aboveground biomass (AGB) prediction models at two locations in the Acadian Forest. Two common statistical techniques, nonlinear mixed effects (NLME) and random forest (RF), were used to fit the prediction models and compared. Finally, this study evaluated the influence of alternative plot radii for LiDAR metrics extraction on model fit and prediction accuracy. AGB models were independently developed at each forest and tested both locally (model applied to same forest used for development) and non-locally (model applied to different forest) using an extensive network of ground-based plots. In general, RF was found to outperform NLME when applied locally, but the differences between the approaches were negligible when applied to the non-local dataset. NLME was found to perform equally well locally and non-locally. LiDAR extraction radius had very little influence on model performance as well. Minimal differences between models developed using fixed- and variable-radius methods were found, while the optimal LiDAR extraction radius was not consistent among forests, statistical technique, or local vs. non-local. Overall, the results highlight the importance of a robust calibration dataset that covers the full range of observed variation for developing accurate prediction models based on remote sensing data.

**Keywords:** LiDAR, random forest, nonlinear mixed effects models, fixed-radius plots, variable-radius plots, Maine, New Brunswick

## 1 INTRODUCTION

For forest inventory attribute predictions, the applicability of light detection and ranging (LiDAR) data has been widely investigated in various forest ecosystems and geographic regions for more than two decades (Hudak et al. 2009). A number of published studies have examined how various factors such as LiDAR sensor specifications (e.g. Ruiz et al. 2014), terrain conditions (e.g. Su and Bork 2007), and stand structures (e.g. Hayashi et al. 2014) influence model development and prediction accuracy. While it is important to investigate these factors, other issues need to be examined

such as statistical modeling techniques, field plot type, and spatial extent of LiDAR covariate extraction.

A number of studies have used parametric modeling techniques, such as multiple linear regression, for developing stem volume and aboveground biomass (AGB) prediction models. For example, Næsset (2004a) used stepwise regression to develop three site and species specific volume prediction models for Norway spruce (*Picea abies* L. Karst.) and Scots pine (*Pinus sylvestris* L.) dominated poor and good site quality mature forests, and birch (*Betula pubescens* Ehrh.) dominated young forests in southeast Norway. These models had a coefficient of determination ( $R^2$ ) value of 0.97, 0.86 and

0.83 for the young forest, the poor and good site quality mature forests, respectively. However, residuals in the young forest and the good site quality mature forest models had significant correlation with the proportion of hardwood species in calibration plots. Likewise, Li et al. (2008) used stepwise regression to develop three forest site specific AGB prediction models for two sites in Washington, USA and a site in Alaska, USA. To effectively parameterize species and structural variations, Chen et al. (2012) developed AGB prediction models using a linear mixed effects (LME) model in a mixed softwood species forest in California, USA. Based on National Agricultural Imagery Program (NAIP), visually classified vegetation types were assigned as random effects in the LME model. This LME model had a  $R^2$  value of 0.83 and improved the root mean square error (RMSE) by 10% when compared with a model without random effects. To account for clones and stands, Packalen et al. (2011) developed volume prediction models using a nonlinear mixed effects (NLME) model in a *Eucalyptus* plantation in Brazil. Clone and stand were assigned as random effects in the NLME model, which had a relative RMSE of 11.8, 8.8, and 7.0% based on only fixed effects, added clone random effects, and added clone and stand random effects, respectively. However, parametric approaches like regression have several limitations including high sensitivity to influential observations, difficulty in handling several highly-correlated variables, and assumed mathematical relationships (e.g. linear or curvilinear). Despite these limitations, parametric methods can extrapolate beyond the range of the fitting dataset.

In contrast, some studies have examined the applicability of nonparametric modeling techniques because they are not generally sensitive to distribution assumptions and collinearity in the data. In addition, determination of optimal number of covariates in a parametric model is a difficult task because there is a need to balance parsimony with prediction accuracy. A common nonparametric technique is random forest (RF) (Breiman 2001) because it can effectively select key covariates and calibrate robust prediction models. For example, Yu et al. (2010) used RF to develop a volume prediction model in Norway spruce and Scots pine dominated boreal forests in Finland. Based on data from 23 independent test plots, this model had a correlation coefficient ( $r$ ) value of 0.79 with a relative RMSE of 20.9%. Using the most similar neighbor (MSN) imputation approach, Kankare et al. (2013) developed an AGB prediction model in Norway spruce and Scots pine dominated boreal forests in Finland. Based on data from 254 independent test plots, this model had an adjusted  $R^2$  value of 0.71 with a relative RMSE of 24.9%. Hayashi et al. (2014) used RF to develop a stem volume predic-

tion model in a multi-layered, mixed hardwood-softwood species forest in central Maine, USA. Based on three selected covariates from a total of 97 potential covariates in LiDAR metrics, RF successfully parameterized a model with a  $R^2$  value of 0.72 with a relative RMSE of 39.7% despite the high species composition and stand structural variability of the forest. However, the effectiveness of nonparametric methods like RF rely on the strength of the underlying training data and the method may not extrapolate well when applied to new locations, but this has not been well tested to date.

While some studies have reported about validity of the nearest neighbor (kNN) approach for various inventory attribute predictions (Bollandsas et al. 2013; Maltamo et al. 2009; Peuhkurinen et al. 2008; Vauhkonen et al. 2010), RF has seemed to show superiority and is generally the recommended nonparametric modeling technique (Hudak et al. 2008). Based on high density LiDAR data, aerial orthoimages, and Landsat Thematic Mapper images, Latifi et al. (2010) examined four different modeling techniques when developing timber volume and AGB prediction models in Scots pine, beech (*Fagus sylvatica* L.), and oak (*Quercus petraea* Liebl. and *Quercus rubra* L.) dominated forests in southwestern Germany. These four techniques were the kNN approaches based on (1) Euclidian distance, (2) Mahalanobis distance, (3) the most similar neighbor (MSN), and (4) RF. The results of this study showed that the RF-based volume and AGB prediction models had the smallest relative RMSE. Using seemingly unrelated regression (SUR), a parametric modeling technique, kNN, and RF, Penner et al. (2013) developed stem volume prediction models in a primarily black spruce (*Picea mariana* Mill. B.S.P.) dominated forest in northeastern Ontario, Canada. While all developed models were validated using independent data, the study reported that the kNN based prediction model had the lowest prediction accuracy, while RF and eight different forest type specific SUR modes had similar prediction accuracy levels. These authors noted that RF would have an advantage over forest type specific SUR models because RF did not require stand stratification data.

In addition to modeling technique issues, the portability of developed prediction models should be examined. Although some studies have evaluated their prediction models using independent test data (e.g. Gobakken et al. 2013; Hawbaker et al. 2009; Maltamo et al. 2009), the test data tend to be collected in areas in close proximity to model calibration plots because of time and budget issues. In contrast, Næsset (2004b) has investigated the portability of developed prediction models in areas apart from where the calibration data were collected. In that study, the calibration plots and independent test plots were approximately 80 km apart,

and the developed models reasonably predicted inventory attributes with respect to stand-level estimates. For aboveground carbon density (ACD) prediction, Asner et al. (2009) reported that a large number of calibration plots were not necessary if the calibration data accounted for a range of forest conditions. Asner et al. (2012b) also found that LiDAR derived mean crown height values were strongly correlated with ACD prediction in four tropical forests in Peru, Panama, Madagascar, and Hawaii; however, regression slopes in each model varied. Consequently, these authors were able to successfully develop a ‘universal’ ACD prediction model accounting for the four different tropical forests using the LiDAR derived mean crown height with two ancillary covariates. Given these findings, LiDAR prediction models appear to be relatively robust when developed using a strong calibration dataset, but this assumption has not been extensively tested in temperate forests of eastern North America, which can be relatively complex due to the number of species present, the range of stand structures, and past silvicultural treatments.

Finally, limited research has been conducted on the influence of the type of ground-based plot and the extraction radius for computing LiDAR metrics. Generally, fixed-radius circular ground plots and a LiDAR metric extraction radii equal to the sample plot radius are suggested (White et al. 2013). Although some research has been done on the effect of fixed-radius plot size on LiDAR calibration (e.g. Frazer et al. 2011; Gobakken and Næsset 2009), less work has been done on alternative sampling methods such as variable-radius plots despite their use in previous LiDAR studies (e.g. Falkowski et al. 2010; Falkowski 2015; Scrinz et al. 2015). In Maine, Hayashi et al. (2014) compared the use of fixed- and variable-radius plots for LiDAR model calibration and found relatively slight differences between model performance ( $R^2$  of 0.82 and 0.76 for fixed- and variable-radius plots, respectively). For ease of LiDAR metrics extraction, Hayashi et al. (2014) recommend the use of fixed-radius plots, but there would be several advantages of using variable-radius plots due to the efficiency of measurement and ability to provide unbiased estimates of forest structural attributes (e.g. Rice et al. 2014). Since co-registration and other georeferencing errors can significantly influence LiDAR model calibration accuracy (e.g. Gobakken and Næsset 2009), the use of multiple LiDAR metric extraction radii could be beneficial, particularly when only low-density LiDAR data and/or variable-radius ground-data are available.

The goal of this analysis was to assess the accuracy and portability of developed average stand-level live AGB (tonnes  $\text{ha}^{-1}$ ) prediction models using different statistical techniques, LiDAR metric extraction radii, and model calibration datasets. The data were from two ex-

perimental forests in the Acadian Forest region, the Noonan Research Forest (NRF) in New Brunswick, Canada, and the Penobscot Experimental Forest (PEF) in Maine, USA. These forests were selected because they cover a range of stand conditions typical for the region, and have a strong network of ground-based plots. The primary objectives of this analysis were to: (1) establish empirical relationships between LiDAR data and AGB using NLME and RF; (2) examine accuracy of the developed prediction models on the local and non-local datasets; and (3) assess various factors (e.g. field plot type, LiDAR metrics extraction radii, forest type) that influence model development and implementation.

## 2 METHODS

**2.1 Study Area** This study was conducted between the Noonan Research Forest (NRF) near Fredericton, New Brunswick, Canada (N 45° 59'12", W 66° 25'15"), and the Penobscot Experimental Forest (PEF), near Orono, Maine, USA (N 44° 49'30", W 68° 39'00") (Figure 1).

The distance between the NRF and the PEF is approximately 220 km. The NRF has been managed by University of New Brunswick since 1985 and is about 1500 ha in size with 271 stands ranging in size from 0.5 to 47 ha. The U.S. Forest Service established the PEF in 1952 for conducting long term research regarding timber management, stand dynamics and biological diversity (Sendak et al. 2003). The total area of the PEF is 1619 ha, and various silvicultural treatments, which are representative of typical northern Maine’s silvicultural practices, have been twice replicated in management units ranging in a size of 0.5 to 23 ha. While both the NRF and the PEF are a part of Acadian Forest (Loo and Ives 2003; Sendak et al. 2003), the NRF is composed of a mixture of relatively pure species, single cohort stands (primarily coniferous species) to mixed species, multi-cohort stands (primarily northern hardwood-softwood species), while the PEF is primarily a mixed northern hardwood-softwood forest.

In both the NRF and the PEF, the major hardwood species are red maple (*Acer rubrum* L.), birch (*Betula* spp.), and aspen (*Populus* spp.), while the major softwood species are spruces (*Picea* spp.), balsam fir (*Abies balsamea* L. (Mill.)), and eastern white pine (*Pinus strobus* L.). The range of elevation above sea level is between 50 and 120 m on the NRF, and 20 and 70 m on the PEF. The terrain at both sites is relatively flat.

### 2.2 NRF AGB Model Calibration Data

NRF model calibration data were collected from a total of 1410 sampling plots established in the summer of 2012. Each calibration plot was established at the inter-

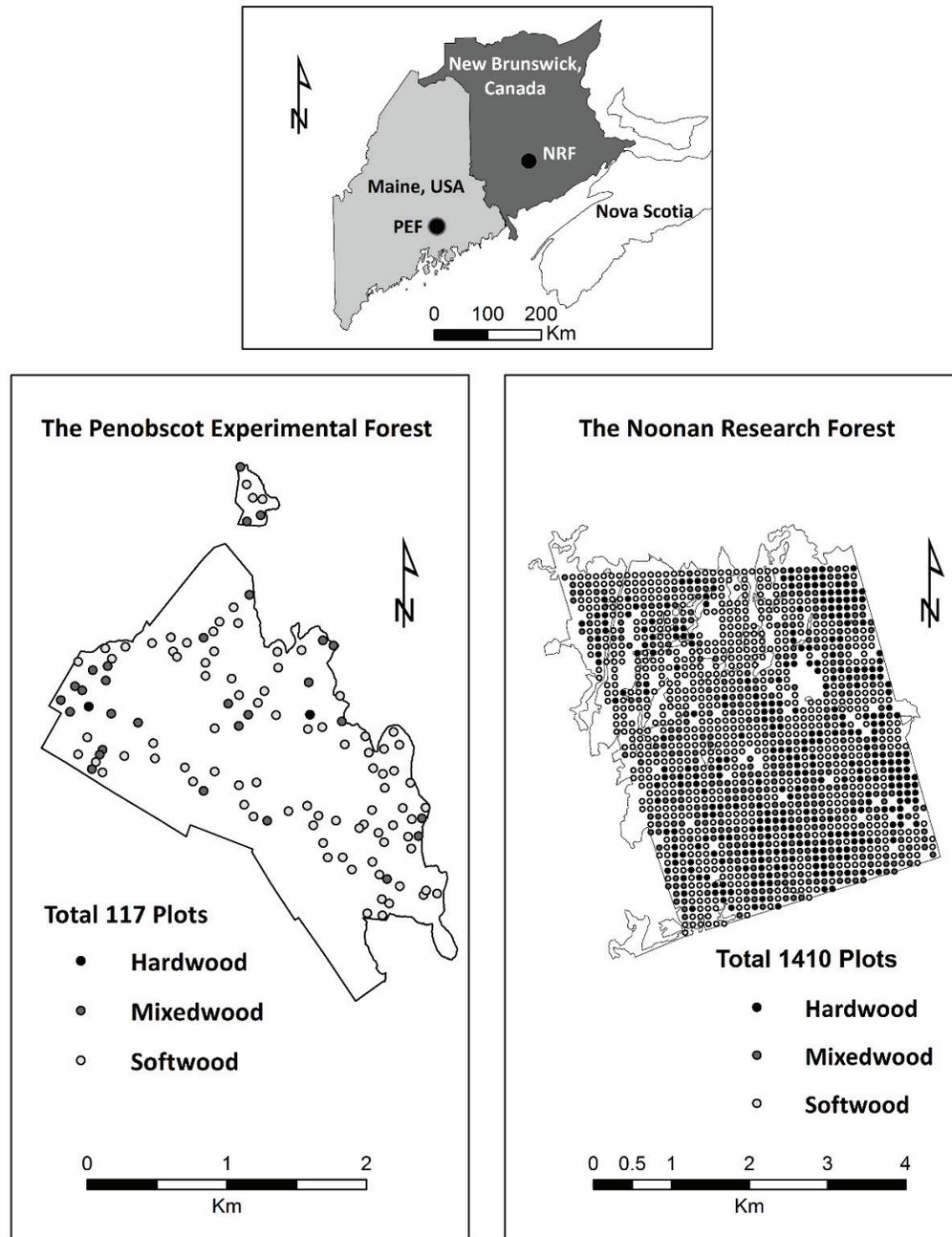


Figure 1: The Noonan Research Forest (NRF) near Fredericton, New Brunswick, Canada ( $N45^{\circ} 59'12''$ ,  $W66^{\circ} 25'12''$ ) and the Penobscot Experimental Forest (PEF) near Orono, Maine, USA ( $N44^{\circ} 49'30''$ ,  $W68^{\circ} 39'00''$ ). The distance between the NRF and the PEF is approximately 220 km.

Table 1: Inventory attributes (mean and standard deviation, in parentheses) by forest type and experimental forest. DBH is diameter at breast height (cm), VOL is stem volume ( $\text{m}^3 \text{ha}^{-1}$ ) and AGB is aboveground biomass (tonnes  $\text{ha}^{-1}$ ).

Forest type	Definition	Plots (#)	DBH		VOL		AGB	
			Mean	(SD)	Mean	(SD)	Mean	(SD)
<b>Noonan Research Forest</b>								
Hardwood	Basal area of hardwood $\geq 70\%$	249	20.19	6.25	141.5	83.9	128.5	61.3
Softwood	Basal area of softwood $\geq 70\%$	682	20.02	7.72	131.1	79.3	93.8	56.6
Mixedwood	Basal area of softwood or hardwood $< 70\%$	479	20.46	5.19	168.8	74.2	133.9	51.7
<b>Overall</b>	-	1410	20.17	6.82	145.8	80.2	111.8	59.5
<b>Penobscot Experimental Forest</b>								
Hardwood	Basal area of hardwood $\geq 70\%$	2	18.37	4.52	25.9	21.8	20.1	16.1
Softwood	Basal area of softwood $\geq 70\%$	29	21.99	9.99	187.3	132.9	108.0	62.3
Mixedwood	Basal area of softwood or hardwood $< 70\%$	86	19.50	7.03	120.4	86.2	76.0	44.7
<b>Overall</b>	-	117	21.26	9.36	167.9	126.2	98.5	60.2

sections of a 100 m  $\times$  100 m grid laid over the study area. Using a 2M basal area factor (BAF;  $2 \text{ m}^2 \text{ha}^{-1} \text{tree}^{-1}$  tallied) angle gauge, species and diameter at breast height (DBH) were recorded for trees greater than 6.0 cm in DBH. AGB for all tallied trees was predicted with a species-specific equation (Jenkins et al. 2003). Using the per ha expansion factor, individual tree AGBs on each calibration plot were aggregated to a per unit area estimate of AGB (tonnes  $\text{ha}^{-1}$ ). Each calibration plot was stratified into one of three forest types based on species basal area (Table 1).

In addition to the primary forest inventory described above, there was a small network of 0.04 ha fixed-area permanent sample plots located on the same 100 m by 100 m grid. On these plots, species DBH, and total height are measured on a 5-year cycle with the last re-measurement occurring in summer 2014. A total of 84 plots were available in this network of plots. These samples will be used to compare field measured and LiDAR predicted AGBs between field plot types.

**2.3 PEF AGB Model Calibration Data** For model calibration data, the PEF's long-term study dataset was used (Brissette et al. 2012). From this dataset, 11 replicated management units (a total of 22 silvicultural treatment units) were selected. Within these 22 management units, a total of 117 permanent fixed area, nested, circular sampling plots were established (3–7 plots per management unit). Species and DBH for trees  $\geq 6.4$  cm (2.5 inches) and  $< 11.2$  cm (4.5 inches) DBH were recorded on 0.02 ha (0.05 acre) plots, and trees  $\geq 11.2$

cm DBH were recorded on 0.08 ha (0.20 acre) plots. The PEF uses a 10-year inventory cycle for the long-term research data collection; therefore, the latest re-measurements were conducted between 2003 and 2009 depending on management unit. The LiDAR data were acquired in 2010, and the Acadian Variant of the Forest Vegetation Simulator (Weiskittel et al. 2012) was used to project DBH to 2010. Based on projected DBH, AGB for all recorded trees were predicted using species-specific equations (Jenkins et al. 2003). With the appropriate per ha expansion factors, individual tree AGBs at each permanent sampling plot were aggregated to plot-level AGB (tonnes  $\text{ha}^{-1}$ ). Each permanent sampling plot was stratified into one of three forest types using the same definitions as the NRF.

**2.4 LiDAR Data** On the NRF, airborne full waveform LiDAR data were acquired under a leaf-off condition on October 21 and 22, 2011, using a Riegl LMS Q680i laser scanner. The mean flying altitude above sea level was about 724 m. This sensor generated the pulse repetition frequency of 180 KHz, and the laser wavelength was 1550 nm with a scan angle of  $< 28.5^\circ$  from the nadir. This scan angle was relatively wider than scan angles suggested in certain LiDAR data acquisition guidelines such as the US Geological Survey National Geospatial Program LiDAR Guideline (Heide-mann 2012). While narrower scan angles would lead to improve accuracy in the creation of digital elevation models, the wider scan angle in this study likely better sensed vegetation structures under dominant and

codominant tree crown positions. In general, LiDAR height metrics like the ones used in this analysis are relatively insensitive to scan angles up to 30° (e.g. Holmgren et al. 2003). The mean pulse density was 3 pulses m<sup>-2</sup> per swath (this was flown at 50% overlap providing a final density of 6 pulses m<sup>-2</sup>) with a footprint of 0.35 m, and the sensor collected up to 8 returns per pulse.

On the PEF, airborne discrete-return LiDAR data were acquired under a leaf-off condition on November 10, 2010, using an Optech Gemini 246 sensor. This LiDAR data were also used and described in detail in Hayashi et al. (2014). Although the data were intended to be acquired under a leaf-off condition, most hardwood trees kept leaves in the PEF due to an abnormal prolonged summer period in 2010. The mean flying altitude above sea level was about 1982 m. The sensor generated the pulse repetition frequency of 5000 KHz, and the laser pulse intensity was 1064 nm with a scan angle of < 20° from the nadir. The mean pulse density was 1.1 pulses m<sup>-2</sup> with a footprint of 30 cm, and the sensor collected up to 4 pulse returns.

Both LiDAR data acquired in the NRF and the PEF were processed using a custom algorithm in R v3.02 (R Development Core Team 2013). Based on 10, 15, 20, 25 and 30 m radius circles, this program extracted LiDAR metrics at matched locations between calibration plots in the field and LiDAR data. While some studies have examined how different sizes of calibration plots in the field influence model development (e.g. Asner and Mascaro 2014; Gobakken and Næsset 2008; Ruiz et al. 2014), this study examined how the extraction area for the LiDAR metrics influenced model fits. Since the calibration data in the NRF were collected based on a variable-radius sampling scheme, they did not have a specific plot area like fixed radius plots (average inclusion zone radii, based on quadratic mean DBHs, were between 10 – 30 m). Similarly, calibration data in the PEF were collected based on the 0.02 ha (7.98 m radius) and 0.08 ha (15.96 m radius) nested circular plots. For both forests, pulse returns below 2 m from the ground were disregarded based on preliminary findings. The algorithm computed a total of 15 potential covariates (Table 2) for developing AGB prediction models. ABG was used for consistency with other similar studies in the remote sensing literature (preliminary analysis indicated that AGB prediction models gave similar results as ones used to estimate total standing volume).

**2.5 LiDAR Metrics** The LiDAR data were read into the R statistical package from LAS-format files. Since this study was focused on the structural attributes, only the X-Y-Z coordinates of the LAS data were used. GPS coordinates of field sample locations were used to extract LiDAR data in 10, 15, 20, 25, and 30 m radii about

the sample location. For each location, two types of LiDAR metrics were extracted: 1) canopy surface metrics; and 2) LiDAR point cloud vertical distribution metrics (Table 2). Several LiDAR density metrics were tested in preliminary analyses, but none were selected as important predictors of AGB. Five canopy surface metrics were calculated using the canopy surface model: maximum canopy height; mean canopy height; and the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> height percentiles of the canopy surface model. Nine point cloud vertical distribution metrics were extracted: mean height of the point cloud and the 25<sup>th</sup> – 95<sup>th</sup> quantiles in 10 percentile increments. The final metric extracted was the height corresponding to maximum LiDAR density. Preliminary analysis indicated that metrics related to the variability of height (skewness, standard deviation) were unimportant and not included in this analysis.

## 2.6 Fixed versus Variable Radius Field Data

The 84 grid intersections at NRF, where both fixed-radius PSPs and variable-radius Inventory plots were available were used to test for effects of field data type on LiDAR AGB predictions. The NLME models described below were fitted separately to the subset of 84 sample locations using field estimates of AGB based on fixed-radius PSPs and variable-radius inventory plots. Field estimated AGBs and the resulting LiDAR predicted AGBs were compared between the fixed-radius PSPs and variable-radius inventory plots using graphical methods and equivalence tests (Robinson and Froese 2004).

## 2.7 Parametric Modeling Technique

NLME was used to develop average stand-level AGB (tonnes ha<sup>-1</sup>) prediction models based on the 1410 and 117 calibration plots collected in the NRF and PEF, respectively. For selecting covariates for the NLME model, boosted regression trees (De’Ath 2007) were used. A major advantage of boosted regression trees is that relative importance between selected covariates is easily determined. Also, unlike indirect methods such as Akaike’s Information Criterion (AIC), boosted regression trees are robust to nonlinearity and multicollinearity within and between covariates. Using boosted regression in the “caret” package (Kuhn 2008) available in R v3.02, two LiDAR covariates were selected for developing AGB prediction models for the NRF and PEF. The models were of the general form:

$$AGB = \beta_0 X_1^{\beta_1} X_2^{\beta_2} \quad (1)$$

Where AGB is aboveground biomass (tonnes ha<sup>-1</sup>); X<sub>1</sub> and X<sub>2</sub> are covariates selected in the boosted regression step described above; β<sub>j</sub> are parameter estimates for the fixed effects. Random effects, using forest type as a nest-

Table 2: Mean, standard deviation (StDev), and range {Min, Max} of potential LiDAR covariates (m).

No.	Potential Covariates	Abbreviation	Mean	StDev	Min.	Max.
<b>Noonan Research Forest</b>						
1	Maximum canopy surface height.	maxCHT	18.99	4.82	0.4	42.4
2	Mean canopy surface height.	meanCHT	9.68	4.34	0.1	4.34
3	25th percentile of canopy surface height.	q25CHT	7.3	4.7	0.0	21.6
4	50th percentile of canopy surface height.	q50CHT	9.98	4.95	0.0	22.9
5	75th percentile of canopy surface height.	q75CHT	12.29	4.93	0.0	24.9
6	Mean point cloud height.	meanHT	7.25	2.71	0.3	16.3
7	Height of 25th percentile of point clouds.	q25LiDAR	4.05	1.99	0.2	12.3
8	Height of 35th percentile of point clouds.	q35LiDAR	5.31	2.42	0.2	14.9
9	Height of 45th percentile of point clouds.	q45LiDAR	6.53	2.83	0.2	16.6
10	Height of 55th percentile of point clouds.	q55LiDAR	7.75	3.24	0.3	18.1
11	Height of 65th percentile of point clouds.	q65LiDAR	8.97	3.61	0.3	19.6
12	Height of 75th percentile of point clouds.	q75LiDAR	10.23	3.9	0.3	21.5
13	Height of 85th percentile of point clouds.	q85LiDAR	11.68	4.14	0.3	23.3
14	Height of 95th percentile of point clouds.	q95LiDAR	13.83	4.34	0.3	26.9
15	Height of maximum point cloud density	HTmaxDens	5.79	4.93	0.3	21.8
<b>Penobscot Experimental Forest</b>						
1	Maximum canopy surface height.	maxCHT	18.85	4.79	8	35.1
2	Mean canopy surface height.	meanCHT	6.85	2.72	1.2	19.6
3	25th percentile of canopy surface height.	q25CHT	3.12	2.77	0.0	16.3
4	50th percentile of canopy surface height.	q50CHT	6.6	3.37	0.1	21.7
5	75th percentile of canopy surface height.	q75CHT	10.25	3.72	1.6	24.3
6	Mean point cloud height.	meanHT	7.37	2.41	2.1	19.4
7	Height of 25th percentile of point clouds.	q25LiDAR	4.13	1.97	0.6	15.3
8	Height of 35th percentile of point clouds.	q35LiDAR	5.29	2.31	1.2	18.9
9	Height of 45th percentile of point clouds.	q45LiDAR	6.44	2.6	1.6	20.5
10	Height of 55th percentile of point clouds.	q55LiDAR	7.65	2.86	1.8	21.6
11	Height of 65th percentile of point clouds.	q65LiDAR	8.95	3.13	2.1	22.2
12	Height of 75th percentile of point clouds.	q75LiDAR	10.35	3.42	2.6	23.9
13	Height of 85th percentile of point clouds.	q85LiDAR	12.01	3.69	3.3	25.8
14	Height of 95th percentile of point clouds.	q95LiDAR	14.45	4.05	4.9	28.7
15	Height of maximum point cloud density	HTmaxDens	5.25	4.25	0.3	22.8

ing factor, were added to each variable and maximum likelihood ratio tests used to assess significance.

The models for the NRF and PEF will be designated as NLME<sub>NRF</sub> and NLME<sub>PEF</sub>, respectively. The notation, NLME<sub>XXX/YYY</sub> will be used to refer to “fitting data/test data” (for example NLME<sub>NRF/PEF</sub> refers to a model developed using NRF calibration data, and validated using PEF test data). In preliminary analysis, we examined the use of additional covariates (up to 5), but  $R^2$  values in these NLME models were marginally improved compared to the NLME models with two covariates, and parameter estimates were often not significantly different from 0. The “nlme” package (Pinheiro et al. 2014) available in R v3.02 was used to fit Equa-

tion 1 to the NRF and PEF data. Five different AGB models were developed corresponding to the 5 different LiDAR extraction radii (10, 15, 20, 25, and 30 m) for each forest.

**2.8 Nonparametric Modeling Technique** RF was used to develop stand-level average estimates of AGB (tonnes ha<sup>-1</sup>) prediction models based on the calibration plot data collected in the NRF and PEF. As its name infers, RF grows a forest based on a number of constructed regression trees. At the beginning, RF randomly separates a given dataset to two groups, one for training and the other for testing. In general, 1/3 of the given data are used for testing. Second, to construct re-

gression trees, RF uses bootstrapped datasets from the training data with replacement. To split at each node, RF estimates mean square error (MSE) using the test data, while trying randomly chosen potential covariates without replacement, and RF disregards covariates with larger MSE. In this study, 3,000 trees were set to grow in RF for RF<sub>NRF</sub> and RF<sub>PEF</sub> models. While the default setting in RF utilizes 1/3 of available covariates (Liaw and Wiener 2002), this study utilized only two covariates and stand type because RF models with additional covariates only marginally improved  $R^2$  values in preliminary analysis. Hereafter, the developed RF models based on the NRF and PEF calibration data with four covariates were called the RF<sub>NRF</sub> model and the RF<sub>PEF</sub> model, respectively. As with NLME, RF<sub>XXX/YYY</sub> refers to “fitting data/test data”. For these models, the “randomForest” package (Liaw and Wiener 2002) available in R v3.02 was used. Five different AGB models were developed corresponding to the 5 different LiDAR extraction radii (10, 15, 20, 25, and 30 m) for each forest.

**2.9 Model Evaluation** To evaluate model performance,  $R^2$ , root mean square error (RMSE), mean bias (MB; observed – predicted), and mean absolute error (|MB|; |observed - predicted|) between field-measured and model predicted AGB were calculated and compared. Negative and positive values in MB indicate overestimation and underestimation by models, respectively. Also, model fits were graphically examined using a one to one plot between field-measured and model predicted AGB with a lowess smoothed line. In addition, similarity between field-measured and model predicted AGB was assessed using equivalence tests (Robinson and Froese 2004), which have a null hypothesis of dissimilarity. Finally, factors influencing model performance were evaluated using model residuals and boosted regression. Factors tested included: model source (NRF vs PEF); model origin (local versus nonlocal); forest type, fitting technique, extraction radius, forest type, and observed AGB.

### 3 RESULTS

**3.1 AGB Distributions by Forest Type** Both NRF and PEF were stratified into the three forest types where the stratification was conducted simply at a per-ha-level based on species basal area (Table 1). Total AGBs across all forest types were slightly greater in the NRF than the PEF, while standard deviations were similar (Table 1). At the forest type level, AGB in the hardwood forest type in the NRF was notably greater than the PEF, though sample sizes were quite small on the PEF. AGB in the NRF was distributed approximately symmetrical, while the PEF was skewed to the right (data not shown).

In the PEF, a limited number of plots had AGBs greater than 200 tonnes ha<sup>-1</sup>.

**3.2 Model Development** Boosted regression trees consistently identified q45LiDAR (covariate #9, Table 2) as an influential covariate at all extraction radii for both forests, except for the 15 m LiDAR extraction radius at PEF. Other covariates identified included q85LiDAR, maxCHT, and q50CHT (see Table 2 for variable definitions). While the second most important covariate varied between forest and LiDAR extraction radius, the pair of variables, q45LiDAR and q50CHT, had the best overall performance of any pair of covariates and were, therefore, used to fit Equation 1 to the observed AGB data from both NRF and PEF at all LiDAR extraction radii. Random effects for forest type were only significant for the intercept term ( $\beta_0$ ). The  $R^2$  for the resulting NLME models ranged from 0.61 to 0.68 at NRF and from 0.40 to 0.48 at PEF (Table 3). RMSEs ranged from 33.4 – 37.0 tonnes ha<sup>-1</sup> at NRF and from 40.6 to 45.0 tonnes ha<sup>-1</sup> at PEF (Table 3). At NRF, the models based on the 10 m LiDAR extraction radius performed best, while the models based on the 20 m LiDAR extraction radius performed best at PEF (Table 3).

Similarly, in preliminary analyses, RF selected q45LiDAR as a predictive covariate in all RF<sub>NRF</sub> models, while q85LiDAR was selected as a predictive covariate in all RF<sub>PEF</sub> models for all LiDAR extraction radii. Again, the secondary variables selected by RF varied by forest and extraction radius, but were similar to those identified by boosted regression, with q45LiDAR and q50CHT being present in most RF models on both forests. To simplify the model development process and to facilitate model comparisons between forests across LiDAR extraction radii, and fitting technique, the two covariates, q45LiDAR and q50CHT (Table 2) were again selected for use in the RF models along with forest type. The resulting RF models had a range of  $R^2$ s of 0.82 – 0.85 on NRF and 0.80 – 0.84 on PEF (Table 3). RMSEs ranged from 22.8 – 25.5 tonnes ha<sup>-1</sup> on NRF and from 23.8 – 26.7 tonnes ha<sup>-1</sup> on PEF (Table 3). For both forests, the RF models performed superior to the NLME models. As with the NLME models, the RF model based on 10 m extraction radius performed best on NRF, while the 10 m extraction radius was the poorest fit on PEF, and all other extraction radii models performed about the same. It should be noted that differences were small for both modeling types at both forests across all extraction radii, and, based on equivalence tests, all predictions were essentially identical to field measurements (Table 4).

Overall prediction bias in AGB was negligible, between -0.31 and -0.14 tonnes ha<sup>-1</sup> for the NLME models on NRF and between -0.50 and 0.33 tonnes ha<sup>-1</sup> (nega-

Table 3: Goodness of fit statistics by fitting and testing data source, LiDAR extraction radius, and modeling approach. NRF is the Noonan Research Forest and PEF is the Penobscot Experimental Forest, while NLME is nonlinear mixed effects models and RF is random forest models. Negative values in mean bias (MB) indicate overestimation by models. RMSE is root mean square error and  $|MB|$  is the absolute mean bias. The units are tonnes ha<sup>-1</sup>.

Data Source		Fit Statistic	NLME					Random Forest				
Model	Test		10	15	20	25	30	10	15	20	25	30
NRF	NRF	R <sup>2</sup>	0.68	0.67	0.64	0.63	0.61	0.85	0.84	0.83	0.82	0.82
		RMSE	33.4	34.4	35.5	36.4	37.0	22.8	23.5	24.3	24.9	25.5
		MB	-0.3	-0.3	-0.2	-0.2	-0.1	-0.11	-0.07	-0.05	-0.01	-0.1
		MB	25.9	26.8	27.8	28.5	29.1	17.7	18.2	19.1	19.6	20.0
NRF	PEF	R <sup>2</sup>	0.4	0.48	0.48	0.47	0.46	0.35	0.37	0.39	0.39	0.39
		RMSE	46.3	43.2	43.1	43.6	44.1	48.3	47.5	46.7	46.6	46.7
		MB	3.56	5.36	7.12	6.93	7.84	14.4	14.51	15.03	14.52	14.36
		MB	36.2	32.7	32.4	32.9	32.9	36.0	35.4	35.4	35.2	35.2
PEF	PEF	R <sup>2</sup>	0.43	0.53	0.54	0.54	0.53	0.8	0.84	0.83	0.84	0.84
		RMSE	45.4	41	40.6	40.8	41.1	26.7	24.3	24.7	23.8	24.3
		MB	0.33	-0.2	-0.5	-0.5	-0.6	-0.07	0.16	0.08	0.15	0.34
		MB	35.5	32.2	31.8	31.8	31.7	18.7	16.6	17.3	16.7	17.1
PEF	NRF	R <sup>2</sup>	0.64	0.58	0.59	0.58	0.58	0.62	0.57	0.53	0.51	0.49
		RMSE	35.8	38.7	38.1	38.5	38.5	36.6	38.9	40.7	41.8	42.5
		MB	5.42	11.7	8.2	7.42	7.42	-5.39	-11.9	-14.1	-15.4	-15.5
		MB	27.8	29.7	29.4	29.8	29.8	28.9	31.4	32.7	33.5	34.0

tive values indicated overestimation) for the PEF. Biases associated with the RF models were smaller in magnitude (1/3 or smaller; Table 4). While overall prediction bias was small, it was not constant across the range of observed AGB (Figs. 2A and 2B for NRF and Figs. 3C and 3D for PEF). For both forests and both modeling approaches, biases became increasingly positive (under-prediction) as observed AGB increased.

### 3.3 Fixed versus Variable Radius Field Plots

The NLME model described above was fitted to the subset of 84 grid points at NRF where both variable-radius inventory plots and fixed-area PSPs were present using a 10 m LiDAR extraction radius. Figure 4 shows AGB from variable radius inventory plots vs AGB from fixed-area PSPs for both observed field values and LiDAR predicted values. On average, the AGB estimates from the fixed-area PSPs were slightly larger than the values obtained from variable-radius inventory plots and this difference increased near the upper end of the range of AGB (failed to reject null hypothesis of dissimilarity at  $p = .05$  and region of similarity = .25; Fig. 4). The LiDAR predicted values showed similar results with less bias near the upper end of the AGB range. As with observed AGB, the predicted AGBs from the two mod-

els were dissimilar (null hypothesis of dissimilarity not rejected at  $p=.05$ ); however, equivalence tests showed similarity between predictions from the NLME based on variable-radius plots and AGB observed on the fixed-area plots and similarity between predictions from the NLME based of fixed-area plots and AGB observed on the variable-radius plots.

### 3.4 Model Evaluation

Although both model fitting procedures performed well locally, the RF model had weaker agreement with the independent test data (Table 3). NLME performance, when applied to the other forest, was similar to the performance of the NLME model fitted locally. R<sup>2</sup>, RMSE, and MB were all nearly equivalent for both forests when comparing local versus non-local model statistics (Table 3). RF performance was lower when applied non-local versus local (Table 3); however, RF performance was nearly on par with the NLME performance when applied non-locally. At the 10 m extraction radius, non-local models performed similar to the local models (Table 4), but all other extraction radii were dissimilar. Even though the fit statistics associated with the RF models were substantially better than those associated with NLME, the predictions for both models were equivalent at all extraction radii.

Table 4: Equivalence test results for mean differences between aboveground biomass (tonnes ha<sup>-1</sup>) at the 95% significance level. The null hypothesis of an equivalence test is that the two values are dissimilar (i.e., rejection of the test indicates similar values). The mean differences, standard deviations (in parentheses), and magnitudes (% standard deviation) of the region of similarity (in square brackets) are also given. Abbreviations are nonlinear mixed effects (NLME), random forest (RF), Noonan Research Forest (N), and Penobscot Experimental Forest (P)

Equivalence Test		Statistic	LiDAR Extraction Radius				
			10	15	20	25	30
Field	NLME <sub>N/N+P/P</sub>	H <sub>0</sub>	Reject	Reject	Reject	Reject	Reject
		Mean Diff.	-0.26	-0.28	-0.25	-0.21	-0.18
		Std Dev	34.5	34.95	35.92	36.74	37.37
		Region of Sim	15%	10%	10%	10%	10%
	NLME <sub>N/P+P/N</sub>	H <sub>0</sub>	Reject	Reject	Reject	Reject	Reject
		Mean Diff.	5.27	11.21	8.12	7.38	6.45
		Std Dev	36.34	37.41	37.66	38.23	38.72
		Region of Sim	25%	40%	30%	30%	25%
	RF <sub>N/N+P/P</sub>	H <sub>0</sub>	Reject	Reject	Reject	Reject	Reject
		Mean Diff.	-0.10	-0.05	-0.04	0.01	-0.07
		Std Dev	23.09	23.57	24.36	24.83	23.58
		Region of Sim	15%	15%	15%	15%	15%
	RF <sub>N/P+P/N</sub>	H <sub>0</sub>	Reject	Reject	Reject	Reject	Reject
		Mean Diff.	-3.87	-9.86	-11.89	-13.01	-13.24
		Std Dev	37.41	38.43	39.44	40.07	40.78
		Region of Sim	20%	35%	40%	45%	45%
NLME <sub>N/N+P/P</sub>	NLME <sub>N/P+P/N</sub>	H <sub>0</sub>	Reject	Not rej.	Not rej.	Not rej.	Not rej.
		Mean Diff.	5.54	11.06	8.37	7.59	6.62
		Std Dev	11.06	11.06	10.91	10.08	9.66
		Region of Sim	50%	≤50%	≤50%	≤50%	≤50%
	RF <sub>N/N+P/P</sub>	H <sub>0</sub>	Reject	Reject	Reject	Reject	Reject
		Mean Diff.	0.16	0.22	0.21	0.21	0.11
		Std Dev	14.96	14.84	14.87	15.18	15.41
		Region of Sim	25%	25%	25%	20%	20%
	RF <sub>N/P+P/N</sub>	H <sub>0</sub>	Reject	Not rej.	Not rej.	Not rej.	Not rej.
		Mean Diff.	-3.60	-9.59	-11.63	-12.89	-13.07
		Std Dev	13.34	17.99	17.99	17.92	18.15
		Region of Sim	40%	≤50%	≤50%	≤50%	≤50%
NLME <sub>N/P+P/N</sub>	RF <sub>N/N+P/P</sub>	H <sub>0</sub>	Reject	Not rej.	Not rej.	Not rej.	Not rej.
		Mean Diff.	-5.38	-11.27	-8.16	-7.38	-6.51
		Std Dev	17.94	19.31	18.24	18.21	18.14
		Region of Sim	45%	≤50%	≤50%	≤50%	≤50%
	RF <sub>N/P+P/N</sub>	H <sub>0</sub>	Not rej.	Not rej.	Not rej.	Not rej.	Not rej.
		Mean Diff.	-9.14	-21.07	-20.00	-20.48	-19.69
		Std Dev	18.04	21.42	20.78	20.22	19.45
		Region of Sim	≤50%	≤50%	≤50%	≤50%	≤50%
	RF <sub>N/N+P/P</sub>	H <sub>0</sub>	Reject	Not rej.	Not rej.	Not rej.	Not rej.
		Mean Diff.	-3.77	-9.81	-11.85	-13.11	-13.17
		Std Dev	19.45	19.84	20.51	20.59	21.24
		Region of Sim	≤50%	≤50%	≤50%	≤50%	≤50%

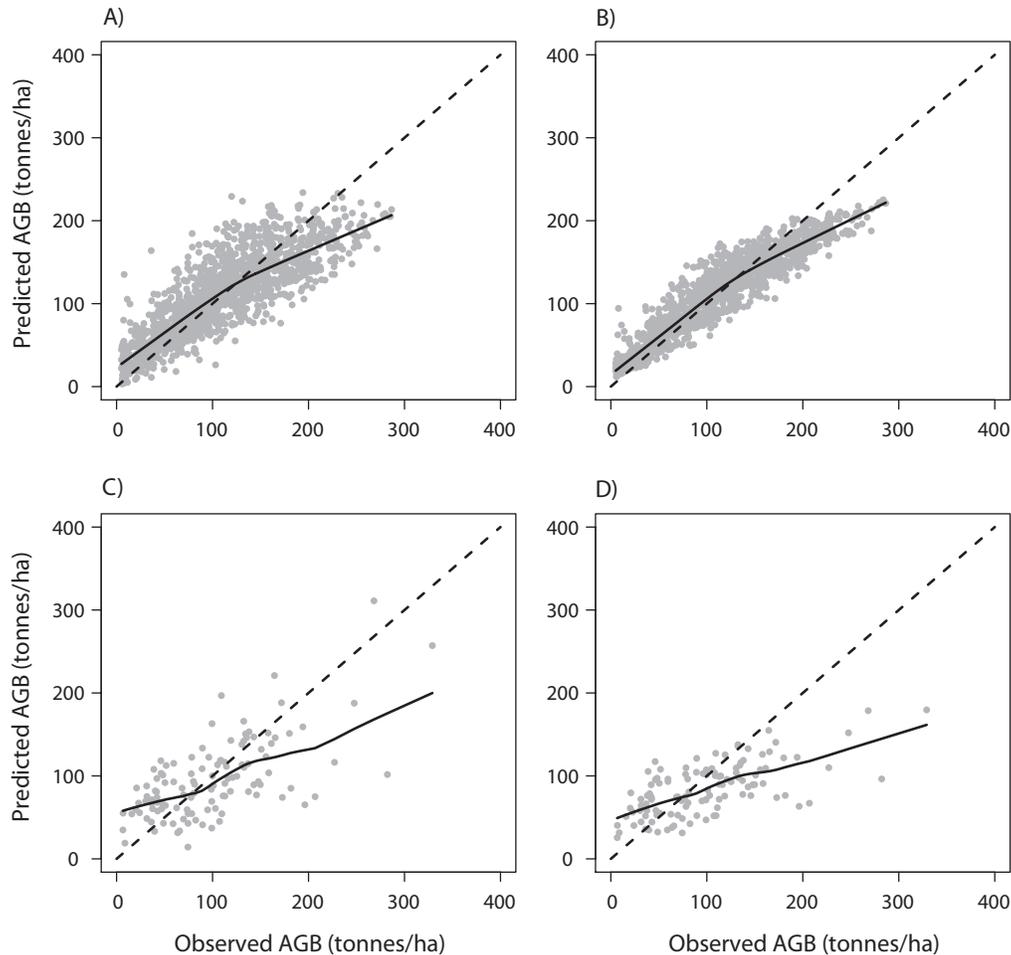


Figure 2: Observed (field-measured) above ground biomass (AGB; tonnes  $\text{ha}^{-1}$ ) versus predicted AGB: A) Noonan Research Forest (NRF) observations versus predictions from nonlinear mixed effects (NLME) model fitted to NRF data; B) NRF observations versus predictions from random forest (RF) model fitted to NRF data; C) Penobscot Experimental Forest (PEF) observations versus predictions from NLME model fitted to NRF data; and D) PEF observations versus predictions from RF model fitted to NRF data. All NRF models based on 10 m extraction radius.

The only factor investigated that influenced model residuals was field observed AGB (Figure 5). Model locale (local versus non-local) had minor influence as did forest type, but extraction radii and location (PEF vs. NRF) had virtually no influence on model residuals.

#### 4 DISCUSSION

This analysis highlights numerous issues associated with properly calibrating a robust AGB prediction model using LiDAR and ground-based plot data. The accuracy of AGB prediction models did not strongly depend on the statistical method, LiDAR metric extraction radius, and type of ground-based data used (Figure 5), but was driven more by the forest type and the application of a non-locally calibrated model. Although there were

key differences between the two study sites used in this analysis including ground-based plot sampling method (fixed- vs. variable-radius), LiDAR point density, and acquisition date that might influence findings, the limited influence of model source in the generalized boosted regression would suggest these factors are of minimal importance (Figure 5). In addition, these study site differences in this analysis were likely reflective of real-world conditions as rarely do research and operational applications perfectly align. At one extreme, this analysis might represent the worst case scenario in terms of model accuracy and bias, but is more likely a reflection of general operating conditions in reality.

Although our study raises many more question than it answers, there is a general need to better evaluate

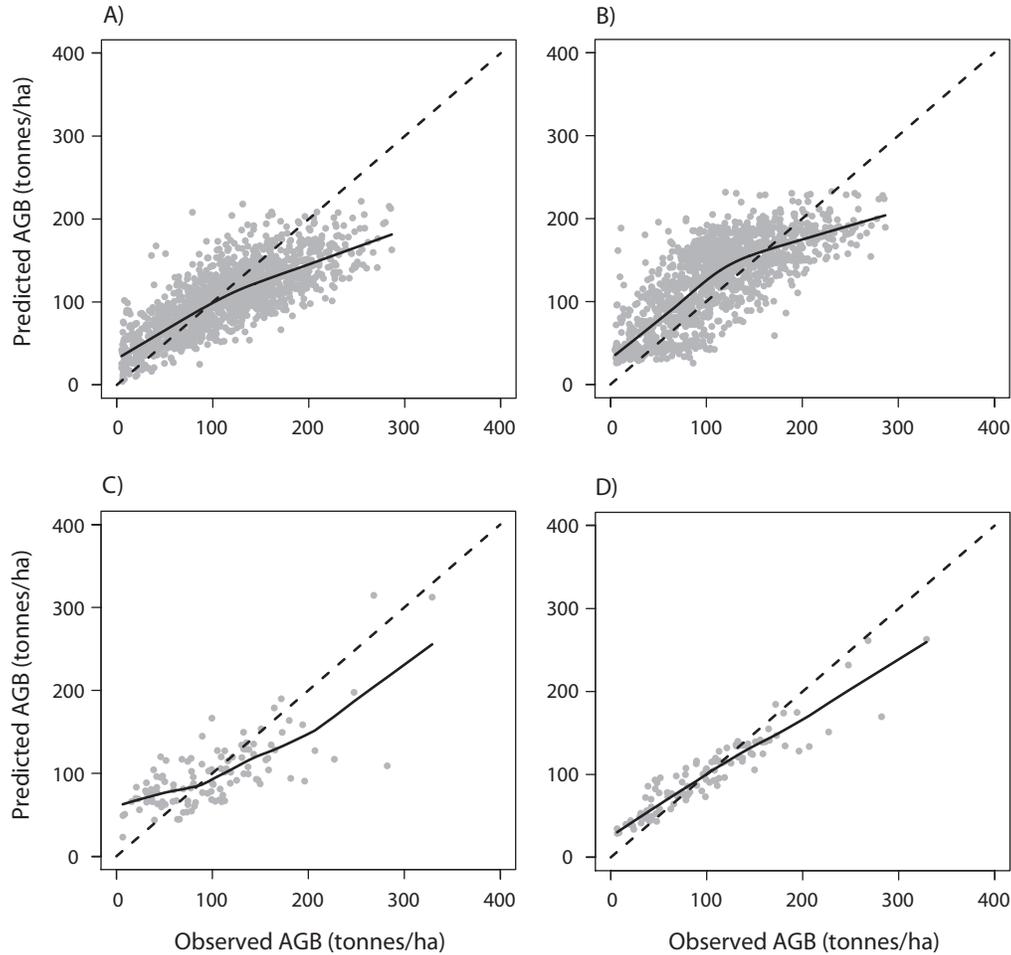


Figure 3: Observed (field-measured) above ground biomass (AGB; tonnes  $\text{ha}^{-1}$ ) versus predicted AGB: A) Noonan Research Forest (NRF) observations versus predictions from nonlinear mixed effects (NLME) model fitted to Penobscot Experimental Forest (PEF) data; B) NRF observations versus predictions from random forest (RF) model fitted to PEF data; C) PEF observations versus predictions from NLME model fitted to PEF data; and D) PEF observations versus predictions from RF model fitted to PEF data. All PEF models based on 20 m extraction radius.

the performance of remote sensing technology on independent datasets as done in this analysis. It is important to note that AGB was used in this analysis as it is consistent with previous remote sensing studies (Chen et al. 2012; Li et al. 2008). AGB is computed based only on DBH and species group, which effectively ignores tree height information. However, we found similar model performance using total stem volume in our analysis and consequently believe that the selection of dependent variable has limited implications for our findings. This might not be the case in regions where stem volume and total AGB are as highly correlated as they are in the Acadian Region ( $r = 0.88$ ).

Overall, using a strong network of ground-based measurement plots, the calibrated prediction models showed a high correlation between field-measured and model

predicted AGB (Table 5) as previously noted in numerous similar studies (Chen et al. 2012; Li et al. 2008). However, to properly use the parametric techniques for model calibration, the data need to be carefully examined as most LiDAR metric-derived covariates tend to be highly correlated and some covariates do not meet the normal distribution criterion (Hayashi et al. 2014; Hudak et al. 2008; Stone et al. 2011), though most parametric techniques are relatively robust for departures in the assumption, especially with large datasets. Relationships between AGB and LiDAR covariates also might not show a linear trend (Packalen et al. 2011). To account for these issues, some previous studies have transformed a dependent and/or independent variables (Li et al. 2008; Means et al. 2000). One drawback to this approach is that predicted AGB needs to be back trans-

formed, which can result in model bias. In this study, we chose a two-covariate NLME for calibrating a parsimonious AGB prediction model to avoid the potential of transformation bias and multicollinearity. In this study, although boosted regression trees tended to select highly correlated covariates, the fitted NLME models generally had better model fits and prediction accuracy when the models were used to predict AGB in another forest. On the other hand, nonparametric techniques, such as RF, generally show robustness against correlation and normality issues.

However, for either modeling technique, other issues remain, particularly robust covariate selection criteria and determination of an optimal number of covariates (Vauhkonen et al. 2010). Regardless of correlation and normality issues, there is a need to balance model prediction accuracy with parsimony to avoid possible model overfit issues. In preliminary analysis, we found that NLME models were marginally improved when using 3, 4, or even 5 covariates. However, NLME might not converge when additional covariates were added in the model. In this study, a random effect was applied to only one coefficient (Equation 1). Additional random effects could have been used; however, convergence would likely again be an issue. In addition, while boosted regression trees tended to select q45LiDAR as a key covariate for

both  $NLME_{NRF}$  and  $NLME_{PEF}$  models, the additional covariates tended to vary every time boosted regression trees were applied to the different forests and LiDAR plot radii. Similarly, the RF process tended to select either q85LiDAR or q45LiDAR as the primary variable with various other metrics selected as secondary.

Interestingly, q45LiDAR tended to be the primary covariate selected irrespective of the forest, LiDAR metric extraction radius, and statistical method. This highlights the robustness of height percentiles when compared to absolute and relative point density measures, even across a wide range of stand structures as used in this analysis. Previous analyses have shown height metrics to be less sensitive to plot size (e.g. Frazer et al. 2011), georeferencing errors (e.g. Gobakken and Næsset 2009), and laser scan angle (e.g. Holmgren and Lindberg 2013), which is further confirmed by this analysis given the differences in study sites and available LiDAR data. When combined with the maximum plot canopy height data, q45LiDAR generally corresponded to a relative canopy height of 0.35–0.40 (data not shown). This value is relatively consistent with the recommended relative height of 0.3–0.4 for importance sampling to determine individual tree volume (see Figure 1 in Wood et al. 1990). This value is considered important because it is related to the centroid of tree volume (Wood et al. 1990), which would suggest that q45LiDAR would have some biological basis for application and should be further examined across a range of forest types.

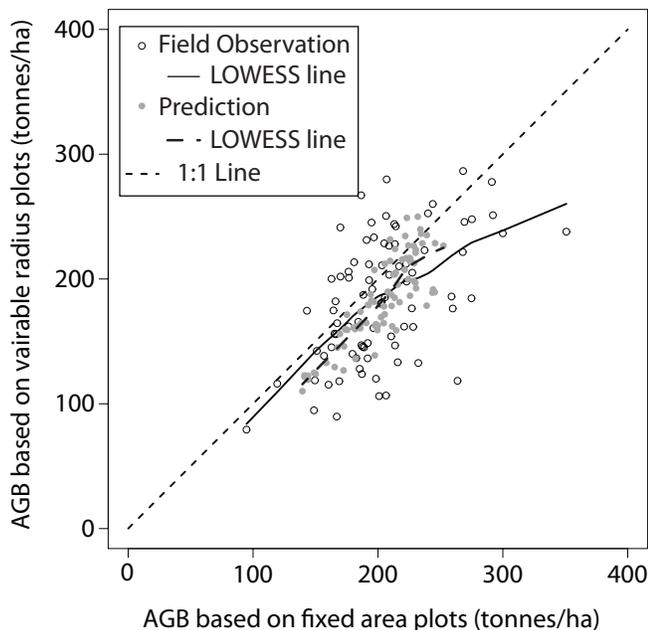


Figure 4: Observed and predicted above ground biomass (AGB;  $\text{tonnes ha}^{-1}$ ) on fixed-area versus variable-radius field plots from Noonan Research Forest. Predictions based on nonlinear mixed effects models and a 10 m extraction radius.

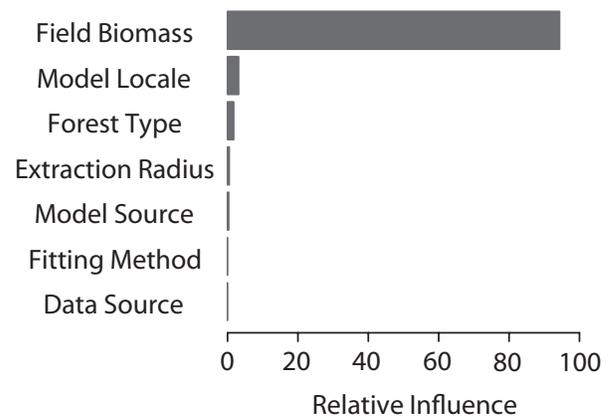


Figure 5: Relative influence by factor based on boosted regressions on model residuals. Factors included observed field biomass, model locale (local vs. non-local), forest type (hardwood, softwood, mixedwood), LiDAR extraction radius (10, 15, 20, 25, 30 m), model source, fitting method (nonlinear mixed effects vs. random forest), and data source (Noonan Research Forest vs. Penobscot Experimental Forest).

Both NLME and RF model performances varied by the plot radius used for the LiDAR metrics extraction. In general, a model based on 10 m LiDAR plot radius had the best model fit among  $NLME_{NRF}$  and  $RF_{NRF}$  models. This is likely because of the high pulse density LiDAR data collected in the NRF. In contrast, the low pulse density LiDAR collected at the PEF would need to use a larger area to obtain comparable LiDAR metrics extraction. For example, a  $NLME_{PEF}$  model based on a 25 m LiDAR plot radius had the best model fit among  $NLME_{PEF}$  models. Also, the  $RF_{NRF-ALL}$  tended to select larger LiDAR plot radii, while  $RF_{PEF-ALL}$  tended to select smaller LiDAR plot radii. However, a primary point from this analysis was that prediction models with higher model fits may not necessarily predict at greater accuracy levels when applied to new locations. This would be another reason that it should be beneficial to utilize as many potential covariates as a modeling technique could handle.

To our knowledge, there has been limited work on optimal LiDAR plot size, but Magnussen and Boudewyn (1998) and Gobakken et al. (2013) did suggest that the plot size between field and LiDAR be same. Although this suggestion is reasonable if field plots are based on fixed-radius plots, the  $NLME_{PEF/PEF}$  with a 20 m LiDAR plot radius had the best  $R^2$  and RMSE, while the radius for the 0.08 ha circular plots was only appropriately 16 m. All extraction radii, other than 10 m, performed equally well. Frazer et al. (2011) reported that smaller field plots tended to be greatly affected by plot positioning errors during calibration data collection, while Dalponte et al. (2011) found that positioning errors up to 5 m led to marginal differences in  $R^2$  values and RMSEs in developed models. Ruiz et al. (2014) indicated that, while a combination of larger field plots and higher pulse density data resulted in greater model fits, field plot sizes generally had greater influence than pulse density on model fits. On the other hand, field plots based on variable-radius sampling like NRF would not have a comparable size for LiDAR metrics extraction, unless truncated angle count samples as suggested by Maltamo et al. (2007), were used. Although the use of variable-radius plots may complicate the model calibration, they had a limited influence on model accuracy as the  $R^2$  values were generally consistent between the NRF and PEF and did not change drastically across the range of LiDAR metric extraction radii evaluated at the NRF. In fact, when variable- and fixed-radius plots were compared at the same forest and LiDAR acquisition, minimal differences were found. As discussed above, much more important factors other than sampling method are influencing the differences between model performances at the two study sites.

Although the use of fixed-radius plots has been much more common for LiDAR model calibration (e.g. Chen et al. 2012; Hayashi et al. 2014; Li et al. 2008), we believe variable-radius are a viable alternative given the strong performance of the NRF model in this analysis, which is consistent with the findings of another recent analysis (e.g. Scrinzi et al. 2015). In general, fixed- and variable-radius both provide unbiased estimates of forest structure and composition, but primarily differ in the time necessary to complete field sampling. For example, Rice et al. (2014) reported that fixed- and variable-radius methods provided similar estimates of plot basal area and volume in diverse stand conditions in northern Maine despite drastic differences in plot measurement time. Previous LiDAR analyses have also used multiple sampling methods for LiDAR calibration and evaluation with no reported confounding influence of sampling method. For example, Falkowski et al. (2010) collected calibration data using fixed-radius plots and the developed models were validated using an independent test data collected using variable-radius sampling. Consequently, we believe the differences in sampling methods used at the NRF and PEF have a limited influence on our study's findings, which is further supported by the relative influence values presented in Figure 5.

## 5 CONCLUSIONS

For future work, evaluating the relationship between field and LiDAR plot sizes to improve model accuracy needs to be investigated. This study showed that LiDAR metrics extraction from a low pulse density data appeared to require larger extraction radii to capture forest structure attributes, while a high density LiDAR data seemed to require a smaller LiDAR plot radius. Also, although calibration data was collected using either fixed- and variable-radius methods, these differences did not seem to be confounded for model development in this study and only minimal differences were observed when both types of plot data were available at the same forest. Further investigation for usability of variable-radius sampling is necessary, but it appears quite viable based on the results of this analysis. For example, Falkowski (2015) suggests that LiDAR metrics be weighted proportionally to return height when using variable-radius sampling.

More importantly, this study investigated different statistical techniques on AGB prediction accuracy at two study sites. In general, both NLME and RF produced well-behaved models that relied on a minimal number of covariates (2) combined with forest type. While the various methods had similar model fit statistics, there were drastic differences between model performances when applied to an independent dataset. For model parame-

terization, RF was generally found superior to NLME. However, for application purposes, NLME models generally showed prediction performances on par with their fitting performances, while RF performance was substantially degraded relative to fit performance. In addition, RF models can never produce values outside the range of values observed in the training dataset, and this greatly reduced the performance of RF<sub>PEF</sub> models on NRF. Irrespective of the statistical method used, a strong network of ground-based plots that cover the full range of observed variation are necessary and although stratification is likely an efficient method for achieving this, forest type was not an effective predictor in this analysis, possibly due to limited numbers of plots in some forest types. Although not suggested by Figure 5, an important limitation of this analysis is that additional factors not accounted for in the assessment may be causing these observed differences between the two locations.

Finally, q45LiDAR was consistently selected as the top covariate across LiDAR metric extraction radii, forest types, and statistical method, which indicates it is a robust and important prediction. The variable does have a biological basis and should be further examined across a greater range of forest types.

## 6 ACKNOWLEDGEMENTS

Thanks to the US Forest Service for providing access to the long-term network of permanent plots. Leading Edge Geomatics, Fredericton, NB provided LiDAR data for Noonan Research Forest. Funding for this work was provided by the School of Forest Resources at University of Maine, the Cooperative Forestry Research Unit, and the NB Innovation Fund Research Assistants Initiative. Margaret Penner and an anonymous reviewer provided helpful feedback on a previous version of this manuscript.

This work is also based upon research supported, in part, by McIntire-Stennis Grant (ME041516) from the USDA National Institute of Food and Agriculture. This is Scientific Contribution no. 3447 of the Maine Agricultural and Forest Experimentation Station.

## REFERENCES

- Asner, G.P., Clark, J.K., Mascaro, J., Galindo Garca, G.A., Chadwick, K.D., Navarrete Encinales, D.A., Paez-Acosta, et al. 2012. High-resolution mapping of forest carbon stocks in the Colombian Amazon. *Biogeosciences* 9(7), 2683–2696.
- Asner, G.P., Hughes, R.F., Varga, T.A., Knapp, D.E., Kennedy-Bowdoin, T., 2009. Environmental and biotic controls over aboveground biomass throughout a tropical rain forest. *Ecosystems* 12, 261–278.
- Asner, G.P., Mascaro, J., 2014. Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. *Remote Sensing of Environment* 140, 614–624.
- Asner, G.P., Mascaro, J., Muller-Landau, H.C., Vieilledent, G., Vaudry, R., Rasamoelina, M., Hall, J.S., Breugel, M.v., 2012. A universal airborne LiDAR approach for tropical forest carbon mapping. *Oecologia* 168, 1147–1160.
- Bollandsås, O.M., Maltamo, M., Gobakken, T., Næsset, E., 2013. Comparing parametric and non-parametric modelling of diameter distributions on independent data using airborne laser scanning in a boreal conifer forest. *Forestry* 86, 493–501.
- Breiman, L., 2001. Random forest. *Machine Learning* 45, 5–32.
- Brissette, J., Kenefic, L., Russell, M., Publick, J., 2012. Overstory tree and regeneration data from the “Silvicultural Effects on Composition, Structure, and Growth” study at Penobscot Experimental Forest. Newtown Square, PA: USDA Forest Service, Northern Research Station.
- Chen, Q., Laurin, G. V, Battles, J.J., Saah, D., 2012. Integration of airborne lidar and vegetation types derived from aerial photography for mapping aboveground live biomass. *Remote Sensing of Environment* 121, 108–117.
- Dalponte, M., Martinez, C., Rodeghiero, M., Gianelle, D., 2011. The role of ground reference data collection in the prediction of stem volume with LiDAR data in mountain areas. *ISPRS Journal of Photogrammetry and Remote Sensing* 66, 787–797.
- De’ath, G., 2007. Boosted trees for ecological modeling and prediction. *Ecology* 88, 243–251.
- Falkowski, M.J. 2015. Increasing the efficiency of LiDAR based forest inventories: A novel approach for integrating variable radius inventory plots with LiDAR Data. Abstract 70590. American Geophysical Union Fall Meeting. 14–18 December 2015. San Francisco, CA.
- Falkowski, M.J., Hudak, A.T., Crookston, N.L., Gessler, P.E., Uebler, E.H., Smith, A.M.S., 2010. Landscape-scale parameterization of a tree-level forest growth model: a k-nearest neighbor imputation approach incorporating LiDAR data. *Canadian Journal of Forest Research* 40, 184–199.

- Frazer, G.W., Magnussen, S., Wulder, M.A., Niemann, K.O., 2011. Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. *Remote Sensing of Environment* 115, 636–649.
- Gobakken, T., Korhonen, L., Næsset, E., 2013. Laser-assisted selection of field plots for an area-based forest inventory. *Silva Fennica* 47, 1–20.
- Gobakken, T., Næsset, E., 2008. Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research* 38, 1095–1109.
- Gobakken, T., Næsset, E., 2009. Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research* 39, 1036–1052.
- Hawbaker, T.J., Keuler, N.S., Lesak, A.A., Gobakken, T., Conruci, 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, 1–11.
- Hayashi, R., Weiskittel, A., Sader, S., 2014. Assessing the feasibility of low-density LiDAR for stand inventory attribute predictions in complex and managed forests of northern Maine, USA. *Forests* 5, 363–383.
- Heidemann, H.K., 2012. Lidar base specification version 1.0: U.S. Geological survey techniques and methods, in: *Book 11, Collection and Delineation of Spatial Data*. p. 63.
- Holmgren, J., Lindberg, E., 2013. Tree crown segmentation based on a geometric tree crown model for prediction of forest variables. *Canadian Journal of Remote Sensing* 39, S86–S98.
- Holmgren, J., Nilsson, M., Olsson, H., 2003. Simulating the effects of lidar scanning angle for estimation of mean tree height and canopy closure. *Canadian Journal of Remote Sensing* 29, 623–632.
- 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.
- Hudak, A.T., Evans, J.S., Smith, A.M.S., 2009. LiDAR utility for natural resource managers. *Remote Sensing* 1, 934–951.
- 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.
- Kankare, V., Vastaranta, M., Holopainen, M., Rätty, M., Yu, X., Hyyppä, J., Hyyppä, H., Alho, P., Viitala, R., 2013. Retrieval of forest aboveground biomass and stem volume with airborne scanning LiDAR. *Remote Sensing* 5, 2257–2274.
- Kuhn, M., 2008. Building predictive models in R using the caret package. *Journal of Statistical Software* 28, 1–26.
- Latifi, H., Nothdurft, A., Koch, B., 2010. Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. *Forestry* 83, 395–407.
- Li, Y.Z., Andersen, H.E., McGaughey, R., 2008. A comparison of statistical methods for estimating forest biomass from light detection and ranging data. *Western Journal of Applied Forestry* 23, 223–231.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R News* 2.
- Loo, J., Ives, N., 2003. The Acadian forest: historical condition and human impacts. *Forestry Chronicle* 79, 462–474.
- Magnussen, S., Boudewyn, P., 1998. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research* 28, 1016–1031.
- Maltamo, M., Korhonen, K.T., Packalén, P., Mehtätalo, L., Suvanto, A., 2007. Testing the usability of truncated angle count sample plots as ground truth in airborne laser scanning-based forest inventories. *Forestry* 80, 73–81.
- Maltamo, M., Packalén, P., Suvanto, A., Korhonen, K.T., Mehtätalo, L., Hyvönen, P., 2009. Combining ALS and NFI training data for forest management planning: a case study in Kuortane, Western Finland. *European Journal of Forest Research* 128, 305–317.
- 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, 1367–1371.
- Næsset, E., 2004a. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research* 19,

- 164–179. Næsset, E., 2004b. Accuracy of forest inventory using airborne laser scanning: evaluating the first Nordic full-scale operational project. *Scandinavian Journal of Forest Research* 19, 554–557.
- Packalen, P., Mehtätalo, L., Maltamo, M., 2011. ALS-based estimation of plot volume and site index in a eucalyptus plantation with a nonlinear mixed-effect model that accounts for the clone effect. *Annals of Forest Science* 68, 1085–1092.
- Penner, M., Pitt, D.G., Woods, M.E., 2013. Parametric vs. nonparametric LiDAR models for operational forest inventory in boreal Ontario. *Canadian Journal of Remote Sensing* 39, 426–443.
- Peuhkurinen, J., Maltamo, M., Malinen, J., 2008. Estimating species-specific diameter distributions and saw log recoveries of boreal forests from airborne laser scanning data and aerial photographs: a distribution-based approach. *Silva Fennica* 42, 625–641.
- Pinheiro, J., Bates, D., Debroy, S., Sarkar, D., 2014. nlme: Linear and nonlinear mixed effects models. R Package Version 3.1–117.
- R Development Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.
- Rice, B., Weiskittel, A.R., Wagner, R.G., 2014. Efficiency of alternative forest inventory methods in partially harvested stands. *European Journal of Forest Research* 133, 261–272.
- Robinson, A.P., Froese, R.E., 2004. Model validation using equivalence tests. *Ecological Modelling* 176, 349–358.
- Ruiz, L.A., Hermosilla, T., Mauro, F., Godino, M., 2014. Analysis of the influence of plot size and LiDAR density on forest structure attribute estimates. *Forests* 5, 936–951.
- Scrini, G., Clementel, F., Floris, A., 2015. Angle count sampling reliability as ground truth for area-based LiDAR applications in forest inventories. *Canadian Journal of Forest Research* 45: 506–514.
- Sendak, P.E., Brissette, J.C., Frank, R.M., 2003. Silviculture affects composition, growth, and yield in mixed northern conifers: 40-year results from the Penobscot Experimental Forest. *Canadian Journal of Forest Research* 33, 2116–2128.
- Stone, C., Penman, T., Turner, R., 2011. Determining an optimal model for processing lidar data at the plot level: results for a *Pinus radiata* plantation in New South Wales, Australia. *New Zealand Journal of Forestry Science* 41, 191–205.
- Su, J.G., Bork, E.W., 2007. Characterization of diverse plant communities in Aspen Parkland rangeland using LiDAR data. *Applied Vegetation Science* 10, 407–416.
- 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.
- Weiskittel, A., Russell, M., Wagner, R., Seymour, R., 2012. Refinement of the Forest Vegetation Simulator Northeast variant growth and yield model: Phase III. In B. Roth (ed.), Cooperative Forest Research Unit. Orono, ME: University of Maine, School of Forest Resources. 96–104.
- White, J.C., Wulder, M.A., Varhola, A., Vastaranta, M., Coops, N., Cook, B.D., Pitt, D., Woods, M., 2013. A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach. Information Report FI-X-010. Canadian Forest Service, Canadian Wood Fibre Centre. Victoria, British Columbia.
- Wood, G.B., Wiant, H.V., Loy, R.J., Miles, J.A., 1990. Centroid sampling: A variant of importance sampling for estimating the volume of sample trees of radiata pine. *Forest Ecology and Management* 36, 233–243.
- Yu, X., Hyypä, J., Holopainen, M., Vastaranta, M., 2010. Comparison of area-based and individual tree-based methods for predicting plot-level forest attributes. *Remote Sensing* 2, 1481–1495.